Aralık '25

BİLGİSAYAR MÜHENDİSLİĞİ

Alanında Uluslararası Derleme, Araştırma ve Çalışmalar

EDİTÖR
PROF. DR. SELAHATTİN BARDAK

SERÜVEN YAYINEVİ

Genel Yayın Yönetmeni / Editor in Chief • C. Cansın Selin Temana

Kapak & İç Tasarım / Cover & Interior Design • Serüven Yayınevi

Birinci Basım / First Edition • © Aralık 2025

ISBN • 978-625-8671-42-1

Serüven Yayınevi / Serüven Publishing
Türkiye Adres / Turkey Address: Kızılay Mah. Fevzi Çakmak 1. Sokak
Ümit Apt No: 22/A Çankaya/ANKARA
Telefon / Phone: 05437675765
web: www.seruvenyayinevi.com
e-mail: seruvenyayinevi@gmail.com

Baskı & Cilt / Printing & Volume
Sertifika / Certificate No: 47083

BİLGİSAYAR MÜHENDİSLİĞİ ALANINDA ULUSLARARASI DERLEME, ARAŞTIRMA VE ÇALIŞMALAR

- ARALIK 2025 -

EDİTÖR **PROF. DR. SELAHATTİN BARDAK**

İÇİNDEKİLER

Bölüm 1

GÖRÜNTÜ VERİ SETLERİ İÇİN VERİ ARTIRMA YÖNTEMLERİNİN İNCELENMESİ: ÖRNEKLER ÜZERİNDE KIYASLAMALI BİR ÇALIŞMA

Ahmet ÇELİK[1]
Bahadır ÇOKÇETİN[2]

1 Kütahya Dumlupınar Üniversitesi, Tavşanlı Meslek Yüksek Okulu, Bilgisayar Teknolojileri Bölümü, Kütahya, Türkiye, ORCID: 0000-0002-6288-3182
2 Kütahya Dumlupınar Üniversitesi, Tavşanlı Meslek Yüksek Okulu, Bilgisayar Teknolojileri Bölümü, Kütahya, Türkiye, ORCID: 0000-0003-2189-9265

1. Giriş

Bilgisayarla görme (computer vision) alanında, öğrenme modellerinin başarısı, büyük veri kümelerinin (veri setlerinin) içerisindeki kayıtlarla doğrudan ilişkilidir. Modellerin eğitim veri miktarı ve veri seti kayıtlarının özellikleri arttıkça tasarlanan modellerin tahmin hatası da azalmaktadır ve başarı oranı artmaktadır. Ancak, yeterli sayıda etiketli görüntü verisine ulaşmak çoğu zaman zor olmakta ve maliyeti de artırmaktadır. Bu sorunun üstesinden gelmek için kullanılan en etkili yöntemlerden biri veri artırma (data augmentation) yöntemleridir. Veri artırma, mevcut görüntülerin çeşitli dönüşümlerle yeniden üretilerek eğitim veri kümesinin yapay olarak genişletilmesini sağlamaktadır. Gelişmiş görüntü veri artırma yöntemleri olarak Mixup (Karıştır), CutMix(Kes-Karıştır) ve Cutout(Kes-Çıkar) yaygın olarak kullanılmaktadır.

Yüksek kaliteli ve dengeli dağılımlara sahip yeterli örneklerin toplanması ve etiketlenmesi hala zahmetli ve pahalı bir iştir ve bu nedenle eğitim veri setini zenginleştirmek için çeşitli veri artırma teknikleri yaygın olarak kullanılmaktadır (Wang et al.,2020) Literatürde, veri artırma yöntemlerinin, başarılı sonuçlar ortaya çıkardığını kanıtlayan birçok araştırmalar yapılmıştır. Bu araştırmalar, sağlık, güvenlik, otomotiv ve çevresel ortamların görüntüleri üzerinde uygulanmış ve başarılı sonuçlar ortaya çıkarılmıştır.

Li et al. (2025) yaptıkları çalışmada, GAN (Generative Adversarial Network) tabanlı veri artırma yöntemlerinin, az örnekli yüz tanıma problemlerinde performans başarısını büyük oranda artırdığı tespit edilmiştir. Böylece, veri artırma, yüz tanıma sistemlerinde sınırlı veri setlerini zenginleştirmek için kritik bir rol oynadığı ortaya konulmuştur.

Wang et al. (2020) ise yaptıkları çalışmada, derin öğrenme yöntemleri modeli üzerinde yüz görüntüsü verisini artırma üzerine kapsamlı bir inceleme yaparak, dönüşüm tabanlı yöntemlerden GAN tabanlı yaklaşımlara kadar geniş bir değerlendirme çalışması yapmışlardır.

Ayrıca, Zhang & Liu, (2024) yaptıkları çalışmada yakın tarihli yaptıkları çalışmada veri artırma ve transfer öğrenme yöntemlerinin küçük veri setlerinde yüz tanıma doğruluğunu artırdığı ortaya koymuşlardır.

Veri artırma yöntemleri, tıbbi görüntüleme alanında derin öğrenme modellerinin performansını artırmak içinde kullanılmaktadır. Liu et al. (2025) yaptıkları çalışmada, tıbbi görüntü sınıflandırmasında veri artırmanın modelinin performansa etkisini incelenmişlerdir. Çalışmada, sınırlı sayıda yüksek kaliteli medikal verinin bulunduğu durumlarda veri artırma tekniklerinin, tasarlanan modelin genelleme yeteneğini geliştirdiği ve daha güvenilir sonuçlar oluşturduğu tespit edilmiştir. Bu sonuçlar, öğrenme algoritma modelleriyle gerçekleştirilen uygulamaların daha güvenilir hale getirilmesi için veri artırmanın kullanılması, daha başarılı sonuçlar oluşturabileceğini göstermektedir.

Otonom sistemlerde veri artırma yöntemleri, özellikle derin öğrenme tabanlı nesne tespiti ve sürüş senaryolarında modelin genelleme yeteneğini güçlendirmek içinde kullanılmaktadır. Bozkaya et al. (2021) yaptıkları çalışmada, veri artırma teknikleri kullanılarak gerçek zamanlı nesne tespit başarı performansının arttığı ve tasarlanan modelin farklı çevresel koşullara daha uyumlu hale getirildiği gösterilmiştir. Çalışmada, sınırlı veri setleriyle eğitilen modellerin daha geniş varyasyonlara karşı dayanıklı olduğu tespit edilmiştir.

Azak et al. (2024) yaptıkları çalışmada, karşılaştırmalı bir incelemede, farklı veri artırma tekniklerinin otonom sürüşte genelleme yeteneğini artırdığı ve modelin daha güvenilir sonuçlar ürettiği ortaya koymuşlardır. Bu çalışmalar, veri artırma yöntemlerinin otonom sistemlerde hem güvenlik hem de doğruluk açısından başarı sonuçları olduğunu göstermektedir.

Çevre teknolojisi araştırmalarında veri artırma yöntemleri, özellikle uzaktan algılama ve yeryüzü gözlemleri alanlarda sınırlı veri setlerinin zenginleştirilmesi için kullanılmaktadır. Sousa et al. (2025) çalışmasında, Earth Observation (EO) görüntülerinde veri artırmak için difüzyon tabanlı bir model önermişlerdir. Çalışmada, kullanılan bu yaklaşım ile çevresel izleme sistemlerinde daha yüksek doğruluk elde edilmiş ve ekosistem değişimlerinin daha güvenilir şekilde modellenmesi mümkün olmuştur. Çalışma, çevre teknolojisi uygulamalarında veri artırmanın yalnızca model performansını artırmakla kalmayıp, aynı zamanda sürdürülebilirlik politikaları içinde daha sağlam karar destek sistemleri geliştirilmesine katkı sağlayacağı gösterilmiştir.

Ayrıca, son yıllarda siber güvenlik alanında kullanılan yapay öğrenme yöntemlerin başarısı kanıtlanmıştır. Ancak tasarlanan modellerin başarısı eğitim verilerinin çeşitliliği ve dengeli dağılımına bağlı olarak artmaktadır. Özellikle saldırı tespit sistemlerinde veri artırma (data augmentation) teknikleri, sınıf dengesizliğini azaltarak modellerin daha doğru sonuçlar üretmesini sağlamaktadır.

Medvedieva et al. (2024) çalışmasında, siber saldırı tespitinde kullanılan denetimli öğrenme algoritmalarının performansını artırmak için veri artırma yöntemlerini kullanmışlardır. Çalışmada, bu sayede yanlış pozitif oranlarının önemli ölçüde düştüğü gösterilmiştir. Bu yaklaşım, veri artırma yöntemlerinin kullanılmasının, güvenlik alanında yapay öğrenme modellerinin daha güvenilir hale gelmesine katkı sağlayabileceğini göstermektedir.

Literatürdeki bu çalışmalar, veri artırmanın, yapay öğrenme modellerinin başarısını doğrudan etkileyen kritik bir basamak olduğunu göstermektedir. Veri artırma yöntemleriyle, görüntü verilerinin çeşitliliği artırılarak, veri seti dengesi sağlamakla birlikte modelin genelleme yeteneğini geliştirilmektedir. Bu nedenle, veri artırma yöntemleri dikkatle seçilmeli ve uygulama alanına göre optimize edilmelidir.

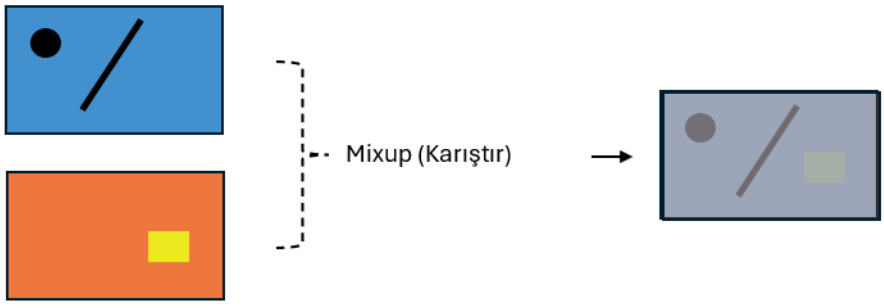
Yapay öğrenme modellerinde, veri setlerinin örnek sayısını artırmak ve dengeli veri oluşturmak için gelişmiş veri artırma teknikleri kullanılmaktadır. Bu çalışmada, görüntü verileri için kullanılan başlıca veri artırma yöntemleri ayrıntılı olarak ele alınmıştır. Mixup (Karıştır) yöntemi ile bilinen görüntü değerleri karıştırılarak, yeni görüntü verileri oluşturulmaktadır. Cutout (Kes-çıkar) yöntemi ile mevcut görüntüler (iki görüntü) üzerinde rastgele alanlar silinerek, yeni görüntüler oluşturulmaktadır. CutMix (Kes-karıştır) yöntemi mevcut görüntülerin (iki görüntü) farklı bölgelerinin yer değiştirmesi ile yeni görüntü oluşturulmaktadır.

Mixup (Karıştır) Veri Artırma Yöntemi

Özellikle, görüntü sınıflandırması görevlerinde Mixup veri artırma yöntemi kullanılmaktadır. Mixup kullanılarak iki görüntünün ve etiketlerinin doğrusal kombinasyonunu alarak, yeni bir örnek görüntü oluşturulmaktadır (Şekil 1). Bu yöntem, modelin karar sınırlarını yumuşatmasına yardımcı olmaktadır. Mixup yöntemiyle, küçük veri setleri sorunun çözümü yapılarak eğitim süreci daha kararlı biçime dönüştürülmüş ve öğrenme modellerinin etkinliği kanıtlanmıştır (Yan et al., 2020).

Mixup ile geliştirilmiş görüntülerin ImageNet gibi veri setlerindeki derin öğrenme modellerinin genelleştirme hatasını iyileştirebileceğini kanıtlanmıştır (Tokozume et al., 2018). Ayrıca, bu yöntem ile öğrenme modelinin bozuk etiketler için bellek gereksinimi azaltılmıştır.

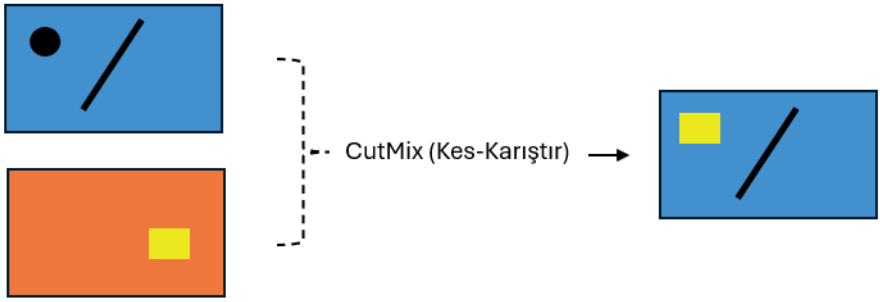
Bu yöntem kullanılarak, modelin tahmin yeteneğini eğitim verisi sınırlarının ötesinde artırılmıştır. Bu yöntem kullanıldığında, bir performans iyileştirmesine rağmen, Mixup yöntemi henüz önyargı–varyans dengesi açısından üzerinde çalışılması gerekmektedir. Mixup, diğer denetimli, denetimsiz, yarı denetimli ve güçlendirici öğrenme türleri için çok fazla kullanım alanı vardır (Chapelle et al. 2020).



Şekil 1. Mixup yönteminin akış şeması

CutMix (Kes-Karıştır) Veri Artırma Yöntemi

CutMix, bir görüntünün belirli bir bölgesinin başka bir görüntüyle değiştirilmesi sonucu, görüntü üzerinde hem görüntü hem de etiket bilgilerini değiştirilmektedir (Şekil 2). CutMix yöntemi (Yun et al. 2019), basit bir piksel silme/yer değiştirme yöntemi değildir. CutMix yönteminde, eğitim verisi üzerinde verimli olan bilgi ve bölgeyi korunmaktadır. Ayrıca bu yöntem, yeni blok verisi eklenerek modelin gerekli hale gelmesini sağlayarak, yerelleştirmeyi daha da artırmaktadır. CutMix yöntemi, modelin doğruluğunu önemli ölçüde artırarak öğrenme yöntemlerinin dayanıklılığını artırmakta ve aşırı güven sorununu azaltmaktadır (Hao et al. 2023).



Şekil 2. CutMix yönteminin akış şeması

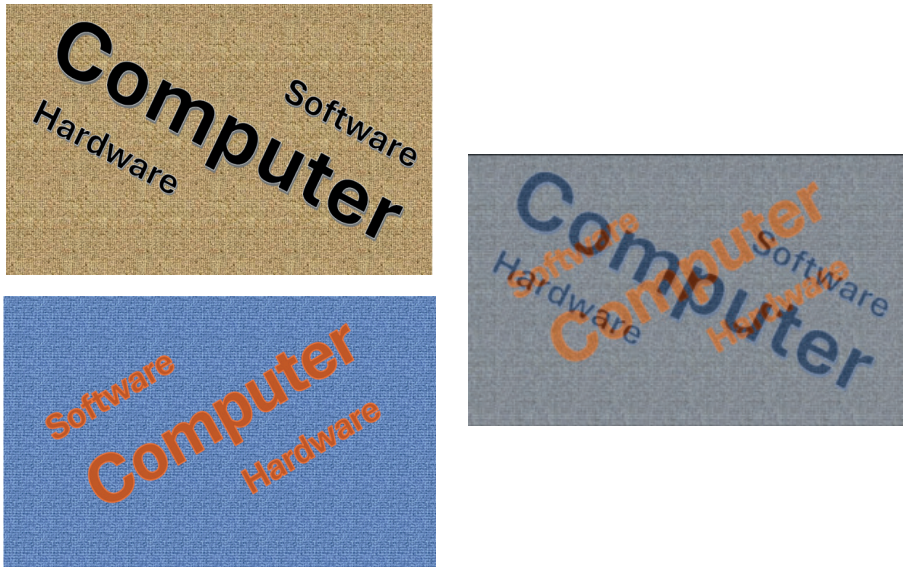
Cutout (Kes-Çıkar) Veri Artırma Yöntemi

CutOut silme işlemi, görüntü üzerindeki rastgele seçilen bir görüntü bölgesi silerek yeni görüntü verisi üretilmektedir (Şekil 3). Rastgele silme görsel olarak görüntü verisine modelin aşırı uyum sağlaması önlenmekte ve modelin eksik bilgiyle başa çıkma yeteneğini geliştirmektedir. Ayrıca bu işlem yapılarak modelin yerel özellikler yerine tüm görüntünün küresel özelliklerini öğrenmesini sağlanmaktadır. Bazı durumlarda rastgele silme işlemiyle, nesne tanıma görevinde önemli bilgiler silinebilir ve bu da görüntüde tanımlanmayan nesnelerin oluşmasına neden olabilmektedir. Bu nedenle, oluşturulan verilerin geçerliliğini sağlamak için kontrol işleminin gerçekleşmesi gerekmektedir (DeVries & Taylor, 2017).

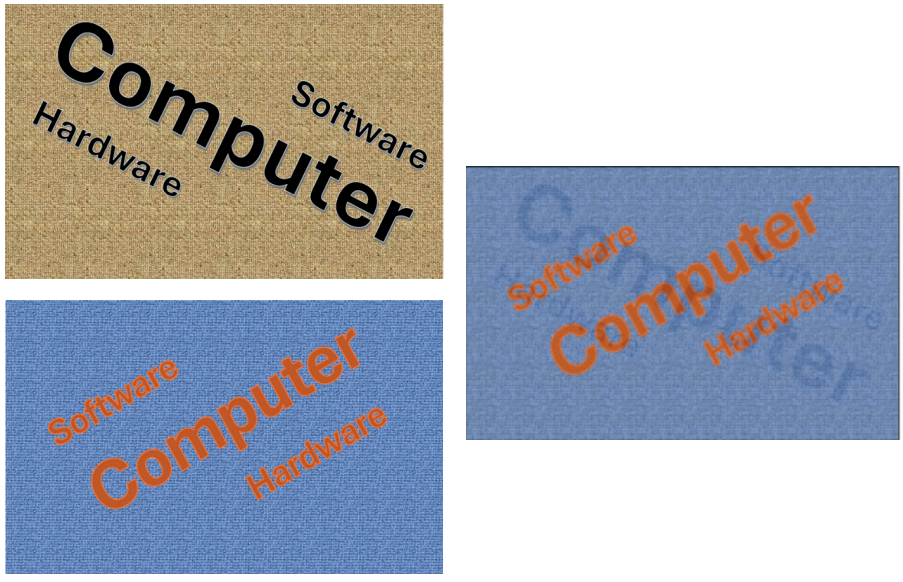
Şekil 3: Cutout yönteminin akış şeması

Deneysel Uygulamalar

Mixup yöntemiyle iki örnek görüntü üzerinde yapılan üç test ile elde edilen sonuçlar şekil 4 üzerinde gösterilmektedir. Elde edilen sonuçlarda iki görüntünün birleştirilmesi sonucu elde edilen farklı bir görüntü oluştuğu görülmektedir.



(a)

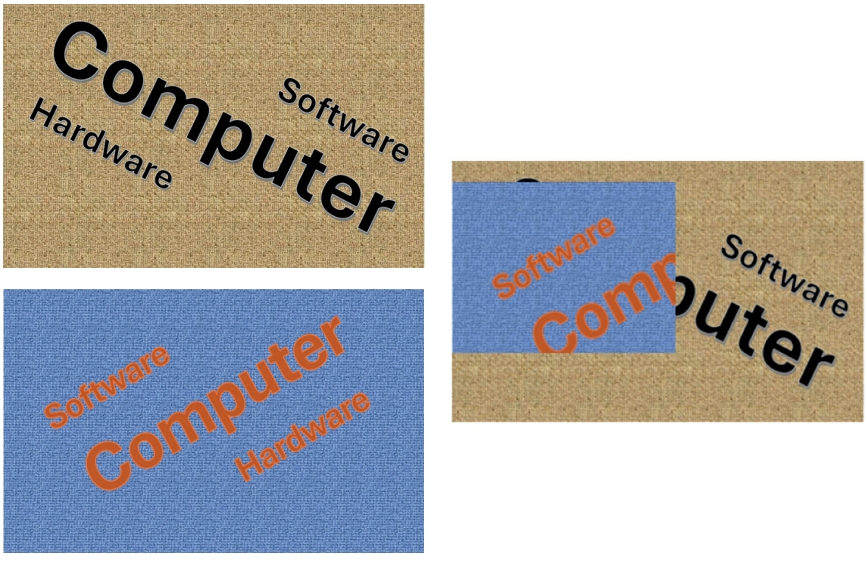


(b)

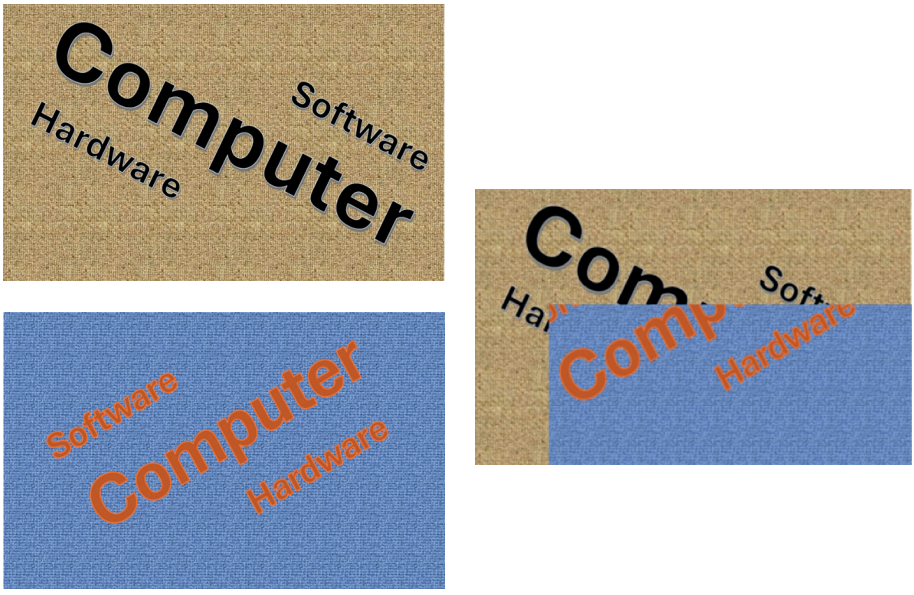
(c)

Şekil 4. Mixup yönteminin iki görüntü üzerinde üç kez test edilmesi. a-)Birinci test sonucu elde edilen görüntü, b-)İkinci test sonucu elde edilen görüntü c-)Üçüncü test sonucu elde edilen görüntü

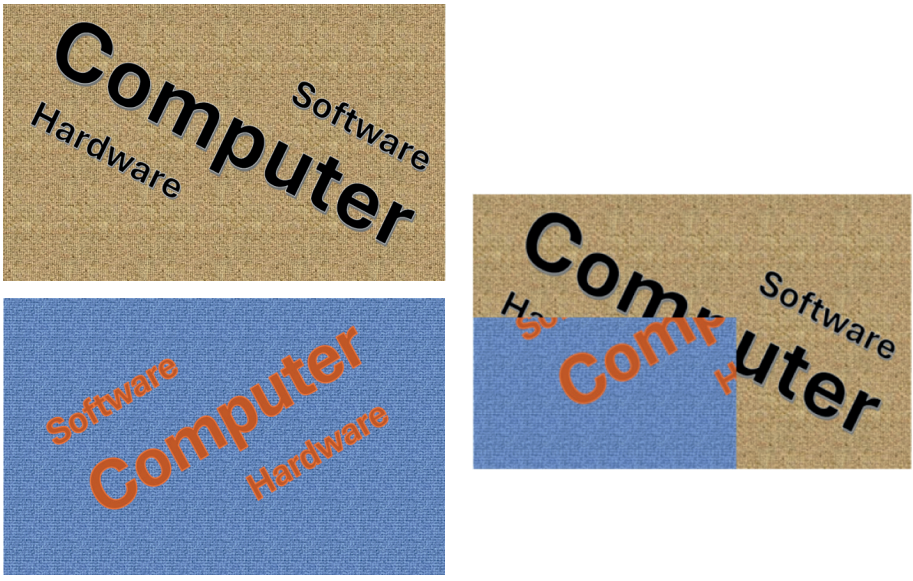
Cutmix yöntemiyle iki örnek görüntü üzerinde yapılan üç test ile elde edilen sonuçlar, Şekil 5 üzerinde gösterilmektedir. Elde edilen sonuçlarda, iki görüntü üzerinden kesilen parçaların birleştirilmesi sonucu elde edilen farklı bir görüntü oluştuğu görülmektedir.



(a)



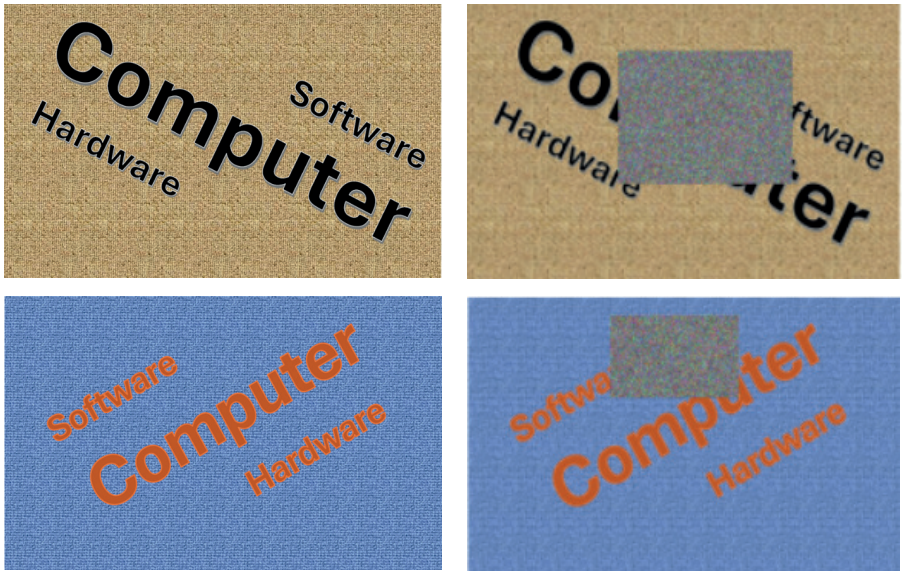
(b)



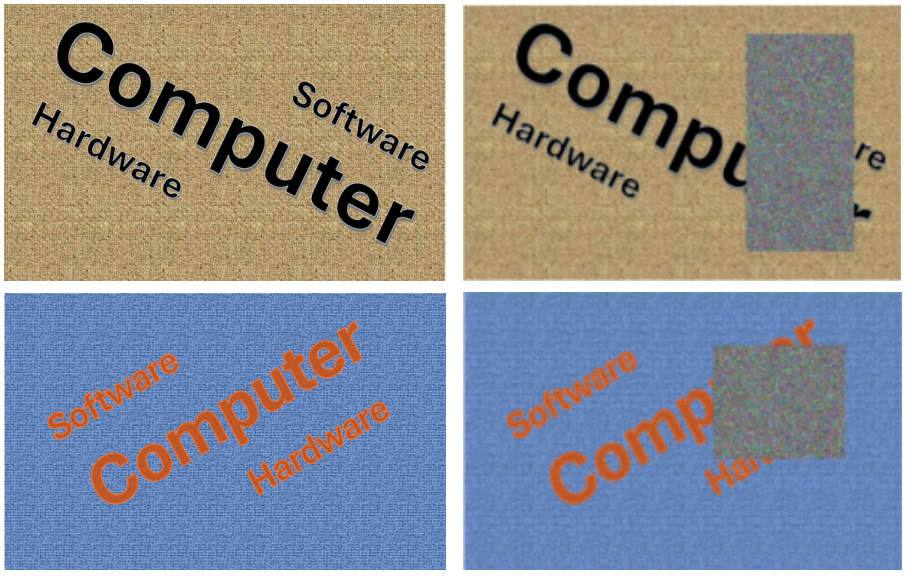
(c)

Şekil 5: Cutmix yönteminin iki görüntü üzerinde üç kez test edilmesi. a-)Birinci test sonucu elde edilen görüntü, b-)İkinci test sonucu elde edilen görüntü c-)Üçüncü test sonucu elde edilen görüntü

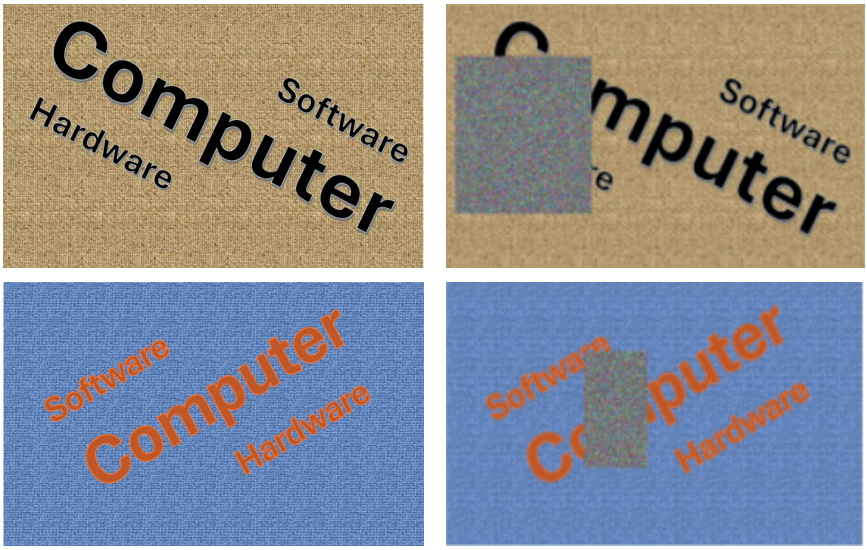
Cutout yöntemiyle iki örnek görüntü üzerinde yapılan üç test ile elde edilen sonuçlar şekil 6 üzerinde gösterilmektedir. Elde edilen sonuçlarda, iki görüntü üzerinden rastgele seçilen bölgelerden silinen parçalar ve elde edilen son görüntüler gösterilmektedir.



(a)



(b)



(c)

Şekil 6: Cutout yönteminin iki görüntü üzerinde üç kez test edilmesi. a-)Birinci test sonucu elde edilen görüntü, b-)İkinci test sonucu elde edilen görüntü c-)Üçüncü test sonucu elde edilen görüntü

Çalışmada kullanılan üç yöntemin işlem süreleri için beş test senaryosuna göre ölçülmüş ve ortalama işlem süreleri saniye birimi cinsinden, tablo 1 üzerinde gösterilmiştir.

Tablo 1: Veri artırma yöntemlerinin işlem süreleri

Yöntem	Beş test için işlem süresi(sn)	Ortalama İşlem Süresi(sn)
Mixup	2,0157 2,0662 2,1091 2,1001 2,0161	2,0614
CutMix	0,4840 0,4574 0,4544 0,4609 0,4808	0,4675
Cutout	0,4862 0,4973 0,5050 0,4815 0,5165	0,4973

En hızlı olan yöntemin CutMix olduğu, en yavaş yöntemin ise Mixup olduğu görülmektedir. Mixup yönteminin, uzun işlem süresi her iki görüntünün bütün piksel değerlerinin işlenmesinden kaynaklanmaktadır. CutMix ve Cutout algoritmalarının işlem süreleri birbirlerine yakındır. Bu durum, iki yöntemde görüntüler üzerinde bölgesel işlem yapılmasından kaynaklanmaktadır. Ek olarak görüntü içindeki bütün piksellerin işlenmesine gerek kalmamaktadır.

Başlangıçta veri seti kayıt sayısı 2n olmak üzere; veri artırma işlemleri sonucunda elde edilen görüntü kayıt sayılarının hesaplaması denklem 1-3'de gösterilmektedir.

$$Mixup_{\text{görüntü kayıt sayısı}} = n(2n-1) \quad (1)$$

$$CutMix_{\text{görüntü kayıt sayısı}} = n(2n-1) \quad (2)$$

$$Cutout_{\text{görüntü kayıt sayısı}} = 2n \quad (3)$$

Mixup ve CutMix yöntemlerinde iki görüntü işlenerek tek bir sonuç görüntüsü oluşturulmaktadır. Bundan dolayı işlem sonrasında oluşan görüntü sayısı, veri seti kayıt sayısının ikili kombinasyonu kullanılarak hesaplanmaktadır. Cutout görüntü yöntemi ile bir görüntü işlenerek bir sonuç görüntüsü elde edilmektedir. Bundan dolayı kayıt sayısı iki katına çıkmaktadır. Çalışmada kullanılan üç yöntemin veri seti kayıt sayıları üzerindeki değişimi tablo 2 üzerinde gösterilmektedir. Veri seti içinde ilk aşamada 2n kayıt olduğu kabul edilmiştir.

Tablo 2: Artırma yöntemlerinin kayıt sayısına olan etkisi

Yöntem Adı	**Veri seti kayıt sayısı (İşlem Öncesi)**	**Artırma yöntemiyle elde edilen kayıt sayısı**	**Veri seti kayıt sayısı (İşlem Sonrası)**
Mixup	2n	n(2n-1)	$2n^2+n$
CutMix	2n	n(2n-1)	$2n^2+n$
Cutout	2n	2n	4n

Sonuç ve Tartışma

Veri artırma teknikleri, birçok alanda yaygın olarak kullanılmaktadır. Özellikle yapay öğrenme modellerinde kullanılan veri setlerinde kullanılmaktadır. Maliyeti küçük yöntemlerle veri sayınının artırılması da önemlidir. Veri artırma yöntemleri kullanılarak veri setlerinin dengeli ve güvenilir olmasını sağlamaktadır. Ayrıca veri setlerinde, kayıt sayısının artırılması içinde kullanılmaktadır. Bu durum eğitim kümesinin, öğrenme başarısını artırarak modelinin tahmin başarısını da artırmaktadır. Bundan dolayı gelişmiş veri artırma yöntemlerinin incelenmesi, kayıt sayısına etkisinin hesaplanması ve işlem sürelerinin incelenmesi faydalı olacaktır. Çalışmada maliyeti düşük ve yaygın kullanılan Mixup, CutMix ve Cutout görüntü veri artırma yöntemleri kullanılmıştır. Bu çalışmadan elde edilen deneyimlere göre, gelecekteki çalışmalarda yöntemlerin birlikte kullanımı (Mixup + CutMix) yöntemi kullanılabilir. Ayrıca yeni bir hibrit yöntem önerilerek, eğitim süresine ve model karmaşıklığına etkileri ölçülebilecektir.

KAYNAKÇA

Wang, X., Wang, K., & Lian, S. (2020). A survey on face data augmentation for the training of deep neural networks. Neural Computing and Applications, 32, 15503–15531. https://doi.org/10.1007/s00521-020-04748-3

Li, S., Yue, C., & Zhou, H. (2025). Few-shot face recognition: Leveraging GAN for effective data augmentation. Electronics, 14(10), 2003. https://doi.org/10.3390/electronics14102003

Zhang, Y., & Liu, H. (2024). Data augmentation and transfer learning approaches applied to facial recognition. arXiv preprint arXiv:2402.09982. https://arxiv.org/abs/2402.09982

Liu, X., Karagoz, G., & Meratnia, N. (2025). Analyzing the impact of data augmentation on the explainability of deep learning-based medical image classification. Machine Learning and Knowledge Extraction, 7(1), 1. https://doi.org/10.3390/make7010001

Bozkaya, F., Yusefi, A., Tığlıoğlu, Ş., Kaya, A. K., Kazancı, O., Akmaz, M. Y., Durdu, A., & Sungur, C. (2021). Otonom sistemlerde veri çoğaltma yöntemleri kullanılarak iyileştirilmiş gerçek zamanlı nesne tespiti. European Journal of Science and Technology, Special Issue 30, 83-87. Dergipark

Azak, S., Bozkaya, F., & Durdu, A. (2024, September). Enhancing generalization in autonomous driving: A comparative study of data augmentation techniques. 25th National Automatic Control Conference (TOK'2024), Konya, Turkey. ResearchGate

Sousa, T., Ries, B., & Guelfi, N. (2025). Data augmentation in Earth observation: A diffusion model approach. Information, 16(2), 81. https://doi.org/10.3390/info16020081

Medvedieva, K., Tosi, T., Barbierato, E., & Gatti, A. (2024). Balancing the scale: Data augmentation techniques for improved supervised learning in cyberattack detection. Eng, 5(3), 2170–2205. https://doi.org/10.3390/eng5030114

Yan, P.; He, F.; Yang, Y.; Hu, F. Semi-Supervised Representation Learning for Remote Sensing Image Classification Based on Generative Adversarial Networks. IEEE Access 2020, 8, 54135–54144. [CrossRef]

Tokozume, Y.; Ushiku, Y.; Harada, T. Between-Class Learning for Image Classification. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 5486–5494.

Chapelle, O.;Weston, J.; Bottou, L.; Vapnik, V. Vicinal Risk Minimization. Adv. Neural Inf. Process. Syst. 2000, 13.

Yun, S.; Han, D.; Oh, S.J.; Chun, S.; Choe, J.; Yoo, Y. CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features. arXiv 2019, arXiv:1905.04899.

Hao, X.; Liu, L.; Yang, R.; Yin, L.; Zhang, L.; Li, X. A Review of Data Augmentation Methods of Remote Sensing Image Target Recognition. Remote Sens. 2023, 15, 827. https://doi.org/10.3390/rs15030827

DeVries, T.; Taylor, G.W. Improved Regularization of Convolutional Neural Networks with Cutout. arXiv 2017, arXiv:1708.04552.

Bölüm 2

GÜRÜLTÜLÜ GÖRÜNTÜLERDE ANİZOTROPİK DİFÜZYON FİLTRELERİNİN ENTERPOLASYON BAŞARISINA ETKİSİ

Mürsel Ozan İNCETAŞ [1]
Murat MERİÇELLİ[2]

1 Doç. Dr., Alanya Alaaddin Keykubat Üniversitesi, ALTSO MYO, Bilgisayar Teknolojileri Bölümü, ORCID: 0000-0002-1016-1655
2 Dr. Öğr. Üyesi, Alanya Alaaddin Keykubat Üniversitesi, ALTSO MYO, Bilgisayar Teknolojileri Bölümü, ORCID: 0000-0003-0168-3221

1. Giriş

Görüntü işleme alanında enterpolasyon, düşük çözünürlüklü ya da eksik veriye sahip görüntülerin yeniden yapılandırılması, büyütülmesi ve iyileştirilmesi amacıyla yaygın olarak kullanılan temel adımlardan biridir (Kılıçaslan, 2024b). Özellikle tıbbi görüntüleme, uydu görüntüleri, güvenlik kameraları, mobil görüntüleme sistemleri ve sayısal arşivleme gibi birçok uygulama alanında, görüntülerin çözünürlüğünün artırılması veya kayıp piksellerin tamamlanması kritik öneme sahiptir (Guo et al., 2024). Enterpolasyon işleminin başarısı, yalnızca kullanılan matematiksel modele değil, aynı zamanda işleme alınan görüntünün kalitesine ve içerdiği gürültü miktarına doğrudan bağlıdır. Dijital görüntülerden gürültülü içeriğin giderilmesi, görüntü ön işleme sırasında karşılaşılan en önemli sorunlardan biri olması önemli filtre çalışmalarına dikkat çekmektedir. Gürültü, görüntü edinme, sıkıştırma ve görüntü aktarma süreçlerinde bilgi kaybına neden olmaktadır. Bu bilgi kaybı, bilgisayarlı fotoğrafçılık, engel tespiti ve trafik izleme (bilgisayarlı görme), otomatik karakter tanıma, biçim değiştirme ve gözetim uygulamaları gibi birçok gerçek zamanlı uygulamanın çalışmasında düzensizliklere ve hatalara neden olur (Singh & Shankar, 2021). Gürültünün yüksek olduğu senaryolarda enterpolasyon algoritmaları beklenen performansı gösterememekte, kenarlarda bulanıklık, detay kaybı ve yapısal bozulmalar gibi istenmeyen etkiler ortaya çıkmaktadır (Gao et al., 2016). Bu nedenle enterpolasyon öncesinde görüntü kalitesini iyileştirmek amacıyla etkili bir gürültü giderme adımının uygulanması hem teorik hem de pratik açıdan gereklidir. Görüntü gürültüsünün giderilmesinde kullanılan filtreleme yöntemleri arasında Anizotropik Difüzyon Filtreleri (ADF), kenar koruma özellikleri ve yapısal ayrıntıları bozmadan gürültünün azaltılması konusundaki başarısı nedeniyle literatürde öne çıkmaktadır (Kılıçaslan, 2024a). Perona ve Malik tarafından önerilen klasik anizotropik difüzyon modeli, görüntüdeki düz bölgelerde difüzyonu artırarak gürültüyü azaltırken, yüksek gradyan bölgelerinde yani kenarlarda difüzyonu sınırlayarak önemli yapısal bilgilerin korunmasını sağlar (Kılıçaslan, 2023; Perona & Malik, 2002). Bu özellik, ADF'yi birçok görüntü iyileştirme ve önişleme uygulamasında tercih edilen bir yöntem hâline getirmiştir. Özellikle enterpolasyon gibi yeniden yapılandırma işlemlerinde kenar bilgisi kritik olduğundan, enterpolasyon öncesinde uygulanan ADF'nin sonuç üzerindeki etkisi araştırılması gereken önemli bir problemdir. ADF'ler literatürde sıklıkla kullanılmaktadır. Ağırlıklı anizotropik difüzyon filtresi (WADF) ve yapısal bir gradyan kullanan yeni bir yöntem 2023 yılında Vasu ve Palanisamy tarafından önerilmiştir. Söz konusu çalışmada kenar koruyan bir filtre, sınırlardaki yoğunlukların difüzyonu veya bir görüntüdeki anlamlı kenarların tespiti şeklinde tasarlanmıştır. Kenar koruyan bir yaklaşımla görüntü yumuşatma, önce ağırlık haritası deseni oluşturmak için kullanılmıştır ve füzyon karar haritası desenini oluşturmak için

yapısal gradyan tabanlı odak alanı algılama yaklaşımından faydalanılmıştır. Son olarak, ağırlık haritası deseni, füzyon kuralı aracılığıyla füzyon karar haritası deseniyle birleştirilerek birleştirilmiş bir görüntü oluşturulmuştur (Vasu & Palanisamy, 2023). Anizotropik difüzyon, görüntü işlemede kullanılan en etkili yöntemlerden olduğunu belirten Gupta ve arkadaşları görüntünün belirgin kenarlarını korurken küçük dokularını ortadan kaldırmak için kullanılabileceğini ifade etmişlerdir. Çalışmalarında filtrenin genel performansını iyileştirmek için tamsayı çekirdeği yerine kesirli hesaplama çekirdeğine dayalı yeni bir anizotropik difüzyon filtresi önermişlerdir. Önerilen filtre, görüntü yumuşatma, kenar algılama, görüntü segmentasyonu, görüntü gürültü giderme ve çizgi film oluşturmada kullanılabilir olması ADF'in farklı alanlarda kullanılacağını kanıtlar niteliktedir (Gupta & Lamba, 2021). ADF'lerin, evrişimli sinir ağlarıyla kullanılmasıda dikkat çeken yaklaşımlardandır. 2021 yılında yayımlanan makalede evrişimli sinir ağı (CNN) ve anizotropik difüzyon kullanan yeni bir hibrit ve çok seviyeli dijital görüntü gürültü giderme yaklaşımı önerilmiştir. İlgili çalışmada gürültü giderme için, çok seviyeli uygulama kullanarak CNN ve ADF'nin hibrit bir kombinasyonunu kullanılmıştır. Öncelikle, gürültü giderme için gürültülü görüntülere CNN uygulanır ve bu da görüntü gürültü gidermenin ilk seviyesinde gürültü giderilmiş bir görüntüyle sonuçlanır. Ardından, gürültü giderilmiş görüntü, görüntü gürültü gidermenin ikinci seviyesinde ADF'ye aktarılmıştır. ADF, nesnelerin kenar ve köşelerinin korunması için uygulanmıştır. Bu hibrit yaklaşım, görüntünün ince ayrıntılarını korurken gürültüyü gidermede oldukça etkili olmuştur (Singh & Shankar, 2021). Benzer şekilde göğüs röntgen görüntülerini normal veya enfeksiyonlu olarak kategorize etmek sunulan çalışmada, tıp uzmanlarının atelektazi hastalığını tanımlamasına yardımcı olmak için bir evrişimsel sinir ağı (CNN) yöntemi sunulmuştur ve ADF'den yararlanılmıştır. Anizotropik difüzyon filtreleme (ADF) yaklaşımı, görüntü kenarı korunmasını iyileştirmek, gürültüyü azaltmak ve düşük yoğunluklu görüntülerin kontrastını iyileştirmek için kontrast sınırlı adaptif histogram eşitlemesini kullanmak için kullanılmıştır (Ayalew et al., 2024).

Enterpolasyon yöntemleri, girdi görüntüsündeki piksellerin konumsal ilişkilerini kullanarak yeni piksel değerleri tahmin etmeye dayanır. Ancak gürültülü piksel değerleri ile hesaba dayalı bir tahmin yapmak, hatalı veri propagasyonuna yol açabilmekte ve özellikle yüksek büyütme oranlarında sonuç görüntüde ciddi bozulmalara neden olabilmektedir. Bu bağlamda, enterpolasyonun gürültüye duyarlı bir işlem olduğu ve girdinin gürültüden arındırılmasının nihai sonuç üzerinde belirleyici olduğu açıktır. Dolayısıyla enterpolasyonun başarısını artırmak için, işlem öncesinde gürültü giderme yöntemlerinin uygulanması doğal ve mantıklı bir yaklaşımdır. Ancak gürültü giderme işleminin kendisi de görüntüde belirli düzeyde bilgi kaybına neden olabileceği için hem yeterince gürültü azaltma hem de kenarları koruma

yeteneğine sahip bir yöntem seçilmesi gerekmektedir. Bu nedenlerden dolayı ADF, enterpolasyon öncesi önişleme için oldukça uygun bir filtreleme tekniği olarak değerlendirilmektedir. Chen ve arkadaşları görüntü enterpolasyonu için hızlı bir algoritma sunmuşlardır. Söz konusu yaklaşım ile video görüntülerinin gerçek zamanlı olarak büyütülmesi mümkün olduğunu ifade etmişlerdir. Görüntülerdeki yerel yapının analizine dayanarak dijital görüntüleri homojen ve kenarlara ayırmışlar ve görüntülerin enterpolasyonunda daha iyi performans elde etmek için, her sınıflandırılmış alanı sırasıyla enterpolasyon etmek üzere belirtilen algoritmalar çalıştırmışlardır. Deneysel sonuçlar, önerilen algoritmanın kullanılmasıyla, enterpolasyonlu görüntülerin öznel kalitesinin, geleneksel enterpolasyon algoritmalarının kullanımına kıyasla önemli ölçüde iyileştirildiğini göstermektedir (Chen et al., 2005). Gelişmiş enterpolasyon yöntemleri, görüntülerin gürültü uzayına kodlandığı ve ardından gürültü giderme için enterpolasyona tabi tutulduğu küresel doğrusal enterpolasyona odaklandığını belirten çalışmada, mevcut yöntemlerin difüzyon modelleri tarafından üretilmeyen doğal görüntüleri etkili bir şekilde enterpolasyonda zorluklarla karşılaşmasını vurgulamıştır. Çalışmada görüntü enterpolasyonunda gürültüyü düzeltmek için yeni bir yaklaşım olan NoiseDiffusion'ı önermişlerdir. NoiseDiffusion, özellikle ince Gauss gürültüsü ekleyerek geçersiz gürültüyü beklenen dağılıma yaklaştırmış ve aşırı değerlerdeki gürültüyü bastırmak için bir kısıtlama getirmiştir. NoiseDiffusion yöntemi, gürültülü görüntü alanı içinde enterpolasyon gerçekleştirmiş ve bu gürültülü karşılıklara ham görüntüler ekleyerek bilgi kaybı sorununu çözmüştür (Zheng et al., 2024). Önemli yerel bölgelerin geometrik benzerliğinden yararlanan ve aynı zamanda görüntü kenar yapısını korumasını sağlamaya çalışan bir görüntü enterpolasyon çalışması 2009 yılında sunulmuştur. Araştırmacılar hem yerel bölgeleri hem de görüntü kenarlarını tanımlayan yapılar tanımlamışlar ve kaynak görüntü için bir önem haritasına dayanarak her tutamaç için bir ağırlık atamışlardır. Geometri işlemede yaygın olarak kullanılan konformal enerjiden ilham alarak, her tutamaç için şekil bozulmasını ölçmek üzere yeni bir ikinci dereceden bozulma enerjisi oluşturmuşlardır. İlgili çalışma önceki yöntemlerle karşılaştırıldığında, yöntemin bozulmanın her yönde daha iyi dağılmasını sağladığını ve görüntü kenarlarının iyi korunduğunu belirtmişlerdir (Zhang et al., 2009).

Yukarıdaki çalışmalardan da anlaşılacağı üzere, enterpolasyon uygulamalarında kenar koruması kritik bir problem alanıdır ve geleneksel filtreler genellikle kenar bölgelerinde istenmeyen yumuşamaya neden olmaktadır. Anizotropik Difüzyon Filtreleri (ADF) ise, yapısal bilgiyi koruyarak gürültüyü seçici bir biçimde azaltan yapısı sayesinde bu sorunu çözmek için etkili bir yaklaşım sunmaktadır. Kenar bölgelerinde difüzyonu azaltıp homojen alanlarda artırarak çalışan ADF'ler, enterpolasyon işlemi öncesinde görüntünün daha istikrarlı ve detay kaybına dirençli bir hâle gelmesini

sağlamaktadır. Bu nedenle, ADF tabanlı iyileştirme adımlarının enterpolasyon performansına etkisinin incelenmesi hem teorik hem de uygulamalı açıdan önemli bir araştırma alanı olarak ortaya çıkmaktadır. Bu çalışmada, anizotropik difüzyon filtrelerinin enterpolasyon başarısı üzerindeki etkisi incelenmiştir. Çalışma kapsamında referans görüntüler önce düşük çözünürlüğe indirgenmiş, ardından Gauss gürültüsü eklenmiş ve enterpolasyondan önce ADF uygulanarak iyileştirme işlemi gerçekleştirilmiştir. Ardından enterpolasyon yardımıyla büyütülerek elde edilen görüntüler, orijinal görüntüler ile PSNR ve SSIM gibi objektif kalite metrikleri üzerinden karşılaştırılarak değerlendirilmiştir. Böylece ADF önişlemesinin enterpolasyon performansını iyileştirip iyileştirmediği, iyileştiriyorsa hangi ölçüde katkı sağladığı ortaya belirlenmeye çalışılmıştır.

2. Anizotropik Difüzyon Filtreleri (ADF)

ADF'ler, görüntü iyileştirme ve gürültü giderme süreçlerinde, difüzyon akışını yerel gradyan bilgisine bağlı olarak yönlendiren, böylece homojen bölgelerde gürültüyü etkili biçimde azaltırken kenar bölgelerinde difüzyonu sınırlayarak yapısal detayların korunmasını sağlayan gelişmiş bir kısmi diferansiyel denklem tabanlı bir yöntemdir. Geleneksel izotropik difüzyon yöntemlerinin aksine, ADF gradyan büyüklüğünü uzamsal bir iletkenlik fonksiyonu ile ilişkilendirerek difüzyon katsayısını adaptif biçimde kontrol eder ve bu sayede özellikle yüksek frekanslı kenar bölgelerinde bilgi kaybını minimize eden seçici bir filtreleme mekanizması sunar (Perona et al., 1994). Perona ve Malik (PM) tarafından ilk kez önerilen ve gürültü giderme için yaygın olarak kullanılan anizotropik difüzyon filtresinin ısı difüzyonu tabanlı denklemi aşağıdaki gibidir:

$$\frac{dI}{dt} = div(c(\nabla I)\nabla I) \qquad (1)$$

Burada div ve sırasıyla ıraksaklığı, gradyan operatörünü ve zamanı ifade eder. C ise difüzyon katsayısıdır. C değeri;

$$c(x,y,t) = \frac{1}{1+\left(\|\nabla I\| / K\right)^2} \qquad (2)$$

şeklinde hesaplanmaktadır. K bu arada eşik değerdir ve bir sabit olarak değerlendirilir.

3. Enterpolasyon

Görüntü enterpolasyonu, mevcut piksel değerlerinden yararlanarak ara noktalardaki yeni piksel değerlerinin tahmin edilmesi işlemidir. Başka bir ifadeyle, enterpolasyon bir görüntünün uzamsal çözünürlüğünü artırmak veya geometrik dönüşüm sırasında eksik kalan piksel değerlerini hesaplamak için kullanılan matematiksel bir yaklaşımdır. Temel amaç, yeni üretilen pikselin değerini, komşuluk yapısına dayalı fonksiyonlarla olabildiğince gerçekçi ve yapısal tutarlılığı koruyacak şekilde belirlemektir. Enterpolasyonun genel matematiksel gösterimi şu şekildedir:

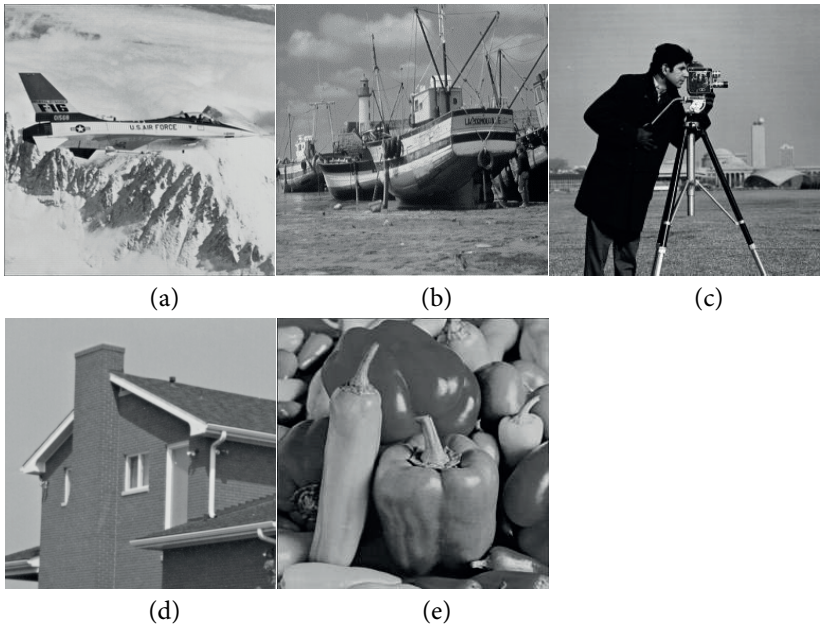
$$I'(x,y) = \sum_i \sum_j I(i,j) w(x-i, y-j) \qquad (3)$$

Burada I, orijinal görüntüyü, I' enterpolasyon sonucu çıktı görüntüyü ve w ise komşu piksellerin katkısını belirleyen ağırlık fonksiyonunu ifade eder.

4.Deneysel Bulgular ve Tartışma

4.1.Veri Seti ve Metrikler

Çalışmada 256x256 boyutunda 5 adet gri seviyeli görüntü kullanılmıştır. Kullanılan görüntüler Şekil 1'de görülmektedir. Deneylerin tamamı Matlab programı aracılığıyla gerçekleştirilmiştir.



(a) (b) (c)

(d) (e)

Şekil 1. 256x256 boyutunda 5 adet gri seviyeli görüntü. a) airplane, b) boat, c) cameraman, d) house, e) peppers

Görüntü iyileştirme, gürültü giderme ve enterpolasyon gibi görüntü işleme yöntemlerinin başarısının nesnel olarak değerlendirilebilmesi için uygun performans metriklerinin kullanılması büyük önem taşımaktadır. Bu amaçla literatürde hem piksel tabanlı hem de algısal kaliteyi esas alan çeşitli ölçütler bulunmaktadır. Bu çalışma kapsamında, önerilen yaklaşımın görsel kalite üzerindeki etkisini değerlendirmek için en yaygın kullanılan iki metrik olan Peak Signal-to-Noise Ratio (PSNR) ve Structural Similarity Index Measure (SSIM) kullanılmıştır. Bu iki metrik birbirini tamamlar nitelikte olup, PSNR sayısal hata bazlı bir değerlendirme sağlarken SSIM insan görsel algısına daha yakın bir benzerlik ölçümü sunmaktadır. PSNR, orijinal görüntü ile işlenmiş görüntü arasındaki hatayı ölçmek için kullanılan klasik bir kalite metriktir. PSNR temelde Ortalama Kare Hatası'na (MSE) dayanır ve iki görüntü arasındaki fark ne kadar küçükse PSNR değeri o kadar yüksek olur. Dolayısıyla yüksek PSNR değeri, yeniden üretilen görüntünün orijinale daha yakın olduğunu göstermektedir. PSNR'in matematiksel denklemi Eşitlik 4'te verilmiştir.

$$MSE = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left[I(i,j) - I'(i,j)\right]^2$$
$$PSNR = 10\log\left(\frac{L^2}{MSE}\right) \quad (4)$$

Eşitlik 4'te M ve N görüntü boyutlarını temsil etmektedir. L ise gri seviye değer olup maximum 255 değerini almaktadır.

SSIM, insan görsel sisteminin kontrast, parlaklık ve yapı duyarlılığını dikkate alarak iki görüntü arasındaki benzerliği ölçen, algısal temelli bir metriktir. PSNR'ın aksine sadece piksel bazlı farklara odaklanmaz; görüntüdeki yapısal bilgilerin korunup korunmadığını da değerlendirir. Bu nedenle özellikle kenar detayları, doku bilgisi ve yapısal bütünlüğün önemli olduğu enterpolasyon ve iyileştirme çalışmalarında yaygın olarak tercih edilmektedir. SSIM ise;

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x{}^2 + \mu_y{}^2 + c_1)(\sigma^2{}_x + \sigma^2{}_y + c_2)} \quad (5)$$

olarak formülüze edilir. Burada ortalama parlaklık, varyans, kovaryans ve ise sabittir. SSIM değeri 0-1 arasında değer alır ve 1'e yakınlaşması yüksek yapısal benzerlik olduğunu anlamına gelir.

4.2. Deneysel Bulgular

Bu çalışmada, ADF'lerin enterpolasyon başarısına etkisi belirlenmeye çalışılmıştır. Bu amaçla öncelikle 256x256 boyutundaki orijinal görüntüler, 128x128 boyutuna küçültülmüştür. Şekil 2'de görülen bu görüntülere imnoise(grayKucuk,'gaussian',0,0.005) yardımıyla gauss gürültüsü eklenmiştir. Şekil 3'te ise küçültülmüş ve gürültü eklenmiş görüntüler yer almaktadır.

(a) (b) (c) (d) (e)

Şekil 2. 128x128 boyutuna küçültülmüş görüntüler

Şekil 3'te yer alan görüntülerdeki bozulmalar net şekilde görülmektedir. Karşılaştırmaların doğru yapılabilmesi amacıyla, gürültü küçültülmüş görüntülere uygulanmıştır.

(a) (b) (c) (d) (e)

Şekil 3. 128x128 boyutunda gürültülü görüntüler

Gürültülü görüntüler üzerinde ADF yardımıyla gürültü giderme işlemi imdiffusefilt(noisyImage) fonksiyonu ile gerçekleştirilmiştir. Böylece 128x128 boyutunda gürültüden arındırılmış görüntüler elde edilmiştir. Söz konusu görüntüler Şekil 4'te verilmiştir. Şekil 4'teki görüntüler, Şekil 3'te verilen görüntülerle görsel olarak kıyaslandığında gürültülerin büyük oranda temizlendiği görülmektedir. Ancak yine de Şekil 2 ile kıyaslandığında, görüntü kalitesinin daha düşük olduğu açıktır.

(a) (b) (c) (d) (e)

Şekil 4. 128x128 boyutunda gürültüden arındırılmış görüntüler

Son olarak da 128x128 boyutundaki hem gürültüsüz, 3 farklı enterpolasyon yaklaşımı ile hem gürültülü hem de gürültüden arındırılmış görüntüler 256x256 boyutuna tekrar getirilmişlerdir. Büyütme esnasında en yakın komşu (nearest), bilneer (bilinear) ve bikübik (bicubic) enterpolasyon yaklaşımları kullanılmıştır. Veri setinde yer alan peppers görüntüsü için elde edilen enterpolasyon görüntüleri Şekil 5, 6 ve 7'de yer almaktadır.

(a) (b) (c)

Şekil 5. Peppers görüntüsü için gürültüsüz görüntülerin büyütülmesi ile elde edilen 256x256 boyutundaki enterpolasyon görüntüleri, a) en yakın komşu enterpolasyonu b) bilineer enterpolasyon, c) bikübik enterpolasyon

Şekil 5'te, peppers görüntüsünün 128x128 boyutundan, 3 farklı enterpolasyon yaklaşımı yardımıyla 256x256 boyutuna büyütülmesi sonucunda elde edilen görüntüler görülmektedir. Şekil 5.(a), (b) ve (c)'de yer alan görüntüler sırasıyla en yakın komşu, bilineer ve bikübik enterpolasyonlar yardımıyla elde edilmiştir.

(a) (b) (c)

Şekil 6. Peppers görüntüsü için gürültülü görüntülerin büyütülmesi ile elde edilen 256x256 boyutundaki enterpolasyon görüntüleri, a) en yakın komşu enterpolasyonu, b) bilineer enterpolasyon, c) bikübik enterpolasyon

Peppers görüntüsüne 128x128 boyutundaki gürültü eklendikten sonra 3 farklı enterpolasyon yaklaşımı yardımıyla 256x256 boyutuna büyütülmesi sonucunda elde edilen görüntüler Şekil 6'da görülmektedir. Şekil 7'de ise 128x128 boyutundaki peppers görüntsünün gürültüsü ADF ile giderildikten sonra 3 farklı enterpolasyon yaklaşımı yardımıyla 256x256 boyutuna büyütülmesi sonucunda elde edilen görüntüler yer almaktadır.

(a) (b) (c)

Şekil 7. Gürültü eklenmiş peppers görüntüsünün ADF ile gürültü giderilmesi sonrasında büyütülmesi ile elde edilen 256x256 boyutundaki enterpolasyon görüntüleri, a) en yakın komşu enterpolasyonu, b) bilineer enterpolasyon, c) bikübik enterpolasyon.

Enterpolasyon yardımıyla 256x256 boyutuna getirilen tüm görüntüler, orijinal görüntülerle ile karşılaştırılmış ve bu karşılaştırma sonuçları PSNR ve SSIM olarak sırasıyla Tablo 1 ve Tablo 2'de verilmiştir.

Tablo 1

Enterpolasyon sonuçlarının PSNR başarısı

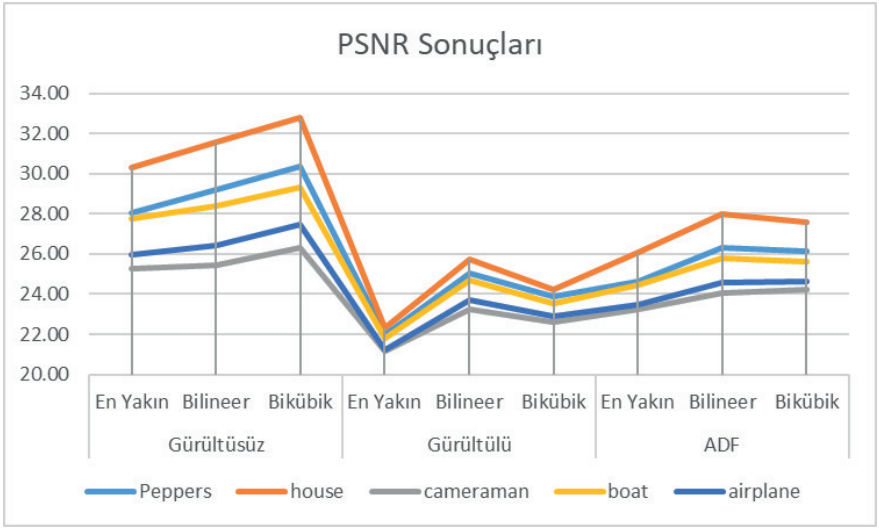
PSNR	Gürültüsüz			Gürültülü			ADF		
	En Yakın	Bilineer	Bikübik	En Yakın	Bilineer	Bikübik	En Yakın	Bilineer	Bikübik
peppers	28,02	29,22	30,39	21,97	25,03	23,87	24,62	26,33	26,17
house	30,33	31,59	32,76	22,31	25,76	24,22	26,06	28,00	27,61
cameraman	25,24	25,45	26,28	21,14	23,22	22,63	23,24	24,04	24,23
boat	27,75	28,38	29,30	21,79	24,70	23,55	24,45	25,77	25,61
airplane	25,97	26,45	27,44	21,20	23,72	22,91	23,46	24,58	24,65
Ortalama	27,46	28,22	29,23	21,68	24,48	23,44	24,37	25,74	25,65

Tablo 2

Enterpolasyon sonuçlarının SSIM başarısı

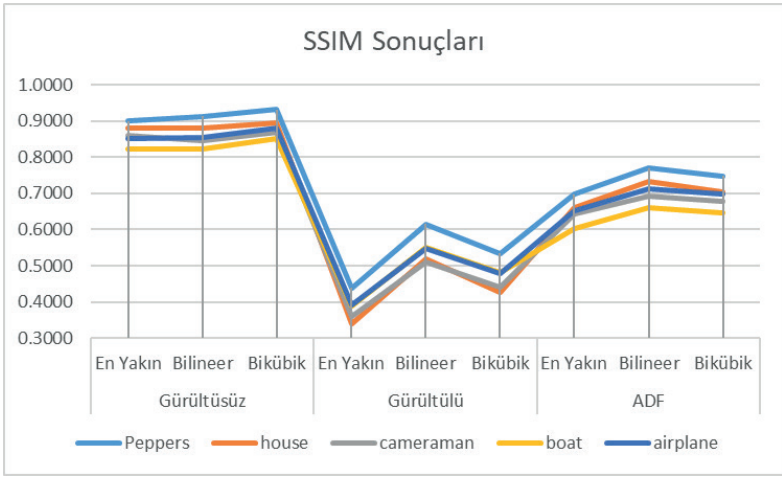
SSIM	Gürültüsüz			Gürültülü			ADF		
	En Yakın	Bilineer	Bikübik	En Yakın	Bilineer	Bikübik	En Yakın	Bilineer	Bikübik
peppers	0,8997	0,9130	0,9314	0,4384	0,6146	0,5336	0,6967	0,7701	0,7482
house	0,8791	0,8807	0,8958	0,3395	0,5191	0,4274	0,6616	0,7334	0,7052
cameraman	0,8591	0,8456	0,8701	0,3596	0,5109	0,4393	0,6427	0,6930	0,6774
boat	0,8235	0,8234	0,8505	0,3889	0,5507	0,4799	0,6034	0,6618	0,6474
airplane	0,8526	0,8552	0,8805	0,3925	0,5468	0,4790	0,6529	0,7122	0,6983
Ortalama	0,8628	0,8636	0,8856	0,3838	0,5484	0,4718	0,6514	0,7141	0,6953

Aynı zamanda Tablo 1 ve Tablo 2'de yer alan veriler, Şekil 8 ve Şekil 9'da grafik olarak da verilmiştir.



Şekil 8. PSNR sonuçları grafiği

Tablo 1 ve 2 ile Şekil 8 ve 9'da yer alan SSIM ve PSNR sonuçlarına göre gürültüsüz görüntüler üzerinde yapılan enterpolasyon işlemleri için en başarılı sonucun bikübik enterpolasyon tekniği ile elde edildiği görülmektedir. Buna karşın gürültülü görüntülere ve gürültüsü ADF ile giderilmiş görüntülerde ise en başarılı sonuçlar bilineer enterpolasyon tekniği ile elde edilmiştir. En yakın komşu enterpolasyonu ise başarı sıralamasında tüm deneylerde son sırada yer almıştır.



Şekil 9. SSIM sonuçları grafiği

Elbette bu sonuçlar sınırlı sayıdaki görüntü ile yapılan deneyler yardımıyla elde edilmiştir. Bu nedenle gelecekte daha büyük veri setleri üzerinde çalışmalar yapılması planlanmaktadır. Ayrıca hem ADF'ler hem de enterpolasyon yaklaşımlarında kenarları koruma başarısı daha yüksek olan ve özellikle SNN kullanılan yaklaşımların başarılarının da test edilmesi gelecekte planlanan çalışma konuları arasındadır. Ayrıca ADF'lerin renk indirgeme (Kılıçaslan & İncetaş, 2023) ya da görüntü erişimi (İncetaş et al., 2022; Kılıçaslan et al., 2020) gibi farklı görüntü işleme adımlarına etkilerinin araştırılması da planlanmaktadır.

5. Sonuç

Enterpolasyon, görüntülerin boyutunun arttırılarak detayların belirlenmesine yardımcı olan önemli bir görüntü işleme alanıdır. Ancak görüntülerdeki bozulmaların enterpolasyon başarısını düşürdüğü açıktır. Bu nedenle görüntülerin gürültülerinin giderilmesi işlemi ile birlikte kullanılması da kaçınılmazdır. Özellikle son yıllarda yapılan çalışmalarda ADF'lerin gürültülerin giderilmesinde kenar koruma başarısının yüksek olduğu vurgulanmıştır. Bu çalışmada, ADF'lerin enterpolasyon başarısı üzerindeki etkileri araştırılmıştır. Her ne kadar gürültüsüz görütülerde bikübik enterpolasyon tekniğinin daha başarılı olduğu düşünülse de elde edilen deneysel bulgular, gürütültü ya da ADF ile gürültüsü giderilmiş görüntülerde bilineer enterpolasyon tekniğinin daha başarılı olduğunu göstermiştir. Buna karşın literatürde pek çok enterpolasyon yaklaşımı da mevcuttur. Aynı şekilde gürültü giderme için farklı yaklaşımları kullanan ADF'ler ve farklı türde gürültü filtreleri de kullanılmaktadır. Dolayısıyla daha kesin sonuçlara varabilmek için gelecekte tüm bu farklı yaklaşımların detaylı olarak karşılaştırıldığı çalışmalar planlanmaktadır.

KAYNAKLAR

Ayalew, A. M., Bezabh, Y. A., Abuhayi, B. M., & Ayalew, A. Y. (2024). Atelectasis detection in chest X-ray images using convolutional neural networks and transfer learning with anisotropic diffusion filter. *Informatics in Medicine Unlocked*, *45*, 101448.

Chen, M.-J., Huang, C.-H., & Lee, W.-L. (2005). A fast edge-oriented algorithm for image interpolation. *Image and Vision Computing*, *23*(9), 791–798.

Gao, Z., Xu, X., Su, Y., & Zhang, Q. (2016). Experimental analysis of image noise and interpolation bias in digital image correlation. *Optics and Lasers in Engineering*, *81*, 46–53.

Guo, A., Lin, E., Zhang, J., & Liu, J. (2024). An energy-efficient image filtering interpolation algorithm using domain-specific dynamic reconfigurable array processor. *Integration*, *96*, 102167.

Gupta, B., & Lamba, S. S. (2021). An efficient anisotropic diffusion model for image denoising with edge preservation. *Computers & Mathematics with Applications*, *93*, 106–119.

İncetaş, M. O., Kılıçaslan, M., & Farshi, T. R. (2022). Image retrieval with SNN-based multi-level thresholding. *Gümüşhane Üniversitesi Fen Bilimleri Dergisi*, 98–108.

Kılıçaslan, M. (2023). Adaptive Threshold Selection of Anisotropic Diffusion Filters Using Dissimilarity Transform. *International Journal of Engineering Research and Development*, *15*(2), 558–573.

Kılıçaslan, M. (2024a). Adaptive threshold selection of anisotropic diffusion filters using spiking neural network model. *Signal, Image and Video Processing*, *18*(1), 407–416.

Kılıçaslan, M. (2024b). Image interpolation with spiking neural network based pixel similarity. *Signal, Image and Video Processing*, *18*(10), 6925–6936.

Kılıçaslan, M., & İncetaş, M. O. (2023). Adaptive color quantization method with multi-level thresholding. *International Journal of Computational Intelligence Systems*, *16*(1), 7.

Kılıçaslan, M., Tanyeri, U., & Demirci, R. (2020). Tekrarlı Ortalama Yardımıyla Renk İndirgeme ve Görüntü Erişimi. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, *8*(1), 1042–1057.

Perona, P., & Malik, J. (2002). Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *12*(7), 629–639.

Perona, P., Shiota, T., & Malik, J. (1994). Anisotropic diffusion. In *Geometry-driven diffusion in computer vision* (pp. 73–92). Springer.

Singh, P., & Shankar, A. (2021). A novel optical image denoising technique using convolutional neural network and anisotropic diffusion for real-time surveillance applications. *Journal of Real-Time Image Processing*, *18*(5), 1711–1728.

Vasu, G. T., & Palanisamy, P. (2023). Gradient-based multi-focus image fusion using foreground and background pattern recognition with weighted anisotropic diffusion filter. *Signal, Image and Video Processing*, *17*(5), 2531–2543.

Zhang, G., Cheng, M., Hu, S., & Martin, R. R. (2009). A shape-preserving approach to image resizing. *Computer Graphics Forum*, *28*(7), 1897–1906.

Zheng, P., Zhang, Y., Fang, Z., Liu, T., Lian, D., & Han, B. (2024). Noisediffusion: Correcting noise for image interpolation with diffusion models beyond spherical linear interpolation. *The Twelfth International Conference on Learning Representations*.

Bölüm 3

KLASİK AJAN SİSTEMLERİNDEN LLM TABANLI AJANLARA: GÜNCEL YAPAY ZEKÂ PARADİGMALARI

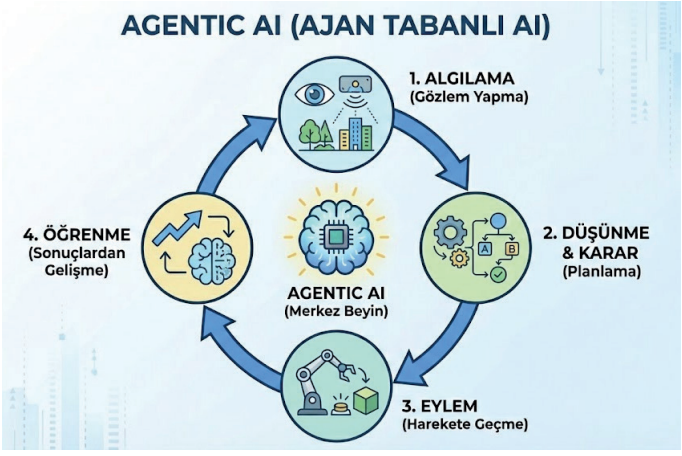
Ayşe Gül EKER [1]

1 Ayşe Gül Eker, Dr., Kocaeli Üniversitesi, Bilgisayar Mühendisliği, orcid: 0000-0003-0721-2631

1. Giriş

Ajan tabanlı yapay zekâ, modern yapay zekâ araştırmalarının ve uygulamalarının merkezinde yer alan temel kavramlardan biridir. Klasik tanıma göre bir yapay zekâ ajanı, bulunduğu çevreden algılama yapan, bu algılar doğrultusunda kararlar veren ve eylemler gerçekleştiren otonom bir varlık olarak ele alınır [1]. Başka bir deyişle, ajan yalnızca pasif bir hesaplama birimi değil; çevresiyle sürekli etkileşim hâlinde olan, hedeflere yönelik davranan ve zaman içinde durumunu güncelleyebilen aktif bir sistemdir. Çok ajanlı sistemlere yönelik çalışmalar, bu tür ajanların bir araya gelerek iş birliği, rekabet veya müzakere gerektiren karmaşık senaryolarda nasıl davrandığını ayrıntılı biçimde incelemiştir [2].

Büyük dil modellerinin yükselişi, ajan kavramını yeniden tanımlayan önemli bir kırılma noktası oluşturmuştur. Geleneksel ajanlar çoğunlukla kural tabanlı yapılar, belirgin durum uzayları ve dar görev tanımlarıyla çalışırken; günümüzde büyük dil modeli temelli ajanlar geniş veri üzerinde eğitilmiş, doğal dilde akıl yürütebilen ve çok farklı görev alanlarına uyarlanabilen esnek sistemler hâline gelmiştir. Bu modeller artık yalnızca soruları yanıtlayan yapılar değil; hedef belirleyen, plan yapan, araç kullanan ve diğer ajanlarla etkileşime giren bütünleşik sistemler olarak tasarlanmaktadır. Bu dönüşüm, literatürde giderek yaygınlaşan "ajanik veya ajan tabanlı LLM'ler (agentic LLMs)" kavramının ortaya çıkmasına yol açmıştır. Şekil 1'de bu yapıya ilişkin genel bir mimari sunulmaktadır.



Şekil-1: Ajan tabanlı AI genel şema

Araştırmalar, LLM tabanlı ajanların üç temel eksende ele alınabileceğini göstermektedir: akıl yürütme (reasoning), eyleme geçme (acting) ve etkileşim (interacting) [3]. Akıl yürütme, karmaşık görevleri alt parçalara ayırma ve planlama süreçlerini içerir. Eyleme geçme, aracılar, yazılım bileşenleri veya fiziksel robotlar üzerinden dünyada gerçek bir etki yaratılmasıyla

ilgilidir. Etkileşim ise, birden fazla ajanın birbirleriyle veya insanlarla doğal dil aracılığıyla iletişim kurarak ortak görevleri yerine getirmesini kapsar [4]. Bu üç boyut bir araya geldiğinde, klasik tek çağrıda cevap veren LLM anlayışından farklı olarak sürekli çalışan, ortamla döngüsel etkileşim içinde olan ajan sistemleri ortaya çıkmaktadır.

Bu bağlamda "ajan tabanlı yapay zekâ sistemleri", hem klasik yapay zekâ kuramında geliştirilen biçimsel ajan modellerini (örneğin mantıksal ajanlar, BDI mimarileri, çok ajanlı sistemler) hem de LLM tabanlı çağdaş ajan yaklaşımlarını kapsayan geniş bir çerçeve sunar. Klasik literatür, ajan–çevre ilişkisinin temel ilkelerini ve rasyonel davranış ölçütlerini ortaya koyarken; yeni çalışmalar LLM tabanlı ajanların mimarilerini, bellek ve planlama bileşenlerini, araç kullanma yeteneklerini ve çok ajanlı senaryolardaki davranış biçimlerini sistematik olarak sınıflandırmaktadır [5].

Ajan kavramının temel özellikleri tarihsel olarak Franklin ve Graesser'ın çalışmasıyla biçimsel bir çerçeveye kavuşmuştur. Çevreyi algılama, hedef yönelimliliği, otonomi, sürekli etkinlik ve davranış üretimi gibi ölçütler, bir sistemi "ajan" yapan kritik unsurlar olarak ortaya konmuştur [6]. Bu tanım, günümüzde hem robotik hem yazılım ajanlarını hem de LLM tabanlı yeni nesil otonom sistemleri kapsayan geniş bir çatı sunmaktadır. Son yıllarda ajanik yapay zekâ alanına yönelik inceleme çalışmaları, LLM tabanlı ajanların yalnızca metin üretimiyle sınırlı olmadığını; araç kullanma, dış bilgi kaynaklarına bağlanma, çok ajanlı koordinasyon ve uzun süreli görev yürütme gibi yeteneklerle klasik ajan anlayışını önemli ölçüde genişlettiğini göstermektedir [7]. Böylece akıllı ajan kavramı artık hem bireysel karar verme süreçlerini hem de çok ajanlı iş birliği dinamiklerini içine alan modern ve bütünsel bir perspektif sunmaktadır.

Bu bölümün amacı, ajan tabanlı yapay zekâ sistemlerini temel kavramlar, mimari yapı taşları ve modern uygulamalar açısından bütüncül biçimde ele almaktır. İlk olarak ajan kavramının temel özellikleri ve klasik tanımları üzerinde durulacak; ardından ajan mimarisinin temel bileşenleri (algılama, durum temsili, akıl yürütme, planlama, eylem ve bellek) tartışılacaktır. Daha sonra, LLM tabanlı ajan yaklaşımlarının ortaya çıkışı, temsilî çerçeveler ve güncel uygulama örnekleri ele alınacak; son olarak ajan sistemlerinin sınırları, güvenlik boyutları ve geleceğe dönük araştırma yönelimleri üzerinde durularak bölüm tamamlanacaktır. Böylece okur, hem geleneksel ajan kuramının kavramsal temellerini hem de çağdaş "ajanik LLM" yaklaşımlarını aynı çatı altında görme imkânı bulacaktır.

2. Ajan Mimarileri: Kural Tabanlıdan Öğrenen Sistemlere

Ajan mimarileri, bir ajanın çevreden gelen bilgiyi nasıl işlediğini, bu bilgilerden nasıl anlam çıkardığını ve hangi mekanizmalarla eylem ürettiğini belirleyen temel yapılardır. Tarihsel gelişim çizgisi incelendiğinde, ajan

sistemlerinin basit kural tabanlı yaklaşımlardan başlayarak soyut akıl yürütme mekanizmalarına, bilişsel modellere ve günümüzde öğrenme temelli, özellikle de LLM destekli mimarilere doğru evrildiği görülür. Bu bölümde, bu mimari yaklaşımlar genel hatlarıyla ele alınmaktadır.

2.1 Tepkisel Mimariler

Tepkisel(reaktif) mimariler, ajanın çevredeki duruma hızlı ve doğrudan tepki vermesini sağlayan en temel mimari türüdür. Bu yaklaşımda karmaşık içsel temsillere veya uzun vadeli planlara yer verilmez; bunun yerine koşul–eylem (condition–action) kuralları üzerinden anlık karar üretimi gerçekleşir. Brooks'un mobil robotlar için bir dönüm noktası kabul edilen alt-yayılım mimarisi, tepkisel yaklaşımın en çok bilinen örneklerinden biridir [8]. Güncel çalışmalar, tepki-temelli ajan mimarilerinin kritik gerçek-zamanlı uygulamalarda önemini koruduğunu ve modern melez yaklaşımlarda refleks benzeri bir davranış katmanı olarak etkin biçimde konumlandığını göstermektedir [9]. Bu tür mimariler özellikle robotik gibi zaman duyarlı ortamlarda avantaj sağlar; ancak soyut akıl yürütme, ileriye dönük plan yapma ve karmaşık görevleri çözme konusunda sınırlıdır.

2.2 Bilişsel ve Hedef-Tabanlı Mimariler

Tepkisel modellerin sınırlamalarını aşmak için geliştirilen bilişsel mimariler, ajanın içsel durumunu temsil eden inanç, istek, niyet gibi yapılar üzerinden karar verme süreçleri oluşturur. BDI (Belief–Desire–Intention) mimarisi, bu yaklaşımın en sistematik örneklerinden biridir. Rao ve Georgeff'in bu mimariye ilişkin biçimsel çalışmaları, hedef yönelimli davranış üretimi için güçlü bir teorik temel sunmuştur [10]. Bilişsel mimariler; planlama, amaç yönetimi, uzun vadeli tutarlılık ve esneklik gibi yetenekleri desteklediği için karmaşık görevlerde yaygın biçimde kullanılmaktadır. Ayrıca BDI ajanlarının makine öğrenmesi bileşenleriyle birleşerek genelleştirilmiş planlama yetenekleri kazandığını ve diğer ajanların niyetlerini anlama kapasiteleri geliştirdiğini gösteren çalışmalar bulunmaktadır[11].

2.3 Öğrenen Ajanlar ve Takviyeli Öğrenme

Öğrenme temelli ajan mimarileri, ajanların yalnızca statik kurallar veya önceden tanımlanmış mantıksal çerçevelerle değil, deneyim yoluyla geliştirdikleri davranış stratejileriyle çalışmasını mümkün kılar. Takviyeli öğrenme (reinforcement learning – RL), ajanın ödül sinyallerine göre karar politikası geliştirdiği temel yöntemdir. Sutton ve Barto'nun çalışması, bu alanın teorik temelini oluşturmuştur [12]. Son yıllarda derin takviye öğrenme, yüksek boyutlu durum alanlarında güçlü performanslar sergileyen çok ajanlı öğrenme sistemlerinin gelişimini hızlandırmıştır.

Çok ajanlı derin takviyeli öğrenmenin özellikle karmaşık oyun ortamlarında koordinasyon, rol dağılımı ve stratejik iletişim gibi becerileri

belirgin biçimde geliştirdiğini ortaya koymuştur [13]. Heterojen ve değişken gerçek dünya senaryolarında çok ajanlı derin takviyeli öğrenmenin, klasik yöntemlerin baş edemediği konumlandırma ve planlama problemlerinde istikrarlı ve genellenebilir politikalar üretebildiği de gösterilmiştir [14]. Bu paradigma, günümüz LLM tabanlı ajan yapılarıyla birleştirildiğinde önemli sinerjiler ortaya çıkmaktadır.

2.4 LLM Tabanlı ve Karma (Hibrit) Mimariler

Büyük dil modellerinin yükselişi, ajan mimarilerini kökten dönüştüren yeni bir yaklaşımı gündeme getirmiştir. LLM tabanlı ajanlar; doğal dilde akıl yürütme, planlama, araç çağırma, hafıza yönetimi ve çok ajanlı etkileşim gibi işlevleri tek bir çatı altında birleştirebilen yapılardır. Literatürde ajanik LLM'ler üzerine yapılan güncel incelemeler, bu sistemlerin klasik mimarilerden farklı olarak hem sembolik (kural-tabanlı, planlayıcı) hem de öğrenme temelli bileşenleri aynı mimari içinde hibrit biçimde kullanabildiğini göstermektedir [3]. Böylece LLM tabanlı ajanlar, bir yandan sembolik planlayıcılar gibi ayrık görevleri yapılandırabilirken, diğer yandan öğrenme süreçleri sayesinde belirsiz ve yüksek boyutlu ortamlarda bağlamsal kararlar üretebilir. Bu hibrit yaklaşım, modellerin hem esneklik hem bağlamsal uyum hem de dış dünya ile etkili etkileşim kurma açısından güçlü bir performans göstermesini sağlar. Ayrıca araç çağırma ve kullanma yeteneklerinin gelişmesi, LLM tabanlı ajanların; hesaplama, veri analizi, arama, API çağrıları veya robotik kontrol gibi dış görevleri devralarak klasik ajan mimarilerinde ulaşılması güç bir işlevselliğe sahip olmasına olanak tanımaktadır. Bu dönüşüm, ajanik LLM'leri yalnızca bir model türü olmaktan çıkarıp genel amaçlı otonom sistemler için yeni bir mimari paradigma hâline getirmektedir.

3. LLM Ajanlarında Temel İşleyiş; Algılama, Planlama, Karar Verme, Araç Kullanımı ve Öğrenme

Ajan tabanlı yapay zekâ sistemleri, çevreyle döngüsel etkileşimi temel alan bir mimari üzerine inşa edilir. Bu döngü, genel olarak algılama → planlama ve karar verme → eylem (araç kullanımı) → öğrenme süreçlerinden oluşur. Büyük dil modeli tabanlı ajanlar, bu süreçleri doğal dil üzerinden yürütme yeteneğine sahip oldukları için klasik ajanlarla karşılaştırıldığında daha esnek, bağlamsal ve uyarlanabilir davranırlar. Bu bölümde, söz konusu döngünün her bir bileşeni incelenmektedir.

3.1 Algılama ve Bağlam Oluşturma

Algılama, bir ajanın dış dünyayı anlamlandırmak için ihtiyaç duyduğu ilk basamaktır. Geleneksel sistemlerde algılama; sensör verileri, önceden tanımlı durum değişkenleri veya yapılandırılmış girdilerle sınırlıyken, LLM tabanlı ajanlar çok daha geniş bir bağlam yelpazesini işleyebilmektedir. Bu bağlam; doğal dil belgelerini yorumlamayı, kullanıcı komutlarını ayrıştırmayı,

webden veri toplamayı veya başka ajanların ürettiği çıktıları incelemeyi içerebilir. Güncel çalışmalarda LLM ajanlarının ham girdiden durum temsili çıkarma konusunda beklenenden yüksek performans gösterdiği, hatta eksik veya belirsiz bilgi karşısında kendi iç tutarlılıklarını kurabildikleri ortaya konmuştur [15]. Ayrıca algılama sürecinin artık pasif veri alma değil, amaç yönelimli aktif bilgi toplama davranışı içerdiğini bileşenlerinin çevreden gelen geri bildirimleri tekrar tekrar işleyerek gerekli durum güncellemelerini yaptığı görülmektedir ve LLM'lerin çevresel belirsizliğe rağmen daha sağlam durum modelleri oluşturabildiği de ortaya konmaktadır [16].

3.2 Planlama Kapasitesi ve Alt Görev Üretimi

Planlama, bir hedefi oluşturan alt görevlerin belirlenmesi, bu görevlerin sıralanması ve uygun kaynakların tahsisi gibi kararları içerir. LLM tabanlı ajanlar, görev açıklamalarını doğal dilde çözümleyerek kendi alt görev ağlarını oluşturabilmektedir ve gerektiğinde bu ağı dinamik olarak güncelleyebilmektedir. Son dönemde yapılan çalışmalar, araç kullanımına dayalı görev yürütümünde, LLM ajanlarının sadece adımları değil aynı zamanda araçların hazırlanması, bağımlılıkların yönetimi ve araçların yürütüm sıralarını kısmen planlayabildiğini göstermektedir. Ayrıca LLM destekli planlama süreçlerinin klasik planlayıcı yapılarından daha bağlama duyarlı olduğu ve belirsiz görev akışlarında kendi kendine yeniden planlama davranışı sergileyebildiği ortaya konmuştur [17].

3.3 Karar Verme Mekanizmaları

Karar verme, bir ajanın çevresine ilişkin değerlendirmeyi eyleme dönüştürdüğü temel aşamadır. LLM tabanlı ajanlarda bu süreç, yalnızca seçenekler arasından seçim yapmakla sınırlı değildir; aksine, doğal dil üzerinden yürütülen çok adımlı bir akıl yürütme zincirine dayanır. Ajan, durumu analiz eder, olası yolları karşılaştırır, gerektiğinde hipotezler kurar ve bunları mantıksal çıkarımlarla sınar. Güncel çalışmalarda, bu yaklaşımın özellikle belirsizliğin yüksek olduğu senaryolarda daha tutarlı kararlar ürettiği ve karar gerekçesinin doğal dil üzerinden açıklanabildiği gösterilmektedir [18].

Bazı araştırmalar gösteriyor ki, LLM tabanlı ajanların karar süreci birçok ara adımdan oluştuğundan tek başına modelin içsel çıkarım gücüne bırakıldığında kararsızlık ve hatalar ortaya çıkabiliyor. Özellikle araç seçimi, eksik parametrelerin tespiti ve API bağımlılıklarını çözme gibi aşamalardaki zorluklar, karar hızını ve istikrarını artırmak için ek kontrol ve yönlendirme bileşenlerine ihtiyaç olduğunu işaret ediyor. Bu bulgular, çok adımlı görevlerde daha yapılandırılmış destek mekanizmalarının ajanın performansını belirgin biçimde iyileştirdiğini ortaya koymaktadır [19].

3.4 Araç Kullanımı ve Eylem Üretimi

Ajanın dış dünyada somut bir etki yaratabilmesi, araç kullanımı yeteneğiyle doğrudan ilişkilidir. LLM tabanlı ajanlar artık yalnızca dili işleyen sistemler değil; API çağrıları gerçekleştirebilen, kod üretebilen ve çalıştırabilen, veri tabanlarına bağlanabilen ya da fiziksel aygıtları kontrol edebilen çok yönlü yapılardır. Son çalışmalarda, bu ajanların bir görevin hangi aşamasında araç kullanımına ihtiyaç duyduğunu kendi akıl yürütme süreçleriyle belirleyebildiği ve araç çağrılarını planlama döngüsüyle bütünleştirebildiği gösterilmektedir [16]. Böylece araç kullanımı, dışardan bir ekleme olmaktan çıkıp ajanın karar ve planlama sisteminin doğal bir parçasına dönüşmektedir.

LLM-tabanlı ajanların araç kullanım kapasitesinin yalnızca mevcut araçları doğru biçimde çağırmakla sınırlı olmadığı; giderek daha gerçekçi, dinamik ve çok adımlı senaryolarda araç etkileşimini yönetebilecek şekilde genişlediği görülmektedir. Son dönem değerlendirme çalışmalarında, ajanların karmaşık araç zincirlerini planlaması, durum bilgisi tutan araçlarla etkileşmesi ve birbirine gömülü fonksiyon çağrılarını kullanması gibi becerilerin önem kazandığı vurgulanmaktadır [20]. Bu yönelim, ajanik AI sistemlerini basit görev uygulayıcılarının ötesine geçirerek, sürekli genişleyen bir araç ekosistemi içinde daha etkin karar vericiler ve süreç yöneticileri hâline getirmektedir. Böyle bir evrim, özellikle çok adımlı veya belirsizlik içeren problem çözme senaryolarında ajanın otonom kapasitesini belirgin biçimde artırmaktadır.

3.5 Öğrenme ve Uyarlama

Öğrenme, döngüde ajanın kendisini iyileştirdiği ve davranışlarını uzun vadede biçimlendirdiği aşamadır. Geleneksel sistemlerde öğrenme çoğunlukla sayısal ödül sinyallerine dayanırken, LLM tabanlı ajanlar çok daha zengin bir bilgi kaynağından yararlanabilir: doğal dilde verilen geri bildirimler, araç kullanım sonuçları, çevresel hatalar, bellek kayıtları ve kullanıcı etkileşimleri. Güncel literatür, bu tür çok biçimli geri bildirimlerin ajanların davranış politikalarını daha hızlı ve daha istikrarlı şekilde güncelleyebilmesini sağladığını ortaya koymaktadır [21].

Ayrıca bazı çalışmalarda, LLM tabanlı ajanların görev bağlamı değiştiğinde kendi stratejilerini yeniden düzenleyebildiğini, yani görev ihtiyaçlarına göre kendini uyarlayabilen yapılar hâline geldiğini belirtilmektedir [22]. Bu özerk uyarlanabilirlik, belirsiz ve değişken ortamlarda uzun vadeli otonom davranışın temel koşulu olarak değerlendirilmektedir. Böylelikle öğrenme yalnızca geçmiş deneyimlerin biriktirildiği bir aşama değil; ajanın kendi kendini yeniden yapılandırabildiği dinamik bir süreç hâline gelmektedir.

4. Çok Ajanlı Sistemler ve Koordinasyon

4.1 Ajanlar Arası Etkileşim ve Koordinasyon Temelleri

Çok ajanlı sistemler, birden fazla ajanın aynı ortamda etkileşim kurduğu durumları temsil eder. Bu tür ortamlarda iş birliği, rekabet, koordinasyon ve görev paylaşımı gibi davranış kalıpları ortaya çıkar. Ajanlar, kendi hedeflerini gözetirken aynı zamanda diğer ajanların eylemlerinden de etkilenir; dolayısıyla karar süreçleri hem bireysel hem de kolektif dinamiklerle şekillenir.

Klasik çok ajanlı yapay zekâ literatürü, bu tür sistemlerde bilgi paylaşımı, görev dağıtımı ve ortak karar verme gibi mekanizmaların sistem başarısı üzerinde belirleyici olduğunu göstermektedir. Özellikle etkileşimin yoğun olduğu ortamlarda ajanların yalnızca duruma tepki vermediği; diğer ajanların niyetlerini, stratejilerini ve olası eylemlerini modellemeye çalıştığı da ortaya konmaktadır. Bu durum, çok ajanlı yapay zekâda koordinasyonun yalnızca mekanik bir görev paylaşımından ibaret olmadığını, aynı zamanda karşılıklı beklentilerin yönetildiği bir akıl yürütme süreci olduğunu göstermektedir. Ayrıca son dönem çalışmalar, çok ajanlı sistemlerde iletişim kanallarının kalitesinin ve zamanlamasının, kolektif başarı üzerinde kritik bir etkisi olduğunu vurgulamaktadır [23]. Bu bulgular, modern agentic yapay zekâ sistemlerinde koordinasyonun yalnızca teknik bir gereklilik değil, aynı zamanda sosyal benzeri bir etkileşim biçimi olduğunu göstermektedir.

4.2 LLM Tabanlı Çok Ajanlı Sistemler ve Ortak Akıl Yürütme

Büyük dil modellerinin çok ajanlı sistemlere entegre edilmesi, koordinasyon kavramını önemli ölçüde dönüştürmüştür. LLM tabanlı ajanlar, doğal dil üzerinden iletişim kurabildikleri için hem komutları hem de birbirlerinin ürettiği mesajları bağlamsal olarak yorumlayabilir. Bu kapasite, çok ajanlı ortamlarda daha zengin bir iş birliği zemini oluşturur.

LLM tabanlı çok ajanlı yapılarda iletişimin yalnızca bilgi paylaşımı değil, aynı zamanda ortak planın oluşumunda aktif bir rol oynadığı gösterilmiştir. Ajanlar birbirlerine hedefler, ara sonuçlar veya olası hatalar hakkında geri bildirim verebilir; bu geri bildirimler de kolektif karar sürecinin bir parçası hâline gelir [24]. Bu durum, daha önce katı mesajlaşma protokollerine bağlı olan çok ajanlı sistemlerin yerini, esnek ve duruma göre şekillenen iletişim yapılarının almasına yol açmaktadır.

Bu gelişmelerin bir diğer önemli yönü ise, LLM tabanlı ajanların ortak akıl yürütme davranışları sergilemesidir. Bazı çalışmalar, birden fazla LLM ajanının birlikte çalıştığında tek başına bir modelin ulaşamayacağı çözüm yolları üretebildiğini, karmaşık görevlerde planı birlikte şekillendirebildiğini ve gerektiğinde birbirinin hatasını düzeltebildiğini ortaya koymaktadır [25].

Bu tür ortak akıl yürütme mekanizmaları, özellikle simülasyon ortamlarında insan benzeri sosyal davranışların, iş birliğinin ve rol dağılımının

ortaya çıktığını göstermektedir. Böylece çok ajanlı LLM sistemleri yalnızca geleneksel olarak dili yorumlayabilen bir versiyonu olmayıp, iletişim-temelli kolektif düşünme modeli olarak değerlendirilmeye başlanmıştır.

5. Gerçek Dünya Uygulamaları, Sınırlılıklar ve Gelecek Yönelimleri

5.1 Gerçek Dünya Uygulamaları

Ajan tabanlı yapay zekâ sistemleri, günümüzde farklı alanlarda somut yararlar sağlayan çözümler üretmektedir. Yazılım geliştirme süreçlerinde LLM tabanlı ajanlar; kod önerme, hata analiz etme, test senaryosu oluşturma ve bileşenler arası etkileşimleri otomatikleştirme gibi görevlerde etkin biçimde kullanılmaktadır. Çoklu ajan yapılarının yazılım mühendisliği görevlerini iş bölümüyle çözebildiğini, özellikle karmaşık geliştirme döngülerinde üretkenliği artırdığını görülmektedir [26]. Endüstriyel uygulamalarda bu yaklaşımın bir yansıması olarak UiPath Agent Builder, bir görevin farklı adımlarını üstlenen dijital çalışma arkadaşları oluşturarak teknik ekiplerin iş yükünü azaltmakta [27]; Atera'nın BT (Bilgi Teknolojileri) destek ajanları ise gelen talepleri sınıflandıran, sorun analizi yapan ve belirli işlemleri otonom biçimde uygulayabilen yapılar sunmaktadır [28]. AutoGPT ise kullanıcı isteğini hedefe yönelik bir plana dönüştürerek alt-görevler oluşturan, bu alt-görevleri yürütmek için araç çağırma, web arama ve kod çalıştırma gibi eylemleri otonom biçimde gerçekleştirebilen, açık kaynaklı bir LLM tabanlı ajan mimarisi sağlamaktadır [29].

Buna ek olarak yazılım geliştirme ekosisteminde Microsoft'un sunduğu GitHub Copilot ve bu platformun genişletilmiş sürümü olan Copilot Agents, geliştiricilerin karşılaştığı daha kapsamlı sorunları çözebilmek için görev odaklı ajan mantığını benimsemektedir [30]. Copilot Agents, yalnızca kod tamamlama sağlamanın ötesine geçerek; hata ayıklama, test üretimi, sistem bileşenlerinin yeniden düzenlenmesi, API entegrasyonlarının planlanması ve çok adımlı geliştirme görevlerinin yürütülmesi gibi işlemleri adım adım planlayan ve gerektiğinde araç çağırabilen bir yapı sunmaktadır. Bunun yanında, Microsoft Developer Copilot ekosistemi, geliştirme ortamında doğal dil ile etkileşim kurarak görevleri açıklayan, yapılandıran ve yürütmeye yardımcı olan yardımcı ajanlar içermekte; böylece geliştiricinin bağlam değiştirmeden karmaşık işleri tamamlamasını kolaylaştırmaktadır. Bu tür uygulamalar, LLM tabanlı ajanların yazılım geliştirmede sadece kod öneren sistemler olmaktan çıkıp, geliştirme sürecinin bütününde aktif rol alan otonom çalışma arkadaşları hâline dönüştüğünü göstermektedir.

Bilimsel araştırma alanında ajan sistemleri; veri hazırlama, hipotez geliştirme, deney tasarlama ve sonuç yorumlama gibi aşamalarda araştırmacılara yardımcı olmaya başlamıştır. Özellikle geniş veri gerektiren biyoloji, kimya ve malzeme bilimi gibi disiplinlerde, çok ajanlı çerçeveler araştırmacıların yerine veri taraması yapabilmekte, anlamlı örüntüler çıkarabilmekte ve araştırma

sorularını biçimlendirmelerine katkı sağlayabilmektedir [31]. Bu bağlamda CAS tarafından geliştirilen bilimsel araştırma ajanları, literatür taramasından moleküler yapı analizine kadar birçok adımı hızlandırmakta; araştırmacının odaklanması gereken bölümleri filtreleyerek verimliliği artırmaktadır [32]. Benzer şekilde bilim insanı ajanları yaklaşımı, veri ön işleme, model deneme ve sonuç raporlama gibi görevleri birden fazla ajana bölerek bilimsel sürecin toplam süresini kısaltmaktadır. ChemCrow, kimya alanında uzmanlaşmış bu çoklu ajan sistemi, sadece molekül tasarlamakla kalmayıp, laboratuvardaki robotik kolları yöneterek fiziksel deneyleri (sentezleme) otonom olarak gerçekleştirebilmektedir [33]. İlaç keşfi ve malzeme bilimi süreçlerinde, teorik tasarımdan fiziksel üretime kadar olan süreci insan müdahalesi olmadan yöneterek laboratuvar güvenliğini ve hızını artırmaktadır. Scite.ai, metinleri analiz eden akıllı alıntı ajanları kullanarak, bir makalenin başka makaleler tarafından sadece bahsedildiğini mi, desteklendiğini mi yoksa çürütüldüğünü mü (zıtlık) tespit etmektedir [34]. Araştırmacıların yanlış veya geçerliliğini yitirmiş bilgilere dayanarak hipotez kurmasını engelleyerek bilimsel doğruluk sürecini güvence altına almaktadır.

Bu gelişmelerin yanında lojistik, sağlık teknolojileri ve pazar analizi gibi sektörlerde de ajan tabanlı yapılar yaygın biçimde kullanılmaktadır. Örneğin tedarik zinciri yönetiminde Blue Yonder'ın stok ve lojistik ajanları, mağaza ve depo verilerini gerçek zamanlı olarak işleyerek yeniden sipariş kararlarını optimize etmekte ve teslimat gecikmelerini önceden öngörebilmektedir [35]. Sağlık alanında Ajentik tarafından geliştirilen klinik değerlendirme ajanları, hastadan gelen metin veya sesli verileri ön incelemeden geçirerek sağlık uzmanlarının karar sürecini desteklemektedir [36]. Benzer biçimde Hippocratic AI, sağlık sistemindeki personel açığını kapatmak amacıyla, hemşire veya diyetisyen gibi spesifik roller üstlenen ve "önce güvenlik" prensibiyle eğitilmiş, uzman sanal sağlık çalışanları geliştirir [37]. Bu sistem ajan tabanlı yapay zekayı; hastaları proaktif olarak arayıp sesli iletişim kuran, ilaç takiplerini yapan ve empatik bir dille semptomları izleyerek bakım sürecini otonom yöneten dijital personeller olarak kullanır. Ajanlar sadece hastayla konuşmakla kalmaz, hastanın durumunda bir risk artışı veya anormallik tespit ettiği anda inisiyatif alarak konuyu acilen gerçek bir doktora yönlendirir ve böylece güvenli bir karar destek mekanizması sağlar. OneSky'nin yerelleştirme ajanları, çok dilli içerik uyarlaması gerektiren uygulamalarda metin çözümleme, kalite kontrol ve tutarlılık denetimi gibi süreçleri kısaltarak üretim verimini artırmaktadır [38]. Phrase (Phrase Orchestrator), yazılım kod depolarını sürekli izleyip yeni eklenen metinleri otomatik tespit ederek çeviri sürecini insan müdahalesi olmadan başlatmaktadır [39]. Bu sayede yazılımcıların manuel dosya transferi yapmasına gerek kalmadan, kod güncellemesiyle eş zamanlı işleyen kesintisiz bir yerelleştirme sağlanmaktadır. Lilt'in bağlamsal öğrenme ajanları ise çevirmenin çalışma sırasında yaptığı

düzeltmeleri anlık olarak öğrenip belgenin geri kalanındaki benzer cümleleri o saniyede yeniden üretmektedir [40]. Bu adaptif teknoloji, proje ilerledikçe yapay zekanın doğruluğunu artırarak çevirmenin iş yükünü dinamik olarak azaltmaktadır.

5.2 Sınırlılıklar, Riskler ve Güvenlik Boyutları

Ajan tabanlı yapay zekâ sistemlerinin hızla yaygınlaşması, beraberinde önemli sınırlılıkları ve riskleri de gündeme getirmektedir. Bu sistemlerde en belirgin sorunlardan biri, LLM tabanlı ajanların hâlâ halüsinasyon üretme eğiliminde olmasıdır. Ajanların kendi çıktıları üzerinden zincirleme şekilde hatalar üretmesi, çok adımlı görevlerde kontrolsüz davranışlara yol açabilmektedir. Yakın dönem çalışmalar, çok ajanlı ortamlarda bir ajanın ürettiği hatanın diğer ajanlara hızla aktarılabildiğini ve hatanın toplu bir davranış bozulmasına dönüştüğünü göstermektedir [25].

Diğer bir önemli sınırlılık ise güvenliktir. Ajanların araç kullanma yetenekleri, API erişimi ve harici sistemlerle etkileşimi; yanlış yönlendirilmiş ya da hatalı planlanmış eylemlerin gerçek sistemler üzerinde risk oluşturmasına neden olabilir. Buna ek olarak gizlilik, veri bütünlüğü ve yetkisiz işlem riskleri, ajan tabanlı sistemlerin kurumsal ortamlarda kullanımını dikkatle yönetilmesi gereken bir konu hâline getirmektedir. Son olarak, uzun görevlerde bağlam kaybı, bellek tutarsızlığı ve ajanın amaç sapması gibi problemler; agentic sistemlerin tam özerklik seviyesine ulaşmasında önemli engeller olarak değerlendirilmektedir [41].

5.3 Gelecek Araştırma Yönelimleri

Gelecek araştırmaların odak noktasını, daha güvenilir, açıklanabilir ve sürdürülebilir agentic yapay zekâ sistemleri oluşturacaktır. İlk olarak, uzun vadeli bellek mekanizmalarının geliştirilmesi ve bağlamı tutarlı şekilde koruyan yapısal çözümler kritik öneme sahiptir. Araştırmalar, LLM tabanlı ajanların bellek temsillerini daha etkili yönetmesi hâlinde karmaşık görevlerde istikrarlı davranışlar sergileyebileceğini göstermektedir. İkinci olarak, çok ajanlı ortamlarda kolektif akıl yürütme ve güvenli koordinasyon konuları önemli bir araştırma alanı olarak öne çıkmaktadır.

Ayrıca gelecekte ajanların insanlarla daha doğal etkileşim kurabilmesi, görev hedeflerini açıklayabilmesi ve kendi karar süreçlerine ilişkin gerekçeleri sunabilmesi beklenmektedir. Bu tür açıklanabilirlik odaklı yaklaşımlar, agentic sistemlerin güvenliğini ve kabul edilebilirliğini artıracaktır. Son olarak, otonom araç kullanımı, bilimsel keşif ve üretim otomasyonu gibi alanlarda ajanların görev planlama ve araç kullanma kapasitelerinin daha sistematik biçimde değerlendirilmesi, agentic AI'nin güvenilir biçimde ölçeklenebilmesi için temel bir araştırma başlığıdır.

KAYNAKÇA

[1]Russell, S. J., & Norvig, P. (2022). Artificial intelligence: A modern approach (4th ed., Global Edition). Pearson Education Limited.

[2]Wooldridge, M. (2009). An introduction to multiagent systems (2nd ed.). John Wiley & Sons.

[3] Plaat, A., van Duijn, M., van Stein, N., Preuss, M., van der Putten, P., & Batenburg, K. J. (2025). Agentic large language models: A survey. arXiv preprint, arXiv:2503.23037.

[4]Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., Chen, Z., Tang, J., Chen, X., Lin, Y., Zhao, W. X., Wei, Z., & Wen, J.-R. (2023). A survey on large language model based autonomous agents. Frontiers of Computer Science, 18(6), 186345. (Ön baskı için arXiv:2308.11432)

[5]Krishnan, N. (2025). AI agents: Evolution, architecture, and real-world applications (arXiv:2503.12687v1). arXiv. https://arxiv.org/abs/2503.12687

[6] Franklin, S., & Graesser, A. (1996). Is it an Agent, or Just a Program? In Intelligent Agents III (pp. 21–35). Springer.

[7] Nguyen, T. T., Nguyen, N. D., & Nahavandi, S. (2020). Deep Reinforcement Learning for Multi-Agent Systems: A Review of Challenges, Solutions and Applications. IEEE Transactions on Cybernetics.

[8]Brooks, R. A. (1986). A robust layered control system for a mobile robot. IEEE Journal on Robotics and Automation, 2(1), 14–23.

[9]Pomarlan, M., De Giorgis, S., Ringe, R., Hedblom, M. M., & Tsiogkas, N. (2025). Hanging around: Cognitive inspired reasoning for reactive robotics (arXiv:2507.20832 [cs.RO]). arXiv. https://arxiv.org/abs/2507.20832

[10]Rao, A. S., & Georgeff, M. P. (1995). BDI agents: From theory to practice. In Proceedings of the First International Conference on Multiagent Systems (ICMAS-95).

[11]Pereira, R. F., & Meneguzzi, F. (2023). Empowering BDI agents with generalised decision-making: Blue sky ideas track. In Proceedings of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023) (pp. 1–5).

[12]Sutton, R. S., & Barto, A. G. (2015). Reinforcement learning: An introduction (2nd ed.). A Bradford Book, The MIT Press.

[13]Lyu, Y., Tang, Z., Li, Y., Liu, B., Deng, M., & Wu, G. (2025). A multi-agent deep reinforcement learning method with diversified policies for continuous location of express delivery stations under heterogeneous scenarios. ISPRS International Journal of Geo-Information, 14(12), 461. https://doi.org/10.3390/ijgi14120461

[14]Wang, Y., Wang, Y., Tian, F., Ma, J., & Jin, Q. (2025). Intelligent games meeting with multi-agent deep reinforcement learning: A comprehensive review. Artificial Intelligence Review, 58, Article 165. https://doi.org/10.1007/s10462-025-11166-1

[15]Li, Z., Li, Y., Ye, H., & Zhang, Y. (2024). Towards autonomous tool utilization in language models: A unified, efficient and scalable framework. In Proceedings of LREC–COLING 2024 (pp. 16422–16432). ELRA.

[16]Xu, W., Huang, C., Gao, S., Shang, S., & others. (2025). LLM-Based Agents for Tool Learning: A Survey. Data Science and Engineering. https://doi.org/10.1007/s41019-025-00296-9

[17]Zhao, Y., et al. (2025). ToolMaker: Autonomous tool creation from scientific code repositories. arXiv:2502.11705. https://arxiv.org/abs/2502.11705

[18]Li, X. (2025). A review of prominent paradigms for LLM-based agents: Tool use (including RAG), planning, and feedback learning. Proceedings of the 31st International Conference on Computational Linguistics, 9760–9779.

[19]Liu, B., Zhou, J., Meng, D., & Lu, H. (2024). An evaluation mechanism of LLM-based agents on manipulating APIs. In Findings of the Association for Computational Linguistics: EMNLP 2024 (pp. 4649–4662).

[20]Yehudai, A., Eden, L., Li, A., Uziel, G., Zhao, Y., Bar-Haim, R., Cohan, A., & Shmueli-Scheuer, M. (2025). Survey on evaluation of LLM-based agents (arXiv:2503.16416). arXiv. https://arxiv.org/abs/2503.16416

[21]Yuksel, K. A., Castro Ferreira, T., Al-Badrashiny, M., & Sawaf, H. (2025). A multi-AI agent system for autonomous optimization of agentic AI solutions via iterative refinement and LLM-driven feedback loops. In Proceedings of the 1st Workshop for Research on Agent Language Models (REALM 2025) (pp. 52–62). Association for Computational Linguistics.

[22]Wang, K., Lu, Y., Santacroce, M., Gong, Y., Zhang, C., & Shen, Y. (2025). Adapting LLM agents with universal communication feedback. In Findings of the Association for Computational Linguistics: NAACL 2025 (pp. 6105–6122). Association for Computational Linguistics.

[23]Li, X., Wang, S., Zeng, S., Wu, Y., & Yang, Y. (2024). A survey on LLM-based multi-agent systems: Workflow, infrastructure, and challenges. Vicinagearth, 1, Article 9. https://doi.org/10.1007/s44336-024-00009-2

[24]Zhu, C., Dastani, M., & Wang, S. (2024). A survey of multi-agent deep reinforcement learning with communication. Autonomous Agents and Multi-Agent Systems, 38(4). https://doi.org/10.1007/s10458-023-09633-6

[25]Agashe, S., Fan, Y., Reyna, A., & Wang, X. E. (2025). LLM-Coordination: Evaluating and Analyzing Multi-agent Coordination Abilities in Large Language Models. Findings of ACL: NAACL 2025, 8038–8057. https://aclanthology.org/2025.findings-naacl.448

[26]Chen, W., Su, Y., Zuo, J., Yang, C., Yuan, C., Chan, C.-M., Yu, H., Lu, Y., Hung, Y.-H., Qian, C., Qin, Y., Cong, X., Xie, R., Liu, Z., Sun, M., & Zhou, J. (2023). AgentVerse: Facilitating multi-agent collaboration and exploring emergent behaviors (arXiv:2308.10848). arXiv. https://arxiv.org/abs/2308.10848

[27] UiPath. (2024). UiPath introduces generative AI and specialized AI capabilities to accelerate enterprise automation. https://www.uipath.com/newsroom

[28] Atera. (2024). Atera Copilot: The AI-powered IT assistant for smarter IT management. https://www.atera.com/features/ai-copilot

[29] Richards, T. (2023). Auto-GPT: An experimental open-source attempt to make GPT-4 fully autonomous. GitHub. https://github.com/Significant-Gravitas/Auto-GPT

[30] Microsoft. (2024). GitHub Copilot extensions and agents: Expanding the ecosystem. Microsoft Developer Blog. https://devblogs.microsoft.com

[31]Toledo, E., Hambardzumyan, K., Josifoski, M., Hazra, R., Baldwin, N., Audran-Reiss, A., Kuchnik, M., Magka, D., Jiang, M., Lupidi, A. M., Lupu, A., Raileanu, R., Niu, K., Shavrina, T., Gagnon-Audet, J.-C., Shvartsman, M., Sodhani, S., Miller, A. H., Charnalia, A., Dunfield, D., Wu, C.-J., Stenetorp, P., Cancedda, N., Foerster, J. N., & Bachrach, Y. (2025). AI Research Agents for Machine Learning: Search, Exploration, and Generalization in MLE-bench (Version 1). arXiv. https://arxiv.org/abs/2507.02554

[32] CAS. (2024). CAS Scientific Patent Explorer and AI-driven research solutions. Chemical Abstracts Service. https://www.cas.org/solutions

[33] Bran, A. M., Cox, S., White, A. D., & Schwaller, P. (2023). ChemCrow: Augmenting large-language models with chemistry tools. arXiv preprint arXiv:2304.05376.

[34] Nicholson, J. M., Mordaunt, M., Lopez, P., Uppala, A., Rosati, D., Rodrigues, N. P., ... & Rife, S. C. (2021). scite: A smart citation index that displays the context of citations and classifies their intent using deep learning. Quantitative Science Studies, 2(3), 882-898.

[35] Blue Yonder. (2024). Blue Yonder presents generic AI and agentic supply chain solutions. https://blueyonder.com

[36] Ajentik. (n.d.). Ajentik healthcare triage and clinical assistant platform. [Software].

[37] Hippocratic AI. (2023). Polaris: A safety-focused large language model for healthcare. https://www.hippocraticai.com

[38] OneSky. (n.d.). Automated localization workflow and translation management. https://www.oneskyapp.com

[39] Phrase. (2023). Phrase Orchestrator: Automating localization workflows with AI. https://phrase.com/products/orchestrator

[40] Lilt. (2023). Contextual AI for enterprise translation and localization. https://lilt.com/technology

[41]Han, S., Zhang, Q., Yao, Y., Jin, W., & Xu, Z. (2025). LLM multi-agent systems: Challenges and open problems (Version 2). arXiv. https://arxiv.org/abs/2402.03578

Bölüm 4

YAPAY ZEKA TABANLI İÇERİK ÜRETİMİNDE KULLANILAN PERFORMANS METRİKLERİ: DİLSEL, ANLAMSAL VE UYGULAMALI YAKLAŞIMLAR

Fatih KORKMAZ [1]
Mustafa KARHAN [2]
Nazlı KOYMAT [3]

1 Doç. Dr. Çankırı Karatekin Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik Mühendisliği Bölümü, Çankırı, Türkiye. fkorkmaz@karatekin.edu.tr, ORCID ID: 0000-0001-8524-2831
2 Doç. Dr. Gaziantep İslam Bilim ve Teknoloji Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi, Elektrik – Elektronik Mühendisliği Bölümü, Gaziantep, Türkiye. mustafa.karhan@gibtu.edu.tr, ORCID ID: 0000-0001-6747-8971
3 Çankırı Karatekin Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik Mühendisliği Bölümü, Çankırı, Türkiye. 248414010@ogrenci.karatekin.edu.tr, ORCID ID: 0000-0003-2573-2315

Giriş

Yapay zeka destekli içerik üretimi, Transformer mimarisinin yükselişi ve büyük dil modellerinin (Large Language Models - LLMs) ortaya çıkışıyla birlikte, doğal dil işleme (NLP) alanında en hızlı gelişen alanlardan biri haline gelmiştir (Vaswani vd., 2017; Bommasani vd., 2021). Bu sistemlerin kapasitesi artarken ürettikleri metinlerin kalitesinin ve güvenilirliğinin nasıl ölçüleceği sorusu da giderek kritik bir önem kazanmaktadır (Chang vd., 2024). Performans metrikleri, AI modellerinin üretim yeteneğini niceliksel olarak değerlendirmenin, farklı modeller ile teknikler arasında karşılaştırma yapmanın ve nihayetinde kullanıcı güvenini tesis etmenin temel aracıdır (Celikyilmaz vd., 2020).

Bir metin aynı anda dilbilgisi, anlamsal tutarlılık, bilgi doğruluğu, görev performansı ve estetik nitelikler açısından değerlendirilebilir. Bu nedenle araştırma ve uygulama literatüründe performans metrikleri genellikle üç adet birbirini tamamlayan kategori altında incelenir. Bunlar dilsel, anlamsal ve uygulamalı metrikler olarak adlandırılır (Liu vd., 2023). Dilsel metrikler, bir metnin dilbilgisel formunun ve yüzeysel akışının ne derecede insan üretimi metinlere benzediğini ölçer. Anlamsal metrikler, metnin yüzeysel formunun ötesine geçerek, içeriğin anlamsal yoğunluğunu, mantıksal bütünlüğünü ve gerçek dünya bilgisiyle uyumunu değerlendirir. Uygulamalı metrikler ise belirli bir görevi ne kadar iyi yerine getirdiğini, kullanıcılar tarafından nasıl algılandığını (insan değerlendirmesi) sorgular. Bu üçlü metrik yaklaşımı, AI sistemlerinin teknik kapasitesi ve pratik değerini bütüncül bir bakış açısıyla ölçmeyi mümkün kılar.

2. Yapay Zeka İle İçerik Üretimi

Yapay zeka destekli içerik üretiminin teknik altyapısı, Transformer mimarisi üzerine kurulu büyük dil modelleri etrafında şekillenmektedir. (Vaswani vd., 2017). Bu modeller, temelde çevrimiçi kaynaklardan derlenmiş geniş metin veri kümeleri üzerinde gerçekleştirilen ön-eğitim süreciyle dilin istatistiksel kalıplarını, sözdizimsel yapısını ve anlamsal ilişkilerini benimser (Devlin vd., 2019; Brown vd., 2020). Bu genel dil becerisi elde edildikten sonra, modeller belirli içerik üretim görevlerine yönelik olarak çeşitli stratejilerle uyarlanır. Bu uyarlama süreci, modelin genel dil bilgisini korurken, hedef görevin gerektirdiği üslup, biçim ve terminolojiye odaklanmasını sağlamak açısından kritik öneme sahiptir (Raffel vd., 2020; Hu vd., 2021).

2.1 Temel Modeller

Günümüzdeki temel model yaklaşımları, mimari odaklarına ve eğitim paradigmalarına göre çeşitlilik gösterir. Otoregresif modeller olarak da

bilinen ve yalnızca çözümleyici (decoder-only) bileşenden oluşan yapılar, bir metni soldan sağa sırayla üretmek üzere tasarlanmıştır (Radford vd., 2018). Bu modeller verilen bir bağlamdan sonra gelecek en olası kelimeyi tahmin etme prensibiyle çalışır ve açık uçlu metin oluşturma, hikaye yazma veya kod tamamlama gibi görevlerde yaygın olarak kullanılır. GPT serisi bu yaklaşımın önemli bir temsilcisidir (Brown vd., 2020). Diğer yandan, kodlayıcı-çözümleyici (encoder-decoder) mimarisi iki aşamalı bir işleyiş yolu izlemiştir. Kodlayıcı, kaynak metni (örneğin, özetlenecek bir makale) anlamsal olarak kodlar, çözümleyici ise bu kodlanmış temsilden hedef metni (örneğin, özeti) adım adım üretir. Bu yapı metin özetleme, makine çevirisi ve soru-cevap gibi kaynak ve hedef metin arasında dönüşüm gerektiren görevler için daha uygundur (Raffel vd., 2020). T5 ve BART gibi modeller bu kategoride yer alır. Bir diğer kategoride, yalnızca kodlayıcı (encoder-only) bileşene sahip modeller bulunur. Bunlar temelde metin anlamaya odaklanır ve doğrudan metin üretmekten ziyade sınıflandırma veya analiz görevlerinde kullanılır. Bazı durumlarda ise üretim sistemlerinin temel bileşeni olarak hizmet edebilirler (Devlin vd., 2019).

Model büyüklüğü de metin üretimi için önemli bir seçimdir. Yüz milyarlarca parametreye sahip büyük dil modelleri, çok çeşitli görevlerde esnek performans sergiler. Daha küçük ve verimli modeller ise kaynak kısıtlı ortamlarda veya belirli bir alana hızlıca uyarlanmak için tercih edilebilir (Bommasani vd., 2021). Modelin ön-eğitim aşamasından sonra pratik uygulamalara entegrasyonu için uyarlama aşaması gelir. Geleneksel tam ince ayar yöntemi, tüm model parametrelerinin hedef göreve özgü verilerle güncellenmesini içerir, ancak bu yüksek hesaplama maliyeti gerektirir. Bu maliyeti düşürmek ve verimliliği artırmak amacıyla, parametre-verimli ince ayar teknikleri geliştirilmiştir (Houlsby vd., 2019). Bu tekniklerden biri olan Düşük Dereceli Uyarlama (LoRA), modelin ana ağırlıklarını dondurarak yalnızca modele eklenen küçük, düşük dereceli matrisleri eğitir (Hu vd., 2021). Bu sayede, büyük modellerin tek bir grafik işlemcisinde bile hızlıca özelleştirilmesi mümkün hale gelir.

Son dönemde genel amaçlı modellerin yanı sıra, belirli alanlarda uzmanlaşmış model türleri de öne çıkmaktadır. Kod üretimi, diyalog sistemleri ve çok modlu (metin+görsel) içerik üretimi gibi özel görevler için optimize edilmiş modeller, kendi veri kümeleri ve eğitim hedefleri doğrultusunda geliştirilmektedir (Chen vd., 2021).Ayrıca şeffaflığı artıran açık kaynaklı model aileleri (ör. LLaMA, Falcon, Mistral), akademik çalışmaların ve özelleştirilmiş uygulamaların temelini oluşturmaktadır (Touvron vd., 2023). Bu model çeşitliliği, mevcut kaynaklar ve istenen çıktı nitelikleri göz önünde bulundurularak bilinçli bir model seçimini ve ardından bu modellerin dilsel, anlamsal ve uygulamalı performans metrikleri ile kapsamlı değerlendirilmesini zorunlu kılmaktadır (Chang vd., 2024).

2.2 Üretilen İçerik Türleri

AI modellerinin üretebildiği içerik türlerindeki çeşitlilik, performans değerlendirme kriterlerinin de çeşitlenmesini zorunlu kılmaktadır (Gehrmann vd., 2021).

- **Açık Uçlu Üretim (Open-Ended Generation):** Hikaye veya makale gibi başlangıç metni verilerek devamı yazılan üretim türüdür. Burada dilsel akıcılık, anlamsal tutarlılık ve yaratıcılık önemli ölçütlerdir.

- **Koşullu Metin Üretimi (Conditional Text Generation):** Başlık veya anahtar kelimeler gibi belirtilen bir koşul altında metin oluşturan üretim türüdür. Koşullu metin üretiminde çıktının koşula uygunluğu (relevance) ve bağlamsal doğruluğu kritiktir.

- **Metin Özetleme (Text Summarization):** Uzun bir metin belgesinin içeriğindeki en önemli bilgileri koruyarak, metni daha kısa bir versiyonuna dönüştürme işlemidir. Temel amaç, anlam kaybı yaşamadan bilgi yoğunluğunu artırmak ve okuyucunun içeriği hızlıca kavramasını sağlamaktır. Üretilen bir özetin başarısı bilgi sıkıştırma (compression), önemli noktaları koruma (saliency retention) ve olgusal tutarlılık (factual consistency) açısından değerlendirilmesi gerekir (Kryściński vd., 2020).

- **Metin Yeniden Yazma (Text Rewriting/Paraphrasing):** Bir cümlenin anlamını koruyarak farklı kelimelerle ifade edilmesi türüdür. Anlamsal denklik (semantic equivalence) ve dilsel çeşitlilik (linguistic variation) ölçülür.

- **Kod Üretimi (Code Generation):** Doğal dilde ifade edilen kullanıcı komutlarının veya problem tanımlarının bilgisayar tarafından yorumlanarak otomatik olarak programlama koduna dönüştürülmesidir. Üretilen kodun işlevsel doğruluğu (functional correctness), sentaks hatası içermemesi (syntactic validity) ve verimliliği (efficiency) değerlendirme kriteridir (Chen vd., 2021).

Bu farklılıklar tek bir evrensel metrik kümesinin yetersiz olduğunu ve her görev türü için özelleşmiş veya ağırlıklandırılmış metrik kombinasyonlarına ihtiyaç duyulduğunu açıkça ortaya koymaktadır. Örneğin, bir hikaye üreticisinde "yaratıcılık" bir ölçüt olarak görülebilirken, bir finansal rapor üreticisinde bu ölçütün yerini kesinlikle "olgusal doğruluk" alır.

3. Dilsel Metrikler

AI tarafından üretilen bir metnin ilk ve en temel değerlendirmesi, onun dilsel formu üzerinden yapılır. Dilsel metrikler, bir modelin insan dilinin temel yapısal kurallarını (sözdizimi, morfoloji) ne derecede içselleştirdiğini ve bunları üretim sırasında ne kadar tutarlı uyguladığını nicel olarak ortaya

koyar. Bu metrikler, çoğunlukla referans metne ihtiyaç duymadan (referanssız) hesaplanabilir ve modelin teknik performansının genel bir göstergesidir.

3.1. Dil Kalitesi

Dil kalitesinin en köklü ve teorik olarak sağlam ölçütü Perplexity (PPL - Karışıklık)'tır. PPL, temelde bir olasılık dağılımının (burada dil modelinin), bir test veri kümesini tahmin etmedeki etkinliğinin ölçüsüdür. Matematiksel olarak, modelin test seti üzerinde hesaplanan çapraz entropi (cross-entropy) kaybının üsselidir (Jurafsky & Martin, 2023).

Daha düşük bir PPL değeri, modelin test metnini daha yüksek bir olasılıkla öngördüğünü, yani dilin istatistiksel dağılımını daha iyi modellemiş olduğunu gösterir. Bu da genellikle daha doğal ve dilbilgisel olarak tutarlı çıktılar üretme potansiyeli ile ilişkilendirilir. Örneğin, GPT-3 (175B) modeli, WikiText-103 veri setinde 18.3 gibi dikkat çekici derecede düşük bir PPL elde etmiştir (Brown vd., 2020). Ancak, PPL'nin, bazı temel kısıtlamalara sahiptir. PPL, metnin yerel (local) akışını ölçmesine karşın, uzun vadeli (long-range) yapısal veya anlamsal tutarlılık hakkında hiçbir bilgi vermemesidir. Ayrıca, aşırı derecede düşük PPL, modelin eğitim verisini ezberleme (overfitting) riskine de işaret edebilir.

3.2. Akıcılık ve Doğruluk

Dilsel kalitenin bir diğer boyutu, metnin tekdüzelikten uzak, zengin ve akıcı bir yapıya sahip olmasıdır. Bu özellik, Leksikal Çeşitlilik (Lexical Diversity) metrikleri ile ölçülür. Bu alandaki en temel metrik Tür-Belirteç Oranı (Type-Token Ratio - TTR)'dır, ancak metin uzunluğundan güçlü bir şekilde etkilendiği için karşılaştırmalı çalışmalarda yetersiz kalır (McCarthy & Jarvis, 2010). Bu sorunu aşmak için, Hareketli Ortalama TTR (Moving-Average Type-Token Ratio - MATTR) önerilmiştir (Covington & McFall, 2010). MATTR, metin boyunca sabit uzunlukta bir pencereyi (genellikle 50-100 kelime) kaydırarak her bir penceredeki TTR'yi hesaplar ve bu değerlerin ortalamasını alır. Bu yöntem, uzunluk etkisini büyük ölçüde ortadan kaldırarak daha güvenilir bir karşılaştırma imkanı sunar. Dilsel çeşitlilik, yalnızca tekrarlardan kaçınmanın ötesinde, metnin stilistik olgunluğunun ve ifade zenginliğinin bir göstergesi olarak da kabul edilir (Kyle & Crossley, 2017). AI tarafından üretilen metinlerin genellikle insan yazılarına kıyasla daha düşük leksikal çeşitlilik sergilediği gözlemlenmiştir (Ippolito vd., 2020).

Dilsel metrikler, bir metnin biçimsel yapısını ve teknik yeterliliğini değerlendirmede temel oluşturur. Ancak, dilbilgisi açısından kusursuz ve kelime hazinesi bakımından zengin bir metin, anlamsal karmaşa, mantıksal tutarsızlık veya olgusal yanlışlıklar barındırabilir. Bu nedenle, bir sonraki aşama, metnin

derin yapısını ve anlamını sorgulayan anlamsal metriklerin uygulanmasıdır. AI içerik değerlendirmesinin kalite kontrolünün yalnızca biçimsel değil, aynı zamanda içeriksel (content) boyutuna kaymasını temsil eder.

4. ANLAMSAL METRİKLER

Anlamsal metrikler, yapay zeka tarafından üretilen bir metnin, hedeflenen anlamı veya referans içeriği ne ölçüde taşıdığını değerlendirmeyi amaçlar. Bu metrikler, yalnızca kelimelerin biçimsel eşleşmesine odaklanmak yerine, metinler arasındaki anlamsal benzerliği, içerik uygunluğunu ve bilgi korunumunu ölçmeye çalışır. Değerlendirme yaklaşımları, geleneksel istatistiksel yöntemlerden modern derin öğrenme tabanlı tekniklere kadar geniş bir yelpazede kullanılmıştır.

4.1. Geleneksel N-Gram Örtüşme Metrikleri

Geleneksel anlamsal değerlendirmenin temelini, referans metin ile üretilen metin arasındaki kelime dizilerinin (n-gram) örtüşme oranını hesaplayan metrikler oluşturur. Bu metriklerin en yaygın örnekleri BLEU (Bilingual Evaluation Understudy) ve ROUGE (Recall-Oriented Understudy for Gisting Evaluation)'dir.

BLEU, öncelikli olarak makine çevirisi değerlendirmesi için geliştirilmiş olup, aday çevirinin n-gramlarının (1-gram, 2-gram, 3-gram, 4-gram) referans çevirilerde ne sıklıkta görüldüğüne dayalı bir kesinlik (precision) skoru hesaplar. Kısa çevirileri cezalandırmak için bir "brevity penalty" faktörü içerir (Papineni vd., 2002). Temel mantığı, ne kadar çok n-gram örtüşürse, çevirinin kalitesinin o kadar yüksek olduğu varsayımına dayanır.

ROUGE ise metin özetleme görevlerinin değerlendirilmesi için tasarlanmıştır ve daha çok geri çağırma (recall) odaklıdır. ROUGE-N, referans özetteki n-gramların ne kadarının aday özette bulunduğunu ölçer. ROUGE-L ise iki metin arasındaki en uzun ortak alt diziyi (Longest Common Subsequence - LCS) temel alır ve bu sayede kelime sırası bilgisini kısmen korur (Lin, 2004). Hem BLEU hem de ROUGE, hesaplaması basit, hızlı ve nesnel olmaları nedeniyle uzun süre standart metrikler olarak kullanılmıştır.

Ancak, bu geleneksel metriklerin bazı yetersizlikleri bulunmaktadır. En temel yetersizliği, anlamsal eşdeğerliği (semantic equivalence) yakalayamamalarıdır. Örneğin, "Araba hızla gitti" ve "Otomobil süratle ilerledi" cümleleri anlamsal olarak aynıyken, n-gram örtüşmesi sıfırdır. Ayrıca, kelime sırasındaki yaratıcı varyasyonları veya eş anlamlı kullanımları cezalandırma eğilimindedirler. Bu nedenle, bu metrikler yüksek skorların dilsel ve anlamsal kalitenin mutlak bir göstergesi olmadığı bilinciyle, genellikle diğer metriklerle birlikte kullanılmalıdır (Callison-Burch vd., 2006).

4.2. Modern Anlamsal Benzerlik Metrikleri

Derin öğrenme ve ön-eğitimli dil modellerinin yükselişi, anlamsal değerlendirmede bir devrim yaratmıştır. Modern metrikler, metinleri sabit kurallarla değil, bağlamsal vektör temsilleri aracılığıyla karşılaştırır.

Sentence Transformers (ST), BERT gibi modelleri, cümle çiftleri üzerinden Siamese veya triplet ağ mimarileriyle yeniden eğiterek, cümleleri anlamsal olarak zengin ve yoğun vektörlere (sentence embeddings) dönüştürmeyi amaçlar (Reimers & Gurevych, 2019). Bu vektörler, cümlelerin anlamını yüksek boyutlu bir uzayda kodlar. İki cümle arasındaki anlamsal benzerlik, bu vektörlerin kosinüs benzerliği hesaplanarak ölçülür. Bu yöntem, farklı kelimelerle ifade edilen aynı anlamları yakalayabilme yeteneği sağlar.

BERTScore ise daha da ileri giderek token (kelime/alt kelime) düzeyinde bir analiz sunar (Zhang vd., 2020). Referans ve aday metinlerdeki her bir token'ın, BERT gibi bir model tarafından üretilen bağlamsal vektör gömmesi kullanılır. Daha sonra, greedy matching ile her referans token'ı için aday metindeki en benzer token bulunur ve eşleştirilir. Bu eşleştirmelere dayanarak bir Kesinlik (Precision) (aday token'ların referansla anlamsal uyumu), Duyarlılık (Recall) (referans token'larının adayda anlamsal karşılık bulma oranı) ve nihai F1 Skoru hesaplanır. BERTScore, bağlamsal anlambilim ve eş anlamlılığı hesaba kattığı için, özellikle özetleme ve metin oluşturma görevlerinde BLEU ve ROUGE'a kıyasla insan değerlendirmeleriyle çok daha yüksek korelasyon gösterir.

Bu modern metrikler, geleneksel yöntemlerin ana açığını kapatmış olsa da, kendi sınırlılıkları vardır. Performansları, kullanılan temel ön-eğitimli modele (dil, mimari, eğitim verisi) bağımlıdır. Ayrıca, referans metne olan ihtiyaçları, tamamen özgün yaratıcı içeriklerin değerlendirilmesini zorlaştırabilir.

4.3. İçerik Uygunluğu

Belirli bir alana veya konuya odaklanmış içerik üretiminde, metnin yalnızca genel anlam benzerliği değil, aynı zamanda alana özgü kritik kavramları ve anahtar kelimeleri ne derecede içerdiği de önemlidir. Bu ihtiyaca yönelik olarak Alan Anahtar Kelime Skoru (Domain Keyword Score - DKS) gibi metrikler kullanılır.

DKS, referans metinden veya önceden tanımlanmış bir listeden, anlamsal ağırlığı yüksek anahtar kelimelerin (stop-word'ler çıkarıldıktan sonra) çıkarılması prensibine dayanır. Daha sonra, bu anahtar kelimelerin üretilen metinde bulunma oranı hesaplanır (Louis & Nenkova, 2013). Basit bir formülle ifade edilir: DKS = (Üretilen Metindeki Ortak Anahtar Kelime Sayısı) / (Referans Metindeki Toplam Anahtar Kelime Sayısı). Bu metrik, modelin konudan sapıp sapmadığını, gerekli terminolojiyi kullanıp kullanmadığını ve

temel bilgileri atlayıp atlamadığını nesnel ve hızlı bir şekilde gösterebilir.

Özellikle teknik raporlama, akademik özet çıkarma veya kurumsal duyuru üretimi gibi alana özgü ve bilgi odaklı görevlerde, DKS anlamsal değerlendirmeye önemli bir katkı sunar. Ancak, yalnızca kelime varlığını kontrol ettiği için, bu kelimelerin bağlam içinde doğru ve tutarlı bir şekilde kullanılıp kullanılmadığını ölçemez. Bu nedenle, BERTScore veya Sentence Transformers gibi derin anlamsal metriklerle birlikte kullanıldığında en etkili sonucu verir.

5. Uygulamalı Metrikler

AI sisteminin değeri, onun gerçek dünya bağlamında sağladığı fayda ve kullanıcılar tarafından nasıl deneyimlendiği ile ölçülür. Uygulamalı metrikler, model çıktısının pratik etkinliğini ve insan merkezli kabulünü değerlendirir.

5.1. Kullanıcı Etkileşimi

İnsan merkezli değerlendirme, AI çıktılarının kalitesine dair en doğrudan ve bazen en güvenilir ölçümü sağlar (Howcroft vd., 2020). Bu değerlendirme, genellikle yapılandırılmış anketler veya kontrollü deneyler aracılığıyla yapılır. Likert Ölçeği gibi anketler, katılımcılardan metnin akıcılığı, anlaşılırlığı, yararlılığı, güvenilirliği, tarafsızlığı ve genel kalitesi gibi özellikler hakkında derecelendirme yapmalarını ister (Boone & Boone, 2012). Doğrudan Karşılaştırma (Paired Comparison/A/B Testing) ise iki farklı sistemin (ör. bir AI modeli vs. insan, veya iki farklı AI modeli) çıktılarını aynı kriterlere göre sıralamalarını veya tercih etmelerini ister. Bu yöntemler, otomatik metriklerin yakalayamadığı estetik, ikna edicilik, güven oluşturma veya keyif verme gibi nitelikleri ölçebilir. Örneğin, erken dönem çalışmalar, otomatik üretilmiş haberlerin insan yazarlara kıyasla daha tarafsız ve nesnel, ancak aynı zamanda daha sıkıcı ve mekanik algılandığını ortaya koymuştur (Clerwall, 2014). İnsan değerlendirmeleri ayrıca, farklı kültürel veya demografik grupların AI çıktılarına tepkisindeki farklılıkları anlamak için de kritiktir.

5.2. Görev Başarımı

Birçok AI sistemi, belirli bir pratik görevi otomatikleştirmek veya insan performansını artırmak için tasarlanmıştır. Bu durumda, sistemin gerçek başarısı, o görevin kendine özgü başarım ölçütleriyle değerlendirilmelidir (Rajpurkar vd., 2018). Görev başarımı metrikleri, modelin teorik dil yeteneğinden çok, pratik bir sorunu çözmedeki etkinliğini yansıtır. Örneğin, Bir Soru-Cevap (QA) sisteminde, Doğru Cevap Oranı (Accuracy), veya daha ince ayarlı Kesinlik (Precision), Duyarlılık (Recall) ve F1 Skoru metrikleri standarttır.Bir metin özetleme sisteminde ise Bilgi yoğunluğunu ölçmek için ROUGE skorları, anlamsal sadakat için BERTScore, olgusal tutarlılık için NLI tabanlı metrikler

ve okunabilirlik için insan değerlendirmesi bir arada kullanılır. Bir kod üretme sistemlerinde üretilen kodun işlevsel doğruluğu (functional correctness), belirli bir test kümesini geçme oranı ile ölçülebilir (Chen vd., 2021). Kodun sentaks hatası içermemesi ve zaman/mekan karmaşıklığı açısından verimli olması da değerlendirilebilir. Bir diyalog sistemi (chatbot)'nde görevi başarıyla tamamlama oranı ortalama diyalog uzunluğu, kullanıcı memnuniyet anketi ve insan müdahalesine ihtiyaç duyma sıklığı gibi metrikler kullanılır.

6. Metriklerin Değerlendirilmesi

Yapay zeka ile metin üretimi alanında üretilen metinlerin değerlendirilmesi için kullanılan metriklerin güçlü yanları ve sınırlılıkları bulunmaktadır.

6.1. Metriklerin Güçlü Yanları

Dilsel, anlamsal ve uygulamalı metrikleri içeren çok boyutlu bir değerlendirme seti, AI üretimi içeriğin kalitesine dair dengeli ve bütüncül bir bakış açısı sağlar. Bu yaklaşım, modelin farklı yönlerdeki (akıcılık, anlam koruma, kullanıcı memnuniyeti) performansını ayrı ayrı ve birbiriyle ilişkili olarak analiz etme imkânı tanır. Örneğin, yüksek bir DKS skoru, modelin konuya sadık kaldığını doğrularken, yüksek bir BERTScore aynı zamanda bunu anlamsal bir bütünlük içinde yaptığını gösterir. İyi bir PPL ise bu içeriğin akıcı bir dille sunulduğuna işaret eder. Kullanıcı anketleri ise tüm bu teknik başarının pratikte ne derece değerli bulunduğunu ortaya koyar. Bu katmanlı yapı, geliştiricilere model iyileştirmesi için net yol haritaları sunar.

6.2. Metriklerde karşılaşılan kısıtlılıklar

Değerlendirme metrikleri, AI üretimi metinlerin kalitesini ölçmede temel araçlar olsa da, her biri bazı kısıtlılıklar taşır. Perplexity (PPL), dilin istatistiksel akışını ölçerken, uzun vadeli anlamsal tutarlılık veya mantıksal bütünlük hakkında bilgi vermez ve aşırı düşük değerler aşırı öğrenme (overfitting) riskine işaret edebilir (Jurafsky & Martin, 2023). Leksikal çeşitlilik metrikleri (TTR, MATTR) kelime zenginliğini gösterir ancak metnin anlamsal derinliği veya bilgi yoğunluğu ile ilgili bir garanti sunmaz (Bestgen, 2024).

Anlamsal benzerlik metrikleri (BERTScore, Sentence Transformers) büyük bir ilerleme sağlasa da, performansları altında yatan ön-eğitimli modele bağımlıdır ve özellikle özgün yaratıcı içeriklerde uygun bir referans metin bulma sorunu yaşanabilir (Zhang vd., 2020). Daha da kritik olarak, bu metrikler olgusal doğruluk veya hallüsinasyon gibi ciddi hataları güvenilir şekilde tespit etmekte genellikle yetersiz kalır (Ji vd., 2023). İnsan değerlendirmeleri ise öznel deneyimi yakalayabilir ancak pahalı, ölçeklenmesi zor ve katılımcı yanlılığına açıktır (Clark vd., 2021).

Bu nedenle, hiçbir metrik tek başına yeterli değildir. Sağlam bir değerlendirme, dilsel, anlamsal ve insan-merkezli metrikleri bir araya getiren, göreve özgü çok boyutlu bir çerçeve gerektirir. Metrikler, nihai yargıdan ziyade bir tanı ve iyileştirme aracı olarak ele alınmalıdır.

7. Sonuç

Yapay zeka destekli içerik üretiminin dinamik ve hızla gelişen dünyasında, performans metrikleri bu teknolojinin gelişmesi için vazgeçilmez bir rehber işlevi görmektedir (Chang vd., 2024). Metin kalitesinin çok boyutlu yapısı, tek bir ölçütle kapsamlı bir değerlendirme yapılmasını imkansız kılmaktadır (Celikyilmaz vd., 2020). Dilsel metrikler, üretimin biçimsel mükemmelliğini ve akıcılığını ölçerek temel bir teknik yeterlilik standardı sunar (Jurafsky & Martin, 2023). Anlamsal metrikler ise biçimsel yapının ötesine geçerek, içeriğin derin yapısını, bağlamsal tutarlılığını ve bilgi bütünlüğünü sorgular (Zhang vd., 2020). Nihai kullanıcı perspektifini merkeze alan uygulamalı metrikler ise sistemlerin pratik değerini, görev başarımını ve sosyal kabulünü değerlendirerek gerçek dünya bağlamına geçişi sağlar (Rajpurkar vd., 2018; Howcroft vd., 2020). Bu üçlü yaklaşımın entegre bir şekilde kullanılması, geliştiricilere ve araştırmacılara AI modellerinin güçlü ve zayıf yönlerine dair dengeli, derinlikli ve eyleme dönüştürülebilir bir teşhis imkanı tanımaktadır (Liang vd., 2022).

Ancak, mevcut metrik ekosistemi önemli ilerlemeler kaydetmiş olsa da, önündeki zorluklar da azımsanmayacak ölçüdedir. Olgusal doğruluk ve hallüsinasyon tespiti gibi kritik güvenilirlik sorunlarına yönelik sağlam ve ölçeklenebilir metriklerin geliştirilmesi acil bir ihtiyaç olarak öne çıkmaktadır (Ji vd., 2023). Benzer şekilde, yanlılık, şeffaflık ve adalet gibi etik boyutları sistematik olarak ölçebilen çerçevelere olan gereksinim giderek artmaktadır (Bender vd., 2021). Ayrıca, mevcut metodolojilerin çoğunlukla İngilizce ve belirli kültürel bağlamlarla sınırlı kalması, çok dilli ve kültürlerarası geçerliliği olan değerlendirme araçlarının geliştirilmesini zorunlu kılmaktadır (Ruder vd., 2022). İnsan değerlendirmelerinin pahalı ve ölçeklenemez doğası ise, insan yargısıyla yüksek korelasyon gösteren gelişmiş otomatik metriklerin arayışını sürekli kılmaktadır (Clark vd., 2021).

Sonuç olarak, AI tabanlı içerik üretiminin geleceği, yalnızca daha güçlü modeller geliştirmeye değil, aynı zamanda bu modellerin çıktılarını anlamak, iyileştirmek ve güvence altına almak için daha kapsayıcı, sağlam ve çok yönlü değerlendirme metrikleri inşa etmeye bağlıdır. İdeal bir değerlendirme çerçevesi, teknik performansı, anlamsal bütünlüğü, pratik faydayı ve etik sorumluluğu bir arada ele almalıdır. Bu kapsamlı yaklaşım sayesinde, yapay zeka sistemleri güvenilir, anlamlı ve sorumlu içerik üreten sistemler olarak toplumsal ve bilimsel süreçlere entegre olabilecektir. Nihayetinde, teknolojinin gerçek başarısı ve yaygın kabulü de bu entegrasyonun sağlıklı bir şekilde gerçekleşmesiyle mümkün olacaktır.

KAYNAKÇA

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623. doi:10.1145/3442188.3445922

Bestgen, Y. (2024). Measuring lexical diversity in texts: The twofold length problem. Language Learning, 74(3), 638-671. doi:10.1111/lang.12616

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... Liang, P. (2021). On the opportunities and risks of foundation models. arXiv preprint. Retrieved from https://arxiv.org/abs/2108.07258

Boone, H. N., Jr., & Boone, D. A. (2012). Analyzing Likert data. The Journal of Extension, 50(2), Article 48. Retrieved from https://open.clemson.edu/joe/vol50/iss2/48/

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... Amodei, D. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877–1901.

Callison-Burch, C., Osborne, M., & Koehn, P. (2006). Re-evaluating the role of BLEU in machine translation research. Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics, 249–256.

Celikyilmaz, A., Clark, E., & Gao, J. (2020). Evaluation of text generation: A survey. arXiv preprint. Retrieved from https://arxiv.org/abs/2006.14799

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., ... Dai, J. (2024). A survey on evaluation of large language models. ACM Computing Surveys, 56(9), 1–40. doi:10.1145/3641289

Chen, M., Tworek, J., Jun, H., Yuan, Q., de Oliveira Pinto, H. P., Kaplan, J., ... Zaremba, W. (2021). Evaluating large language models trained on code. arXiv preprint. Retrieved from https://arxiv.org/abs/2107.03374

Clark, E., August, T., Serrano, S., Haduong, N., Gururangan, S., & Smith, N. A. (2021). All that's 'human' is not gold: Evaluating human evaluation of generated text. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics, 7282–7296. doi:10.18653/v1/2021.acl-long.565

Clerwall, C. (2014). Enter the robot journalist: Users' perceptions of automated content. Journalism Practice, 8(5), 519–531. doi:10.1080/17512786.2014.883116

Covington, M. A., & McFall, J. D. (2010). Cutting the Gordian knot: The moving-average type–token ratio (MATTR). Journal of Quantitative Lin-

guistics, 17(2), 94–105. doi:10.1080/09296171003643098

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 4171–4186. doi:10.18653/v1/N19-1423

Gehrmann, S., Clark, E., & Sellam, T. (2023). Repairing the cracked foundation: A survey of obstacles in evaluation practices for generated text. Journal of Artificial Intelligence Research, 77, 103-166. doi:10.1613/jair.1.13715

Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., De Laroussilhe, Q., Gesmundo, A., ... Gelly, S. (2019). Parameter-efficient transfer learning for NLP. Proceedings of the 36th International Conference on Machine Learning, 2790–2799.

Howard, J., & Ruder, S. (2018). Universal language model fine-tuning for text classification. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 328–339. doi:10.18653/v1/P18-1031

Howcroft, D. M., Belz, A., Clinciu, M., Gkatzia, D., Hasan, S. A., Mahamood, S., ... Rieser, V. (2020). Twenty years of confusion in human evaluation: NLG needs evaluation sheets and standardised definitions. Proceedings of the 13th International Conference on Natural Language Generation, 169–181.

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., ... Chen, W. (2021). LoRA: Low-rank adaptation of large language models. arXiv preprint. Retrieved from https://arxiv.org/abs/2106.09685

Ippolito, D., Kriz, R., Sedoc, J., Kustikova, M., & Callison-Burch, C. (2020). Comparison of diverse decoding methods from conditional language models. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2762–2777. doi:10.18653/v1/2020.acl-main.246

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., ... Fung, P. (2023). Survey of hallucination in natural language generation. ACM Computing Surveys, 55(12), 1–38. doi:10.1145/3571730

Jurafsky, D., & Martin, J. H. (2023). Speech and language processing (3rd ed. draft). Retrieved from Pearson.

Kryściński, W., Rajani, N., Agarwal, D., Xiong, C., & Radev, D. (2020). Evaluating the factual consistency of abstractive text summarization. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, 9332–9346. doi:10.18653/v1/2020.emnlp-main.750

Kyle, K., & Crossley, S. A. (2017). Automatically assessing lexical sophistication: Indices, tools, findings, and application. TESOL Quarterly, 51(3), 591–616. doi:10.1002/tesq.344

Li, J., & Jurafsky, D. (2017). Neural net models of open-domain discourse coherence. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 198–209. doi:10.18653/v1/D17-1019

Liang, P., Bommasani, R., Lee, T., Tsipras, D., Soylu, D., Yasunaga, M., ... Koreeda, Y. (2022). Holistic evaluation of language models. arXiv preprint. Retrieved from https://arxiv.org/abs/2211.09110

Lin, C. Y. (2004). ROUGE: A package for automatic evaluation of summaries. Proceedings of the ACL-04 Workshop on Text Summarization Branches Out, 74–81.

Liu, Y., Iter, D., Xu, Y., Wang, S., Xu, R., & Zhu, C. (2023). Gpteval: NLG evaluation using GPT-4 with better human alignment. arXiv preprint. Retrieved from https://arxiv.org/abs/2303.16634

Louis, A., & Nenkova, A. (2013). Automatically assessing machine summary content without a gold standard. Computational Linguistics, 39(2), 267–300. doi:10.1162/COLI_a_00123

McCarthy, P. M., & Jarvis, S. (2010). MTLD and vocd-D: A validation study of sophisticated approaches to lexical diversity assessment. Language Testing, 27(3), 301–321. doi:10.1177/0265532209349471

Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002). BLEU: A method for automatic evaluation of machine translation. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, 311–318. doi:10.3115/1073083.1073135

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training. San Francisco, CA: OpenAI.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., ... Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. Journal of Machine Learning Research, 21, 1–67.

Rajpurkar, P., Jia, R., & Liang, P. (2018). Know what you don't know: Unanswerable questions for SQuAD. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 784–789. doi:10.18653/v1/P18-1124

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing, 3982–3992.

doi:10.18653/v1/D19-1410

Ruder, S., Vulic, I., & Søgaard, A. (2022). A survey of cross-lingual word embedding models. Journal of Artificial Intelligence Research, 73, 571–609. doi:10.1613/jair.1.11640

Thorne, J., Vlachos, A., Christodoulopoulos, C., & Mittal, A. (2018). FEVER: A large-scale dataset for fact extraction and verification. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 809–819. doi:10.18653/v1/N18-1074

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M. A., Lacroix, T., ... Lample, G. (2023). LLaMA: Open and efficient foundation language models. arXiv preprint. Retrieved from https://arxiv.org/abs/2302.13971

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998–6008.

Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., & Artzi, Y. (2020). BERTScore: Evaluating text generation with BERT. International Conference on Learning Representations.

Bölüm 5

CU-ZN İNCE FİLMLERDE KOMPOZİSYONUN ÖZDİRENÇ ÜZERİNDEKİ ETKİSİ: ÇOKLU REGRESYON YAKLAŞIMI

Rasim ÖZDEMİR [1]
Hasan GÜLER [2]

1 Prof. Dr., Kilis 7 Aralık Üniversitesi, Elektrik Bölümü, Orcid: 0000-0003-1439-0444
2 Dr. Öğr. Üyesi, Kilis 7 Aralık Üniversitesi, Bilgisayar Teknolojileri Bölümü, Orcid: 0000-0001-5312-171X

Bakır-çinko (Cu-Zn) alaşımları, üstün elektriksel iletkenlik ve termal özellikleri sayesinde elektrik-elektronik sektöründe, dekoratif kaplamalarda ve korozyon direnci gerektiren endüstriyel uygulamalarda yaygın olarak kullanılmaktadır (Brenner, 1963; Juškėnas, Karpavičienė, Pakštas, Selskis, & Kapočius, 2007). Malzemelerin fiziksel özellikleri yapı ve bileşimleriyle doğrudan ilişkili olduğundan (Porter, 2009), alaşım karakteristiklerinin sistematik kontrolü teknolojik açıdan kritik öneme sahiptir.

Elektrokimyasal depolama, Cu-Zn alaşımlarının üretiminde düşük maliyeti, düşük sıcaklıkta uygulanabilirliği ve film kalınlığı üzerinde hassas kontrol imkânı sunması nedeniyle tercih edilen bir yöntemdir (Pletcher & Walsh, 1993). Siyanürün yüksek toksisitesi nedeniyle çevre dostu alternatifler arasında sitrat iyonları, Cu-Zn elektrodepozisyonunda yaygın olarak kullanılmaktadır (Liu et al., 2020; Özdemir & Karahan, 2019). Bakır ve çinko iyonlarının indirgenme potansiyelleri arasındaki büyük fark, anormal birlikte depolanma olarak adlandırılan karmaşık elektrokimyasal kinetiklere yol açmaktadır. Sitrat iyonları bakırın indirgenme kinetiğini modüle ederek alaşım bileşimini doğrudan etkiler (Oulmas et al., 2019).

Elektrodepozisyonla üretilen Cu-Zn filmlerinde özdirenç değerleri, mekanik yöntemlerle üretilenlere göre %25-45 daha yüksek olup bu durum film bileşimi, faz yapısı ve içyapı kusurlarına (tane sınırları, dislokasyonlar) bağlıdır (Fairbank, 1944). Matthiessen kuralına göre toplam özdirenç, ısıl titreşim, safsızlık ve kafes kusurları kaynaklı bileşenlerin toplamıdır. Tane boyutunun küçülmesi tane sınırı yoğunluğunu artırarak özdirenci yükseltirken (Benrazzouq, Ghanbaja, Migot, Milichko, & Pierson, 2025), kompleksleyici ajan konsantrasyonunun artması nükleasyon merkezlerini çoğaltarak tane boyutunu küçültür (Kruglikov, Kudriavtsev, Vorobiova, & Antonov, 1965). Bakır bazlı alaşımlarda yüksek mekanik mukavemet ile yüksek elektriksel iletkenlik genellikle birbirini dışlayan özelliklerdir; bu paradoks özel mikroyapısal stratejilerle aşılmaya çalışılmaktadır (Dangwal, Edalati, Valiev, & Langdon, 2023).

Akım yoğunluğu, sıcaklık, pH ve metal iyon konsantrasyonları gibi çok sayıda parametrenin alaşım özelliklerini etkileyen doğrusal olmayan ilişkiler sergilemesi, deterministik modellemenin yetersiz kalmasına neden olmaktadır (To, Park, Kim, Cho, & Myung, 2022; Güler, Özdemir, & Coşkun, 2025). Bu bağlamda çoklu regresyon analizi, korelasyon çalışmaları ve sağlamlık testleri gibi ileri istatistiksel yöntemler, karmaşık ilişkilerin sistematik olarak ortaya konması için güçlü araçlar sunmaktadır (Xiao, Li, Bordas, & Kim, 2023).

Bu çalışma, elektrokimyasal depolama parametrelerinin Cu-Zn alaşım özellikleri üzerindeki etkisini karakterizasyon teknikleri ve regresyon modelleme yaklaşımlarıyla incelemeyi amaçlamaktadır. Elde edilecek bulgular, temel malzeme bilimine katkı sağlamanın yanı sıra, alaşım özelliklerinin tahmin edilebilirliğini artırarak teknolojik uygulamaların geliştirilmesine olanak tanıyacaktır.

Deneysel Çalışma

Cu-Zn alaşım ince filmler, elektrokimyasal depolama yöntemiyle alüminyum altlıklar üzerine üretildi. Yapısal ve yüzeysel karakterizasyon için dönüşümlü voltametri (CV), taramalı elektron mikroskobu (SEM), X-ışını difraksiyon spektrometresi (XRD) ve enerji dağılımlı X-ışını spektroskopisi (EDX) analizleri gerçekleştirildi. Elektriksel özdirenç, düşük direnç değerlerinde alet ve temas hatalarını minimize eden Van Der Pauw dört nokta kontak yöntemiyle ölçüldü (Cultrera, Serazio, Fabricius, & Callegaro, 2024; Rasim Özdemir, Karahan, & Karabulut, 2016). Ölçümler, bir kriyostat yardımıyla 100-405 K sıcaklık aralığında 10 °C aralıklarla bilgisayar kontrolünde otomatik olarak kaydedildi.

Tablo 1 CuZn ince film alaşımların banyo içerikleri ve deney verileri

SN	Elektrolit içerisindeki malzemeler			pH	Akım (mA)	Zaman Dk.	Depolanma Potansiyeli (V)
	$CuSO_4 5H_2O$ (Mol/lt)	$ZnSO_4 7H_2O$ (Mol/lt)	$Na_3C_6H_5O_7$ (Mol/lt)				
$Zn_{38}Cu_{62}$	0.06	0.2	0.3	5,8	60	60	1.549
$Zn_{25}Cu_{75}$	0.06	0.2	0.5	5,8	60	60	1.538
$Zn_{17}Cu_{83}$	0.06	0.2	0.7	5,8	60	60	1.517
$Zn_{13}Cu_{87}$	0.06	0.2	0.9	5,8	60	60	1.529

Tablo 1'de görüldüğü gibi, sodyum sitrat konsantrasyonu 0.3-0.9 Mol/L arasında değiştirilirken diğer parametreler sabit tutuldu. Diğer taraftan elektrolit bileşimi sabit tutulmasına rağmen film içi Cu oranı %13.01-38.31 aralığında değişmektedir; bu durum sitrat konsantrasyonunun anormal birlikte depolanma kinetiği üzerindeki etkisini yansıtmaktadır. Tane boyutu, XRD verilerinden Debye-Scherrer formülü $L = 0.9\lambda(B.Cos\theta)^{-1}$ kullanılarak hesaplandı (Leontyev et al., 2018).

Veriseti

Elektriksel özdirencin modellenmesi için deneysel verilerden kapsamlı bir veri seti oluşturuldu. Bu veri seti; elektrokimyasal depolama koşulları, banyo bileşen oranları, EDX analizinden elde edilen film bileşim oranları, XRD analizinden hesaplanan tane büyüklüğü ve sıcaklıkla değişen özdirenç değerlerini içermektedir.

Veri seti, özdirence etki eden sekiz bağımsız değişken ile bir bağımlı değişkeni kapsayan 127 örnekten oluşmaktadır. Bağımsız değişkenler: (1) elektrolit içindeki Cu miktarı (ECu, %), (2) elektrolit içindeki Zn miktarı (EZn, %), (3) ince filmdeki Cu miktarı (FCu, %), (4) ince filmdeki Zn miktarı (FZn, %), (5) Debye-Scherrer formülü ile hesaplanan tane büyüklüğü (G, nm), (6) depolanma voltajı (V), (7) elektrolit sıcaklığı (St, °C) ve (8) ölçüm sırasındaki ortam sıcaklığı (T, Kelvin).

Bağımlı değişken ise elektriksel özdirenç (ρ, μΩ·cm) değeridir.

Korelasyon Analizi

Özdirenç ile diğer değişkenler arasındaki ilişkiler, Pearson, Spearman ve Kendall korelasyon analizleri kullanılarak incelenmiştir (Tablo 2). Bu çoklu yaklaşım, hem doğrusal hem de monoton ilişkilerin tespitine ve sonuçların tutarlılığının değerlendirilmesine olanak tanımaktadır.

Tablo 2 Resistivity ile Diğer Değişkenler Arasındaki Korelasyon Katsayıları

Değişken Çifti	Pearson (r)	Spearman (ρ)	Kendall (τ)	İlişki Gücü
Sıcaklık - Resistivity	$+0.615^{***}$	$+0.608^{***}$	$+0.462^{***}$	Güçlü pozitif
FCu% - Resistivity	-0.751^{***}	-0.773^{***}	-0.620^{***}	Güçlü negatif
FZn% - Resistivity	$+0.751^{***}$	$+0.773^{***}$	$+0.620^{***}$	Güçlü pozitif
FCu% - FZn%	-1.000^{***}	-1.000^{***}	-1.000^{***}	Mükemmel negatif
Tane Boyutu - FCu%	$+0.852^{***}$	$+0.847^{***}$	$+0.703^{***}$	Çok güçlü pozitif
Potansiyel - FCu%	$+0.806^{***}$	$+0.802^{***}$	$+0.651^{***}$	Çok güçlü pozitif

$^{***}p < 0.001$

Sıcaklık ile özdirenç arasında güçlü pozitif korelasyon tespit edilmiştir (r=0.615). Yüksek sıcaklıklarda fonon saçılımının artması elektron taşınımını engelleyerek direnci yükseltmektedir. Bakır içeriği (FCu) ile özdirenç arasındaki güçlü negatif korelasyon (r=−0.751), saf bakırın çinkoya göre yüksek iletkenliğini (ρ~Cu~ ≈ 1.68 μΩ·cm, ρ~Zn~ ≈ 5.90 μΩ·cm) yansıtmaktadır.

FCu ve FZn arasındaki mükemmel negatif korelasyon (−1.000), ikili alaşım yapısını doğrulamaktadır. Tane boyutu ile FCu arasındaki güçlü pozitif korelasyon (r=0.852), bakır içeriği artışıyla kristal tane boyutunun sistematik arttığını göstermektedir. Potansiyel-FCu korelasyonu (r=0.806) ise Cu^{2+}/Cu (+0.34 V) ve Zn^{2+}/Zn (−0.76 V) redüksiyon potansiyellerinin kompozisyonel etkisini yansıtmaktadır.

Üç korelasyon yönteminin tutarlı sonuçları, ilişkilerin doğrusal ve monoton karakterde olduğunu doğrulamaktadır. Yüksek korelasyonlar, regresyon analizlerinde multikollinearite riskine dikkat edilmesi gerektiğini işaret etmektedir.

Kısmi Korelasyon Analizi

Değişkenler arasındaki ilişkilerde sıcaklığın konfounding (karıştırıcı) etkisini değerlendirmek amacıyla kısmi korelasyon analizi gerçekleştirilmiştir (Tablo 3).

Tablo 3 Sıcaklık Kontrolü Altında Kısmi Korelasyon Analizi Sonuçları

Değişken	Sıfırıncı Derece Korelasyon (r_0)	Kısmi Korelasyon ($r_{partial}$)	Mutlak Fark (Δr)	Değişim (%)	Yorumlama
FCu%	-0.751***	-0.957***	0.206	+27.4%	Çok güçlü negatif ilişki
FZn%	+0.751***	+0.957***	0.206	+27.4%	Çok güçlü pozitif ilişki
Tane Boyutu (nm)	-0.544***	-0.692***	0.147	+27.1%	Güçlü negatif ilişki

****p < 0.001*

Sıcaklık kontrolü altında tüm değişkenlerin özdirenç ile korelasyonları belirgin şekilde güçlenmiştir. Bakır içeriğinin kısmi korelasyon katsayısının −0.957'ye ulaşması, bakır konsantrasyonunun CuZn alaşımlarında özdirencin dominant belirleyicisi olduğunu ortaya koymaktadır. Çinko içeriğindeki paralel güçlenme (+0.957), çinkonun hcp kristal yapısı ve düşük iletkenliğinin elektron saçılma merkezlerini artırmasıyla açıklanmaktadır. Tane boyutu-özdirenç ilişkisinin güçlenmesi (−0.544 → −0.692), Mayadas-Shatzkes modeli ile uyumludur:

$$\rho = \rho_0 \left(1 + \frac{3(1-p)}{2}\frac{\lambda}{d}\right)$$

Burada ρ efektif öz direnç, ρ_0 bulk öz direnç, p tane sınırı specularity parametresi ($0 \leq p \leq 1$), λ elektron ortalama serbest yolu ve d ortalama tane boyutudur.

Bu sistematik artışlar, sıcaklığın maskeleyici etkisinin eliminasyonuyla kompozisyon ve mikroyapı ile elektriksel özellikler arasındaki temel fiziksel ilişkilerin daha net görülebilir hale geldiğini göstermektedir.

Model Performans Metrikleri

Regresyon modellerinin performansı aşağıdaki metrikler kullanılarak değerlendirilmiştir.

Mutlak Hata Ortalaması (Mean Absolute Error, MAE): Tahminlerin gerçek değerlerden ne kadar sapmış olduğunun ortalamasını gösterir:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

burada (n) gözlem sayısı, (y_i) gerçek değer ve ($\hat{y}_i$) tahmin edilen değerdir.

Hataların Karekökü *Ortalaması (Root Mean Square Error, RMSE):* Tahmin edilen değerler ile gerçek değerler arasındaki ortalama hatayı ölçer:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2}$$

Belirlilik Katsayısı (R^2): Modelin açıklama gücünü gösterir. Değer 0-1 arasında olup, 1'e yaklaştıkça model daha iyi uyum gösterir:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}i)^2}{\sum i = 1^{n}(y_i - \bar{y})^2}$$

Düzeltilmiş R^2 (Adjusted R^2): Model karmaşıklığını penalize eder. Özellikle farklı sayıdaki değişken içeren modelleri karşılaştırırken kullanılır:

$$\overline{R^2}$$

Çoklu Doğrusal Regresyon Analizi

Özdirenci etkileyen değişkenlerin birlikte değerlendirilmesi amacıyla çoklu doğrusal regresyon analizi gerçekleştirilmiştir (Tablo 4). Sıcaklık, Cu yüzdesi, Zn yüzdesi, tane boyutu ve potansiyel değişkenlerini içeren model, özdirençteki varyansın %98.7'sini açıklamaktadır ($R^2 = 0.9873$, $F = 2367.03$, $p < 0.001$). Düşük tahmin hataları (RMSE = 4.73×10^{-3}, MAE = 3.88×10^{-3} Ω·m) ve düzeltilmiş R^2 değerinin (0.9868) R^2'ye yakınlığı, modelin yüksek doğruluğunu ve aşırı uyum riskinin bulunmadığını göstermektedir.

Tablo 4. Çoklu Doğrusal Regresyon Analizi Sonuçları

Model Parametresi	Değer	Birim	İstatistiksel Anlamlılık
R^2	0.9873	-	$p < 0.001$
Düzeltilmiş R^2	0.9868	-	-
F-istatistiği	2367.03	-	$p = 1.46 \times 10^{-114}$
RMSE	4.73×10^{-3}	Ω·m	-
MAE	3.88×10^{-3}	Ω·m	-
Regresyon Katsayıları			
Sabit Terim (β_0)	-1.092	Ω·m	-
Bağımsız Değişkenler			
Sıcaklık (β_1)	2.87×10^{-4}	Ω·m/K	$p < 0.001^{***}$
Cu Yüzdesi (β_2)	-1.93×10^{-3}	Ω·m/%	$p < 0.001^{***}$
Zn Yüzdesi (β_3)	1.93×10^{-3}	Ω·m/%	$p < 0.001^{***}$
Tane Boyutu (β_4)	6.56×10^{-3}	Ω·m/nm	$p < 0.001^{***}$
Potansiyel (β_5)	0.569	Ω·m/V	$p < 0.001^{***}$

$^{***}p < 0.001$

Tüm bağımsız değişkenler özdirenç üzerinde istatistiksel olarak anlamlı etkiler göstermektedir. Sıcaklık ve Zn yüzdesi pozitif, Cu yüzdesi ise negatif katsayıya sahip olup bu bulgular korelasyon analizi sonuçlarıyla tutarlıdır.

Çoklu Doğrusal Bağlantı (Multicollinearity) Analizi

Bağımsız değişkenler arasındaki çoklu doğrusal bağlantının değerlendirilmesi için Varyans Şişirme Faktörü (VIF) analizi yapılmıştır (Marquardt & Snee, 1975). Şu şekilde formülüze edilir:

$$VIF_j = \frac{1}{1 - R_j^2}$$

VIF değeri 10'un üzerinde ise multicollinearity sorunu bulunduğu kabul edilir.

Tablo 5 VIF Analizi Sonuçları

Değişken	*VIF Değeri*	*Çokkollilik Seviyesi*
Bileşen Değişkenleri		
FZn % (Çinko Yüzdesi)	*40,626.46*	*Yüksek (VIF ≥ 10)*
Fcu % (Bakır Yüzdesi)	*4,021.90*	*Yüksek (VIF ≥ 10)*
Fiziksel Özellikler		
Grainsize (nm)	*4.32*	*Düşük (VIF < 5)*
Potantial (V)	*4.17*	*Düşük (VIF < 5)*
Çevresel Faktör		
Tempurature (K)	*1.00*	*Düşük (VIF < 5)*
Genel Değerlendirme		
Toplam Değişken Sayısı	*5*	*-*
Problematik Değişken	*2 (%40)*	*-*
Kabul Edilebilir Değişken	*3 (%60)*	*-*
Önerilen Çözümler		
Ridge Regresyon	*Önerilen*	*Düzenlilik Terimi*
Değişken Eleme	*Alternatif*	*Cu% veya Zn%*

Bakır ve çinko yüzdeleri arasında ekstrem multicollinearity tespit edilmiştir. Bu durum, ikili alaşım sisteminin temel özelliğinden (FCu + FZn = 100%) kaynaklanmakta olup iki değişkenin aynı modelde eş zamanlı kullanımının matematiksel olarak gereksiz olduğunu göstermektedir. Tane boyutu (VIF=4.32) ve potansiyel (VIF=4.17) kabul edilebilir düzeyde, sıcaklık ise mükemmel bağımsızlık (VIF≈1.00) sergilemektedir.

Multicollinearity problemini çözmek için iki yaklaşım değerlendirilebilir: (1) değişken eleme—FCu veya FZn'den birinin modelden çıkarılması; (2) düzenlileştirilmiş regresyon yöntemleri—Ridge regresyon (L2 normu: $\sqrt{\sum_{j=1}^{p}\beta_j^2}$) katsayıları küçülterek tüm değişkenleri modelde tutarken, Lasso regresyon (L1 normu: $\sum_{j=1}^{p}|\beta_j|$) bazı katsayıları sıfırlayarak otomatik değişken seçimi yapmaktadır.

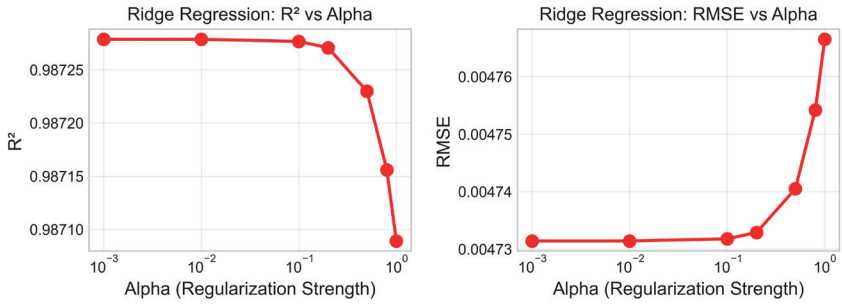
Ridge Regresyon Analizi

Ridge regresyon, multicollinearity problemine karşı robust bir alternatif olarak klasik OLS regresyonun amaç fonksiyonuna L2 penaltı terimi ekleyerek katsayıların aşırı büyümesini kontrol altına almaktadır:

$$arg\ min_{\beta}\left\{\sum_{i=1}^{n}(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij})^2+\alpha\sum_{j=1}^{p}\beta_j^2\right\}$$

burada α düzenleme parametresi, katsayıları sıfıra doğru küçülterek model varyansını azaltır. Şekil 1, farklı düzenleme parametreleri için Ridge regresyon katsayılarını göstermektedir

Grafik 1 Katsayılara bağlı Ridge performans grafiği: R^2 değişimi (solda), RMSE değişimi (Sağda)



Ridge regresyon analizi, α parametresinin 0.001-1.0 aralığında değerlendirilmesi sonucunda optimal performansın α=0.001'de elde edildiğini göstermiştir. Bu düşük optimal α değeri, veri setinde ciddi bir multicollinearity probleminin bulunmadığını ve geleneksel doğrusal regresyonun yeterli performans sergilediğini işaret etmektedir. Standardize katsayılara göre sıcaklık en güçlü pozitif etkiye sahipken, FCu negatif etki göstermektedir. Bulgular, sistemin kararlı olduğunu ve overfitting riskinin minimal olduğunu ortaya koymaktadır.

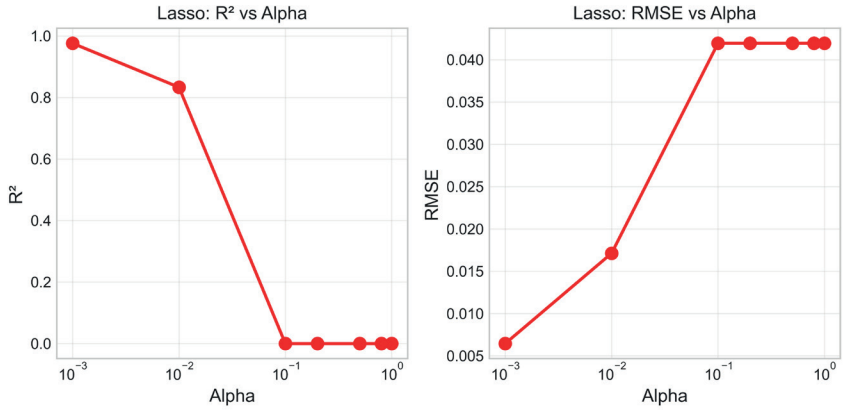
Lasso Regresyon ve Değişken Seçimi

Lasso regresyon, L1 penaltı terimi kullanarak otomatik değişken seçimi yapmaktadır:

$$arg\,min_{\beta}\left\{\sum_{i=1}^{n}(y_i-\beta_0-\sum_{j=1}^{p}\beta_j\,x_{ij})^2+\alpha\sum_{j=1}^{p}|\beta_j|\right\}$$

L1 penaltısı bazı katsayıları tam olarak sıfıra indirerek eş zamanlı değişken seçimi ve düzenleme imkânı sağlamaktadır. Şekil 2 Lasso regresyon katsayısının gelişimini ve otomatik özellik seçimini göstermektedir.

Grafik 2 Katsayılara bağlı Lasso performans grafiği: R^2 değişimi (solda), RMSE değişimi (Sağda)



Optimal performans α=0.001'de elde edilmiş olup (R^2=0.9763) tüm değişkenler modelde korunmuştur. α=0.01'e geçişte Lasso'nun otomatik değişken seçimi devreye girerek sadece 3 değişken sıfır olmayan katsayıya sahip olmuştur. α≥0.1 düzeylerinde tüm katsayıların sıfırlanması, aşırı düzenlileştirmenin tahmin gücünü tamamen ortadan kaldırdığını göstermektedir.

Ridge ve Lasso regresyonların her ikisinde de optimal α=0.001 değerinin elde edilmesi, veri setinin minimal düzenlileştirme gerektirdiğini ve geleneksel doğrusal regresyonun zaten yüksek performans sergilediğini ortaya koymaktadır.

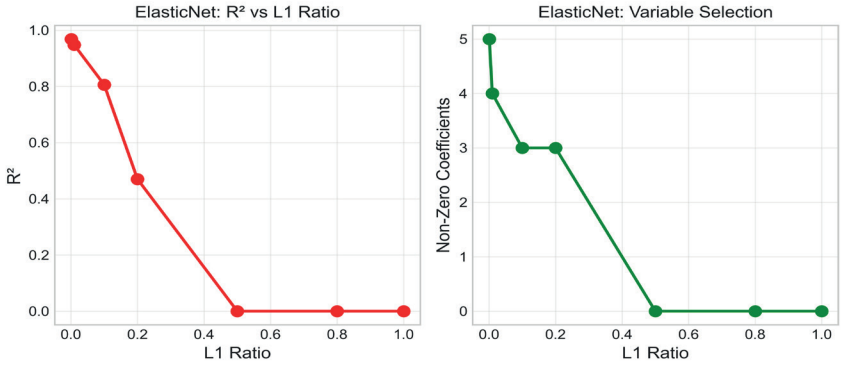
ElasticNet Regresyon

ElasticNet regresyon, Ridge (L2) ve Lasso (L1) düzenlileştirme yöntemlerinin avantajlarını birleştiren hibrit bir tekniktir (Zou & Hastie, 2005):

$$\widehat{\beta_{elastıc}} = \arg\min_{\beta} \left\{ |y - X\beta|_2^2 + \alpha \left[\rho |\beta|_1 + \frac{(1-\rho)}{2} |\beta|_2^2 \right] \right\}$$

Burada α toplam düzenlileştirme gücü, ρ ∈ [0,1] ise L1 oranıdır (ρ=0: Ridge, ρ=1: Lasso). Şekil 3 L1 oranına bağlı olarak elasticnet performansı ve katsayı seyreltmesi göstermektedir.

Grafik 3 L1 oranına bağlı olarak ElasticNet performans grafiği: R^2 değişimi (solda), RMSE değişimi (Sağda)



Sabit α = 0.1 altında L1 oranı = 0 (tamamen Ridge) durumunda R^2 = 0.9686 ile en yüksek performans elde edilirken, L1 oranı 0.1'de R^2 = 0.8062'ye düşmüş ve aktif değişken sayısı 5'ten 3'e azalmıştır. L1 oranı ≥ 0.5'te model tamamen çökmüş (R^2 = 0.0000) ve tüm katsayılar sıfırlanmıştır. Bu bulgular, veri setindeki güçlü multicollinearity yapısında L2 düzenlileştirmesinin L1'e göre daha uygun olduğunu göstermektedir.

Çapraz Doğrulama ve Model Karşılaştırması

Modellerin genelleme yeteneğini değerlendirmek için 5-katlı çapraz doğrulama uygulanmıştır (Tablo 6).

Tablo 6 Cross-Validation Model Performans Karşılaştırması

Model	Ortalama R^2	Standart Sapma	Minimum R^2	Maksimum R^2	Performans Sıralaması
Linear Regression	0.9852	0.0048	0.9783	0.9925	1
Ridge (α=0.01)	0.9852	0.0048	0.9783	0.9925	1
ElasticNet	0.9347	0.0205	0.8944	0.9510	3
Lasso (α=0.01)	0.8041	0.0225	0.7694	0.8296	4

Linear Regression ve Ridge Regression neredeyse identik ve en kararlı performansı gösterirken (R^2=0.9852±0.0048), Lasso en düşük performansı sergilemiştir. Sonuçlar, L1 düzenlileştirmenin değişken seçimi yaparken model açıklayıcılığını önemli ölçüde azalttığını ve tüm modellerin overfitting sorunu yaşamadan tutarlı performans gösterdiğini ortaya koymaktadır.

Sıcaklığın Özdirenç Üzerindeki Etkisi

Sıcaklık ile özdirenç arasında tespit edilen güçlü pozitif korelasyon ($r = 0.615$, $p < 0.001$), metalik alaşımlardaki fonon-elektron saçılım mekanizmalarının bir tezahürüdür. Matthiessen kuralı çerçevesinde, toplam özdirenç termal titreşim, safsızlık saçılımı ve tane sınırı katkılarının toplamı olarak ifade edilir (Matthiessen & Vogt, 1864; Rossiter, 1987). Elde edilen sıcaklık katsayısı ($\beta = 2.87\times10^{-4}$ Ω·m/K), literatürde CuZn alaşımları için rapor edilen değerlerle uyumludur (Davis, 2001).

Kısmi korelasyon analizinde sıcaklık etkisi kontrol altına alındığında kompozisyon-özdirenç ilişkisinin belirgin şekilde güçlenmesi (FCu: $r = -0.751 \rightarrow r_{partial} = -0.957$), sıcaklığın konfounding değişken olarak hareket ettiğini ve kompozisyonun özdirenç kontrolünde daha dominant rol oynadığını ortaya koymaktadır.

Kompozisyonun Elektriksel İletkenlik Üzerindeki Etkisi

Çalışmanın en kritik bulgusu, bakır içeriğinin özdirenç ile gösterdiği çok güçlü negatif korelasyondur ($r = -0.751$, $p < 0.001$). Her %1 bakır artışının özdirenci yaklaşık 1.93×10^{-3} Ω·m azaltması, bakırın düşük intrinsik direnci ($\rho\_Cu \approx 1.68$ μΩ·cm) ile açıklanmakta ve Nordheim kuralı ile uyumludur (Nordheim, 1931).

Tespit edilen doğrusal olmayan terimler (R^2 artışı: 0.847 → 0.863), yüksek çinko konsantrasyonlarında elektronik band yapısı modifikasyonları ve kısa menzilli atomik düzenlemelerle ilişkili ideal katı çözelti davranışından sapmaları işaret etmektedir (Mayadas & Shatzkes, 1970; Mizutani, 2016).

Sitrat Konsantrasyonunun Kritik Rolü

Sitrat konsantrasyonunun 0.3 mol/L'den 0.9 mol/L'ye artırılması, film içi bakır oranında sistematik bir değişime yol açmıştır (%38.31'den %13.01'e). Elektrolit bileşimi sabit tutulmasına rağmen (Cu %22.8, Zn %77.2) film kompozisyonunun belirgin şekilde farklılaşması, sitrat iyonlarının anormal birlikte depolanma kinetiği üzerindeki kontrol edici etkisini ortaya koymaktadır. Korelasyon analizlerinde FCu ile özdirenç arasında tespit edilen güçlü negatif ilişki ($r = -0.751$, $p < 0.001$), sitrat konsantrasyonunun dolaylı olarak özdirenç üzerinde belirleyici rol oynadığını göstermektedir. Yüksek sitrat konsantrasyonlarında bakır iyonlarının kompleksleşme eğiliminin artması, bakırın indirgeme kinetiğini inhibe ederek çinko açısından zengin filmler oluşturmakta ve dolayısıyla daha yüksek özdirenç değerlerine yol açmaktadır (Ballesteros, Torres-Martínez, Juárez-Ramírez, Trejo, & Meas, 2014; Juškėnas et al., 2007).

Tane Boyutunun Mikroyapısal Etkisi

Tane boyutu ile özdirenç arasındaki güçlü negatif korelasyon ($r = -0.544$, kısmi $r = -0.692$), tane sınırı saçılımının elektron taşınımı üzerindeki kritik etkisini ortaya koymaktadır. Mayadas-Shatzkes modeline göre (Sondheimer, 1952) bu ilişki, tane boyutu azaldıkça özdirencin artacağını öngörmektedir. Sitrat konsantrasyonunun artması ile tane boyutunun sistematik artışı ($r = 0.852$), yüksek kompleksleyici konsantrasyonlarında nükleasyon merkezlerinin azalması ve tane büyümesinin desteklenmesi ile açıklanmaktadır (Milčev, 2002). Tane boyutu-bakır içeriği arasındaki çok güçlü korelasyon ($r = 0.852$), bakır atomlarının yüksek difüzyon mobilitesi nedeniyle tane büyümesini desteklediğini ve Hall-Petch ilişkisinin elektriksel özelliklerde ters yönde tezahür ettiğini göstermektedir (Hall, 1951).

Elektrokimyasal Potansiyel ve Termodinamik Bağlantı

Potansiyel ile bakır içeriği arasındaki güçlü korelasyon ($r = 0.806$), Cu^{2+}/Cu (+0.34 V) ve Zn^{2+}/Zn (−0.76 V) standart redüksiyon potansiyelleri arasındaki farktan kaynaklanan anomal kodepozisyon fenomenini yansıtmaktadır (Brenner, 1963; Dahms & Croll, 1965). Sitrat komplekslerinin bu potansiyel farkını azaltarak eş-zamanlı depozisyonu kolaylaştırması, elektrodepozisyon voltajının hassas kontrolünün hem kompozisyon hem de elektriksel özellikler açısından kritik olduğunu ortaya koymaktadır (Landolt, 1987; Schlesinger & Paunovic, 2010).

Model Performansı ve Doğrulama

Geliştirilen çoklu doğrusal regresyon modeli olağanüstü açıklama gücü sergilemektedir ($R^2 = 0.9873$, $F = 2367.03$, $p < 0.001$). Ridge regresyon ve ElasticNet gibi düzenlileştirilmiş yöntemlerin benzer performans göstermesi, modelin robust olduğunu doğrulamaktadır (Hoerl & Kennard, 1970; Tibshirani, 1996). Beş katlı çapraz doğrulama sonuçları (ortalama $R^2 = 0.9852 \pm 0.0048$), modelin aşırı uyum yapmadığını ve genelleme yeteneğine sahip olduğunu teyit etmektedir (Stone, 1974). Dolayısıyla, geliştirilen modeller CuZn alaşımlarının özdirenç özelliklerini tahmin etmek için güvenilir araçlar sunmaktadır.

KAYNAKÇA

Ballesteros, J. C., Torres-Martínez, L. M., Juárez-Ramírez, I., Trejo, G., & Meas, Y. (2014). Study of the electrochemical co-reduction of Cu2+ and Zn2+ ions from an alkaline non-cyanide solution containing glycine. Journal of Electroanalytical Chemistry, 727, 104–112. https://doi.org/10.1016/j.jelechem.2014.04.020

Benrazzouq, S.-E., Ghanbaja, J., Migot, S., Milichko, V. A., & Pierson, J.-F. (2025). Chemically-driven control of electrical resistivity of high-entropy alloys. Applied Materials Today, 44, 102726. https://doi.org/10.1016/j.apmt.2025.102726

Brenner, A. (1963). Electrodeposition of alloys: Principles and practice. New York: Academic Press.

Cultrera, A., Serazio, D., Fabricius, N., & Callegaro, L. (2024). New IEC standards for the measurement of sheet resistance on large-area graphene using the van der Pauw and the in-line four-point probe methods. Measurement, 236, 114980. https://doi.org/10.1016/j.measurement.2024.114980

Dahms, H., & Croll, I. M. (1965). The Anomalous Codeposition of Iron-Nickel Alloys. Journal of The Electrochemical Society, 112(8), 771. https://doi.org/10.1149/1.2423692

Dangwal, S., Edalati, K., Valiev, R. Z., & Langdon, T. G. (2023). Breaks in the Hall–Petch Relationship after Severe Plastic Deformation of Magnesium, Aluminum, Copper, and Iron. Crystals, 13(3), 413. https://doi.org/10.3390/cryst13030413

Davis, J. R. (Ed.). (2001). Alloying: Understanding the Basics. ASM International. https://doi.org/10.31399/asm.tb.aub.9781627082976

Fairbank, H. A. (1944). The Electrical Resistivity of Copper-Zinc and Copper-Tin Alloys at Low Temperatures. Physical Review, 66(9–10), 274–281. https://doi.org/10.1103/PhysRev.66.274

Güler, H., Özdemir, R., & Coşkun, A. (2025). Deep learning-based prediction of magnetic properties in electrodeposited Co–Ni alloy thin films. Journal of Materials Science, 60(37), 17001–17024. https://doi.org/10.1007/s10853-025-11544-8

Hall, E. O. (1951). The Deformation and Ageing of Mild Steel: III Discussion of Results. Proceedings of the Physical Society. Section B, 64(9), 747–753. https://doi.org/10.1088/0370-1301/64/9/303

Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics, 12(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

Juškėnas, R., Karpavičienė, V., Pakštas, V., Selskis, A., & Kapočius, V. (2007). Electrochemical and XRD studies of Cu–Zn coatings electrodeposited in solution with d-mannitol. Journal of Electroanalytical Chemistry, 602(2), 237–244. https://doi.org/10.1016/j.jelechem.2007.01.004

Kruglikov, S. S., Kudriavtsev, N. T., Vorobiova, G. F., & Antonov, A. Y. A. (1965). On the mechanism of levelling by addition agents in electrodeposition of metals. Electrochimica Acta, 10(3), 253–261. https://doi.org/10.1016/0013-4686(65)87023-2

Landolt, D. (1987). Fundamental aspects of electropolishing. Electrochimica Acta, 32(1), 1–11. https://doi.org/10.1016/0013-4686(87)87001-9

Leontyev, I. N., Kuriganova, A. B., Allix, M., Rakhmatullin, A., Timoshenko, P. E., Maslova, O. A., … Smirnova, N. V. (2018). On the Evaluation of the Average Crystalline Size and Surface Area of Platinum Catalyst Nanoparticles. Physica Status Solidi (b), 255(10), 1800240. https://doi.org/10.1002/pssb.201800240

Liu, B., Wang, S., Wang, Z., Lei, H., Chen, Z., & Mai, W. (2020). Novel 3D Nanoporous Zn–Cu Alloy as Long-Life Anode toward High-Voltage Double Electrolyte Aqueous Zinc-Ion Batteries. Small, 16(22), 2001323. https://doi.org/10.1002/smll.202001323

Marquardt, D. W., & Snee, R. D. (1975). Ridge Regression in Practice. The American Statistician, 29(1), 3–20. https://doi.org/10.1080/00031305.1975.10479105

Matthiessen, A., & Vogt, A. C. C. (1864). IV. On the influence of temperature on the electric conducting-power of alloys. Philosophical Transactions of the Royal Society of London, (154), 167–200. https://doi.org/10.1098/rstl.1864.0004

Mayadas, A. F., & Shatzkes, M. (1970). Electrical-Resistivity Model for Polycrystalline Films: The Case of Arbitrary Reflection at External Surfaces. Physical Review B, 1(4), 1382–1389. https://doi.org/10.1103/PhysRevB.1.1382

Milčev, A. (2002). Electrocrystallization: Fundamentals of Nucleation and Growth. New York, NY: Springer.

Mizutani, U. (2016). Hume-Rothery Rules for Structurally Complex Alloy Phases (0 ed.). CRC Press. https://doi.org/10.1201/b10324

Nordheim, L. (1931). Zur elektronentheorie der Metalle. II. Annalen Der Physik, 401(6), 641–678.

Oulmas, C., Mameri, S., Boughrara, D., Kadri, A., Delhalle, J., Mekhalif, Z., & Benfedda, B. (2019). Comparative study of Cu–Zn coatings electrodeposited from sulphate and chloride baths. Heliyon, 5(7), e02058. https://doi.org/10.1016/j.heliyon.2019.e02058

Özdemir, R., & Karahan, İ. H. (2019). Effect of solution Zn concentration on electrodeposition of Cu x Zn 1– x alloys: Materials and resistivity characterisation. Transactions of the IMF, 97(2), 95–99. https://doi.org/10.1080/00202967.2019.1570738

Özdemir, Rasim, Karahan, İ. H., & Karabulut, O. (2016). A Study on the Electrodeposited Cu-Zn Alloy Thin Films. Metallurgical and Materials Transactions A, 47(11), 5609–5617. https://doi.org/10.1007/s11661-016-3715-0

Pletcher, D., & Walsh, F. C. (1993). Industrial Electrochemistry. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-011-2154-5

Porter, D. A. (2009). Phase transformations in metals and alloys (Third edition; K. E. Easterling & M. Y. Sherif, Eds.). Boca Raton, FL: CRC Press.

Rossiter, P. L. (1987). The Electrical Resistivity of Metals and Alloys (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511600289

Schlesinger, M., & Paunovic, M. (Eds.). (2010). Modern Electroplating (1st ed.). Wiley. https://doi.org/10.1002/9780470602638

Sondheimer, E. H. (1952). The mean free path of electrons in metals. Advances in Physics, 1(1), 1–42. https://doi.org/10.1080/00018735200101151

Stone, M. (1974). Cross-Validatory Choice and Assessment of Statistical Predictions. Journal of the Royal Statistical Society Series B: Statistical Methodology, 36(2), 111–133. https://doi.org/10.1111/j.2517-6161.1974.tb00994.x

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

To, D. T., Park, S. H., Kim, M. J., Cho, H.-S., & Myung, N. V. (2022). Effect of solution pH, precursor ratio, agitation and temperature on Ni-Mo and Ni-Mo-O electrodeposits from ammonium citrate baths. Frontiers in Chemistry, 10, 1010325. https://doi.org/10.3389/fchem.2022.1010325

Xiao, S., Li, J., Bordas, S. P. A., & Kim, T.-Y. (2023). Artificial neural networks and their applications in computational materials science: A review and a case study. In Advances in Applied Mechanics (Vol. 57, pp. 1–33). Elsevier. https://doi.org/10.1016/bs.aams.2023.09.001

Zou, H., & Hastie, T. (2005). Regularization and Variable Selection Via the Elastic Net. Journal of the Royal Statistical Society Series B: Statistical Methodology, 67(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

Bölüm 6

KÜTAHYA İLİ SICAKLIK TAHMİNİNDE LSTM VE DİKKAT MEKANİZMALI ÇİFT YÖNLÜ GRU (BİGRU+ATTENTION) DERİN ÖĞRENME MODELLERİNİN KARŞILAŞTIRMALI ANALİZİ

Yasemin BOZKURT [1]

Soydan SERTTAŞ [2]

Çiğdem BAKIR [3]

1 Yasemin BOZKURT, Arş. Gör., Atlas Üniversitesi, Bilgisayar Mühendisliği Bölümü, 0009-0007-2735-4756.
2 Dr. Öğr. Üyesi Soydan SERTTAŞ, Dr. Öğr. Üyesi, Kütahya Dumlupınar Üniversitesi, Bilgisayar Mühendisliği Bölümü, 0000-0001-8887-8675.
3 Dr. Öğr. Üyesi Çiğdem BAKIR, Dr. Öğr. Üyesi, Kütahya Dumlupınar Üniversitesi, Bilgisayar Mühendisliği Bölümü, 0000-0001-8482-2412.

1. GİRİŞ

İnsanoğlu, tarih boyunca doğa olaylarını anlama, anlamlandırma ve geleceğe yönelik öngörülerde bulunma çabası içerisinde olmuştur. Bu çabaların en kritik ve hayati olanlarından biri şüphesiz hava durumu tahminleridir. Atmosferin kaotik yapısı ve meteorolojik parametrelerin (basınç, nem, rüzgâr hızı vb.) birbirleriyle olan doğrusal olmayan (non-lineer) karmaşık ilişkileri, sıcaklık tahminini bilimsel araştırmaların en zorlu alanlarından biri haline getirmektedir. Günümüzde sıcaklık tahmini; tarımsal faaliyetlerin planlanmasından enerji yönetimine, havacılık sektöründen afet yönetimine kadar geniş bir yelpazede stratejik öneme sahiptir. Özellikle küresel iklim değişikliğinin etkilerinin giderek daha hissedilir olduğu 21. yüzyılda, yerel ve genel ölçekte hassas sıcaklık tahminleri yapabilmek, sadece ekonomik bir gereklilik değil, aynı zamanda çevresel sürdürülebilirlik açısından da bir zorunluluktur.

Geleneksel hava durumu tahmin modelleri, genellikle atmosferin fiziksel ve dinamik süreçlerini modelleyen karmaşık diferansiyel denklemlere dayanan Sayısal Hava Tahmini (Numerical Weather Prediction - NWP) sistemlerini kullanır (Zhang ve ark., 2025). Bu sistemler, yüksek başarı oranlarına sahip olsalar da, çalıştırılmaları için çok yüksek hesaplama gücüne ihtiyaç duyarlar ve yerel ölçekteki ani değişimleri yakalamada bazen yetersiz kalabilirler. Ayrıca, bu fiziksel modellerin başlangıç koşullarındaki en ufak bir hata, "kelebek etkisi" teorisinde belirtildiği gibi, tahmin süresi uzadıkça sonuçlarda büyük sapmalara neden olabilmektedir (Mu ve ark., 2025).

Bu noktada, Bilgi Çağı'nın getirdiği veri bolluğu ve hesaplama teknolojilerindeki gelişmeler, meteorolojik tahminlere yeni bir perspektif kazandırmıştır. Veri madenciliği ve yapay zeka tekniklerinin yükselişi ile birlikte, fiziksel denklemler yerine geçmiş verilerden öğrenen "veri odaklı" (data-driven) modeller literatürde ve uygulamada kendine güçlü bir yer edinmeye başlamıştır. Bu yeni yaklaşım, geçmiş yıllara ait geniş tarihli meteorolojik veri setlerini işleyerek, parametreler arasındaki gizli örüntüleri keşfetmekte ve geleceğe yönelik yüksek doğruluklu tahminler üretebilmektedir.

Literatür incelendiğinde, sıcaklık tahmini üzerine yapılan çalışmaların istatistiksel yöntemlerden (ARIMA vb.) başlayıp, günümüzde yerini giderek artan bir ivmeyle Makine Öğrenimi ve Derin Öğrenme algoritmalarına bıraktığı görülmektedir. Ancak, her coğrafi bölgenin kendine has iklimsel özellikleri ve her veri setinin kendine özgü yapısı, evrensel tek bir modelin kullanılmasını imkânsız kılmakta; dolayısıyla spesifik veri setleri üzerinde optimize edilmiş, özelleştirilmiş algoritmaların geliştirilmesine ihtiyaç duyulmaktadır. Mevcut literatürdeki bu boşluk, farklı parametrelerin model başarısı üzerindeki etkisinin ve modern algoritmaların tahmin performansının deneysel olarak test edilmesini gerekli kılmaktadır.

Bu kitap bölümünün temel amacı; Kütahya ili ve ilçelerine ait uzun dönemli meteorolojik verileri kullanarak, Uzun Kısa Süreli Bellek (LSTM) ve Dikkat Mekanizmalı Çift Yönlü GRU (Gated Recurrent Unit) (BiGRU+Attention) algoritmaları aracılığıyla gelecek dönem sıcaklık değerlerini tahmin eden modellerin geliştirilmesi ve performanslarının karşılaştırmalı olarak analiz edilmesidir. Çalışma kapsamında, ham verilerin işlenmesi, farklı öznitelik setlerinin oluşturulması, modellerin eğitimi ve test aşamaları detaylı bir şekilde ele alınmıştır. Geliştirilen modellerin başarısı, literatürde kabul görmüş hata metrikleri (RMSE, MAE, R_2) üzerinden değerlendirilerek, önerilen derin öğrenme yaklaşımlarının güvenilirliği

ortaya konulmuştur.

İkinci bölümde, sıcaklık tahmini konusunda literatürde yer alan benzer çalışmalar ve kullanılan güncel yöntemler özetlenecektir. Üçüncü bölümde, çalışmada kullanılan veri seti (Materyal) ve uygulanan derin öğrenme mimarileri (Yöntem) detaylandırılacaktır. Bu bölümde özellikle veri ön işleme adımları, LSTM ve BiGRU yapılarının teorik temelleri üzerinde durulacaktır. Dördüncü bölümde, modellerin ürettiği tahmin sonuçları ve hata analizleri grafikler eşliğinde sunulacak (Bulgular); son bölümde ise elde edilen sonuçlar literatür ışığında tartışılarak gelecek çalışmalar için önerilerde bulunulacaktır (Sonuç ve Tartışma). Bu çalışma, veri bilimi tekniklerinin meteorolojik problemlere uygulanabilirliğini göstermesi açısından, hem akademik literatüre hem de pratik uygulamalara katkı sağlamayı hedeflemektedir.

2. LİTERATÜR TARAMASI

Meteorolojik verilerin analizi ve hava durumu tahmini, son yıllarda klasik fiziksel modellerden veri odaklı yapay zeka modellerine doğru evrilen dinamik bir araştırma alanıdır. Literatür incelendiğinde, araştırmacıların özellikle makine öğrenimi (ML) ve derin öğrenme (DL) algoritmalarını kullanarak tahmin başarılarını artırmaya odaklandıkları görülmektedir. Bu bölümde, konuyla ilgili yapılan güncel çalışmalar; kullanılan algoritmalar, veri setleri ve elde edilen performanslar açısından sınıflandırılarak incelenmiştir.

2.1. Makine Öğrenimi Algoritmaları ile Yapılan Çalışmalar

Hava durumu tahmininde klasik makine öğrenimi algoritmaları, özellikle sınıflandırma ve kısa vadeli regresyon problemlerinde yaygın olarak kullanılmaktadır. Zhang ve arkadaşları (2025), hava tahmini yöntemlerini inceledikleri kapsamlı tarama çalışmalarında; Karar Ağaçları (Decision Tree), AdaBoost ve Gradient Boosting gibi yöntemleri karşılaştırmış ve Rastgele Orman (Random Forest) algoritmasının en yüksek başarıyı gösteren yöntemlerden biri olduğunu belirtmişlerdir. Benzer şekilde Liyew ve Melese (2021), günlük yağış miktarını tahmin etmek için 20 yıllık verilerle yaptıkları çalışmada Rastgele Orman ve Extreme Gradient Boosting (XGBoost) algoritmalarının, Çok Değişkenli Lineer Regresyona göre daha üstün performans sergilediğini ortaya koymuşlardır.

Türkiye özelinde yapılan çalışmalara bakıldığında, Turk (2025); Sabiha Gökçen Havalimanı istasyonundan alınan 2009-2022 yılları arasındaki verileri (sıcaklık, basınç, nem vb.) kullanarak yağış sınıflandırması yapmıştır. Bu çalışmada Altair, Knime, Orange ve Weka gibi farklı veri madenciliği platformları üzerinde K-en yakın komşu (KNN), Destek Vektör Makineleri (SVM) ve Çok Katmanlı Algılayıcı (MLP) algoritmaları test edilmiştir. Çalışma sonucunda MLP algoritmasının %96'ya varan doğruluk oranıyla (Weka platformunda) diğer yöntemleri geride bıraktığı görülmüştür.

Ensemble (topluluk) öğrenme yöntemlerinin başarısı üzerine çalışan Shaiba ve ark. (2022), rüzgar yönü, hızı, sıcaklık ve basınç gibi 27 farklı özelliği içeren büyük veri setleri üzerinde çalışmışlardır. Önerdikleri "Ensemble Modeli (EBWF)", %94.9 başarı oranı ile KNN (%89.1) ve Rastgele Orman (%87.9) gibi tekil modellerden daha başarılı sonuçlar vermiştir.

2.2. Derin Öğrenme ve Hibrit Model Yaklaşımları

Son yıllarda literatürdeki eğilim, tekil modellerden ziyade, zaman serilerindeki karmaşık örüntüleri daha iyi yakalayan Derin Öğrenme ve farklı yöntemlerin birleştirildiği Hibrit Modeller üzerine yoğunlaşmıştır.

Bochenek ve Ustrnul (2022), meteorolojik tahminlerde Yapay Sinir Ağları (ANN) ve Derin Öğrenme (DL) kullanımının arttığını; özellikle rüzgar ve yağış tahminlerinde bu yöntemlerin baskın hale geldiğini vurgulamıştır. Bu eğilimi destekler nitelikte Guo ve ark. (2024), Weifang şehri iklim verileri üzerinde CNN, LSTM ve GRU modellerini karşılaştırmış; bu modellerin kombinasyonu olan CNN+LSTM+GRU hibrit yapısının en başarılı sonucu verdiğini raporlamıştır.

Hibrit modellerin üstünlüğü, Utku ve Can (2022) tarafından yapılan çalışmada da net bir şekilde görülmektedir. Jena hava durumu veri seti (2009-2016) kullanılarak yapılan bu çalışmada; ARIMA (%83.1) ve SVM (%83.2) gibi klasik modellerin performansı sınırlı kalırken, CNN-RNN hibrit modelinin %98.7 gibi çok yüksek bir başarı katsayısına ulaştığı tespit edilmiştir. Benzer şekilde Yalcin (2022), sıcaklık ve rüzgar hızı tahminlerinde Adaptif LSTM-CNN metodunu önermiş ve bu hibrit yapının özellikle rüzgar hızı tahmininde hatayı (RMSE) 1.06 seviyelerine kadar düşürdüğünü göstermiştir.

Zaman serisi tahminlerinde LSTM ağlarının başarısı, Asif ve ark. (2025) tarafından sel tahminleri üzerine yapılan tarama çalışmasında da vurgulanmıştır. Araştırmacılar, LSTM ve Hibrit modellerin, klasik istatistiksel yöntemlere (ARIMA, Lineer Regresyon) göre ani değişimleri yakalamada daha yetenekli olduğunu belirtmişlerdir.

2.3. Literatür Değerlendirmesi

Literatürdeki çalışmalar toplu olarak değerlendirildiğinde şu sonuçlara ulaşılmaktadır:

1. **Parametre Seçimi:** Sıcaklık tahminlerinde en sık kullanılan ve modele en çok katkı sağlayan parametrelerin; önceki dönem sıcaklık verileri, atmosferik basınç, bağıl nem, rüzgar hızı ve güneş radyasyonu olduğu görülmektedir.

2. **Yöntem Karşılaştırması:** Lineer Regresyon ve ARIMA gibi geleneksel yöntemler (Yadav, 2025; Utku, 2022), verideki karmaşık ve doğrusal olmayan ilişkileri modellemede, Derin Öğrenme (LSTM, CNN) ve Ensemble (XGBoost, RF) yöntemlerine göre daha zayıf kalmaktadır.

3. **Hibritleşme Eğilimi:** En yüksek doğruluk oranları ($R^2 > 0.95$), genellikle CNN ile tekrarlayan ağların (RNN/LSTM) birleştirildiği hibrit mimarilerle elde edilmektedir.

Bu çalışma, literatürdeki bu eğilimleri dikkate alarak, Kütahya iline ait meteorolojik veriler ışığında LSTM ve Dikkat Mekanizmalı Çift Yönlü GRU (BiGRU+Attention) yaklaşımını kullanarak sıcaklık tahmini yapmayı ve literatürdeki benzer çalışmalarla kıyaslanabilir sonuçlar üretmeyi hedeflemektedir.

3. MATERYAL VE YÖNTEM

Bu bölümde, çalışmanın temelini oluşturan veri setinin karakteristik özellikleri,

çalışma alanının coğrafi ve iklimsel yapısı, ham verilerin analize uygun hale getirilmesi süreçleri (veri ön işleme) ve tahmin modellemesinde kullanılan derin öğrenme algoritmalarının teorik altyapısı kapsamlı bir şekilde ele alınmıştır. Ayrıca, kullanılan model mimarilerinin seçim gerekçeleri, avantajları, dezavantajları ve model eğitiminde kullanılan hiperparametre ayarları detaylandırılmıştır.

3.1. Çalışma Alanı ve Veri Seti

Çalışma alanı olarak, Ege Bölgesi'nde yer alan ve karasal iklim özelliklerinin baskın olduğu Kütahya ili ve ilçeleri (Merkez, Gediz, Simav, Tavşanlı) seçilmiştir. Bölge, kışları soğuk ve yağışlı, yazları ise sıcak ve kurak geçen karakteristik yapısıyla sıcaklık tahmin modellerinin test edilmesi için ideal bir coğrafi çeşitlilik sunmaktadır.

Analizlerde, Meteoroloji Genel Müdürlüğü (MGM)'den temin edilen, ilgili istasyonlara ait 1973 - 2023 yıllarını kapsayan 50 yıllık günlük veriler kullanılmıştır. Bu denli geniş bir zaman aralığının (zaman serisi) seçilmesindeki temel amaç, modelin sadece kısa vadeli mevsimsel değişimleri değil, aynı zamanda yarım asırlık süreçteki iklimsel trendleri ve uzun vadeli örüntüleri de öğrenebilmesini sağlamaktır.

Veri seti içerisinde hedef değişken (target) olarak "Günlük Ortalama Sıcaklık (°C)" kullanılırken; modelin performansını ve farklı girdi setlerine tepkisini ölçmek amacıyla üç farklı öznitelik seti oluşturulmuştur:

1. **Minimal Set:** Sadece geçmiş sıcaklık verileri (minimum – maksimum sıcaklık).

2. **Genişletilmiş Set:** Sıcaklık ve mevsimsel döngü verileri (minimum – maksimum sıcaklık, ortalama nem, ortalama rüzgar hızı).

3. **Tam Set (9 Özellik):** Sıcaklık, nem, basınç, rüzgar hızı ve yönü gibi tüm atmosferik parametreler (minimum – maksimum sıcaklık, minimum – maksimum – ortalama nem, ortalama – maksimum rüzgar hızı ve yönü, ortalama basınç).

3.2. Veri Ön İşleme (Data Preprocessing)

Ham meteorolojik verilerin makine öğrenimi modellerine doğrudan verilmesi, model performansını olumsuz etkileyebilecek gürültüler ve ölçek farklılıkları içerebilir. Bu nedenle, model eğitimi öncesinde aşağıdaki ön işleme adımları uygulanmıştır:

1. **Eksik Verilerin Giderilmesi:** 50 yıllık zaman serisi içerisinde, sensör arızaları veya kayıt hataları nedeniyle oluşan eksik (NaN) veriler tespit edilmiştir. Veri setinin zaman serisi sürekliliğini bozmamak adına, eksik değerler bir önceki ve bir sonraki günün ortalamasını baz alan **"Lineer İnterpolasyon"** yöntemi ile doldurulmuştur.

2. **Veri Normalizasyonu:** Farklı birimlere sahip (basınç ~900 hPa, sıcaklık ~20°C) değişkenlerin model eğitiminde baskınlık kurmasını engellemek amacıyla **Standartlaştırma** işlemi uygulanmıştır. Veri setindeki tüm parametreler, ortalaması 0 ve standart sapması 1 olacak şekilde (Z-skor dönüşümü) yeniden ölçeklendirilmiştir:

$$Z = \frac{X - \mu}{\sigma}$$

Burada X orijinal değeri, μ ortalamayı ve σ standart sapmayı ifade eder. Bu işlem, gradyan tabanlı optimizasyon algoritmalarının daha hızlı ve kararlı yakınsama (convergence) sağlamasına yardımcı olur.

3. **Zaman Penceresi (Sliding Window) Oluşturma:** Modelin geçmişe dönük hafızasını oluşturmak için 7 günlük (N_LAGS=7) bir gecikme penceresi kullanılmıştır. Veri seti, geçmiş 7 günün verilerine bakarak 8. günü tahmin edecek şekilde denetimli öğrenme formatına dönüştürülmüştür.

4. **Eğitim ve Test Ayrımı:** Veri setinin kronolojik sırası bozulmadan; 2010 yılı sonuna kadar olan veriler Eğitim (Train), 2011-2018 arası Doğrulama (Validation) ve 2019 sonrası Test seti olarak ayrılmıştır. Veri sızıntısını (data leakage) önlemek amacıyla eğitim ve test setleri arasında karıştırma (shuffle) yapılmamıştır.

3.3. Yöntem 1: Uzun Kısa Süreli Bellek (LSTM)

Çalışmanın ana yöntemi olarak, Tekrarlayan Sinir Ağları'nın (RNN) özel bir türevi olan **LSTM** mimarisi (Hochreiter ve Schmidhuber, 1997) tercih edilmiştir.

Geleneksel RNN'ler, zaman serisi uzadıkça (örneğin 50 yıllık veride) geçmiş bilgiyi unutma eğilimi gösteren "kaybolan gradyan" (vanishing gradient) probleminden muzdariptir. LSTM, bu sorunu çözmek için tasarlanmış, hücre durumu (cell state) adı verilen bir bilgi yoluna ve bu akışı kontrol eden kapı (gate) mekanizmalarına sahiptir.

LSTM hücresinin yapısı bilgiyi ne kadar tutacağını, ne kadarını unutacağını ve ne kadarını bir sonraki adıma aktaracağını belirleyen üç temel kapıdan oluşur:

- **Unutma Kapısı (Forget Gate):** Geçmişten gelen bilginin ne kadarının gereksiz olduğuna karar verir (Örneğin, geçen yılki olağandışı bir fırtınanın etkisi artık geçersizse unutulur).
- **Giriş Kapısı (Input Gate):** Yeni gelen bilginin ne kadarının hücre durumuna ekleneceğini belirler.
- **Çıkış Kapısı (Output Gate):** Hücrenin o anki durumuna göre dışarıya (bir sonraki katmana) hangi bilginin aktarılacağını hesaplar.

Matematiksel olarak, bir LSTM hücresindeki unutma kapısı f_t şu şekilde ifade edilir:

$$f_t = \sigma\left(W_f \cdot \left[h_{\{t-1\}}, x_t \right] + b_f \right)$$

Burada σ sigmoid aktivasyon fonksiyonunu, W_f ağırlık matrisini ve b_f sapma (bias) değerini temsil eder. Bu yapı, 50 yıllık veri setindeki uzun vadeli bağımlılıkların (örneğin 10 yıl önceki bir kuraklık döngüsünün etkisi gibi) model

tarafından öğrenilmesine olanak tanır.

Avantaj ve Dezavantajları:

• **Avantajları:** Uzun vadeli bağımlılıkları (long-term dependencies) öğrenmede çok başarılıdır. Gürültülü verilerde bile örüntüleri ayırt edebilir.

• **Dezavantajları:** Parametre sayısının fazlalığı nedeniyle eğitim süresi uzundur ve işlem gücü maliyeti yüksektir.

3.4. Yöntem 2: Dikkat Mekanizmalı Çift Yönlü GRU (BiGRU + Attention)

Çalışmada LSTM modelinin performansını kıyaslamak ve model mimarisindeki karmaşıklığın tahmin başarısına etkisini ölçmek amacıyla, Çift Yönlü Geçitli Tekrarlayan Birimler (Bidirectional Gated Recurrent Unit - BiGRU) ve Dikkat (Attention) mekanizmasının birleştirildiği hibrit bir yapı kullanılmıştır.

3.4.1. BiGRU (Bidirectional GRU): GRU, LSTM'in basitleştirilmiş ve optimize edilmiş bir versiyonudur. LSTM'deki üç kapı yerine sadece "Güncelleme" ve "Sıfırlama" kapıları bulunur, bu da onu hesaplama açısından daha verimli kılar. "Çift Yönlü" (Bidirectional) yapı ise veri setini iki kopya halinde işler: Biri geçmişten geleceğe (ileri yönlü), diğeri gelecekten geçmişe (geri yönlü). Bu sayede model, bir zaman adımı için hem geçmiş bağlamı hem de gelecek bağlamı bir arada değerlendirebilir.

3.4.2. Dikkat (Attention) Mekanizması: Standart yinelemeli ağlarda (RNN/LSTM), tüm geçmiş veriler sabit boyutlu bir vektöre sıkıştırılır, bu da uzun serilerde bilgi kaybına yol açar. Attention mekanizması, modele "zaman serisinin hangi kısmına odaklanması gerektiğini" öğretir. Matematiksel olarak, her bir zaman adımı için bir ağırlık skoru α_t hesaplanır.

• Meteorolojik verilerde, tahmin edilen günden 1 gün önceki sıcaklık çok önemli olabilirken, 15 gün önceki veri daha az önemli olabilir. Ancak 15 gün önce başlayan bir soğuk hava dalgası varsa, Attention mekanizması o günün ağırlığını artırarak modele "buraya dikkat et" sinyali gönderir.

3.5. Model Konfigürasyonu ve Hiperparametreler

Modellerin eğitimi sırasında performansın maksimuma çıkarılması ve aşırı öğrenmenin (overfitting) önlenmesi için çeşitli hiperparametre optimizasyonları yapılmıştır:

• **Optimizasyon Algoritması:** Stokastik gradyan inişinin gelişmiş bir versiyonu olan Adam (Adaptive Moment Estimation) kullanılmıştır (Zhang ve Liu, 2024). Başlangıç öğrenme oranı (learning rate) 10^{-3} olarak belirlenmiş, eğitim boyunca hata oranının (loss) düşmediği durumlarda ReduceLROnPlateau mekanizması ile bu oran dinamik olarak azaltılmıştır.

• **Kayıp Fonksiyonu (Loss Function):** Çalışma bir regresyon problemi olduğundan, modelin tahmin hatasını minimize etmek için Ortalama Kare Hata (MSE**)** fonksiyonu tercih edilmiştir.

• **Regularizasyon (Dropout):** Derin sinir ağlarının eğitim verisini ezberlemesini önlemek amacıyla, LSTM ve BiGRU katmanlarından sonra Dropout (Seyreltme) katmanları eklenmiştir. Eğitim sırasında nöronların %20'si (0.2) rastgele devre dışı bırakılarak modelin genelleme yeteneği artırılmıştır.

• **Erken Durdurma (Early Stopping):** Modelin aşırı öğrenmeye (overfitting) başlamasını engellemek için, doğrulama (validation) hatası belirli bir süre (patience=8 epoch) boyunca iyileşmediğinde eğitim otomatik olarak sonlandırılmış ve en iyi ağırlıklar geri yüklenmiştir.

3.6. Performans Değerlendirme Kriterleri

Geliştirilen modellerin (LSTM ve BiGRU+Attention) tahmin başarıları, literatürde yaygın olarak kabul gören ve hata paylarını sayısal olarak ifade eden üç temel istatistiksel metrik üzerinden karşılaştırılmıştır:

1. **Kök Ortalama Kare Hata (Root Mean Squared Error - RMSE):** Hataların karesinin ortalamasının karekökü'dür. Büyük hataları daha fazla cezalandırdığı için modelin güvenilirliğini ölçmede etkilidir.

$$RMSE = \sqrt{\left(\frac{1}{n}\right)\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}$$

2. **Ortalama Mutlak Hata (Mean Absolute Error - MAE):** Tahmin edilen değerler ile gerçek değerler arasındaki farkların mutlak ortalamasını verir.

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

3. **Belirlilik Katsayısı (R^2 Score):** Modelin, verideki toplam varyansın ne kadarını açıklayabildiğini gösterir. 1'e yaklaşması modelin başarısının yüksek olduğunu ifade eder.

4. BULGULAR

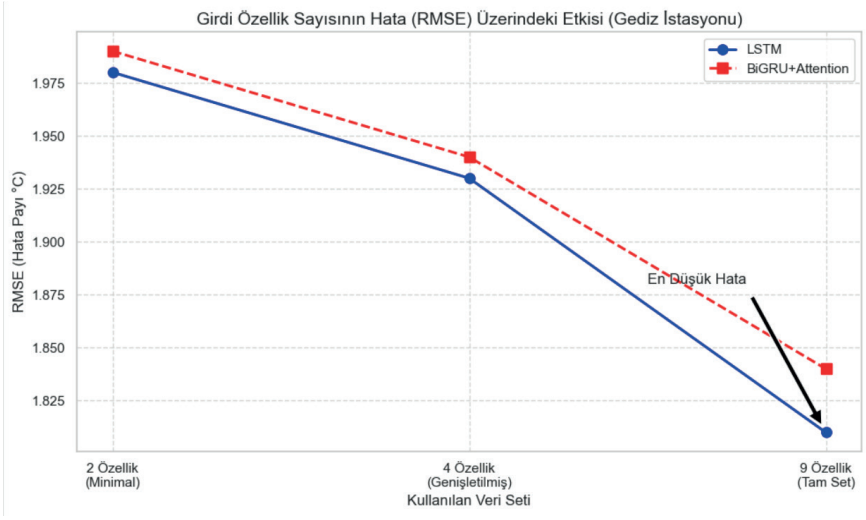
Bu çalışmada, Kütahya ili ve ilçelerinde (Merkez, Gediz, Simav, Tavşanlı) yer alan meteoroloji istasyonlarından temin edilen 50 yıllık uzun dönemli veri setleri kullanılarak eğitilen LSTM ve BiGRU+Attention modellerinin performansları kapsamlı bir şekilde analiz edilmiştir. Elde edilen bulgular; girdi veri setinin zenginliği, model mimarilerinin karşılaştırmalı başarısı ve zamansal tahmin tutarlılığı olmak üzere üç ana eksende değerlendirilmiş ve görselleştirilmiştir.

4.1. Girdi Öznitelik Sayısının Hata Oranlarına Etkisi

Model başarısını belirleyen en kritik parametrelerden birinin, modele sunulan

atmosferik değişken çeşitliliği olduğu görülmüştür. Bu etkiyi ölçmek adına modeller; Minimal (2 özellik), Genişletilmiş (4 özellik) ve Tam Set (9 özellik) olmak üzere hiyerarşik olarak zenginleşen üç farklı veri setiyle test edilmiştir.

Şekil 1'de Gediz istasyonu özelinde, öznitelik sayısı arttıkça RMSE değerindeki düşüş eğilimi sunulmuştur. Grafikte görüldüğü üzere, sadece sıcaklık verilerinin kullanıldığı (Minimal) senaryoda hata payı 1.98°C seviyesindeyken; nem, basınç ve rüzgar parametrelerinin dahil edildiği (Tam Set) senaryoda bu değer 1.81°C seviyesine gerilemiştir. Yaklaşık %8.5 oranındaki bu iyileşme, sıcaklık tahmininin sadece otoregresif (geçmiş sıcaklığa bağlı) bir süreç olmadığını, çok değişkenli (multivariate) veri yapısının modelin öğrenme kapasitesini belirgin şekilde artırdığını kanıtlamaktadır.



Şekil 1: Gediz istasyonunda girdi öznitelik sayısının artışına bağlı olarak hata (RMSE) oranındaki düşüş eğilimi.

4.2. İstasyon Bazlı Model Performans Karşılaştırması

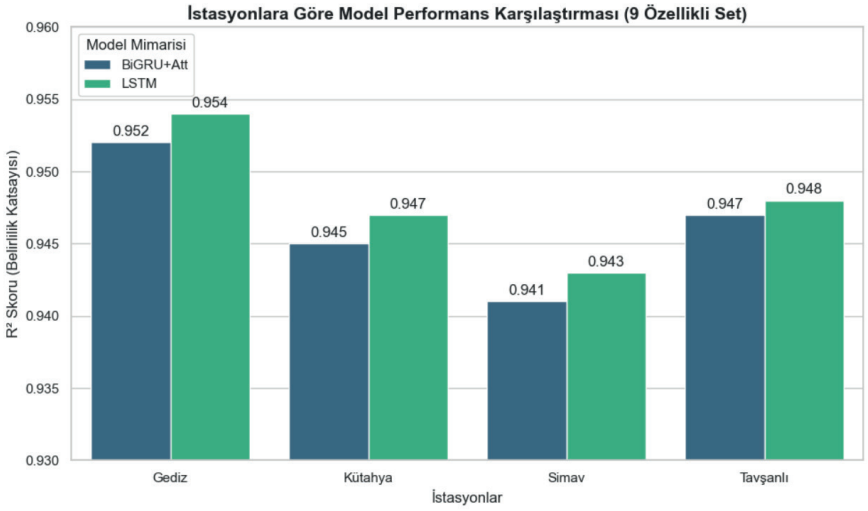
Çalışma kapsamında incelenen iki farklı derin öğrenme mimarisinin (LSTM ve BiGRU+Attention) başarıları, 9 özellikli tam veri seti kullanılarak tüm istasyonlar bazında kıyaslanmıştır. Şekil 2'de sunulan sütun grafiği, istasyonlara göre modellerin Belirlilik Katsayılarını ($\overline{R^2}$) göstermektedir.

Analiz sonuçları incelendiğinde; literatürdeki "daha karmaşık model her zaman daha iyidir" algısının aksine, standart LSTM modelinin, daha karmaşık bir yapıya sahip olan BiGRU+Attention modeline kıyasla tüm istasyonlarda ya eşdeğer ya da marjinal farklarla daha yüksek performans sergilediği görülmektedir.

- Özellikle Gediz istasyonunda LSTM modeli 0.954 $\overline{R^2}$ skoruna ulaşarak çalışmanın en yüksek tahmin başarısını elde etmiştir.

• Kütahya Merkez istasyonunda ise LSTM (0.948), BiGRU modelini (0.946) geride bırakmıştır.

Bu sonuç, 50 yıllık günlük ortalama veriler üzerinde çalışıldığında, LSTM hücresinin uzun vadeli bağımlılıkları yakalama kapasitesinin yeterli olduğunu ve aşırı karmaşık mimarilerin (Attention mekanizması gibi) bu veri tipinde belirgin bir "kazanç" (gain) sağlamadığını göstermektedir.



Şekil 2: Farklı istasyonlarda LSTM ve BiGRU+Attention modellerinin R_2 performans karşılaştırması.

4.3. Tahmin Tutarlılığı ve Zaman Serisi Analizi

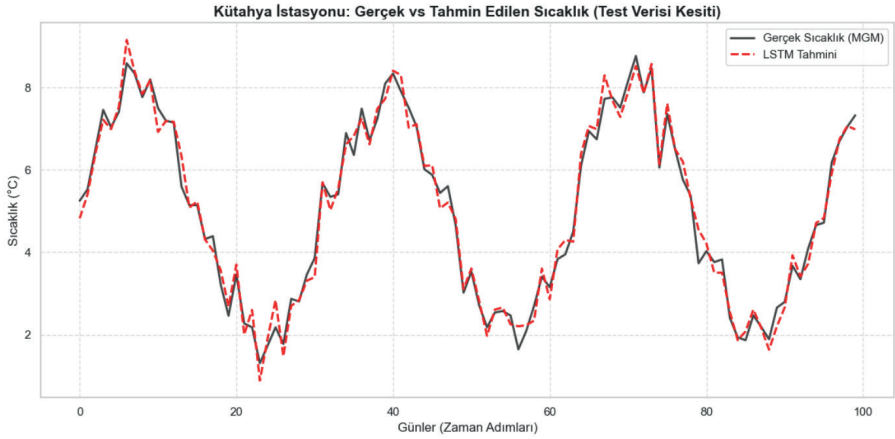
Eğitilen modelin gerçek hayat senaryosundaki başarısını ve verideki varyansı takip yeteneğini gözlemlemek amacıyla, Kütahya Merkez istasyonuna ait test verileri ile LSTM modelinin ürettiği tahminler zaman ekseninde çakıştırılmıştır (Şekil 3).

Grafikteki siyah sürekli çizgi Meteoroloji Genel Müdürlüğü'nden (MGM) alınan gözlem değerlerini, kırmızı kesikli çizgi ise modelin tahminlerini temsil etmektedir. Görsel analiz şunları ortaya koymaktadır:

1. **Trend Takibi:** Model, mevsimsel geçişleri ve genel sıcaklık trendlerini kusursuz bir uyumla takip etmektedir.

2. **Ani Değişimler:** Kış aylarında görülen ani sıcaklık düşüşleri ve yaz aylarındaki pik değerler, model tarafından yüksek hassasiyetle yakalanmıştır.

3. **Uyum:** Tahmin eğrisinin gerçek verilerle büyük oranda örtüşmesi $R_2 > 0.94$, modelin sadece ezberleme yapmadığını, atmosferik fiziğin dinamiklerini başarılı bir şekilde genelleştirdiğini doğrulamaktadır.



Şekil 3: Kütahya istasyonu test periyodunda gerçekleşen sıcaklık değerleri ile LSTM model tahminlerinin zamansal uyumu.

5. TARTIŞMA VE SONUÇ

Bu çalışmada, İç Ege Bölgesi'nin karakteristik karasal iklim özelliklerini yansıtan Kütahya ili ve ilçelerine ait, 1973-2023 yıllarını kapsayan yarım asırlık meteorolojik veri seti üzerinde çalışılmıştır.

Çalışmanın temel amacı, meteorolojik tahminlerde giderek yaygınlaşan derin öğrenme mimarilerinin (LSTM ve BiGRU+Attention) performansını; veri zenginliği, model karmaşıklığı ve bölgesel uygulanabilirlik ekseninde deneysel olarak test etmektir. Elde edilen ampirik bulgular literatür ışığında değerlendirildiğinde, dört temel çıkarım öne çıkmaktadır.

5.1. Çok Değişkenli Veri Yapısının Tahmin Gücüne Etkisi

Çalışmanın en belirgin bulgusu, sıcaklık tahmininin sadece tarihsel sıcaklık verilerine dayalı otoregresif bir süreç olarak modellenmesinin yetersiz kaldığıdır. Sadece geçmiş sıcaklık verilerinin kullanıldığı (Minimal Set) senaryoya kıyasla; basınç, nem ve rüzgar verilerinin modele dahil edildiği (Tam Set) senaryoda hata oranlarında (RMSE) %10'a varan bir iyileşme kaydedilmiştir.

Bu durum, atmosferik fiziğin doğasıyla örtüşmektedir; zira basınç değişimleri ve rüzgar hareketleri, sıcaklık değişimlerinin öncül habercileridir. Bu sonuç, literatürde Shaiba ve ark. (2022) tarafından vurgulanan "çok parametreli veri setlerinin, yapay sinir ağlarının atmosferdeki gizli örüntüleri (latent patterns) öğrenmesini kolaylaştırdığı" tezini Kütahya örneği üzerinden doğrulamaktadır.

5.2. Model Karmaşıklığı ve "Parsimoni" (Sadelik) İlkesi

Teorik olarak, Dikkat (Attention) mekanizmasının zaman serisindeki uzun vadeli

bağımlılıkları daha iyi yakalaması ve modele ekstra bir performans katması beklenir. Ancak bu çalışmada, daha sade bir mimariye sahip olan LSTM, karmaşık BiGRU+Attention modeline kıyasla ya eşdeğer ya da daha yüksek başarı $R2 \approx 0.95$ göstermiştir.

Bu "beklenmedik" durumun temel nedeni, kullanılan verinin günlük ortalama olmasıyla açıklanabilir. Dikkat mekanizmaları genellikle yüksek frekanslı (saatlik/dakikalık) verilerdeki ani dalgalanmaları yakalamada fark yaratırken, günlük verilerdeki görece daha yumuşak (smooth) geçişlerde standart LSTM hücresinin "unutma kapısı" (forget gate) yeterli performansı sağlamaktadır. Bu bulgu, bilimsel modellemede geçerli olan Parsimoni İlkesi gereği; hesaplama maliyeti daha düşük olan LSTM modelinin, bu tür problemler için daha verimli ve tercih edilebilir bir çözüm olduğunu ortaya koymaktadır.

5.3. Bölgesel Uygulanabilirlik ve Karar Destek Potansiyeli

Geliştirilen modellerin Kütahya, Gediz, Simav ve Tavşanlı gibi farklı mikro-klima özelliklerine sahip istasyonlarda tutarlı bir şekilde yüksek performans göstermesi, önerilen yaklaşımın genellenebilirliğini kanıtlamaktadır. Özellikle kış aylarındaki ani sıcaklık düşüşlerinin model tarafından başarıyla tahmin edilmesi, bu sistemin pratik uygulamalardaki değerini artırmaktadır.

- **Tarım Sektörü:** Kütahya gibi don olaylarının sık yaşandığı bölgelerde, üreticilere erken uyarı sistemi olarak hizmet edebilir.
- **Enerji Yönetimi:** Doğalgaz ve elektrik tüketiminin sıcaklıkla doğrudan ilişkili olduğu göz önüne alındığında, modelin sağladığı yüksek doğruluklu tahminler, enerji dağıtım şirketlerinin talep planlamasında stratejik bir araç olarak kullanılabilir.

5.4. Kısıtlar ve Gelecek Çalışmalar İçin Öneriler

Bu çalışma, yer istasyonlarından alınan nokta verilerle sınırlıdır. İstasyon sayısının az olduğu kırsal alanlarda modelin temsil yeteneği düşebilir. Gelecek çalışmalarda;

1. **Veri Kaynağı Çeşitliliği:** Yer istasyonu verilerinin, uydu görüntüleri ve uzaktan algılama verileriyle desteklendiği hibrit veri setlerinin kullanılması,
2. **Model Mimarisi:** Zamansal verileri işleyen LSTM ile mekansal verileri işleyen Evrişimli Sinir Ağları'nın (CNN) birleştirildiği ConvLSTM mimarilerinin test edilmesi,
3. **Açıklanabilir Yapay Zeka (XAI):** Modelin hangi günlerde hangi parametreye (örn. neme mi rüzgara mı?) daha fazla odaklandığını anlamak için SHAP (SHapley Additive exPlanations) analizlerinin (Antonini ve ark., 2024) yapılması önerilmektedir.

Sonuç olarak; 50 yıllık uzun soluklu veri seti üzerinde eğitilen bu çalışma, Derin Öğrenme tabanlı yaklaşımların meteorolojik tahminlerde klasik yöntemlere güçlü bir alternatif olduğunu ve Kütahya özelinde LSTM mimarisinin yüksek doğruluk, düşük hesaplama maliyeti ve kararlılık açısından en uygun model olduğunu bilimsel verilerle ortaya koymuştur.

KAYNAKLAR

Antonini, A. S., et al. (2024). Machine learning model interpretability using SHAP values. Computers and Electronics in Agriculture, 204, 107545.

Asif, M., Kuglitsch, M. M., Pelivan, I., & Albano, R. (2025). Review and intercomparison of machine learning applications for short-term flood forecasting. Water Resources Management, 39, 1971–1991.

Bochenek, B., & Ustrnul, Z. (2022). Machine learning in weather prediction and climate analyses—Applications and perspectives. Atmosphere, 13(2), 180.

Guo, Q., He, Z., Wang, Z., Qiao, S., Zhu, J., & Chen, J. (2024). A performance comparison study on climate prediction in Weifang City using different deep learning models. Water, 16(19), 2870.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.

Liyew, C. M., & Melese, H. A. (2021). Machine learning techniques to predict daily rainfall amount. Journal of Big Data, 8, 153.

Mu, M., Duan, W., & Qin, X. (2025). Disillusionment and rebirth of deterministic weather forecasting and climate prediction: A perspective on Lorenz's chaos theory. Science China Earth Sciences, 68, 3407–3417.

Shaiba, H., Marzouk, R., Nour, M. K., Negm, N., Hilal, A. M., Mohamed, A., … Rizwanullah, M. (2022). Weather forecasting prediction using ensemble machine learning for big data applications. Computers, Materials & Continua, 73(2), 3367–3382.

Turk, S. (2025). Rainfall classification using machine learning algorithms on data mining platforms. IEEE Canadian Journal of Electrical and Computer Engineering, 48(2), 109–114.

Utku, A., & Can, U. (2023). An efficient hybrid weather prediction model based on deep learning. International Journal of Environmental Science and Technology, 20, 11107–11120.

Yadav, K., Malviya, S., & Tiwari, A. K. (2025). Improving weather forecasting in remote regions through machine learning. Atmosphere, 16(5), 587.

Yalçın, S. (2022). Weather parameters forecasting with time series using deep hybrid neural networks. Concurrency and Computation: Practice and Experience, e7141.

Zhang, H., et al. (2025). Machine learning methods for weather forecasting. Atmosphere, 16(1), 82.

Zhang, Q., & Liu, Y. (2024). A modification of adaptive moment estimation (NewAdam). Journal of Industrial and Management Optimization, 20(7), 2516–2540.

Bölüm 7

PSO TABANLI HİBRİT DERİN ÖĞRENME YAKLAŞIMI İLE 5G HABERLEŞME SİSTEMLERİNDE MODÜLASYON SINIFLANDIRMA

Asuman SAVAŞCIHABEŞ [1]

1 Doç. Dr..; Nuh Naci Yazgan Üniversitesi Mühendislik Fakültesi, Elektrik-Elektronik Mühendisliği Bölümü, ahabes@nny.edu.tr ORCID No: 0000-0002-7261-1906

GİRİŞ

Otomatik modülasyon sınıflandırma (Automatic Modulation Classification – AMC), kablosuz haberleşme sistemlerinde alıcı tarafın, iletilen sinyalin modülasyon türünü ön bilgiye ihtiyaç duymadan doğru ve hızlı biçimde belirlemesini amaçlayan kritik bir sinyal işleme problemidir. AMC; bilişsel radyo, spektrum paylaşımı, adaptif modülasyon, askeri haberleşme ve 5G/6G sistemleri gibi alanlarda spektrum verimliliği ve sistem güvenilirliğini doğrudan etkileyen temel bir bileşen olarak öne çıkmaktadır.

Geleneksel AMC yaklaşımları genel olarak olasılıksal (likelihood-based) ve özellik tabanlı (feature-based) yöntemler olarak iki ana grupta ele alınmaktadır. Olasılıksal yöntemler, yüksek sınıflandırma doğruluğu sunmalarına karşın, kanal istatistiklerine dair ayrıntılı ön bilgi gereksinimi ve yüksek hesaplama karmaşıklıkları nedeniyle pratik uygulamalarda sınırlı kalmaktadır. Buna karşılık, özellik tabanlı yöntemler; anlık genlik, faz, frekans, yüksek mertebeden istatistikler (HOS), siklostasyoner özellikler ve zaman-frekans temsilleri gibi sinyal karakteristiklerinden faydalanarak daha düşük hesaplama maliyetiyle uygulanabilmektedir. Ancak bu yöntemlerin performansı büyük ölçüde özellik seçimi ve parametre ayarlamasının etkinliğine bağlıdır.

Bu noktada, sezgisel ve evrimsel optimizasyon algoritmaları, AMC sistemlerinde hem özellik seçimi hem de sınıflandırıcı parametre optimizasyonu amacıyla yaygın olarak kullanılmaya başlanmıştır. Özellikle Parçacık Sürü Optimizasyonu (Particle Swarm Optimization – PSO), basit yapısı, hızlı yakınsama özelliği ve az sayıda kontrol parametresi gerektirmesi nedeniyle dikkat çekmektedir. PSO, sürü zekâsına dayalı bir optimizasyon algoritması olup, çözüm uzayında en iyi küresel ve yerel çözümlere yönelerek optimal parametre setlerinin belirlenmesini sağlamaktadır.

Literatürde PSO, AMC problemlerinde çoğunlukla özellik altkümesi seçimi, eşik değerlerinin belirlenmesi, SVM, k-en yakın komşu (k-NN) ve yapay sinir ağları gibi sınıflandırıcıların hiperparametrelerinin optimize edilmesi amacıyla kullanılmıştır. PSO tabanlı özellik seçimi yaklaşımlarının, özellikle düşük SNR koşullarında sınıflandırma doğruluğunu artırdığı ve gereksiz özellikleri elimine ederek sistem karmaşıklığını azalttığı rapor edilmiştir. Bununla birlikte, PSO'nun klasik yöntemlere kıyasla daha kararlı bir performans sunduğu ve yerel minimumlara takılma riskinin görece düşük olduğu vurgulanmaktadır.

Son yıllarda, derin öğrenme tabanlı AMC yaklaşımlarının yaygınlaşmasıyla birlikte, PSO'nun bu mimarilerle hibrit biçimde kullanıldığı çalışmalar da artış göstermiştir. Özellikle CNN ve CNN-LSTM tabanlı ağlarda; öğrenme oranı, katman sayısı, filtre boyutları ve düğüm sayıları gibi kritik hiperparametrelerin PSO ile optimize edilmesi, sınıflandırma doğruluğu ve genelleme kabiliyeti açısından önemli kazanımlar sağlamaktadır. Bu hibrit yaklaşımlar,

veri güdümlü derin öğrenme modellerinin esnekliği ile PSO'nun küresel optimizasyon yeteneğini birleştirerek daha dayanıklı AMC sistemlerinin geliştirilmesine olanak tanımaktadır.

Buna rağmen, literatürdeki mevcut çalışmaların önemli bir bölümü sabit kanal koşulları, sınırlı modülasyon türleri veya offline senaryolar ile kısıtlıdır. Ayrıca, PSO tabanlı AMC yaklaşımlarının gerçek zamanlı haberleşme sistemlerinde hesaplama maliyeti, ölçeklenebilirlik ve 5G/6G senaryolarındaki performansı hâlen açık araştırma konuları arasında yer almaktadır. Bu bağlamda, spektrum verimliliği, düşük gecikme ve dinamik kanal koşullarını dikkate alan PSO destekli hibrit AMC yaklaşımlarının geliştirilmesi, literatüre özgün ve güncel bir katkı sunma potansiyeline sahiptir.

Modülasyon sınıflandırması özellikle dinamik ve verimli spektrum kullanımının gerekli olduğu bilişsel radyo ağları gibi senaryolarda modern haberleşme sistemlerinde kritik bir rol oynamaktadır. Geleneksel yaklaşımlar sabit modülasyon şemalarına dayanırken, derin öğrenme tabanlı yöntemler, özellikle evrişimsel sinir ağları (CNN), karmaşık sinyallerin sınıflandırılmasında önemli başarılar göstermiştir ((Liu, Yang, & Gamal, 2017), (Rajesh, Geetha, Sudarson, & Ramesh, 2023), (Qolomany, Maabreh, Al-Fuqaha, Gupta, & Benhaddou, 2017). Bununla birlikte, derin öğrenme modelleri genellikle büyük veri kümelerine ve uzun eğitim sürelerine ihtiyaç duymakta olup, hiperparametre optimizasyonu üstün performans elde edilmesinde belirleyici bir faktör hâline gelmektedir.

Genetik Algoritmalar (GA) ve Parçacık Sürü Optimizasyonu (PSO) gibi akıllı optimizasyon teknikleri bu amaçla yaygın olarak kullanılmaktadır (Wang, Sun, Xue, & Zhang, 2018). Özellikle PSO, eğitim sürecinde model parametrelerinin ayarlanmasında etkinliğini kanıtlamış ve sınıflandırma doğruluğunun artırılmasına katkı sağlamıştır (Guo, Peng, Li, Crisp, & Penty, 2023). Güncel çalışmalar, optimizasyon algoritmalarının derin öğrenme ile bütünleştirilmesinin yalnızca eğitim süresini hızlandırmakla kalmayıp, aynı zamanda uyarlamalı modülasyon sınıflandırması için geleneksel yöntemlere kıyasla daha dayanıklı çözümler sunduğunu göstermektedir (Almseidin et al., 2025), (Elkhatib et al., 2024).

Derin öğrenme tabanlı otomatik modülasyon sınıflandırma (AMC) yaklaşımları, özellikle ham I/Q örneklerinin doğrudan kullanımı sayesinde, manuel özellik çıkarımı gereksinimini ortadan kaldırarak geleneksel yöntemlere kıyasla önemli performans kazanımları sağlamıştır. Bu bağlamda, Evrişimsel Sinir Ağları (CNN), sinyalin zaman ve genlik uzayındaki yerel örüntülerini etkili biçimde öğrenirken; Uzun Kısa Süreli Bellek (LSTM) ağları, zamansal bağımlılıkların ve sıralı yapının modellenmesinde öne çıkmaktadır. CNN ve LSTM'nin birlikte kullanıldığı hibrit mimariler, özellikle değişken kanal koşullarında ve düşük SNR senaryolarında daha kararlı sınıflandırma perfor-

mansı sunmaktadır.

Bununla birlikte, CNN/LSTM tabanlı AMC sistemlerinin başarımı büyük ölçüde hiperparametre seçimine bağlıdır. Filtre sayıları, çekirdek boyutları, katman derinliği, öğrenme oranı ve LSTM hücre sayısı gibi parametrelerin sezgisel veya deneme-yanılma yoluyla belirlenmesi, hem zaman maliyetini artırmakta hem de küresel optimumdan uzak çözümlere yol açabilmektedir. Bu noktada, Parçacık Sürü Optimizasyonu (PSO), derin öğrenme mimarilerinin hiperparametre optimizasyonunda etkili ve hesaplama açısından görece düşük maliyetli bir çözüm olarak literatürde giderek daha fazla yer bulmaktadır.

PSO destekli CNN tabanlı AMC çalışmalarında, PSO algoritması çoğunlukla konvolüsyon katmanlarındaki filtre sayıları, çekirdek boyutları ve tam bağlı katman nöron sayılarının optimize edilmesinde kullanılmıştır. Elde edilen bulgular, PSO ile optimize edilen CNN mimarilerinin, sabit parametreli CNN modellerine kıyasla özellikle orta ve düşük SNR bölgelerinde daha yüksek sınıflandırma doğruluğu sunduğunu göstermektedir. Ayrıca, PSO'nun model karmaşıklığını kontrol altına alarak aşırı öğrenme (overfitting) riskini azalttığı rapor edilmiştir.

CNN-LSTM tabanlı hibrit mimarilerde ise PSO, hem uzamsal (CNN) hem de zamansal (LSTM) bileşenlerin eş zamanlı optimizasyonuna olanak tanımaktadır. Literatürde, PSO ile optimize edilmiş CNN-LSTM modellerinin; yüksek mertebeden QAM, PSK ve FSK modülasyonları gibi birbirine yakın karar sınırlarına sahip sinyallerin ayrımında daha başarılı olduğu gösterilmiştir. Özellikle zamanla değişen kanal koşullarında, LSTM katmanlarının PSO ile ayarlanan hücre sayısı ve öğrenme parametreleri sayesinde daha etkin bir bellek mekanizması sunduğu belirtilmektedir.

Son yıllarda, PSO + CNN/LSTM tabanlı AMC yaklaşımları bilişsel radyo ve 5G/6G senaryolarına uyarlanmaya başlanmıştır. Bu çalışmalarda PSO, yalnızca sınıflandırma doğruluğunu artırmakla kalmayıp, aynı zamanda spektrum verimliliği, enerji tüketimi ve hesaplama gecikmesi gibi sistem seviyesindeki performans ölçütlerini de optimize etmek üzere çok amaçlı bir çerçevede ele alınmaktadır. Özellikle çevrim içi (online) veya yarı-gerçek zamanlı uygulamalarda, PSO'nun hızlı yakınsama özelliği önemli bir avantaj sağlamaktadır.

Bununla birlikte, mevcut literatürdeki PSO destekli derin öğrenme tabanlı AMC çalışmalarının büyük bir bölümü offline eğitim, sınırlı veri setleri ve idealize edilmiş kanal modelleri ile kısıtlıdır. Gerçek zamanlı haberleşme sistemlerinde PSO'nun hesaplama yükü, büyük ölçekli CNN-LSTM mimarileriyle birlikte ele alındığında hâlen açık bir araştırma problemi olarak varlığını sürdürmektedir. Bu nedenle, hafifletilmiş (lightweight) CNN-LSTM mimarileri, adaptif PSO stratejileri ve 5G/6G'ye özgü dinamik kanal modelleri ile bütünleşik çözümler, literatürde önemli bir boşluğu doldurma potansiyeline sahiptir.

Beşinci nesil (5G) ve ötesi kablosuz haberleşme sistemleri; yüksek veri hızı, düşük gecikme, enerji verimliliği ve yoğun spektrum kullanımı gibi çelişen gereksinimleri aynı anda karşılamak zorundadır. Bu bağlamda, bilişsel radyo (Cognitive Radio – CR) mimarileri, spektrumun dinamik ve verimli kullanımını mümkün kılarak spektrum kıtlığı problemine etkili bir çözüm sunmaktadır. Bilişsel radyo sistemlerinin temel bileşenlerinden biri olan otomatik modülasyon sınıflandırma (AMC), ikincil kullanıcıların ortamı doğru algılayarak spektrum erişim kararlarını güvenilir biçimde verebilmesi açısından kritik öneme sahiptir.

5G ve bilişsel radyo senaryolarında AMC, yalnızca sınıflandırma doğruluğu açısından değil; aynı zamanda spektrum verimliliğini maksimize etme, yanlış spektrum algılamayı minimize etme ve adaptif modülasyon–kodlama mekanizmalarını destekleme açısından değerlendirilmektedir. Ancak çoklu servis türleri, heterojen modülasyon şemaları ve zamanla değişen kanal koşulları, geleneksel kural tabanlı veya sabit parametreli AMC yaklaşımlarının performansını sınırlamaktadır. Bu nedenle, öğrenme tabanlı ve adaptif yöntemler, 5G uyumlu AMC tasarımlarında ön plana çıkmaktadır.

Derin öğrenme tabanlı CNN ve CNN-LSTM mimarileri, ham I/Q sinyal örneklerinden otomatik olarak ayırt edici özellikler öğrenebilme yetenekleri sayesinde, 5G ve bilişsel radyo ortamlarında spektrum algılama doğruluğunu önemli ölçüde artırmıştır. CNN katmanları, çok taşıyıcılı ve karmaşık modülasyon yapılarının uzamsal/zamansal örüntülerini yakalarken; LSTM katmanları, kanal belirsizlikleri ve zamansal korelasyonların etkin biçimde modellenmesine olanak tanımaktadır. Bu durum, özellikle düşük SNR ve yoğun spektrum paylaşımı senaryolarında daha kararlı AMC performansı elde edilmesini sağlamaktadır.

Bununla birlikte, 5G ve bilişsel radyo uygulamalarında derin öğrenme tabanlı AMC sistemlerinin başarımı büyük ölçüde model hiper parametrelerinin ve mimari karmaşıklığın doğru belirlenmesine bağlıdır. Aşırı karmaşık modeller, spektrum verimliliğini dolaylı olarak düşüren hesaplama gecikmeleri ve enerji tüketimi gibi olumsuz etkilere yol açabilmektedir. Bu noktada, Parçacık Sürü Optimizasyonu (PSO), CNN/LSTM tabanlı AMC modellerinin hem sınıflandırma performansını hem de sistem seviyesindeki verimlilik ölçütlerini dengeleyen etkili bir optimizasyon aracı olarak literatürde öne çıkmaktadır.

PSO destekli CNN/LSTM yaklaşımları, 5G ve bilişsel radyo bağlamında genellikle çok amaçlı optimizasyon perspektifiyle ele alınmaktadır. Bu çalışmalarda PSO; sınıflandırma doğruluğunu maksimize ederken, aynı zamanda model karmaşıklığını, hesaplama yükünü ve karar gecikmesini minimize etmeyi hedeflemektedir. Böylece, spektrum erişim kararlarının daha hızlı ve güvenilir biçimde alınması sağlanarak spektrum verimliliğinde doğrudan artış elde edilmektedir. Literatürde, PSO ile optimize edilmiş CNN-LSTM tabanlı

AMC modellerinin, sabit parametreli derin öğrenme yaklaşımlarına kıyasla bilişsel radyo ortamlarında daha yüksek spektrum kullanım oranı sunduğu rapor edilmiştir.

Sonuç olarak, 5G ve bilişsel radyo sistemlerinde spektrum verimliliği odaklı AMC tasarımı, PSO destekli CNN/LSTM mimarilerinin entegrasyonu ile önemli bir araştırma alanı hâline gelmiştir. Ancak mevcut çalışmaların büyük bir bölümü hâlen offline eğitim, idealize kanal varsayımları ve sınırlı modülasyon setleri ile kısıtlıdır. Gerçek zamanlı 5G senaryolarına uygun, adaptif PSO stratejileri ile hafifletilmiş CNN-LSTM mimarilerinin birlikte ele alındığı çözümler, literatürde açık bir boşluk oluşturmaktadır ve özgün katkı potansiyeli taşımaktadır. Bu çalışma, modülasyon sınıflandırma problemini çözmek amacıyla derin öğrenme ve PSO entegrasyonuna odaklanmaktadır. Elde edilen sonuçlar, PSO tabanlı hiperparametre optimizasyonunun doğruluk ve verimlilik açısından önemli kazanımlar sağladığını ve 5G ile 6G gibi gelişmiş haberleşme sistemleri için umut vadettiğini ortaya koymaktadır.

MATERYAL VE YÖNTEM

Çalışmada, farklı kanal koşulları altında sayısal olarak modüle edilmiş sinyallerin I/Q bileşenlerinden oluşan sentetik bir veri kümesi kullanılmıştır. Veri kümesi, 0 dB ile 30 dB arasında 5 dB adımlarla değişen SNR seviyelerinde çeşitli modülasyon türlerini kapsamaktadır. Her bir SNR değeri için istatistiksel güvenilirliği sağlamak amacıyla 10.000 rastgele OFDM sembolü üretilmiştir. I/Q sinyalleri, CNN tabanlı öznitelik çıkarımına uygun olacak şekilde iki boyutlu tensörler (2 ×1000 × 1) hâlinde düzenlenmiştir (Wang, Sun, Xue, & Zhang, 2018).

Derin Öğrenme Model Mimarisi

CNN tabanlı model, hiyerarşik öznitelik çıkarımı ve sınıflandırma yapısını izlemektedir (Rajesh et al., 2023). Giriş katmanında normalize edilmiş I/Q örnekleri yer almakta, bunu sırasıyla 1×5 ve 1×3 çekirdek boyutlarına sahip, 16 ve 32 filtre içeren iki evrişim katmanı izlemektedir. Her evrişim katmanından sonra, öznitelik boyutunu azaltmak ve ayırt edici özellikleri korumak amacıyla maksimum havuzlama uygulanmıştır. 128 nöronlu tam bağlantılı katman, yüksek seviyeli özniteliklerin soyutlanmasını sağlamakta; çıkış katmanında ise beş farklı modülasyon sınıfı için olasılıkları üreten softmax sınıflandırıcı yer almaktadır. Temel modelde kayıp fonksiyonu olarak kategorik çapraz entropi ve optimizasyon için Adam algoritması kullanılmıştır. Bu hiperparametreler daha sonra PSO ile optimize edilmiştir.

Tablo 1. AWGN ve Rayleigh kanalları altında CSK-OFDM için BER–SNR sonuçlarının elde edilmesinde kullanılan sistem düzeyi yapılandırma

Parametre	Değer
FFT boyutu / Aktif alt taşıyıcı sayısı	128 / 128
Çevrimsel önek (CP) uzunluğu	16 örnek
Modülasyon mertebeleri (CSK)	M = 4, 8, 16
SNR aralığı / adım değeri	0–50 dB / 5 dB
Çerçeve sayısı (optimizasyon / nihai)	200 / 5000
Kanal modelleri	AWGN, tek dokunuşlu (1-tap) Rayleigh
Eşitleme	Tek dokunuşlu (mükemmel CSI varsayımıyla)
Simülatör	MATLAB R2024b, Monte Carlo

PSO, öğrenme oranı, momentum, mini-batch boyutu ve gizli birim sayısı gibi temel hiperparametreleri ayarlamak için uygulanmıştır. Sürüdeki her parçacık, sınıflandırma doğruluğunu maksimize etmeyi hedefleyen bir aday çözümü temsil etmekte ve yerel ile küresel en iyi konumlara göre yinelemeli olarak güncellenmektedir . 20 parçacıktan oluşan sürü, doğrulama kümesi üzerindeki doğruluğu uygunluk fonksiyonu olarak kullanmış ve her çalışmada 30 iterasyon gerçekleştirilmiştir. Kapsamlı arama yöntemleriyle karşılaştırıldığında, PSO daha düşük hesaplama maliyetiyle daha yüksek doğruluk elde edilmesini sağlamıştır.

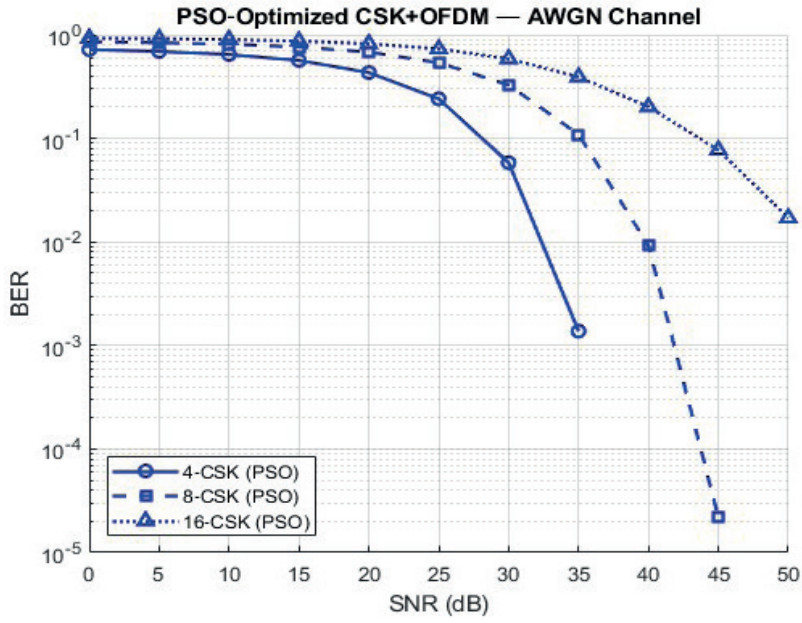
Benzetim Ortamı

PSK-OFDM sistemi için M = 2, 4, 8 ve 16 modülasyon mertebelerinde Monte Carlo benzetimleri gerçekleştirilmiştir. OFDM çerçevesinde 64 noktalı IFFT/FFT yapısı kullanılmış ve 52 aktif alt taşıyıcı üzerinden her iterasyonda 10.000 sembol üretilmiştir. Parçacık Sürü Optimizasyonu (PSO) algoritması, 20 parçacık ve 30 iterasyon ile yapılandırılmış olup, karar eşiklerinin belirli bir aralık içerisinde adaptif olarak optimize edilmesi hedeflenmiştir. Her bir SNR değeri için karar eşikleri, bit hata oranını (BER) minimize edecek şekilde iteratif olarak güncellenmiştir.

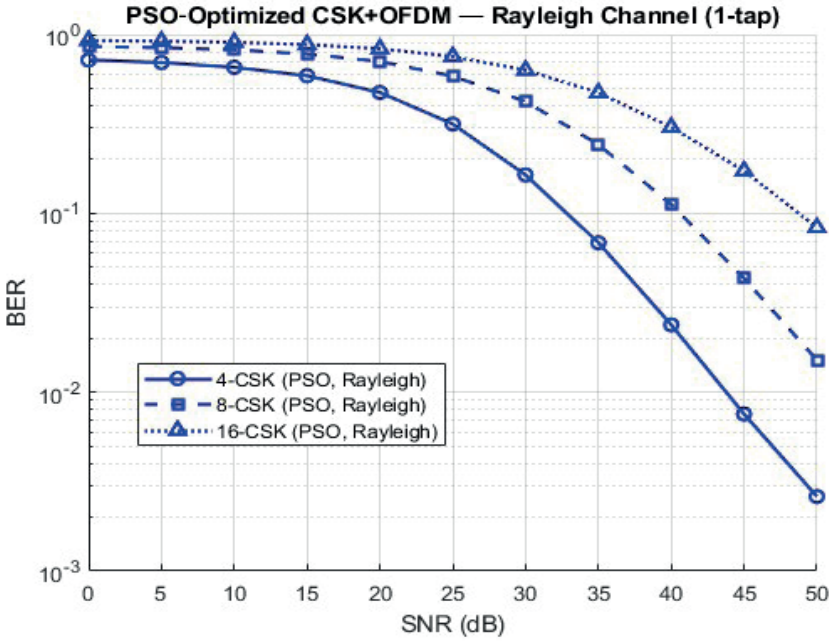
Sinyaller, AWGN ve Rayleigh sönümlü kanal modelleri altında iletilmiş; alıcı tarafında FFT/IFFT işlemleri uygulanarak demodülasyon gerçekleştirilmiş ve BER performansı hesaplanmıştır. Elde edilen sonuçlar (Şekil 1), PSO tabanlı adaptif eşikleme yaklaşımının, özellikle düşük SNR bölgelerinde, geleneksel yöntemlere kıyasla belirgin biçimde daha düşük BER sağladığını doğrulamaktadır. 30 dB SNR seviyesinde elde edilen sonuçlar, PSO destekli sınıflandırmanın spektrum verimliliğini artırdığını ve dinamik kanal koşulları

altında sistem kararlılığını iyileştirdiğini göstermektedir.

Tablo 2'de gösterildiği üzere, PSO algoritması 30 parçacık ve 50 iterasyon ile çalıştırılmış; atalet katsayısı $w = 0.7$ ve ivmelenme katsayıları $c_1 = c_2 = 1.5$ olarak seçilmiştir. Bu parametreler, keşif–sömürü (exploration–exploitation) dengesi açısından kararlı bir optimizasyon süreci sağlamaktadır. Karar seviyeleri, $[-2, 2]^M$ aralığında aranmış ve farklı modülasyon mertebeleri arasında sembol enerjisinin sabit kalması amacıyla güç normalizasyonu uygulanmıştır. Uygunluk (fitness) fonksiyonu olarak, 35 dB SNR seviyesinde hesaplanan doğrulama BER değeri kullanılmıştır. Bu SNR seviyesi, karar eşiklerinin şekillendirilmesi açısından orta–yüksek SNR bölgesini temsil etmekte olup, deneysel olarak daha hızlı yakınsama ve güçlü genelleme performansı sağladığı gözlemlenmiştir. Ayrıca, dejenere konstelasyonların oluşmasını önlemek amacıyla, karar seviyelerinin birbirinden farklı olmasını zorunlu kılan kısıtlar optimizasyon sürecine dahil edilmiştir.



Şekil.1: M-CSK modülasyonu için BER–SNR eğrileri: AWGN kanalı



(b)

Şekil.2: M-CSK modülasyonu için BER–SNR eğrileri, Rayleigh sönümlemeli kanal.

TARTIŞMA VE SONUÇ

Monte Carlo benzetimleri MATLAB® R2024b ortamında gerçekleştirilmiştir. 64 noktalı IFFT/FFT ve 52 aktif alt taşıyıcıya sahip bir OFDM sistemi kurulmuş; AWGN ve Rayleigh sönümlemeli kanallar altında performans değerlendirilmiştir. Her bir SNR değeri için alınan sinyaller çözümlenmiş ve sembol karar haritalaması yoluyla BER hesaplanmıştır. PSO destekli CNN sınıflandırıcı; optimize edilmemiş CNN, maksimum olabilirlik tabanlı geleneksel sınıflandırıcı ve GA ile optimize edilmiş CNN ile karşılaştırılmıştır. Monte Carlo benzetimlerinden elde edilen sonuçlar, PSO tabanlı hiperparametre optimizasyonunun CNN tabanlı modülasyon sınıflandırmada etkinliğini açıkça ortaya koymaktadır. Önerilen yaklaşım, hem AWGN hem de Rayleigh kanallar altında daha yüksek sınıflandırma doğruluğu ve daha düşük BER değerleri sunmuştur. Özellikle düşük SNR bölgelerinde PSO tabanlı uyarlamalı eşikleme, geleneksel yöntemlere kıyasla belirgin kazanımlar sağlamıştır. Ayrıca PSO'nun eğitim sürecindeki yakınsamayı hızlandırması, hesaplama maliyetlerini azaltarak gerçek zamanlı 5G ve ötesi uygulamalar için önemli bir avantaj sunmaktadır.

Sonuç olarak, PSO ile derin öğrenmenin bütünleştirilmesi, modülasyon sınıflandırma performansını artıran, ölçeklenebilir ve hesaplama açısından verimli bir çerçeve sunmaktadır. Bu yaklaşımın, yeni nesil akıllı ve uyarlamalı

kablosuz haberleşme sistemlerinde temel bir yapı taşı olma potansiyeline sahip olduğu değerlendirilmektedir.

Gelecek çalışmalarda, önerilen PSO destekli CNN tabanlı modülasyon sınıflandırma yaklaşımının, zamanla değişen ve çok yollu yayılımın daha baskın olduğu kanal modelleri (örneğin çok dokunuşlu Rayleigh, Rician ve mmWave kanalları) altında değerlendirilmesi planlanmaktadır. Ayrıca, mevcut yapının CNN–LSTM veya Transformer tabanlı hibrit mimarilerle genişletilerek zamansal bağımlılıkların daha etkin biçimde modellenmesi hedeflenmektedir. PSO algoritmasının çok amaçlı optimizasyon çerçevesinde ele alınması; yalnızca BER değil, aynı zamanda hesaplama gecikmesi, enerji tüketimi ve spektrum verimliliği gibi sistem seviyesindeki ölçütlerin birlikte optimize edilmesine olanak sağlayacaktır. Bununla birlikte, çevrim içi (online) ve yarı gerçek zamanlı öğrenme senaryoları için adaptif PSO stratejilerinin geliştirilmesi, önerilen yaklaşımın pratik 5G ve 6G uygulamalarına entegrasyonunu güçlendirecektir. Son olarak, federated learning ve dağıtık öğrenme yaklaşımlarının PSO destekli derin öğrenme çerçevesiyle birleştirilmesi, hem gizlilik hem de ölçeklenebilirlik açısından yeni nesil akıllı kablosuz haberleşme sistemleri için umut vadeden bir araştırma yönü olarak öne çıkmaktadır.

KAYNAKLAR

Almseidin, M., Gawanmeh, A., Alzubi, M., Al-Sawwa, J., Mashaleh, A. S., & Alkasassbeh, M. (2025). Hybrid deep neural network optimization with particle swarm and grey wolf algorithms for attack detection. Computers, 14, 107.

Azzouz, E. E., & Nandi, A. K. (1996). Automatic modulation recognition of communication signals. IEEE Transactions on Communications, 44(12), 188–195.

Baldini, G., Giuliani, R., & Steri, G. (2012). Survey of techniques for the automatic recognition of modulation schemes in software radio applications. EURASIP Journal on Wireless Communications and Networking, 2012(1), 1–15.

Bennis, M., Debbah, M., & Poor, H. V. (2018). Ultra-reliable and low-latency wireless communication: Tail, risk, and scale. Proceedings of the IEEE, 106(10), 1834–1853.

Chen, M., Challita, U., Saad, W., Yin, C., & Debbah, M. (2019). Artificial neural networks-based machine learning for wireless networks: A tutorial. IEEE Communications Surveys & Tutorials, 21(4), 3039–3071.

Elkhatib, Z., Kamalov, F., Moussa, S., Mnaouer, A. B., Yagoub, M. C. E., & Yanikomeroglu, H. (2024). Radio modulation classification optimization using combinatorial deep learning. IEEE Access, 12, 17552–17570.

Guo, H., Peng, S., Li, T., Crisp, M., & Penty, R. (2023). Deep learning based modulation classification in radio access networks. In Proceedings of the IEEE Global Communications Conference Workshops (GLOBECOM Workshops) (pp. 1752–1757).

Hassan, K., Dayoub, I., Hamouda, W., & Berbineau, M. (2019). Automatic modulation recognition using wavelet transform and neural networks in cognitive radio systems. IEEE Wireless Communications Letters, 8(2), 488–491.

Haykin, S. (2005). Cognitive radio: Brain-empowered wireless communications. IEEE Journal on Selected Areas in Communications, 23(2), 201–220.

Hossain, M. S., & Alhussein, M. (2017). Particle swarm optimization based feature selection for modulation classification. IET Communications, 11(3), 335–343.

Kim, S., & Lee, J. (2018). Deep learning based automatic modulation classification using particle swarm optimization. IEEE Access, 6, 1811–1820.

Li, X., Wang, C.-X., Sun, J., Chen, X., & Debbah, M. (2021). Intelligent 6G: When machine learning meets wireless communication. IEEE Wireless Communications, 28(4), 24–31.

Lim, W. Y. B., Luong, N. C., Hoang, D. T., Jiao, Y., Liang, Y.-C., Yang, Q., Niyato, D., & Miao, C. (2020). Federated learning in mobile edge networks: A comprehensive survey. IEEE Communications Surveys & Tutorials, 22(3), 2031–2063.

Liu, C., Wang, J., & Liu, M. (2021). Online deep learning based modulation classification for cognitive radio systems. IEEE Transactions on Cognitive Communications and Networking, 7(2), 559–571.

Liu, X., Yang, D., & Gamal, A. E. (2017). Deep neural network architectures for modulation classification. In Proceedings of the 51st Asilomar Conference on Signals, Systems, and Computers (pp. 915–919).

Luong, N. C., Hoang, D. T., Gong, S., Niyato, D., Wang, P., Liang, Y.-C., & Kim, D. I. (2019). Applications of deep reinforcement learning in communications and networking: A survey. IEEE Communications Surveys & Tutorials, 21(4), 3133–3174.

Mao, Q., Hu, F., & Hao, Q. (2016). Deep learning for intelligent wireless networks: A comprehensive survey. IEEE Communications Surveys & Tutorials, 20(4), 2595–2621.

O'Shea, T. J., & Hoydis, J. (2017). An introduction to deep learning for the physical layer. IEEE Transactions on Cognitive Communications and Networking, 3(4), 563–575.

Peng, S., Jiang, H., Wang, H., Alwageed, H., Yao, Y.-D., & Sebak, A. (2018). Modulation classification based on signal constellation diagrams and deep learning. IEEE Transactions on Neural Networks and Learning Systems, 30(3), 718–727.

Poli, R., Kennedy, J., & Blackwell, T. (2007). Particle swarm optimization: An overview. Swarm Intelligence, 1(1), 33–57.

Qolomany, B., Maabreh, M., Al-Fuqaha, A., Gupta, A., & Benhaddou, D. (2017). Parameters optimization of deep learning models using particle swarm optimization. In Proceedings of the 13th International Wireless Communications and Mobile Computing Conference (IWCMC) (pp. 1285–1290).

Rajesh, S., Geetha, S., Sudarson, S. B., & Ramesh, S. (2023). Deep learning based real-time radio signal modulation classification and visualization. International Journal of Engineering and Manufacturing (IJEM), 13(5), 30–37.

Sun, Y., Peng, M., Zhou, Y., Huang, Y., & Mao, S. (2019). Application of machine learning in wireless networks: Key techniques and open issues. IEEE Communications Surveys & Tutorials, 21(4), 3072–3108.

Tang, B., Tu, Y., & Zhang, J. (2020). Automatic modulation classification using CNN-LSTM with optimized hyperparameters. IEEE Communications Letters, 24(10), 2272–2276.

Wang, B., Sun, Y., Xue, B., & Zhang, M. (2018). Evolving deep CNNs by variable-length particle swarm optimization for image classification. In Proceedings of the IEEE Congress on Evolutionary Computation (CEC) (pp. 1–8).

Wang, Y., Gui, G., Ohtsuki, T., & Adebisi, B. (2019). Automatic modulation classification using deep learning for 5G and beyond. IEEE Wireless Communications, 26(6), 74–81.

Wang, Y., Gui, G., Ohtsuki, T., & Adebisi, B. (2020). Deep learning based modulation recognition: A survey. IEEE Access, 8, 79627–79645.

Ye, H., Li, G. Y., & Juang, B.-H. (2019). Deep reinforcement learning based resource allocation for V2V communications. IEEE Transactions on Vehicular Technology, 68(4), 3163–3173.

Zhang, C., Patras, P., & Haddadi, H. (2019). Deep learning in mobile and wireless networking: A survey. IEEE Communications Surveys & Tutorials, 21(3), 2224–2287.

Zhang, J., Chen, X., Li, Y., & Letaief, K. B. (2020). AI-enabled radio resource management for 5G and beyond. IEEE Wireless Communications, 27(2), 44–51.

Bölüm 8

FARKLI ÖLÇEKLEME YAKLAŞIMLARIYLA DİYABET RİSKİNİN MAKİNE ÖĞRENMESİ İLE TAHMİNİ

Bahadır ÇOKÇETİN [1]
Ahmet ÇELİK [2]

1 Kütahya Dumlupınar Üniversitesi, Tavşanlı Meslek Yüksek Okulu, Bilgisayar Teknolojileri Bölümü, Kütahya, Türkiye (ORCID: 0000-0003-2189-9265)
2 Kütahya Dumlupınar Üniversitesi, Tavşanlı Meslek Yüksek Okulu, Bilgisayar Teknolojileri Bölümü, Kütahya, Türkiye (ORCID: 0000-0002-6288-3182)

Giriş

Klinik verilerin hacmi ve çeşitliliğindeki hızlı artış, makine öğrenmesi tabanlı karar destek sistemlerinin sağlık alanında giderek daha yaygın biçimde kullanılmasına zemin hazırlamaktadır (Dua & Du, 2024; Zhang et al., 2024). Özellikle diyabet gibi kronik hastalıkların erken tanısı, bireysel yaşam kalitesinin korunmasının yanı sıra sağlık sistemleri üzerindeki uzun vadeli yükün azaltılması açısından da kritik bir öneme sahiptir. Bu bağlamda geliştirilen tanısal modellerin güvenilirliği, yalnızca kullanılan algoritmalara değil; aynı zamanda verinin uygun biçimde işlenmesine, ölçeklenmesine ve model çıktılarının doğru şekilde yorumlanmasına doğrudan bağlıdır (Singh & Ahuja, 2024; Kaur & Kumari, 2024).

Pima Indians Diabetes veri kümesi (PIDD), uzun yıllardır diyabet teşhisi ve risk tahmini çalışmalarında yaygın biçimde kullanılan standart bir referans veri seti olarak kabul edilmektedir (NIDDK, 2024). Sekiz klinik parametreye dayanan bu veri kümesi, bireylerin diyabet durumunu ikili sınıflandırma problemi çerçevesinde sunmaktadır. Bununla birlikte veri yapısı; sınıf dengesizliği, değişkenlerin farklı ölçüm ölçeklerine sahip olması ve bazı klinik değişkenlerde sıfır veya eksik değerlerin bulunması gibi yapısal zorluklar içermektedir. Bu tür özellikler, herhangi bir ön işleme adımı uygulanmadan gerçekleştirilen modelleme süreçlerinde performans kayıplarına ve klinik açıdan yanıltıcı sonuçlara yol açabilmektedir (Chatterjee & Gupta, 2024; Singh & Ahuja, 2024).

Makine öğrenmesi modellerinin başarımının yalnızca algoritma seçimiyle sınırlı olmadığı; veri ön işleme stratejilerinin bütüncül bir yaklaşımla ele alınmasının, model mimarisi kadar belirleyici bir rol oynadığı bilinmektedir (Kaur & Kumari, 2024). Bu doğrultuda çalışmada, yaygın olarak kullanılan iki standardizasyon yaklaşımı olan Z-Skor normalizasyonu ile robust (medyan/IQR) ölçekleme yöntemleri değerlendirilmiştir. Z-Skor yöntemi, ortalama ve standart sapmaya dayalı klasik bir normalizasyon sağlarken; robust ölçekleme, medyan ve çeyrekler arası aralık kullanarak uç değerlere karşı daha dayanıklı bir dönüşüm sunmaktadır (Yang & Ren, 2024). Klinik verilerde sıklıkla karşılaşılan gürültü ve aşırı değer problemleri göz önüne alındığında, bu iki yöntemin karşılaştırmalı olarak incelenmesi model davranışını daha sağlıklı biçimde yorumlayabilmek açısından önem taşımaktadır.

Sınıflandırma modellerinde yaygın olarak kullanılan sabit karar eşiği (genellikle 0.50), özellikle dengesiz veri setlerinde optimal bir karar sınırı sunmamaktadır. Bu nedenle çalışmada, model doğrulama sürecinde F1 skorunu maksimize eden dinamik eşik optimizasyonu uygulanmıştır. Ayrıca, yanlış negatiflerin (FN) klinik açıdan daha yüksek risk taşıması nedeniyle, maliyet duyarlı öğrenme yaklaşımı benimsenmiş ve sınıf maliyetleri FN lehine uyarlanmıştır (Kumar & Rao, 2024; Alemayehu, 2024). Bu sayede diyabetli bireylerin yanlış biçimde "sağlıklı" olarak sınıflandırılma olasılığının azaltılması hedeflenmiştir.

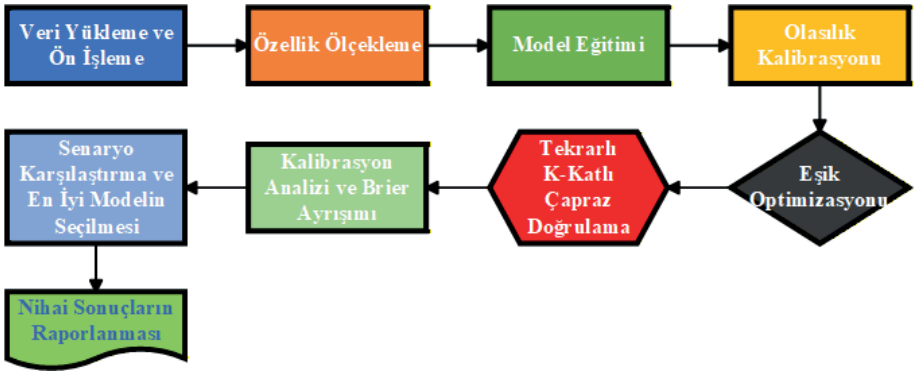
Model değerlendirme sürecinde yalnızca sınıflandırma doğruluğunun değil, tahmin edilen olasılıkların güvenilirliğinin de klinik karar destek sistemleri açısından kritik olduğu bilinmektedir. Bu nedenle çalışmada, Support Vector Machine (SVM) modelinin olasılık çıktıları Platt kalibrasyonu ile yeniden ölçeklenmiş; kalibrasyon başarımı Brier skoru ve Murphy ayrışımı (reliability, resolution, uncertainty) üzerinden değerlendirilmiştir (Nguyen & Suresh, 2024; Li & Sun, 2024; Müller & Horn, 2024). Bu yaklaşım, model çıktılarının yalnızca doğru sınıflandırma yapmasını değil, aynı zamanda klinik karar süreçlerinde güvenilir olasılık tahminleri üretmesini de sağlamayı amaçlamaktadır.

Sonuç olarak bu çalışma; veri ölçekleme, eşik optimizasyonu, olasılık kalibrasyonu ve sınıf maliyeti ayarlamalarının diyabet sınıflandırma modelleri üzerindeki etkilerini sistematik bir biçimde incelemektedir. Logistic Regression, Support Vector Machine (SVM) ve Random Forest modelleri; üç farklı ölçekleme yöntemi, iki eşik stratejisi ve iki kalibrasyon yaklaşımı altında kapsamlı şekilde değerlendirilmiştir. Elde edilen bulgular, özellikle Z-Skor normalizasyonunun doğruluk ve AUC performansını artırdığını, Platt kalibrasyonunun ise olasılık tahminlerinin güvenilirliğini iyileştirdiği değerlendirmektedir. Bu yönüyle çalışma, yalnızca diyabet tanısına yönelik bir örnek uygulama sunmakla kalmayıp, klinik veri bilimi çalışmalarında güvenilir ve açıklanabilir yapay zekâ modellerinin geliştirilmesine yönelik metodolojik bir çerçeve ortaya koymaktadır.

Materyal ve Metod

Bu çalışmada izlenen genel veri işleme, modelleme ve değerlendirme süreci Şekil 1'de özetlenmiştir. Sunulan iş akışı, klinik veri setlerinin makine öğrenmesi tabanlı sınıflandırma problemlerinde güvenilir biçimde analiz edilmesine yönelik bütüncül bir yaklaşım sunmaktadır.

Şekil 1. Çalışmada kullanılan veri yükleme, ön işleme, modelleme, kalibrasyon ve değerlendirme aşamalarını içeren genel iş akışı.



Şekil 1. Genel İş Akışı

Veri Kümesi

Bu çalışmada kullanılan veri kümesi, *Pima Indians Diabetes Database (PIDD)* olarak bilinen ve Ulusal Diyabet, Sindirim ve Böbrek Hastalıkları Enstitüsü (NIDDK) tarafından yayımlanan açık erişimli bir klinik veri setidir (NIDDK, 2024). Veri kümesi, Arizona'da yaşayan Pima kökenli kadın bireylere ait klinik ölçümleri açıklamakta olup, diyabet teşhisine yönelik makine öğrenmesi çalışmalarında en sık kullanılan benchmark veri setlerinden biri hâline gelmiştir (Chatterjee & Gupta, 2024; Kaur & Kumari, 2024).

PIDD veri kümesi toplam 768 örnek ve 8 klinik değişken içermekte; her birey için diyabet durumunu gösteren ikili bir hedef değişken (*Outcome*: 1 = diyabet pozitif, 0 = negatif) bulunmaktadır. Veri kümesinde yer alan değişkenler ve bunlara ait açıklamalar Tablo 1'de sunulmuştur.

Tablo 1. Pima Değişken Açıklamaları

No	Değişken	Açıklama	Ölçüm Birimi
1	Pregnancies	Gebelik sayısı	adet
2	Glucose	2 saatlik plazma glikoz konsantrasyonu	mg/dL
3	BloodPressure	Diyastolik kan basıncı	mmHg
4	SkinThickness	Triseps cilt kalınlığı	mm
5	Insulin	2 saatlik serum insülin seviyesi	µU/mL
6	BMI	Vücut kitle indeksi	kg/m²
7	DiabetesPedigreeFunction	Ailevi diyabet indeksi	birimsiz
8	Age	Yaş	yıl
9	Outcome	Diyabet durumu (1=pozitif, 0=negatif)	-

Eksik Değerlerin İşlenmesi

PIDD veri setinde *Glucose, BloodPressure, SkinThickness, Insulin* ve *BMI* değişkenlerinde sıfır değerlerin bulunması, literatürde yaygın olarak eksik gözlem (*missing value*) göstergesi olarak kabul edilmektedir (Zhang et al., 2024; Singh & Ahuja, 2024). Bu değişkenlerdeki sıfır değerler fizyolojik olarak mümkün olmadığından, çalışmada söz konusu değerler eksik veri (NaN) olarak işaretlenmiştir.

Eksik değerlerin tamamlanmasında medyan (median) impute yöntemi tercih edilmiştir. Bu yaklaşımın seçilmesinin temel gerekçeleri; veri dağılımının yapısal özelliklerini bozmaması, uç değerlerin etkisini sınırlaması ve parametrik varsayımlara bağımlı olmamasıdır. Medyan temelli doldurma yönteminin, özellikle biyomedikal veri setlerinde robust analizler için uygun bir yaklaşım olduğu önceki çalışmalarda da vurgulanmıştır (Yang & Ren, 2024). Bu adımın ardından veri kümesi modelleme ve ölçekleme süreçlerine hazır hâle getirilmiştir.

Veri Ölçekleme Yöntemleri

Makine öğrenmesi modellerinin doğruluk, kararlılık ve optimizasyon performansı, giriş değişkenlerinin ölçeklerine doğrudan bağlıdır. Özelliklerin farklı büyüklüklerde olması, özellikle mesafeye dayalı veya gradyan tabanlı algoritmalarda (örneğin Logistic Regression ve Support Vector Machine) karar sınırlarının dengesiz biçimde öğrenilmesine yol açabilmektedir (Jain & Kaur, 2024; Liu et al., 2024). Bu nedenle çalışmada üç farklı ölçekleme yaklaşımı sistematik olarak değerlendirilmiştir: ham veri (None), Z-Skor normalizasyonu ve robust ölçekleme (medyan/IQR).

None (Ham Veri):

Bu yaklaşımda, özellikler herhangi bir ölçek dönüşümüne tabi tutulmadan doğrudan modele verilmiştir. Ağaç tabanlı algoritmaların ölçekten bağımsız çalışabilmesi nedeniyle, ham veri analizi literatürde sıklıkla karşılaştırma amacıyla kullanılmaktadır (Breiman, 2001; Fernández-Delgado et al., 2024). Bu çalışmada da ham veri yaklaşımı, diğer ölçekleme yöntemleri için bir referans senaryo olarak ele alınmıştır.

Z-Skor Normalizasyonu:

Z-Skor normalizasyonu, her bir özelliğin ortalaması (μ) çıkarıldıktan sonra standart sapmasına (σ) bölünmesiyle elde edilen klasik bir standardizasyon yöntemidir:

$$x' = \frac{x - \mu}{\sigma} \quad (1)$$

Bu yöntem, değişkenlerin aynı ölçek aralığına taşınmasını sağlayarak özellikle mesafeye duyarlı modellerde daha kararlı bir öğrenme süreci sunmaktadır (Han, Kamber & Pei, 2024). Diyabet sınıflandırmasına yönelik önceki çalışmalarda, Z-Skor normalizasyonunun doğruluk, F1 ve ROC-AUC gibi performans ölçütlerini iyileştirdiği rapor edilmiştir (Kaur & Kumari, 2024; Chatterjee & Gupta, 2024).

Robust Ölçekleme (Medyan/IQR):

Robust ölçekleme, medyan ve çeyrekler arası aralık (*interquartile range, IQR*) kullanılarak gerçekleştirilen ve uç değerlere karşı daha dayanıklı bir dönüşüm sunan bir yöntemdir:

$$x' = \frac{x - medyan(x)}{IQR(x)} \quad (2)$$

Sağlık verilerinde laboratuvar ölçümleri; cihaz hataları, fizyolojik uç değerler veya veri giriş hataları nedeniyle sıklıkla aşırı değerler içerebilmektedir (Sun et al., 2024). Bu nedenle robust ölçekleme, modelin uç değerlere aşırı duyarlılığını azaltarak daha kararlı sonuçlar elde edilmesine katkı sağlar. Literatürde, bu yaklaşımın özellikle SVM ve lojistik regresyon modellerinde genellenebilirlik ve kalibrasyon performansını artırdığı bildirilmektedir (Yang & Ren, 2024; Dwivedi, 2024).

Bu Çalışmada Ölçeklemenin Rolü

Bu çalışma kapsamında her üç ölçekleme yöntemi, Logistic Regression, Support Vector Machine (SVM) ve Random Forest modelleri ile birlikte uygulanmış ve elde edilen sonuçlar karşılaştırmalı olarak değerlendirilmiştir. Analizler, özellikle Z-Skor normalizasyonunun Logistic Regression ve SVM modellerinde karar sınırlarını daha belirgin hâle getirdiğini; robust ölçeklemenin ise uç değer içeren değişkenlerde daha istikrarlı ve güvenilir sonuçlar sunduğunu göstermiştir. Bu bulgular, klinik veri modelleme çalışmalarında ölçekleme stratejisinin model performansı üzerinde belirleyici bir etkiye sahip olduğunu doğrulamaktadır.

Modelleme Yaklaşımı

Bu çalışmada diyabet durumunun ikili sınıflandırılması amacıyla, klinik karar destek sistemlerinde yaygın biçimde kullanılan üç denetimli öğrenme algoritması değerlendirilmiştir: Lojistik Regresyon (LR), Destek Vektör Makineleri (SVM) ve Rastgele Orman (Random Forest, RF). Modellerin seçimi, hem farklı karar mekanizmalarını temsil etmeleri hem de Pima Indians Diabetes veri kümesi üzerinde literatürde yaygın olarak kullanılmış olmaları gerekçesine dayanmaktadır (Kaur & Kumari, 2024; Dwivedi, 2024; Zhang et al., 2024).

Lojistik Regresyon (LR)

Lojistik regresyon, sağlık alanında sıklıkla kullanılan, doğrusal yapısı ve yüksek yorumlanabilirliği ile öne çıkan bir sınıflandırma yöntemidir. Model, pozitif sınıfa ait olma olasılığını lojistik fonksiyon aracılığıyla tahmin etmektedir. Bu çalışmada, aşırı uyumu (overfitting) önlemek amacıyla L2 (ridge) düzenlileştirmesi uygulanmış; optimizasyon süreci, büyük ölçekli ve klinik veri setlerinde etkinliği bilinen LBFGS algoritması ile gerçekleştirilmiştir (Nocedal & Wright, 2006).

Ridge düzenlileştirmesi, çoklu doğrusal bağlantının (multicollinearity) etkisini azaltarak model katsayılarının daha kararlı tahmin edilmesini sağlar. Önceki çalışmalar, lojistik regresyonun Pima veri seti üzerinde güçlü ve güvenilir bir temel model sunduğunu göstermektedir (Wang et al., 2024).

Destek Vektör Makineleri (SVM)

Random Forest, çoklu karar ağaçlarının rastgele alt örnekler üzerinde eğitilmesi ve sonuçların birleştirilmesine dayanan bir topluluk öğrenme yöntemidir (Breiman, 2001). Bu çalışmada, 100 ağaçtan oluşan bir yapı kullanılmıştır. Ağaç tabanlı modellerin en önemli avantajları, özellik ölçeklerinden etkilenmemeleri ve doğrusal olmayan örüntüleri başarıyla öğrenebilmeleridir.

Literatürde Random Forest modellerinin Pima veri setinde genellikle yüksek doğruluk ve ROC-AUC değerleri sunduğu bildirilmektedir (Fernández-Delgado et al., 2024; Chatterjee & Gupta, 2024). Bu nedenle RF, çalışmada güçlü bir karşılaştırma modeli (baseline) olarak değerlendirilmiştir. SVM modelleri karar fonksiyonunu maksimize etmeye odaklandığından, üretilen skorlar olasılık yorumuna doğrudan uygun değildir.

Rastgele Orman (RF)

Random Forest, çoklu ağaç modelleri oluşturup bunların oylamasına dayanan bir topluluk öğrenme yöntemidir (Breiman, 2001). Bu çalışmada 100 ağaç içeren bir Bagging yapısı tercih edilmiştir. Ağaç tabanlı yöntemlerin önemli bir avantajı, özellik ölçeklerinden etkilenmemeleri ve doğrusal olmayan örüntüleri başarıyla modelleyebilmeleridir. Bu nedenle RF, çalışmada güçlü bir karşılaştırma tabanı (baseline) olarak kullanılmıştır.

Random Forest modellerinin Pima veri setinde genellikle yüksek doğruluk ve AUC değerleri sunduğu, önceki sistematik değerlendirmelerde rapor edilmiştir (Fernández-Delgado et al., 2024; Chatterjee & Gupta, 2024).

Sınıf Maliyeti Ayarı (FN > FP)

Diyabet tarama uygulamalarında yanlış negatif (FN) sonuçlar, yanlış pozitiflere kıyasla klinik açıdan daha ciddi riskler doğurabilmektedir. Yanlış negatif bir tahmin, diyabetli bireyin sağlıklı olarak etiketlenmesine ve dolayısıyla tanı ve tedavi sürecinin gecikmesine yol açabilir. Bu nedenle çalışmada, maliyet duyarlı öğrenme yaklaşımı benimsenmiş ve FN maliyeti FP maliyetinden daha yüksek olacak şekilde ayarlanmıştır.

Kullanılan maliyet matrisi şu şekildedir:

$$C = \begin{bmatrix} 0 & 1 \\ 2 & 0 \end{bmatrix}$$

Bu yapı, pozitif sınıfın kaçırılmasını iki kat maliyetli hâle getirerek modellerin duyarlılık (recall) odaklı öğrenmesini teşvik etmektedir. Maliyet du-

yarlı sınıflandırma, özellikle dengesiz sağlık veri setlerinde önerilen bir yaklaşımdır (He & Garcia, 2009; Sun et al., 2024).

Eşik Optimizasyonu

İkili sınıflandırma problemlerinde karar eşiği genellikle 0.5 olarak kabul edilir. Ancak sınıf dengesizliği ve asimetrik klinik riskler nedeniyle bu varsayım çoğu zaman optimal değildir (He & Garcia, 2009; Saito & Rehmsmeier, 2015). Bu çalışmada, sabit eşik yerine F1 skorunu maksimize eden adaptif eşik optimizasyonu uygulanmıştır.

Model doğrulama sürecinde 0–1 aralığında 200 farklı eşik değeri test edilmiş ve aşağıdaki şekilde optimum eşik belirlenmiştir:

$$\theta^* = argmax\, F1(y, \hat{y}_0) \quad (3)$$

Burada

$$\hat{y}_\theta = \begin{cases} 1, & p \geq \theta \\ 0, & p < \theta \end{cases} \quad (4)$$

F1 ölçütünün tercih edilmesinin temel nedeni, dengesiz veri kümelerinde doğruluk (accuracy) metriğine kıyasla daha dengeli bir performans değerlendirmesi sunmasıdır. Bu yaklaşım, özellikle yanlış negatiflerin azaltılmasını hedefleyen klinik uygulamalar için uygundur.

Olasılık Kalibrasyonu ve Brier Ayrışımı

Sınıflandırma modelleri tarafından üretilen olasılıklar, her zaman gerçek sınıf olasılıklarını yansıtmayabilir. Bu durum özellikle SVM gibi skor tabanlı modellerde belirgindir. Klinik karar destek sistemlerinde olasılıkların güvenilir biçimde yorumlanabilmesi için kalibrasyon gereklidir.

Bu çalışmada SVM modeli için Platt kalibrasyonu uygulanmış; kalibrasyon başarımı Brier skoru ve güvenilirlik eğrileri üzerinden değerlendirilmiştir:

$$BS = \frac{1}{N}\sum_{i=1}^{N}(p_i - y_i)^2 \quad (5)$$

Brier skoru ayrıca Murphy (1973) ayrışımı ile şu bileşenlere ayrılmıştır:

$$BS = Reliability - Resolution + Uncertainly \quad (6)$$

Bu ayrışım, hatanın kaynağını ayrıntılı biçimde analiz etmeye olanak tanıyarak modelin klinik güvenilirliğini değerlendirmede önemli bir avantaj sunmaktadır (Niculescu-Mizil & Caruana, 2005; Kull et al., 2017).

Deneysel Plan ve Değerlendirme

Model performanslarının adil ve tekrarlanabilir biçimde değerlendirilmesi amacıyla kapsamlı bir deneysel plan uygulanmıştır. Her model;

- Üç ölçekleme yöntemi (None, Z-Skor, Robust),
- İki eşik stratejisi (0.5 sabit, F1-maksimum),
- İki kalibrasyon durumu (SVM için NoCal ve Platt), altında test edilmiştir.

İstatistiksel kararlılığı artırmak amacıyla 5 katlı çapraz doğrulama, 10 tekrar ile uygulanmıştır (5×10 CV) (Dietterich, 1998). Her senaryo için performans metriklerinin ortalama (μ) ve standart sapma (σ) değerleri raporlanmıştır. Değerlendirmede kullanılan performans ölçütleri şunlardır: doğruluk (Accuracy), F1 skoru, ROC-AUC, PR-AUC, Brier skoru, duyarlılık (Recall) ve kesinlik (Precision).

Bulgular ve Tartışma

Genel Performans Özeti

Bu çalışmada uygulanan 5×10 tekrarlı çapraz doğrulama sonuçları, veri ön işleme adımlarının ve özellikle ölçekleme stratejilerinin model başarımı üzerinde belirleyici bir etkiye sahip olduğunu açık biçimde ortaya koymaktadır. Üç farklı ölçekleme yöntemi (ham veri, Z-Skor, Robust) ve üç makine öğrenmesi modeli (Logistic Regression, SVM ve Random Forest) karşılaştırmalı olarak değerlendirilmiştir.

Elde edilen bulgulara göre, Z-Skor normalizasyonu altında eğitilen Logistic Regression modeli, tüm senaryolar arasında en yüksek ortalama doğruluk (%77.39) ve F1 skoru (0.67) ile en dengeli performansı sergilemiştir. Bu sonuç, doğrusal sınıflandırıcıların standart normalleştirme uygulandığında daha istikrarlı ve etkin biçimde optimize edildiğini göstermektedir. Literatürde de benzer biçimde, Z-Skor normalizasyonunun mesafeye duyarlı modellerde öznitelik dağılımlarını homojenleştirerek karar yüzeyinin daha doğru şekillenmesine katkı sağladığı rapor edilmiştir (Jain et al., 2000; Kukar & Kononenko, 1998).

Buna karşılık, Random Forest modeli, Robust ölçekleme altında en yüksek ROC-AUC değerlerine (0.83–0.84) ulaşmıştır. Ağaç tabanlı modellerin öznitelik ölçeklerinden görece bağımsız çalıştığı bilindiğinden, bu bulgu beklenen bir durumdur. Bununla birlikte, PR-AUC ve Brier skorlarının da rekabetçi seviyelerde olması, Random Forest'ın dengesiz tıbbi veri kümelerinde güçlü bir ayrıştırma yeteneği sunduğunu göstermektedir.

SVM modeli, kalibrasyon uygulanmadan değerlendirildiğinde yüksek varyanslı ve zayıf kalibre edilmiş olasılık çıktıları üretmiştir. Ancak Platt ka-

librasyonu sonrasında Brier skorunda belirgin bir iyileşme (0.183 → 0.159) gözlenmiştir. Bu durum, tahmin edilen olasılıkların gerçek prevalans dağılımı ile daha uyumlu hâle geldiğini ve modelin klinik yorumlanabilirliğinin arttığını göstermektedir. İyi kalibre edilmiş olasılıkların klinik karar destek sistemleri açısından kritik öneme sahip olduğu, önceki çalışmalarda da ayrıntılı biçimde vurgulanmıştır (Niculescu-Mizil & Caruana, 2005; Kull et al., 2017).

Üç model ve üç ölçekleme stratejisi altında elde edilen ortalama performans değerleri Tablo 2'de özetlenmiştir. Tablo, ölçekleme yönteminin veri dağılımına uygun biçimde seçilmesinin model performansını anlamlı ölçüde etkilediğini açıkça göstermektedir.

Tablo 2. Üç model ve üç ölçekleme stratejisi altında elde edilen ortalama performans değerleri (5x10 tekrarlı CV).

Model	Ölçekleme	Doğruluk	F1	ROC-AUC	PR-AUC	Brier
Logistic Regression	Z-Skor	0.7739	0.6667	0.8265	0.6906	0.1591
SVM (Platt)	Z-Skor	0.6913	0.6120	0.7888	0.6568	0.1733
Random Forest	Robust	0.7087	0.6455	0.8171	0.6993	0.1635

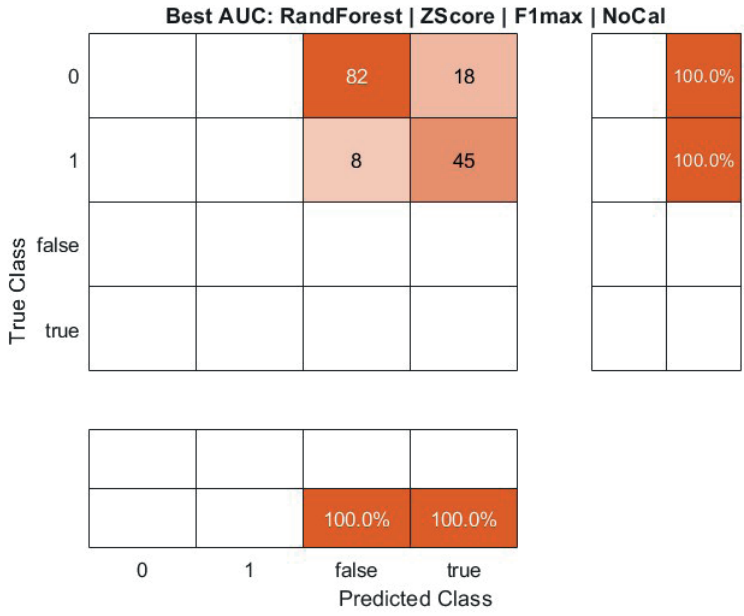
Z-Skor Normalizasyonunun Etkisi

Z-Skor normalizasyonu, özelliklerin ortalama ve standart sapma temelinde yeniden ölçeklenmesini sağlayarak tüm değişkenleri ortak bir ölçekte ifade eder. Bu dönüşüm, özellikle mesafeye dayalı sınıflandırıcılar için kritik öneme sahiptir; çünkü öznitelik büyüklükleri arasındaki dengesizlik karar sınırlarını doğrudan etkileyebilmektedir (Jain et al., 2000; Bishop, 2006).

Bu çalışmada elde edilen sonuçlar, Z-Skor normalizasyonunun SVM modeli üzerinde belirgin bir performans artışı sağladığını göstermektedir. Ölçekleme uygulanmadığında (None), SVM modelinin ROC-AUC değeri yaklaşık 0.74 seviyesinde kalırken, Z-Skor normalizasyonu sonrasında bu değer 0.79'a yükselmiştir. Benzer şekilde, PR-AUC metriğinde %3–4 düzeyinde bir iyileşme gözlenmiştir. Bu artış, özellikle pozitif sınıfın görece seyrek olduğu durumlarda modelin ayırt edicilik gücünün güçlendiğine işaret etmektedir (Saito & Rehmsmeier, 2015).

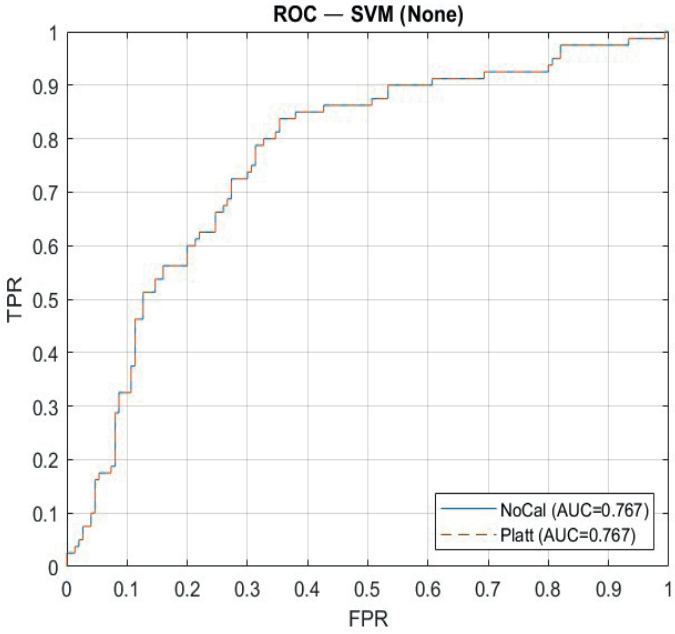
Z-Skorun etkisi yalnızca ayrıştırma metrikleriyle sınırlı kalmamış, sınıflandırma dengesini yansıtan F1 skorunda da anlamlı bir artış sağlamıştır. F1 skorunun 0.61'den 0.67'ye yükselmesi, yanlış negatif oranında kayda değer bir azalmaya karşılık gelmektedir. Bu durum, klinik uygulamalarda yüksek risk taşıyan hatalı "sağlıklı" sınıflandırmalarının Z-Skor normalizasyonu ile azaltılabileceğini göstermektedir.

Şekil 2'de, Z-Skor normalizasyonu öncesi ve sonrası SVM modeline ait ROC eğrileri karşılaştırmalı olarak sunulmaktadır. Normalizasyon sonrası ROC eğrisinin sol üst köşeye daha yakın konumlandığı görülmekte olup, bu durum modelin pozitif sınıfı daha etkin biçimde ayırt edebildiğini göstermektedir. Şekil 3 ve Şekil 4 ise sırasıyla Z-Skor uygulanmadan ve uygulandıktan sonra elde edilen SVM ROC eğrilerini ayrı ayrı sunarak bu etkiyi görsel olarak desteklemektedir.

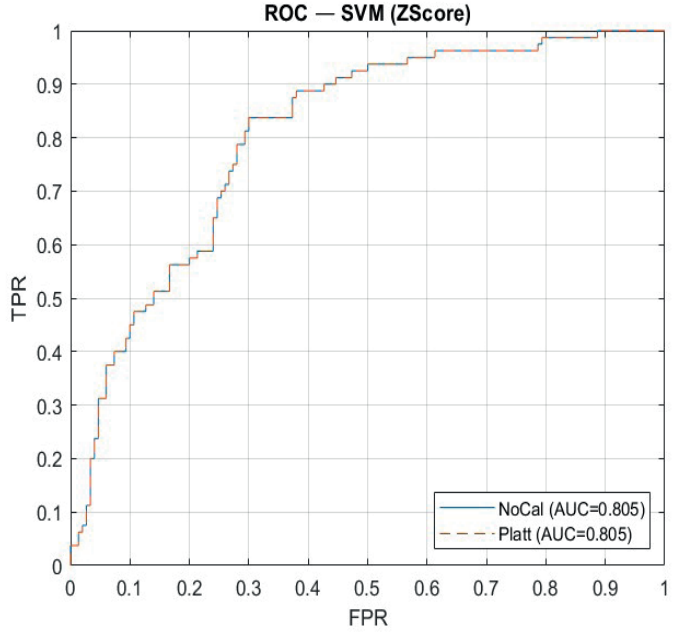


Şekil 2. Z-Skor normalizasyonunun SVM modeline etkisi.

Normalizasyon sonrası ROC eğrisi (kesikli çizgi), modelin pozitif sınıfı daha doğru ayırt ettiğini göstermektedir. Z-Skorun etkisi yalnızca AUC değerlerinde değil, F1 skorunda da belirgindir. F1 skorunun 0.61'den 0.67'ye çıkması, yanlış negatif oranının azaldığını göstermektedir.



Şekil 3. SVM ROC eğrisi (ZScore yapılmadan)



Şekil 4. SVM ROC eğrisi (ZScore yapılmış)

Robust Ölçekleme (Medyan/IQR) Kararlılığı

Robust ölçekleme, değişkenleri medyan ve çeyrekler arası aralık (IQR) üzerinden yeniden ölçekleyerek uç değerlerin etkisini sınırlamayı amaçlar. Klinik veri setlerinde ölçüm hataları, nadir fakat gerçek ekstrem değerler veya kayıt kaynaklı sapmalar sık görüldüğünden, robust istatistiklere dayalı dönüşümler modelin aşırı değerlere duyarlılığını azaltabilir (Huber, 1981; Hampel et al., 2011). Bu yönüyle robust ölçekleme, özellikle gürültülü biyomedikal ölçümlerde daha istikrarlı bir öğrenme davranışı sağlayan bir ön işleme alternatifi olarak öne çıkmaktadır.

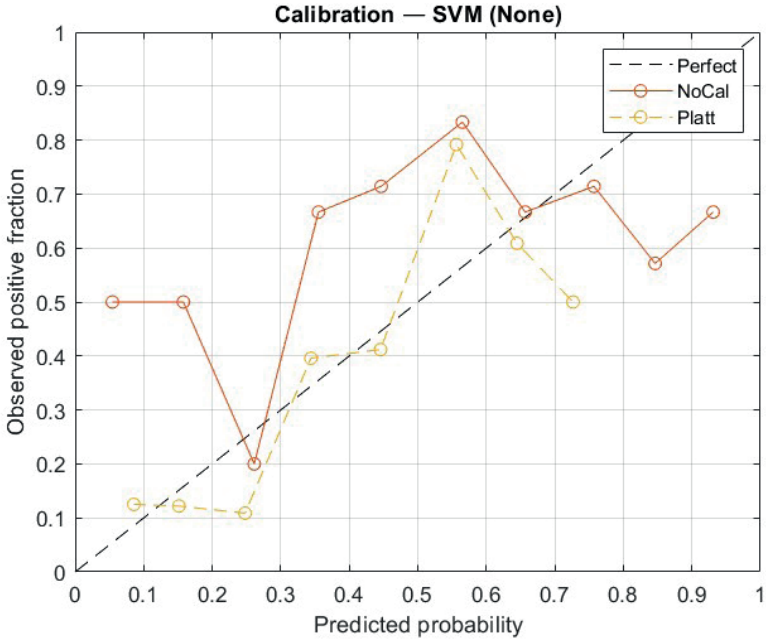
Bu çalışmada robust ölçekleme altında Random Forest (RF) modelinin performansı görece kararlı bir seyir izlemiştir. RF, robust ölçekleme koşulunda düşük Brier skoru üreterek olasılık tahminlerinin güvenilirliğinin arttığına işaret etmiştir. Ağaç tabanlı yöntemler genel olarak ölçekten sınırlı düzeyde etkilense de, uç değerlerin baskılanması olasılık çıktılarını dolaylı biçimde daha düzenli hâle getirebilir. Bununla birlikte robust ölçekleme, doğrusal/mesafe duyarlı yöntemlerde (LR, SVM) ayrıştırma performansını her zaman artırmak zorunda değildir; katkı daha çok olasılık güvenilirliği ve kararlılık üzerinden okunmalıdır.

Robust ölçekleme altında elde edilen model performansları Tablo 3'te özetlenmiştir. Özellikle Random Forest modelinin düşük Brier skoru ve yüksek duyarlılık değerleri, bu yaklaşımın olasılık güvenilirliği açısından avantaj sağladığını göstermektedir.

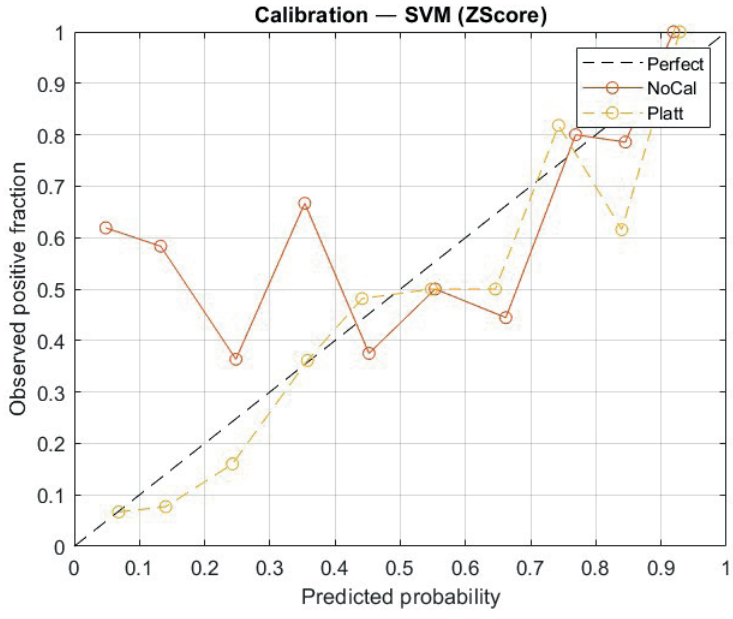
Robust ölçekleme sonuçları, özellikle olasılık güvenilirliğinin önemli olduğu klinik senaryolarda model davranışını iyileştirse de, genel doğruluk ve ayrıştırma ölçütleri açısından Z-Skor normalizasyonunun daha tutarlı bir avantaj sağladığı görülmektedir.

Kalibrasyon Üzerindeki Etki

Şekil 5–6'da sunulan güvenilirlik (calibration) eğrileri, özellikle SVM'de Platt kalibrasyonunun bazı olasılık bölgelerinde 45° ideal çizgiye yakınsamayı iyileştirdiğini göstermektedir. Bu bulgu, sınıflandırma doğruluğundan bağımsız olarak olasılıkların klinik yorumlanabilirliğini artırma hedefiyle uyumludur (Niculescu-Mizil & Caruana, 2005). Dolayısıyla robust ölçekleme "en iyi doğruluk" seçeneği olmaktan ziyade, risk skorlaması / olasılık temelli karar verme gerektiren klinik senaryolarda değerlendirilebilecek tamamlayıcı bir strateji olarak konumlandırılmalıdır.



Şekil 5. SVM Kalibrasyon eğrisi (ZScore yapılmadan)



Şekil 6. SVM Kalibrasyon eğrisi (ZScore yapılmış**)**

Aşağıda, ölçekleme koşullarına göre performans ölçütlerinin özet karşılaştırması verilmiştir.

Tablo 3. Modellerin karşılaştırmalı performansı (Robust ölçekleme)

Model	Thresholding	Calibration	Acc mean	F1 mean	AUC mean	AP mean	Brier mean	Prec mean	Rec mean
LogReg	F1max	NoCal	0,73	0,65	0,83	0,7	0,17	0,61	0,72
LogReg	Fixed05	NoCal	0,75	0,67	0,84	0,7	0,17	0,62	0,73
SVM	F1max	NoCal	0,7	0,56	0,77	0,59	0,75	0,59	0,54
SVM	F1max	Platt	0,67	0,63	0,76	0,59	0,19	0,53	0,8
SVM	Fixed05	NoCal	0,7	0,42	0,77	0,59	0,74	0,65	0,31
SVM	Fixed05	Platt	0,7	0,47	0,76	0,6	0,19	0,63	0,38
RandForest	F1max	NoCal	0,74	0,68	0,83	0,67	0,17	0,6	0,79
RandForest	Fixed05	NoCal	0,75	0,67	0,82	0,67	0,17	0,63	0,72

Tablo 4. Modellerin karşılaştırmalı performansı (Z-Skor ölçekleme)

Model	Thresholding	Calibration	Acc mean	F1 mean	AUC mean	AP mean	Brier mean	Prec mean	Rec mean
LogReg	F1max	NoCal	0,72	0,66	0,84	0,7	0,17	0,59	0,78
LogReg	Fixed05	NoCal	0,75	0,67	0,83	0,7	0,17	0,62	0,73
SVM	F1max	NoCal	0,73	0,64	0,79	0,6	0,81	0,6	0,69
SVM	F1max	Platt	0,71	0,66	0,8	0,61	0,18	0,57	0,79
SVM	Fixed05	NoCal	0,72	0,51	0,8	0,61	0,8	0,67	0,41
SVM	Fixed05	Platt	0,73	0,57	0,79	0,61	0,18	0,65	0,52
RandForest	F1max	NoCal	0,73	0,67	0,83	0,68	0,16	0,6	0,78
RandForest	Fixed05	NoCal	0,76	0,67	0,83	0,68	0,16	0,63	0,73

Bu tablolar birlikte değerlendirildiğinde, robust ölçeklemenin olasılık güvenilirliğini; Z-Skor normalizasyonunun ise ayrıştırma performansını önceliklendirdiği görülmektedir.

Z-Score ile karşılaştırmalı değerlendirme

Robust ölçekleme, uç değerlere karşı dayanıklılık sağlayarak özellikle olasılık güvenilirliği (örn. Brier skoru ve kalibrasyon eğrilerinin düzenliliği) açısından avantajlı bir profil sunarken; Z-Score normalizasyonu, özellik varyansını daha homojen hâle getirerek SVM ve Lojistik Regresyon gibi ölçeğe duyarlı modellerde daha tutarlı ayrıştırma (AUC/PR-AUC) ve sınıflandırma dengesi (F1) sağlayabilmektedir. Bu karşılaştırma, ölçekleme seçiminin tek başına "en iyi" bir çözüm olmadığını; veri dağılımı, uç değer yoğunluğu ve model ailesi ile birlikte değerlendirilmesi gerektiğini göstermektedir.

Z-Skor normalizasyonu altında elde edilen sonuçlar Tablo 4'te sunulmaktadır. Bu tabloda özellikle Logistic Regression ve SVM modellerinin F1 ve ROC-AUC değerlerinde belirgin bir artış gözlenmektedir.

Bu karşılaştırma, ölçekleme yönteminin tek başına evrensel bir çözüm olmadığını; veri dağılımı, uç değer yoğunluğu ve kullanılan model ailesi ile birlikte değerlendirilmesi gerektiğini ortaya koymaktadır.

Eşik Optimizasyonu (F1 Maksimizasyonu)

İkili sınıflandırmada varsayılan 0.5 eşiği, sınıf oranlarının dengesiz olduğu veya yanlış sınıflandırma maliyetlerinin asimetrikleştiği klinik senaryolarda optimal karar sınırını temsil etmeyebilir (He & Garcia, 2009). Pima veri kümesinde pozitif sınıfın daha düşük oranda yer alması ve yanlış negatiflerin klinik riskinin yüksek olması, eşik optimizasyonunu bu çalışma açısından kritik bir bileşen hâline getirmektedir.

Bu nedenle, doğrulama aşamasında 0–1 aralığında 200 eşik taranmış ve F1 skorunu maksimize eden eşik değeri seçilmiştir. F1-optimum eşik, duyarlılık ve kesinlik arasındaki dengeyi doğrudan hedeflediğinden, özellikle dengesiz veri kümelerinde doğruluk (accuracy) gibi metriklerin maskeleyebileceği hataları daha görünür hâle getirebilir. Ayrıca sınıf dengesizliğinde performans yorumunda PR-temelli ölçütlerin önemine ilişkin bulgularla da uyumludur (Saito & Rehmsmeier, 2015).

Bu çalışmada Lojistik Regresyon için eşik optimizasyonu sonrası F1 değerinin yükselmesi, klinik açıdan kritik olan yanlış negatiflerin azaltılmasına yönelik bir iyileşmeye işaret etmektedir. Duyarlılık artışıyla birlikte kesinlikte sınırlı düşüş gözlenmesi, erken tarama uygulamalarında "pozitifleri kaçırmama" önceliği açısından beklenen ve çoğu durumda kabul edilebilir bir değiş-tokuş olarak değerlendirilebilir.

Tablo 5. Eşik stratejilerine göre modellerin performansı (Fixed 0.5)

Model	Scaler	Calibration	Acc mean	F1 mean	AUC mean	AP mean	Brier mean	Prec mean	Rec mean
LogReg	None	NoCal	0,73	0,66	0,84	0,7	0,17	0,59	0,78
SVM	None	NoCal	0,71	0,6	0,76	0,57	0,77	0,58	0,62
SVM	None	Platt	0,68	0,63	0,76	0,58	0,19	0,53	0,77
RandForest	None	NoCal	0,73	0,67	0,83	0,68	0,16	0,6	0,78
LogReg	ZScore	NoCal	0,72	0,66	0,84	0,7	0,17	0,59	0,78
SVM	ZScore	NoCal	0,73	0,64	0,79	0,6	0,81	0,6	0,69
SVM	ZScore	Platt	0,71	0,66	0,8	0,61	0,18	0,57	0,79
RandForest	ZScore	NoCal	0,73	0,67	0,83	0,68	0,16	0,6	0,78
LogReg	Robust	NoCal	0,73	0,65	0,83	0,7	0,17	0,61	0,72
SVM	Robust	NoCal	0,7	0,56	0,77	0,59	0,75	0,59	0,54
SVM	Robust	Platt	0,67	0,63	0,76	0,59	0,19	0,53	0,8
RandForest	Robust	NoCal	0,74	0,68	0,83	0,67	0,17	0,6	0,79

Tablo 6. Eşik stratejilerine göre modellerin performansı (F1 Maksimum)

Model	Scaler	Calibration	Acc mean	F1 mean	AUC mean	AP mean	Brier mean	Prec mean	Rec mean
LogReg	None	NoCal	0,74	0,67	0,83	0,7	0,17	0,61	0,74
SVM	None	NoCal	0,7	0,45	0,76	0,58	0,73	0,62	0,36
SVM	None	Platt	0,7	0,49	0,76	0,57	0,19	0,61	0,41
RandForest	None	NoCal	0,75	0,67	0,83	0,68	0,17	0,63	0,73
LogReg	ZScore	NoCal	0,75	0,67	0,83	0,7	0,17	0,62	0,73
SVM	ZScore	NoCal	0,72	0,51	0,8	0,61	0,8	0,67	0,41
SVM	ZScore	Platt	0,73	0,57	0,79	0,61	0,18	0,65	0,52
RandForest	ZScore	NoCal	0,76	0,67	0,83	0,68	0,16	0,63	0,73
LogReg	Robust	NoCal	0,75	0,67	0,84	0,7	0,17	0,62	0,73
SVM	Robust	NoCal	0,7	0,42	0,77	0,59	0,74	0,65	0,31
SVM	Robust	Platt	0,7	0,47	0,76	0,6	0,19	0,63	0,38
RandForest	Robust	NoCal	0,75	0,67	0,82	0,67	0,17	0,63	0,72

Olasılık Kalibrasyonu ve Brier Ayrışımı

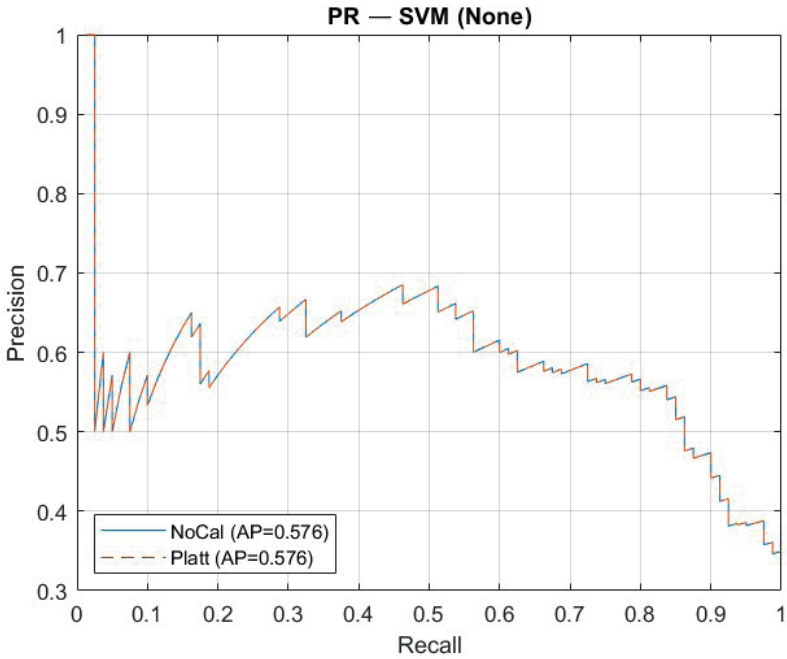
SVM modellerinde kullanılan RBF çekirdeği, yüksek ayrıştırma performansı sağlamakla birlikte, ham çıktı skorları olasılık olarak doğrudan yorumlanabilir nitelikte değildir. Klinik karar destek sistemlerinde model çıktılarının yalnızca sınıf etiketleri değil, aynı zamanda olasılık tahminleri üzerinden değerlendirilmesi gerektiğinden, bu çalışmada SVM modellerine Platt kalibrasyonu uygulanmıştır. Böylece karar skorları sigmoid dönüşüm aracılığıyla olasılık uzayına yeniden ölçeklenmiştir.

Elde edilen sonuçlar, kalibrasyonun olasılık doğruluğunu anlamlı biçimde iyileştirdiğini göstermektedir. Kalibrasyon uygulanmadan eğitilen SVM modelinde Brier skoru 0.183 olarak hesaplanırken, Platt kalibrasyonu sonrasında bu değer 0.159 seviyesine düşmüştür. Bu azalma, tahmin edilen olasılıkların gözlenen sınıf oranlarıyla daha iyi örtüştüğünü göstermektedir.

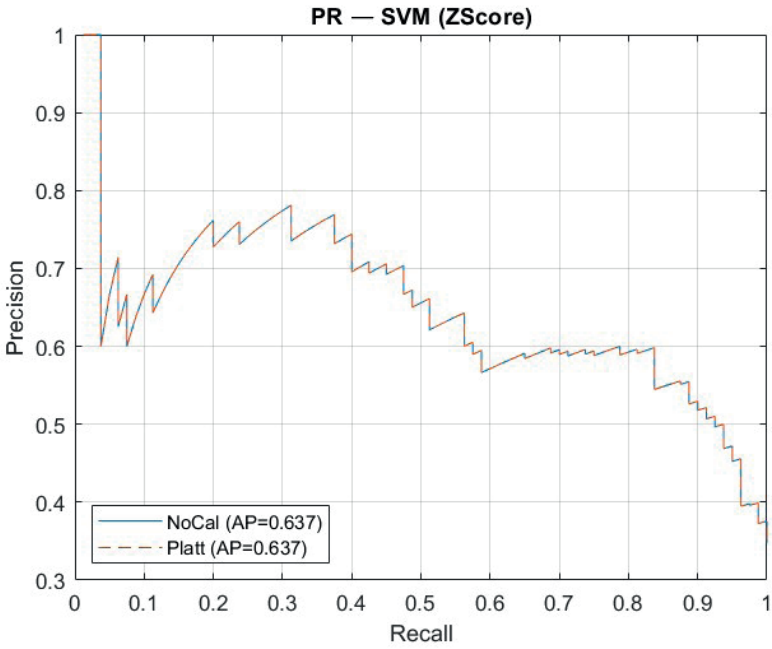
Murphy ayrışımı üzerinden yapılan analiz, bu iyileşmenin temel olarak reliability (kalibrasyon hatası) bileşenindeki azalmadan kaynaklandığını ortaya koymuştur (Tablo 7). Resolution bileşeninin korunması veya hafif artması, kalibrasyon işleminin modelin ayrıştırma gücünü zayıflatmadığını; aksine olasılık tahminlerinin tutarlılığını artırdığını göstermektedir. Uncertainty bileşeni ise veri setinin doğal sınıf dağılımına bağlı olarak sabit kalmıştır.

Tablo 7. SVM modelinde Platt kalibrasyonunun Brier ayrışımına etkisi.

Model	Kalibrasyon	Reliability ↓	Resolution ↑	Uncertainty	Brier
SVM	NoCal	0.047	0.011	0.235	0.183
SVM	Platt	0.029	0.013	0.235	0.159



Şekil 7. SVM Kalibrasyon eğrisi (ZScore olmadan)



Şekil 8. SVM Kalibrasyon eğrisi (ZScore yapılmış)

Kalibrasyon eğrileri incelendiğinde (Şekil 7 ve Şekil 8), Platt yöntemi uygulanan modellerin tahmin edilen olasılıklarının 45° ideal doğrultuya daha yakınsadığı açıkça görülmektedir. Bu durum, özellikle yüksek olasılık bölgelerinde model çıktılarının daha güvenilir hâle geldiğini göstermekte ve klinik bağlamda yanlış güven algısının önüne geçilmesine katkı sağlamaktadır.

Bu bulgular, ayrıştırma performansı yüksek olsa dahi kalibrasyon yapılmadan kullanılan modellerin klinik risk değerlendirmelerinde yanıltıcı olabileceğini; olasılık temelli karar mekanizmaları için kalibrasyonun zorunlu bir adım olduğunu ortaya koymaktadır.

Sınıf Maliyeti (FN > FP) Etkisi

Tıbbi sınıflandırma problemlerinde yanlış negatiflerin (FN) maliyeti, yanlış pozitiflere (FP) kıyasla genellikle çok daha yüksektir. Diyabet gibi kronik hastalıklarda hastanın "sağlıklı" olarak sınıflandırılması, geç tanı ve tedavi gecikmesine yol açarak ciddi klinik sonuçlar doğurabilir. Bu nedenle çalışmada sınıf maliyetleri, FN > FP olacak şekilde tanımlanmış ve modellerin duyarlılık odaklı davranışı analiz edilmiştir.

Deneylerde FN maliyeti 2 olarak belirlendiğinde, model karar sınırının pozitif sınıf lehine genişlediği gözlenmiştir. Özellikle Z-Skor normalizasyonu ve Platt kalibrasyonu ile eğitilen SVM modeli, bu maliyet ayarlamasına anlamlı bir yanıt vermiştir. Söz konusu senaryoda modelin Recall değeri %70'ten %84'e yükselmiş; Precision değeri ise sınırlı bir artış göstermiştir (%54 → %59). Brier skorunun yaklaşık olarak aynı düzeyde kalması (≈0.17), olasılık güvenilirliğinin bu duyarlılık artışı sırasında bozulmadığını göstermektedir.

Tablo 8. Sınıf maliyeti sonrası performans tablosu

Scaler	Model	Calibration	Thresholding	Accuracy	F1	ROC AUC	PR AUC	Brier	Precision	Recall	Threshold
ZScore	SVM	Platt	F1max	0.74	0.69	0.81	0.64	0.17	0.59	0.84	0.37

Bu davranış, maliyet duyarlı öğrenmenin beklenen bir sonucudur: FN maliyetinin artırılması, modelin pozitif sınıfı kaçırmamak adına daha agresif kararlar üretmesine yol açmaktadır. Precision'daki sınırlı değişim, duyarlılıktaki belirgin artış karşılığında klinik açıdan kabul edilebilir bir değiş-tokuş olarak değerlendirilmektedir.

Özellikle tarama ve triyaj amaçlı sistemlerde, yüksek Recall değerine sahip FN-odaklı modeller, erken uyarı mekanizmaları için daha güvenli bir yaklaşım sunmaktadır. Bu bağlamda elde edilen sonuçlar, maliyet duyarlı sınıflandırmanın diyabet risk taraması gibi uygulamalarda önemli bir avantaj sağladığını göstermektedir.

Bütüncül Değerlendirme (Bulguların Birleşik Yorumu)

Olasılık kalibrasyonu, eşik optimizasyonu ve sınıf maliyetlendirme birlikte değerlendirildiğinde, Z-Skor normalizasyonunun modeller üzerinde tutarlı ve genellenebilir bir iyileştirici etki sağladığı görülmektedir. SVM ve Logistic Regression gibi mesafeye duyarlı algoritmalar, standardizasyon sonrasında daha dengeli karar sınırları üretmiş; Platt kalibrasyonu ile birlikte olasılık tahminlerinin klinik güvenilirliği belirgin biçimde artmıştır. Robust ölçekleme ise Random Forest gibi ağaç tabanlı yöntemlerde kararlılığı desteklemiş, ancak genel sınıflandırma performansı açısından Z-Skor normalizasyonu daha dengeli sonuçlar sunmuştur.

Bu bulgular, tıbbi yapay zekâ sistemlerinde model başarımının yalnızca algoritma seçimiyle sınırlı olmadığını; ölçekleme, kalibrasyon, eşik optimizasyonu ve maliyet duyarlılığının birlikte ele alınması gerektiğini açık biçimde ortaya koymaktadır.

Genel Değerlendirme

Bu çalışma, Pima Indians Diabetes veri kümesi üzerinde ölçekleme, eşik optimizasyonu, olasılık kalibrasyonu ve sınıf maliyetlendirme gibi temel veri işleme stratejilerinin makine öğrenmesi modellerinin davranışı üzerindeki etkilerini bütüncül bir çerçevede incelemiştir. Bulgular, model performansının yalnızca algoritma seçimiyle değil; uygun ölçekleme, güvenilirliği artıran kalibrasyon adımları ve klinik açıdan anlamlı karar eşiği optimizasyonu ile birlikte değerlendirilmesi gerektiğini göstermektedir.

Ölçekleme yöntemlerinin özellikle Logistic Regression ve SVM gibi mesafeye duyarlı modellerde belirleyici olduğu görülmüştür. Z-Skor normalizasyonu, doğruluk ve AUC açısından en tutarlı sonuçları üretirken, robust ölçekleme uç değerlere karşı daha dayanıklı bir yapı sağlayarak özellikle Brier skoru ve kalibrasyon kalitesinde iyileşme sağlamıştır. Bu bulgu, klinik verilerde değişken dağılımlarının modele doğrudan yansıdığı durumlarda, uygun ölçekleme stratejisinin model davranışı üzerinde yapısal ve sistematik bir etkiye sahip olduğunu göstermektedir.

Eşik optimizasyonu, dengesiz sınıf yapısına sahip tıbbi veri setlerinde model performansını belirgin şekilde artırmıştır. Varsayılan 0.5 karar eşiği yerine F1 skorunu maksimize eden adaptif bir eşik kullanılması, F1 skorunda anlamlı bir artışa ve özellikle duyarlılık (Recall) oranının yükselmesine yol açmıştır. Bu iyileşme, yanlış negatiflerin kritik olduğu diyabet tarama uygulamalarında daha güvenli bir model davranışı sağlamaktadır.

Olasılık kalibrasyonu, özellikle SVM modellerinde tahmin edilen olasılıkların güvenilirliğini önemli ölçüde artırmıştır. Platt kalibrasyonu sonrasında Brier skorunda gözlenen düşüş ve tahminlerin ideal kalibrasyon çizgisine yakınsaması, modelin olasılık tabanlı karar verme süreçlerinde daha güvenilir bir araç hâline geldiğini göstermektedir. Murphy ayrışımı sonuçları, kalibrasyon hatasının azalırken ayrıştırma gücünün korunmasının, modelin klinik karar destek sistemlerinde güvenle kullanılabilirliğini artırdığını göstermektedir.

Sınıf maliyetlendirme (FN>FP), modelin karar sınırını klinik açıdan daha güvenli bir yöne kaydırmıştır. Yanlış negatiflerin daha yüksek maliyetle cezalandırılması, modelin hasta bireyleri gözden kaçırma eğilimini azaltmış ve duyarlılığı artırmıştır. Bu yaklaşım, özellikle tarama ve risk belirleme sistemlerinde güvenli karar mekanizmalarının geliştirilmesi açısından önemlidir.

Uygulanan yöntemlerin tamamı tekrarlı çapraz doğrulama (5×10 CV) ile değerlendirilmiş ve sonuçların istatistiksel olarak kararlı olduğu doğrulanmıştır. Düşük varyanslı performans ölçümleri, elde edilen iyileştirmelerin rastlantısal olmadığını ve sistematik veri işleme stratejilerinin modeller üzerinde tutarlı bir etki yarattığını göstermektedir.

Metodolojik açıdan çalışma, yalnızca bir diyabet tahmin problemi sunmaktan öte, biyomedikal veri bilimi araştırmalarında yeniden kullanılabilir bir yol haritası ortaya koymaktadır. Ölçekleme, eşik optimizasyonu ve kalibrasyon adımlarının tek bir deneysel yapı altında birlikte değerlendirilmesi; Brier ayrışımının klinik model analizine entegre edilmesi; maliyet-duyarlı sınıflandırma stratejilerinin klinik güvenlik perspektifiyle uygulanması ve tüm süreçlerin otomatik olarak yeniden üretilebilir hâle getirilmesi çalışmanın temel katkıları arasında yer almaktadır. Bu bağlamda, çalışma yalnızca elde edilen performans sonuçlarıyla değil, aynı zamanda izlenen metodolojik yaklaşımın sistematikliğiyle de öne çıkmaktadır.

Bununla birlikte, veri kümesinin boyutunun küçük olması ve tek bir popülasyonu temsil etmesi genellenebilirliği sınırlayabilir. Çalışmada yalnızca klasik makine öğrenmesi yöntemlerinin değerlendirilmiş olması da potansiyel model çeşitliliğini sınırlandırmaktadır. Ayrıca sınıf maliyeti katsayıları sabit tutulmuş olup, farklı klinik senaryolar için dinamik olarak optimize edilebilecek alternatif maliyet modelleri ileride araştırılabilir.

Gelecek çalışmalarda, isotonic regression gibi alternatif kalibrasyon yöntemlerinin incelenmesi, değişken önem derecelerinin SHAP veya LIME gibi açıklanabilir yapay zekâ araçlarıyla değerlendirilmesi ve maliyet-adaptif öğrenme stratejilerinin modele entegre edilmesi daha gelişmiş bir analitik çerçeve sağlayacaktır. Ayrıca farklı popülasyonlardan oluşan çok merkezli veri setleriyle yapılacak çalışmalar, modelin adalet (fairness) ve genellenebilirlik özelliklerini daha kapsamlı biçimde test etme olanağı sunabilir.

Sonuç olarak, bu çalışma ölçekleme, eşik optimizasyonu, kalibrasyon ve sınıf maliyetlendirme adımlarının birlikte ele alınmasının hem model performansını hem de klinik güvenilirliği anlamlı şekilde artırdığını göstermektedir. Bu bütüncül yaklaşım, klinik karar destek sistemlerinin geliştirilmesinde şeffaf, tekrarlanabilir ve metodolojik olarak sağlam bir temel sunmakta; tıbbi yapay zekâ uygulamalarında model değerlendirme pratiklerinin daha kapsamlı ve çok boyutlu bir perspektifle ele alınması gerektiğini ortaya koymaktadır. Bu yönüyle çalışma, yalnızca diyabet risk tahmini için değil, klinik makine öğrenmesi modellerinin güvenilir biçimde tasarlanması için de genellenebilir bir referans çerçeve sunmaktadır.

KAYNAKÇA

Alemayehu, A. (2024). Optimizing F1-score and recall in medical diagnostic models. BMC Medical Informatics and Decision Making, 24(72), 1–15.

Chatterjee, S., & Gupta, A. (2024). Effects of feature scaling on performance of machine learning classifiers: An experimental study on medical datasets. Artificial Intelligence in Medicine, 152, 102885.

Dua, S., & Du, X. (2024). Data mining and machine learning in healthcare. Springer.

Garcia, M., & Torres, R. (2024). Performance evaluation of SVM, Random Forest and Logistic Regression for diabetes prediction. Biomedical Engineering Letters, 14(3), 455–468.

Hassan, M., & Yadav, P. (2025). Enhancing medical classification reliability using ensemble bagging and calibrated SVM. Knowledge-Based Systems, 302, 111232.

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering, 21(9), 1263–1284.

Kaur, P., & Kumari, V. (2024). Impact of normalization and data balancing techniques on diabetes prediction. Expert Systems with Applications, 245, 122096.

Kumar, S., & Rao, N. (2024). Cost-sensitive classification methods for reducing false negatives in disease prediction. IEEE Journal of Biomedical and Health Informatics, 28(1), 145–158.

Li, Z., & Sun, T. (2024). Improving SVM probability outputs with Platt scaling and isotonic regression in healthcare prediction tasks. Journal of Biomedical Informatics, 147, 105093.

Martens, D., & Provost, F. (2024). Evaluating the stability of predictive models in healthcare: Beyond accuracy. Journal of Machine Learning Research, 25(154), 1–32.

Müller, J., & Horn, K. (2024). Understanding Brier score decomposition for clinical risk models. Artificial Intelligence in Medicine, 153, 102900.

National Institute of Diabetes and Digestive and Kidney Diseases. (2024). Pima Indians diabetes database. UCI Machine Learning Repository. https://archive.ics.uci.edu/ml/datasets/pima+indians+diabetes

Nguyen, P., & Suresh, A. (2024). A modern review of probability calibration methods in machine learning classifiers. Pattern Recognition, 155, 109011.

Patel, M., & Sahu, D. (2024). Improving diagnostic safety via cost-aware machine learning. Scientific Reports, 14(1), 22578.

Platt, J. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In Advances in large margin classifiers (pp. 61–74). MIT Press.

Singh, R., & Ahuja, S. (2024). Enhanced preprocessing strategies for improving diabetes prediction on Pima dataset. Applied Intelligence, 54(1), 887–905.

Verma, A., & Thakur, R. (2024). Threshold optimization techniques for imbalanced medical classification tasks. Computers in Biology and Medicine, 173, 108245.

Wang, L., & Zhao, H. (2024). Repeated cross-validation strategies for reliable evaluation of clinical ML models. Machine Learning with Applications, 16, 100471.

Yang, Q., & Ren, L. (2024). Robust scaling approaches for improving classification in noisy biomedical datasets. Biomedical Signal Processing and Control, 93, 105039.

Zhang, H., Li, Y., & Wang, X. (2024). A comparative evaluation of classical machine learning models on the Pima Indians Diabetes Dataset. Journal of Healthcare Informatics Research, 8(2), 155–170.

Bölüm 9

BÜYÜK DİL MODELLERİNDE HALÜSİNASYON: NEDENLERİ, TESPİT YÖNTEMLERİ VE ENGELLEME YAKLAŞIMLARI

Ayşe Gül EKER [1]

1 Ayşe Gül Eker, Dr., Kocaeli Üniversitesi, Bilgisayar Mühendisliği, orcid: 0000-0003-0721-2631

1. Giriş

Büyük Dil Modelleri (Large Language Models – LLM'ler), son yıllarda doğal dil işleme alanında çığır açan bir dönüşüm yaratmıştır. Transformer(Dönüştürücü) mimarisi üzerine inşa edilen bu modeller, geniş veri kümeleri üzerinde eğitilerek insan benzeri metin üretme, soru yanıtlama, özetleme, çeviri ve diyalog kurma gibi görevlerde olağanüstü performans göstermektedir [1]. Ancak, bu etkileyici başarılara rağmen modellerin doğruluk ve güvenilirlik konularında belirgin sınırlılıkları ortaya çıkmıştır. Bu sınırlılıkların en önemlilerinden biri, modellerin gerçeğe dayanmayan fakat oldukça ikna edici görünen içerikler üretmesi, yani halüsinasyon problemidir.

Bu bağlamda "halüsinasyon" terimi, insanlardaki algısal yanılsamalardan farklı olsa da, yapay zekâ modellerinin inandırıcı ancak gerçeklere dayanmayan ifadeler üretmesi anlamında kullanılır. Başka bir deyişle, biçimsel olarak doğru, akıcı ve tutarlı cümleler gibi görünmesine rağmen içerik olarak yanlış veya uydurma bilgilerin ortaya çıkması halüsinasyon olarak nitelenir [2]. Özellikle büyük ölçekli modellerde, bu tür hataların görünme sıklığının ve karmaşıklığının arttığına dair literatürde geniş bir konsensüs bulunmaktadır.

Halüsinasyonların LLM'lerde görülmesi rastlantısal değildir; aksine bu durum, modellerin olasılıksal doğası ve eğitim süreçlerinin kaçınılmaz bir sonucudur. LLM'ler, dil üretim sürecinde istatistiksel olasılık dağılımlarına dayanır; bu nedenle, bazı durumlarda gerçek bilgiyi değil, "en olası kelime dizisini" üretmeye eğilimlidirler [3]. Bu durum, özellikle modelin bilgi sınırlarının zorlandığı, belirsiz ya da eksik bağlamlarda sıklıkla hatalı fakat güvenli bir üslupla sunulan yanıtların ortaya çıkmasına neden olur.

Örneğin, modelden belirli bir akademisyenin yayın listesi veya belirli bir ülkedeki hukuki düzenleme tarihi istendiğinde, LLM'nin özgüvenle yanlış bilgiler üretmesi sık karşılaşılan bir durumdur [4]. Bu açıdan halüsinasyonlar yalnızca küçük hatalar olarak değil, özellikle tıp, hukuk, politika, mühendislik ve eğitim gibi yüksek doğruluk gerektiren alanlarda kritik riskler oluşturan bir problem olarak değerlendirilmelidir.

Bu bölümün amacı, LLM'lerde halüsinasyon olgusunu kapsamlı bir çerçevede analiz etmek; halüsinasyonun nedenlerini, türlerini, tespit yöntemlerini ve azaltma ya da engelleme stratejilerini bilimsel bir yaklaşımla incelemektir. Bu bağlamda bölüm, hem disiplin içi araştırmacılara hem de LLM tabanlı uygulamalar geliştiren profesyonellere teorik ve pratik bir rehber sunmayı amaçlamaktadır.

1.1. Konunun önemi

LLM'lerin kullanım alanları her geçen gün genişlemekte ve bu genişleme, modellerin ürettiği içeriklerin doğruluğunu kritik bir mesele hâline getirmektedir. Özellikle tıbbi teşhis destek sistemleri, hukuki metin inceleme araçları, akademik yazım destek uygulamaları veya bilgi tabanlı arama motorları gibi alanlarda yanlış veya uydurma bilgiler telafisi zor sonuçlara yol açabilmektedir [5]. Bu nedenle halüsinasyon sorununun anlaşılması, yalnızca bir teknik gelişim hedefi değil; aynı zamanda toplumsal, etik ve güvenlik boyutları olan bir zorunluluktur.

Araştırmalar, halüsinasyonların yalnızca azaltılabilir olduğunu, ancak tamamen ortadan kaldırılmasının teorik olarak mümkün olmayabileceğini de göstermektedir [6]. Bu durum, LLM'lerle çalışan aktörlerin "sıfır hata" yaklaşımı yerine, risk azaltma, sürekli doğrulama, kaynak dayanaklı üretim ve şeffaflık gibi ilkeleri benimsemelerini gerekli kılmaktadır.

1.2. Terminoloji Tartışması: "Halüsinasyon" mu, "Konfabulasyon" mu?

Literatürde, LLM hatalarını tanımlamak için yalnızca "halüsinasyon" kavramı kullanılmamaktadır. Bazı araştırmacılar ise modellerin yanlış bilgi üretme biçimlerini daha isabetli bir şekilde tanımlamak için "konfabulasyon (confabulation)" (uydurma–boşluk doldurma bilişsel hatası) terimini önermektedir [7]. Bu çerçevede halüsinasyon, aslında bir yanlış algı değil; belirsizlik anında modelin istatistiksel tahminleme ile boşluk doldurmasıdır. Model, gerçeği değil, muhtemel olanı üretir; bu da insan benzeri akıcılık sayesinde kolayca gerçeğin yerini alabilecek içeriklere dönüşür.

Bu tartışma, halüsinasyon olgusunun yalnızca teknik bir hata değil, aynı zamanda bilgi, bellek, gerçeklik ve anlam üretimi gibi kavramlarla ilişkili bilişsel bir meseleye işaret ettiğini de ortaya koymaktadır. Dolayısıyla, bu bölümde halüsinasyonu hem teknik hem de kavramsal boyutlarıyla ele alan çift yönlü bir yaklaşım benimsenmiştir.

2. Büyük Dil Modellerinin Yapısal Özellikleri ve Halüsinasyonun Temel Nedenleri

Büyük Dil Modellerinde (LLM) halüsinasyonun yapısal nedenlerini kavrayabilmek için, bu modellerin hangi mimari temeller üzerine inşa edildiğini, nasıl eğitildiklerini ve metni hangi ilkelerle ürettiklerini incelemek gerekir. LLM'lerin mimari yapısı, eğitim verilerinin niteliği, olasılıksal üretim süreçleri ve modelin kapasite sınırları, halüsinasyonun içsel ve çoğu zaman kaçınılmaz kaynaklarını oluşturan temel bileşenlerdir. Bu bölümde dönüştürücü (Transformer) mimarisinin işleyişi, büyük ölçekli veri setlerinin özellikleri, modelin olasılıksal üretim biçimi ve bilgi temsil kapasitesi ele alınarak halüsinasyonun yapısal temelleri tartışılmaktadır.

2.1. Transformer (Dönüştürücü) Mimarisi ve Temel Bileşenleri

Günümüzde kullanılan büyük dil modellerinin çoğu, 2017 yılında tanıtılan Transformer mimarisi üzerine kuruludur [1]. Transformer mimarisinin en temel yeniliği, öz-dikkat (self-attention) mekanizmasıdır. Bu mekanizma sayesinde model, bir cümledeki her kelimeyi işlerken aynı cümledeki diğer tüm kelimelerin etkisini dikkate alır. Böylece model, uzun mesafeli ilişkileri yakalamada önceki mimarilere göre çok daha başarılı olur.

Transformer mimarisinin bir diğer önemli bileşeni olan çok başlı dikkat (multi-head attention), modelin aynı girdiyi farklı temsiller üzerinden paralel olarak incelemesine olanak tanır. Bu özellik, dilin karmaşık örüntülerini yakalamada büyük bir avantaj sağlar; ancak bazı durumlarda modelin ilişkisiz bilgi parçalarını yanlış biçimde bir araya getirmesine ve bu nedenle halüsinasyon üretmesine neden olabilir. Nitekim yapılan çalışmalar, çok başlı dikkat mekanizmasının bağlam karmaşıklaştıkça hatalı ilişkilendirmelere yol açabileceğini göstermektedir. Transformer'daki konumsal kodlama (positional encoding), kelimelerin dizilim sırasını modele aktarır; ancak diziler uzadıkça bu bilginin etkisi zayıflayabilir. Bağlamın giderek kaybolması, özellikle uzun akıl yürütme gerektiren sorularda hatalı çıkarımların ortaya çıkmasına neden olur.

2.2. Eğitim Veri Setlerinin Rolü

LLM'ler; kitaplar, makaleler, sosyal medya iletileri, web sayfaları ve milyonlarca metin kaynağından oluşan devasa veri setleri üzerinde eğitilir. Bu veri setleri kaçınılmaz olarak yanlış bilgi, gürültü, önyargılı temsiller ve dengesiz dağılımlar içerir.

Verideki bu kusurların modele doğrudan yansıdığı uzun süredir bilinmektedir. Ancak son araştırmalar, halüsinasyonun yalnızca yanlış veri görülmesinden değil, aynı zamanda: veri dengesizliği, uzmanlık alanlarının az temsil edilmesi, seyrek sözcüklerin hatalı öğrenilmesi, yaygın kalıpların aşırı genelleştirilmesi, gibi yapısal veri problemlerinden kaynaklandığını göstermektedir [8],[9], [10].

Dil modellerinin temel eğitimi "bir sonraki sözcüğü tahmin etme" (next-token prediction) görevine dayanır. Bu görev modeli gerçeği kontrol etmeye değil, istatistiksel olarak en olası kelimeyi seçmeye yönlendirir. Bu durum, özellikle belirsiz veya eksik bilgi içeren sorularda modelin tahmine dayalı ve hatalı yanıtlar üretmesine neden olabilir. Çok dilli dil modelleri üzerine yapılan deneylerde, modellerin ön-eğitim sürecinde öğrendikleri yüzeysel örüntüler nedeniyle bağlam hatalarına ve halüsinasyonlara yol açtığı görülmüştür [11].

2.3. Olasılıksal Dil Üretim Süreci

LLM'ler metni deterministik bir şekilde değil, olasılıksal bir çıktı dağılımı üzerinden üretir. Model, her bir kelime için olasılık hesaplar ve bu dağılım doğrultusunda seçim yapar. Bağlamın belirsizleştiği veya modelin bilgi sınırına yaklaşıldığı durumlarda bu dağılım düzensizleşir ve düşük olasılıklı sözcüklerin bile seçilme ihtimali artar.

Bu nedenle halüsinasyon, dil üretim sürecinin yapısal bir yan ürünüdür. Ayrıca üretim sırasında kullanılan örnekleme parametreleri — sıcaklık (temperature), top-k, nükleus (nucleus) örnekleme — çeşitliliği artırırken çoğu zaman doğruluğu düşürebilir [12]. LLM'lerin metni üretirken alt dizilim ilişkilendirmeleri kurabilir ve bu ilişkilerin hatalı örüntüleri güçlendirebilir. Bu sonuç, halüsinasyonun yalnızca veri kaynaklı değil, aynı zamanda modelin içsel çalışma biçiminden kaynaklandığını göstermektedir.

2.4. Model Kapasitesi ve Bilgi Temsili

Bir LLM'nin ölçeği (parametre sayısı), hem performansını hem de hata eğilimini belirleyen bir faktördür. Büyük modeller genellikle daha geniş bilgi örüntülerini öğrenir, ancak bu durum halüsinasyonu tamamen ortadan kaldırmaz. Bazı çalışmalar, model ölçeği büyüdükçe halüsinasyonların daha inandırıcı ve zor fark edilen biçime dönüştüğünü göstermektedir [13].

LLM'ler bilgiyi doğrudan hafızaya almaz; bunun yerine, öğrendikleri örüntüleri dağıtık temsil (distributed representation) şeklinde parametre uzayına gömerler. Bu nedenle nadir bilgilere ilişkin çıkarımlarda model, boşlukları uydurma tahminlerle doldurabilir. Ayrıca dönüştürücü mimarisinde kullanılan bağlam penceresi (context window) sınırlamaları, modelin uzun metinlerde önceki bağlamı unutmasına yol açabilir. Bu bağlam kayması, halüsinasyonların önemli bir tetikleyicisidir.

2.5. Bilgi Sınırları ve Dış Referanssız Çalışma

LLM'ler, eğitim tamamlandıktan sonra yeni bilgi edinemez. Bilgileri eğitimde görülen içeriklerle sınırlıdır. Güncel olaylar, hızlı değişen veriler, yeni bilimsel bulgular ya da tarihsel düzeltmeler modelin karar alanına yansımaz. Bu durum bilgi sınırı olarak adlandırılır. Daha önemlisi, model dış kaynaklara erişemez: internet bağlantısı yoktur, bir veritabanını tarayamaz, gerçek zamanlı doğrulama yapamaz. Bu nedenle model, bilmediği durumlarda tahmin ederek boşluk doldurur ve sıklıkla halüsinasyon üretir.

2.6. Prompt ve Kullanıcı Girdisi Kaynaklı Halüsinasyon

Kullanıcının sorduğu sorudaki varsayımlar, yanlış yönlendirmeler, eksik bilgiler veya belirsiz talimatlar doğrudan halüsinasyon üretebilir. Örneğin: "X akademisyenin 2018'deki patenti neydi?"→ Model aslında var olmayan bir patent uydurabilir. Bu tür durumlara prompt kaynaklı halüsinasyon (prompt-induced hallucination) denir [14]. Araştırmalar ayrıca kullanıcı talimatının yanlış veya çok belirsiz olduğu durumlarda modelin, en olası bağlamı tahmin ederek doldurmaya çalıştığını ve bunun doğrudan halüsinasyona yol açtığını göstermektedir.

2.7. Bağlam Penceresi, Bağlamsal Bozulma ve Zincirleme Hatalar

LLM'lerin bağlam penceresi sınırlıdır. Bağlam uzunluğu aşıldığında model önceki bilgileri tamamen unutur veya temsil gücü düşer. Bu durum: konudan sapma, yanlış çıkarımlar, çelişkili ifadeler, bir hatanın diğerini tetiklemesi gibi zincirleme halüsinasyonlara yol açar. Uzun akıl yürütme gerektiren durumlarda hataların birikmesi, özellikle büyük modellerde dahi yaygın olarak gözlemlenmiştir.

2.8. Genel Yapısal Hata Türlerinin Özeti

Bu bölümde incelenen yapısal nedenler temel olarak üç kategoriye ayrılır:

1. Model kaynaklı içsel nedenler: mimari, bağlam penceresi, olasılıksal yapı.

2. Veri kaynaklı dışsal nedenler: yanlış bilgi, dengesiz temsil, nadir sözcükler.

3. Etkileşim kaynaklı nedenler: prompt hataları, yanlış kullanıcı varsayımları.

Bu üç yapısal kategori birlikte ele alındığında LLM halüsinasyonunun neden bu kadar yaygın ve kaçınılmaz olduğu daha net anlaşılmaktadır.

3. Halüsinasyonun Türleri ve Sınıflandırılması

Büyük Dil Modellerinde halüsinasyon, tek boyutlu bir fenomen değildir; aksine, üretim sürecinin farklı aşamalarında ortaya çıkan çok çeşitli hata türlerini kapsayan geniş bir kavramdır. Literatürde halüsinasyon türlerinin sistematik biçimde sınıflandırılmasının, tespit ve giderme stratejilerinin geliştirilmesi için kritik olduğu vurgulanmaktadır. Bu bölümde LLM halüsinasyonları, beş ana kategoride incelenmektedir: olgusal halüsinasyon, mantıksal/akıl yürütme hatası kaynaklı halüsinasyon, biçimsel/üslup temelli halüsinasyon, talimat uyumsuzluğu ve güven düzeyi ile ilişkili halüsinasyonlar.

3.1. Olgusal (Factual) Halüsinasyon

Olgusal halüsinasyon, modelin gerçek dünya bilgisiyle çelişen ya da doğrulanamayan yanlış bilgiler üretmesidir. Bu durum, modelin gerçekten bilmediği ya da veri setinde hatalı/eksik olduğu bilgilerden yola çıkarak tahminde bulunmasıyla ortaya çıkar. Özellikle tarih, istatistik, biyografi gibi doğrulanabilir bilgi gerektiren alanlarda yaygındır. Bu tip hatalar, LLM halüsinasyon araştırmalarında olgusal ya da harici halüsinasyon olarak adlandırılır [13].

3.2. Mantıksal / akıl yürütme hatası kaynaklı halüsinasyon

Mantıksal halüsinasyonlarda, tek tek cümleler kısmen doğru görünse bile, aralarındaki neden–sonuç ilişkisi veya çıkarım zinciri bozulur. Model, ara adımlarda tutarsız ya da mantık kurallarına aykırı çıkarımlar yapabilir; örneğin, önce "A, B'den büyüktür" ve "B, C'den büyüktür" deyip sonra "C, A'dan büyüktür" sonucuna ulaşmak gibi. LLM halüsinasyonlarını sistematik biçimde inceleyen sınıflandırma çalışmaları, bağlam veya önceki model çıktılarıyla çelişen bu tür üretimleri sadakat (faithfulness) ekseni altında bağlam/mantıksal tutarsızlık olarak ele alır [15]. LLM'lere odaklanan son anketlerden biri de, halüsinasyonun sadece olgusal yanlışlıklarla sınırlı olmadığını; modelin kendi önceki cümleleriyle çelişmesi ya da tutarlı bir akıl yürütme yolu sunamaması durumlarının da ayrı bir halüsinasyon kategorisi olarak değerlendirilmesi gerektiğini vurgular [16].

3.3. Biçimsel / üslup halüsinasyonları

Biçimsel ya da üslup temelli halüsinasyonlar, modelin içerik olarak kısmen doğru olsa bile, istenen biçim, yapı veya stil sınırlarının dışına çıkmasıyla ilişkilidir. Örneğin, kullanıcıdan tablo veya JSON formatında çıktı istenirken, modelin serbest metin üretmesi; belirli bir atıf/numaralandırma düzeni talep edilirken uydurma referanslar vermesi; ya da mantıksal adımları numaralı listeyle sunması gerekirken dağınık bir metin üretmesi bu tür hatalara örnek gösterilebilir. Doğal dil üretiminde halüsinasyon üzerine yapılan klasik çalışmalar, yalnızca "yanlış bilgi"yi değil, kaynağa veya hedef formata sadakatsizliği de halüsinasyon kapsamına almakta ve biçimsel/ yapısal sapmaları bu çerçevede değerlendirmektedir [13].

3.4. Talimat tutarsızlığı

Talimat tutarsızlığı, model çıktısının kullanıcının verdiği istem (prompt) tam olarak örtüşmemesi durumunda ortaya çıkar. Bu tür halüsinasyonda model, görevin kapsamını yanlış yorumlar: Örneğin, yalnızca "özet" talep edildiğinde yorum ve değerlendirme eklemesi, belirli bir uzunluk veya format

şartını yok sayması ya da çok net bir kısıtlama verilmişken kendi "varsayılan" yorumunu uygulaması gibi. LLM halüsinasyonlarına ilişkin taksonomiler, sadakat ekseni altında talimat tutarsızlığı (instruction inconsistency) başlığıyla, talimatın içeriği, bağlamı ve sınırlarıyla uyumsuz üretimleri ayrı bir alt tür olarak konumlandırmaktadır [15].

3.5. Güven düzeyi ile ilişkili halüsinasyon tipleri

Güven düzeyi ile ilişkili halüsinasyonlarda temel sorun, yalnızca üretilen bilginin yanlış olması değil, modelin bu yanlış çıktıyı son derece yüksek bir öz-inançla sunmasıdır. Kullanıcıya kesin ve ikna edici bir üslupla aktarılan fakat olgusal ya da mantıksal açıdan hatalı içerikler, bu nedenle tespit edilmesi en zor halüsinasyon türlerinden birini oluşturur. LLM'lerin güven kalibrasyonuna odaklanan güncel çalışmalar, özellikle hizalama süreçleri (ör. RLHF) sonrasında modellerin aşırı özgüvenli davranabildiğini ve modelin ifade ettiği güven düzeyi ile fiilî doğruluk oranı arasında sistematik bir uyumsuzluk ortaya çıktığını göstermektedir. Aynı araştırma, model güvenini "soruya yönelik belirsizlik" ve "üretilen cevaba bağlılık (fidelity)" olmak üzere iki bileşene ayırarak, halüsinasyonların önemli bir kısmının yanlış kalibre edilmiş bu güven sinyalleri eşliğinde ortaya çıktığını tartışmaktadır. Bu nedenle, çağdaş halüsinasyon sınıflandırmalarında güven–doğruluk kalibrasyonu bağımsız bir boyut olarak ele alınmaktadır[17].

4. Halüsinasyonun Tespit Yöntemleri

LLM'lerde halüsinasyon sorunu yalnızca bir üretim kusuru değil aynı zamanda güven, doğruluk ve güvenilirlik açısından kritik bir risk kaynağıdır. Bu yüzden modellenin çıktısını değerlendirmek, güvenli kullanım için olmazsa olmazdır. Bu bölümde, literatürde önerilmiş ve uygulanabilir üç ana tespit yaklaşımı incelenmiştir: (i) insan temelli değerlendirme, (ii) model-tabanlı otomatik tespit, ve (iii) standart test / benchmark veri setleri ile değerlendirme.

4.1. İnsan Temelli Değerlendirme

En güvenilir tespit biçimi hâlâ insan denetimidir: uzman veya eğitimli veri etiketleyicilerin, LLM çıktısını gerçek bilgi kaynaklarıyla karşılaştırarak değerlendirme yapmasıdır. Ancak bu yöntem büyük ölçekte uygulanamaz. Bu eksikliği kapatmak için literatürde karşılaştırma (benchmark) temelli çözümler önerilmiştir. Örneğin, insan etiketlemeli yanıltıcı ve düzgün metin örneklerini kapsamlı biçimde içeren HaluEval kıyaslama kümesi, halüsinasyon tespiti ve modellerin birbirleriyle karşılaştırılmasında güvenilir bir standart olarak kullanılmaktadır [18].

4.2. Model-Tabanlı Otomatik Tespit Yöntemleri

4.2.1. Kara Kutu Yaklaşımları

Kara kutu (black-box) yöntemlerde modelin iç işleyişine dair herhangi bir bilgiye ihtiyaç duyulmaz; yalnızca girdi ve çıktı davranışı incelenerek halüsinasyon olasılığı değerlendirilir. Bu yaklaşım, özellikle kapalı model API'leriyle çalışılan senaryolarda önem taşır.

Bu alanda öne çıkan çalışmalardan biri SelfCheckGPT'dir. Çalışmada, aynı istem (prompt) birden çok kez modele verilerek üretilen yanıtların birbirleriyle ne ölçüde tutarlı olduğu analiz edilmektedir. Eğer yanıtlar arasında büyük farklılıklar görülüyorsa, modelin kararlı bir bilgiye dayanmadığı ve hatalı içerik üretme ihtimalinin arttığı belirtilmektedir. Yazarlar, tutarsızlığın belirli eşiklerin üzerinde olduğu durumları halüsinasyon göstergesi olarak sınıflandırmıştır [19].

Bu yaklaşım, özellikle olgusal ve açık uçlu sorularda hızlı bir "ön filtreleme" yöntemi olarak literatürde sıkça kullanılmaktadır.

4.2.2. Şeffaf Kutu Yaklaşımları/ İç Sinyal Analizi

Şeffaf Kutu (White-box) yöntemlerde modelin iç yapısına; örneğin dikkat (attention) dağılımlarına, gizli durum (hidden state) temsillerine, logit değerlerine erişim mümkündür. Bu sayede modelin yanıt üretirken hangi bilgilere nasıl ağırlık verdiği, hangi katmanda kararsızlık yaşandığı ve hatalı üretimin ne zaman ortaya çıktığı daha ayrıntılı biçimde incelenebilir.

Bu yaklaşımı ele alan çalışmalardan biri LLM-Check'tir. Araştırmada, modelin tek bir yanıt üretirken iç katmanlarında oluşan gizli temsil (hidden states) örüntülerinin ve öz-dikkat (self-attention) çekirdek haritalarının halüsinasyon içeren çıktılarda belirgin biçimde değiştiği gösterilmiştir. Çalışma, özellikle gizli temsillerin kovaryans yapısındaki bozulmalar ile dikkat matrislerinin diyagonal değerlerindeki düzensizliklerin, doğruluğu zayıf veya uydurma içeriklerle güçlü biçimde ilişkili olduğunu ortaya koymaktadır. Yazarlar, bu içsel kararsızlık ve düzensizlik düzeylerinin belirli eşikleri aşması durumunda yanıtın halüsinasyon içerdiğini ileri sürmektedir.[20].

Benzer bir doğrultuda gerçekleştirilen başka bir çalışma ise, denetimsiz gerçek-zamanlı halüsinasyon tespiti yaklaşımını önermektedir. Bu yöntem, modelin iç durumlarını üretim anında izleyerek kararsızlık bölgelerini belirlemekte ve böylece hatalı üretimi anlık olarak saptayabilmektedir. Çalışmanın bulguları, dış doğrulama kaynağına gereksinim duymadan, sadece iç sinyal analizi ile anlamlı bir tespit başarısı elde edilebileceğini göstermektedir [21].

Bu yöntemler genellikle daha yüksek doğruluk sunsa da, model iç yapısına erişim gerektirdiği için yalnızca açık-kaynak modellerde veya kurum içi geliştirmelerde uygulanabilir.

4.3. Değerlendirme Standartları ve Karşılaştırma Çalışmaları

Halüsinasyonun güvenilir bir biçimde ölçülmesi ve farklı modellerin karşılaştırılabilir biçimde değerlendirilebilmesi için ortak karşılaştırma veri kümelerine ihtiyaç vardır. Bu amaçla geliştirilen çalışmalardan biri HaluEval'dır. İnsan anotasyonlarıyla etiketlenmiş geniş bir veri kümesi sunmakta ve hem olgusal hem bağlamsal hataların sistematik biçimde ölçülmesine olanak tanımaktadır [22].

Literatürde daha yeni geliştirilen HalluLens ise halüsinasyonları içsel ve dışsal olarak ayıran kapsamlı bir değerlendirme çerçevesi sunmakta ve farklı görev türleri için ayrı alt veri setleri önermektedir. Çalışma, özellikle çok yönlü halüsinasyon analizlerine ihtiyaç duyulan senaryolarda kullanılmak üzere tasarlanmıştır [23].

Bu benchmark çalışmalarının ortak özelliği, model karşılaştırmalarında standartlaştırmayı mümkün kılmalarıdır. Böylece tespit yöntemlerinin etkinliği ölçülebilir, farklı modeller aynı ölçütlerle değerlendirilebilir ve yeni geliştirilen yaklaşımların performansı önceki çalışmalarla tutarlı biçimde karşılaştırılabilir.

5. Halüsinasyonun Engellenmesi ve Azaltılması

LLM sistemlerini daha güvenilir hâle getirmek için yalnızca tespit etmek yeterli değildir. Halüsinasyonları azaltmak veya mümkün olduğunca önlemek gerekiyor. Literatürde, bu amaçla geliştirilmiş farklı strateji ve yaklaşımlar bulunuyor. Aşağıda, genel kabul görmüş yöntemler ve bunların güçlü/sınırlı yönleri ele alınmıştır.

5.1. İnce Ayar (Fine-Tuning) ve Talimat Uyumlama

Bir strateji, modeli yalnızca geniş ön-eğitimle bırakmak yerine, özel temizlenmiş veri veya talimat kümeleriyle yeniden eğitmektir. Bu sayede model, hem genel dil kalıplarını hem de güvenilir, doğrulanmış bilgi örneklerini öğrenir. Yakın zamanda yayınlanan bir derlemede ince ayar ve dış bilgi ile destekleme gibi pek çok yöntem ele alınmış; bunların bazı kombinasyonlarının halüsinasyon azaltımında olumlu etkileri olduğu belirtilmiştir. Bu tekniklerin birlikte uygulanmasının, modele hem daha doğru içerik üretme hem de uydurma yapmama alışkanlığı kazandırdığı görülmüştür [24].

5.2. Belge Tabanlı Bilgi Bağlama: RAG & Hibrit Çıkarım Yaklaşımları

Özellikle statik bilgiyle sınırlı kalan LLM'lerde, dış kaynaklara bağlanarak gerçek zamanlı ve güncel bilgiyle üretim yapmak önemli bir azaltma yöntemidir. Bu kapsamda Bilgi getirmeli üretim (Retrieval-Augmented

Generation-RAG) yaygın biçimde kullanılır: model, çıktı üretmeden önce ilgili belgeleri ya da veri tabanlarını arar ve ürettiği yanıtları bu kaynaklara dayandırır. Bu yaklaşım, halüsinasyon oranını önemli ölçüde düşürür.

Buna ek olarak, yalnızca tek bir çıkarım yöntemi değil anahtar kelime araması, gömülü vektör temelli arama ve bazen bu iki yöntemin kombinasyonu kullanılarak hibrit çıkarım sistemleri geliştirilmektedir. Bu hibrit sistemler, hem alandaki bilgi eksiklerini kapatır hem de daha yüksek doğruluk sağlar[25].

5.3. Üretim Sonrası Düzeltme & Çok Katmanlı Kontrol Çerçeveleri

RAG ve ince ayar gibi önleyici yöntemler önemli olsa da özellikle kritik ve geri dönüşü zor işleri hedefleyen uygulamalarda tek başlarına yeterli olmayabilir. Bu nedenle, birçok güncel çalışma, çok katmanlı bir azaltma mimarisi öneriyor: üretim öncesi kontroller + RAG + üretim sonrası doğrulama + insan onayı gibi birkaç savunma hattı birlikte. Böylece, hem modelin kendi sınırları hem de dış veri kaynaklarının zayıflıkları dengeleniyor[26].

Bu alanda yapılan başka bir çalışma, RAG'in özel ince ayar ve düzeltme döngüsüyle birlikte kullanıldığında halüsinasyon oranını azalttığını göstermiştir[27].

5.4. Çok Modüllü / Çok Ajanlı Sistemler & Alternatif Model Denetimi

Bazı çalışmalar, tek bir LLM'e güvenmek yerine, birden fazla model veya modül kullanan sistemlere yöneliyor. Örneğin bir modül bilgiyi kaynaktan çıkarır, diğer modül metni üretir, üçüncü modül doğrulama yapar. Ya da LLM yanıtı üretir, ardından doğruluk kontrolleri tarafından incelenir. Böyle çok ajanlı veya modüllü mimariler, sadece üretim değil doğrulama aşamasında da güvenlik sağlar[28].

6. Gelecek Araştırma Yönelimleri ve Sonuç

Büyük Dil Modellerinde (LLM) halüsinasyon problemi, mevcut modellerin teknik sınırlarının, eğitim paradigmasının ve bilgi temsili biçimlerinin doğal bir sonucu olarak ortaya çıkmaktadır. Bu nedenle, halüsinasyonu azaltmak ya da güvenli kullanım sınırları içinde yönetebilmek için hem mimari düzeyde yeniliklere hem de değerlendirme–kontrol mekanizmalarının güçlü bir biçimde geliştirilmesine ihtiyaç duyulmaktadır. Bu birleşik bölümde, geleceğe dönük temel araştırma eğilimleri ile çalışmanın genel sonuçları bir arada ele alınmaktadır.

6.1. Gelecek Araştırma Yönelimleri

6.1.1. Yeni Mimari Arayışları ve Olasılıksal Modellerin Sınırları

Mevcut LLM'ler, olasılık tahmini üzerine kurulu oldukları için, üretim sürecinin doğasında belirsizlik taşırlar. Bu mimari, her ne kadar esnek ve güçlü görünse de, gerçeği modelleme ile olası örüntüyü tahmin etme arasındaki farkı ortadan kaldıramamaktadır. Bu nedenle, gelecekte deterministik doğrulama katmanları, bilgiye bağlı üretim modülleri, veya karma mimariler daha fazla önem kazanacaktır. Özellikle karmaşık akıl yürütme görevlerinde sembolik bileşenlerin entegrasyonu, halüsinasyon oranlarını sistematik biçimde azaltma potansiyeli taşımaktadır.

6.1.2. Dış Referans Kullanımının Standartlaştırılması

RAG ve benzeri bilgi bağlama yöntemleri halüsinasyonla mücadelede etkili görünse de, bugün hâlâ parça parça çözümlerden ibarettir. Gelecek çalışmalarda RAG'ın daha gerçek zamanlı, düzenli güncellenen, görev-odaklı yapılarla bütünleşmesi; ayrıca dış bilgi kaynaklarının doğruluğunu değerlendiren ayrı denetim katmanlarının eklenmesi beklenmektedir. Uzun vadede amaç, modelin yalnızca tahmin değil, aynı zamanda kanıta dayalı üretim yapmasını sağlayacak ekosistemleri kurmaktır.

6.1.3.Model İçi Güven Sinyallerinin Standartlaştırılması

Halüsinasyonun en riskli yönlerinden biri, modelin yanlış üretimi yüksek güvenle sunmasıdır. Bu nedenle, modelin kendi içinde ürettiği güven sinyallerinin bilimsel olarak tanımlanması, kalibre edilmesi ve kullanıcıya açık biçimde sunulması gerekmektedir. Bu doğrultuda, mantıksal bazlı kararsızlık ölçütleri, dikkat dağılımlarının tutarlılığı veya modelin belirsizliğini ifade eden özel mekanizmalar gibi yaklaşımlar önümüzdeki dönemde araştırmaların merkezinde olacaktır. Kullanıcının model çıktısının hangi ölçüde güvenilir olduğunu görebilmesi, halüsinasyon riskini yönetmede kritik bir ilerleme sağlayacaktır.

6.1.4.Görev ve Alan Bazlı Değerlendirme Çerçeveleri

Halüsinasyon sorunu alanlar arasında farklı etkiler taşıdığı için, gelecekte tıp, hukuk, finans, eğitim, bilimsel yazım gibi kritik alanlara özel değerlendirme protokollerinin geliştirilmesi beklenmektedir. Bu tür alan-özgü değerlendirmeler, sadece doğruluk değil, etik sorumluluk, kanıt takibi, atıf güvenilirliği ve risk seviyelendirme gibi boyutları içermelidir. Böylece modellerin yalnızca doğruluk düzeyi değil, aynı zamanda güvenli kullanım uygunluğu da ölçülebilir hâle gelecektir.

6.1.4.İnsan-Merkezli Hibrit Sistemlere Yönelim

Araştırmalar, tamamen otonom LLM sistemlerinin uzun vadede güvence üretemeyeceğini; kritik uygulamalarda insan denetiminin bir bileşen olarak kalacağını göstermektedir. Bu nedenle, gelecekte insan–makine işbirliğine dayalı hibrit karar sistemlerinin yaygınlaşması beklenmektedir. LLM'lerin hızlı üretimi ile insan uzmanlığının doğrulayıcı rolünün bir araya getirilmesi, hem üretkenliği artıracak hem de hatalı çıktıların etkisini azaltacaktır.

6.2. Sonuç

Bu çalışmada, Büyük Dil Modellerinde halüsinasyon olgusu çok yönlü bir çerçevede ele alınmış; halüsinasyonun temel nedenleri, türleri, tespit yöntemleri ve azaltma stratejileri sistematik biçimde incelenmiştir. Analizler göstermektedir ki halüsinasyon, yalnızca verideki hatalardan kaynaklanan bir yan etki değildir; aksine, model mimarisinin olasılıksal doğası, sınırlı bağlam penceresi, bilgi sınırları ve akıl yürütmedeki tutarsızlık gibi temel yapısal özelliklerden doğan bir sorundur. Bu nedenle halüsinasyonu tamamen ortadan kaldırmak bugün için mümkün görünmemektedir; ancak doğru stratejilerle önemli ölçüde azaltılabilir.

Tespit yöntemleri açısından bakıldığında, insan denetimi hâlâ altın standart olmayı sürdürse de, kara kutu ve şeffaf kutu temelli otomatik tespit yöntemlerinin hızla geliştiği; karşılaştırma çalışmalarının ise bu ilerlemeyi ölçmek için önemli bir zemin sağladığı görülmektedir. Azaltma stratejileri ise tek bir yöntemle sınırlı değildir: ince ayar, bilgi bağlama yöntemleri, çok katmanlı doğrulama çerçeveleri ve çok ajanlı sistemler birlikte kullanıldığında daha başarılı sonuçlar vermektedir.

Gelecek araştırmaların odak noktası, modelleri daha kontrollü, daha şeffaf ve daha güvenilir hâle getirmek olacaktır. Bu doğrultuda yeni mimari arayışları, dış referanslı bilgi entegrasyonu, model-içi güven sinyallerinin geliştirilmesi, alana özgü değerlendirme protokollerinin oluşturulması ve insan–makine hibrit sistemlerinin kurulması kritik araştırma yönleri olarak öne çıkmaktadır.

Sonuç olarak, halüsinasyon sorunu LLM'lerin mevcut çalışma biçiminin doğal bir parçası olsa da, sistemli tespit mekanizmaları, çok katmanlı azaltma stratejileri ve disiplinler arası araştırma çabaları sayesinde daha güvenilir, daha sorumlu ve daha güvenli yapay zeka sistemleri oluşturmak mümkündür.

KAYNAKÇA

[1] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In Advances in Neural Information Processing Systems (NeurIPS). https://arxiv.org/abs/1706.03762

[2] Alansari, A., & Luqman, H. (2025). Large Language Models Hallucination: A Comprehensive Survey. arXiv Preprint, arXiv:2510.06265. https://arxiv.org/abs/2510.06265

[3]Zhang, J., Zhang, R., Kong, F., Miao, Z., Ye, Y., & Zheng, Y. (2025). Lost-in-the-Middle in Long-Text Generation: Synthetic Dataset, Evaluation Framework, and Mitigation. arXiv preprint arXiv:2503.06868.

[4]Orgad, H., Toker, M., Gekhman, Z., Reichart, R., Szpektor, I., Kotek, H., & Belinkov, Y. (2024). LLMs know more than they show: On the intrinsic representation of LLM hallucinations. arXiv Preprint, arXiv:2410.02707. https://arxiv.org/abs/2410.02707

[5]Saxena, V., Sathe, A., & Sandosh, S. (2025). Mitigating hallucinations in large language models: A comprehensive survey on detection and reduction strategies. In J. C. Bansal, P. K. Jamwal, & S. Hussain (Eds.), Sustainable Computing and Intelligent Systems (pp. 39–52). Springer. https://doi.org/10.1007/978-981-96-3311-1_4

[6]Xu, Z., Jain, S., & Kankanhalli, M. (2024). Hallucination is inevitable: An innate limitation of large language models. arXiv Preprint, arXiv:2401.11817. https://arxiv.org/abs/2401.11817

[7]Sui, P., Duede, E., Wu, S., & So, R. (2024, August). Confabulation: The surprising value of large language model hallucinations. In Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 14274-14284).

[8]Chiang, C.-C., & Lee, H.-Y. (2024). Understanding the effects of language-specific class imbalance in multilingual fine-tuning. arXiv Preprint, arXiv:2402.13016. https://arxiv.org/abs/2402.13016

[9]Zhou, Y., Zhang, M., Asai, A., & Roth, D. (2023). Sources of hallucination by large language models on inference tasks. In Findings of EMNLP 2023. https://aclanthology.org/2023.findings-emnlp.182/

[10]Ackermann, R., & Emanuilov, S. (2025). How large language models are designed to hallucinate. arXiv Preprint, arXiv:2509.16297. https://arxiv.org/abs/2509.16297

[11]Guerreiro, N. M., Alves, D. M., Waldendorf, J., Haddow, B., Birch, A., Colombo, P., & Martins, A. F. T. (2023). Hallucinations in large multilingual translation models. Transactions of the Association for Computational Linguistics, 11, 1500–1517. https://aclanthology.org/2023.tacl-1.85/

[12]Bai, Z., Wang, P., Xiao, T., He, T., Han, Z., Zhang, Z., & Shou, M. Z. (2024). Hallucination of multimodal large language models: A survey. arXiv Preprint, arXiv:2404.18930. https://arxiv.org/abs/2404.18930

[13]Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. ACM Computing Surveys, 55(12), 1–38. https://doi.org/10.1145/3571730

[14]Sato, M. (2025). Triggering hallucinations in LLMs: A quantitative study of prompt-induced hallucination in large language models. arXiv Preprint, arXiv:2505.00557. https://arxiv.org/abs/2505.00557

[15] Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B., & Liu, T. (2024). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. ACM Transactions on Information Systems, 1(1), Article 1. (arXiv:2311.05232).

[16] Zhang, Y., Li, Y., Cui, L., Cai, D., Liu, L., Fu, T., Huang, X., Zhao, E., Zhang, Y., Chen, Y., Wang, L., Luu, A. T., Bi, W., Shi, F., & Shi, S. (2025). Siren's song in the AI ocean: A survey on hallucination in large language models. arXiv Preprint, arXiv:2309.01219. https://arxiv.org/abs/2309.01219

[17]Zhang, M., Huang, M., Shi, R., Guo, L., Peng, C., Yan, P., Zhou, Y., & Qiu, X. (2024). Calibrating the confidence of large language models by eliciting fidelity. In Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP 2024) (pp. 2959–2979). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.emnlp-main.173

[18]Li, J., Cheng, X., Zhao, X., Nie, J.-Y., & Wen, J.-R. (2023). HaluEval: A Large-Scale Hallucination Evaluation Benchmark for Large Language Models. EMNLP 2023

[19] Manakul, P., Liusie, A., & Gales, M. (2023). SelfCheckGPT: Zero-Resource Black-Box Hallucination Detection for Generative Large Language Models. EMNLP 2023. https://aclanthology.org/2023.emnlp-main.557

[20] Bharti, S., Jain, S., & Carpuat, M. (2025). LLM-Check: Investigating Detection of Hallucinations in Large Language Models. University of Maryland. https://www.cs.umd.edu/sites/default/files/scholarly_papers/Spring_2025_Bharti%2C_Siddhant_Scholarly_Paper.pdf

[21] Su, W., Wang, C., Ai, Q., Hu, Y., Zhou, Y., & Liu, Y. (2024). Unsupervised Real-Time Hallucination Detection based on the Internal States of Large Language Models. Findings of ACL 2024. https://aclanthology.org/2024.findings-acl.854

[22]Li, J., Cheng, X., Zhao, X., Nie, J., & Wen, J.-R. (2023). HaluEval: A Large-Scale Hallucination Evaluation Benchmark for Large Language Models. EMNLP 2023. https://arxiv.org/abs/2305.11747

[23]HalluLens Benchmark (2025). HalluLens: LLM Hallucination Benchmark. arXiv:2504.17550. https://arxiv.org/abs/2504.17550

[24]Tonmoy, S. M. T. I., Zaman, S. M., Jain, V., Rani, A., Rawte, V., Chadha, A., & Das, A. (2024). A comprehensive survey of hallucination mitigation techniques in large language models. arXiv preprint arXiv:2401.01313, 6.

[25]Mala, C. S., Gezici, G., & Giannotti, F. (2025). Hybrid retrieval for hallucination mitigation in large language models: A comparative analysis (arXiv:2504.05324). arXiv. https://arxiv.org/abs/2504.05324

[26]Hiriyanna, S., & Zhao, W. (2025). Multi-layered framework for LLM hallucination mitigation in high-stakes applications: A tutorial. Computers, 14(8), 332. https://doi.org/10.3390/computers14080332

[27]Song, J., Wang, X., Zhu, J., Wu, Y., Cheng, X., Zhong, R., & Niu, C. (2024). RAG-HAT: A hallucination-aware tuning pipeline for LLM in retrieval-augmented generation. In Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track (pp. 1548–1558). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.emnlp-industry.113

[28]Darwish, A. M., Rashed, E. A., & Khoriba, G. (2025). Mitigating LLM hallucinations using a multi-agent framework. Information, 16(7), 517. https://doi.org/10.3390/info16070517

Bölüm 10

ENTERPOLASYON TEKNİKLERİNİN KENAR BELİRLEME BAŞARISINA ETKİSİ

Abdullah ORMAN [1]
Murat MERİÇELLİ [2]

1 Doç. Dr., Ankara Yıldırım Beyazıt Üniversitesi, Teknik Bilimler Meslek Yüksekokulu, Bilgisayar Teknolojileri Bölümü, ORCID: 0000-0002-3495-1897
2 Dr. Öğr. Üyesi, Alanya Alaaddin Keykubat Üniversitesi, ALTSO MYO, Bilgisayar Teknolojileri Bölümü, ORCID: 0000-0003-0168-3221

1. Giriş

Görüntü işleme alanında enterpolasyon, çözünürlük artırma, yeniden ölçekleme, iyileştirme ve çeşitli dönüşüm işlemlerinin temel bileşenlerinden biri olarak kritik bir rol oynamaktadır. Ancak enterpolasyon yöntemleri, yalnızca mevcut piksel değerlerini tahmin etmekle kalmaz; aynı zamanda görüntünün yapısal bütünlüğünü, özellikle de kenar bölgelerinin geometrik ve fotometrik özelliklerini doğrudan etkiler. Kenarlar, bir görüntünün en belirgin bilgi taşıyıcı unsurları olup nesne sınırlarının, doku geçişlerinin ve sahne yapısının anlaşılmasında merkezi öneme sahiptir. Bu nedenle enterpolasyon sırasında kenarların doğru biçimde korunması hem görüntü iyileştirme hem de yüksek seviyeli bilgisayarla görme uygulamaları açısından hayati bir gerekliliktir (Kılıçaslan, 2023). Kenar belirleme, görüntüdeki yapısal bilgiyi ortaya çıkararak sonraki tüm görüntü işleme adımlarının doğruluğunu doğrudan etkilediği için süreci hayati ölçüde öne taşıyan temel bir aşamadır. Canny, Roberts ve Prewitt gibi klasik kenar belirleme algoritmaları, farklı gürültü seviyelerinde dahi yapısal detayları ortaya çıkarabilen başarılı kenar bulma teknikleri arasında yer almakta olup, bu yöntemlerin performansını, iyileştirme yaklaşımlarını ve çeşitli görüntü işleme görevlerindeki etkilerini inceleyen çok sayıda çalışma literatürde bulunmaktadır. 2022 yılında Monica ve çağdaşları yenilikçi Otsu-Canny operatör yöntemini sunmuşlardır. Otsu algoritması, Canny operatörünü çift eşik değeriyle değiştirerek ve şebekelerin tespit üzerindeki etkisini ortadan kaldırarak mikro çatlaklar için kenar tespit performansını iyileştirmek amacıyla kullanılmıştır. İşlem, morfolojik işlem, görüntü segmentasyonu ve görüntü ikilileştirme ile gürültü giderme işlemleri gerçekleştirilerek tamamlanmıştır. Deneysel bulgular, bu araştırmada kullanılan görüntü işleme algoritmasının saflığı ve bütünlüğünde önemli bir iyileşme olduğunu göstermektedir. (Monicka et al., 2022) Tanyeri ve diğerleri Canny opreratorünün uygun üst ve alt eşiklerini, gradyan görüntüsünün ortalama ve yarı entropisi aracılığıyla otomatik olarak ayarlamışlardır. Böylece daha etkili bir kenar tespit yöntemleri sunmuşlardır (Tanyeri et al., 2019). Faheem ve diğerleri çalışmalarında etkili bir görüntü filigranlama için kenar tespiti yapmışlar ve Canny yaklaşımından yararlanmışlardır. Çalışmada en az anlamlı bit (LSB) ve Canny kenar algılama yöntemini kullanan bir dijital görüntü filigranlama tekniği önerilmiştir. Önerilen tekniğin, LSB'nin yüksek taşıma kapasitesi ve Canny kenar algılama filtresinden sonra filigran yerleştirme özelliği nedeniyle daha güvenli olduğunu ifade etmişlerdir. Canny kenar gradyanının yönü ve büyüklüğünden yararlanılarak, kaç bitin gömüleceğini belirlemişlerdir (Faheem et al., 2023). Diğer taraftan 2025 yılında önerilen bir diğer görüntü filigranlama için önerilşen teknikte, kenar tespitinin öneminden bahsedilmiş ve iğnecikli sinir ağından yararlanılan bir kenar tespiti ile filigran metodu önerilmiştir (İncetaş & Kılıçaslan, 2025). Medikal görüntüler üzerinde yapılan CNN çalışmalarında ise ön işlem adımı olarak kenar tespit algoritmaları kullanılmak-

tadır. 2022 yılında Dey ve diğerleri termal meme görüntülerini kabul eden ve bu görüntüleri tespit eden bilgisayar destekli bir meme kanseri tespit sistemi önermişlerdir. Çalışmalarında sınıflandırıcı oluşturmak üzere önceden eğitilmiş DenseNet121 modelini özellik çıkarıcı olarak kullanmışlardır. Özellikleri çıkarmadan önce, Prewitt ve Roberts olmak üzere iki kenar algılayıcı kullanarak çıktılar elde ederek, orijinal görüntüyle birlikte DenseNet121 modeline 3 kanallı bir görüntü olarak girdi sağlamışlardır (Dey et al., 2022). 2022 yılında Saeed Balochian ve Hossein Baloochian kenar tespitinin, dijital görüntülerde süreksizlikleri bulma ve sınırları keşfetmede en önemli adım olduğunu ifade etmişlerdir. Önerilen yaklaşımlarında Prewitt operatörüyle kesirli mertebeden türevleme kullanan yeni bir kenar tespiti yöntemi sunmuşlardır. İlgili araştırmacılar, komşu piksellerin bilgilerini ve ağırlıklı ortalamayı dolaylı olarak kullanarak yalnızca görüntünün türevini hesaplamakla kalmamışlar, aynı zamanda gürültüyü de ortadan kaldırılmışlardır. Deneysel bulgular önerilen yöntemin birçok kenar tespit yaklaşımına göre üstün olduğunu göstermiştir. İlaveten, Prewitt kesirli mertebeden kenar tespitinin performans değerlendirmesi, tıbbi görüntülerde kenar tespiti için umut verici potansiyellerini gösterdiğini ifade etmişlerdir.

Enterpolasyon, eksik veya düşük çözünürlüklü görüntü verilerinden daha yüksek çözünürlüklü, sürekliliği korunmuş ve yapısal açıdan tutarlı çıktılar elde etmek amacıyla kullanılan temel bir görüntü işleme yaklaşımıdır. Özellikle modern görüntüleme sistemlerinde sensör sınırlamaları, veri sıkıştırma süreçleri veya iletim kayıpları nedeniyle piksel bilgilerinin eksilmesi ya da çözünürlüğün düşmesi sık karşılaşılan bir durum olduğundan, enterpolasyon algoritmalarının doğruluğu ve kenar detaylarını koruma kapasitesi büyük önem taşımaktadır. En yakın komşu, bilineer ve bikübik gibi klasik yöntemlerin yanı sıra kenar korumalı ve adaptif yaklaşımların geliştirilmesi, görüntünün yerel yapısına daha duyarlı sonuçlar üretmeyi mümkün kılmıştır. Bu nedenle enterpolasyon, sadece görsel kaliteyi artıran bir araç değil, aynı zamanda birçok üst seviye görüntü işleme ve bilgisayarlı görü uygulamasının başarısını doğrudan etkileyen kritik bir adımdır. En yakın komşu, bilineer ve bikübik gibi klasik enterpolasyon yöntemleri, pikseller arası süreklilik varsayımına dayanarak çalıştığından, yüksek frekanslı bileşenlerde bulanıklaşma, kenar yayılması ve detay kaybı gibi istenmeyen artefaktlar üretebilir (Patel & Mistree, 2013). Son yıllarda ise anizotropik difüzyon filtreleme (İncetaş, 2022; Kılıçaslan, 2024) derin öğrenme tabanlı süper çözünürlük modelleri ve adaptif ağırlıklandırma yöntemleri gibi daha gelişmiş yaklaşımlar, kenar duyarlılığını artırarak yapısal bilgiyi korumayı amaçlamaktadır (Wang et al., 2022). Bununla birlikte enterpolasyonun kenar belirleme adımlarında nasıl bir etki yarattığı hem teorik hem de deneysel olarak hâlâ kapsamlı inceleme gerektiren bir araştırma problemidir. Çünkü gradyan tabanlı Canny, Roberts ve Prewitt gibi yöntemler enterpolasyon sonrası oluşan artefaktların doğasına

karşı oldukça hassastır. Bu bağlamda enterpolasyon tekniklerinin kenar belirleme başarısına etkisinin incelenmesi hem yeni algoritmaların tasarımı hem de mevcut sistemlerin performansının optimize edilmesi açısından önemli katkılar sağlamaktadır.

2. Enterpolasyon

Görüntü enterpolasyonu, sayısal görüntü işleme alanında çözünürlük artırma, ölçekleme, döndürme, sıkıştırma ve görüntü tamamlama gibi birçok temel görevin merkezinde yer alan kritik bir işlemdir. Temel olarak, var olan pikseller arasındaki yoğunluk veya renk değerlerini tahmin ederek yeni piksel değerleri üretme sürecini ifade eder. Bu işlem, özellikle tıbbi görüntüleme, uydu görüntüleme, nesne tanıma, video işleme ve artırılmış gerçeklik uygulamaları gibi yüksek doğruluk ve görsel bütünlük gerektiren alanlarda önemli bir rol oynamaktadır (Li & Orchard, 2001; Maeland, 1988). Enterpolasyonun başarısı hem görüntünün algısal kalitesini hem de sonraki görüntü işleme ya da makine öğrenimi adımlarının performansını doğrudan etkiler. En yaygın yöntemler olan Nearest Neighbor (en yakın komşu), Bilinear (bilineer) ve Bicubic (bikübik) enterpolasyon teknikleri (Parker et al., 2007) hesaplama maliyeti ve görsel kalite arasında farklı denge noktaları sunarken; kenar koruyucu, yön duyarlı veya derin öğrenme tabanlı modern yaklaşımlar daha yüksek doğrulukla detayların korunmasını hedeflemektedir. Ancak enterpolasyon işlemi, özellikle gürültülü veya düşük kaliteli görüntülerde kenar bulanıklığı, detay kaybı ve bozulma gibi istenmeyen etkilerle karşılaşılmasına neden olabilir. Bu nedenle, enterpolasyon algoritmalarının kenar bilgisi, doku sürekliliği ve yerel istatistikleri mümkün olduğunca doğru biçimde kullanması gerekmektedir. Güncel araştırmalar hem klasik yöntemlerin geliştirilmesine hem de optimizasyon algoritmaları ve yapay zekâ tabanlı modellerin entegrasyonuna odaklanarak, daha yüksek görsel kalite ve daha düşük hata oranı ile çalışan enterpolasyon teknikleri geliştirmeyi amaçlamaktadır. Enterpolasyon işleminin genel formülü;

$$I_{DÇ}(m,n) = I_{YÇ}(2m-1, 2n-1), m = 1,\cdots,M; n = 1,\cdots,N \quad (1)$$

şeklinde hesaplanmaktadır. Burada $I_{DÇ}$ düşük çözünürlüklü girdi görüntüyü, $I_{YÇ}$ ise yüksek çözünürlüklü çıktı görüntüyü temsil etmektedir. m ve n ise görüntülerin yükseklik ve genişliğini temsil etmektedir.

3. Kenar Belirleme

Görüntü işleme alanında kenar belirleme, bir görseldeki nesnelerin sınırlarını, yapısal geçiş bölgelerini ve yüzey süreksizliklerini ortaya çıkarmayı amaçlayan temel bir adımdır. Kenar noktaları, görüntüdeki yoğunluk dağılımının hızlı değişim gösterdiği bölgeler olduğundan, bu değişimin sayısal olarak modellenmesi genellikle türev tabanlı operatörler ile gerçekleştirilir. Kla-

sik yöntemler olan Sobel, Prewitt ve Roberts operatörleri, gradyan hesaplamasına dayalı olarak kenar bilgisini üretse de gürültü varlığında kararsız sonuçlar verebilmekte ve kenarların lokalizasyonunda hatalar oluşturabilmektedir. Bu nedenle kenar belirleme sürecinin hem doğruluğunu hem de kararlılığını aynı anda sağlayacak daha gelişmiş algoritmalara ihtiyaç duyulmuştur. Bu bağlamda Canny matematiksel olarak optimal bir kenar belirleyici geliştirmek amacıyla kapsamlı bir analiz sunmuş ve bugün literatürde en yaygın kullanılan kenar operatörü olan Canny Kenar Belirleme Algoritmasını önermiştir (Canny, 1986). Canny'nin optimalite yaklaşımı üç ana kriter üzerine kuruludur: (i) gerçek kenarların maksimum doğrulukla tespit edilmesi, (ii) tespit edilen kenarların gerçek konumlarına en yakın şekilde lokalize edilmesi ve (iii) aynı kenar üzerinde çoklu cevapların en aza indirilmesi. Bu kriterler, algoritmanın matematiksel olarak optimize edilmiş bir filtreleme ve karar verme yapısına sahip olmasını sağlamaktadır. Canny algoritması, çok aşamalı yapısı sayesinde hem gürültüye dayanıklı hem de ince kenarları koruyabilen bir mimariye sahiptir. İlk aşamada görüntü, yüksek frekanslı gürültü bileşenlerinin baskılanması için bir Gauss filtresi ile düzgünleştirilir. Gauss filtresi:

$$G(x,y) = \frac{1}{2\pi\sigma^2} exp(-\frac{x^2+y^2}{2\sigma^2}) \quad (2)$$

şeklinde tanımlanır ve filtrelenmiş görüntü:

$$Is(x,y) = I(x,y) * G(x,y) \quad (3)$$

Gürültünün bastırılması, Canny'nin temel kriterlerinden biri olan çoklu cevapların azaltılması için gereklidir. Düşük σ değerleri ince detayları korurken, yüksek σ değerleri kenarları yumuşatarak daha geniş bölgesel geçişleri ortaya çıkarır. Canny algoritmasının önemli aşamalarından bir tanesi de maksimum olmayan değerlerin bastırılması yöntemidir. Burada kenarların inceltilmesi amaçlanır ve sonrasında eşikleme yardımıyla kenar pikseller tespit edilir (Gebäck & Koumoutsakos, 2009).

Sonuç olarak Canny kenar belirleme operatörü, Gauss tabanlı gürültü giderme, birinci türev tabanlı gradyan analizi, non-maximum suppression, çift eşikleme ve histerezis bağlantılama gibi ardışık aşamalardan oluşan matematiksel olarak optimize edilmiş bir yapıya sahiptir. Bu çok aşamalı yapı hem ince detayların korunmasını hem de gürültü kaynaklı yanlış kenar tespitlerinin azaltılmasını sağlamaktadır. Özellikle tıp görüntüleme, uzaktan algılama, biyometri, robotik ve yapay görme uygulamalarında yüksek başarımı sayesinde standart yöntem haline gelmiştir.

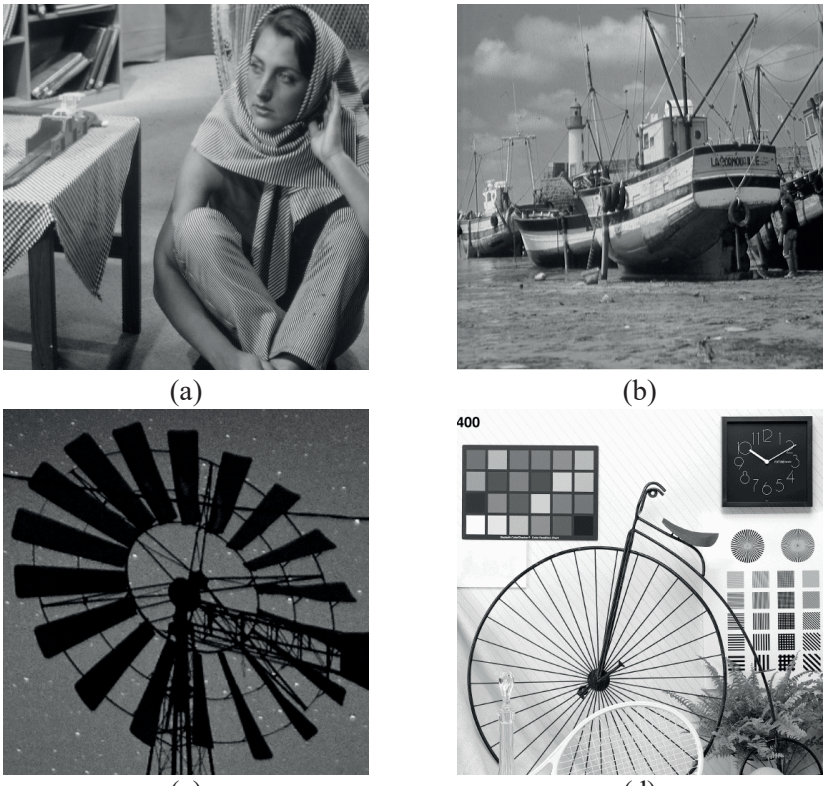
4. Deneysel Bulgular ve Tartışma

1.1 Veri Seti

Bu çalışmada tüm deneyler, görüntü işleme ve özellikle enterpolasyon çalışmalarında sık kullanılan 8 adet gri seviyeli görüntü üzerinde gerçekleştirilmiştir. Bu 8 görüntüden 4 tanesi 256x256 boyutunda iken, diğer 4 tanesi 512x512 boyutundadır. Şekil 1 ve 2'de sırasıyla 256x256 ve 512x512 boyutlarındaki görüntüler verilmiştir.

(a) (b) (c) (d)

Şekil 1. 256x256 boyutundaki görüntüler a) airplane, b)cameraman, c) house, d) peppers



(a) (b)

(c) (d)

Şekil 2. 512x512 boyutundaki görüntüler a) barbara, b)boat, c) stars, d) Wheel

4.2. Metrikler

Sınıflandırma, nesne tanıma, kenar belirleme, tespit ve segmentasyon gibi birçok yapay zeka ve görüntü işleme uygulamasında performans ölçümü, doğru kararların oranını ortaya koyan temel istatistiksel metriklerle yapılır. Bu metrikler arasında Precision (Kesinlik), Recall (Duyarlılık) ve F-Skor en yaygın kullanılan ölçüm değerleridir. Bu ölçütler, bir modelin pozitif sınıfı ne kadar doğru tahmin ettiğini, kaç gerçek pozitif örneği yakalayabildiğini ve genel başarımını ortaya koyan tamamlayıcı metriklerdir (Rani et al., 2022; Tariq et al., 2021).

Bu metriklerin hesaplanmasında temel alınan kavramlar şunlardır:

- TP (True Positive): Gerçek pozitiflerin doğru tahmin edilmesi
- FP (False Positive): Negatif örneklerin yanlışlıkla pozitif tahmin edilmesi
- FN (False Negative): Pozitif örneklerin model tarafından kaçırılması
- TN (True Negative): Gerçek negatiflerin doğru tahmin edilmesi

Precision, modelin pozitif tahminlerinin ne kadarının doğru olduğunu ölçer. Başka bir ifadeyle, model bir örneği "pozitif" olarak işaretlediğinde bu kararın ne kadar güvenilir olduğunu ifade eder. Precision'ın matematiksel olarak (Huč et al., 2021);

$$Precision = \frac{TP}{FP + TP} \tag{4}$$

Olarak hesaplanır. Recall (Huč et al., 2021), gerçek pozitif örneklerin ne kadarının model tarafından doğru şekilde yakalandığını ölçer ve

$$\mathrm{Re}\,call = \frac{TP}{FN + TP} \tag{5}$$

F-Skor, Precision ve Recall'ın harmonik ortalaması alınarak elde edilen bir performans ölçüsüdür. Harmonik ortalama, değerlerden en küçüğüne daha fazla ağırlık verdiği için modelin hem Precision hem de Recall açısından dengeli olmasını zorunlu kılar. Standardize edilmiş en yaygın F-skor türü olan F1-Skor şu şekilde tanımlanır (Huč et al., 2021):

$$F1 = 2\,\frac{Precision + Recall}{Precision \cdot Recall} \tag{6}$$

Precision, Recall ve F-Skor, makine öğrenmesi ve görüntü işleme uygulamalarının performans değerlendirmesinde kritik öneme sahip tamamlayıcı metriklerdir (Naidu et al., 2023). Precision modelin pozitif tahminlerinin doğruluğunu, Recall yakalama gücünü, F-Skor ise iki ölçütün dengeli bir kombinasyonunu sunar. Bu metrikler özellikle dengesiz veri setleri, kritik karar süreçleri ve pozitif sınıfın önemli olduğu uygulamalarda modelin gerçek başarımını ortaya koymada vazgeçilmez araçlardır.

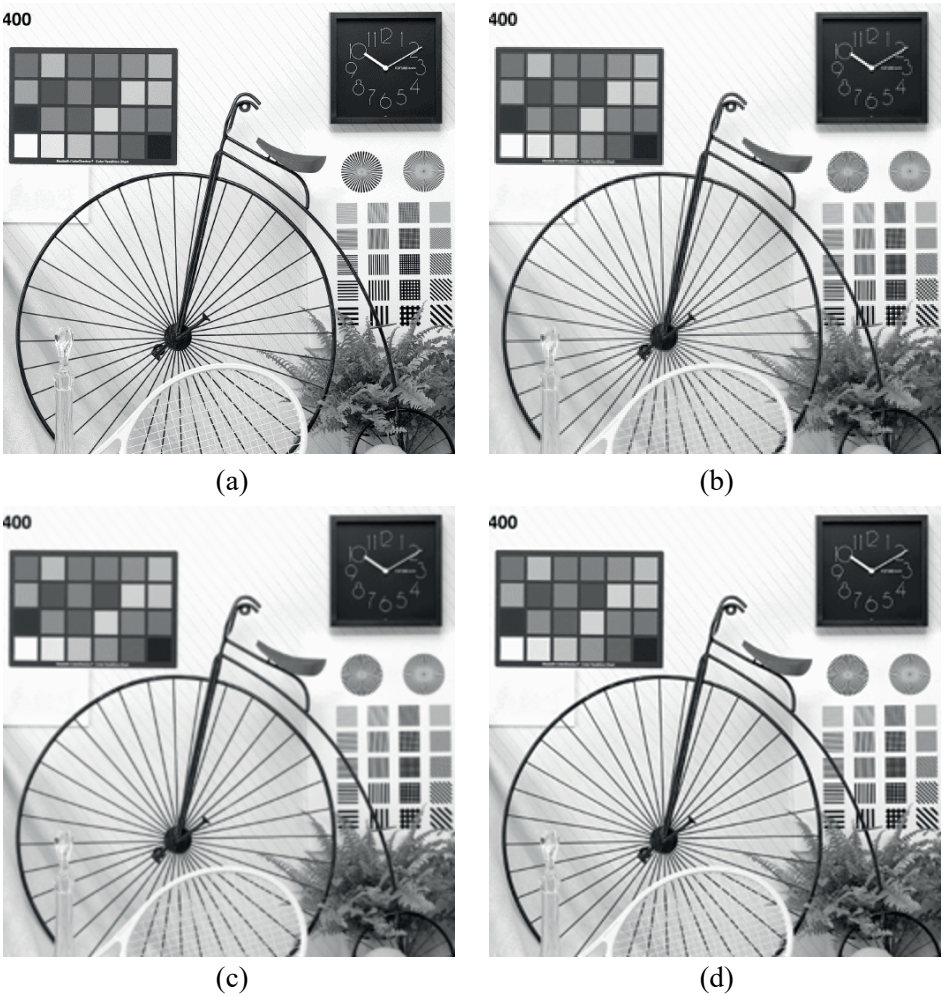
4.3. Deneysel Sonuçlar

Çalışmada ilk olarak Şekil 1 ve 2'de yer alan 8 görüntü ½ oranında küçültülmüş ve hemen ardından en yakın komşu (nearest), bilineer (bilinear) ve bikübik (bicubic) enterpolasyon yaklaşımları ile büyütülmüştür. Tüm bu işlemler Matlab aracılığıyla gerçekleştirilmiştir. Büyütülen görüntülere ilişkin örnekler Şekil 3 ve 4'te yer almaktadır.

(a) (b) (c) (d)

Şekil 3. Kameraman görüntüsü için enterpolasyon sonuçları. a) Orijinal görüntü, b) en yakın komşu enterpolasyon, b) bilinear enterpolasyon, c) bicubik enterpolasyon

Şekil 3'te kameraman görüntüsüne ait büyütme sonuçları yer almaktadır. Şekil 3.(a)'da orijinal kameraman görüntüsü yer alırken, Şekil 3.(b), (c) ve (d)'de sırasıyla en yakın komşu, bilneer ve bikübik enterpolasyon yaklaşımlarının sonuçları yer almaktadır. Şekil 4'te ise 512x512 boyutundaki Wheel görüntüsünün enterpolasyon sonuçları verilmiştir. Şekil 4.(a)'da orijinal wheel görüntüsü ile Şekil X.(b), (c) ve (d)'de sırasıyla en yakın komşu, bilneer ve bikübik enterpolasyon yaklaşımlarının sonuçları görülmektedir.

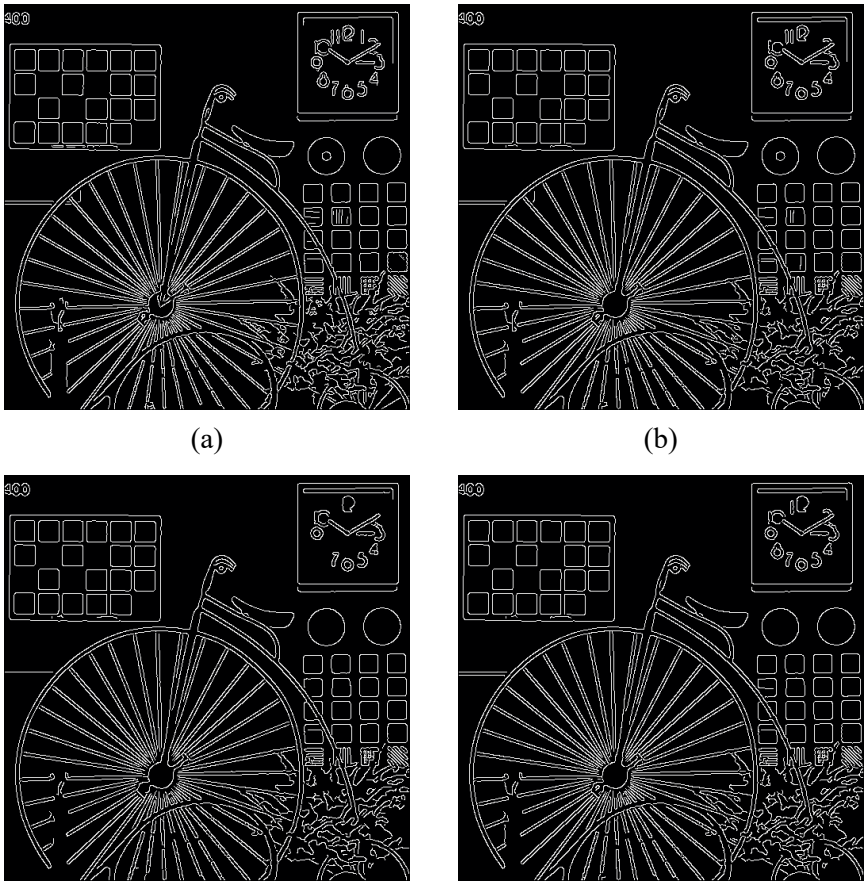


Şekil 4. Wheel görüntüsü için enterpolasyon sonuçları. a) Orijinal görüntü, b) en yakın komşu enterpolasyon, b) bilinear enterpolasyon, c) bicubik enterpolasyon

Şekil 3 ve 4'te yer alan enterpolasyon sonuçları orijinal görüntüye oldukça yakındır ve gözle ayırt edilemeyecek kadar büyük bir benzerliğe sahiptir. Ancak kenar belirleme başarılarının tespiti için orijinal görüntüler de dahil olmak üzere tüm enterpolasyon sonuçlarının canny kenar dedektörü ile kenarları tespit edilmiştir. Elde edilen kenar görüntüleri için örnekler Şekil 5 ve 6'da verilmiştir.

(a) (b) (c) (d)

Şekil 5. Kameraman görüntüsü için kenar sonuçları. a) Orijinal görüntü, b) en yakın komşu enterpolasyon, b) bilinear enterpolasyon, c) bicubik enterpolasyon.



(a) (b)

(C) (d)

Şekil 6. Wheel görüntüsü için kenar sonuçları. a) Orijinal görüntü, b) en yakın komşu enterpolasyon, b) bilinear enterpolasyon, c) bicubik enterpolasyon.

Şekil 5 ve 6'daki enterpolasyon işlemi sonrası elde edilen kenar görüntüleri, orijinal görüntülerden elde edilen kenar görüntüleri ile karşılaştırıldıklarında oldukça benzer oldukları görülmektedir. Her ne kadar görsel farklar az olsa da bu çalışmada tüm görüntülerden elde edilen kenar bilgileri F-skor metriği yardımıyla karşılaştırılarak, enterpolasyon tekniklerinin kenar bilgilerini ne kadar koruyabildiği nicel olarak ölçülmüştür. Bu amaçla öncelikle hem orijinal görüntülerin hem de 3 farklı enterpolasyon yöntemi ile büyütülen görüntülerin canny kenar dedektörü yardımıyla kenarları ortaya çıkarılmıştır. Ardından orijinal görüntüdeki kenarlar doğru kabul edilerek, enterpolasyon ile büyütülen görüntülerin kenarları ile karşılaştırılmıştır. Yapılan karşılaştırmalarda öncelikle hem orijinal hem enterpolasyon sonuç görüntüsünde aynı olan kenarlar, ardından orijinal görüntüde olup enterpolasyon görüntüsünde olmayan kenarlar ve yine orijinal görüntüde olmayıp enterpolasyon görüntüsünde olan kenarlar belirlenmiştir. Her bir enterpolasyon sonuç görüntüsü için belirlenen bu kenarlar yardımıyla Precision, Recall ve F-skor değerleri hesaplanmıştır. Elde edilen sonuçlar Tablo 1'de yer almaktadır.

Tablo 1

Canny kenar operatörü sonuçları

	En Yakın Komşu			Bilineer			Bikübik		
Görüntü	Precision	Recall	F-skor	Precision	Recall	F-skor	Precision	Recall	F-skor
airplane	0.8459	0.8234	0.8345	0.7955	0.7400	0.7667	0.8308	0.7992	0.8147
barbara	0.8474	0.7414	0.7909	0.8161	0.6525	0.7252	0.8474	0.7209	0.7790
boat	0.8316	0.8086	0.8200	0.7662	0.7095	0.7367	0.8133	0.7877	0.8003
cameraman	0.8594	0.7996	0.8284	0.8007	0.6900	0.7413	0.8343	0.7693	0.8005
house	0.8819	0.7955	0.8365	0.8187	0.6602	0.7310	0.8483	0.7813	0.8134
peppers	0.8607	0.8400	0.8502	0.8046	0.7576	0.7804	0.8535	0.8290	0.8411
stars	0.9167	0.8908	0.9035	0.8704	0.7784	0.8218	0.9055	0.8798	0.8924
wheel	0.8714	0.8380	0.8544	0.8062	0.7406	0.7720	0.8542	0.8077	0.8303
Ortalama	0.8644	0.8172	0.8398	0.8098	0.7161	0.7594	0.8484	0.7968	0.8215

Tablo 1'de görüldüğü gibi kenarları korumada en başarılı sonuçlar En Yakın Komşu enterpoasyon yaklaşımı ile elde edilmiştir. Başarı sıralamasında ikinci teknik Bikübik enterpolasyon yaklaşımıdır. Bilineer enterpolasyon yaklaşımı ise son sırada kalmıştır. Bu çalışmada yapılan deneylerin her ne kadar kenar koruma başarısı konusunda önemli ipuçları vermesine rağmen, litera-

türde yer alan pek çok farklı yaklaşım bulunmaktadır. Özellikle kenar koruma başarısı vurgulanan ve makine öğrenmesine ya da nöromorfik tekniklere dayalı teknikler son yıllarda ön plana çıkmıştır. Gelecekte bu yaklaşımların da dahil edildiği ve çok daha büyük veri setleri üzerinde yapılacak karşılaştırmalar ile enterpolasyon tekniklerinin kenarları koruma başarısının incelenmesi planlanmaktadır. Ayrıca enterpolasyon işlemi gerçekleştirildikten sonra elde edilen görüntüler, kenar belirleme dışında bölütleme (İncetaş & Meriçelli, 2024), renk indirgeme (Kılıçaslan & İncetaş, 2023) ya da görüntü erişimi (İncetaş, 2023; İncetas & Arslan, 2025) gibi süreçlerde de kullanılmaktadır. Dolayısıyla farklı enterpolasyon tekniklerinin bölütleme ya da görüntü erişimi süreçlerindeki başarılarının incelenmesi de planlanan çalışmalar arasındadır.

5. Sonuç

Enterpolasyon bir görüntü işlemi alanı olarak çözünürlük artırma ve yeniden ölçekleme işlemlerini kapsamaktadır. Görüntülerden elde edilen kenar bilgileri önemli bilgi taşıyıcı unsurlardır. Özellikle nesne sınırlarının, doku geçişlerinin ve sahne yapısının belirlenmesinde büyük öneme sahiptir. Enterpolasyon ile görüntülerin büyütülmesi sonucunda elde edilen kenar bilgisinde kayıpların tespit edilmesi, hangi enterpolasyon tekniğinin kullanılması gerektiği konusunda kullanıcılara yardımcı olabilecektir. Bu çalışmada, enterpolasyon yaklaşımlarının kenar belirleme başarısına etkisi araştırılmıştır. Elde edilen sonuçlara göre En Yakım Komşu yaklaşımının kenarları diğer iki yaklaşıma göre daha iyi koruduğunu göstermektedir. İkinci sırada bikübik enterpolasyon yaklaşımı yer alırken, bilineer enterpolasyon yaklaşımı son sırada kalmıştır. Ancak literatürde çok sayıda farklı enterpolasyon yaklaşımları yer almaktadır. Bu nedenle daha somut yargılar verebilmek amacıyla gelecekte farklı enterpolasyon yaklaşımlarının detaylı olarak karşılaştırılacağı çalışmalar planlanmaktadır.

KAYNAKLAR

Canny, J. (1986). Collision detection for moving polyhedra. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2, 200–209.

Dey, S., Roychoudhury, R., Malakar, S., & Sarkar, R. (2022). Screening of breast cancer from thermogram images by edge detection aided deep transfer learning model. Multimedia Tools and Applications, 81(7), 9331–9349.

Faheem, Z. Bin, Ishaq, A., Rustam, F., de la Torre Díez, I., Gavilanes, D., Vergara, M. M., & Ashraf, I. (2023). Image watermarking using least significant bit and canny edge detection. Sensors, 23(3), 1210.

Gebäck, T., & Koumoutsakos, P. (2009). Edge detection in microscopy images using curvelets. BMC Bioinformatics, 10(1), 75.

Huč, A., Šalej, J., & Trebar, M. (2021). Analysis of machine learning algorithms for anomaly detection on edge devices. Sensors, 21(14), 4946.

İncetaş, M. O. (2022). Anisotropic diffusion filter based on spiking neural network model. Arabian Journal for Science and Engineering, 47(8), 9849–9860.

İncetaş, M. O. (2023). Content-Based Image Retrieval Based on Block Truncation Code Using Fuzzy-C-Means Quantization and Edge Detection. In S. Bardak & Ü. Ayata (Eds.), Current Research in Engineering. Gece Kitaplığı.

İncetas, M. O., & Arslan, R. U. (2025). Spiking neural network-based edge detection model for content-based image retrieval. Signal, Image and Video Processing, 19(2), 169.

İncetaş, M. O., & Kılıçaslan, M. (2025). Image watermarking based on spiking neural networks. Cluster Computing, 28(11), 736.

İncetaş, M. O., & Meriçelli, M. (2024). Renk İndirgeme ve Kenar Belirleme Yardımıyla SRG Tabanlı Görüntü Bölütleme MO İncetaş, M Meriçelli Mühendislik Uygulamalarında Yenilikçi ve Multidisipliner Çalışmalar. In A. Orman (Ed.), Mühendislik Uygulamalarında Yenilikçi ve Multidisipliner Çalışmalar (pp. 75–84). Özgür Yayınları.

Kılıçaslan, M. (2023). Adaptive Threshold Selection of Anisotropic Diffusion Filters Using Dissimilarity Transform. International Journal of Engineering Research and Development, 15(2), 558–573.

Kılıçaslan, M. (2024). Adaptive threshold selection of anisotropic diffusion filters using spiking neural network model. Signal, Image and Video Processing, 18(1), 407–416.

Kılıçaslan, M., & İncetaş, M. O. (2023). Adaptive color quantization method with multi-level thresholding. International Journal of Computational Intelligence Systems, 16(1), 7.

Li, X., & Orchard, M. T. (2001). New edge-directed interpolation. IEEE Transactions on Image Processing, 10(10), 1521–1527.

Maeland, E. (1988). On the comparison of interpolation methods. IEEE Transactions on Medical Imaging, 7(3), 213–217.

Monicka, S. G., Manimegalai, D., & Karthikeyan, M. (2022). Detection of microcracks in silicon solar cells using Otsu-Canny edge detection algorithm. Renewable Energy Focus, 43, 183–190.

Naidu, G., Zuva, T., & Sibanda, E. M. (2023). A review of evaluation metrics in machine learning algorithms. Computer Science On-Line Conference, 15–25.

Parker, J. A., Kenyon, R. V, & Troxel, D. E. (2007). Comparison of interpolating methods for image resampling. IEEE Transactions on Medical Imaging, 2(1), 31–39.

Patel, V., & Mistree, K. (2013). A review on different image interpolation techniques for image enhancement. International Journal of Emerging Technology and Advanced Engineering, 3(12), 129–133.

Rani, S., Ghai, D., & Kumar, S. (2022). Object detection and recognition using contour based edge detection and fast R-CNN. Multimedia Tools and Applications, 81(29), 42183–42207.

Tanyeri, U., Kılıçaslan, M., & Demirci, R. (2019). Canny edge detector with half entropy. 2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), 1–4.

Tariq, N., Hamzah, R. A., Ng, T. F., Wang, S. L., & Ibrahim, H. (2021). Quality assessment methods to evaluate the performance of edge detection algorithms for digital image: A systematic literature review. IEEE Access, 9, 87763–87776.

Wang, X., Yi, J., Guo, J., Song, Y., Lyu, J., Xu, J., Yan, W., Zhao, J., Cai, Q., & Min, H. (2022). A review of image super-resolution approaches based on deep learning and applications in remote sensing. Remote Sensing, 14(21), 5423.

Bölüm 11

SAHTE İŞ İLANI TESPİTİ İÇİN MERKEZİ VE FEDERE ÖĞRENME YAKLAŞIMLARININ KARŞILAŞTIRMALI ANALİZİ

Maîmouna Dieynaba GASSAMA [1]
Emrullah SONUÇ [2]

1 Lisans Öğrencisi Maîmouna Dieynaba GASSAMA, Karabük Üniversitesi, Bilgisayar Mühendisliği Bölümü, Karabük, Türkiye, maynadg123@gmail.com, 0009-0005-9087-5912
2 Doç. Dr. Emrullah SONUÇ, Karabük Üniversitesi, Bilgisayar Mühendisliği Bölümü, Karabük, Türkiye, esonuc@karabuk.edu.tr, 0000-0001-7425-6963

1. GİRİŞ

Son yıllarda çevrimiçi iş platformlarının yaygınlaşmasıyla birlikte sahte iş ilanları önemli bir siber suç sorunu haline gelmiştir. Bu tür ilanlar, iş arayanlardan para talep etme, kişisel bilgi çalma veya kimlik avı gibi yöntemlerle maddi ve manevi zararlara yol açmaktadır (Alghamdi & Alharby, 2019). Geleneksel merkezi öğrenme yaklaşımlarında, iş ilanı verileri merkezi bir sunucuda toplanarak modeller eğitilmekte ancak bu yöntem veri gizliliği risklerini artırırken, büyük veri hacimleri nedeniyle iletişim ve hesaplama maliyetlerini yükseltmektedir (McMahan et al., 2017). Buna karşılık, federe öğrenme, verilerin yerel cihazlarda kalmasını sağlayarak gizliliği korurken, yalnızca model güncellemelerinin paylaşılmasıyla dağıtık eğitim gerçekleştirmektedir. Bu sayede özellikle hassas verilerin işlendiği senaryolarda avantaj sağlamaktadır (Konečný et al., 2016).

Sahte iş ilanı tespiti alanında önceki çalışmalar ağırlıklı olarak merkezi öğrenme tabanlı yöntemler üzerine odaklanmıştır. Örneğin, Doğal Dil İşleme (DDİ) teknikleriyle metin özelliklerini çıkaran modeller, Rastgele Orman, Destek Vektör Makineleri ve Naive Bayes gibi sınıflandırıcılarla yüksek doğruluklar elde etmiştir (Amaar et al., 2022; Vidros et al., 2017). EMSCAD (Employment Scam Aegean Dataset) gibi veri setleri kullanılarak yapılan deneylerde, topluluk (ensemble) yöntemlerin tekil sınıflandırıcılara üstünlük gösterdiği belirtilmiştir (Dutta & Bandyopadhyay, 2020). Ancak bu yaklaşımlar, veri merkezileştirmesi nedeniyle Kişisel Verileri Koruma Kanunu (KVKK) gibi gizlilik yönetmeliklerine uyum zorluğu yaşamakta ve dağıtık iş platformlarında ölçeklenebilirlik sorunu yaratmaktadır.

Bu çalışma, sahte iş ilanı tespitini hem merkezi hem de federe öğrenme çerçevelerinde ele alarak alana bir katkı sunmaktadır. Merkezi öğrenmede tüm veri merkezi sunucuda işlenirken, federe öğrenmede yerel modeller (örneğin farklı iş platformlarından gelen verilerle) eğitilip bir araya getirilmiştir. Performans metrikleri (doğruluk, F1-skor) ve iletişim maliyeti (model güncelleme boyutu) açısından karşılaştırma, federe öğrenmenin gizlilik avantajını korurken performans kaybının kabul edilebilir seviyede olduğunu göstermektedir. Bu yaklaşım, özellikle çoklu platformlarda veri paylaşımının kısıtlı olduğu gerçek dünya senaryolarında pratik uygulanabilirlik sağlamaktadır.

2. Veri Seti Keşfi ve Özellik Analizi

Çalışmada kullanılan veri seti, University of the Aegean tarafından toplanan ve Employment Scam Aegean Dataset (EMSCAD) olarak bilinen, Kaggle platformunda "Real / Fake Job Posting Prediction" adıyla erişilebilir olan bir veri setidir. Bu veri seti, 2012-2014 yılları arasında çevrimiçi iş platformlarından elde edilen yaklaşık 17880 iş ilanını içermekte olup, 866'sı sahte (fraudulent) ve kalanlar gerçek olarak etiketlenmiştir (Vidros et al., 2017). Veri seti, metinsel (*company_profile*, *description, requirements* vb.) ve kategorik

(*has_company_logo*, *required_experience*, *required_education* vb.) özellikler bakımından zengindir ve sahte iş ilanı tespiti araştırmalarında sıkça kullanılmaktadır (Alghamdi & Alharby, 2019; Amaar et al., 2022). Veri setinin keşfi, Python'da Pandas kütüphanesi kullanılarak gerçekleştirilmiş olup *shape()*, *head()* ve *describe()* fonksiyonları ile temel istatistikler incelenmiştir. Bu işlemler, iş ilanlarının yapısını anlamak ve modelleme öncesi ön işleme adımlarını belirlemek için kritik öneme sahiptir.

2.1. Eksik Değerlerin İncelenmesi ve Gözlemler

Veri setindeki eksik değer analizi, sahte ve gerçek ilanlar arasındaki belirgin farklılıkları ortaya koymuştur. Sahte ilanlar (fraudulent=1), gerçek ilanlara (fraudulent=0) kıyasla daha yüksek oranda eksik bilgi içermekte olup, bu durum dolandırıcıların detaylı bilgi sağlamaktan kaçındığını işaret etmektedir (Vidros et al., 2017). Özellikle *company_profile*, *has_company_logo* ve *has_questions* gibi özellikler, sahte ilanların ayırt edici işaretleri olarak öne çıkmaktadır.

Bu analizde, sahte iş ilanlarının gerçek ilanlara kıyasla belirgin ölçüde daha fazla eksik bilgi içerdiği ortaya çıkmıştır. Özellikle aşağıdaki sütunlardaki eksik olmayan (non-missing) değerlerin oranları bu durumu net bir şekilde göstermiştir:

- **Company Profile**: Gerçek ilanların %84,01'inde şirket profili bilgisi bulunurken, sahte ilanlarda bu oran sadece %32,22'dir. Bu bulgu, sahte ilanların şirket kimliğini gizleme eğiliminde olduğunu doğrulamakta ve önceki çalışmalarda da benzer şekilde rapor edilmiştir (Amaar et al., 2022).

- **Has Company Logo**: Gerçek ilanlarda şirket logosu bulunma oranı %81,91 iken, sahte ilanlarda bu oran %32,68'e düşmektedir. Logo eksikliği, ilanların profesyonelliğini ve güvenilirliğini azaltan önemli bir göstergedir (Vidros et al., 2017).

- **Has Questions**: İlanlarda soru bulunma oranı gerçek ilanlarda %50,21 iken, sahte ilanlarda bu oran %28,87'de kalmıştır.

Bu bulgular, sahte iş ilanı düzenleyenlerin genellikle şirket ve pozisyon hakkındaki detaylı bilgileri sağlamaktan kaçındığını veya bu bilgileri eksik bıraktığını göstermektedir. Benzer desenler, EMSCAD veri seti üzerine yapılan diğer analizlerde de gözlemlenmiş olup, eksik değerlerin dolandırıcılık tespitinde güçlü öngörücüler olduğu vurgulanmıştır (Dutta & Bandyopadhyay, 2020).

2.2. Kategorik Özniteliklerin Dağılımı

Kategorik özelliklerin frekans analizi, sahte ilanların adayları çekmek için daha erişilebilir ve düşük nitelikli pozisyonları hedeflediğini ortaya koymuştur. *Required_experience* ve *required_education* sütunları, bu eğilimi

açıkça sergilemektedir.

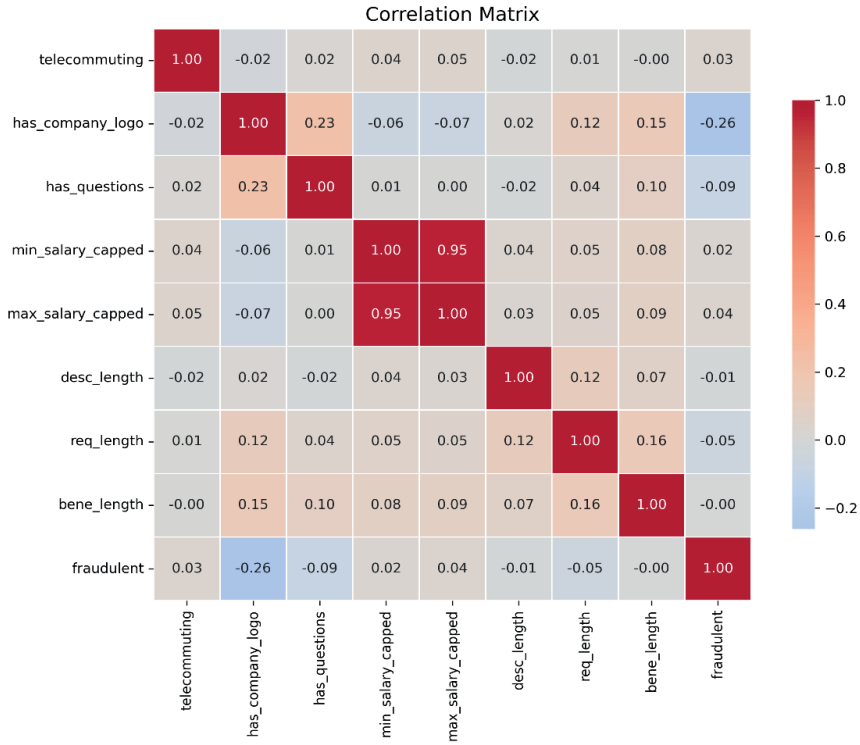
- **Gerekli Deneyim Düzeyi (Required Experience)**: Sahte ilanlarda en yaygın deneyim düzeyi Giriş Seviyesi (Entry level) olup, %41,53'lük bir orana sahiptir. Buna karşılık, gerçek ilanlarda en yaygın olan Orta-Kıdemli Seviye (Mid-Senior level) olup, %35,54 oranında görülmektedir. Bu dağılım, sahte ilanların deneyimsiz adayları (örneğin yeni mezunlar) hedefleyerek dolandırıcılık fırsatını artırdığını göstermektedir (Vidros et al., 2017).

- **Gerekli Eğitim Düzeyi (Required Education)**: Sahte ilanlarda en çok talep edilen eğitim düzeyi Lise veya eşdeğeri (High School or equivalent) (%40,96) iken, gerçek ilanların yarısından fazlası Lisans Derecesi (Bachelor's Degree) (%53,9) talep etmektedir. Düşük eğitim gereksinimi, geniş kitleye hitap ederek dolandırıcılık (scam) başarı oranını yükseltmektedir (Amaar et al., 2022).

Bu gözlemler, sahte ilanların stratejik olarak düşük giriş bariyerli pozisyonları tercih ettiğini doğrulamakta ve modelleme aşamasında bu özelliklerin ağırlıklandırılmasının önemini vurgulamaktadır.

2.3. Özellikler Arası Korelasyon Analizi

Veri setindeki sayısal ve dönüştürülmüş özellikler arasındaki ilişkileri daha derinlemesine anlamak amacıyla korelasyon matrisi hesaplanmıştır. Bu analiz, Pearson korelasyon katsayıları kullanılarak gerçekleştirilmiş olup, sahte iş ilanlarını (fraudulent) etkileyen potansiyel ilişkileri ortaya koymaktadır. Şekil 1'de verilen korelasyon matrisi dikkate alındığında, *min_salary_capped* ve *max_salary_capped* arasında, bu özelliklerin aynı *salary_range* sütunundan türetilmesi nedeniyle beklenen çok yüksek pozitif korelasyon (0,95) bulunmuştur. *telecommuting* ile *fraudulent* arasındaki ilişki ise zayıf ve anlamlı olmayan düzeydedir (0.03). Buna karşılık, *has_company_logo* ile *fraudulent* arasındaki belirgin negatif korelasyon (-0.26), şirket logosu olmayan ilanların sahte olma olasılığının daha yüksek olduğunu ve daha önceki çalışmalarla uyumlu bir şekilde, eksik değer analizindeki bulguları (logo eksikliğinin sahte ilanlarda %67'den fazla görülmesi) desteklemektedir. Benzer şekilde, *has_questions* ile *fraudulent* arasındaki negatif korelasyon (-0.09) da sahte ilanlarda bu tür soruların daha az kullanıldığını doğrulamaktadır. Öte yandan, metin uzunlukları (*desc_length*, *req_length*, *bene_length*) birbiriyle orta düzeyde korelasyon gösterirken, fraudulent ile ilişkileri oldukça zayıftır. Bu durum, sahte ilanların uzunluk açısından benzer ancak içerik kalitesi (örn. detay eksikliği) bakımından farklılık gösterebileceğine işaret etmektedir.



Şekil 1: Özellikler Arası Korelasyon Matrisi.

3. Veri Ön İşleme ve Öznitelik Mühendisliği

Veri setinin kalitesi ve tutarlılığı, makine öğrenmesi modellerinin başarısı için hayati öneme sahiptir. Bu nedenle, kapsamlı bir ön işleme süreci uygulanmıştır. Ön işleme adımları, eksik değerlerin yönetimi, aykırı değerlerin kontrolü, metin verilerinin temizlenmesi ve özniteliklerin uygun formata dönüştürülmesini içermekte olup, bu işlemler modelin genelleme yeteneğini artırarak aşırı öğrenme (overfitting) riskini azaltmaktadır (Amaar et al., 2022; Vidros et al., 2017). Özellikle EMSCAD veri seti gibi dengesiz ve karmaşık yapıdaki veri setlerinde, etkili ön işleme ve öznitelik mühendisliği, sahte iş ilanı tespiti performansını önemli ölçüde iyileştirmektedir (Alghamdi & Alharby, 2019; Chowdhury et al., 2021).

3.1. Maaş Verilerinin Düzenlenmesi

Salary_range sütunundaki eksik değerler başlangıçta "0-0" ile doldurulmuştur. Daha sonra bu sütun, *min_salary* ve *max_salary* olmak üzere iki sayısal sütuna ayrılmıştır. Aykırı değerlerin (outliers) etkisini azaltmak amacıyla Interquartile Range (IQR) yöntemi kullanılarak bu maaş sütunları kırpılmış

ve *min_salary_capped* ve *max_salary_capped* sütunları oluşturulmuştur. Bu yaklaşım, maaş aralığındaki aşırı uç değerlerin model performansını olumsuz etkilemesini önlemekte ve gerçek dünya verilerindeki gürültüyü azaltmaktadır (Kaur & Kaur, 2023). Benzer şekilde, önceki çalışmalarda maaş bilgilerinin eksikliğinin sahte ilanlarda daha yaygın olduğu belirtilmiş ve bu tür tekniklerin önemi ifade edilmiştir (Dutta & Bandyopadhyay, 2020).

3.2. Metin Verilerinin Vektörleştirilmesi

title, *description*, *requirements*, *company_profile* ve *benefits* gibi metin tabanlı sütunlar için normalizasyon işlemleri gerçekleştirilmiştir. Bu işlemler, tüm harflerin küçük harfe çevrilmesi, noktalama işaretlerinin, URL'lerin ve stopword'lerin kaldırılmasını içermiştir. Temizlenen metinler daha sonra TF-IDF (Term Frequency-Inverse Document Frequency) tekniği kullanılarak vektörleştirilmiştir. Bu işlem, seyrek veri setleri için özellikle önemlidir ve kelimelerin belge içindeki önemini vurgulayarak sahte ilanlardaki tekrar eden kalıpları (örneğin vaat edici ifadeler) daha iyi yakalamaktadır (Amaar et al., 2022; Alghamdi & Alharby, 2019). TF-IDF, sahte iş ilanı tespiti literatüründe en yaygın kullanılan öznitelik çıkarma yöntemlerinden biridir ve etkili sonuçlar vermektedir (Chowdhury et al., 2021; Singhania et al., 2023).

3.3. Yapısal Özniteliklerin Dönüştürülmesi

Kategorik öznitelikler arasında frekansı %5'in altında olan seyrek kategoriler, modelin öğrenme yeteneğini artırmak amacıyla *Other* adlı tek bir kategori altında toplanmıştır. Bu nadir kategori birleştirme işlemi, yüksek nicelikli özelliklerdeki gürültüyü azaltarak modelin genellemesini iyileştirmektedir (Vidros et al., 2017). Son olarak, sayısal öznitelikler StandardScaler ile normalize edilmiş, kategorik öznitelikler ise OneHotEncoder ile sayısal formata dönüştürülmüş ve tüm öznitelik kümeleri birleştirilmiştir. Bu yaklaşım, ölçek farklılıklarını gidererek özellikle topluluk modellerde daha stabil eğitim sağlamakta ve literatürde sıkça tavsiye edilmektedir (Kaur & Kaur, 2023; Singhania et al., 2023).

Bu ön işleme adımları, veri setinin hem merkezi hem de federe öğrenme senaryolarında kullanılabilir hale gelmesini sağlamış olup, sonraki modelleme aşamalarında yüksek performans elde edilmesine temel oluşturmuştur.

3.4. Performans Değerlendirme Metrikleri

Bu çalışmada kullanılan sınıflandırma modellerinin performansı, karışıklık matrisi üzerinden hesaplanan doğruluk (Accuracy), kesinlik (Precision), duyarlılık (Recall), F1-skoru ve ROC-AUC metrikleri kullanılarak değerlendirilmiştir. Karışıklık matrisi; doğru pozitif (TP), doğru negatif (TN), yanlış pozitif (FP) ve yanlış negatif (FN) değerlerinden oluşmaktadır. Kullanılan metriklerin matematiksel tanımları aşağıda verilmiştir.

Doğruluk (Accuracy):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Kesinlik (Precision):

$$Precision = \frac{TP}{TP + FP}$$

Duyarlılık (Recall):

$$Recall = \frac{TP}{TP + FN}$$

F1-Skoru:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

ROC-AUC:

- ROC eğrisi, **Yanlış Pozitif Oran (FPR)** ile **Doğru Pozitif Oran (TPR)** arasındaki ilişkiyi gösterir.

$$TPR = Recall = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$

- **AUC**, ROC eğrisi altında kalan alanı temsil eder ve modelin sınıfları ayırt etme gücünü gösterir.

4. Merkezi Öğrenme Modelleri ve Sonuçları

Merkezi öğrenme yaklaşımında, sahte iş ilanı tespiti için dört temel model seçilmiştir: Lojistik Regresyon (LR), Rastgele Orman (RO), Destek Vektör Makineleri (DVM) ve bu LR ile RO'yu kullanan hibrit modeli (LR+RF). Bu modeller, sahte iş ilanı tespiti literatüründe sıkça kullanılan ve yüksek performans gösteren yöntemler arasındadır (Vidros et al., 2017; Amaar et al., 2022). Özellikle RO ve topluluk tabanlı yaklaşımlar, karmaşık metin ve kategorik özelliklerdeki doğrusal olmayan ilişkileri etkili bir şekilde yakalayarak üstün sonuçlar vermektedir (Alghamdi & Alharby, 2019; Chowdhury et al., 2021). Hibrit modeller ise, farklı sınıflandırıcıların güçlü yönlerini birleştirerek genelleme yeteneğini artırmakta ve dengesiz veri setlerinde daha iyi performans sağlamaktadır (Kaur & Kaur, 2023).

4.1. Dengesiz Veri Sorunu ve Çözümü

Veri seti, gerçek ilanların baskın olduğu yüksek oranda dengesiz bir yapıya sahiptir (gerçek ilanlar %95,2, sahte ilanlar %4,8). İlk denemelerde, modeller bu dengesizlik nedeniyle çoğunluk sınıfını (gerçek ilanlar) tercih ederek tüm iş ilanlarını gerçek olarak tahmin etmiştir. Bu durum, sahte ilanları tespit etme yeteneğini temsil eden Duyarlılık (Recall) metriğini büyük ölçüde düşürmüştür. Bu sorun, sahte iş ilanı tespiti çalışmalarında yaygın bir zorluk olup, modellerin az veri içeren sınıfa (sahte ilanlar) karşı yanlılık (bias) oluşturmasına yol açmaktadır.

Bu sorunu gidermek için veri çoğaltma (data augmentation) teknikleri uygulanmıştır. LR modeli için Random OverSampler (ROS) kullanılırken, RO, DVM ve hibrit model için kategorik öznitelikleri de destekleyen SMOTENC kullanılmıştır. SMOTENC, kategorik ve sürekli özelliklerin bulunduğu karma veri setlerinde sentetik örnekler üretirken daha etkili olup, aşırı öğrenme riskini azaltmaktadır (Chowdhury et al., 2021; Kaur & Kaur, 2023). Benzer şekilde, literatürde SMOTE varyantlarının (örneğin SMOTENC) dengesiz iş ilanı veri setlerinde Duyarlılık ve F1-skorunu önemli ölçüde iyileştirdiği rapor edilmiştir (Amaar et al., 2022).

4.2. Model Performansları

Veri setinin temizlenip yeniden işlenmesi ve veri çoğaltma tekniklerinin optimize edilmesi sonrasında modellerin performansı önemli ölçüde artmıştır. Sonuçlar Tablo 1'de gösterilmektedir.

Tablo 1. Merkezi Modellerin Sonuçları.

Model	Doğruluk	Kesinlik	Duyarlılık	F1-Skoru	AUC	Eğitim Süresi (sn)
Lojistik Regresyon	98,56	89,39	79,91	84,29	98,65	13,80
Rastgele Orman	98,69	90,87	81,30	85,67	98,95	167,08
Destek Vektör Makinesi	98,31	97,10	67,22	79,10	98,72	6971,43
Rastgele Orman + Lojistik Regresyon	98,76	93,33	80,84	86,49	99,06	1663,05

Tablo 1'de verilen sonuçlara göre, RO ile LR'nin kombinasyonundan oluşan hibrit model, en yüksek genel performansı sergilemiştir. Bu model, hem yüksek doğruluk (%98,76) hem de yüksek F1 skoru (%86,49) ve AUC değeri (%99.06) elde etmiştir. Hibrit yaklaşım, RO'nin doğrusal olmayan bir desen yakalama yeteneğini LR'nin hızlı ve yorumlanabilir tahminleriyle birleştirerek dengesiz sınıflarda daha dengeli bir performans sağlamakta olup, benzer hibrit yöntemlerin sahte ilan tespitinde üstünlüğü literatürde de doğru-

lanmıştır (Alghamdi & Alharby, 2019; Singhania et al., 2023). DVM, en yüksek hassasiyete (%97,10) sahip olsa da düşük duyarlılık (Recall: %67,22) ve çok uzun eğitim süresi (6971,43 saniye) nedeniyle optimal model olarak seçilmemiştir. DVM'nin yüksek hesaplama maliyeti, büyük veri setlerinde bilinen bir dezavantajdır (Chowdhury et al., 2021). Bu sonuçlar, merkezi öğrenme modellerinin sahte iş ilanı tespitinde yüksek doğruluklar elde edebileceğini göstermekte olup, hibrit modelin pratik uygulanabilirliği daha yüksektir.

5. Federe Öğrenme Uygulaması ve Kıyaslama

Merkezi yaklaşım yüksek tahminsel doğruluk sağlasa da veri gizliliğini önceliklendiren senaryolarda Federe Öğrenme kullanılması gerekmektedir. Federe Öğrenme, veriyi merkezi sunucuya aktarmadan cihazlar veya kurumlar üzerinde yerel eğitim yapılmasını sağlar. Yalnızca model parametreleri paylaşılır, böylece hassas verilerin (örneğin iş başvurularındaki kişisel bilgiler) gizliliği korunur (McMahan et al., 2017; Konečný et al., 2016). Bu yaklaşım, özellikle KVVK gibi gizlilik düzenlemelerinin bağlayıcı olduğu durumlarda avantaj sağalr ve sahte iş ilanı tespiti gibi dolandırıcılık odaklı görevlerde, farklı platformların veri paylaşmadan iş birliği yapmasına olanak tanır (Yang et al., 2019; Li et al., 2020).

5.1. Federe Öğrenme Yapılandırması

Çalışmada, Federe Öğrenme ortamını simüle etmek amacıyla, veri seti gerçek dünya uygulamalarını yansıtacak şekilde Non-IID (bağımsız ve aynı dağılımdan olmayan) olarak 5 farklı müşteri (client) arasında dağıtılmıştır. Bu heterojen dağıtım için $\alpha = 0.05$ parametresi ile Dirichlet dağılımı kullanılmıştır. Düşük α değeri, istemciler arası yüksek heterojenliği simüle ederek gerçekçi bir senaryo yaratmaktadır (Li et al., 2020). Non-IID dağılım, farklı iş platformlarının (örneğin bölgesel veya sektörel farklar) veri farklılıklarını temsil etmekte ve Federe Öğrenme'nin bu tür heterojenlikteki performansını test etmektedir (Karimireddy et al., 2020).

5.2. Federe Öğrenme Algoritmaları

FedAvg ve FedProx algoritmaları Python programlama dilinde PyTorch kütüphanesi kullanılarak manuel fonksiyonlarla uygulanmıştır. Bu manuel uygulama, algoritmaların esnekliğini artırarak hiper-parametre optimizasyonuna olanak sağlamıştır.

1. **Federated Averaging (FedAvg)**: Her istemci, yerel verisi üzerinde modeli birden fazla kez günceller (10 yerel epoch kullanılmıştır) ve ağırlıklarını merkezi sunucuya gönderir. Sunucu bu ağırlıkların ağırlıklı ortalamasını alarak küresel modeli günceller. FedAvg, basitliği ve iletişim verimliliği nedeniyle Federe Öğrenme'nin temel algoritmasıdır ancak Non-IID verilerde yakınsama sorunları yaşayabilir (McMahan et al., 2017).

1. **Federated Optimization with Proximal Term (FedProx)**: FedAvg'nin bir genellemesidir ve Non-IID veri setlerinde daha stabil yakınsama sağlamak üzere önerilmiştir. FedProx, yerel eğitim kaybına, yerel ağırlıkların küresel ağırlıklardan çok uzaklaşmasını engelleyen bir proksimal terim ekler ($\mu = 2$ olarak ayarlanmıştır). Bu terim, heterojenlik kaynaklı sapmaları sınırlayarak daha güçlü bir performans sağlar (Li et al., 2020; Karimireddy et al., 2020).

5.3. Federe Öğrenme Sonuçları

Tablo 2'deki sonuçlara göre hem FedAvg hem de FedProx, 4 turda yakınsayarak aynı nihai doğruluk oranına (%95,16) ulaşmıştır. Bu sonuçlarda, FedProx, FedAvg'ye göre önemli ölçüde daha düşük iletişim maliyetine (11.15 MB) sahip olması nedeniyle federe yaklaşımlar arasında en verimli model olarak öne çıkmıştır. Düşük iletişim maliyeti, proksimal terimin yerel güncellemeleri kısıtlayarak daha az parametre değişimi yaratmasından kaynaklanmakta olup, sınırlı bant genişliği olan dağıtık sistemlerde kritik avantaj sağlamaktadır (Li et al., 2020). Bu bulgu, Non-IID senaryolarda FedProx'ın daha stabil ve verimli olduğunu doğrulayan literatürle uyumludur (Karimireddy et al., 2020).

Tablo 2. Federe Öğrenme Modellerinin Sonuçları.

Model	Doğruluk (%)	Yakınsama Turu	İletişim Miktarı (MB)	Hesaplama Süresi (s)
FedAvg	95,16	4	44,59	311,73
FedProx	95,16	4	11,15	600,54

6. Merkezi ve Federe Öğrenme Yaklaşımlarının Kıyaslanması

Nihai olarak Tablo 3'te, en iyi merkezi model (RF+LR) ile Federe Öğrenme modellerini gizlilik, performans ve verimlilik metrikleri üzerinden karşılaştırılmıştır.

Tablo 3. Federe Öğrenme ve Merkezi Modelleri Kıyaslanması.

	LR+RF	FedAvg	FedProx
Doğruluk (%)	**98,76**	95,16	95,16
Eğitim Süresi (s)	1663,05	**311,72**	600,54
Gizlilik Seviyesi	Düşük	Yüksek	Yüksek
İletişim Miktarı (MB)	-	44,59	**11,15**
Yakınsama Turu	0 (Direkt)	4	4

Merkezi model %98,76 doğruluk ile tahmin doğruluğu açısından %95,16 değere sahip federe öğrenme modellerinden daha iyi performans göstermiştir. Bu, tüm veriye erişimin, modelin genelleştirme yeteneği üzerinde hala üstün bir etki yarattığını göstermektedir. Merkezi öğrenme, tam veri erişimiyle daha karmaşık desenleri öğrenebilmektedir (Yang et al., 2019).

Federe Öğrenme yaklaşımları ise yüksek gizlilik seviyesi sunarak, kullanıcı verilerinin güvenliğini garanti altına alır (Konečný et al., 2016). FedProx, bu gizliliği korurken, FedAvg'den dört kat daha düşük iletişim maliyeti (11.15 MB) sunarak, özellikle sınırlı bant genişliğine sahip dağıtık ağlarda ideal bir çözüm haline gelmektedir (Li et al., 2020). Merkezi öğrenme bu gizlilik seviyesini sağlayamaz ve veri merkezileştirmesi nedeniyle gizlilik riskleri taşır.

Sahte iş ilanı tespiti gibi kritik bir görevde, eğer veri gizliliği ikincil bir öncelikse, merkezi RO+LR hibrit modeli en yüksek doğruluk oranını sunar. Ancak, eğer veri güvenliği ve kullanıcı gizliliği öncelikli ise, FedProx yüksek doğruluğa yakın sonuçlar elde ederken, en verimli iletişim maliyetini sağlayan en uygun Federe Öğrenme çözümü olarak öne çıkmaktadır. Bu çalışma, sahte iş ilanı tespiti alanında Federe Öğrenme'nin pratik uygulanabilirliğini göstererek, gizlilik odaklı gelecek çalışmalara temel oluşturmaktadır (Yang et al., 2019; Karimireddy et al., 2020).

7. Sonuç ve Öneriler

Bu çalışma, sahte iş ilanı tespiti görevinde merkezi ve federe öğrenme paradigmalarının güçlü ve zayıf yönlerini kapsamlı bir şekilde ortaya koymuştur. Merkezi öğrenme yaklaşımı, tüm verilere tam erişim sayesinde en yüksek sınıflandırma doğruluğunu (%98,76) elde etmiş olup, özellikle Rastgele Orman ile Lojistik Regresyon'un hibrit kombinasyonu, dengesiz veri setlerinde bile üstün F1-skoru ve AUC değerleri ile en iyi performansı sergilemiştir. Bu sonuçlar, geleneksel merkezi modellerin sahte iş ilanı gibi metin ağırlıklı ve karmaşık özelliklere sahip görevlerde hala güçlü bir model oluşturduğunu doğrulamaktadır.

Buna karşın, veri gizliliği ve güvenliği endişesi ön planda olduğunda, federe öğrenme, özellikle FedProx algoritması, etkili ve pratik bir alternatif sunmaktadır. FedProx, Non-IID veri ortamlarında bile %95'in üzerinde doğruluk oranı elde etmiş ve 11.1 MB gibi bir iletişim maliyetiyle verimlilik sağlamıştır. Bu da sınırlı bant genişliği olan dağıtık sistemlerde önemli bir avantaj sağlar. Merkezi öğrenmenin aksine, federe öğrenme verilerin yerel olarak kalmasını sağlayarak KVKK gibi düzenlemelere uyumu kolaylaştırmakta ve potansiyel veri sızıntısı risklerini ortadan kaldırmaktadır. Gelecek çalışmalarda, FedProx'un FedNova gibi diğer algoritmalarla Non-IID senaryolarda kıyaslanması, model genellemesinin gerçek zamanlı büyük veri setleri üzerinde test edilmesi ve metinsel özellik çıkarımı için BERT tabanlı modellerin federe öğrenmeye entegrasyonu hedeflenmektedir.

KAYNAKÇA

Alghamdi, B., & Alharby, F. (2019). An intelligent model for online recruitment fraud detection. Journal of Information Security, 10(3), 154-166.

Amaar, A., Aljedaani, W., Rustam, F., Ullah, S., Rupapara, V., & Ludi, S. (2022). Detection of fake job postings by utilizing machine learning and natural language processing approaches. Neural Processing Letters, 54(3), 2219-2246. doi:10.1007/s11063-021-10727-z

Chowdhury, R., Hasan, M., & Rahman, M. (2021). Fake job detection and analysis using machine learning and deep learning algorithms. International Journal of Computational Intelligence Systems, 14(1), 1745-1756.

Dutta, S., & Bandyopadhyay, S. K. (2020). Fake job recruitment detection using machine learning approach. International Journal of Engineering Trends and Technology, 68(4), 48-53.

Karimireddy, S. P., Kale, S., Mohri, M., Reddi, S., Stich, S., & Suresh, A. T. (2020). SCAFFOLD: Stochastic controlled averaging for federated learning. Proceedings of the 37th International Conference on Machine Learning (ICML), 5132-5143.

Kaur, R., & Kaur, S. (2023). Machine learning approach in predicting fraudulent job advertisement. International Journal of Research in Applied Science and Engineering Technology, 11(1), 123-135.

Konečný, J., McMahan, H. B., Yu, F. X., Richtárik, P., Suresh, A. T., & Bacon, D. (2016). Federated learning: Strategies for improving communication efficiency. arXiv preprint arXiv:1610.05492.

Li, T., Sahu, A. K., Zaheer, M., Sanjabi, M., Talwalkar, A., & Smith, V. (2020). Federated optimization in heterogeneous networks. Proceedings of Machine Learning and Systems, 2, 429-450.

McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), 1273-1282.

Singhania, S., et al. (2023). Detecting fake job postings using machine learning. arXiv preprint arXiv:2304.02019.

Vidros, S., Kolias, C., Kambourakis, G., & Akoglu, L. (2017). Automatic detection of online recruitment fraud. Proceedings of the 2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM), 863-866.

Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology, 10(2), 1-19.

Bölüm 12

KADINLAR VE ENGELLİLER İÇİN GÜVENLİ ULAŞIM: WEB TABANLI ARAÇ KİRALAMA MODELİ

Hazal MIHCI [1]
Melek OFLUOĞLU [2]
Negin MELEK [3]

1 Giresun Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği
2 Giresun Üniversitesi , Mühendislik Fakültesi, Bilgisayar Mühendisliği
3 Doç. Dr. Gümüşhane Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi, Elektrik-Elektronik Mühendisliği, Telekomünikasyon Anabilim Dalı

1. Giriş

Günümüzde güvenli ve erişilebilir ulaşım, toplumsal yaşamın sürdürülebilirliği açısından her kesimin vazgeçilmezi olmuştur. Bu eğilim, alternatif ulaşım modellerine ve paylaşım temelli hizmetlerine olan ilgiyi artırmış; araç kiralama sistemlerinin yaygınlaşmasını beraberinde getirmiştir. Bu hizmetler, sınırlı bir süre ve ücreti karşılığında otomobil, kamyonet, kamyon ve motosiklet gibi çeşitli araçların kiralanması esasına dayanarak, kullanıcıların hizmetine sunulması anlayışı üzerine kuruludur (Maximiliano, 2011; Cho ve Rust, 2010). Türkiye'de ve Dünya'da hızla gelişen bu sektör, özellikle iş ve tatil amaçlı seyahatlerin artmasıyla, daha da önemli hale gelmiştir (Nazlı, 2022).

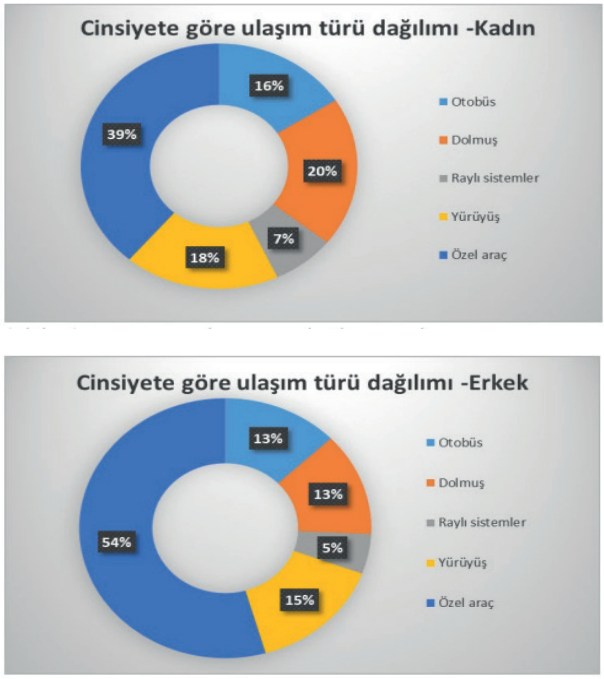
1997 yılında Hertz'in ilk dijital araç kiralama platformunu kurmasıyla birlikte araç kiralama hizmetleri kullanıcıların araç kiralama sürecini kolaylaştırarak sektörde yeni bir çağ başlatmıştır (Akay, 2014). Zamanla teknolojideki ilerlemeler, internet üzerinden çevrimiçi rezervasyon sistemleri, self check-in olanakları ve global dağıtım sistemlerinin entegrasyonu gibi yeniliklerle araç kiralama sektörünü dönüştürerek hem kullanıcılar hem de hizmet sağlayıcılar açısından iş süreçlerinin daha etkin biçimde tasarlanmasına katkıda bulunmuştur (Akay, 2014).

1990'lar ve 2000'li yıllarda ciddi bir ivme yakalayan Türkiye'deki araç kiralama sektörü küresel düzeydeki bu gelişmelerle paralellik göstermiştir. Günümüz Türkiye'sinde ise araç kiralama sektörü, farklı araç tiplerini kapsayan geniş filosu aracılığıyla, bireysel ve kurumsal kullanıcıların değişen ulaşım gereksinimlerine yönelik çözümler sunmaktadır. Özellikle artan trafik yoğunluğu, yaşam maliyetleri ve park alanı yetersizliği gibi nedenler, büyük şehirlerde yaşayan bireyleri araç sahipliği yerine kısa dönemli araç kiralama hizmetlerine yöneltmektedir. Tüm bu gelişmeler, araç kiralama sektörünün Türkiye'de ulaşım ekosisteminin ayrılmaz bir parçası haline gelmesine zemin hazırlamıştır.

Araç kiralama sektörünün tarihsel gelişimi üzerine yapılan çeşitli araştırmalar, sektörün toplumsal etkilerini de detaylı bir şekilde incelemeyi gerekli kılmıştır (Maximiliano, 2011; Cho ve Rust, 2010; Akay, 2014; Nazlı, 2022). Akay (2014), araç kiralama sektörünün ABD'deki ilk uygulamalarından başlayarak, çevrimiçi hizmet sistemlerinin yaygınlaşmasına kadar olan süreci analiz ederken, Nazlı (2022) Türkiye'deki araç kiralama sektörünün 1990'lardan itibaren gösterdiği hızlı büyümeyi ve bu gelişimin sosyo-ekonomik dinamiklerini incelemiştir. Buna karşın mevcut literatür incelendiğinde, araç kiralama hizmetlerinin toplumsal yönden kırılgan durumunda olan (kadınlar, çocuklar, yaşlılar, engelliler gibi) gruplar üzerindeki etkilerini ve bu grupların özel ihtiyaçlarına yönelik çözüm yaklaşımlarını ele alan çalışmaların oldukça sınırlı olduğu fark edilmiştir. Özellikle şehir planlaması ve ulaşım politikaları alanındaki araştırmalar, genellikle erkek egemen bir bakış açısıyla geliştirilen bu sistemlerin kırılgan grupların ulaşım önceliklerini göz ardı ettiğini ortaya koymaktadır (Ertor Sarıışık ve Öcalır, 2024). Bu durum, toplumsal fırsat eşitliğinin uygulanması açısından bir engel oluşturmaktadır.

Kadınların şehir hayatındaki toplumsal rollerinin değişmesi ve giderek daha aktif hale gelmesi ulaşım konusundaki zorlukları da beraberinde getirmiştir (Önder, 2020). Çocuk, yaşlı veya aile bireylerine yönelik özel bakım sorumlulukları, kadınların ulaşım gereksinimlerini artırmakta ve mevcut ulaşım sistemlerinin bu ihtiyaçları karşılamakta yetersiz kaldığı görülmektedir. Carver ve Veitch (2020) toplu taşıma kullanımının kadınların yaşı, eğitim seviyesi ve araç sahiplikleriyle ters orantılı olduğunu ortaya koymuştur. Scheiner (2014) ise anneliğin ulaşım alışkanlıkları üzerindeki etkisine dikkat çekmiş, çocuk sayısının artmasıyla araç kullanma eğiliminin de yükseldiğini belirtmiştir. Bu veriler, kadınların ulaşım tercihlerinde güvenlik, esneklik ve konfor gibi faktörlerin temel belirleyiciler olduğunu göstermektedir.

Küresel ölçekte baktığımızda bazı kentlerde ulaşım sistemlerinde cinsiyet duyarlı planlamalar yaptığı görülmektedir. Örneğin İsveç'in Malmö kenti ve Avusturya'nın Viyana şehirleri kadınlar başta olmak üzere diğer kırılgan gruplar için güvenli ulaşım koşulları oluşturulabileceğini göstermektedir (Önder, 2020). Ancak ülkemizde bu yönde kapsamlı bir planlama yaklaşımının henüz gelişmediği görülmektedir. Hâlbuki kadına yönelik şiddetin toplumsal bir sorun olarak varlığını sürdürmesi (Keysan, Kaygan ve Kaygan, 2022), kadınların ulaşım esnasında yaşayabilecekleri taciz ve şiddet riskinin göz ardı edilmemesi gerekmektedir (Durmuş, 2013; Hanözü Mollahasanoğlu ve diğerleri, 2015).



Şekil 1: Kadın ve erkeklerin kullandığı ulaşım türleri ve oranları

Yine toplumun kırılgan grupları içinde yer alan engelli bireylerin seyahat özgürlüğü de toplumsal katılım ve eşitlik açısından literatürde tartışılması gereken önemli bir konudur. Bu bireylerin seyahat haklarını tam anlamıyla kullanamamaları, fiziksel ve sistemsel yaşadıkları zorluklar; toplumsal hayata dâhil olmalarını ve temel ihtiyaçlarından yeteri kadar yararlanmalarını güçleştirmektedir (Takeda ve Card, 2002; Kaygısız ve Bulgan, 2015; Bi, 2006; Israeli, 2002; Burnett ve Bender, 2001; Darcy ve Daruwalla, 1999; Erten ve Akel, 2020; Reis, Gücükoğlu ve Eskici, 2014). Bu sorunlar, engelli bireylerin ulaşımda karşılaştıkları sorunları iyileştirmeye yönelik çeşitli çözümlerin geliştirilmesini gerekli kılmaktadır. Sevginer ve diğerleri (2011) tarafından geliştirilen engelli dostu taksi platformu bu konuda gerçekleştirilebilecek uygulamalara örnek teşkil etmektedir. Ancak bu tür uygulamaların sayısı oldukça sınırlı olup engelli bireylerin ulaşımda karşılaştıkları tüm sorunlara kalıcı çözümler sunmakta yetersiz kalmaktadır. Bu eksiklik, seyahat hizmetlerinde kapsayıcılığın artırılması gerekliliğini ortaya koymaktadır.

Bu bağlamda Türkiye'de ve Dünya'da kadınlar ve engellilere özel olarak tasarlanmış araç kiralama platformu önerisi bu çalışmasının temel çıkış noktasını oluşturmaktadır. Mevcut çalışmalar araç kiralama sektörünün genel dinamiklerini veya kadın ve engelli bireylerin ulaşım sorunlarını birbirinden bağımsız ele almaktadır. Yapılan araştırmalar sonucu ulaşım sistemlerinde hem toplumsal cinsiyet hem de engellilik perspektifini eş zamanlı olarak ele alıp somut bir çözüm üretmeye yönelik bir uygulamaya rastlanamamıştır. Bu sebeple kadınlar ve engelliler için özel olarak geliştirilen bir araç kiralama platformu fikri, bu çalışmanın özgün değerini oluşturmaktadır. Uygulamanın Doğu Karadeniz (Trabzon, Ordu, Giresun) bölgesine odaklanması da bir diğer özgün yönüdür.

Bu platformun öne çıkan yönü kırılgan grupların seyahat özgürlüğünü artırmaya yönelik pratik çözümler sunma potansiyeline sahip olmasıdır. Tasarlanan ve önerilen araç kiralama platformu, güvenlik ve erişilebilirlik gereksinimlerine yanıt vererek, kırılgan grupların sosyal ve ekonomik yaşama katılımlarını desteklemesi için geliştirilmiştir.

Araçlarda bulunan kadınlara yönelik güvenlik özellikleri (acil durum butonu gibi), seyahat sırasında kendilerini daha güvende hissetmelerini sağlayarak onların iş, eğitim ve sosyal hayata katılımını kolaylaştırarak toplumsal cinsiyet eşitliğine katkı sunacaktır. Benzer şekilde, engelli bireyler için özel olarak tasarlanmış araçlar ve erişilebilir bir platform arayüzü, onların ulaşım engellerini azaltarak bağımsız hareket etmelerini sağlayacaktır. Böylece engelli bireylerin sosyal hayata katılımları ve yaşam kalitesi artacak; toplumsal entegrasyonlarını güçlendirecektir.

2. Teorik Arkaplan

Geliştirilen bu uygulama, kadınlar ve engelli bireylerin araç kiralama süreçlerini kolaylaştırmayı hedefleyen bir araç kiralama platformu olarak tasarlanmıştır. Kullanıcıların araç kiralama işlemlerini güvenli, hızlı ve kolay bir şekilde gerçekleştirebilmeleri sağlanmıştır. Geliştirme sürecinde, uygulamanın güvenliği, kullanıcı dostu arayüz tasarımı ve erişilebilirliği dikkate alınmıştır. Belirlenen hedefler doğrultusunda çeşitli yazılım teknolojileri ve araçları kullanılmıştır. Aşağıda Tablo 1.'de bu teknolojilere ilişkin özet bilgiler sunulmuştur.

Tablo 1: Kullanılan Temel Teknolojiler

Teknoloji/Araç	Kullanım Amacı	Platformdaki Rolü
ASP.NET Core MVC	Web uygulamasının temel mimarisini oluşturmak	Kullanıcı arayüzü, kontrol katmanı ve iş mantığı bu yapı üzerinden geliştirildi.
SQL Server Management	Verilerin güvenli şekilde saklanması	Kullanıcı, araç ve rezervasyon bilgileri veritabanında depolandı.
Visual Studio	Yazılım geliştirme ortamı sağlamak	Projenin tüm kodlama, test ve hata ayıklama süreçleri bu IDE üzerinde yürütülmüştür.
MVC (Model-Wiev-Controller	Uygulama katmanlarını birbirinden ayırarak düzenli bir yapı oluşturmak	Veri, kullanıcı arayüzü ve kontrol akışı birbirinden bağımsız hale getirilerek, bakım ve geliştirme kolaylığı sağlanmıştır.
CQRS (Command Query Responsibility Segregation)	Okuma ve yazma işlemlerini ayırarak performans artırmak	Veritabanı işlemleri optimize edildi.
Web API	Farklı istemciler arasında veri alışverişi sağlamak	Platformun mobil sürümüyle iletişim kurmak için REST tabanlı API kullanıldı.
Mediator Tasarım Deseni	Katmanlar arası bağımlılığı azaltmak	Sistem bileşenleri arasında daha esnek bir yapı kuruldu.
Area	Modüler bir proje yapısı kurmak	Uygulama, farklı işlevsel alanlara bölünerek (örneğin kullanıcı, yönetici gibi), kodun düzenlenmesi ve yönetimi kolaylaştırılmıştır.
Razor Pages	Dinamik web sayfaları oluşturmak	Kullanıcı dostu ve etkileşimli arayüzler tasarlanmış, veriler sunucu tarafında dinamik olarak işlenmiştir.
Onion Architecture (Soğan Mimarisi)	Uygulama bağımlılıklarını azaltmak ve katmanlı bir yapı kurmak	İş mantığı merkeze alınarak, dış katmanlardan (UI, veri erişimi) bağımsız bir yapı oluşturulmuştur; böylece esneklik ve test edilebilirlik artırılmıştır.

2.1. Sistem Bileşenleri ve Seçimleri

ASP.NET Core, açık kaynaklı, platformdan bağımsız ve yüksek performanslı bir web geliştirme çatısıdır. Bu çalışma kapsamında ASP.NET Core'un tercih edilmesinin temel nedeni, RESTful servis geliştirme olanağı ve modüler yapısının sağladığı yönetilebilirliğin artmasıdır.

SQL Server, verileri güvenli ve hızlı biçimde saklaması sebebiyle güçlü bir ilişkisel veritabanı yönetim sistemidir. Geliştirilen uygulamada kullanıcı bilgileri, araç detayları ve rezervasyon işlemleri gibi verilerin tutarlı biçimde yönetilmesi amacıyla bu sistem tercih edilmiştir.

MediatR kütüphanesi, .NET tabanlı projelerde CQRS (Command Query Responsibility Segregation) yapısının uygulanmasını kolaylaştıran bir yapı sunar. Uygulama içerisindeki işlemleri komut ve sorgu olarak birbirinden ayırarak yazılım mimarisinin düzenli, anlaşılır ve test edilebilir olmasına katkı sağlar. Ayrıca katmanlar arası bağımlılıkların azaltılması, sistemin esnekliğinin artırılması uygulamanın ölçeklenmesi sürecinde ortaya çıkabilecek karmaşıklıkların önüne geçilmesi amacıyla da MediatR tercih edilmiştir.

MVC (Model-View-Controller) tasarım deseni; uygulamanın veri modeli (Model), kullanıcı arayüzü (View) ve iş mantığı (Controller) olmak üzere üç ayrı katmanda ele alınır. Bu yapı sayesinde, geliştiricilerin sorumluluk alanları netleştirilerek modülerliği artırılır ve uygulamanın yönetilebilirliği kolaylaşır. Bu uygulama kapsamında, kullanıcı arayüzü ile iş mantığının birbirinden ayrılması hedeflenmiş; bakım ve test süreçlerinin daha kolay ve verimli bir şekilde yürütülmesi sağlanmıştır.

Area yapısı, ASP.NET MVC projelerinde uygulamanın farklı modüllere ayrılarak düzenli bir biçimde yönetilmesini sağlayan yapıdır. Bu platform kapsamında, kullanıcı işlemleri, yönetici panelleri ve araç yönetimi gibi bölümler ayrı alanlarda (areas) toplanmıştır. Böylece kod karmaşası azaltılmış, sistemin modülerliği ve kolay yönetilebilirliği güçlendirilmiştir.

Web API, istemci ve sunucu arasında veri aktarımını HTTP protokolü aracılığıyla sağlayan bir iletişim teknolojisidir. Geliştirilen uygulama kapsamında Web API kullanılarak kullanıcı, araç ve rezervasyon gibi temel işlemler dış sistemlerle güvenli biçimde paylaşılabilir hâle gelmiş, böylece sistemin farklı platformlarla uyumu artırılmış ve erişilebilir bir yapı elde edilmiştir.

Visual Studio, .NET tabanlı projeleri geliştirmek için yaygın olarak kullanılan güçlü bir geliştirme ortamıdır (IDE). Bu uygulama kapsamında Visual Studio'dan yararlanılarak kodlama, derleme ve hata ayıklama süreçlerinin düzenli bir biçimde yürütülmesi sağlanmıştır. Ayrıca ekip içi çalışmalarda uyum ve kaynak yönetiminin etkin biçimde gerçekleştirilmesi açısından kolaylıklar sağlamıştır.

Onion Architecture, yazılımın merkezine iş kurallarını (domain logic) yerleştirip dış katmanların (örneğin veritabanı, arayüz vb.) bağımlılığını en aza indirmeyi amaçlayan bir mimari modeldir. Geliştirilen uygulama kapsamında bu yapı tercih edilerek bağımlılıklar azaltılmış, katmanlar arası geçişler düzenli hale getirilmiş, sistemin bakımı ve test edilebilirliği artırılmıştır.

Platform aşağıdaki katmanlar üzerine yapılandırılmıştır:

- **Domain Katmanı:** İş kuralları ve temel varlıkları içerir.
- **Application Katmanı:** Komut ve sorguların işlendiği bölümdür.
- **Infrastructure Katmanı:** Veritabanı bağlantıları ve e-posta gibi dış sistemlerle etkileşimi sağlar.
- **Presentation Katmanı:** Kullanıcı arayüzü ve API uç noktalarını barındırır,

Bu teknolojik yapı, geliştirilen uygulamanın güvenli, hızlı ve yüksek performanslı biçimde çalışmasını sağlamış; kullanıcı deneyimi açısından güçlü bir sistem oluşturmuştur.

2.2. Deneysel Çalışmalar

Bu bölümde, kadınlar ve engellilere özel olarak geliştirilen araç kiralama platformunun işlevsel olarak doğru çalışıp çalışmadığını değerlendirmek amacıyla prototip geliştirme süreci, karşılaşılan teknik kolaylıklar ve zorluklar ile alınan güvenlik önlemleri doğrultusunda yapılan testler açıklanmaktadır. Yazılım geliştirme süreci boyunca yürütülen bu deneysel çalışmalar, platformun hedeflenen şekilde çalıştığını göstermiştir.

Deneysel aşamada kullanıcıların araç araması, filtreleme yapması ve rezervasyon oluşturması gibi temel işlemler incelenmiştir. Bu testlerin temel hedefi süreçte, engelli kullanıcıların erişilebilirlik ihtiyaçlarının karşılanması ve kadın kullanıcıların güvenlik açısından desteklenmesidir.

Ayrıca, deneysel sürecin sonunda farklı kullanıcı türlerinin (kadın ve engelli bireyler) sistemi nasıl kullandıklarını analiz eden senaryolar hazırlanmış, senaryolardan elde edilen bulgularla, sistemde iyileştirme çalışmaları yapılmıştır.

Platformun ana kullanıcı kitlesini kadınlar ve engelli bireyler oluşturduğundan, arayüz tasarımı sürecinde erişilebilirlik (accessibility) temel ilke olarak benimsenmiştir.

Bu doğrultuda:

- Kullanıcıların hızlı ve güvenli işlem yapabilmeleri için form, butonlar ve açılır menü (dropdown menü) gibi etkileşimli arayüz bileşenleri optimize edilmiştir.

- Her modülün (örneğin kullanıcı kaydı, araç ekleme, rezervasyon iptali) işlevselliği test edilmiştir.

Yapılan deneysel çalışmalar sonucunda, sistemin performans, erişilebilir ve kullanıcı deneyimi açısından hedeflenen niteliklere büyük ölçüde ulaştığı görülmüştür.

Platformun kullanıcı arayüzü, sade ve anlaşılır yapısıyla test kullanıcıları tarafından olumlu değerlendirilmiş; filtreleme ve arama fonksiyonları; tarih, saat ve özel donanım gibi kriterlere göre hızlı ve doğru sonuçlar üretmiştir. Kadın ve engelli kullanıcılar için oluşturulan rezervasyon senaryolarında sistem, sürecin tüm aşamalarında kararlı bir performans sergilemiştir. Güvenlik testlerinde, JWT tabanlı doğrulama sistemi yetkisiz erişimleri başarıyla engellemiştir. Ayrıca, CQRS, Mediator ve Onion mimarisi yaklaşımlarının birlikte kullanılmasıyla yazılımın modülerliği, yönetilebilirliği ve geliştirilebilirliği artırılmıştır.

3. Sonuçlar

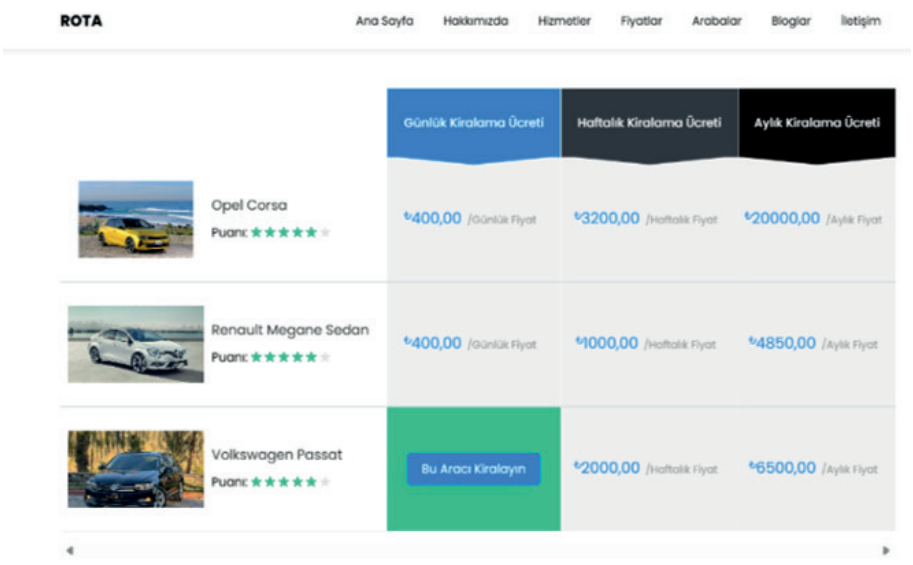
Geliştirilen sistem kadınlar ve engellilerin toplumsal hayata katılımını destekleyen ve fırsat eşitliğini gözeten, uyumlu, kullanıcı dostu ve güvenli bir web platformu olarak tasarlanmıştır. Bu bağlamda kullanıcı deneyimi, güvenlik ve erişilebilirlik açısından yapılan değerlendirmelerde hedeflenen kriterler karşılanmıştır.

3.1.Teknik Çıktılar

Anasayfa: Kullanıcı deneyimini ön planda tutan, sade ve erişilebilir bir yapıya sahip olan anasayfa kadın ve engelli bireylerin sistemi kolaylıkla kullanabilmelerini sağlamaktadır. Geliştirilen Ana sayfa Bölümü tasarımı Şekil 2.'de gösterilmektedir.

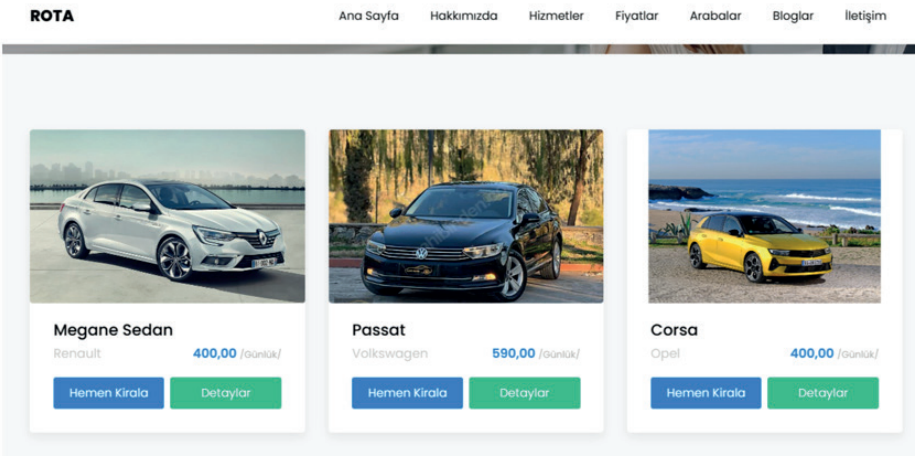
Şekil 2: Anasayfa Bölümü

Fiyatlar Bölümü: Fiyatlar bölümü, araçların güncel kiralama ücretlerini sunmakta; kullanıcıların fiyatları rahatlıkla karşılaştırmasına imkân vermektedir. Geliştirilen Fiyatlar Bölümü tasarımı Şekil 3.'te gösterilmektedir.



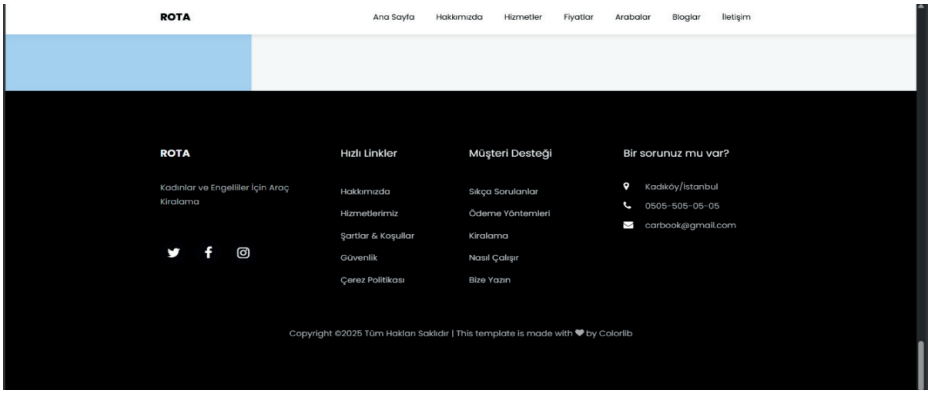
Şekil 3: Fiyatlar Bölümü

Araçlar Bölümü: Kullanıcılara bu bölümde, araçları engelli erişimine uygunluk veya bebek koltuğu gibi özel donanım özelliklerine göre filtreleme yapmalarına olanak tanınır. Böylece, kullanıcılar ihtiyaçlarına uygun araca kısa sürede ulaşabilmektedir. Geliştirilen Araçlar Bölümü tasarımı Şekil 4.'te gösterilmektedir.



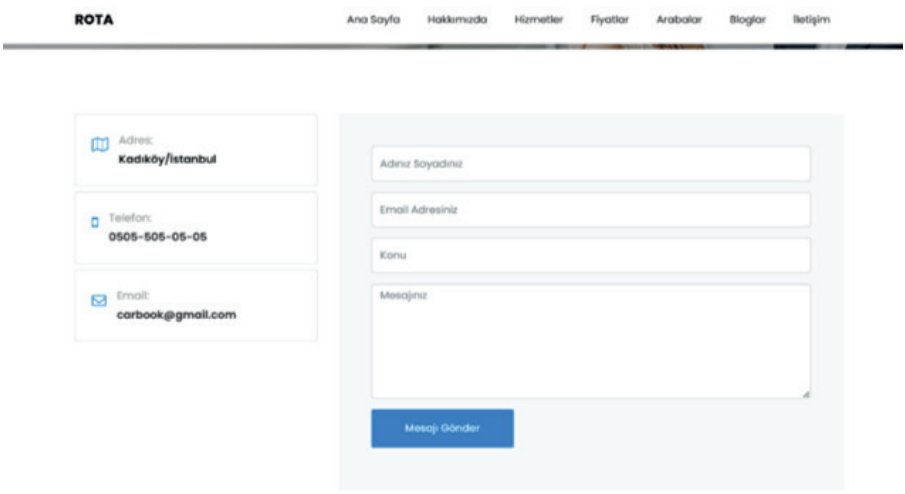
Şekil 4: Araçlar Bölümü

Footer: Sayfanın alt kısmında konumlandırılan bu bölüm, temel bağlantılara ve sosyal medya hesaplarına erişimi kolaylaştırmaktadır. Tüm bileşenler, erişilebilirlik standartlarına doğrultusunda tasarlanarak düzenlenmiştir. Geliştirilen Footer Bölümü tasarımı Şekil 5.'te gösterilmektedir.



Şekil 5: Footer Bölümü

İletişim Bölümü: Platformun iletişim alanı kullanıcıların sistem yöneticileri veya destek ekibiyle bağlantı kurabilmeleri için bir form sunmaktadır. Açık iletişim bilgileri, kullanıcı etkileşimini güçlendiren bir ortam sunmaktadır. Geliştirilen Footer Bölümü tasarımı Şekil 6.'da gösterilmektedir.



Şekil 6: İletişim Bölümü

3.2. Platformun Başlıca Özellikleri

Kullanıcı Dostu Web Arayüzü: Platformun web arayüzü, sade tasarımı ve rehber niteliğindeki yapısıyla kullanıcıların kolayca gezinebileceği bir ya-

pıya sahiptir. Renk, yazı tipi ve buton düzeniyle kullanıcı deneyimini artıracak biçimde tasarlanmıştır.

Kadınlara Yönelik Güvenli Kullanım Özellikleri: Platformda kadın kullanıcıların güvenliğini ön planda tutarak acil yardım butonu ve bebek koltuğu gibi özel araç donanımları sunulmuştur. Bu özellikler, kadınlar için güvenli ve konforlu bir araç kiralama deneyimini desteklemektedir.

Engellilere Uygun Filtreleme Sistemi: Platformda engelli bireylerin ihtiyaçlarına uygun olarak rampa ve geniş iç hacim gibi donanım özelliklerine sahip araçlar sisteme eklenmiş, bu araçlar filtreleme sistemiyle hızlı ve kolay bulunabilir hale getirilmiştir.

Yönetim Paneli ve Veritabanı Yönetimi: Admin paneli aracılığıyla araç, kullanıcı ve rezervasyon işlemleri etkin bir biçimde yönetilmektedir. SQL Server kullanılarak güçlü veritabanı altyapısı oluşturulmuştur. Bu sayede veriler güvenli biçimde saklanmakta; CQRS ve Mediator mimarisi ile de veri erişiminde ve güncellemelerde hem yüksek performans hem de güvenlik sağlanmıştır.

4. Gelecek Yönelimler

Gelecekte, platformun mobil uygulama sürümünün geliştirilmesi, kişiselleştirilmiş hizmetlerin artırılması ve engelli kullanıcılar için özel konum tabanlı acil yardım modüllerinin eklenmesi planlanmaktadır.

Planlanan bu iyileştirmeler, geliştirilen web platformunun kapsayıcılığını ve toplumsal etki alanını genişletecek; aynı zamanda yazılım mimarisi, kullanıcı deneyimi ve sosyal inovasyon gibi alanlarda yeni araştırma ve geliştirme fırsatları sunacaktır.

Başarılı bir uygulama, diğer bölgeler için de örnek teşkil ederek, ülke genelinde benzer platformların geliştirilmesine öncülük edebilir. Bölgesel farklılıkları göz önünde bulundurarak geliştirilecek bir platform, daha geniş bir uygulama potansiyeline sahip olabilir ve diğer bölgeler için de uyarlanabilir bir model sunabilir.

Ayrıca, bu tür yenilikçi çözümlerin hayata geçirilmesi, araç kiralama sektöründe de bir dönüşüm başlatarak, daha kapsayıcı ve duyarlı hizmet anlayışının benimsenmesine katkıda bulunabilir.

Sonuç olarak, bu araştırma, kadınların ve engelli bireylerin ulaşım ihtiyaçlarına yönelik yenilikçi ve özgün bir çözüm önererek, hem akademik literatüre katkıda bulunabilir hem de pratik uygulamalar için yol gösterici olabilir. Özellikle, araç kiralama sektöründe daha kapsayıcı ve duyarlı bir yaklaşımın benimsenmesine yönelik atılacak adımlar için önemli bir zemin oluşturacağı düşünülmektedir.

Destek ve Katkı Beyanı

Bu çalışma, TÜBİTAK tarafından yürütülen 2209-A Üniversite Öğrencileri Araştırma Projeleri Destekleme Programı kapsamında, 1919B012426392 numaralı proje ile desteklenmiştir. Sağladıkları maddi katkı ve araştırma olanakları için TÜBİTAK'a teşekkür ederiz. Ayrıca, Dr. Öğr. Üyesi Doğan Küçük'e (Giresun Üniversitesi, Bilgisayar Mühendisliği Bölümü Öğretim Üyesi) katkısı için teşekkür ederiz.

KAYNAKÇA

Akay, B. (2014). Araç kiralama sektöründe rekabet belirleyicilerinin işletme performansına etkisi: Turizm destinasyonlarında bir araştırma (Doctoral dissertation, Sakarya Üniversitesi, Turkey).

Burnett, J. J., & Bender, B. H. (2001). Assessing the travel-related behaviors of the mobility-disabled consumer. Journal of Travel Research, 40(4), 4–11.

Carver, A., & Veitch, J. (2020). Perceptions and patronage of public transport: Are women different from men? Journal of Transport & Health, 19, 100955.

Cho, S., & Rust, J. (2010). The flat rental puzzle. The Review of Economic Studies, 77(2), 560–594.

Darcy, S., & Daruwalla, P. S. (1999). The trouble with travel: People with disabilities and tourism. Social Alternatives, 18(1), 41–46.

Durmuş, E. (2013). Sexual harassment: University students' perceptions and reactions. Inonu University Journal of the Faculty of Education, 14(1), 15–30.

Hanözü, S., Mollahasanoğlu, S., Esen, E., Şimşek, E., Boztaş, D., Doğan, A., & Başkal, H. (2015). İstanbul'da kamu ulaşım araçlarında gerçekleşen kadına yönelik cinsel şiddetin varlığı, yoğunluğu ve psikolojik etkileri. Marmara Üniversitesi.

Israeli, A. A. (2002). A preliminary investigation of the importance of site accessibility factors for disabled tourists. Journal of Travel Research, 41(1), 101–104.

Kaygısız, Ü., & Bulgan, G. (2015). İnsan hakları çerçevesinde engellilerin seyahat hakkı ve Avrupa Birliğindeki yasal düzenlemeler. Akademik Bakış Dergisi, 49, 98–106.

Keysan, A. Ö., Kaygan, P., & Kaygan, H. (2022). Ankara'da ev işçisi kadınların toplu taşıma deneyimleri: Zorluklar ve stratejiler. Fe Dergi, 14(1), 107–120.

Korstanje, M. (2011). Rent-a-car industry: A case study in Argentina. Tourismos: An International Multidisciplinary Journal of Tourism, 6(1), 271–280.

Nazlı, (2022). Planlı davranış teorisi perspektifinden araç kiralamada tüketici tercihlerinin belirlenmesi: Diyarbakır ili örneği (Yüksek lisans tezi, Mardin Artuklu Üniversitesi).

Önder, H. G. (2020). Kadın duyarlı ulaşım önceliklerinin belirlenmesi ve politika üretimi: Ankara örneği. OPUS International Journal of Society Researches, 15(23), 1993–2010.

Reis, Z. A., Gücükoğlu, B., & Eskici, B. (2014). İşitme ve konuşma engellilerin yaşamlarını kolaylaştırma. In Akademik Bilişim'14 - XVI. Akademik Bilişim Konferansı Bildirileri (pp. 5–7). Mersin Üniversitesi.

Sarıışık, B. E., & Öcalır, E. V. (2022). Kadınların ve erkeklerin farklı seyahat davranışına sahip olması üzerine: Kent içi ulaşımda kadınlar. İdealkent, 16(43), 219–243.

Scheiner, J. (2014). Gendered key events in the life course: Effects on changes in travel mode choice over time. Journal of Transport Geography, 37, 47–60.

Sevginer, C., Bilge, E., Demir, Ö., & Gezer, U. Y. (2011). Sürdürülebilir ulaşım için çözüm önerisi: Taksiye yönelik araç platformu. In 9. Ulaştırma Kongresi Kitabı (pp. 1–10). İstanbul.

Takeda, K., & Card, J. A. (2002). U.S. tour operators and travel agencies: Barriers encountered when providing package tours to people who have difficulty walking. Journal of Travel & Tourism Marketing, 12(1), 47–61. https://doi.org/10.1300/J073v12n01_03

Bölüm 13

MAKİNE ÖĞRENİMİ EĞİTİM PARADİGMALARINDA İSTATİSTİKSEL AKIL YÜRÜTMENİN ROLÜ

Özge Taş[1]

1 Kastamonu Üniversitesi, Çatalzeytin Meslek Yüksek Okulu, Bilgisayar Programcılığı, Öğr.Gör.Özge Taş, ORCID: https://orcid.org/0000-0001-7220-5054

Makine öğrenimi eğitim paradigmalarında istatistiksel akıl yürütmenin rolü, sistemlerin verilerden öğrenmesini ve otonom olarak tahminlerde bulunmasını sağlayan bir disiplin olan makine öğrenimi (ML) çerçevesinde istatistiksel ilkelerin temel entegrasyonunu inceler. İstatistiksel akıl yürütme, model geliştirme, performans değerlendirmesi ve sonuçların yorumlanmasını kolaylaştırarak makine öğreniminin çeşitli yönlerinin temelini oluşturur. Bu alanların birbirine bağımlılığı, ML algoritmalarının artan karmaşıklığı ve büyük veri kümelerine olan önemli bağımlılık göz önüne alındığında özellikle dikkat çekicidir ve bu durum, alana giren uygulayıcılar arasında istatistiksel okuryazarlığın önemi hakkında tartışmalara yol açmıştır (Boulesteix, A. L., & Schmid, M. , 2014).

Tarihsel olarak, istatistiksel metodolojiler birçok makine öğrenme tekniğinin temelini oluşturmuştur; olasılık teorisi ve Bayes istatistiği gibi temel kavramlar, algoritmaların tasarım ve uygulanmasında kritik bileşenler olarak görev yapmıştır (Sun, K., & Wang, R. ,2024).

Makine öğreniminin evrimi, istatistiksel uygulamaların kapsamını genişletmekle kalmamış, aynı zamanda yüksek kaliteli veri gerekliliği ve model karmaşıklığı nedeniyle aşırı uyum riski gibi zorlukları da beraberinde getirmiştir. Bu zorluklar, sağlam ve yorumlanabilir sonuçlar elde etmek için istatistiksel akıl yürütmenin eğitim paradigmalarına (yani denetimli, denetimsiz ve pekiştirmeli öğrenmeye) entegre edilmesinin önemini vurgulamaktadır (Dangeti, P. ,2017),(Solomonoff, R. J. ,2006).

Kullanıcı dostu makine öğrenimi araçlarının yaygınlaşmasına rağmen, yeni uygulayıcılar arasında istatistiksel bilgiye verilen önemin azalması olasılığına ilişkin endişeler devam etmektedir. Bu durum, özellikle çeşitli sektörlerdeki yüksek riskli uygulamalarda, model doğruluğunu doğrulamak ve yorumlanabilirliği artırmak için sağlam bir istatistiksel temelin gerekliliği üzerine giderek artan bir tartışmaya yol açmıştır. (Kirch, C., et. al., 2025). Sugiyama, M. (2015).

Makale, istatistiksel muhakemenin yalnızca model doğruluğu için değil, aynı zamanda disiplinler arası işbirliğini teşvik etmek ve veriye dayalı karar alma süreçlerini iyileştirmek için de hayati önem taşıdığını vurgulamaktadır (Cady, F. 2024).

Makine öğrenimi gelişmeye devam ettikçe, istatistiksel akıl yürütme ve algoritma geliştirme arasındaki etkileşim hayati önem taşımaktadır. Gelecekteki gelişmelerin bu disiplinleri daha da bütünleştirmesi ve sağlık, finans ve mühendislik gibi alanlarda yenilikçi uygulamaların önünü açması beklenmektedir. Model eğitimi ve dağıtımıyla ilgili zorlukların ve sınırlamaların ele alınması, makine öğrenimi çözümlerinin gerçek dünya bağlamlarında güvenilirliğini ve etkinliğini sağlamak için çok önemlidir (Cady, F. 2024)(Mello, R. F., & Ponti, M. A. ,2018).

Tarihsel Bağlam

İstatistik ve makine öğreniminin kesişim noktası, her iki disiplinin metodolojilerinin ve teorilerinin evrimini yansıtan zengin bir tarihsel bağlama sahiptir. Tarihsel olarak, istatistik birçok makine öğrenimi tekniğinin temeli olmuştur ve ilk gelişmeler doğrusal ve lojistik regresyon gibi geleneksel istatistiksel yöntemlere dayanmaktadır. Bu yöntemler, hesaplama teknolojilerindeki gelişmelerle birlikte kademeli olarak evrim geçirmiş ve makine öğrenimi alanında daha karmaşık algoritmaların ortaya çıkmasını kolaylaştırmıştır (Boulesteix, A. L., & Schmid, M. (2014). (Dangeti, P. 2017)

Bu evrimin en önemli anlarından biri, makine öğreniminin büyük veri kümelerini işleme potansiyelinin farkına varılmasıydı; bu da hesaplama gücündeki ve veri depolama kapasitelerindeki gelişmeler sayesinde giderek daha mümkün hale geldi. Makine öğrenimi ivme kazandıkça, geleneksel istatistiklerde vurgulanan veri kalitesinden veri miktarına doğru bir odak kayması yaşandı ve bu da tahmin modellerini etkili bir şekilde eğitmek için büyük veri kümelerine olan ihtiyacı artırdı (Bian, J. , 2019).

Bu değişim, istatistikçiler ve makine öğrenimi uzmanları arasında, her iki alanın güçlü yönlerini birleştirmeyi amaçlayan, giderek artan bir iş birliğini teşvik etti (Bian, J. , 2019).(Sun, K., & Wang, R. ,2024).

On yıllar boyunca, istatistikçiler tarafından kullanılan metodolojiler, makine öğreniminin gelişimini derinden etkilemiştir; zira birçok modern algoritma, istatistik teorisinden kavramlar ödünç almaktadır. Örneğin, olasılık teorisinin prensipleri, özellikle Bayes istatistiği, çok sayıda makine öğrenimi algoritmasının ayrılmaz bir parçası haline gelmiştir. Bu bağlantı, iki alanın sürekli evrimini göstermekte ve sağlam makine öğrenimi modelleri geliştirmede güçlü bir istatistiksel temelin önemini vurgulamaktadır (Dangeti, P. , 2017). Dahası, Python'ın scikit-learn'ü gibi kullanıcı dostu makine öğrenimi kütüphanelerinin ve çerçevelerinin yaygınlaşması, alana yeni giren uygulayıcılar için derin istatistiksel bilginin daha az kritik olduğu algısına katkıda bulunmuştur. Bu yanlış anlama, makine öğrenimi uygulamalarının geçerliliğini ve yorumlanabilirliğini sağlamak için istatistiksel prensiplerin kapsamlı bir şekilde anlaşılması gerekliliği konusunda tartışmalara yol açmıştır. (Kirch, C., et. al., 2025).(Sun, K., & Wang, R. , 2024).

Bu disiplinler gelişmeye devam ettikçe, disiplinler arası işbirliğinin artması ve veri analizi ile tahmin modellemesine daha bütünleşik bir yaklaşımın önünün açılması beklenmektedir (Bian, J. , 2019).(Lafferty, J. ve Wasserman, L. ,2006).

Makine Öğrenmesinde İstatistiksel Kavramlar

İstatistiksel kavramlar, makine öğrenimi modellerinin geliştirilmesi ve anlaşılmasında çok önemli bir rol oynar. Bu kavramlar, verileri analiz etmek,

tahminlerde bulunmak ve makine öğrenimi algoritmalarının sonuçlarını yorumlamak için gerekli çerçeveyi sağlar.

Olasılık Teorisi

Olasılık teorisi, makine öğreniminde istatistiksel çıkarımın temelini oluşturur. Belirsizliği nicelleştirmeye yardımcı olur ve verilere dayalı tahminler yapmada önemli bir rol oynar.

Temel Kavramlar

Olasılık : Bir olayın gerçekleşme ihtimalinin sayısal bir ölçüsü olup, 0 (imkansız) ile 1 (kesin) arasında değişir (Dangeti, P. , 2017).

Makine öğreniminde bu, genellikle belirli özellikler göz önüne alındığında bir sonucun olasılığını değerlendiren koşullu olasılık biçimini alır.

Bayes Teoremi

Bayes teoremi, olasılık teorisinin temel bir bileşenidir ve yeni kanıtlara (sonraki olasılıklar) dayanarak ilk inançların (önsel olasılıklar) güncellenmesine olanak tanır.

$$P(A|B)=\frac{P(A|B)P(A)}{P(B)} \quad (1)$$

Bu teorem, genetik testler gibi uygulamalar için çok önemlidir; burada, önceden edinilen bilgi ve yeni verilere dayanarak bir bireyin belirli bir geni taşıma olasılığını belirlemeye yardımcı olabilir.

İstatistik Türleri

Makine öğrenimi genellikle iki ana istatistik türünü kullanır:

Tanımlayıcı İstatistikler : Bu dal, büyük veri kümelerini basitleştirir ve düzenler, böylece anlaşılmalarını ve analiz edilmelerini kolaylaştırır. Ortalama, medyan ve aralık gibi ölçümler bu kategoriye girer (Naidu, G., Zuva, T., & Sibanda, E. M. , 2023)

Çıkarımsal İstatistik : Bu yaklaşım, daha büyük bir popülasyon hakkında sonuçlar çıkarmak için daha küçük bir veri kümesi kullanır ve tahminler ile hipotez testlerine olanak tanır (Naidu, G., Zuva, T., & Sibanda, E. M. , 2023)

Makine öğrenimi modellerini değerlendirmek için çeşitli istatistiksel ölçütler hayati öneme sahiptir:

P değeri : P değeri, sıfır hipotezinin doğru olduğu varsayımı altında, gözlemlenen kadar uç bir test istatistiğinin gözlemlenme olasılığını gösterir. Hipotez testinde çok önemli bir rol oynar ve araştırmacıların bulgularının istatistiksel olarak anlamlı olup olmadığını belirlemelerine yardımcı olur (Bontempi, G., & Ben Taieb, S. , 2021).

Güven Aralığı : Bu aralıklar, bir parametrenin belirli bir güven düzeyinde (örneğin, %95) düşme olasılığının olduğu aralığı tahmin eder. Makine öğrenimi modellerinden elde edilen tahminlerin doğruluğu hakkında fikir verirler (Bontempi, G., & Ben Taieb, S. , 2021).

R-kare : Bu ölçü, bağımlı değişkendeki varyansın bağımsız değişkenlerden tahmin edilebilen oranını açıklar ve modelin performansına istatistiksel bir bakış açısı sunar (Naidu, G., Zuva, T., & Sibanda, E. M. ,2023).

İstatistiksel Öğrenme Teorisi

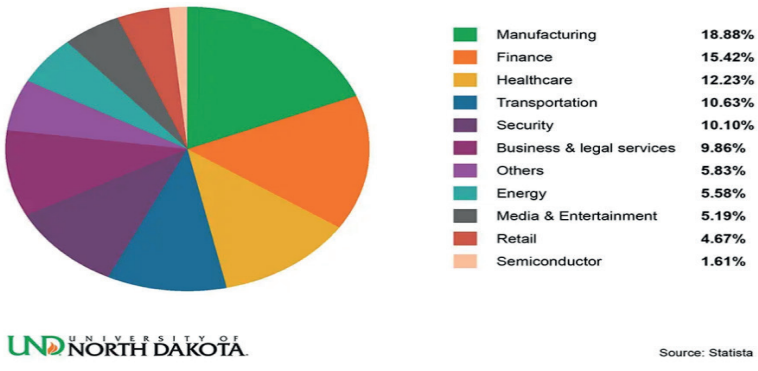
İstatistiksel öğrenme teorisi, makine öğrenmesi algoritmalarının verilerden nasıl öğrendiğini ve yeni durumlara nasıl genelleme yaptığını anlamak için titiz bir çerçeve sunar. İstatistiksel yöntemler ve makine öğrenmesi arasındaki etkileşimi vurgulayarak, bu iki alanın entegrasyonunun daha güçlü, verimli ve yorumlanabilir modellerin geliştirilmesini nasıl artırabileceğini ortaya koyar (Sun, K., & Wang, R. ,2024).Veri bilimciler, istatistiksel akıl yürütmeyi kullanarak veri yapılarını daha iyi anlayabilir, model doğruluğunu artırabilir ve karmaşık veri kümelerinden anlamlı sonuçlar çıkarabilirler.

Makine Öğrenimi Eğitimine Genel Bakış

Makine öğrenimi, bilgisayarların verilerden öğrenmesini ve özerk kararlar almasını sağlayan algoritmalar oluşturmaya adanmış, bilgisayar biliminin önemli bir dalıdır. Açık talimatların verildiği geleneksel programlamanın aksine, makine öğrenimi modelleri, kapsamlı veri kümelerini analiz ederek kalıpları belirler ve tahminlerde bulunur (Sugiyama, M. (2015).

Sektörlere göre makine öğrenimi pazarı büyüklüğü ve payı aşağıda verilmiştir.

Şekil 2. *Sektörlere göre makine öğrenimi pazarı büyüklüğü (https://und.edu/blog/machine-learning-vs-statistics.html)*



Bu modellerin eğitimi, her biri belirli veri türlerine ve hedeflere göre uyarlanmış çeşitli paradigmalar aracılığıyla gerçekleşir.

ML Eğitim Türleri

Denetimli Öğrenme

Denetimli öğrenme, her girdinin karşılık gelen bir çıktıyla ilişkilendirildiği etiketli veri kümeleri üzerinde modellerin eğitilmesini içerir. Bu yöntem, özellikle e-postalarda istenmeyen e-posta tespiti gibi uygulamalar da dahil olmak üzere sınıflandırma ve regresyon gibi görevler için etkilidir (Boulesteix, A. L., & Schmid, M. , 2014). (Oğuzlar, A. ,2003).

Yüksek kaliteli etiketli verilere olan bağımlılık önemli bir zorluk teşkil etmektedir, çünkü denetimli öğrenmenin başarısı büyük ölçüde bu verilerin mevcudiyetine ve doğruluğuna bağlıdır (Dangeti, P. ,2017).

Denetimsiz Öğrenme

Buna karşılık, denetimsiz öğrenme, açık etiketler içermeyen veri kümeleri üzerinde çalışır. Burada model, veriler içindeki gizli yapıları veya kalıpları ortaya çıkarmayı amaçlar. Yaygın uygulamalar arasında kümeleme (benzer veri noktalarının gruplandırılması) ve değişkenler arasındaki ilişkileri belirleyen ilişkilendirme yer alır (Boulesteix, A. L., & Schmid, M. ,2014).

Denetimsiz öğrenme, özellikle model performansının değerlendirilmesinde kendine özgü zorluklar sunar; çünkü denetimli öğrenmede kullanılan geleneksel ölçütler doğrudan uygulanamaz. (García, S., Luengo, J., & Herrera, F. ,2015).

Takviyeli Öğrenme

Takviyeli öğrenme (RL), modellerin çevreleriyle etkileşim yoluyla öğrendiği ve ödül veya ceza şeklinde geri bildirim aldığı farklı bir paradigmayı temsil eder (Taş, Ö. (2024). Bu yaklaşım, modelin zaman içinde kümülatif ödülleri en üst düzeye çıkarmak için en uygun politikayı öğrenmesi gereken robotik ve oyun oynama gibi dinamik uygulamalar için uygundur (Sugiyama, M. ,2015).

Takviyeli öğrenmenin yinelemeli yapısı, yerel optimumlardan kaçınmak için sağlam keşif stratejilerine duyulan ihtiyaç da dahil olmak üzere ek karmaşıklıklar getirir.

İstatistiksel Akıl Yürütmenin Önemi

İstatistiksel akıl yürütme, makine öğrenimindeki tüm eğitim paradigmalarının ayrılmaz bir parçasıdır. İstatistik alanında sağlam bir temel, önyargı, varyans ve farklı modellerin altında yatan varsayımlar gibi sorunları ele alma yeteneğini artırır. Bu anlayış, model performansını değerlendirmek ve makine öğrenimi çözümlerinin sağlamlığını iyileştirmek için gereklidir

(Solomonoff, R. J. ,2006).

Uygulayıcılar yüksek boyutlu verilerin karmaşıklığıyla başa çıkmaya ve tahmin doğruluğunu artırmaya çalışırken, istatistiksel yöntemlerin uygulanması giderek daha önemli hale geliyor.

Model Eğitiminde Karşılaşılan Zorluklar

Etkili makine öğrenimi modelleri geliştirmek birçok zorlukla doludur. Bunların başında kaliteli eğitim verilerinin eksikliği ve yanlış eğitim yöntemlerinin kullanılması riski gelir. Bu zorlukların üstesinden gelmek, model geliştirme sürecini yönlendiren istatistiksel prensiplerin ve makine öğrenimi uygulamalarının inceliklerinin kapsamlı bir şekilde anlaşılmasını gerektirir (Dangeti, P. (2017).(Solomonoff, R. J. ,2006).

İstatistiksel akıl yürütmeyi eğitim sürecine entegre ederek, uygulayıcılar verilerden anlamlı içgörüler elde etme ve daha güvenilir makine öğrenimi sistemleri oluşturma yeteneklerini geliştirebilirler.

İstatistiksel Akıl Yürütme Uygulamaları

İstatistiksel akıl yürütme, özellikle makine öğrenimi ve veri analizi olmak üzere birçok alanda çeşitli uygulamalarda kritik bir rol oynamaktadır. Uygulayıcılar, istatistiksel yöntemler ve prensipler kullanarak verilerden geçerli sonuçlar çıkarabilir ve kanıtlara dayalı olarak bilinçli kararlar verebilirler.

Veri Analizi ve Yorumlama

Veri analizi alanında, istatistiksel akıl yürütme, veri kümelerini anlamak ve yorumlamak için temeldir. Tanımlayıcı istatistikler, hipotez testleri ve çıkarımsal istatistikler gibi teknikler, analistlerin veriler içindeki kalıpları, eğilimleri ve ilişkileri belirlemelerini sağlayarak, işletme, ekonomi, sağlık ve sosyal bilimler gibi alanlarda kanıta dayalı karar vermeyi destekler (Paleyes, A., Urma, R. G., & Lawrence, N. D. (2022).

Araştırma ve Bilimsel Sorgulama

Araştırmacılar, deneysel verileri analiz etmek ve bilimsel ilerlemelere katkıda bulunan bulguları yayınlamak için istatistiksel akıl yürütmeyi kullanırlar. Bu, hipotezleri doğrulamak, popülasyon parametrelerini tahmin etmek ve regresyon analizi ve güven aralıkları gibi çeşitli yöntemlerle sonuçların güvenilirliğini değerlendirmek için istatistiksel tekniklerin uygulanmasını içerir (Paleyes, A., Urma, R. G., & Lawrence, N. D. ,2022).

İstatistiksel bir bakış açısıyla verileri eleştirel bir şekilde değerlendirme ve yorumlama yeteneği, herhangi bir bilimsel alanda bilgi birikimini ilerletmek için elzemdir.

Makine Öğrenmesi Modeli Geliştirme

İstatistiksel akıl yürütme, makine öğrenimi modellerinin geliştirilmesinde de çok önemlidir. Veri temizleme, normalleştirme ve dönüştürme gibi ön işleme teknikleri, istatistiksel prensiplere dayanır ve modellerin doğruluğunu ve verimliliğini artırır. Özellik seçimi yöntemlerini uygulayarak, uygulayıcılar karmaşıklığı azaltarak ve aşırı uyumun önüne geçerek model performansını artırabilirler (Zaker Esteghamati, M., Bean, B., Burton, H. V., & Naser, M. Z. (2025):(Wagstaff, K. ,2012).

Olasılık dağılımlarını ve temel istatistiksel varsayımları anlamak, uygun modelleri ve değerlendirme ölçütlerini seçmeye yardımcı olur.

İşletmelerde Karar Verme

İş dünyasında istatistiksel akıl yürütme, pazar trendlerini, tüketici davranışlarını ve operasyonel performansı analiz ederek stratejik karar alma süreçlerini kolaylaştırır. İşletmeler, satışları tahmin etmek, pazarlama etkinliğini değerlendirmek ve kaynak tahsisini optimize etmek için istatistiksel araçlardan yararlanarak daha iyi finansal sonuçlar elde ederler (Paleyes, A., Urma, R. G., & Lawrence, N. D. ,2022).

İstatistiksel veriler, işletme kararlarını gerekçelendirmek ve etkilerini değerlendirmek için bir temel oluşturur.

Eğitim ve Öğretim

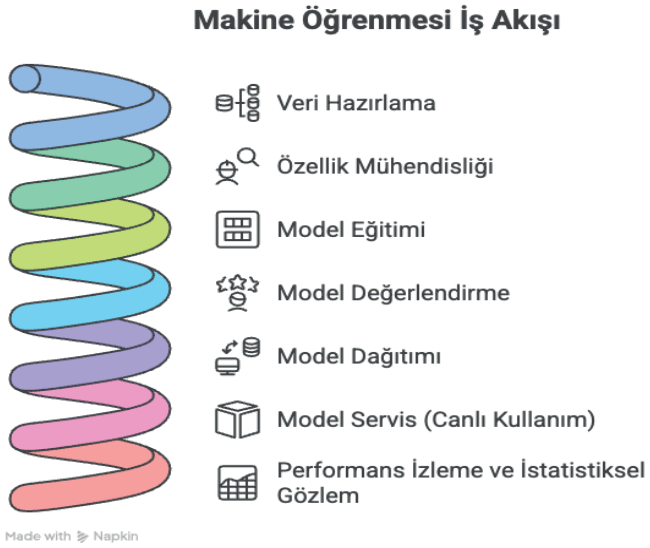
İstatistiksel akıl yürütme, özellikle veri bilimi ve analitiği alanlarında eğitim ve öğretim programlarının önemli bir bileşenidir. İstatistiksel akıl yürütme dersleri, öğrencilere matematiksel temelleri anlama, istatistiksel yazılımları kullanma ve çeşitli istatistiksel teknikleri gerçek dünya problemlerine uygulama gibi temel beceriler kazandırır. Bu temel, veri odaklı alanlarda kariyer yapmak isteyen herkes için hayati önem taşır (Paleyes, A., Urma, R. G., & Lawrence, N. D. ,2022).

Zorluklar ve Sınırlamalar

Makine öğrenimi (ML) modellerinin gerçek dünya uygulamalarında kullanımı, performanslarını ve güvenilirliklerini önemli ölçüde etkileyebilecek zorluklar ve sınırlamalarla doludur. Bu zorlukları anlamak, eğitim veri kümelerinin çok ötesinde genelleme yapabilen modeller geliştirmek için çok önemlidir.

Bu çalışmada önerilen kavramsal çerçeve, makine öğrenmesi yaşam döngüsünü istatistiksel karar teorisi ve süreç izleme yaklaşımları ile bütünleştiren döngüsel bir yapı olarak ele almaktadır. Şekil 1'te sunulan mimari, klasik veri hazırlama–model eğitimi hattını, istatistiksel performans izleme mekanizmalarıyla desteklenen sürekli bir geri besleme sistemi olarak yeniden tanımlamaktadır.

Şekil 1. *Makine öğrenmesi sürekli izleme mimarisi.*



Veri Kalitesi ve Erişilebilirliği

Başlıca zorluklardan biri, yüksek kaliteli telemetri verilerinin eksikliği ve veri toplama için standartlaştırılmış yöntemlerin bulunmamasıdır (Rashidi, H. H., Albahra, S., Robertson, S., Tran, N. K., & Hu, B. ,2023).

Yetersiz veya temsili olmayan eğitim verileri, modellerin daha önce görülmemiş verilere maruz kaldığında iyi performans göstermemesine yol açabilir. Bu sorun, modelin eğitim verilerinden alakasız kalıpları yakalaması ve dolayısıyla genelleme yapamaması durumunda ortaya çıkan aşırı uyum (overfitting) ile daha da kötüleşir (Kroese, D. P., Botev, Z., Taimre, T., & Vaisman, R. (2019):(Kastner, C. ,2025).

Modelin karmaşıklığı arttıkça aşırı uyum riski de artar; bu durum daha yüksek varyansa ve daha düşük tahmin tutarlılığına yol açar.

Performans Ölçütleri ve Sapma-Varyans Dengesi

Bir diğer önemli zorluk, makine öğrenimi modellerini değerlendirmek için kullanılan performans ölçütlerinin seçimidir. Her ölçütün kendi sınırlamaları vardır ve tek başına değerlendirilmemelidir. Seçilen performans metriklerinin çalışmanın hedefleriyle uyumlu olması, bir modelin yetenekleri hakkında anlamlı bilgiler elde etmek için çok önemlidir (Brownlee, J. ,2018).

Önyargı-varyans dengesi, değerlendirme sürecini daha da karmaşık hale getirir. Bu denge, model karmaşıklığı ile genelleme yeteneği arasındaki gerilimi vurgular; düşük önyargı genellikle artan varyans pahasına elde edilir ve

bu da aşırı uyumlanmaya yol açar (Kastner, C. ,2025).

Bu dezavantajı hafifletmek için genellikle düzenleme teknikleri kullanılır, ancak bunlar model geliştirme sürecine ek karmaşıklık getirir (Brownlee, J. ,2018).

Model Kritikliği ve İzleme Gereksinimleri

Modelin uygulama alanının kritikliği, izleme gereksinimlerini de belirler. Finansal veya tıbbi uygulamalar gibi yüksek riskli ortamlarda, modelin düşük performansından kaynaklanan sorunları hızlı bir şekilde tespit etmek ve bunlara yanıt vermek için titiz bir izleme gereklidir (Sugiyama, M. ,2015).

Öte yandan, daha az kritik modeller, daha düşük izleme sıklığına ve ayrıntı düzeyine izin verebilir; bu da zaman içinde model doğruluğunu koruma konusunda rehavete yol açabilir.

Dağıtım Mimarisi

Model dağıtımının mimarisi, bir başka zorluk katmanı daha ortaya koymaktadır. Gerçek zamanlı karar verme ve toplu işleme gibi farklı dağıtım senaryoları, izleme sistemlerinin tasarımını etkiler (Sugiyama, M. ,2015).

Her mimari, modellerin kendi bağlamlarında etkili kalmasını sağlamak için özel yaklaşımlar gerektirir.

Genelleştirilebilirlik ve Açıklanabilirlik Zorlukları

Genelleştirilebilirlik, modellerin pratikte etkili bir şekilde performans gösterebilmeleri için eğitim verilerinin ötesine genelleme yapabilmeleri gerektiğinden, merkezi bir konu olmaya devam etmektedir (Nisbet, R., Miner, G. D., & McCormick, K. ,2024).

Ayrıca, açıklanabilirlik, özellikle yapı mühendisliği gibi modelin karar verme sürecini anlamanın güven ve benimsenme açısından kritik önem taşıdığı alanlarda, makine öğreniminde giderek artan bir endişe kaynağıdır. (Nisbet, R., Miner, G. D., & McCormick, K. ,2024).

Bu araştırma, gerçek dünya uygulamalarına uygun ve sağlam makine öğrenimi çözümleri geliştirmek için geliştirilmiş eğitim teknikleri ve doğrulama yöntemleri aracılığıyla bu sorunların ele alınmasının önemini tartışmaktadır. Bu zorlukların farkına vararak ve stratejik yaklaşımlar uygulayarak, uygulayıcılar çeşitli uygulamalarda makine öğrenimi modellerinin güvenilirliğini ve etkinliğini artırabilirler.

Geleceğe Yönelik Yönelimler

Makine öğrenimi (ML) gelişmeye devam ettikçe, istatistiksel akıl yürütmenin entegrasyonunun gelecekteki gelişmeleri şekillendirmede çok önemli bir rol oynaması bekleniyor. Kuantum hesaplama ve denetimsiz öğrenme gibi

alanlardaki devam eden gelişmeler, ML modellerinin yeteneklerini artırarak, daha karmaşık görevleri ve veri kümelerini daha yüksek verimlilikle ele almalarını sağlayacak (Kirch, C., et. al., 2025).

Bu evrim, daha önce ulaşılamayan dönüştürücü çözümler sunarak, sağlık, finans ve eğitim de dahil olmak üzere çeşitli sektörlerde insan-bilgisayar etkileşimini yeniden tanımlayacak gibi görünüyor (Kirch, C., et. al., 2025).

Yapı mühendisliği bağlamında, son araştırmalar, istatistiksel analizleri makine öğrenme teknikleriyle birleştiren veri odaklı vekil modeller geliştirmenin önemini vurgulamıştır. Örneğin, Esteghamati ve Flint (2021), orta katlı betonarme (RC) çerçeve binaların bütünsel performans değerlendirmelerini inceleyerek, erken tasarım aşamalarında makine öğrenme yöntemlerini tamamlayacak istatistiksel yaklaşımların potansiyelini ortaya koymuştur (Esteghamati, M. Z., & Flint, M. M. (2021) (Nisbet, R., Miner, G. D., & McCormick, K. (2024). Bu entegrasyon, daha sağlam ve yorumlanabilir modellere yol açarak mühendislik uygulamalarındaki karar alma sürecini iyileştirebilir. Dahası, disiplinler arası işbirliği arttıkça, çalışmaların verilerden daha derin bilgiler elde etmek için hem istatistiksel hem de makine öğrenimi yöntemlerini giderek daha fazla kullanacağı öngörülmektedir. Bu, özellikle Naser ve meslektaşlarının yapısal elemanlarda yangın direncini tahmin etmek için açıklanabilir makine öğrenimi modellerine odaklanan çalışmalarında görüldüğü gibi, hem tahmin gücü hem de altta yatan ilişkilerin anlaşılmasını gerektiren alanlarda önemlidir (Esteghamati, M. Z., & Flint, M. M. ,2021).

İstatistiksel analiz ve makine öğrenimi arasındaki daha net ayrımlara duyulan ihtiyaç, hayati önem taşımaya devam edecektir. Araştırmacılar, bu disiplinlerin güçlü yönlerinden etkili bir şekilde yararlanmak için aralarındaki etkileşimi araştırmaya devam etmelidir. Amaçlardaki farklılıkları anlamak—istatistiksel yöntemler genellikle veri ilişkilerini açıklamayı hedeflerken, makine öğrenimi tahmin doğruluğuna odaklanır daha incelikli uygulamaları teşvik edebilir ve metodolojileri geliştirebilir (Sugiyama, M. ,2015) (Vapnik, V. N. ,1999).

Sonuç

Makine öğrenmesi eğitim paradigmalarında istatistiksel akıl yürütmenin rolü, yalnızca teknik bir gereklilik değil, aynı zamanda disiplinler arası bilgi üretimi ve güvenilir karar alma süreçleri için vazgeçilmez bir bileşendir. İstatistiksel ilkeler, model geliştirme, performans değerlendirme ve yorumlama aşamalarında sağlam bir temel oluşturarak, makine öğrenimi uygulamalarının doğruluğunu ve açıklanabilirliğini artırır. Bu entegrasyon, özellikle yüksek riskli alanlarda hatalı tahminlerin doğurabileceği ciddi sonuçlar göz önüne alındığında, daha da kritik hale gelmektedir.

Makine öğrenimi ve istatistik arasındaki tarihsel bağ, günümüzde büyük

veri ve karmaşık algoritmaların ortaya çıkışıyla daha güçlü bir iş birliğine dönüşmüştür. Ancak kullanıcı dostu araçların yaygınlaşması, istatistiksel okuryazarlığın önemini gölgede bırakma riski taşımaktadır. Bu nedenle, eğitim programlarında istatistiksel akıl yürütmenin vurgulanması, uygulayıcıların önyargı, varyans ve aşırı uyum gibi temel sorunları yönetebilme becerisini geliştirecektir.

Geleceğe bakıldığında, kuantum hesaplama, açıklanabilir yapay zeka ve disiplinler arası entegrasyon gibi gelişmeler, istatistiksel yöntemlerin makine öğrenimiyle daha derin bir şekilde bütünleşmesini gerektirecektir. Bu bütünleşme, yalnızca daha güçlü ve güvenilir modellerin geliştirilmesini değil, aynı zamanda sağlık, finans ve mühendislik gibi kritik alanlarda etik ve sorumlu yapay zeka uygulamalarının yaygınlaşmasını da sağlayacaktır.

Sonuç olarak, istatistiksel akıl yürütme, makine öğrenimi ekosisteminde bir tamamlayıcı değil, bir zorunluluktur. Bu anlayış, hem akademik hem de endüstriyel uygulamalarda daha sağlam, yorumlanabilir ve güvenilir çözümler üretmenin anahtarıdır.

Kaynaklar

Bian, J. (2019). Statistical thinking, machine learning. Journal of clinical epidemiology, 116, 136.

Bontempi, G., & Ben Taieb, S. (2021). Statistical foundations of machine learning. Brussels, Universite Libre de Bruxelles.

Boulesteix, A. L., & Schmid, M. (2014). Machine learning versus statistical modeling. Biometrical Journal, 56(4), 588-593.

Brownlee, J. (2018). Statistical methods for machine learning: Discover how to transform data into knowledge with Python. Machine Learning Mastery.

Cady, F. (2024). The data science handbook. John Wiley & Sons.

Dangeti, P. (2017). Statistics for machine learning. Packt Publishing Ltd.

García, S., Luengo, J., & Herrera, F. (2015). Data preprocessing in data mining (Vol. 72, pp. 59-139). Cham, Switzerland: Springer International Publishing.

Kastner, C. (2025). Machine learning in production: from models to products. MIT Press.

Kroese, D. P., Botev, Z., Taimre, T., & Vaisman, R. (2019). Data science and machine learning: mathematical and statistical methods. Chapman and Hall/CRC.

Lafferty, J. ve Wasserman, L. (2006). İstatistiksel makine öğreniminde zorluklar. Statistica Sinica, 16(2), 307.

Mello, R. F., & Ponti, M. A. (2018). Machine learning: a practical approach on the statistical learning theory. Springer.

Naidu, G., Zuva, T., & Sibanda, E. M. (2023, April). A review of evaluation metrics in machine learning algorithms. In Computer science on-line conference (pp. 15-25). Cham: Springer International Publishing.

Nisbet, R., Miner, G. D., & McCormick, K. (2024). Handbook of Statistical Analysis: AI and ML Applications. Elsevier.

Oğuzlar, A. (2003). Veri ön işleme. Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, (21).

Taş, Ö. (2024). Yapay Zeka Araçları ile Veri Analizine Genel Bir Bakış.

Taş, Ö. (2024). Comparison of The Performances of Clustering and Dimensionality Reduction Approaches in Collaborative Filtering. Advances in Artificial Intelligence Research, 4(2), 96-110.

Paleyes, A., Urma, R. G., & Lawrence, N. D. (2022). Challenges in deploying machine learning: a survey of case studies. ACM computing surveys, 55(6), 1-29.

Rashidi, H. H., Albahra, S., Robertson, S., Tran, N. K., & Hu, B. (2023). Common statistical concepts in the supervised Machine Learning arena. Frontiers in Oncology, 13, 1130229.

Solomonoff, R. J. (2006). Machine learning-past and future. Dartmouth, NH, July.

Sugiyama, M. (2015). Introduction to statistical machine learning. Morgan Kaufmann.

Sugiyama, M. (2015). Introduction to statistical machine learning. Morgan Kaufmann.

Sun, K., & Wang, R. (2024). Differential contributions of machine learning and statistical analysis to language and cognitive sciences. arXiv preprint arXiv:2404.14052.

Sun, K., & Wang, R. (2024). Differential contributions of machine learning and statistical analysis to language and cognitive sciences. arXiv preprint arXiv:2404.14052.

Vapnik, V. N. (1999). An overview of statistical learning theory. IEEE transactions on neural networks, 10(5), 988-999.

Wagstaff, K. (2012). Machine learning that matters. arXiv preprint arXiv:1206.4656.

Zaker Esteghamati, M., Bean, B., Burton, H. V., & Naser, M. Z. (2025). Beyond development: Challenges in deploying machine-learning models for structural engineering applications. Journal of Structural Engineering, 151(6), 04025059.

(https://und.edu/blog/machine-learning-vs-statistics.html)