Bilgisayar Mühendisliğinde Güncel Konular:

Kuram, Teknoloji ve Uygulamalar

Editör:
Dr. Öğr. Üyesi Harun AKBULUT

DUVAR

BİLGİSAYAR MÜHENDİSLİĞİNDE GÜNCEL KONULAR:

KURAM, TEKNOLOJİ VE UYGULAMALAR

Editör

Dr. Öğr. Üyesi Harun AKBULUT

Bilgisayar Mühendisliğinde Güncel Konular:
Kuram, Teknoloji ve Uygulamalar
Editör: Dr. Öğr. Üyesi Harun AKBULUT

Genel Yayın Yönetmeni: Berkan Balpetek
Kapak ve SayfaTasarımı Duvar DESIGN
Basım Tarihi: Eylül 2026
Yayıncı Sertifika No: 49837
E-ISBN: 978-625-8936-80-3

853 Sokak No:13 P.10 Kemeraltı-Konak/İzmir
Tel: 0 232 484 88 68

www.duvaryayinlari.com
duvarkitabevi@gmail.com

IÇINDEKILER

1. Bölüm

Histopatolojik Görüntülerde Çekirdek Segmentasyonu: Yöntemler, Veri Setleri, Değerlendirme Ölçütleri ve Güncel Yaklaşımlar

Abdulkadir GÜLŞEN[1]

Özet

Dijital patolojinin yaygınlaşmasıyla birlikte histopatolojik görüntülerin bilgisayar destekli analizi; hastalıkların tanısı, derecelendirilmesi, prognozun belirlenmesi ve nicel biyobelirteçlerin geliştirilmesi açısından giderek daha önemli hale gelmektedir. Bu süreçte hücre çekirdeklerinin doğru biçimde belirlenmesi ve sınırlarının ayrıştırılması; hücresel yoğunluk, morfoloji, fenotip ve mekânsal organizasyon gibi nicel özelliklerin güvenilir biçimde elde edilmesi için temel bir aşamadır. Bununla birlikte çekirdeklerin birbirine temas etmesi veya üst üste gelmesi, boyut ve şekil bakımından yüksek morfolojik değişkenlik göstermesi, boyama (stain) ve görüntüleme koşullarındaki farklılıklar, doku ve organ heterojenliği ile yüksek kaliteli piksel düzeyi anotasyonların sınırlı olması, çekirdek segmentasyonunu zorlu bir görüntü analizi problemi haline getirmektedir.

Bu bölümde histopatolojik çekirdek segmentasyonu, geleneksel görüntü işleme yaklaşımlarından derin öğrenme, transformer ve güncel temel (foundation) model tabanlı yöntemlere uzanan gelişim çizgisi içerisinde bütüncül olarak ele alınmaktadır. Anlamsal segmentasyon, örnek tabanlı segmentasyon ve panoptik segmentasyon kavramları açıklanmakta; farklı yöntem ailelerinin temel çalışma prensipleri, güçlü yönleri ve sınırlılıkları karşılaştırmalı olarak değerlendirilmektedir. Ayrıca çekirdek segmentasyonunda yaygın olarak kullanılan veri setleri ile bölgesel örtüşme, sınır doğruluğu ve örnek düzeyindeki başarıyı ölçen değerlendirme ölçütleri incelenmektedir. Son olarak alan kayması (domain shift), organlar arası genellenebilirlik, sınırlı ve belirsiz anotasyonlar, değerlendirme standardizasyonu, hesaplama verimliliği, insan destekli (human-in-the-loop) öğrenme ve temel modellerin histopatolojiye uyarlanması gibi güncel araştırma problemleri ve gelecekteki çalışma yönleri tartışılmaktadır.

Anahtar Kelimeler: Histopatoloji, çekirdek segmentasyonu, dijital patoloji, derin öğrenme, örnek tabanlı segmentasyon, transformer.

[1] Dr. Öğr. Üyesi, Yapay Zeka ve Makine Öğrenmesi Bölümü, Kayseri Üniversitesi, ORCID: 0000-0002-4250-2880

1. Giriş

Histopatolojik inceleme, hastalıkların tanı ve karakterizasyonunda doku morfolojisinin mikroskobik düzeyde değerlendirilmesine dayanmaktadır. Dijital tarayıcıların gelişmesiyle cam preparatların yüksek çözünürlüklü tüm slayt görüntülerine (whole-slide image, WSI) dönüştürülebilmesi, patolojik değerlendirmenin bilgisayar destekli görüntü analiziyle desteklenmesini mümkün hale getirmiştir. Bununla birlikte WSI'ların gigapiksel düzeyinde görüntü boyutlarına ulaşabilmesi ve çok sayıda hücresel yapı içermesi, manuel analizin ölçeklenebilirliğini sınırlamakta ve otomatik yöntemlere duyulan gereksinimi artırmaktadır (Bankhead vd., 2017; Nunes vd., 2025).

Histopatolojik görüntülerde hücre çekirdekleri, dokunun hücresel organizasyonunu yansıtan temel morfolojik yapılardır. Çekirdek sayısı, boyutu, şekli ve mekânsal dağılımı gibi özellikler çeşitli patolojik süreçler hakkında nicel bilgi sağlayabilmektedir. Bu nedenle çekirdek segmentasyonu; hücre sayımı, nükleer morfometri, hücre tipi sınıflandırması ve tümör mikroçevresinin analizi gibi birçok ileri görevin başlangıç aşamasını oluşturmaktadır (Kumar vd., 2017; Graham vd., 2019, 2024).

Çekirdek segmentasyonunda ilk yaklaşımlar; eşikleme, renk ayrıştırma, matematiksel morfoloji, watershed ve aktif contour gibi geleneksel görüntü işleme yöntemlerine dayanmaktadır (Otsu, 1979; Kass vd., 1988; Vincent ve Soille, 1991). Ancak çekirdek morfolojisindeki heterojenlik, temas eden çekirdeklerin ayrıştırılmasındaki güçlükler ve hematoksilen-eozin (H&E) boyamasındaki renk değişkenliği, bu yöntemlerin farklı doku ve görüntüleme koşullarına genellenebilirliğini sınırlandırmıştır.

Derin öğrenmenin gelişmesiyle çekirdek segmentasyonu evrişimsel sinir ağı (convolutional neural network, CNN) tabanlı anlamsal segmentasyona, ardından temas eden çekirdekleri ayrı nesneler olarak belirlemeyi amaçlayan örnek tabanlı yöntemlere doğru ilerlemiştir. U-Net bu dönüşümün temel mimarilerinden biri olurken, HoVer-Net ve StarDist gibi yaklaşımlar çekirdeklerin geometrik ve konumsal özelliklerinden yararlanarak örnek tabanlı segmentasyona önemli katkılar sağlamıştır (Ronneberger vd., 2015; Schmidt vd., 2018; Graham vd., 2019). Son yıllarda transformer ve temel model tabanlı yaklaşımlar, geniş ölçekli ön eğitimden elde edilen temsillerin çekirdek analizine aktarılmasını mümkün hale getirmiştir (Hörst vd., 2024; Kirillov vd., 2023). Buna paralel olarak çok-organlı ve çok sınıflı veri setlerinin yaygınlaşması, yöntemlerin farklı doku ve görüntüleme koşullarındaki genellenebilirliğinin değerlendirilmesini önemli hale getirmiştir (Kumar vd., 2020; Nunes vd., 2025).

Bununla birlikte, modellerin farklı organ, laboratuvar, tarayıcı ve boyama koşullarında performanslarını koruyabilmesi, çekirdek segmentasyonu

alanındaki temel araştırma sorunlarından biri olmaya devam etmektedir. Bu doğrultuda güncel çalışmalar, farklı veri setleri arasında (cross-dataset) genellenebilirliğin değerlendirilmesine ve alan genellemesi (domain generalization) yaklaşımlarının geliştirilmesine giderek daha fazla odaklanmaktadır (Mahbod vd., 2024b; Nunes vd., 2025). Buna ek olarak, veri setleri ve değerlendirme protokollerinin standardizasyonu ile sınırlı anotasyon gereksinimiyle etkili öğrenmeyi amaçlayan yöntemler de giderek önem kazanmaktadır (Lin vd., 2023; Torbati vd., 2026).

Bu bölümün amacı, histopatolojik çekirdek segmentasyonunun temel kavramlarını, yöntemlerini, veri setlerini ve değerlendirme ölçütlerini bütüncül biçimde incelemek; güncel transformer ve temel model yaklaşımlarını mevcut yöntemsel gelişim içerisinde değerlendirmek ve alanın gelecekteki araştırma gereksinimlerini ortaya koymaktır.

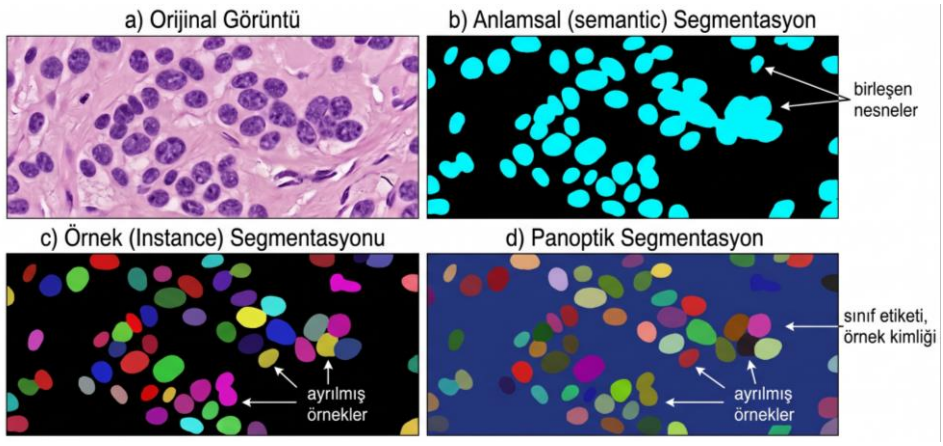
2. Histopatolojik Görüntülerde Çekirdek Segmentasyonunun Temelleri

2.1. Histopatolojik Görüntülerin Özellikleri

Hematoksilen-eozin (H&E), histopatolojik incelemelerde en yaygın kullanılan boyama yöntemlerinden biridir. Hematoksilen, hücre çekirdeklerini ağırlıklı olarak mavi-mor tonlarda görünür hale getirirken, eozin sitoplazma ve ekstrasellüler doku bileşenlerini pembe tonlarda boyamaktadır. Bununla birlikte H&E görüntülerindeki renk dağılımı; laboratuvar protokolleri, kullanılan reaktifler, kesit kalınlığı, tarayıcı özellikleri ve preparat hazırlama süreçleri gibi çeşitli faktörlerden etkilenebilmektedir (Ruifrok ve Johnston, 2001; Macenko vd., 2009; Hoque vd., 2024). Farklı boya bileşenlerinin ayrıştırılması amacıyla renk ayrıştırma (color deconvolution) yöntemlerinden yararlanılmakta, farklı merkez ve laboratuvarlardan elde edilen görüntüler arasındaki renk değişkenliğini azaltmak için ise Reinhard, Macenko ve Vahadane gibi boyama normalizasyonu yaklaşımları kullanılmaktadır (Reinhard vd., 2001; Macenko vd., 2009; Vahadane vd., 2016). Histopatolojik görüntülerin bir diğer temel özelliği, çok yüksek uzamsal çözünürlüğe sahip olmalarıdır. Tüm slayt görüntülerinin (whole-slide images, WSI) gigapiksel ölçeğine ulaşabilmesi nedeniyle, bu görüntülerin doğrudan işlenmesi önemli bellek ve hesaplama gereksinimleri oluşturmaktadır. Bu nedenle segmentasyon modelleri çoğunlukla görüntünün tamamı yerine sabit boyutlu görüntü yamaları (patch) üzerinde çalıştırılmaktadır. Kullanılan görüntü çözünürlüğü, büyütme oranı ve yama boyutu ise modelin erişebildiği morfolojik ayrıntı ve bağlamsal bilgi düzeyini belirleyen önemli deneysel değişkenler arasında yer almaktadır.

2.2. Anlamsal, Örnek Tabanlı ve Panoptik Segmentasyon

Çekirdek segmentasyonu, üretilen çıktıların yapısına göre farklı yaklaşımlar çerçevesinde ele alınabilmektedir. Anlamsal segmentasyonda görüntüdeki her piksel, çekirdek veya arka plan gibi önceden tanımlanmış bir sınıfa atanmaktadır. Bununla birlikte, birbirine temas eden çekirdeklerin tek bir bağlantılı bölge olarak segmentlenmesi durumunda, bireysel çekirdeklerin ayrı kimliklerle temsil edilmesi mümkün olmamaktadır. Örnek tabanlı segmentasyon ise her çekirdeği bağımsız bir nesne olarak ele alarak komşu veya temas eden çekirdeklerin ayrı örnekler şeklinde tanımlanmasını sağlamaktadır. Bu özellik, özellikle hücre sayımı ve çekirdek düzeyindeki morfolojik analizlerin güvenilir biçimde gerçekleştirilmesi açısından önem taşımaktadır. Panoptik segmentasyon ise anlamsal ve örnek tabanlı segmentasyon yaklaşımlarını bütünleşik bir çerçevede birleştirmektedir. Bu yaklaşımda görüntüdeki tüm pikseller anlamsal sınıflara atanırken, nesne niteliğindeki yapılar ayrıca kendilerine özgü örnek kimlikleriyle temsil edilmektedir (Kirillov vd., 2019).



Şekil 1. Anlamsal, örnek tabanlı ve panoptik çekirdek segmentasyonunun karşılaştırması.

2.3. Çekirdek Segmentasyonundaki Zorluklar

Çekirdek segmentasyonundaki temel sorunlardan biri, birbirine temas eden veya yoğun biçimde kümelenmiş çekirdeklerin ayrı nesneler olarak belirlenmesidir. Komşu çekirdekler arasında belirgin bir arka plan bulunmadığında anlamsal segmentasyon modelleri bu yapıları tek bir nesne olarak tahmin edebilmektedir. Bu nedenle uzaklık haritaları, kontur bilgisi, yön vektörleri ve geometrik temsiller gibi çekirdek içi yapıyı modelleyen ek çıktılardan yararlanılmaktadır (Naylor vd., 2019; Zhou vd., 2019; Graham vd.,

2019). Bir diğer önemli sorun morfolojik heterojenliktir. Çekirdeklerin boyutu, şekli ve görünümü farklı organ, hastalık ve hücre tipleri arasında önemli ölçüde değişebilmektedir. Buna ek olarak farklı laboratuvar ve tarayıcı koşullarından kaynaklanan boyama değişkenliği ile yüksek kaliteli piksel düzeyi anotasyonların maliyeti de model performansını ve genellenebilirliği etkilemektedir (Tellez vd., 2019; Mahbod vd., 2024a, 2024b). Sınırlı anotasyon, alan kayması ve tahmin belirsizliğine yönelik güncel yaklaşımlar Bölüm 5'te ayrıntılı olarak ele alınmaktadır.

3. Çekirdek Segmentasyon Yöntemleri

3.1. Geleneksel Görüntü İşleme Tabanlı Yöntemler

Derin öğrenme öncesindeki çekirdek segmentasyonu yöntemleri çoğunlukla renk, yoğunluk, gradyan ve geometrik özelliklere dayanmaktadır. Otsu (1979) tarafından geliştirilen otomatik eşikleme yöntemi, sınıflar arası varyansı en yüksek yapan yoğunluk eşiğini belirleyerek görüntüyü farklı bölgelere ayırmaktadır. H&E görüntülerinde hematoksilen kanalının ayrıştırılmasının ardından benzer eşikleme yöntemleri çekirdek bölgelerinin belirlenmesinde kullanılabilmektedir. Watershed yaklaşımı, özellikle birbirine temas eden nesnelerin ayrılmasında kullanılan temel yöntemlerden biridir. Görüntüyü topografik bir yüzey olarak ele alarak farklı minimumlardan gelişen bölgelerin ayrılmasını sağlar (Vincent ve Soille, 1991). Bununla birlikte marker seçimine bağlı olarak aşırı veya yetersiz segmentasyon oluşabilmektedir. Aktif contour yöntemleri ise nesne sınırını görüntü ve geometrik özelliklerden elde edilen bir enerji fonksiyonunun optimizasyonuyla belirlemektedir (Kass vd., 1988). Bu yöntemler belirli koşullarda etkili olsa da başlangıç konturu, parametreler ve görüntü gürültüsüne duyarlıdır. Histopatolojik görüntülerdeki yüksek morfolojik ve kromatik çeşitlilik, öğrenilebilir özelliklere dayanan yöntemlerin ön plana çıkmasına neden olmuştur.

3.2. CNN Tabanlı Segmentasyon Yöntemleri

Derin öğrenme yaklaşımlarının yaygınlaşmasıyla birlikte, çekirdek segmentasyonunda elle tasarlanan özelliklerin yerini büyük ölçüde veriden otomatik olarak öğrenilen hiyerarşik temsiller almıştır. Bu dönüşümün temel mimarilerinden biri olan U-Net, özellik çıkarımını gerçekleştiren kodlayıcı (encoder) ile segmentasyon haritasını yeniden oluşturan çözücü (decoder) yapılarını bir araya getirmekte ve yüksek çözünürlüklü özelliklerin atlama bağlantıları (skip connections) aracılığıyla çözücü katmanlarına aktarılmasını sağlamaktadır (Ronneberger vd., 2015). U-Net sonrasında geliştirilen mimariler, özellikle çok ölçekli özelliklerin etkin biçimde birleştirilmesine ve kodlayıcı-

çözücü arasındaki bilgi aktarımının geliştirilmesine odaklanmıştır. U-Net++, iç içe geçmiş yoğun bağlantılar aracılığıyla kodlayıcı ve çözücü özellikleri arasındaki anlamsal farklılığı azaltmayı amaçlamaktadır (Zhou vd., 2018). DeepLabV3+ ise seyreltilmiş evrişim (atrous convolution) mekanizmasından yararlanarak farklı ölçeklerdeki bağlamsal bilgiyi modellemektedir (Chen vd., 2018). Benzer biçimde Micro-Net, çoklu çözünürlük seviyelerinden elde edilen özellikleri bütünleştirirken (Raza vd., 2019), Triple U-Net RGB görüntü bilgisi ile hematoksilen bileşenini birlikte kullanarak histopatolojik görüntülere özgü boyama bilgisini segmentasyon sürecine dâhil etmektedir (Zhao vd., 2020).

Anlamsal segmentasyon yöntemlerinin, özellikle birbirine temas eden çekirdekleri bağımsız nesneler olarak ayırt etmedeki sınırlılıkları, örnek tabanlı segmentasyon yaklaşımlarının geliştirilmesini teşvik etmiştir. Bu kapsamda Naylor vd. (2019), çekirdek merkezlerine göre oluşturulan uzaklık haritalarından yararlanırken; CIA-Net çekirdek bölgeleri ile kontur bilgisini birlikte modellemektedir (Zhou vd., 2019). HoVer-Net ise çekirdek piksellerinin yatay ve düşey konumsal ilişkilerini temsil eden haritalar kullanarak komşu çekirdeklerin birbirinden ayrıştırılmasını amaçlamaktadır (Graham vd., 2019). StarDist, çekirdekleri yıldız-konveks poligonlar biçiminde temsil ederek geometrik özelliklere dayalı bir örnek tabanlı segmentasyon yaklaşımı sunmaktadır (Schmidt vd., 2018). Daha genel amaçlı bir yöntem olan Cellpose ise farklı hücre türleri ve mikroskopi görüntülerinde kullanılmak üzere uzamsal akış alanlarına dayalı bir segmentasyon yaklaşımı geliştirmiştir (Stringer vd., 2021). Yaklaşımın sonraki sürümleri, insan destekli yeniden eğitim ve görüntü restorasyonu gibi özelliklerle genişletilmiştir (Pachitariu ve Stringer, 2022; Stringer ve Pachitariu, 2025). NuHTC ise çok seviyeli özellik temsilleri ile üstten alta öneri (top-down proposal) mekanizmasını birleştirerek çekirdeklerin örnek tabanlı segmentasyonu ve sınıflandırılmasına yönelik bir yaklaşım sunmaktadır (Li vd., 2025).

3.3. Transformer Tabanlı Yaklaşımlar

Transformer mimarileri, öz-dikkat (self-attention) mekanizması sayesinde görüntünün farklı bölgeleri arasındaki uzun menzilli ilişkileri modelleyebilmekte ve yerel morfolojik özelliklerle daha geniş bağlamsal bilginin birlikte değerlendirilmesine olanak sağlamaktadır. SegFormer, hiyerarşik bir transformer kodlayıcıyı hafif bir çok katmanlı algılayıcı (multilayer perceptron, MLP) çözücü yapısıyla birleştirerek farklı çözünürlük seviyelerinden elde edilen özellikleri segmentasyon çıktısında bütünleştirmektedir (Xie vd., 2021). Bununla birlikte, genel amaçlı bir anlamsal segmentasyon mimarisi olması nedeniyle temas eden çekirdeklerin bağımsız örnekler olarak ayrıştırılması için ek mekanizmalara

gereksinim duyabilmektedir. Histopatolojik görüntü analizine daha doğrudan uyarlanan CellViT ise Vision Transformer tabanlı kodlayıcıları, HoVer-Net benzeri çekirdek segmentasyonu ve hücre sınıflandırma bileşenleriyle bir araya getirmektedir. PanNuke veri kümesi üzerinde gerçekleştirilen deneyler, önceden eğitilmiş transformer temsillerinin çekirdek segmentasyonu ve hücre tipi sınıflandırmasında etkili biçimde kullanılabileceğini göstermiştir (Hörst vd., 2024). Genel olarak transformer tabanlı yaklaşımlar, geniş bağlamsal ilişkilerin modellenmesi ve önceden eğitilmiş temsillerden yararlanılması açısından önemli avantajlar sunmaktadır. Bununla birlikte, yüksek çözünürlüklü histopatolojik görüntülerde artan bellek gereksinimi ve hesaplama maliyeti, bu yöntemlerin önemli sınırlılıkları arasında yer almaktadır.

3.4. Temel (Foundation) Modeller ve Prompt Tabanlı Güncel Yaklaşımlar

Geniş ölçekli veri setleri üzerinde önceden eğitilen temel modeller, belirli bir veri setine özgü olarak sıfırdan model geliştirme gereksinimini azaltmayı ve öğrenilen temsillerin farklı görevlere aktarılmasını amaçlamaktadır. Bu yaklaşımın öne çıkan örneklerinden biri olan Segment Anything Model (SAM), nokta, sınırlayıcı kutu (bounding box) veya maske gibi farklı istem (prompt) türleriyle yönlendirilebilen genel amaçlı bir segmentasyon modeli sunmaktadır (Kirillov vd., 2023). Bununla birlikte, modelin ağırlıklı olarak doğal görüntüler üzerinde geliştirilmiş olması, histopatolojik görüntülerde bulunan küçük, yoğun ve birbirine temas eden çekirdeklerin doğrudan segmentasyonunda alan uyuşmazlığına (domain mismatch) yol açabilmektedir.

SAM tabanlı yaklaşımların farklı tıbbi ve biyolojik görüntüleme alanlarına uyarlanmasına yönelik çeşitli çalışmalar geliştirilmiştir. MedSAM genel tıbbi görüntü segmentasyonuna (Ma vd., 2024), µSAM mikroskopi görüntülerine (Archit vd., 2025), CellSAM ise farklı hücresel görüntüleme ortamlarında genel hücre segmentasyonuna odaklanmaktadır (Marks vd., 2025). Bununla birlikte, H&E görüntülerindeki çekirdeklerin küçük boyutlu, yoğun ve sıklıkla birbirine temas eden yapısı, histopatolojiye özgü uyarlamaların gerekliliğini sürdürmektedir. Bu doğrultuda CellViT++, önceden eğitilmiş kodlayıcılardan elde edilen hücresel temsillerin yeni hücre tiplerine daha sınırlı veri ve yeniden eğitim gereksinimiyle aktarılmasını amaçlamaktadır (Hörst vd., 2026). PathoSAM, çekirdeklerin otomatik ve etkileşimli örnek tabanlı segmentasyonunu histopatolojiye özgü bir temel model çerçevesinde ele almaktadır (Griebel vd., 2026). AMA-SAM ise birden fazla histopatoloji veri kümesinden kaynaklanan alan farklılıklarını doğrudan modelleyerek çekişmeli özellik hizalama (adversarial feature alignment) yoluyla daha genellenebilir temsiller öğrenmeyi

hedeflemekte; ayrıca ince çekirdek sınırlarının daha iyi korunabilmesi amacıyla yüksek çözünürlüklü bir çözücü yapısından yararlanmaktadır (Qian vd., 2026).

Genel olarak bu gelişmeler, histopatolojik çekirdek segmentasyonunun veri setine özgü modellerden, farklı veri dağılımlarına ve analiz görevlerine daha etkin biçimde uyarlanabilen önceden eğitilmiş temel model tabanlı yaklaşımlara doğru yöneldiğini göstermektedir. Literatürde öne çıkan yöntemlerin temel özellikleri ve karşılaştırmalı değerlendirmeleri Tablo 1'de sunulmuştur.

Tablo 1. Histopatolojik çekirdek segmentasyonunda kullanılan başlıca yöntemlerin karşılaştırmalı özeti

Yaklaşım	Yöntem	Segmentasyon Türü	Özellik
Geleneksel	Otsu+Watershed	Anlamsal/örnek	Düşük veri ihtiyacı
CNN	U-Net	Anlamsal	Basit ve güçlü temel model
CNN	U-Net++	Anlamsal	Gelişmiş çok ölçekli özellik aktarımı
CNN	DeepLabV3+	Anlamsal	Çok ölçekli bağlam modelleme
Histopatoloji-özel CNN	Triple U-Net	Anlamsal	H&E boyama bilgisinden yararlanma
Kontur-duyarlı	CIA-Net	Örnek	Çekirdek ve sınır bilgisini birlikte öğrenme
Örnek-duyarlı	HoVer-Net	Örnek+ sınıflandırma	Temas eden çekirdeklerin etkili ayrımı
Geometrik	StarDist	Örnek	Yoğun çekirdek kümelerinde etkili
Genel amaçlı	Cellpose	Örnek	Farklı mikroskopi alanlarına uygulanabilirlik
Top-down	NuHTC	Örnek + sınıflandırma	Çok seviyeli özellik kullanımı
Transformer	SegFormer	Anlamsal	Global ve çok ölçekli bağlam
Transformer	CellViT	Örnek + sınıflandırma	Güçlü önceden eğitilmiş temsiller
Temel model	CellSAM	Örnek	Geniş hücresel veri alanına genellenebilirlik
Temel model	CellViT++	Örnek + sınıflandırma	Veri-verimli adaptasyon
Temel model	PathoSAM	Örnek/etkileşimli	Histopatolojiye özgü SAM yaklaşımı
Çoklu-alan temel model	AMA-SAM	Çekirdek segmentasyonu	Alan hizalama (domain alignment) ve genellenebilirlik

4. Veri Setleri ve Değerlendirme Ölçütleri

4.1. Yaygın Çekirdek Segmentasyonu Veri Setleri

Histopatolojik çekirdek segmentasyonundaki gelişmeler, farklı organları, hücre tiplerini ve görüntüleme koşullarını temsil eden standart veri setlerinin oluşturulmasıyla yakından ilişkilidir. Bu alandaki güncel karşılaştırma veri setleri, başlangıçta sınırlı kapsam ve tek görev odaklı yapılardan, çok organlı, çok sınıflı ve birden fazla analiz görevini destekleyen daha kapsamlı yapılara doğru gelişmiştir. MoNuSeg, çok organlı çekirdek segmentasyonu için temel referans veri setlerinden biri olarak kullanılırken, MoNuSAC bu kapsamı çekirdek sınıflandırmasını da içerecek biçimde genişletmiştir (Kumar vd., 2020; Verma vd., 2021). CoNSeP, kolorektal dokularda çekirdek segmentasyonu ile hücre sınıflandırmasını birlikte ele alırken, PanNuke farklı kanser ve doku türlerini kapsayan geniş yapısıyla daha genel çekirdek temsillerinin geliştirilmesine katkı sağlamaktadır (Graham vd., 2019; Gamper vd., 2019, 2020). Bunun yanında CryoNuSeg, farklı preparat koşulları ile gözlemciler arası anotasyon değişkenliğinin incelenmesine olanak sağlarken, NuInsSeg geniş organ çeşitliliğine ek olarak anotasyon belirsizliğinin değerlendirilmesini desteklemektedir (Mahbod vd., 2021, 2024a). Lizard ve CoNIC, büyük ölçekli çekirdek segmentasyonu ve sınıflandırmasının yanı sıra sonraki analiz görevlerinin değerlendirilmesine yönelik kapsamlı veri yapıları sunmaktadır (Graham vd., 2021, 2024). PUMA ise çekirdek anotasyonlarını doku ve tümör mikroçevresine ilişkin bilgilerle birlikte sunarak çekirdek düzeyi analizlerin daha geniş histopatolojik bağlam içerisinde değerlendirilmesine olanak tanımaktadır (Schuiveling vd., 2025).

Bu veri setleri arasındaki organ kapsamı, anotasyon biçimi, hücre sınıfları ve görev tanımlarındaki farklılıklar, tek bir veri setinde elde edilen yüksek performansın bir yöntemin genel genellenebilirliğini göstermek için yeterli olmadığını ortaya koymaktadır. Bu nedenle güncel çekirdek segmentasyonu yöntemlerinin farklı veri setleri ve görüntüleme koşulları üzerinde dış doğrulama yoluyla değerlendirilmesi önem taşımaktadır. Yaygın olarak kullanılan veri setlerinin temel özellikleri Tablo 2'de karşılaştırmalı olarak özetlenmiştir.

Tablo 2. Yaygın histopatolojik çekirdek segmentasyonu veri setlerinin genel özellikleri

Veri Seti	Yıl	Doku/Organ Kapsamı	Temel Görev	Özellik
MoNuSeg	2017	Çok organ	Örnek tabanlı segmentasyon	Eğitimde görülmeyen organları içeren test yapısı
MoNuSAC	2021	Çok organ	Segmentasyon ve sınıflandırma	Birden fazla çekirdek/hücre tipi
CoNSeP	2019	Kolorektal	Segmentasyon ve sınıflandırma	24.319 anotasyonlu çekirdek
PanNuke	2019/ 2020	19 doku tipi	Segmentasyon ve sınıflandırma	Yaklaşık 200.000 çekirdek ve pan-kanser yapı
CryoNuSeg	2021	10 organ	Örnek tabanlı segmentasyon	Frozen-section H&E ve çoklu manuel anotasyon
Lizard	2021	Kolorektal	Segmentasyon ve sınıflandırma	Yaklaşık yarım milyon sınıflandırılmış çekirdek
NuInsSeg	2024	31 insan/fare organı	Örnek tabanlı segmentasyon	30.698 manuel çekirdek ve belirsizlik maskeleri
CoNIC	2024	Kolorektal	Tespit, segmentasyon, sınıflandırma ve sayım	Büyük ölçekli challenge ve WSI tabanlı downstream analiz
PUMA	2025	Melanoma	Çekirdek ve doku segmentasyonu	Çekirdek ile doku mikroçevresinin birlikte anotasyonu

4.2. Anlamsal Segmentasyon Ölçütleri

Histopatolojik çekirdek segmentasyonunda anlamsal performansın değerlendirilmesinde en yaygın kullanılan ölçütler Dice benzerlik katsayısı ve Intersection over Union (IoU) değeridir. Her iki ölçüt de tahmin edilen çekirdek bölgesi ile gerçek (ground-truth) maske arasındaki örtüşmeyi değerlendirmektedir.

Tahmin maskesi P ve gerçek maske G olmak üzere Dice katsayısı,

$$Dice = \frac{2|P \cap G|}{|P| + |G|}$$

şeklinde tanımlanmaktadır. Dice değeri 0 ile 1 arasında değişmekte ve 1 tahmin ve gerçek maskeler arasında tam örtüşmeyi ifade etmektedir. IoU veya Jaccard indeksi ise,

$$IoU = \frac{|P \cap G|}{|P \cup G|}$$

şeklinde hesaplanmaktadır. Her iki ölçütte yüksek değer daha başarılı segmentasyonu göstermektedir. Kesinlik (precision) ve duyarlılık (recall) değerleri de segmentasyon performansını tamamlayıcı biçimde değerlendirmektedir:

$$Precision = \frac{TP}{TP + FP}$$
$$Recall = \frac{TP}{TP + FN}$$

Burada doğru pozitif (True Positive, TP), çekirdek olarak doğru sınıflandırılan pikselleri; yanlış pozitif (False Positive, FP), arka plan olmasına rağmen çekirdek olarak sınıflandırılan pikselleri; yanlış negatif (False Negative, FN) ise çekirdek olmasına rağmen arka plan olarak sınıflandırılan pikselleri ifade etmektedir. Kesinlik, özellikle yanlış pozitif tahminlerden etkilenirken, duyarlılık gerçek maskede yer alan ancak model tarafından tespit edilemeyen çekirdek bölgelerine karşı duyarlıdır. Bununla birlikte bu ölçütler bireysel çekirdek kimliklerini ve çekirdekler arasındaki ayrımı doğrudan dikkate almamaktadır. Bu nedenle histopatolojik çekirdek segmentasyonunun daha kapsamlı biçimde değerlendirilmesi için bölgesel ölçütlerin sınır ve örnek tabanlı değerlendirme ölçütleriyle birlikte kullanılması önerilmektedir (Kumar vd., 2020; Foucart vd., 2023).

4.3. Sınır Tabanlı Değerlendirme Ölçütleri

Çekirdek sınırlarının doğru biçimde belirlenmesi; alan, çevre ve şekil gibi morfometrik özelliklerin güvenilir olarak hesaplanması açısından önem taşımaktadır. Bu nedenle çekirdek segmentasyonu performansının değerlendirilmesinde Dice ve IoU gibi bölgesel örtüşme ölçütlerinin yanı sıra sınır doğruluğunu doğrudan dikkate alan ölçütlerden de yararlanılmaktadır.

Boundary F1 (BF1), tahmin edilen ve referans sınırların belirli bir tolerans θ içerisinde eşleşmesine dayanan sınır tabanlı bir değerlendirme ölçütüdür (Csurka vd., 2013). Tahmin ve referans maske sınır noktaları sırasıyla S_P ve S_G ile gösterildiğinde sınır kesinliği ve sınır duyarlılığı,

$$P_B = \frac{|\{p \in S_P : d(p, S_G) \leq \theta\}|}{|S_P|}$$
$$R_B = \frac{|\{g \in S_G : d(g, S_P) \leq \theta\}|}{|S_G|}$$

şeklinde tanımlanmaktadır. Burada *d(x,S)*, *x* noktası ile *S* sınır kümesindeki en yakın nokta arasındaki Öklidyen mesafeyi ifade etmektedir. BF1 değeri ise sınır kesinliği ile sınır duyarlılığının harmonik ortalaması olarak,

$$BF1 = \frac{2P_B R_B}{P_B + R_B}$$

şeklinde hesaplanmaktadır. BF1 değerinin yüksek olması, tahmin edilen sınırlar ile gerçek çekirdek sınırları arasındaki uzamsal uyumun yüksek olduğunu göstermektedir.

Sınır doğruluğunun değerlendirilmesinde kullanılan bir diğer ölçüt Hausdorff uzaklığıdır (HD). HD, iki sınır kümesindeki noktaların karşı kümedeki en yakın noktalarına olan uzaklıklarını dikkate alarak bu uzaklıkların en büyüğünü,

$$HD(S_P, S_G) = \max\left(\max_{a \in S_P} \min_{b \in S_G} |a - b|, \max_{b \in S_G} \min_{a \in S_P} |b - a|\right)$$

şeklinde tanımlamaktadır. Bununla birlikte HD, aykırı değerlere duyarlı olduğundan maksimum uzaklık yerine yüzdelik tabanlı Hausdorff uzaklıklarından yararlanılabilmektedir (Taha ve Hanbury, 2015). Bu yaklaşımın yaygın bir biçimi olan %95 Hausdorff uzaklığı (HD95), tahmin ve referans sınır noktaları arasındaki karşılıklı en yakın-komşu mesafelerinin birleşik dağılımının %95 yüzdelik değerini,

$$HD95 = percentile_{95}\left(\{\min_{b \in S_G} |a - b| \;\middle|\; a \in S_P\} \cup \{\min_{a \in S_P} |b - a| \;\middle|\; b \in S_G\}\right)$$

şeklinde dikkate almaktadır. Buna göre yüksek BF1 ve düşük HD95 değerleri, sınır segmentasyonu açısından daha başarılı bir performansa işaret etmektedir.

4.4. Örnek Tabanlı Segmentasyon Ölçütleri

Örnek tabanlı segmentasyonda her çekirdeğin yalnızca doğru biçimde segmentlenmesi değil, aynı zamanda bağımsız bir nesne olarak tespit edilmesi ve diğer çekirdeklerden ayrıştırılması gerekmektedir. Bu nedenle performans değerlendirmesinde, bölgesel örtüşmenin yanı sıra gerçek ve tahmin edilen çekirdek örnekleri arasındaki eşleşmeleri dikkate alan ölçütlerden yararlanılmaktadır.

Aggregated Jaccard Index (AJI), gerçek ve tahmin edilen çekirdek örneklerini eşleştirerek bunların kesişim ve birleşim alanlarını toplu biçimde değerlendiren

bir ölçüttür. G_i gerçek çekirdek örneğini, $P_{j(i)}$ bununla eşleşen tahmin edilen örneği ve U eşleşmeyen tahmin kümesini göstermek üzere AJI,

$$AJI = \frac{\sum_{i=1}^{N} |G_i \cap P_{j(i)}|}{\sum_{i=1}^{N} |G_i \cup P_{j(i)}| + \sum_{P_k \in U} |P_k|}$$

şeklinde tanımlanmaktadır. Eşleşmeyen tahminlerin de paydaya dâhil edilmesi, AJI'nin çekirdeklerin hatalı biçimde birleştirilmesi, ayrıştırılması veya gerçekte bulunmayan çekirdeklerin tahmin edilmesi gibi örnek düzeyindeki hatalara duyarlı olmasını sağlamaktadır (Kumar vd., 2017).

Panoptic Quality (PQ), örneklerin tespit ve segmentasyon başarısını birlikte değerlendiren bir ölçüt olarak panoptik segmentasyon kapsamında geliştirilmiştir (Kirillov vd., 2019). Çekirdek segmentasyonu literatüründe PQ'nun tespit bileşeni Detection Quality (DQ), segmentasyon bileşeni ise Segmentation Quality (SQ) olarak yaygın biçimde kullanılmaktadır (Graham vd., 2019). Eşleşen çekirdek çiftleri doğru pozitif (TP), eşleşmeyen tahminler yanlış pozitif (FP) ve eşleşmeyen gerçek çekirdekler yanlış negatif (FN) olarak tanımlandığında DQ ve SQ,

$$DQ = \frac{|TP|}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|}$$

$$SQ = \frac{\sum_{(g,p) \in TP} IoU(g,p)}{|TP|}$$

şeklinde hesaplanmaktadır. PQ ise bu iki bileşenin çarpımı olarak,

$$PQ = DQ \times SQ$$

şeklinde elde edilmektedir (Graham vd., 2019). DQ, çekirdeklerin doğru biçimde tespit edilmesi ve eşleştirilmesine ilişkin performansı; SQ ise doğru eşleşen çekirdek çiftleri arasındaki ortalama segmentasyon kalitesini ifade etmektedir.

Hücre tiplerinin de değerlendirmeye dâhil edildiği çalışmalarda, sınıf bazında hesaplanan PQ değerlerinin ortalaması alınarak ortalama Panoptic Quality (mPQ),

$$mPQ = \frac{1}{C} \sum_{c=1}^{C} PQ_c$$

şeklinde hesaplanabilmektedir. Bununla birlikte PQ'nun örnek eşleştirme eşiğine duyarlı olması ve farklı hata türlerini tek bir bileşik skor içerisinde birleştirmesi, tek başına model sıralaması amacıyla kullanılmasında bazı sınırlılıklar oluşturmaktadır (Foucart vd., 2023). Bu nedenle histopatolojik çekirdek segmentasyonunun daha kapsamlı ve güvenilir biçimde değerlendirilmesi için bölgesel, sınır tabanlı ve örnek tabanlı ölçütlerin birlikte raporlanması uygun bir yaklaşım olarak değerlendirilmektedir. Bu bölümde ele alınan değerlendirme ölçütlerinin temel özellikleri Tablo 3'te özetlenmiştir.

Tablo 3. Histopatolojik çekirdek segmentasyonunda kullanılan temel değerlendirme ölçütleri

Ölçüt	Değerlendirme Düzeyi	Temel Değerlendirme Amacı
Dice	Bölge/piksel	Tahmin ve gerçek maskelerin örtüşmesi
IoU	Bölge/piksel	Kesişim alanının birleşim alanına oranı
Precision	Piksel	Yanlış pozitif tahminlere duyarlılık
Recall	Piksel	Kaçırılan çekirdek bölgelerine duyarlılık
BF1	Sınır	Çekirdek sınırlarının uyumu
HD95	Sınır	Tahmin ve gerçek sınırları arasındaki mesafe
AJI	Örnek	Çekirdeklerin ayrı nesneler olarak eşleşmesi ve örtüşmesi
DQ	Örnek	Doğru örnek eşleştirme/tespit başarısı
SQ	Örnek	Eşleşmiş örneklerin segmentasyon kalitesi
PQ	Panoptik/örnek	Tespit ve segmentasyon başarısının ortak değerlendirilmesi
mPQ	Sınıflı örnek	Hücre sınıfları genelindeki panoptik performans

5. Mevcut Zorluklar ve Gelecek Araştırma Yönleri

Histopatolojik çekirdek segmentasyonunda önemli ilerlemeler sağlanmış olmakla birlikte, standart veri setlerinde elde edilen yüksek performansın farklı veri dağılımlarında ve gerçek klinik koşullarda sürdürülebilmesi halen önemli bir araştırma problemidir. Modellerin genellenebilirliği, sınırlı ve belirsiz anotasyonlar, temel modellerin alan-özel uyarlanması, hesaplama gereksinimleri ve değerlendirme süreçlerinin standardizasyonu, gelecekte ele alınması gereken başlıca araştırma konuları arasında yer almaktadır.

5.1. Alan Kayması ve Genellenebilirlik

Belirli bir veri seti üzerinde eğitilen çekirdek segmentasyonu modelleri, farklı organ, merkez, tarayıcı veya preparat koşullarında benzer performansı sürdüremeyebilmektedir. Boyama özellikleri, çekirdek morfolojisi, görüntü çö2zünürlüğü ve anotasyon biçimlerindeki farklılıklar, alan kaymasının başlıca kaynakları arasında yer almaktadır (Mahbod vd., 2024b; Nunes vd., 2025). Çok organlı eğitim verilerinin kullanılması bu sorunun etkisini azaltabilse de eğitim

sürecinde birden fazla organın yer alması, modelin daha önce görülmemiş bir organa başarıyla genellenebileceğini garanti etmemektedir. Bu nedenle aynı veri dağılımı içerisindeki değerlendirmelerin (in-domain) yanı sıra organlar arası (cross-organ), merkezler arası (cross-center) ve veri setleri arası (cross-dataset) doğrulama protokollerine daha fazla önem verilmesi gerekmektedir. AMA-SAM alan farklılıklarını model düzeyinde alan hizalama yoluyla ele alırken, NucFuseRank farklı çekirdek veri kümelerini ortak bir değerlendirme yapısında birleştirerek veri kümeleri arası karşılaştırılabilirliği geliştirmeyi amaçlamaktadır (Qian vd., 2026; Torbati vd., 2026).

5.2. Sınırlı Anotasyon, Anotasyon Belirsizliği ve Güvenilirlik

Piksel düzeyinde çekirdek anotasyonlarının oluşturulması önemli ölçüde uzman zamanı ve iş gücü gerektirdiğinden, tam anotasyon gereksinimini azaltmayı amaçlayan zayıf denetimli (weakly-supervised), yarı denetimli (semi-supervised) ve kendinden denetimli (self-supervised) öğrenme yaklaşımları giderek önem kazanmaktadır. Nokta anotasyonlarının ve otomatik olarak oluşturulan zayıf anotasyonların kullanılması, ayrıntılı segmentasyon maskelerine olan bağımlılığı azaltabilecek yaklaşımlar arasında değerlendirilmektedir (Zhao ve Yin, 2021; Lin vd., 2023; Shrestha vd., 2023).

Bununla birlikte, gerçek maskeler her zaman kesin ve hatasız bir temel gerçeklik olarak değerlendirilememektedir. Çoklu uzman anotasyonları ve belirsizlik maskeleri, özellikle çekirdek sınırlarının belirlenmesinde uzmanlar arasında farklılıklar bulunabileceğini göstermektedir (Mahbod vd., 2021, 2024a). Ayrıca kusurlu veya hatalı anotasyonların model eğitimi ve segmentasyon performansı üzerinde olumsuz etkiler oluşturabildiği gösterilmiştir (Gálvez Jiménez ve Decaestecker, 2024). Bu nedenle gelecekte anotasyon kalitesi ile model belirsizliğinin birlikte değerlendirilmesi; güvenilirlik haritaları, uzmanlar arası uyum ve olasılıksal referans maskeler gibi yaklaşımların geliştirilmesi açısından önem taşımaktadır.

5.3. Temel Modeller, Adaptasyon ve Hesaplama Verimliliği

Temel modeller, geniş ölçekli ön eğitim süreçlerinde öğrenilen temsillerin yeni histopatolojik veri setlerine daha sınırlı anotasyon gereksinimiyle aktarılabilmesi açısından önemli bir potansiyel sunmaktadır. Bununla birlikte farklı organ ve merkezlere genellenebilirlik, histopatolojiye özgü uyarlama ve sınırlı veriyle etkili ince ayar (fine-tuning) halen açık araştırma problemleri arasında yer almaktadır (Marks vd., 2025; Hörst vd., 2026; Griebel vd., 2026). Bu modeller açısından bir diğer önemli sınırlılık hesaplama maliyetidir. Büyük ölçekli transformer ve temel model kodlayıcılarının, WSI düzeyindeki

görüntülerde yüksek bellek gereksinimi, uzun çıkarım (inference) süresi ve artan enerji tüketimi oluşturması mümkündür. Bu nedenle gelecek çalışmalarda yalnızca segmentasyon doğruluğunun değil; parametre sayısı, grafik işlem birimi (graphics processing unit, GPU) bellek kullanımı, çıkarım süresi ve enerji tüketimi gibi hesaplama kaynaklarına ilişkin göstergelerin de raporlanması önem taşımaktadır. Daha hafif çözücü mimarileri, model sıkıştırma ve bilgi damıtma (knowledge distillation) gibi yaklaşımlar, doğruluk ile hesaplama verimliliği arasında daha dengeli çözümler geliştirilmesi açısından önem kazanmaktadır.

5.4. Değerlendirme Standardizasyonu ve Karşılaştırılabilirlik

Çalışmalar arasında eğitim-test ayrımlarının, görüntü çözünürlüğünün, yama boyutunun, ön işleme ve son işleme adımlarının ve değerlendirme ölçütlerinin uygulanma biçimlerinin farklılık göstermesi, elde edilen sonuçların doğrudan karşılaştırılmasını güçleştirmektedir (Chen vd., 2025; Nunes vd., 2025). Gelecekte geliştirilecek karşılaştırmalı değerlendirme çerçevelerinin yalnızca veri biçimini değil; görüntü ölçeğini, veri bölme protokollerini, ön ve son işleme adımlarını, değerlendirme ölçütlerini ve hesaplama maliyetine ilişkin raporlama ilkelerini de mümkün olduğunca standartlaştırması gerekmektedir. Bu tür ortak değerlendirme protokolleri, yöntemler arasında daha güvenilir, tekrarlanabilir ve adil karşılaştırmaların yapılmasına katkı sağlayacaktır.

5.5. Klinik Uygulamaya Aktarım ve İnsan Destekli (Human-in-the-Loop) Sistemler

Klinik uygulama açısından standart veri kümelerinde yüksek segmentasyon performansı elde edilmesi tek başına yeterli değildir. Bir modelin farklı merkez ve görüntüleme koşullarında güvenilir biçimde çalışabilmesi, hatalarının uzmanlar tarafından kolaylıkla belirlenip düzeltilebilmesi ve yeni verilere sınırlı anotasyonla uyarlanabilmesi gerekmektedir. Cellpose 2.0, PathoSAM ve güncel insan destekli yaklaşımlar, uzman geri bildiriminin model uyarlama süreçlerine dâhil edilmesinin bu açıdan önemli bir potansiyel taşıdığını göstermektedir (Pachitariu ve Stringer, 2022; Guo vd., 2025; Griebel vd., 2026). Bunun yanında yüksek Dice veya PQ değerleri doğrudan klinik faydayı göstermemektedir. Segmentasyon sonucunda elde edilen çekirdek morfolojisi, hücre tipi ve mekânsal organizasyon gibi özelliklerin tanı, prognoz ve biyobelirteç analizlerine sağladığı katkının da ortaya konulması gerekmektedir. Bu nedenle klinik kullanıma yönelik çekirdek segmentasyonu sistemlerinin yalnızca segmentasyon performansı açısından değil; genellenebilirlik, uzman-model etkileşimi, hesaplama verimliliği ve sonraki klinik analizlere sağladığı katkı açısından da bütüncül biçimde değerlendirilmesi gerekmektedir.

6. Sonuç ve Tartışma

Histopatolojik görüntülerde çekirdek segmentasyonu; hücre sayımı, nükleer morfometri, hücre tipi sınıflandırması ve tümör mikroçevresinin analizi gibi birçok ileri görüntü analizi görevinin temelini oluşturmaktadır (Graham vd., 2019, 2024; Nunes vd., 2025). Bu alandaki yöntemsel gelişim, geleneksel görüntü işleme yaklaşımlarından CNN tabanlı anlamsal segmentasyon yöntemlerine, ardından temas eden çekirdekleri bağımsız nesneler olarak belirlemeyi amaçlayan örnek tabanlı segmentasyon yaklaşımlarına doğru ilerlemiştir. Buna paralel olarak çok organlı ve çok sınıflı veri setlerinin geliştirilmesi, çekirdek segmentasyonu yöntemlerinin daha heterojen doku ve görüntüleme koşullarında değerlendirilmesine olanak sağlamıştır. Performans değerlendirmesinde yalnızca Dice ve IoU gibi bölgesel örtüşme ölçütlerinin kullanılması, çekirdek sınırlarının doğruluğunu ve bireysel çekirdeklerin birbirinden ayrıştırılma başarısını tam olarak yansıtmayabilmektedir. Bu nedenle BF1 ve HD95 gibi sınır tabanlı ölçütler ile AJI ve PQ gibi örnek tabanlı ölçütlerin birlikte kullanılması, segmentasyon performansının daha kapsamlı biçimde değerlendirilmesini sağlamaktadır (Taha ve Hanbury, 2015; Kumar vd., 2017; Foucart vd., 2023).

Güncel Transformer ve temel model tabanlı yaklaşımlar, önceden öğrenilmiş temsillerin farklı histopatolojik görevlere aktarılması, sınırlı anotasyonla uyarlama ve farklı veri dağılımlarına genellenebilirliğin artırılması üzerine yoğunlaşmaktadır (Hörst vd., 2024, 2026; Marks vd., 2025; Griebel vd., 2026; Qian vd., 2026). Bununla birlikte boyama ve görüntüleme koşullarındaki farklılıklar, organlar arası morfolojik değişkenlik, anotasyon maliyeti, alan kayması ve WSI ölçeğindeki yüksek hesaplama gereksinimleri, güncel yöntemlerin klinik uygulamalara aktarımını sınırlayan başlıca sorunlar arasında yer almaktadır.

Sonuç olarak, gelecekteki çalışmaların yalnızca tek bir veri setinde yüksek segmentasyon performansı elde etmeye odaklanması yeterli değildir. Modellerin farklı organ, merkez ve görüntüleme koşullarındaki genellenebilirliği, sınırlı anotasyonla uyarlanabilirliği, hesaplama verimliliği ve klinik kullanıma uygunluğu birlikte değerlendirilmelidir. Bu doğrultuda histopatolojik çekirdek segmentasyonu araştırmalarının, yalnızca model performansının iyileştirilmesine odaklanan yaklaşımlardan daha genellenebilir, güvenilir ve klinik açıdan uygulanabilir çekirdek analizi sistemlerinin geliştirilmesine doğru ilerlemesi beklenmektedir.

7. Kaynakça

Archit, A., Freckmann, L., Nair, S., Khalid, N., Hilt, P., Rajashekar, V., Freitag, M., Teuber, C., Spitzner, M., Tapia Contreras, C., Buckley, G., von Haaren, S., Gupta, S., Grade, M., Wirth, M., Schneider, G., Dengel, A., Ahmed, S., ve Pape, C. (2025). Segment Anything for Microscopy. Nature Methods, 22, 579–591.

Bankhead, P., Loughrey, M. B., Fernández, J. A., Dombrowski, Y., McArt, D. G., Dunne, P. D., McQuaid, S., Gray, R. T., Murray, L. J., Coleman, H. G., James, J. A., Salto-Tellez, M., ve Hamilton, P. W. (2017). QuPath: Open source software for digital pathology image analysis. Scientific Reports, 7, 16878.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., ve Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. European Conference on Computer Vision, 801–818.

Chen, Y., Huang, Q., Geng, M., Wang, Z., ve Han, Y. (2025). A systematic review on cell nucleus instance segmentation. IET Image Processing, 19, e70129.

Csurka, G., Larlus, D., ve Perronnin, F. (2013). What is a good evaluation measure for semantic segmentation? Proceedings of the British Machine Vision Conference, 32.1–32.11.

Foucart, A., Debeir, O., ve Decaestecker, C. (2023). Panoptic quality should be avoided as a metric for assessing cell nuclei segmentation and classification in digital pathology. Scientific Reports, 13, 8614.

Gálvez Jiménez, L., ve Decaestecker, C. (2024). Impact of imperfect annotations on CNN training and performance for instance segmentation and classification in digital pathology. Computers in Biology and Medicine, 177, 108586.

Gamper, J., Koohbanani, N. A., Benet, K., Khuram, A., ve Rajpoot, N. (2019). PanNuke: An open pan-cancer histology dataset for nuclei instance segmentation and classification. European Congress on Digital Pathology, 11–19.

Gamper, J., Koohbanani, N. A., Benes, K., Graham, S., Jahanifar, M., Khurram, S. A., Azam, A., Hewitt, K., ve Rajpoot, N. (2020). PanNuke dataset extension, insights and baselines. arXiv preprint arXiv:2003.10778.

Graham, S., Vu, Q. D., Raza, S. E. A., Azam, A., Tsang, Y.-W., Kwak, J. T., ve Rajpoot, N. (2019). HoVer-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. Medical Image Analysis, 58, 101563.

Graham, S., Jahanifar, M., Azam, A., Nimir, M., Tsang, Y.-W., Dodd, K., Hero, E., Sahota, H., Tank, A., Benes, K., Wahab, N., Minhas, F., Raza, S. E. A., El Daly, H., Gopalakrishnan, K., Snead, D., ve Rajpoot, N. (2021). Lizard: A large-scale dataset for colonic nuclear instance segmentation and classification. Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 684–693.

Graham, S., Vu, Q. D., Jahanifar, M., Weigert, M., Schmidt, U., Zhang, W., Zhang, J., Yang, S., Xiang, J., Wang, X., Rumberger, J. L., Baumann, E., Hirsch, P., Liu, L., Hong, C., Aviles-Rivero, A. I., Jain, A., Ahn, H., Hong, Y., Azzuni, H., Xu, M., Yaqub, M., Blache, M.-C., Piégu, B., Vernay, B., Scherr, T., Böhland, M., Löffler, K., Li, J., Ying, W., Wang, C., Snead, D., Raza, S. E. A., Minhas, F., ve Rajpoot, N. M. (2024). CoNIC Challenge: Pushing the frontiers of nuclear detection, segmentation, classification and counting. Medical Image Analysis, 92, 103047.

Griebel, T., Archit, A., ve Pape, C. (2026). Segment Anything for Histopathology. Proceedings of the 8th International Conference on Medical Imaging with Deep Learning, Proceedings of Machine Learning Research, 301, 515–546.

Guo, J., Lu, S., Cui, C., Deng, R., Yao, T., Tao, Z., Lin, Y., Lionts, M., Liu, Q., Xiong, J., Wang, Y., Zhao, S., Chang, C., Wilkes, M., Fogo, A., Yin, M., Yang, H., ve Huo, Y. (2025). Evaluating cell AI foundation models in kidney pathology with human-in-the-loop enrichment. Communications Medicine, 5, 495.

Hoque, M. Z., Keskinarkaus, A., Nyberg, P., ve Seppänen, T. (2024). Stain normalization methods for histopathology image analysis: A comprehensive review and experimental comparison. Information Fusion, 102, 101997.

Hörst, F., Rempe, M., Heine, L., Seibold, C., Keyl, J., Baldini, G., Ugurel, S., Siveke, J., Grünwald, B., Egger, J., ve Kleesiek, J. (2024). CellViT: Vision Transformers for precise cell segmentation and classification. Medical Image Analysis, 94, 103143.

Hörst, F., Rempe, M., Becker, H., Heine, L., Keyl, J., ve Kleesiek, J. (2026). CellViT++: Energy-efficient and adaptive cell segmentation and classification using foundation models. Computer Methods and Programs in Biomedicine, 277, 109206.

Kass, M., Witkin, A., ve Terzopoulos, D. (1988). Snakes: Active contour models. International Journal of Computer Vision, 1, 321–331.

Kirillov, A., He, K., Girshick, R., Rother, C., ve Dollár, P. (2019). Panoptic segmentation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 9404–9413.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., ve Girshick, R. (2023). Segment Anything. Proceedings of the IEEE/CVF International Conference on Computer Vision, 4015–4026.

Kumar, N., Verma, R., Sharma, S., Bhargava, S., Vahadane, A., ve Sethi, A. (2017). A dataset and a technique for generalized nuclear segmentation for computational pathology. IEEE Transactions on Medical Imaging, 36(7), 1550–1560.

Kumar, N., Verma, R., Anand, D., Zhou, Y., Onder, O. F., Tsougenis, E., Chen, H., Heng, P. A., Li, J., Hu, Z., Wang, Y., Koohbanani, N. A., Jahanifar, M., Tajeddin, N. Z., Gooya, A., Rajpoot, N., Ren, X., Zhou, S., Wang, Q., Shen, D., Yang, C. K., Weng, C. H., Yu, W. H., Yeh, C. Y., Yang, S., Xu, S., Yeung, P. H., Sun, P., Mahbod, A., Schaefer, G., Ellinger, I., Ecker, R., Smedby, O., Wang, C., Chidester, B., Ton, T. V., Tran, M. T., Ma, J., Do, M. N., Graham, S., Vu, Q. D., Kwak, J. T., Gunda, A., Chunduri, R., Hu, C., Zhou, X., Lotfi, D., Safdari, R., Kascenas, A., O'Neil, A., Eschweiler, D., Stegmaier, J., Cui, Y., Yin, B., Chen, K., Tian, X., Gruening, P., Barth, E., Arbel, E., Remer, I., Ben-Dor, A., Sirazitdinova, E., Kohl, M., Braunewell, S., Li, Y., Xie, X., Shen, L., Ma, J., Baksi, K. D., Khan, M. A., Choo, J., Colomer, A., Naranjo, V., Pei, L., Iftekharuddin, K. M., Roy, K., Bhattacharjee, D., Pedraza, A., Bueno, M. G., Devanathan, S., Radhakrishnan, S., Koduganty, P., Wu, Z., Cai, G., Liu, X., Wang, Y., ve Sethi, A. (2020). A multi-organ nucleus segmentation challenge. IEEE Transactions on Medical Imaging, 39(5), 1380–1391.

Li, B., Liu, Z., Zhang, S., Liu, X., Sun, C., Liu, J., Qiu, B., ve Tian, J. (2025). NuHTC: A hybrid task cascade for nuclei instance segmentation and classification. Medical Image Analysis, 103, 103595.

Lin, Y., Qu, Z., Chen, H., Gao, Z., Li, Y., Xia, L., Ma, K., Zheng, Y., ve Cheng, K.-T. (2023). Nuclei segmentation with point annotations from pathology images via self-supervised learning and co-training. Medical Image Analysis, 89, 102933.

Ma, J., He, Y., Li, F., Han, L., You, C., ve Wang, B. (2024). Segment anything in medical images. Nature Communications, 15, 654.

Macenko, M., Niethammer, M., Marron, J. S., Borland, D., Woosley, J. T., Guan, X., Schmitt, C., ve Thomas, N. E. (2009). A method for normalizing

histology slides for quantitative analysis. IEEE International Symposium on Biomedical Imaging, 1107–1110.

Mahbod, A., Schaefer, G., Bancher, B., Löw, C., Dorffner, G., Ecker, R., ve Ellinger, I. (2021). CryoNuSeg: A dataset for nuclei instance segmentation of cryosectioned H&E-stained histological images. Computers in Biology and Medicine, 132, 104349.

Mahbod, A., Polak, C., Feldmann, K., Khan, R., Gelles, K., Dorffner, G., Woitek, R., Hatamikia, S., ve Ellinger, I. (2024a). NuInsSeg: A fully annotated dataset for nuclei instance segmentation in H&E-stained histological images. Scientific Data, 11, 295.

Mahbod, A., Dorffner, G., Ellinger, I., Woitek, R., ve Hatamikia, S. (2024b). Improving generalization capability of deep learning-based nuclei instance segmentation by non-deterministic train time and deterministic test time stain normalization. Computational and Structural Biotechnology Journal, 23, 669–678.

Marks, M., Israel, U., Dilip, R., Li, Q., Yu, C., Laubscher, E., Iqbal, A., Pradhan, E., Ates, A., Abt, M., Brown, C., Pao, E., Li, S., Pearson-Goulart, A., Perona, P., Gkioxari, G., Barnowski, R., Yue, Y., ve Van Valen, D. (2025). CellSAM: A foundation model for cell segmentation. Nature Methods, 22, 2585–2593.

Naylor, P., Laé, M., Reyal, F., ve Walter, T. (2019). Segmentation of nuclei in histopathology images by deep regression of the distance map. IEEE Transactions on Medical Imaging, 38(2), 448–459.

Nunes, J. D., Montezuma, D., Oliveira, D., Pereira, T., ve Cardoso, J. S. (2025). A survey on cell nuclei instance segmentation and classification: Leveraging context and attention. Medical Image Analysis, 99, 103360.

Otsu, N. (1979). A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics, 9(1), 62–66.

Pachitariu, M., ve Stringer, C. (2022). Cellpose 2.0: How to train your own model. Nature Methods, 19, 1634–1641.

Qian, J., Fang, Y., Hao, J., ve Zhou, B. (2026). AMA-SAM: Adversarial multi-domain alignment of Segment Anything Model for high-fidelity histology nuclei segmentation. Medical Image Analysis, 110, 104010.

Raza, S. E. A., Cheung, L., Shaban, M., Graham, S., Epstein, D., Pelengaris, S., Khan, M., ve Rajpoot, N. M. (2019). Micro-Net: A unified model for segmentation of various objects in microscopy images. Medical Image Analysis, 52, 160–173.

Reinhard, E., Ashikhmin, M., Gooch, B., ve Shirley, P. (2001). Color transfer between images. IEEE Computer Graphics and Applications, 21(5), 34–41.

Ronneberger, O., Fischer, P., ve Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention, 234–241.

Ruifrok, A. C., ve Johnston, D. A. (2001). Quantification of histochemical staining by color deconvolution. Analytical and Quantitative Cytology and Histology, 23(4), 291–299.

Schmidt, U., Weigert, M., Broaddus, C., ve Myers, G. (2018). Cell detection with star-convex polygons. Medical Image Computing and Computer-Assisted Intervention, 265–273.

Schuiveling, M., Liu, H., Eek, D., Breimer, G. E., Suijkerbuijk, K. P. M., Blokx, W. A. M., ve Veta, M. (2025). A novel dataset for nuclei and tissue segmentation in melanoma with baseline nuclei segmentation and tissue segmentation benchmarks. GigaScience, 14, giaf011.

Shrestha, P., Kuang, N., ve Yu, J. (2023). Efficient end-to-end learning for cell segmentation with machine generated weak annotations. Communications Biology, 6, 232.

Stringer, C., Wang, T., Michaelos, M., ve Pachitariu, M. (2021). Cellpose: A generalist algorithm for cellular segmentation. Nature Methods, 18, 100–106.

Stringer, C., ve Pachitariu, M. (2025). Cellpose3: One-click image restoration for improved cellular segmentation. Nature Methods, 22, 592–599.

Taha, A. A., ve Hanbury, A. (2015). Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. BMC Medical Imaging, 15, 29.

Tellez, D., Litjens, G., Bándi, P., Bulten, W., Bokhorst, J.-M., Ciompi, F., ve van der Laak, J. (2019). Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. Medical Image Analysis, 58, 101544.

Torbati, N., Meshcheryakova, A., Woitek, R., Hatamikia, S., Mechtcheriakova, D., ve Mahbod, A. (2026). NucFuseRank: Dataset fusion and performance ranking for nuclei instance segmentation. Bioengineering, 13(8), 886.

Vahadane, A., Peng, T., Sethi, A., Albarqouni, S., Wang, L., Baust, M., Steiger, K., Schlitter, A. M., Esposito, I., ve Navab, N. (2016). Structure-preserving color normalization and sparse stain separation for histological images. IEEE Transactions on Medical Imaging, 35(8), 1962–1971.

Verma, R., Kumar, N., Patil, A., Kurian, N. C., Rane, S., Graham, S., Vu, Q. D., Zwager, M., Raza, S. E. A., Rajpoot, N., Wu, X., Chen, H., Huang, Y., Wang, L., Jung, H., Brown, G. T., Liu, Y., Liu, S., Jahromi, S. A. F., Khani, A. A., Montahaei, E., Baghshah, M. S., Behroozi, H., Semkin, P., Rassadin, A., Dutande, P., Lodaya, R., Baid, U., Baheti, B., Talbar, S., Mahbod, A., Ecker, R., Ellinger, I., Luo, Z., Dong, B., Xu, Z., Yao, Y., Lv, S., Feng, M., Xu, K., Zunair, H., Hamza, A. B., Smiley, S., Yin, T.-K., Fang, Q.-R., Srivastava, S., Mahapatra, D., Trnavska, L., Zhang, H., Narayanan, P. L., Law, J., Yuan, Y., Tejomay, A., Mitkari, A., Koka, D., Ramachandra, V., Kini, L., ve Sethi, A. (2021). MoNuSAC2020: A multi-organ nuclei segmentation and classification challenge. IEEE Transactions on Medical Imaging, 40(12), 3413–3423.

Vincent, L., ve Soille, P. (1991). Watersheds in digital spaces: An efficient algorithm based on immersion simulations. IEEE Transactions on Pattern Analysis and Machine Intelligence, 13(6), 583–598.

Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., ve Luo, P. (2021). SegFormer: Simple and efficient design for semantic segmentation with Transformers. Advances in Neural Information Processing Systems, 34, 12077–12090.

Zhao, B., Chen, X., Li, Z., Yu, Z., Yao, S., Yan, L., Wang, Y., Liu, Z., Liang, C., ve Han, C. (2020). Triple U-Net: Hematoxylin-aware nuclei segmentation with progressive dense feature aggregation. Medical Image Analysis, 65, 101786.

Zhao, T., ve Yin, Z. (2021). Weakly supervised cell segmentation by point annotation. IEEE Transactions on Medical Imaging, 40(10), 2736–2747.

Zhou, Y., Onder, O. F., Dou, Q., Tsougenis, E., Chen, H., ve Heng, P.-A. (2019). CIA-Net: Robust nuclei instance segmentation with contour-aware information aggregation. Information Processing in Medical Imaging, 682–693.

Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., ve Liang, J. (2018). UNet++: A nested U-Net architecture for medical image segmentation. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, 3–11.

2. Bölüm

Biyoinformatikte Grafik Öğrenimi: Grafik Sinir Ağı Mimarileri, Biyolojik Grafik Oluşturma ve Biyoinformatik Uygulamalarına Genel Bakış

Ayşenur SOYTÜRK PATAT[1]

Özet

Bu bölüm, grafik tabanlı derin öğrenme yaklaşımlarının biyoinformatik alanındaki kuramsal temellerini ve güncel uygulamalarını içermektedir. Giriş bölümünde biyolojik sistemlerin grafik yapılarına neden uygun olduğu açıklanmaktadır. İkinci bölümde ise protein-protein etkileşim ağları, gen düzenleyici ağlar, moleküler grafikler, biyomedikal bilgi grafikleri, çoklu omik ilişki ağları gibi temel biyolojik grafik yapıları açıklanmaktadır. Ayrıca biyolojik verilerin modellenmesinde karşılaşılan heterojenlik, eksik veri ve gürültü gibi temel modelleme problemleri açıklanmaktadır. Üçüncü bölümde Grafik Sinir Ağlarının (Graph Neural Networks, GNN) temel çalışma prensipleri mesaj iletimi (message passing) çerçevesi temelinde açıklanmakta; Grafik Evrişimsel Ağları (GCN), GraphSAGE, Grafik Dikkat Ağları (GAT) gibi biyoinformatik analizlerde kullanılan yaygın mimariler kuramsal olarak karşılaştırılmaktadır. Bunun yanı sıra heterojen grafik sinir ağları, grafik transformatörleri, geometrik derin öğrenme ve kendi kendine denetimli grafik öğrenmesi gibi güncel yaklaşımlar, biyolojik problemlerin çözümündeki avantajları ve sınırlılıkları açısından değerlendirilmektedir. Son bölümde ise grafik tabanlı yöntemlerin biyoinformatik uygulamalarındaki kullanımı genel değerlendirme ise güncel literatür bilgileriyle açıklanmaktadır. Bunun yanı sıra GNN tabanlı çözümlerde karşılaşılan temel zorluklar, model yorumlanabilirliği, ölçeklenebilirlik ve biyolojik güvenilirlik gibi konular ışığında gelecekteki çalışmaların yönelimlerine ilişkin genel bir perspektif sunulmaktadır.

[1] Dr. Öğretim Üyesi, Yapay Zekâ ve Makine Öğrenmesi, Bilgisayar ve Bilişim Bilimleri Fakültesi, Kayseri Üniversitesi, Türkiye, ORCID: 0000-0002-1086-3913

1. Giriş

Biyoinformatik, biyolojik sistemlerden elde edilen yüksek hacimli ve karmaşık verilerin analizi ve biyolojik olarak anlamlı bilgiye dönüştürülmesi olarak tanımlanır [1]. Burada biyolojik verinin analizi için bilgisayar bilimleri, matematik, istatistik gibi alanlardan yararlanılır. Biyoinformatik başlangıçta biyolojik dizilerin karşılaştırılması, veri tabanı yönetimi ve filogenetik analizler gibi sınırlı uygulama alanlarına sahipti ancak son yirmi yılda yüksek verimli deneysel teknolojilerde yaşanan gelişmeler, elde edilen veri miktarını arttırmış ve biyoinformatiğin biyomedikal araştırmaların merkezinde yer almasını sağlamıştır. Özellikle Yeni Nesil Dizileme (Next Generation Sequencing, NGS) teknolojilerinin yaygınlaşması, genom, transkriptom, epigenom, proteom ve metabolom gibi omik verilerde yüksek çıktıda veri üretebilmeyi sağlamıştır [2]. Yaşanan bu gelişmeler sadece verinin hacmini arttırmakla kalmamış, aynı zamanda çeşitliliği, heterojen yapısı ve çok katmanlı organizasyonu da önemli ölçüde artmıştır [3]. Böylelikle biyoinformatik günümüzde yalnızca biyolojik verilerin analiz edildiği yardımcı bir alan olmaktan çıkmış; hassas tıp, ilaç geliştirme, sistem biyolojisi ve kişiselleştirilmiş tedavi yaklaşımlarını destekleyen temel bir araştırma disiplinine dönüşmüştür [4].

Son yıllarda üretilen biyolojik bilgi sadece NGS teknolojisine bağlı değildir. Üçüncü nesil dizileme, tek hücre dizileme, transkriptomik, proteomik ve metabolomik teknolojilerinin de gelişmesi biyolojik sistemlerin aydınlatılmasını sağlamaktadır. Tek hücre düzeyinde gerçekleştirilen analizler, aynı doku içerisinde yer alan hücre popülasyonlarının moleküler heterojenliğini ortaya koyarken; transkriptomik teknolojileri gen ekspresyonunun hem miktarını hem de doku içerisindeki konumsal organizasyonunu incelemeye imkan tanımaktadır. Benzer şekilde kütle spektrometrisine dayalı proteomik ve metabolomik yaklaşımlar, hücresel süreçlerin fonksiyonel çıktılarının çok boyutlu olarak değerlendirilmesine olanak sağlamaktadır. Böylelikle farklı biyolojik katmanlardan (genomik, transkriptomik, epigenomik, proteomik ve metabolomik) gelen verilerin birlikte analiz edilmesini gerektiren çoklu omik yaklaşımları benimsenmektedir [5]. Ancak elde dilen bu veriler yüksek boyut, eksik gözlemler, deneysel gürültü ve karmaşık biyolojik ilişkileri de içermektedir [5,6]. Bu nedenle geleneksel istatistiksel yöntemler ya da yalnızca özellik tabanlı methodlar bu büyük ve ilişkisel veri yapılarında biyolojik olarak anlamlı örüntüleri ortaya çıkarmakta çoğu zaman yetersiz kalmaktadır [6]. Bu doğrultuda gelişen yapay zeka yöntemleri biyoinformatik araştırmalarında giderek daha önemli bir yer edinmektedir. Destek Vektör Makineleri (Support Vector Machines, SVM), Rastgele Ormanlar (Random Forest), lojistik regresyon ve gradyan artırmalı karar ağaçları gibi klasik makine öğrenmesi

algoritmaları uzun yıllar boyunca gen ekspresyonu sınıflandırması, protein fonksiyon tahmini, biyobelirteç keşfi ve hastalık tanısı gibi birçok biyoinformatik probleminde başarılı biçimde kullanılmıştır [7]. Daha sonra derin öğrenmenin gelişmesiyle birlikte Evrişimsel Sinir Ağları (Convolutional Neural Networks, CNN) görüntü tabanlı biyomedikal analizlerde, Tekrarlayan Sinir Ağları (Recurrent Neural Networks, RNN) ve bunların türevleri ise biyolojik dizilerin modellenmesinde önemli başarılar göstermiştir [8]. Ancak bu yöntemler verilerin düzenli Öklid uzaylarında (ızgara yapıları veya sıralı diziler) temsil edildiğini varsayar. Ancak biyolojik sitemler çoğu zaman doğrusal olmayan genler, proteinler, metabolitler, ilaçlar ve hastalıklar arasındaki verilerden oluşur. Protein-protein etkileşim ağları, gen düzenleyici ağlar, metabolik yolaklar ve biyomedikal bilgi grafikleri gibi yapılar; değişken komşuluk ilişkileri, heterojen düğüm tipleri ve çok ölçekli topolojik organizasyonları nedeniyle CNN veya RNN gibi modeller tarafından doğal biçimde temsil edilememektedir. Sonuç olarak geleneksel makine öğrenmesi ve klasik derin öğrenme yöntemleri biyolojik sistemlerdeki uzun menzilli bağımlılıkları, ağ topolojisini ve çok katmanlı ilişkisel bilgiyi birlikte modelleme konusunda önemli sınırlılıklar içermektedir. Biyolojik sistemler sürekli birbiri ile etkileşim halinde olan moleküler bileşenlerden oluşan karmaşık bir ağ yapısına sahiptir. Genler birbirleriyle etkileşerek ekspresyonlarını düzenlerler ya da proteinler çok sayıda fiziksel ve fonksiyonel etkileşim aracılığıyla biyolojik süreçlerini gerçekleştirir, metabolitler biyokimyasal reaksiyon ağlarını oluşturur ve hastalık fenotipleri bu çok katmanlı moleküler etkileşimlerin sonucu olarak ortaya çıkar. Bu nedenle biyolojik bilgi tek tek ve ayrı ayrı moleküllerin özelliklerinden değil, bu moleküller arasındaki ilişkilerden oluşur. Günümüzde protein-protein etkileşim (Protein-Protein Interaction, PPI) ağları, gen düzenleyici ağlar (Gene Regulatory Networks, GRN), moleküler grafikler, metabolik ağlar, biyomedikal bilgi grafikleri ve çoklu omik ilişki ağları biyoinformatikte yaygın olarak kullanılan grafik temsillerini oluşturmaktadır. Bu grafiklerde düğümler biyolojik varlıkları temsil ederken, kenarlar fiziksel etkileşimleri, düzenleyici ilişkileri, fonksiyonel benzerlikleri veya biyokimyasal bağlantıları ifade etmektedir. Böylece biyolojik sistemler yalnızca özellik tabanlı değil, aynı zamanda topolojik ve ilişkisel bağlam içerecek şekilde kodlanır [9].

Graf yapılarının biyolojik verileri temsil edebilmesi, bu yapılardan yeni nesil derin öğrenme modellerinin gelişmesine zemin hazırlamıştır. İlk olarak Grafik Sinir Ağları (Graph Neural Networks, GNN), düzensiz ve Öklid dışı veri yapıları üzerinde öğrenme gerçekleştirebilen derin öğrenme modelleri olarak ortaya çıkmış ve Kipf ve Welling [10] tarafından geliştirilen Grafik Evrişimsel

Ağ (Graph Convolutional Network, GCN) mimarisi bu alanda kullanılmaya başlanmıştır. Daha sonraki yıllarda Hamilton ve arkadaşlarının geliştirdiği GraphSAGE [11] modeli büyük ölçekli grafiklerde örnekleme tabanlı öğrenmeyi mümkün kılmış; Veličković ve arkadaşlarının [12] Grafik Dikkat Ağları (Graph Attention Networks, GAT) ise komşu düğümlerin önemini dikkat mekanizmaları aracılığıyla öğrenerek ilişkisel modellemeyi daha esnek hale getirmiştir. Bu mimarilerin temelinde ise mesaj iletimi (message passing) yoluyla sadece ilgili düğümün özelliklerinden değil, komşularından ve ağ topolojisinden de bilgi öğrenebilmesi yer almaktadır. Protein üç boyutlu yapısı tahmininde çok yüksek başarı elde eden AlphaFold'un [13] başarısında da bu mekanizma yer almaktadır. Özellikle protein yapı ve fonksiyon tahmininde kullanılan araçlardan DeepFRI'nin [14] protein fonksiyon tahmini ve MOGONET'in [15] çoklu omik entegrasyonu, GNN tabanlı yöntemlerin biyoinformatikte yalnızca yüksek doğruluk sağlayan tahmin araçları olmadığını, aynı zamanda biyolojik mekanizmaların yorumlanmasına katkı sunan güçlü hesaplamalı modeller olduğunu göstermektedir. Günümüzde hastalık-gen ilişki tahmini, ilaç keşfi, protein mühendisliği, çoklu omik entegrasyonu ve biyomedikal bilgi grafikleri gibi birçok araştırma alanında GNN tabanlı yaklaşımlar hızla yaygınlaşmaktadır. Bu eğilim, biyolojik sistemlerin ilişkisel doğası ile GNN mimarilerinin ilişkisel öğrenme kapasitesi arasındaki güçlü uyumun doğal bir sonucudur.

2. Biyoinformatikte Temel Grafik Türleri

Giriş bölümünde biyolojik sistemlerin oluşturan unsurların ilişkisel doğası ile Grafik Sinir Ağları arasındaki ilişki anlatılmıştır. Ancak GNN tabanlı öğrenmenin arkasındaki başarı sadece kullanılan mimariye ait değildir, biyolojik sistemlerin grafik olarak nasıl temsil edildiğine de bağlıdır. Bu amaçla bu bölümde biyoinformatikte yaygın olarak kullanılan temel grafik türleri, matematiksel gösterimleri ve biyolojik anlamları açıklanmıştır. Biyoinformatikte var olan grafik temsilleri, araştırılan biyolojik probleme, ölçeğine bu problemin veri kaynağına ve hedeflenen öğrenme problemine göre ayrışır. Bir biyolojik grafik temelde şu şekilde tanımlanır:

$$G = (V, E) \tag{1}$$

Burada $V = \{v_1, v_2, \ldots\ldots, v_n\}$ biyolojik varlıklardan oluşan düğüm kümesini, $E \subseteq V \times V$ ise düğümler arasındaki ilişkileri tanımlayan kenar kümesini temsil eder. Grafiğin yapısı ise, $A \in \mathbb{R}^{n \times n}$ komşuluk matrisi ile ifade edilir:

$$A_{ij} = \begin{cases} 1, & (v_i, v_j) \in E \\ 0, & (v_i, v_j) \notin E \end{cases} \quad (2)$$

Ağırlıklı grafiklerde A_{ij}, yalnızca bağlantının varlığını değil; ilişkinin gücünü, deneysel güven skorunu, moleküler benzerliği veya etkileşim olasılığını da temsil edebilir. Her düğüm ayrıca gen ekspresyonu, protein dizisi, atom türü veya klinik özellikler gibi tanımlayıcıları içeren:

$$X = [x_1, x_2, \ldots\ldots, x_n]^T \in \mathbb{R}^{n \times d} \quad (3)$$

özellik matrisiyle ilişkilendirilebilir. Burada x_i, v_i düğümün d boyutlu özellik vektörüdür. Biyolojik grafik türleri arasındaki temel farklılıklar; düğüm ve kenarların biyolojik anlamı, grafiğin yönlü veya yönsüz oluşu, kenar ağırlıklarının tanımı ve tek ya da birden fazla varlık türü içermesiyle belirlenmektedir. PPI ağları, gen düzenleyici ağlar, moleküler grafikler ve heterojen bilgi grafikleri biyoinformatikteki başlıca grafik yapıları olarak ele alınmaktadır.

2.1. Protein-Protein Etkileşim Ağları

Hücresel süreçlerin temelini oluşturan fiziksel veya fonksiyonel protein ilişkilerini grafik biçiminde temsil edilmesine Protein-protein etkileşim ağları (Protein-Protein Interaction Networks, PPI) denir. PPI'lerde düğümler proteinlere, kenarlar ise protein arasındaki deneysel olarak gözlemlenen ya da hesaplamalı yöntemlerle tahmin edilen etkileşimlere karşılık gelir. Bir PPI ağı şu şekilde ifade edilir:

$$G_{PPI} = (V_P, E_{PP}) \quad (4)$$

Burada V_P protein kümesini, E_{PP} ise protein çiftleri arasındaki etkileşimleri göstermektedir. İki protein arasındaki etkileşim çoğu uygulamada simetrik kabul edildiğinden, PPI grafikleri yönsüzdür. Kenar ağırlıkları olarak ise deneysel kanıt düzeyi, birlikte ekspresyon yapısal uyumluluk veya veri tabanı güven skoru gibi ölçütlerle atanabilir. PPI ağları protein işlevini sadece protein dizisinden ya da yapısından değil, içinde bulunduğu etkileşim bağlamından çıkarır ve bu yüzden biyolojik olarak oldukça önemli bir yer tutar. PPI ağlarında aynı biyolojik sürecin içinde bulunan proteinler ağ içerisinde de sıklıkla birbirine yakın konumlarda yer alırlar ve fonksiyonel modüller oluştururlar. Bir hastalığın moleküler süreci araştırılırken hastalıkla ilişkili proteinler de

etkileşim ağı üzerinde belirli alt ağlarda kümelenir. Böylelikle oluşan ağ tıbbi yaklaşımlar için temel oluşturur. Ağ tabanlı hastalık araştırmaları, tek bir hastalık genini belirlemekten ziyade, birbirine bağlı moleküllerden oluşan düzensiz biyolojik modülleri tanımlamayı amaçlamaktadır. Bir protein için bağlantı sayısı sadece derece şeklinde ifade edilir:

$$k_i = \sum_{j=1}^{n} A_{ij} \tag{5}$$

Burada derecesi yüksek proteinler ağın merkezi düğümleridir. Ancak sadece derece bilgisi biyolojik işlevi açıklamak için yeterli değildir; ara merkezliği, yakınlık merkeziliği ve topluluk yapısı gibi topolojik ölçütler de kullanılır. Protein fonksiyon tahmini, hastalık geni önceliklendirme, biyolojik yolak analizi, protein komplekslerinin belirlenmesi ve ilaç hedefi keşfi gibi görevlerde PPI ağları sıklıkla kullanılır. Ancak yanlış pozitiflerin bulunması deneysel eksiklikler PPI modellerinin başarısını sınırlamaktadır.

2.2. Gen Düzenleyici Ağlar

Genleri ve gen ekspresyonunu kontrol eden düzenleyici ilişkilerin grafiksel gösterimine Gen düzenleyici ağlar (Gene Regulatory Networks, GRN) denilir. Bu ağlarda düğümler genleri, transkripsiyon faktörlerini veya düzenleyici elementleri; kenarlar ise aktivasyon ve baskılama gibi yönlü düzenleyici etkileri temsil etmektedir. Bir GRN,

$$G_{GRN} = (V_G, E_R) \tag{6}$$

Şeklinde tanımlanır düzenleyici etkinin asimetrik olması nedeniyle GRN'ler çoğunlukla yönlü grafiklerdir. Düzenleyici etkinin büyüklüğü biliniyorsa A_{ij} sürekli değerli bir katsayı olarak kullanılabilir. Zamana bağlı gen ekspresyonu dikkate alındığında basitleştirilmiş bir düzenleyici sistem:

$$\frac{dx_j(t)}{dt} = f_j\left(\sum_{i=1}^{n} A_{ij}\, x_i(t)\right) - \gamma_j x_j(t) \tag{7}$$

Şeklinde modellenir. Burada $x_j(t)$, $j.$ genin t zamanındaki ekspresyon düzeyini; A_{ij}, $i.$ genin $j.$ gen üzerindeki düzenleyici etkisini; $\gamma_j$ise bozunma terimini gösterir. GRN'ler, PPI ağlarından fiziksel birliktelikten çok bilgi akışını

ve yönlü düzenlemeyi temsil etmeleri yönüyle farklılaşırlar. Hücre farklılaşması, gelişimsel süreçler, stres yanıtı, hücrenin durumu ve hastalıkla ilişkili düzenleyici bozulmaların incelenmesi sebebiyle önem taşımaktadır. Ancak gen düzenleyici ilişkiler hücre tipine, dokuya, gelişim evresine ve çevresel koşullara bağlı olarak değişebildiğinden, tek bir sabit GRN çoğu zaman biyolojik gerçekliğin yalnızca belirli bir bağlamını temsil etmektedir.

2.3. Moleküler Grafikler

Moleküler grafikler, küçük molekülleri, ilaç adaylarını, nükleik asitleri veya protein yapılarını atomik ya da kalıntı düzeyinde temsil etmektedir. Küçük molekül grafiğinde düğümler atomları, kenarlar ise kovalent bağları göstermektedir. Her atom düğümüne element türü, formal yük, hibritleşme durumu, aromatiklik, valans ve kiralite gibi özellikler atanabilir. Kenar özellikleri ise bağ türü, bağ derecesi, aromatik bağ durumunu içerebilir. Bir moleküler grafiğin komşuluk matrisi bağ derecelerine bağlı olarak ağırlıklı olarak tanımlanabilir. Moleküler özelliklerin yerel kimyasal ortam, bağ yapısı ve üç boyutlu geometriden kaynaklanması, grafik gösterimlerini moleküler öğrenme açısından doğal bir gösterimi haline getirir. Bu grafikler moleküler özellik tahmini, toksisite ve aktivite kestirimi, ilaç-hedef etkileşim tahmini, protein yapı analizi ve protein tasarımı gibi çeşitli biyolojik görevlerde kullanılır.

2.4. Biyomedikal Bilgi Grafikleri

Biyomedikal bilgi grafikleri, farklı biyolojik varlıkları ve bunlar arasındaki çok çeşitli iliş türlerini inceleyen heterojen bir grafiktir. Bu heterojen grafikte düğümleri genler, proteinler, hastalıklar, ilaçlar, fenotipler, biyolojik yolaklar ve yan etkiler gibi farklı varlık sınıfları oluşturabilir. Matematiksel olarak heterojen bir bilgi grafiği,

$$G_{KG} = (V, E, T_V, T_E) \tag{8}$$

Şeklinde ifade edilir. Burada T_V düğüm türlerine ait kümeyi, T_E ise ilişki türlerine ait kümedir. Her ilişki çoğunlukla bir üçlüyle temsil edilir. Örneğin $(\text{İlaç } A, hedefler, Protein\ B)$, $(Gen\ C, ilişkilidir, Hastalık\ D)$ bunlar birer bilgi grafiği üçlüsüdür. Homojen grafiklerde tek bir komşuluk matrisi yeterliyken, biyomedikal bilgi grafiklerinde her ilişki türü için ayrı bir komşuluk matrisi tanımlanır. Böylelikle "ilaç–hedef", "gen–hastalık", "protein–protein" ve "ilaç–yan etki" ilişkileri aynı sistemde korunurken birbirinden ayrıştırılabilir. Farklı deneysel veri tabanlarını, kontrollü sözlükleri ve

literatürden çıkarılan ilişkileri ortak bir anlamsal yapı altında birleştirilebilmesi bilgi grafiklerinin temel avantajıdır. Bu nedenle ilaç yeniden konumlandırma, hedef keşfi, hastalık mekanizmalarının aydınlatılması, biyomedikal soru yanıtlama ve eksik bağlantıların tahmini gibi uygulamalarda kullanılmaktadır. Bununla birlikte bilgi grafiğindeki her kenarın doğrudan nedensel bir ilişkiyi göstermediği; bazı kenarların literatür ortaklığına, istatistiksel ilişkiye veya farklı güven düzeylerine dayanabileceği göz önünde bulundurulmalıdır.

2.5. Çoklu Omik Grafikler

Genomik, epigenomik, transkriptomik, proteomik, metabolomik ve mikrobiyom gibi veriler, omik verilerini oluşturur. Çoklu omik grafikleri ise omik verilerinde var olan bu birden fazla moleküler katmanı bütünleşik olarak modellemek amacıyla oluşturulmaktadır. Bu grafikler hasta, örnek bazlı ve molekül merkezli ya da çok katmanlı biçimde olabilir. Molekül merkezli çoklu omik grafiklerde düğümler gen, miRNA, protein veya metabolit gibi farklı varlık türlerine karşılık gelebilir. Bu durumda grafik, heterojen ve çok katmanlı bir yapı kazanır. Çoklu omik grafiklerin temel avantajı, her veri katmanını ayrı olarak analiz etmek yerine biyolojik sistemin farklı moleküler düzeyleri arasındaki eşgüdümlü değişimleri modelleyebilmesidir. Bu yapılar kanser alt tipi sınıflandırması, hasta tabakalandırması, biyobelirteç keşfi, sağkalım tahmini ve tedavi yanıtının modellenmesinde giderek daha fazla kullanılmaktadır. Güncel çalışmalar, grafik tabanlı çoklu omik modellerin hem katman içi hem de katmanlar arası ilişkileri doğrudan temsil edebilmesini önemli bir üstünlük olarak değerlendirmektedir. Bu grafiklerin en büyük eksikliği ise grafiklerde kullanılan ilişkilerin çoğu istatistiksel birlikteliğe dayanmasıdır. Oysaki öğrenilen bağlantıların doğrudan moleküler nedensellik olarak yorumlanmaması ve bağımsız deneysel verilerle doğrulanması gereklidir.

Sonuç olarak aynı matematiksel çerçeveyi içermelerine rağmen biyoinformatikte kullanılan grafik türleri farklı biyolojik varsayımlarda çalışırlar. PPI ağları proteinlerin fiziksel ve fonksiyonel birlikteliğini, GRN'ler yönlü düzenleyici bilgi akışını, moleküler grafikler kimyasal bağları ve üç boyutlu yapıyı, bilgi grafikleri heterojen biyomedikal semantiği, çoklu omik grafikler ise farklı moleküler katmanlar arasındaki bütünleşik ilişkileri temsil etmektedir. Bu yüzden hangi grafik türünün seçileceği sadece teknik bir veri ön işleme adımıyla değil, modelin hangi biyolojik ilişkileri öğrenebildiğine ve oluşan sonuçların nasıl yorumlanacağına göre belirlenmelidir.

3. Grafik Sinir Ağlarının Temelleri

3.1. Mesaj İletim Mekanizması

Bir önceki bölümde anlatılan biyolojik sistemlerdeki grafik yapılarında farklı biyolojik problemlere bağlı olarak çeşitli grafik türlerinin oluşturulabildiği açıklanmıştır. Ancak biyolojik verilerin grafik biçiminde modellenmesi, tek başına biyolojik örüntülerin öğrenilebilmesi için yeterli değildir. Grafik üzerinde tanımlanan düğümler, bu düğümlerin durumu ve aralarındaki ilişkilerden anlamlı temsiller elde etmek önemlidir. Bu yüzden bu bilgilerin doğrudan ağ topolojisinin öğrenme sürecine doğrudan dahil edilmesi gerekmektedir. Bu gereksinim doğrultusunda geliştirilen Grafik Sinir Ağları (Graph Neural Networks, GNN), düğümlerin yalnızca kendi özelliklerini değil, komşuluk yapısından elde edilen bilgileri de dikkate alarak öğrenme gerçekleştirmektedir. Günümüzde geliştirilen Grafik Evrişimsel Ağları (GCN), GraphSAGE, Grafik Dikkat Ağları (GAT), Grafik İzomorfizm Ağları (GIN) ve Mesaj İletimli Sinir Ağları (MPNN) gibi birçok mimari farklı hesaplama stratejileri kullanıyor olmasına rağmen ortak bir öğrenme prensibini paylaşmaktadır. Mesaj iletim mekanizması olarak adlandırılan bu yaklaşım, grafik üzerindeki bir bilginin komşu düğümler arasında iteratif olarak aktarılması ve her katmanda düğüm temsillerinin güncellenmesi işlemidir. Mesaj iletiminde temal amaç, her düğümün hem kendi özniteliklerinden öğrenmesi hem de komşu düğümlerden elde edilen bilgileri de kullanarak daha zengin ve biyolojik açıdan daha anlamlı bir temsil oluşturmasını sağlamaktır. Bu süreç iki temel aşamadan oluşmaktadır: mesaj toplama (aggregation) ve düğüm güncelleme (update). İlk aşamada her düğüm, komşularından gelen bilgileri belirli bir toplama fonksiyonu yardımıyla bir araya getirir. Genel mesaj toplama işlemi aşağıdaki şekilde ifade edilmektedir:

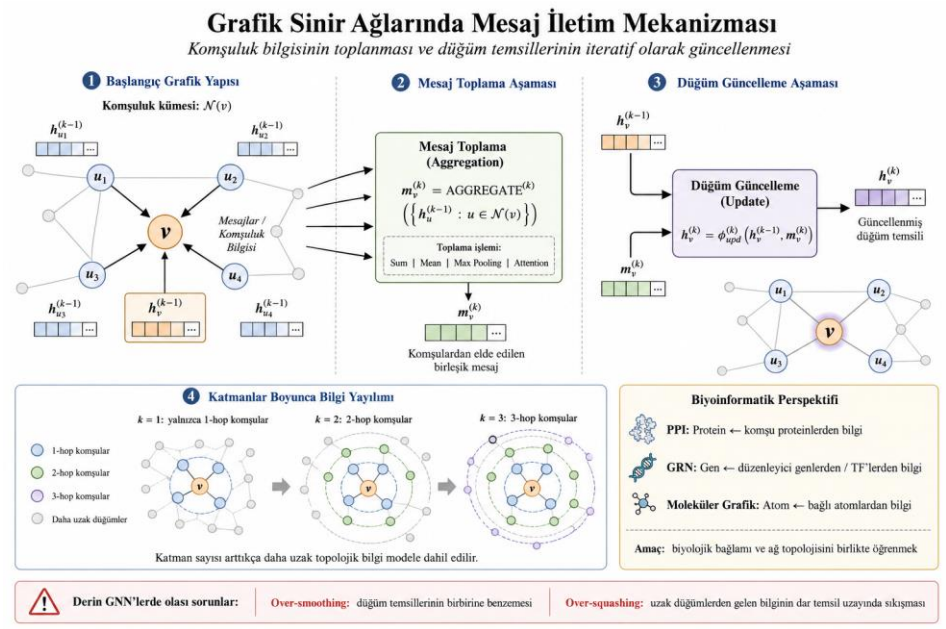
$$m_v^{(k)} = \mathrm{AGGREGATE}^{(k)}\left(\left\{h_u^{(k-1)} : u \in \mathcal{N}(v)\right\}\right) \tag{9}$$

Burada $\boldsymbol{m_v^{(k)}}, k.$ katmanda düğüm v için elde edilen mesaj vektörünü, $\boldsymbol{h_u^{(k-1)}}$ komşu düğümlerin bir önceki katmandaki gizli gösterimlerini, $\mathcal{N}(v)$ ise v'nin komşuluk kümesini ifade etmektedir. AGGREGATE fonksiyonu, komşulardan gelen bilgilerin tek bir temsil altında nasıl birleştirileceğini tanımlamaktadır. Kullanılan GNN mimarisine bağlı olarak bu işlem toplama (sum), ortalama (mean), maksimum seçme (max pooling) veya dikkat (attention) tabanlı ağırlıklandırma yöntemleri kullanılarak gerçekleştirilebilmektedir. Komşulardan elde edilen mesajlar daha sonra

düğümün mevcut gösterimi ile birleştirilerek yeni düğüm temsili oluşturulur. Bu güncelleme işlemi genel olarak aşağıdaki biçimde ifade edilmektedir:

$$h_v^{(k)} = \phi_{\text{upd}}^{(k)}\left(h_v^{(k-1)}, m_v^{(k)}\right) \tag{10}$$

Burada UPDATE fonksiyonu doğrusal dönüşümler, doğrusal olmayan aktivasyon fonksiyonları veya öğrenilebilir sinir ağı katmanları kullanılarak gerçekleştirilmektedir. Böylece her katmanda düğüm gösterimleri yeniden hesaplanmakta ve komşuluk bilgisini de içerecek şekilde giderek daha zengin temsil vektörleri elde edilmektedir. Katman sayısı arttıkça her düğüm daha uzak komşulardan bilgi alabilmekte, dolayısıyla grafik üzerindeki daha geniş yapısal ilişkiler öğrenme sürecine dahil edilmektedir (Şekil 1). Mesaj iletim mekanizması, Grafik Sinir Ağlarının biyolojik veriler üzerinde başarılı olmasının temel nedenlerinden biri olarak kabul edilmektedir. Bununla birlikte katman sayısının gereğinden fazla artırılması bazı yapısal problemlere yol açabilmektedir. Bunlardan ilki aşırı düzleşme (over-smoothing) problemidir. Bu durumda düğüm temsilleri ardışık mesaj iletim işlemleri sonucunda birbirine giderek benzemekte ve farklı biyolojik varlıkları ayırt edebilme yeteneği azalmaktadır. İkinci önemli problem ise aşırı sıkışma (over-squashing) olarak bilinmektedir. Bu problemde ise uzak düğümlerden gelen çok miktardaki bilgi, sınırlı boyuttaki temsil vektörleri içerisine sıkıştırılmaya çalışıldığı için önemli biyolojik ilişkilerin bir kısmı öğrenme sürecinde kaybolabilmektedir. Günümüzde geliştirilen birçok yeni GNN mimarisi, dikkat mekanizmaları, artık bağlantılar (residual connections) ve gelişmiş mesaj iletim stratejileri kullanarak bu problemlerin etkisini azaltmayı amaçlamaktadır.



Şekil 1. Grafik Sinir Ağlarında mesaj iletim mekanizmasının şematik gösterimi.

Şekilde komşu düğümlerden bilgi toplanması (aggregation), düğüm temsillerinin güncellenmesi (update) ve katmanlar boyunca bilginin yayılımı gösterilmektedir. Ayrıca biyoinformatik uygulamalarındaki kullanım alanları ile over-smoothing ve over-squashing problemleri özetlenmektedir.

3.2. Grafik Sinir Ağları Erken Dönem Mimarileri

Grafik yapılarındaki düğümler ve aralarındaki ilişkilerin daha etkin öğrenilebilmesi için farklı mimariler geliştirilmeye başlanmıştır. Günümüze kadar çok fazla GNN modeli önerilmiştir ancak bunların çoğu aynı mesaj iletim mekanizmasını temel almakta ve yalnızca komşuluk bilgisinin nasıl işlendiği konusunda farklılaşmaktadır. Bu bölümde Biyoinformatik çalışmalarında en yaygın kullanılan temel GNN mimarileri yer almaktadır.

Grafik Evrişimsel Ağları (Graph Convolutional Networks, GCN): Kipf ve Welling [10] tarafından önerilmiş olup günümüzde en yaygın kullanılan GNN mimarilerinden biridir. GCN'nin temel amacı, klasik evrişim (convolution) işlemini düzensiz grafik yapıları üzerine uyarlayarak her düğümün komşuluk bilgisinden yararlanmasını sağlamaktır. GCN, biyolojik ağlardaki topolojik bilgiyi doğrudan yansıtabilmesi sebebiyle sıklıkla tercih edilir ancak sadece sabit komşuluk ağırlıklarını kullanması tüm komşuların eşit biyolojik öneme sahip olduğunu varsayar.

GraphSAGE: Hamilton ve arkadaşları tarafından geliştirilen GraphSAGE (Graph Sample and Aggregate), özellikle büyük ölçekli grafiklerde karşılaşılan hesaplama maliyetini azaltmak amacıyla önerilmiştir. GCN'den farklı olarak tüm komşular yerine rastgele örneklenen komşular üzerinden öğrenme gerçekleştirilmektedir. Bu durumundan dolayı uzak bağlantıları da ağa dahil edebilir ve milyonlarca düğüm içeren protein etkileşim ağları, biyomedikal bilgi grafiklerinde kullanılır.

Grafik Dikkat Ağları (Graph Attention Networks, GAT): Veličković ve arkadaşları [12] tarafından geliştirilmiştir. Komşu düğümlerin öğrenmeye olan katkısını dikkat mekanizmasıyla belirler. Böylelikle tüm komşular eşit ağırlıkla değerlendirilmez ve biyolojik olarak daha önemli komşular modele daha fazla katkı sağlar. Bu mekanizmasıyla biyoinformatikte GAT'lar birçok problemin çözümünde sıklıkla tercih edilir. GAT mimarisi, gen düzenleyici ağlar ve ilaç-hedef etkileşimleri gibi tüm bağlantıların aynı biyolojik öneme sahip olmadığı problemlerde başarılı sonuçlar vermektedir.

Grafik İzomorfizm Ağları (Graph Isomorphism Networks, GIN): GIN, Xu ve arkadaşları [16] tarafından geliştirilmiştir grafik yapılarının ayırt ediciliğini arttırmayı amaçlar. Temel mekanizması, komşu düğüm özelliklerini toplama (sum aggregation) işlemiyle birleştirerek daha güçlü düğüm temsilleri oluşturmaktır. GIN özellikle moleküler grafiklerde izomer ayrımı, molekül sınıflandırması ve kimyasal özellik tahmini gibi görevlerde güçlü temsil öğrenmesi sağlamaktadır.

Mesaj İletimli Sinir Ağları (Message Passing Neural Networks, MPNN): Bölüm başlangıcında anlatıldığı gibi MPNN, birçok modern Grafik Sinir Ağı mimarisinin temelini oluşturan genel bir öğrenme çerçevesidir. Bu yaklaşımda düğümler arasındaki bilgi aktarımı mesaj fonksiyonları aracılığıyla gerçekleştirilirken, kenar özellikleri de öğrenme sürecine dahil edilir. MPNN'ler atomlar arasındaki bağ özelliklerini doğrudan kullanabilmeleri sebebiyle molekül özellik tahmini, ilaç keşfi ve protein tasarımı çalışmalarında yaygın olarak kullanılmaktadır.

3.3. Güncel Grafik Sinir Ağı Yaklaşımları

Grafik Sinir Ağlarının biyoinformatikte sıklıkla kullanılmasıyla birlikte, klasik GCN ve GraphSAGE gibi ilk nesil mimarilerin sınırlılıkları göze çarpmaya başlamıştır. Özellikle heterojen biyolojik ağların modellenmesi, uzun menzilli düğüm ilişkilerinin öğrenilmesi, etiketlenmiş verilerin sınırlı olması ve biyomoleküllerin üç boyutlu yapılarının dikkate alınması gibi gereksinimler doğrultusunda yeni nesil GNN yaklaşımları geliştirilmiştir. Bu bölümde biyoinformatikte öne çıkan güncel GNN mimarileri kısaca özetlenmektedir.

Heterojen Grafik Sinir Ağları: Klasik GNN modelleri genellikle tek tip düğüm ve kenardan oluşan homojen grafik yapılarıdır. Ancak biyolojik sistemler genler, proteinler, ilaçlar ve hastalıklar gibi farklı biyolojik varlıklardan oluşur. Heterojen Grafik Sinir Ağları (Heterogeneous Graph Neural Networks, HGNN), farklı düğüm ve ilişki tiplerini ayrı ayrı modelleyerek bu yapıları daha gerçekçi biçimde temsil etmektedir:

$$G = (V, E, \tau, \phi) \tag{11}$$

Burada, $\tau(\mathrm{v})$ düğüm tipini, $\phi(\mathrm{e})$ ise ilişki tipini göstermektedir. Özellikle Heterogeneous Graph Transformer (HGT) gibi modeller, farklı ilişki türleri için ayrı dikkat mekanizmaları kullanarak daha başarılı temsil öğrenimi gerçekleştirmektedir. Biyoinformatik uygulamalarda HGNN'ler biyomedikal bilgi grafikleri, ilaç yeniden konumlandırma ve çoklu omik veri entegrasyonu gibi heterojen biyolojik problemlerde sıklıkla kullanılır.

Grafik Dönüştürücü Mimarileri: Transformer mimarilerinin başarısı grafik verilerine de uygulandığında Graph Transformer modelleri oluşturulur. Klasik GNN'lerde bilgi çoğunlukla komşu düğümler arasında aktarılırken, Graph Transformer modelleri dikkat mekanizması sayesinde uzak düğümler arasındaki ilişkileri de öğrenebilmektedir.

$$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V \tag{12}$$

Dikkat mekanizması özellikle büyük ve karmaşık biyolojik ağlarda uzun menzilli ilişkilerin modellenmesini kolaylaştırır.

Kendinden Denetimli Grafik Öğrenmesi: Biyoinformatikte özellikle etiketlenmiş verinin nispeten az olduğu uygulamalarda tercih edilmektedir. Kendinden denetimli öğrenme (Self-supervised Graph Learning) yönteminde model, etiket bilgisine ihtiyaç duymadan grafik yapısından temsil öğrenmektedir. GraphCL, Deep Graph Infomax (DGI), GraphMAE ve BGRL bu alandaki önemli yaklaşımlar arasında yer almaktadır [17,18,19,20].

Geometrik Grafik Sinir Ağları: Proteinler ve küçük moleküller gibi biyolojik yapılar yalnızca bağlantılarıyla değil, üç boyutlu geometrileriyle de karakterize edilmektedir. Geometrik Grafik Sinir Ağları (Geometric GNN), mesaj iletim sürecine uzaysal koordinat bilgisini de dahil ederek bu yapıları daha doğru modellemektedir.

Sonuç olarak Grafik Sinir Ağları biyolojik problemin karmaşıklığına göre farklı yaklaşımlar sunan zengin bir mimari ailesine sahiptir bu nedenle de

biyoinformatik uygulamalarında da sıklıkla tercih edilmeye başlanmıştır. Temel GNN modelleri biyolojik ağların topolojik yapısını başarıyla öğrenirken, güncel yaklaşımlar heterojen veri yapıları, uzun menzilli ilişkiler, sınırlı etiketli veri ve üç boyutlu biyomoleküler yapılar gibi daha karmaşık problemlere çözüm üretmektedir. Bu yüzden kullanılan GNN mimarisi seçimi hem tahmin performansına hem de biyolojik ilişkileri doğru temsil ederek yorumlanabilirliğine göre belirlenmelidir.

4. Biyoinformatikte Grafik Sinir Ağı Uygulamaları

Bir önceki bölümde Grafik Sinir Ağlarının temel çalışma mantığı ile biyolojik problemlerin özelliklerine göre geliştirilen çeşitli GNN mimarileri ele alınmıştır. Bununla birlikte, bu mimarilerin gerçek değeri sadece kuramsal özellikleriyle değil, biyoinformatikte karşılaşılan ve bu alanda kullanılan karmaşık problemlerin çözümüne sağladıkları katkılarla ortaya çıkmaktadır. Protein yapısı ve fonksiyon tahmini, ilaç keşfi, hastalık biyolojisi, çoklu omik veri entegrasyonu ve biyomedikal bilgi grafikleri gibi birçok uygulama alanında GNN tabanlı yöntemler, biyolojik varlıklar arasındaki ilişkisel yapıyı öğrenebilme yetenekleri sayesinde önemli başarılar elde etmiştir. Bu bölümde, Grafik Sinir Ağlarının biyoinformatikte öne çıkan başlıca uygulama alanları ve bu problemlerde yaygın olarak kullanılan temsili modeller özetlenmektedir. GNN'ler, biyolojik sistemlerin doğal olarak ağ yapısında organize olması nedeniyle son yıllarda biyoinformatik araştırmalarında en yaygın kullanılan derin öğrenme yaklaşımlarından biri hâline gelmiştir. Genler, proteinler, ilaçlar, hastalıklar ve biyolojik yolaklar arasındaki karmaşık ilişkilerin grafik yapıları ile temsil edilebilmesi, GNN tabanlı yöntemlerin yalnızca düğümlerin özelliklerini değil, aynı zamanda biyolojik sistemlerin topolojik organizasyonunu da öğrenebilmesini sağlamaktadır. Bu özellik, protein fonksiyon tahmini, ilaç keşfi, hastalık biyolojisi, çoklu omik veri entegrasyonu ve tek hücre analizleri gibi çok sayıda biyoinformatik probleminde klasik makine öğrenmesi ve geleneksel derin öğrenme yöntemlerine göre önemli avantajlar sunmaktadır. Günümüzde geliştirilen birçok biyoinformatik aracı, araştırılan biyolojik problemin yapısına bağlı olarak GCN, GAT, MPNN, Heterogeneous GNN ve Geometric GNN gibi farklı grafik sinir ağı mimarilerinden yararlanmaktadır. Tablo 1'de, Grafik Sinir Ağlarının biyoinformatikte öne çıkan uygulama alanları ile bu alanlarda yaygın olarak kullanılan temsili model ve araçlar özetlenmiştir. Tablo, farklı biyolojik problemlerin çözümünde tercih edilen GNN mimarilerinin ve temel kullanım amaçlarının genel bir görünümünü sunmaktadır.

Tablo 1 Biyoinformatikte Grafik Sinir Ağlarının Başlıca Uygulama Alanları ve Temsili Modeller

Uygulama alanı	Biyoinformatik problem	Temsili model	GNN yaklaşımı
Protein biyolojisi	Protein fonksiyon tahmini	**DeepFRI [14]**	GCN
	Protein yapı analizi	**GVP-GNN [21]**	Geometrik GNN
	Yapı-temelli protein tasarımı	**ProteinMPNN [22]**	Mesaj iletimli GNN
	Ters katlama (yapı → dizi)	**ESM-IF1 [23]**	Geometrik GNN
İlaç keşfi	Moleküler özellik tahmini	**AttentiveFP [24]**	Dikkat tabanlı GNN
	İlaç-hedef etkileşimi	**GraphDTA [25]**	GCN
	İlaç yeniden konumlandırma	**Decagon [26]**	Çok ilişkili GCN
	Yapı-temelli sanal tarama	**SIGN [27]**	Geometrik/etkileşim tabanlı GNN
Hastalık biyolojisi ve çoklu omik	Hastalık sınıflandırması ve çoklu omik entegrasyonu	**MOGONET [15]**	GCN
	Hastalık–gen–ilaç ilişkilerinin modellenmesi	**PrimeKG / [28]Hetionet [29]**	Heterojen bilgi grafiği*
	Tek hücre analizi	**scGNN [30]**	Grafik otoenkoderi
Uzamsal omik	Uzamsal transkriptomik analizi	**SpaGCN [31]**	GCN

Tablo 1 incelendiğinde, Grafik Sinir Ağlarının biyoinformatikte oldukça geniş bir uygulama yelpazesine sahip olduğu görülmektedir. Bununla birlikte, farklı biyolojik problemlerin çözümünde tek bir GNN mimarisi yerine problemin doğasına uygun modeller tercih edilmektedir. Örneğin protein yapı analizi ve moleküler modelleme çalışmalarında üç boyutlu uzaysal bilgiyi doğrudan işleyebilen Geometric GNN tabanlı yaklaşımlar öne çıkarken, ilaç-hedef etkileşimleri ve moleküler özellik tahmini çalışmalarında MPNN ve GAT tabanlı modeller daha yaygın olarak kullanılmaktadır. Benzer şekilde, gen, protein, ilaç ve hastalık gibi farklı biyolojik varlıkların birlikte analiz edildiği bilgi grafikleri ve çoklu omik entegrasyonu çalışmalarında ise Heterogeneous GNN mimarileri önemli avantajlar sağlamaktadır.

5. Değerlendirme, Gelecek Perspektifi ve Sonuç

5.1. Avantaj ve Sınırlılıklar

Grafik Sinir Ağlarının birçok avantajı vardır ancak bunlardan en önemlisi, biyolojik sistemlerin ilişkisel doğasını doğrudan modelleyebilmesidir. Düğümlerin yalnızca kendi özelliklerini değil, komşuluk yapısını ve ağ topolojisini de öğrenme sürecine dahil etmesi sayesinde klasik makine öğrenmesi yöntemlerinin göz ardı ettiği ve modellemekte başarısız kaldığı biyolojik ilişkiler, GNN ile temsil öğrenimine aktarılabilmektedir. Özellikle protein fonksiyon tahmini, ilaç-hedef etkileşimlerinin modellenmesi ve çoklu omik veri entegrasyonu gibi karmaşık biyolojik problemlerde bu özellik önemli performans artışları sağlamaktadır [32]. Ayrıca farklı veri kaynaklarının tek bir grafik yapısı altında bütünleştirilebilmesi, GNN tabanlı modelleri

biyoinformatik açısından oldukça esnek ve güçlü bir öğrenme yöntemi haline getirir [33].

Tüm bu avantajlarının yanı sıra GNN modelleri çeşitli sınırlılıklara da sahiptir. Bunlardan ilki yorumlanabilirlik (Explainability) problemidir. Derin öğrenme tabanlı diğer yöntemlerde olduğu gibi GNN modelleri de çoğu zaman "kara kutu" olarak değerlendirilmekte ve modelin belirli bir biyolojik kararı hangi düğüm veya hangi ilişkiyi temel alarak verdiğini açıklamak güçleşmektedir [34]. Bu durum biyoinformatik problemlerde özellikle klinik karar destek sistemleri ve hassas tıp gibi uygulamalarında önemli bir güvenilirlik problemi oluşturmaktadır. Bir diğer önemli sınırlılığı ise aşırı düzleşme (Over-smoothing) olarak söyelenebilir. Katman sayısı arttıkça mesaj iletim mekanizması nedeniyle düğüm temsilleri giderek birbirine benzemekte ve farklı biyolojik varlıkların ayırt edilebilirliği azalmaktadır [35]. Özellikle büyük biyolojik ağlarda bu durum model performansını olumsuz etkileyebilmektedir. Grafik Sinir Ağlarının karşılaştığı bir diğer zorluk ölçeklenebilirlik (Scalability) problemidir [36]. Günümüzde biyomedikal bilgi grafikleri ve protein etkileşim ağları milyonlarca düğüm ve kenar içerebilmektedir. Bu büyüklükteki grafiklerin eğitimi yüksek bellek gereksinimi ve uzun eğitim süreleri nedeniyle önemli hesaplama maliyetleri oluşturmaktadır. Bu nedenle büyük ölçekli biyolojik ağlar üzerinde daha verimli örnekleme ve mini-batch tabanlı öğrenme stratejileri geliştirilmeye devam etmektedir.

Son olarak, GNN modellerinin başarısı büyük ölçüde oluşturulan grafiğin kalitesine bağlıdır. Grafik kalitesi (Graph Quality); düğümlerin doğru tanımlanması, kenarların biyolojik olarak anlamlı ilişkileri temsil etmesi ve kullanılan veri tabanlarının güvenilirliği ile doğrudan ilişkilidir. Eksik etkileşimler, yanlış pozitif bağlantılar veya biyolojik açıdan zayıf ilişkiler, öğrenilen düğüm temsillerini olumsuz etkileyerek model performansını düşürebilmektedir. Bu nedenle grafik oluşturma süreci, GNN tabanlı biyoinformatik çalışmalarının en önemli aşamasıdır denilebilir.

5.2. Gelecek Araştırma Alanları

Grafik Sinir Ağları alanında son yıllarda yaşanan gelişmelerden biri biyoinformatikte daha güçlü ve biyolojik açıdan daha güvenilir modellerin geliştirilmesine de olanak tanıyan Graph Transformer tabanlı yaklaşımlardır. Dikkat mekanizmasını grafik öğrenmesine uyarlayan bu modeller, yalnızca yerel komşuluk bilgisine bakmaz aynı zamanda grafik üzerindeki uzak düğümler arasındaki ilişkileri de öğrenir ve böylelikle özellikle büyük ölçekli biyolojik ağlarda umut verici sonuçlar verebilmektedir. Bir diğer önemli araştırma yönü Geometric Deep Learning yaklaşımıdır. Proteinler, nükleik

asitler ve küçük moleküller gibi biyolojik yapılar üç boyutlu uzaysal organizasyona sahip olduğundan, sadece ağ topolojisine bakarak değil, geometrik özellikleri de göz önüne alarak öğrenme sürecine dahil edilmesi büyük önem taşımaktadır. Özellikle protein tasarımı, yapı tahmini ve moleküler modelleme çalışmalarında geometrik GNN mimarilerinin önümüzdeki yıllarda daha yaygın kullanılması beklenmektedir. Son yıllarda dikkat çeken bir diğer yaklaşım ise Nedensel Grafik Sinir Ağları (Causal GNN) olarak öne çıkmaktadır. Günümüzde birçok GNN modeli düğümler arasındaki istatistiksel ilişkileri öğrenirken, biyolojik süreçlerin temelini oluşturan nedensel mekanizmaları doğrudan modelleyememektedir. Causal GNN yaklaşımları ise korelasyonun ötesine geçerek biyolojik sistemlerde neden-sonuç ilişkilerini ortaya koymayı hedeflemektedir. Bu durum özellikle hastalık mekanizmalarının açıklanması ve ilaç hedeflerinin belirlenmesi açısından önemli bir araştırma alanı oluşturmaktadır. Biyoinformatikte gelecekte öne çıkması beklenen bir diğer konu ise Foundation Models yaklaşımıdır. Büyük ölçekli biyolojik veri kümeleri üzerinde ön eğitim alan genel amaçlı modeller, farklı biyoinformatik problemlerine yeniden uyarlanabilmekte ve sınırlı etiketli veri bulunan uygulamalarda önemli avantajlar sağlamaktadır. Her ne kadar bu modeller henüz gelişim aşamasında olsa da protein dizileri, moleküler grafikler ve biyomedikal bilgi grafikleri üzerinde eğitilen temel modellerin önümüzdeki yıllarda GNN tabanlı biyoinformatik çalışmalarında önemli bir yer edinmesi beklenmektedir. Son olarak gelişme ise, Açıklanabilir Yapay Zeka (Explainable Artificial Intelligence, XAI) yaklaşımlarının GNN araştırmalarının temel odak noktalarından biri haline getirilmesidir. Özellikle GNNExplainer, PGExplainer ve benzeri yöntemler, model kararlarını etkileyen düğüm ve kenarların belirlenmesine olanak sağlayarak biyolojik yorumlanabilirliği artırmayı amaçlar. Özellikle klinik uygulamalarda güvenilir bir yapay zeka sistemigeliştirilmesi açısından açıklanabilirlik, gelecekte üzerinde en fazla çalışılacak araştırma alanlarından biri olarak değerlendirilmektedir.

5.3. Sonuç

Grafik Sinir Ağları, biyolojik sistemlerin ilişkisel yapısını doğrudan modelleyebilme yetenekleri sayesinde biyoinformatikte kullanılmaya başlanmış ve son zamanlarda yeni bir araştırma paradigması oluşturmuştur. Protein biyolojisi, ilaç keşfi, hastalık biyolojisi, çoklu omik veri entegrasyonu ve biyomedikal bilgi grafikleri gibi farklı uygulama alanlarında elde edilen başarılı sonuçlar, GNN tabanlı yöntemlerin sadece yüksek tahmin performansı sağlamasıyla değil, aynı zamanda biyolojik süreçlerin daha kapsamlı biçimde anlaşılmasına katkı sağlayan güçlü hesaplamalı araçlar olduğunu

göstermektedir. Bununla birlikte yorumlanabilirlik, ölçeklenebilirlik, grafik kalitesi ve model güvenilirliği gibi mevcut sınırlılıkların giderilmesi, bu modellerin klinik uygulamalara daha etkin biçimde aktarılabilmesi açısından kritik öneme sahiptir. Gelecekte Graph Transformer, Geometric Deep Learning, Causal GNN, Foundation Models ve Explainable AI gibi yeni yaklaşımların gelişmesiyle birlikte grafik tabanlı derin öğrenme yöntemlerinin biyoinformatik araştırmalarında daha etkin bir rol üstlenmesi ve hassas tıp başta olmak üzere birçok biyomedikal uygulamada standart hesaplama araçları arasında yer alması beklenmektedir.

Kaynakça

[1] Baykal, P.I., Łabaj, P.P., Markowetz, F. et al. Genomic reproducibility in the bioinformatics era. Genome Biol 25, 213 (2024). https://doi.org/10.1186/s13059-024-03343-2

[2] Satam, H., Joshi, K., Mangrolia, U., Waghoo, S., Zaidi, G., Rawool, S., Thakare, R. P., Banday, S., Mishra, A. K., Das, G., & Malonia, S. K. (2023). Next-Generation Sequencing Technology: Current Trends and Advancements. Biology, 12(7), 997. https://doi.org/10.3390/biology12070997

[3] Baysoy, A., Bai, Z., Satija, R., & Fan, R. (2023). The technological landscape and applications of single-cell multi-omics. Nature reviews. Molecular cell biology, 24(10), 695–713. https://doi.org/10.1038/s41580-023-00615-w

[4] Yadav, D., Patil-Takbhate, B., Khandagale, A., Bhawalkar, J., Tripathy, S., & Khopkar-Kale, P. (2023). Next-Generation sequencing transforming clinical practice and precision medicine. Clinica chimica acta; international journal of clinical chemistry, 551, 117568. https://doi.org/10.1016/j.cca.2023.117568

[5] Wekesa, J. S., & Kimwele, M. (2023). A review of multi-omics data integration through deep learning approaches for disease diagnosis, prognosis, and treatment. Frontiers in genetics, 14, 1199087. https://doi.org/10.3389/fgene.2023.1199087

[6] Flores JE, Claborne DM, Weller ZD, Webb-Robertson B-JM, Waters KM and Bramer LM (2023) Missing data in multi-omics integration: Recent advances through artificial intelligence. Front. Artif. Intell. 6:1098308. doi: 10.3389/frai.2023.1098308

[7] Ng, S., Masarone, S., Watson, D. et al. The benefits and pitfalls of machine learning for biomarker discovery. Cell Tissue Res 394, 17–31 (2023). https://doi.org/10.1007/s00441-023-03816-z

[8] Iqbal, S., N. Qureshi, A., Li, J. et al. On the Analyses of Medical Images Using Traditional Machine Learning Techniques and Convolutional Neural Networks. Arch Computat Methods Eng 30, 3173–3233 (2023). https://doi.org/10.1007/s11831-023-09899-9

[9] Muzio, G., O'Bray, L., & Borgwardt, K. (2021). Biological network analysis with deep learning. Briefings in bioinformatics, 22(2), 1515–1530. https://doi.org/10.1093/bib/bbaa257

[10] Kipf, T. N., & Welling, M. (2017). Semi-Supervised Classification with Graph Convolutional Networks. In Proceedings of the 5th International Conference on Learning Representations (ICLR 2017), Toulon, France.

[11] Hamilton, W. L., Ying, Z., & Leskovec, J. (2017). Inductive Representation Learning on Large Graphs. Advances in Neural Information Processing Systems (NeurIPS 2017), 30, 1024–1034.

[12] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2018). Graph Attention Networks. International Conference on Learning Representations (ICLR 2018).

[13] Jumper, J., Evans, R., Pritzel, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

[14] Gligorijević, V., Renfrew, P. D., Kosciolek, T., et al. (2021). Structure-based protein function prediction using graph convolutional networks. Nature Communications, 12, 3168. https://doi.org/10.1038/s41467-021-23303-9

[15] Wang, T., Shao, W., Huang, Z., Tang, H., Zhang, J., Ding, Z., & Huang, K. (2021). MOGONET integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification. Nature communications, 12(1), 3445. https://doi.org/10.1038/s41467-021-23774-w

[16] Xu, K., Hu, W., Leskovec, J., & Jegelka, S. (2019). How Powerful are Graph Neural Networks? International Conference on Learning Representations (ICLR 2019).

[17] You, Y., Chen, T., Sui, Y., Chen, T., Wang, Z., & Shen, Y. (2020). Graph Contrastive Learning with Augmentations. Advances in Neural Information Processing Systems (NeurIPS 2020), 33, 5812–5823.

[18] Veličković, P., Fedus, W., Hamilton, W. L., Liò, P., Bengio, Y., & Hjelm, R. D. (2019). Deep Graph Infomax. International Conference on Learning Representations (ICLR 2019).

[19] Hou, Z., Liu, X., Cen, Y., Dong, Y., Yang, H., Wang, C., & Tang, J. (2022). GraphMAE: Self-Supervised Masked Graph Autoencoders. Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2022).

[20] Thakoor, S., Tallec, C., Azar, M. G., Azabou, M., Dyer, E., & Valko, M. (2022). Bootstrapped Representation Learning on Graphs. International Conference on Learning Representations (ICLR 2022).

[21] Jing, B., Eismann, S., Suriana, P., Townshend, R. J. L., & Dror, R. O. (2021). Learning from protein structure with geometric vector perceptrons. International Conference on Learning Representations (ICLR).

[22] Dauparas, J., Anishchenko, I., Bennett, N., Bai, H., Ragotte, R. J., Milles, L. F., Wicky, B. I. M., Courbet, A., de Haas, R. J., Bethel, N., Leung, P. J. Y., Huddy, T. F., Pellock, S., Tischer, D., Chan, F., Koepnick, B., Nguyen, H., Kang, A., Sankaran, B., Bera, A. K., King, N. P., & Baker, D. (2022). Robust deep learning–based protein sequence design using ProteinMPNN. Science, 378(6615), 49–56. https://doi.org/10.1126/science.add2187

[23] Hsu, C., Verkuil, R., Liu, J., Lin, Z., Hie, B., Sercu, T., Lerer, A., & Rives, A. (2022). Learning inverse folding from millions of predicted structures. Proceedings of the 39th International Conference on Machine Learning, 162, 8946–8970.

[24] Xiong, Z., Wang, D., Liu, X., Zhong, F., Wan, X., Li, X., Li, Z., Luo, X., Chen, K., Jiang, H., & Zheng, M. (2020). Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. Journal of Medicinal Chemistry, 63(16), 8749–8760.

[25] Nguyen, T., Le, H., Quinn, T. P., Nguyen, T., Le, T. D., & Venkatesh, S. (2021). GraphDTA: Predicting drug–target binding affinity with graph neural networks. Bioinformatics, 37(8), 1140–1147. https://doi.org/10.1093/bioinformatics/btaa921

[26] Zitnik, M., Agrawal, M., & Leskovec, J. (2018). Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics, 34(13), i457–i466. https://doi.org/10.1093/bioinformatics/bty294

[27] Li, S., Zhou, J., Xu, T., Huang, L., Wang, F., Xiong, H., Huang, W., Dou, D., & Xiong, H. (2021). Structure-aware interactive graph neural networks for the prediction of protein-ligand binding affinity. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (pp. 975–985). Association for Computing Machinery. https://doi.org/10.1145/3447548.3467311

[28] Chandak, P., Huang, K., & Zitnik, M. (2023). Building a knowledge graph to enable precision medicine. Scientific Data, 10, 67. https://doi.org/10.1038/s41597-023-01960-3

[29] Himmelstein, D. S., Lizee, A., Hessler, C., Brueggeman, L., Chen, S. L., Hadley, D., Green, A., Khankhanian, P., & Baranzini, S. E. (2017). Systematic integration of biomedical knowledge prioritizes drugs for repurposing. eLife, 6, e26726. https://doi.org/10.7554/eLife.26726

[30] Wang, J., Ma, A., Chang, Y., Gong, J., Jiang, Y., Qi, R., Wang, C., Fu, H., Ma, Q., & Xu, D. (2021). scGNN is a novel graph neural network framework for single-cell RNA-Seq analyses. Nature Communications, 12, 1882. https://doi.org/10.1038/s41467-021-22197-x

[31] Hu, J., Li, X., Coleman, K., Schroeder, A., Ma, N., Irwin, D. J., Lee, E. B., Shinohara, R. T., & Li, M. (2021). SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. Nature Methods, 18(11), 1342–1351. https://doi.org/10.1038/s41592-021-01255-8

[32] Cheng-Pei Lin, Yann-Jen Ho, Yen-Peng Chiu, Yun Tang, You Sheng Paik, Guan-Ting Chen, Wei-Chih Huang, Tzong-Yi Lee, MoAGNN: a multi-omics hierarchical graph neural network for subtype classification and prognosis prediction in lung adenocarcinoma, Briefings in Bioinformatics, Volume 27, Issue 1, January 2026, bbaf735, https://doi.org/10.1093/bib/bbaf735

[33] Valous, N.A., Popp, F., Zörnig, I. et al. Graph machine learning for integrated multi-omics analysis. Br J Cancer 131, 205–211 (2024). https://doi.org/10.1038/s41416-024-02706-7

[34] Yuan, H., Yu, H., Gui, S., & Ji, S. (2023). Explainability in Graph Neural Networks: A Taxonomic Survey. IEEE transactions on pattern analysis and machine intelligence, 45(5), 5782–5799. https://doi.org/10.1109/TPAMI.2022.3204236

[35] Li, Q., Han, Z., & Wu, X.-M. (2018). Deeper Insights into Graph Convolutional Networks for Semi-Supervised Learning. Proceedings of the AAAI Conference on Artificial Intelligence, 32(1).

[36] Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., & Sun, M. (2020). Graph Neural Networks: A Review of Methods and Applications. AI Open, 1, 57–81.

3. Bölüm

Kalıcı Mıknatıslı Senkron Makinelerde Parametre Tahmini: Sistem Tanımlanabilirliği ve Metasezgisel Optimizasyon Yaklaşımları

Ertuğrul ATEŞ[1]

Özet

Bu bölüm, kalıcı mıknatıslı senkron motorların (KMSM) değişken hız kontrolü altında çoklu parametre tahminini ele almaktadır. Geleneksel parametre tahmin yöntemlerinin sıralı oran yetmezliği (rank-deficiency) sorunu ve gerilim kaynaklı evirici (VSI) doğrusal olmama etkisinin parametre doğruluğuna etkileri ayrıntılı biçimde incelenmektedir. Liu ve Zhu (2015) tarafından önerilen Kuantum Genetik Algoritma (QGA) tabanlı çerçeve esas alınarak, sistem tanımlanabilirliği sağlayan tam sıralı (full-rank) bir tahmin modelinin nasıl kurulduğu açıklanmaktadır. Ayrıca bu yöntemin diğer evrimsel algoritmalarla (PSO, DE, ACO) karşılaştırmalı değerlendirmesi sunulmakta ve güncel literatürdeki gelişmeler tartışılmaktadır.

[1] Dr. Öğr. Üyesi, Veri Bilimi ve Analitiği Bölümü, Kayseri Üniversitesi, ORCID: 0000-0001-6628-5350

1. Giriş

Kalıcı mıknatıslı senkron makineler (KMSM), yüksek güç yoğunluğu, geniş hız aralığı ve yüksek verimi nedeniyle servo sürücüler, rüzgar enerji türbinleri, elektrikli araçlar ve havacılık uygulamalarında giderek daha yaygın bir kullanım alanı bulmaktadır. Bu uygulamalarda stator sargı direnci (R_s), d ve q ekseni endüktansları (L_d, L_q) ve rotor mıknatıs akı değeri (λ_{pm}) gibi parametrelerin güvenilir biçimde bilinmesi hem kontrol algoritmasının performansı hem de durum izleme sistemlerinin etkinliği açısından kritik öneme sahiptir.

Parametre tahmini alanındaki zorluğun özü, motorun kararlı durum denklemlerinin sıra sayısının iki olmasına karşın dört bilinmeyeni ($R_s, L_d, L_q, \lambda_{pm}$) barındırmasından kaynaklanmaktadır. Bu durum doğrudan bir sıralı oran yetmezliği sorununa yol açmakta; doğrusal olmayan optimizasyon algoritmalarının bile sapma veya yanlış noktaya yakınsama göstermesine neden olmaktadır. Öte yandan gerçek bir sürücü sisteminde gerilim kaynaklı evirici (Voltage Source Inverter, VSI) anahtarlama ölü zamanı (dead-time) ve güç yarı iletkenlerinin iletim düşümleri, referans gerilim sinyallerine ilave bozucu bir voltaj terimi (V_{dead}) eklemektedir. Bu bozucu etki ihmal edildiğinde, stator direnci ve rotor akı bağının tahmini ciddi ölçüde hata içermektedir.

Bu bölüm, sistem tanımlanabilirliği ve gerilim kaynaklı eviricilerden kaynaklı doğrusal olmama etkisi gibi iki temel sorunu eş zamanlı olarak ele alan Liu ve Zhu (2015) tarafından yayınlanmış Kuantum Genetik Algoritma (QGA) tabanlı tahmin çerçevesini merkeze alarak PMSM parametre tahmin yöntemlerinin geniş bir perspektiften incelenmesini amaçlamaktadır. Ayrıca Genetik Algoritma (GA), Parçacık Sürüsü Optimizasyonu (PSO), Diferansiyel Evrim (DE) ve Karınca Kolonisi Optimizasyonu (ACO) gibi diğer metasezgisel yöntemlerle karşılaştırmalı bir çerçeve sunulmaktadır.

2. KMSM Parametre Tahmininin Temelleri

2.1. *dq*-Ekseni Durum Modeli

Vektör kontrolünde yaygın biçimde kullanılan *dq*-ekseni referans eksenindeki KMSM durum denklemi aşağıdaki şekilde ifade edilmektedir.

$$\begin{aligned} \frac{di_d}{dt} &= -\left(\frac{R_s}{L_d}\right) i_d + \left(\frac{L_q}{L_d}\right) \omega_e i_q + \frac{u_d}{L_d} \\ \frac{di_q}{dt} &= -\left(\frac{R_s}{L_d}\right) i_q - \left(\frac{L_d}{L_q}\right) \omega_e i_d + \frac{u_q}{L_q} - \frac{\omega_e \lambda_{pm}}{L_q} \end{aligned} \tag{1}$$

Kararlı durumda (steady-state) bu denklemler Euler formülü ile ayrık zamanlı hale getirilir ve aşağıdaki doğrusal biçime indirgenebilir.

$$\begin{aligned} u_d[k] &= R_s i_d[k] - L_q \omega_e[k] i_q[k] \\ u_q[k] &= R_s i_q[k] + L_d \omega_e[k] i_d[k] + \lambda_{pm} \omega_e[k] \end{aligned} \tag{2}$$

Burada k, sürücü sistemindeki ayrık örnekleme anlarının indeksidir. Bu denklem sisteminin katsayılar matrisi incelendiğinde, dört bilinmeyen parametreye $(R_s, L_d, L_q, \lambda_{pm})$ karşı sıra sayısının yalnızca iki olduğu görülmektedir. Dolayısıyla bu parametreler eş zamanlı tahmin edilmek istendiğinde sonsuz sayıda çözümün bulunduğu sıralı oran yetmezliği problemi ortaya çıkmaktadır.

2.2. Sıralı Oran Yetmezliği ve Tanımlanabilirlik

Sistem tanımlanabilirliği (identifiability), tahmin modelinin belirli bir çalışma koşulunda biricik çözüme sahip olup olmadığını ifade eder. Boileau vd. (2011), KMSM için çevrimiçi parametre tanımlanabilirliği kavramını kapsamlı biçimde ele almış ve tahmin algoritması tasarımından önce tanımlanabilirlik analizinin yapılması gerektiğini vurgulamıştır. Kararlı durum koşulunda dört parametrenin eşzamanlı tahmini, geleneksel maliyet fonksiyonu kullanılarak denendiğinde, maliyet yüzeyinin birden fazla global minimuma veya geniş bir düşük gradyan düzlemine sahip olması nedeniyle algoritmanın yanlış bir çözüme yakınsaması ya da salınım içinde kalması kaçınılmazdır.

Bu sorunun aşılması için literatürde temel üç yaklaşım önerilmektedir:

- Nominal değer sabitleme: Bazı parametreler (örn. L_d) nominal değerlere sahiplenip bilinmeyen sayısı azaltılır. Basit uygulanabilirliği avantaj sağlasa da nominal değer ile gerçek değer arasındaki uyumsuzluk tahmin hatasına dönüşür.
- Pertürbasyon sinyali enjeksiyonu: $i_d \neq 0$ gibi işaret enjeksiyonları ile ek KMSM denklemleri elde edilir. Ancak bu yaklaşım VSI kaynaklı doğrusal olmama etkisine ve gürültüye karşı hassasiyeti artırabilmektedir.
- Değişken hız kontrolü altında veri toplama: Farklı hız noktalarındaki ölçümler kullanılarak model sıra sayısı etkin biçimde artırılır. Liu ve Zhu (2015), bu yaklaşımı VSI kaynaklı doğrusal olmama telafisiyle birleştirerek tam sıralı bir tahmin modeli oluşturmuştur.

2.3. VSI Doğrusal Olmama Etkisi ve Tanımlanabilirlik Analizi

Geleneksel maliyet fonksiyonu yaygın olarak asenkron makine ve KMSM'lerde parametre kestirimi için kullanılır. Teoride, eğer akımların, gerilimlerin ve rotor hızının ölçümleri ideal ortamda doğru bir şekilde ölçülüyorsa, tüm makine parametreleri doğru bir şekilde tanımlanabilir.

$$P = \sum_{k=1}^{n} (K_d \left| i_d[k] - \hat{i}_d[k] \right| + K_q \left| i_q[k] - \hat{i}_q[k] \right|)$$

Ya da (3)

$$P = \sum_{k=1}^{n} (K_d \left(i_d[k] - \hat{i}_d[k] \right)^2 + K_q \left(i_q[k] - \hat{i}_q[k] \right)^2)$$

Burada P maliyet fonksiyonudur. Aynı zamanda uygunluk fonksiyonu olarak bilinir. K_d, K_q, $\hat{i}_d$, $\hat{i}_q$ ifadeleri sırasıyla maliyet fonksiyonlarının kazanç katsayılarını ve kestirilen parametreleri kullanılarak hesaplanan *dq*-eksenindeki akımları temsil etmektedir. Ne var ki, gerçek dünyada bu yöntemin gerilim kaynaklı eviricilerdeki doğrusal olmama durumundan etkilenmesi kolaydır ve bu kestirim modelinin tanımlanabilirliğinin farklı çalışma koşullarında dikkate alınması gerekmektedir.

Endüstriyel sürücü sistemlerinde KMSM, genellikle IGBT tabanlı bir gerilim kaynaklı evirici aracılığıyla beslenmektedir. Ölü zaman gecikmesi, güç yarı iletkeni iletim düşümleri ve anahtarlama gerilim hataları, gerçek uçtan uca voltajlar ile referans gerilimler arasında bir sapmaya neden olur. Bu etki *dq*-referans ekseninde tek bir distorsiyon gerilimi terimi V_{dead} olarak modellenebilir. Buradan yola çıkarak, gerilim kaynaklı eviricilerdeki doğrusal olmama durumu da hesaba katılarak (2) numaralı eşitlik aşağıdaki gibi yazılabilir:

$$\begin{aligned} u_d^* &= R_s i_d - L_q \omega_e i_q - D_d V_{dead} \\ u_q^* &= R_s i_q + L_d \omega_e i_d + \lambda_{pm} \omega_e - D_q V_{dead} \end{aligned} \quad (4)$$

Burada, $u_d = u_d^* + D_d V_{dead}$ ve $u_q = u_q^* + D_q V_{dead}$ olarak temsil edilir. D_d ve D_q rotor açısı ve faz akımlarının işaretlerine bağlı kazanç katsayılarıdır. $i_d \neq 0$ kontrolünde d-eksenindeki $D_d V_{dead}$ bileşeni sıfır ortalamalı bir bozucu gerilimi özelliği taşıyarak L_q tahminini kayda değer ölçüde etkilemezken, q-eksenindeki $D_q V_{dead}$ bileşeni doğru akım bileşeni içerdiğinden R_s ve λ_{pm} tahminini ciddi biçimde bozabilmektedir.

3. Önerilen Tam Sıralı Tahmin Modeli

3.1. Genel Çerçeve

Liu ve Zhu (2015), sıralı oran yetmezliği ve gerilim kaynaklı eviricilerdeki doğrusal olmama sorunlarını birlikte çözen sıralı (sequential) bir tahmin çerçevesi önermektedir. Motor değişken hız kontrolü altında çalıştırılmakta; $i_d = 0$ koşulunda dq-ekseni akımları sabit tutulmakta ve düşük hız ve orta/yüksek hız olmak üzere iki farklı rotor hızı bölgesine ait veri kümeleri kaydedilmektedir. Bu yapı, modelin sıra sayısını etkin biçimde artırarak tanımlanabilirliğini güvence altına almaktadır.

Tahmin süreci beş adımdan oluşmaktadır:

1. Değişken hız kontrolü altında rotor hızı, rotor açısı, dq-ekseni akımları ve referans gerilimleri ölçülür. dq-ekseni akımları sabit tutularak L_q ve λ_{pm}'in veri toplama süresince değişmemesi sağlanır.
2. L_q endüktansı, d-ekseni denklemi üzerinden P1(L_q, V_{dead}) maliyet fonksiyonu minimize edilerek tahmin edilir.
3. L_q endüktansı, tahmin değerine sabitlenir; V_{dead}, P2(V_{dead}) maliyet fonksiyonu minimize edilerek tahmin edilir.
4. λ_{pm} akı değeri, P3(λ_{pm}) maliyet fonksiyonu minimize edilerek tahmin edilir.
5. V_{dead} ve λ_{pm} tahmin değerlerine sabitlenir; R_s direnci, P4(R_s) maliyet fonksiyonu minimize edilerek tahmin edilir.

Bu sıralı yapı, olası lokal minimum tuzaklarını azaltmakta her alt problemin tekil bir çözüme sahip olmasını sağlamaktadır.

3.2. Maliyet Fonksiyonları ve Tasarım İlkeleri

Tahmin sürecinde kullanılan beş maliyet fonksiyonu aşağıda özetlenmektedir. Bu fonksiyonlar ile sırasıyla L_q, V_{dead}, λ_{pm}, R_s ve L_d parametrelerinin tahmini hedeflenmektedir.

Tablo 1 Önerilen Tahmin Çerçevesinde Kullanılan Maliyet Fonksiyonları

Fonksiyon	Tahmin Edilen	Açıklama
P1	L_q ve V_{dead}	d-ekseni denklemini minimize ederek L_q ve V_{dead} değerini eşzamanlı olarak tahmin eder.
P2	V_{dead}	L_q sabitlenip sadece V_{dead} tahmin edilir; doğruluğu artırır.
P3	λ_{pm}	Verinin iki yarısı arasındaki fark alınarak gerilim kaynaklı eviricilerdeki doğrusal olmama etkisi yok edilir.
P4	R_s	V_{dead} ve λ_{pm} sabit tutularak stator sargı direnci hassas biçimde tahmin edilir.
P5	L_d	Sabit yük momenti altında $\Delta i_d \neq 0$ enjeksiyonu ile d-ekseni endüktansı tahmin edilir.

P3 maliyet fonksiyonunun en özgün yönü, ölçüm verisini iki eşit parçaya bölerek q-ekseni referans voltajları arasındaki farkı kullanılmasıdır:

$$\begin{aligned}&\sum_{k=1}^{0.5n}\left(u_q^*[k]-u_q^*[k+0.5n]\right)\\ \approx&\sum_{k=1}^{0.5n}\Big(\lambda_{pm}(\omega_e[k]-\omega_e[k+0.5n])\Big)\\ &+V_{dead}\left(\sum_{k=1}^{0.5n}D_q[k+0.5n]-\sum_{k=1}^{0.5n}D_q[k]\right)\end{aligned} \tag{5}$$

i_d ve i_q değerleri sabit tutulduğunda D_q'nun ortalama değeri iki veri yarısında aynı kalacağından gerilim kaynaklı eviricilerdeki doğrusal olmama terimi kendiliğinden iptal olmaktadır. Bu sayede λ_{pm} tahmini gerilim kaynaklı evirici etkisinden bağımsız hale gelmekte ve (3) numaralı eşitlikteki geleneksel maliyet fonksiyonu ile karşılaştırıldığında önemli ölçüde daha doğru bir sonuç elde edilmektedir.

3.3. *d*-Ekseni Endüktans Tahmini

$i_d = 0$ kontrolünde devre denkleminden d-ekseni endüktansı terimi ortadan kalkmaktadır. L_d'yi tahmin edebilmek için yük momenti anlık olarak sabit tutulurken d-ekseni akımına küçük bir doğru akım ofseti ($\Delta i_d \neq 0$) geçici olarak eklenmektedir. Bu kabulden yola çıkarak, sabit yük altında $\Delta i_d \neq 0$ eklemesi

yapılarak ve yapılmadan elektromanyetik tork ifadesi aşağıdaki gibi ifade edilebilir:

$$T_e = 1.5p\lambda_{pm}i_q = 1.5p[\lambda_{pm}i_{q2} + (L_d - L_q)i_{q2}i_{d2}] \quad (6)$$

Burada, p, i_{d2}, i_{q2} sırasıyla KMSM'deki rotor kutup çiftini ve $i_{d2} = \Delta i_d$ olduğu durumdaki *dq*-ekseni akımlarını ifade etmektedir. Böylece (6) numaralı denklem tekrar aşağıdaki gibi türetilebilir.

$$(L_d - L_q)i_{q2}i_{d2} = \lambda_{pm}(i_q - i_{q2}) \quad (7)$$

Buradan L_d için P5 maliyet fonksiyonu tanımlanmakta ve gerilim kaynaklı eviricilerdeki doğrusal olmama etkisi bu denklemde yer almadığından tahmin bu etkiden muaf olmaktadır. Liu ve Zhu (2015), $\Delta i_d = -0.2\ A$ (anma akımının %5'i) değerinin hem sinyal-gürültü oranı hem de parametre değişimi bakımından uygun bir denge sağlığını raporlamaktadır.

4. Kuantum Genetik Algoritma

4.1. Genetik Algoritmanın Temelleri

Genetik Algoritma, doğal seçilim ve genetik kalıtım ilkelerinden esinlenen, popülasyon tabanlı bir küresel optimizasyon yöntemidir. Bir genetik algoritma döngüsü genel olarak üç aşamadan oluşmaktadır: rastgele üretilen bir başlangıç popülasyonu oluşturma, seçim-çaprazlama-mutasyon operatörleriyle yinelenen evrim ve durdurma koşulunun sağlanmasıyla sonlandırma. Genetik algoritma, KMSM parametre tahmininde erken dönemden itibaren uygulanmış ve asenkron motor, doğru akım motoru ile endüktans belirleme çalışmalarında başarılı sonuçlar vermiştir.

Bununla birlikte, klasik ikili kodlama kullanan genetik algoritmanın yüksek boyutlu arama uzaylarında yerel minimumlara sıkışma eğilimi gösterdiği ve büyük popülasyon boyutlarına ihtiyaç duyduğu bilinmektedir. Bu kısıtlamaların aşılması amacıyla Han ve Kim (2002) Kuantumdan esinlenerek ortaya çıkarılan evrimsel algoritma çerçevesini, ardından Malossini vd. (2008) Kuantum Genetik Optimizasyon yaklaşımını önermiştir.

4.2. Kuantum Bit Kodlaması

Kuantum genetik algoritmanın klasik genetik algoritmadan en temel farkı, çözüm temsilinde kullanılan kodlama şemasıdır. Klasik Genetik algoritmadan her bit yalnızca 0 veya 1 değerini alabilirken kuantum genetik algoritmada her

gen birimi, kuantum mekaniğinin süperpozisyon ilkesinden yararlanarak aşağıdaki kuantum bit yapısıyla temsil edilmektedir:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle, |\alpha|^2 + |\beta|^2 = 1 \quad (8)$$

Burada, $|\alpha|^2$ ve $|\beta|^2$, ilgili bitin sırasıyla 0 ve 1 durumunda bulunma olasılığını ifade etmektedir. Bu yapı, bir kuantum bitin 0 ve 1 arasındaki tüm olası değerleri eşzamanlı olarak temsil edebilmesine olanak tanıyarak arama uzayını önemli ölçüde zenginleştirmektedir.

Bu önerilen yöntemde, kuantum genetik algoritmanın maksimum nesil sayısı M ve mevcut nesil numarasının X olduğu varsayılmaktadır. Evrim süreci boyunca, kuantum rotasyon kapısı operatörü, kuantum bit genleri mevcut en iyi çözüme doğru yönlendirmektedir. Liu ve Zhu (2015) uygulamasında, evrimin ilk yarısında ($X < 0.5M$) dönüş açısı 0.05π olarak ayarlanarak hızlı küresel arama desteklenirken evrimin ikinci yarısında ($X \geq 0.5M$) 0.0025π değerine indirilerek yerel arama hassasiyeti artırılmaktadır.

4.3. Kuantum Genetik Algoritmanın KMSM Parametrelerine Kodlanması

KMSM parametre tahmininde V_{dead}, L_d, L_q, R_s ve λ_{pm} değerlerinin her biri için 8 bitlik kuantum bir gen dizisi kullanılmaktadır. Parametre değer aralıkları ve her genetik birime karşılık gelen çözünürlükler şu şekilde belirlenmiştir. R_s için 0.05 Ω, λ_{pm} için 0.002 Wb, V_{dead} için 0.01 V, L_d ve L_q için 0.5 mH'dir. Bu değer aralıkları, gerçek sargı direnci ve rotor mıknatıs akı bağı değişim aralıklarını kapsayacak şekilde tasarlanmıştır.

Hem Genetik Algoritma, hem de Kuantum Genetik algoritma için 50 bireylik popülasyon boyutu, %50 çaprazlama ve %6 mutasyon olasılığı kullanılmıştır. Deneysel bulgular, kuantum genetik algoritmanın yaklaşık 40 nesil sonra global optimuma yakınsadığını gösterirken klasik Genetik Algoritmanın alt optimum bir noktada kalmaya devam ettiğini ortaya koymuştur.

5. Deneysel Doğrulama

5.1. Test Düzeneği

Önerilen yöntemin performansı, dSPACE DS1006 tabanlı vektör kontrollü bir sürücü sistemine bağlı içten yerleştirmeli KMSM üzerinde test edilmiştir. Evirici olarak SEMIX 71GD12E4s kullanılmış; tüm veriler dSPACE DS1006 ile kaydedilerek Borland C++ 6.0 ile programlanan parametre tahmin

algoritmalarına aktarılmıştır. Test motorunun temel tasarım parametreleri aşağıda sunulmaktadır:

Tablo 2 Test için kullanılan KMSM tasarım parametreleri ve nominal değerleri

Parametre	Değer	Parametre	Değer
Anma akımı	4 A	Nominal faz direnci (T=20°C)	6.00 Ω
Anma hızı	1000 d/dk	Terminal direnci (anma)	0.20 Ω
DC bara gerilimi	158 V	Nominal q-ekseni endüktansı	58.5 mH
Kutup çifti sayısı	3	Nominal d-ekseni endüktansı	38.1 mH
Rotor mıknatıs akı bağı (yüksüz)	236 mWb	PM türü / Yapısı	Gömülü KMSM

Deneysel veri kümesi üç çalışma bölgesine ayrılmıştır: S0 (sabit hız/moment kontrolü, t < 2.5 s), S1 (değişken hız-düşük hız bölgesi) ve S2 (değişken hız-orta /yüksek hız bölgesi). $i_d = 0$ kontrolü boyunca $i_q = 2A$ sabit tutulmuştur.

5.2. Sıralı Oran Yetmezliğinin Gözlemlenmesi

Klasik maliyet fonksiyonu (3) kullanılarak S0 bölgesindeki sabit hız verisi ile yürütülen tahmin denemesinde, R_s ve λ_{pm}'e ait değerlerin beklenen aralıkta doğru sonuçlara yakınsamak yerine dengesiz seyrettiği gözlemlenmiştir. Bu durum, kararlı durumda eşzamanlı tahmin yapıldığında tahmin modelinin sonsuz sayıda çözüm içerdiği analizini doğrulamaktadır.

Değişken hız bölgesi (S1) verisinin klasik maliyet fonksiyonuyla durumunda ise kuantum genetik algoritma yaklaşık 40 nesil sonra bir optimuma yakınsarken klasik genetik algoritma alt optimumda kalmaktadır. Öte yandan gerilim kaynaklı eviricilerdeki doğrusal olmama etkisi ihmal edildiğinde kuantum genetik algoritmanın ürettiği R_s (10.9Ω) ve λ_{pm} (0.26 Wb) değerleri nominal değerlerden (6.2 Ω ve 0.236 Wb) belirgin biçimde saparak gerilim kaynaklı evirici etkisinin dikkate alınmamasının önemini açıkça ortaya koymaktadır.

5.3. Önerilen Yöntemin Sonuçları

Önerilen kuantum genetik algoritma tabanlı sıralı tahmin yöntemi, hem düşük hız (S1) hem de orta/yüksek hız (S2) bölgelerinde uygulanmıştır. Temel bulgular aşağıdaki gibi özetlenebilir:

- L_q tahmini: $L_q \approx 58\,mH$ olarak bulunmuş olup nominal değere ($58.5\,mH$) çok yakındır. Bu sonuç, $i_d = 0$ kontrolünde $D_d V_{dead}$'in sıfır ortalamalı bir bozucu olduğu ve L_q tahminini etkilemediği analizini doğrulamaktadır.
- V_{dead} tahmini: L_q sabitlenip P2 minimize edildikten sonra V_{dead} düşük hız bölgesinde $\approx 1.32\,V$, yüksek hız bölgesinde $\approx 1.46\,V$ olarak tahmin edilmiştir.
- λ_{pm} tahmini: P3 maliyet fonksiyonu ile gerilim kaynaklı evirici etkisi iptal edilerek $\lambda_{pm} \approx 0.224\,Wb$ olarak bulunmuştur. Nominal yüksüz değere ($\backslash 0.236\,Wb$) kıyasla bu düşüş, yük koşulunda sıcaklık artışı ve akı doygunluğuyla tutarlıdır.
- R_s tahmini: Gerilim kaynaklı evirici telafisi olmadan $R_s \approx 9.2\Omega$ iken gerilim kaynaklı evirici telafisi ile $R_s \approx 6.65\Omega$ bulunmuştur. Nominal değer (6.2Ω) ile arasındaki 0.45Ω farkı, sargı sıcaklık artışı, çekirdek kaybı direnci ve IGBT iletim direnci kaynaklarına bağlanabilmektedir.
- L_d tahmini: Sabit yük momenti altında $\Delta i_d = -0.2\,A$ enjeksiyonu ile $L_d \approx 37.0\,mH$ bulunmuştur ve nominal değer ($38.1\,mH$) ile uyum iyidir.

Sonlu elemanlar (2B FE, Ansoft Maxwell) simülasyonuyla karşılaştırıldığında, λ_{pm} ve L_q'nun i_q akımına bağlı değişim eğrileri tahmin sonuçlarıyla uyumlu seyretmektedir. $i_q > 3A$ bölgesinde λ_{pm} eğrisinin sonlu elemanlar sonuçlarından uzaklaşması, rotor mıknatısının sıcaklığının artışına bağlanmaktadır.

6. Metasezgisel Yöntemlerin Karşılaştırmalı Değerlendirmesi

6.1. Genetik Algoritma ve Kuantum Genetik Algoritma Karşılaştırması

Liu ve Zhu (2015)'in kapsamlı deneysel karşılaştırması, Kuantum Genetik Algoritmanın klasik Genetik Algoritmaya kıyasla KMSM parametre tahmini bağlamında belirgin üstünlükler sergilediğini göstermektedir. Aşağıda her iki algoritmanın temel özellikleri karşılaştırmalı olarak sunulmaktadır:

Tablo 3 Genetik algoritma ve Kuantum Genetik Algoritmanın KMSM parametre tahmini açısından karşılaştırılması

Özellik	GA (Genetik Algoritma)	QGA (Kuantum GA)
Kodlama türü	İkili (binary)	Kuantum (Q-bit)
Global arama yeteneği	Orta	Yüksek
Yakınsama hızı	Yavaş (~100+ nesil)	Hızlı (~40 nesil)
Yerel minimuma takılma	Sık görülür	Nadir görülür
Hesaplama yükü	Düşük	Orta
R_s tahmini (Ω)	Yanıltıcı sonuç	~6.65 Ω (doğru)
λ_{pm} tahmini (Wb)	Yanıltıcı sonuç	~0.224 Wb (doğru)
L_q tahmini (mH)	~57.5 mH	~58.0 mH

6.2. Parçacık Sürüsü Optimizasyonu (PSO)

Parçacık Sürüsü Optimizasyonu, sürü zekasından ilham alan bir metasezgisel yöntemdir. KMSM parametre tahmini için PSO uygulamaları Liu vd. (2010) ve Liu vd. (2008) tarafından raporlanmıştır. PSO, türev hesabı gerektirmeyen yapısı ve az sayıda ayar parametresiyle dikkat çekmektedir. Ancak PSO'nun doğrudan klasik maliyet fonksiyonu (3) ile kullanılması, sıralı oran yetmezliği sorununun çözülmediği durumlarda yanlış yakınsamaya yol açabilmektedir. Liu ve Zhu (2015)'in önerdiği gibi sıralı tahmin çerçevesiyle birleştirildiğinde PSO da geçerli bir alternatif oluşturmaktadır. Huang vd. (2012), PSO tabanlı genişletilmiş Kalman filtresi kombinasyonuyla λ_{pm} tahmininde iyileşmiş yakınsama rapor etmiştir.

6.3. Difernsiyel Evrim (DE)

Diferansiyel Evrim, deney ve fark vektörleri yoluyla popülasyonu güncelleyen güçlü bir evrimsel algoritmadır. KMSM parametre tahmini literatüründe DE uygulamalarına ilişkin çalışmalar görece sınırlı kalmakla birlikte asenkron motor parametre tayininde DE'nin Genetik Algoritma ve PSO'ya kıyasla daha hızlı yakınsama gösterdiği rapor edilmiştir (Kim ve Kim, 2005). Karmaşık arama uzaylarındaki dayanıklılığı DE'yi gelecekteki KMSM çalışmaları için ilgi çekici bir aday kılmaktadır.

6.4. Bağışıklık Klonal Algoritması ve Diğer Yaklaşımlar

Liu ve Zhu (2015), önerilen sıralı tahmin modelinin yalnızca Kuantum Genetik Algoritma ile sınırlı olmadığını, diğer evrimsel algoritmaların

(bağışıklık klonal algoritması, PSO vb.) de bu çerçeve içinde kullanılabileceğini vurgulamaktadır. Bu bakış açısı, yöntemin genel geçerliğini ortaya koymakta ve farklı evrimsel algoritmalarla performans iyileştirme olanaklarına kapı aralamaktadır. Morimoto vd. (2006) ve Piippo vd. (2009) gibi çalışmalar, farklı hız bölgelerinde parametre tahminini ele almış ve önerilenle uyumlu bulgular sunmuştur.

7. Güncel Literatür ve Gelişmeler

7.1. Adaptif Gözlemci ve Birleşik Yöntemler

Geleneksel evrimsel algoritmalar çevrimdışı (offline) parametre tahminine odaklanırken son dönem araştırmaları çevrimiçi (online) ve uyarlanabilir (adaptive) yöntemlere yönelmiştir. Ichikawa vd. (2006), sistem tanımlama teorisine dayalı çevrimiçi parametre tanımlamasını sürücü sistemiyle bütünleştirmiştir. Rashed vd. (2007), model referanslı uyarlanabilir sistem (MRAS) kullanarak stator direncini sensörsüz hız kontrolüyle eşzamanlı olarak tahmin etmiştir. Underwood ve Husain (2010) ise özyinelemeli en küçük kareler yöntemiyle KMSM parametrelerini çevrimiçi güncelleyen bir yaklaşım sunmuştur.

Shi vd. (2012), genişletilmiş Kalman filtresi (EKF) aracılığıyla gömülü KMSM'de kalıcı mıknatıs akı bağını sensörsüz konum kontrolüyle birlikte çevrimiçi olarak tahmin etmiştir. Bu çalışma, sıralı oran yetmezliği sorununu farklı bir sistem gözlemci yapısıyla ele almaktadır.

7.2. Sıcaklık ve Yaşlanma Etkilerinin Modellenmesi

KMSM'nin durum izleme (condition monitoring) uygulamaları açısından parametre tahmini, özellikle stator sargı sıcaklığı R_s(üzerinden) ve rotor mıknatıs sıcaklığı/akı azalması (λ_{pm} üzerinden) bilgisini dolaylı yoldan elde etmek için giderek daha fazla kullanılmaktadır. Liu vd. (2011), farklı çalışma akım noktaları için çevrimiçi çoklu parametre tahminini sıcaklık değişim takibiyle birleştirmiştir. Liu ve Zhu (2014), sadece λ_{pm} değil gerilim kaynaklı eviricilerdeki doğrusal olmama etkisini de içeren kapsamlı bir çevrimiçi tahmin yöntemi geliştirmiştir. Reigosa vd. (2010), yüzey yerleştirmeli kalıcı mıknatıs makinelerinde yüksek frekanslı işaret enjeksiyonu yoluyla mıknatıs sıcaklığını tahmin etmiştir.

7.3. Derin Öğrenme ve Hibrit Yaklaşımlar

Son yıllarda parametre tahmini alanına derin öğrenme ve yapay sinir ağı tabanlı yaklaşımlar da dahil olmuştur. Yapay sinir ağları, motor parametreleri ile ölçülebilen elektriksel büyüklükler arasındaki karmaşık doğrusal olmayan

ilişkileri yaklaşık olarak öğrenme kapasitesi sunmaktadır. Bununla birlikte, bu tür yöntemlerin de eğitim verisinin çeşitliliği ve sıralı oran yetmezliğinin yol açtığı doğrusal bağımlılık sorununa dikkat etmesi gerektiği çeşitli araştırmacılar tarafından vurgulanmıştır. Metasezgisel algoritmalar ile derin öğrenme modellerinin hibrit biçimde kullanımı, özellikle çevrimiçi ayarlama senaryolarında araştırma ilgisini korumaktadır.

8. Sonuç ve Değerlendirme

Bu bölümde, KMSM'lerin değişken hız kontrolü altında parametre tahminine yönelik kapsamlı bir inceleme sunulmuştur. Temel bulgular şu şekilde özetlenebilir:

- Kararlı durum koşulunda dört KMSM parametresinin (R_s, L_d, L_q, λ_{pm}) eş zamanlı olarak tahmin edilmesi, sıralı oran yetmezliği nedeniyle tanımlanamaz bir problem oluşturmaktadır. Bu sorunun farkında olmadan tasarlanan tahmin algoritmaları sapma veya yanlış yakınsama riski taşımaktadır.
- Liu ve Zhu (2015) tarafından önerilen tam sıralı tahmin çerçevesi, değişken hız kontrolü altında toplanan verileri kullanarak hem tanımlanabilirlik sorununu hem de gerilim kaynaklı eviricilerdeki doğrusal olmama etkisini sistematik biçimde ele almaktadır. P3 maliyet fonksiyonundaki akıllı fark alma yapısı sayesinde λ_{pm} tahmini gerilim kaynaklı evirici etkisinden bağımsız kılınmakta; P4 ile R_s tahmini ise tahmin edilen V_{dead}'in telafisiyle yüksek doğruluğa ulaşmaktadır.
- Kuantum Genetik Algoritma, klasik Genetik Algoritmaya kıyasla küresel arama kapasitesi, yakınsama hızı ve yerel minimuma takılmama performansı bakımından üstünlük sergilemiştir. Ancak yöntemin sıralı çerçevesiyle birleştirildiğinde PSO, DE ve diğer evrimsel algoritmaların da eşdeğer doğruluk sağlayabileceği öngörülmektedir.
- Deneysel sonuçlar, önerilen yöntemin stator sargı direnci ve rotor mıknatıs akı bağını nominal parametre değerlerine ihtiyaç duymaksızın yüksek doğrulukla tahmin edebildiğini ve stator sargısı ile rotor mıknatıslarının durum izleme uygulamaları için kullanılabileceğini ortaya koymaktadır.

Gelecekteki araştırmalar için $i_d \neq 0$ değişken hız kontrolü senaryosu, PSO ve bağışıklık klonal algoritması gibi alternatif evrimsel algoritmalarla entegrasyon ve gerçek zamanlı gömülü sistem uygulamalarında hesaplama verimliliğinin artırılması önemli açık araştırma alanları olarak öne çıkmaktadır.

Kaynakça

Alonge, F., D'Ippolito, F., Ferrante, G. ve Raimondi, F.M. (1998). Parameter identification of induction motor model using genetic algorithms. IEE Proceedings – Control Theory and Applications, 145(6), 587–593.

Boileau, T., Leboeuf, N., Nahid-Mobarakeh, B. ve Meibody-Tabar, F. (2011). Online identification of PMSM parameters: Parameter identifiability and estimator comparative study. IEEE Transactions on Industry Applications, 47(4), 1944–1957.

Han, K.H. ve Kim, J.H. (2002). Quantum-inspired evolutionary algorithm for a class of combinatorial optimization. IEEE Transactions on Evolutionary Computation, 6(6), 580–593.

Holland, J.H. (1975). Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbor.

Huang, L., Shi, Y., Sun, K. ve Li, Y. (2012). Online identification of permanent magnet flux based on extended Kalman filter for IPMSM drive with position sensorless control. IEEE Transactions on Industrial Electronics, 59(11), 4169–4178.

Ichikawa, S., Tomita, M., Doki, S. ve Okuma, S. (2006). Sensorless control of permanent-magnet synchronous motors using online parameter identification based on system identification theory. IEEE Transactions on Industrial Electronics, 53(2), 363–372.

Kim, H.W., Youn, M.J. ve Cho, K.Y. (2005). New voltage distortion observer of PWM VSI for PMSM. IEEE Transactions on Industrial Electronics, 52(4), 1188–1192.

Kim, S.Y., Lee, W., Rho, M.S. ve Park, S.Y. (2010). Effective dead-time compensation using a simple vectorial disturbance estimator in PMSM drives. IEEE Transactions on Industrial Electronics, 57(5), 1609–1614.

Kim, J.W. ve Kim, S.W. (2005). Parameter identification of induction motors using dynamic encoding algorithm for searches (DEAS). IEEE Transactions on Energy Conversion, 20(1), 16–24.

Liu, K., Sun, J., Hu, Y.S. ve Zhu, Z.Q. (2011). Online multiparameter estimation of non-salient pole PM synchronous machines with temperature variation tracking. IEEE Transactions on Industrial Electronics, 58(5), 1776–1788.

Liu, K. ve Zhu, Z.Q. (2014). Online estimation of rotor flux linkage and voltage source inverter nonlinearity in permanent magnet synchronous machine drives. IEEE Transactions on Power Electronics, 29(1), 418–427.

Liu, K. ve Zhu, Z.Q. (2015). Quantum genetic algorithm-based parameter estimation of PMSM under variable speed control accounting for system

identifiability and VSI nonlinearity. IEEE Transactions on Industrial Electronics, 62(4), 2363–2371.

Liu, K., Zhu, Z.Q. ve Stone, D.A. (2013). Parameter estimation for condition monitoring of PMSM stator winding and rotor permanent magnets. IEEE Transactions on Industrial Electronics, 60(12), 5902–5913.

Liu, K., Zhu, Z.Q., Zhang, J. ve Zhang, Q. (2010). Multi-parameter estimation of nonsalient pole permanent magnet synchronous machines by using evolutionary algorithms. Proceedings of the IEEE 15th International Conference on Bio-Inspired Computing, 766–774.

Liu, L., Liu, W.X. ve Cartes, D.A. (2008). Permanent magnet synchronous motor parameter identification using particle swarm optimization. International Journal of Computational Intelligence Research, 4(2), 211–218.

Malossini, A., Blanzieri, E. ve Calarco, T. (2008). Quantum genetic optimization. IEEE Transactions on Evolutionary Computation, 12(2), 231–241.

Morimoto, S., Sanada, M. ve Yakeda, Y. (2006). Mechanical sensorless drives of IPMSM with online parameter identification. IEEE Transactions on Industry Applications, 42(5), 1241–1248.

Piippo, A., Hinkkanen, M. ve Luomi, J. (2009). Adaptation of motor parameters in sensorless PMSM drives. IEEE Transactions on Industry Applications, 45(1), 203–212.

Pillay, P., Nolan, R. ve Haque, T. (1997). Application of genetic algorithms to motor parameter determination for transient torque calculations. IEEE Transactions on Industry Applications, 33(5), 1273–1282.

Rashed, M., Macconnell, P.F.A., Stronach, A.F. ve Acarnley, P. (2007). Sensorless indirect-rotor-field-orientation speed control of a permanent-magnet synchronous motor with stator-resistance estimation. IEEE Transactions on Industrial Electronics, 54(3), 1664–1675.

Reigosa, D.D., Briz, F., Garcia, P., Guerrero, J.M. ve Degner, M.W. (2010). Magnet temperature estimation in surface PM machines using high-frequency signal injection. IEEE Transactions on Industry Applications, 46(4), 1468–1475.

Underwood, S. ve Husain, I. (2010). Online parameter estimation and adaptive control of permanent magnet synchronous machines. IEEE Transactions on Industrial Electronics, 57(7), 2435–2443.

4. Bölüm

Makine Öğrenmesi Modellerinde Açıklanabilirlik ve Karşıt Saldırılar: Riskler, Zafiyetler ve Gelecek Araştırma Yönelimleri

Ertuğrul ATEŞ[1]

1. Giriş

Derin öğrenme tabanlı modeller günümüzde sağlıktan finansa, güvenlikten otonom sistemlere kadar birçok kritik uygulama alanında yaygın biçimde kullanılmaktadır. Ancak bu modellerin güvenilir bir şekilde konuşlandırılmasını sınırlayan iki temel zafiyet öne çıkmaktadır: kara kutu (black-box) doğaları nedeniyle karar süreçlerinin şeffaf olmaması ve karşıt örneklere karşı kırılganlıklarıdır (Vadillo, Santana ve Lozano, 2025). Bu iki sorunu ayrı ayrı ele alan araştırma hatları, son yıllarda birbiriyle kesişmeye başlamış; açıklanabilir modellerin de karşıt saldırılara ne ölçüde dayanıklı olduğu ciddi biçimde sorgulanır hale gelmiştir.

Bir yandan sonradan yapılan (post-hoc) açıklama yöntemleri, modellerin girdilerde hangi öğeleri kararlarında en çok dikkate aldığını ortaya koymaya çalışırken (Ghorbani, Abid ve Zou, 2019; Kim vd., 2018; Yosinski vd., 2015; Zeiler ve Fergus, 2014), öte yandan doğuştan yorumlanabilir modeller tasarlanarak şeffaf bir akıl yürütme hedeflenmektedir (Alvarez-Melis ve Jaakkola, 2018a; Chen, Li vd., 2019; Hase vd., 2019; Li vd., 2018; Saralajew vd., 2019). Diğer taraftan, Szegedy vd. (2014) tarafından tanımlanan karşıt örnekler, insana görünmez kalacak biçimde girdilerde yapılan küçük değişikliklerin modelin çıktısını yanlış yönde değiştirebildiğini göstermiştir. Bu durum; sağlık (Bortsova vd., 2021; Finlayson vd., 2019), gözetim ve güvenlik (Sharif vd., 2016; Thys, Ranst ve Goedemé, 2019), makine çevirisi (Belinkov ve Bisk, 2018; Zhao, Dua ve Singh, 2018) ve finans (Ballet vd., 2019; Cartella vd., 2021) gibi pek çok alanda ciddi riskler doğurmaktadır.

Bu bölümün amacı, Vadillo, Santana ve Lozano (2025) tarafından WIREs Data Mining and Knowledge Discovery dergisinde yayımlanan "Adversarial Attacks in Explainable Machine Learning: A Survey of Threats Against Models and Humans" başlıklı kapsamlı derlemeyi esas alarak; açıklanabilir makine öğrenmesi modellerine yönelik karşıt saldırıların kavramsal çerçevesini,

[1] Dr. Öğr. Üyesi, Veri Bilimi ve Analitiği Bölümü, Kayseri Üniversitesi, ORCID: 0000-0001-6628-5350

taksonomisini, senaryo temelli tehdit modelini, açıklama türüne özgü zafiyetlerini ve ileriye dönük araştırma yönelimlerini ayrıntılı biçimde sunmaktır. Anılan çalışma, insanın yalnızca girdiyi ve çıktıyı değil, aynı zamanda modelin sunduğu açıklamayı da değerlendirdiği bir denetim sürecini merkeze alarak, karşıt örnek kavramını literatürde ilk kez bu denli sistematik biçimde genişletmektedir.

2. Açıklanabilir Makine Öğrenmesine Giriş

2.1. Açıklamaların Kapsamı, Amacı ve Etkisi

Bir açıklamanın amacı, modelin davranışını insan tarafından kolayca anlaşılabilir biçimde gerekçelendirmektir (Vadillo, Santana ve Lozano, 2025). Açıklamalar, kapsamları bakımından yerel (local) ve küresel (global) olmak üzere ikiye ayrılabilir (Zhang, Tiňo, Leonardis ve Tang, 2021): yerel açıklamalar belirli bir girdiye ilişkin modelin tahminini karakterize etmeyi hedeflerken, küresel açıklamalar modelin genel akıl yürütme sürecini; örneğin belirli bir sınıfın hangi koşullarda tahmin edildiğini ortaya koymaya çalışır. Karşıt saldırılar özelinde belirli girdi-çıktı çiftleri söz konusu olduğundan, bu alandaki araştırmalar ağırlıklı olarak yerel açıklamalara odaklanmaktadır.

Aynı model için farklı kullanıcılar farklı amaçlarla açıklama talep edebilir: bir kredi başvurusunda bulunan kullanıcı yalnızca kendi vakasının açıklamasıyla ilgilenirken, bir geliştirici modelin hangi girdilerde hatalı sınıflandırma yaptığını anlamak isteyebilir; bir analist ise modelin belirli bir sosyal gruba karşı önyargılı olup olmadığını sorgulayabilir (Guo, Daly, Alkan, Mattetti, Cornec ve Knijnenburg, 2022; Ross, Hughes ve Doshi-Velez, 2017; Schramowski vd., 2020; Teso, Alkan, Stammer ve Daly, 2023; Teso ve Kersting, 2019). Açıklamanın amacına ek olarak, etkisi de kritik bir faktördür; özellikle sağlık gibi alanlarda yanlış bir açıklamanın sonuçları ağır olabilir (Doshi-Velez ve Kim, 2018). Vadillo, Santana ve Lozano (2025), bu faktörlerin hem açıklama yöntemlerinin tasarımında hem de bu yöntemlere yönelik karşıt saldırıların tasarımında sıklıkla göz ardı edildiğini vurgulamaktadır.

2.2. Açıklama Türleri

Söz konusu derleme, literatürde yer alan dört temel açıklama türünü tanımlamaktadır:

Özellik tabanlı açıklamalar (feature-based): Girdideki her bir özelliğe, çıktı sınıflandırmasındaki göreceli önemine göre bir puan atanır (Baehrens vd., 2010; Lundberg ve Lee, 2017; Ribeiro, Singh ve Guestrin, 2016; Robnik-Šikonja ve Kononenko, 2008; Štrumbelj ve Kononenko, 2010). Görüntü alanında en yaygın biçimi, girdinin en ilgili bölgelerini vurgulayan aktivasyon ya da belirginlik

(salience) haritalarıdır (Selvaraju vd., 2017; Simonyan, Vedaldi ve Zisserman, 2014; Zeiler ve Fergus, 2014). Bununla birlikte, bu tür açıklamaların güvenilmez ve yanıltıcı olabileceğine dair önemli eleştiriler de bulunmaktadır (Kim vd., 2018; Lipton, 2018; Rudin, 2019).

Örnek tabanlı (prototip tabanlı) açıklamalar: Girdinin, tahmin edilen sınıfı temsil eden prototipik girdilerle olan benzerliğine dayanır. Bazı yaklaşımlar, prototip tabanlı akıl yürütmeyi doğrudan modelin öğrenme sürecine entegre etmektedir (Alvarez-Melis ve Jaakkola, 2018a; Chen, Li vd., 2019; Gautam vd., 2022; Li vd., 2018; Nauta, Van Bree ve Seifert, 2021).

Kural tabanlı açıklamalar: Modelin akıl yürütmesini, basitleştirilmiş ve insan tarafından anlaşılabilir kurallar (örn. eğer-o hâlde (if-then) kuralları) biçiminde ortaya koyar (Guidotti vd., 2019; Lakkaraju, Kamar, Caruana ve Leskovec, 2019; Ribeiro, Singh ve Guestrin, 2018).

Karşı olgusal (counterfactual) açıklamalar: Girdide, farklı ve genellikle daha olumlu bir çıktı sınıfının elde edilmesi için yapılması gereken değişiklikleri önerir (Guidotti vd., 2019; Wachter, Mittelstadt ve Russell, 2017).

3. Karşıt Saldırılara Genel Bakış ve Taksonomi

Karşıt örnekler, kötü niyetli bir aktörün modeli yanlış tahminlere yönlendirmek amacıyla kasıtlı olarak değiştirdiği, ancak bu değişikliklerin insan gözüyle algılanamaz kalmasını hedeflediği girdilerdir (Szegedy vd., 2014). Vadillo, Santana ve Lozano (2025), literatürdeki taksonomiyi dört temel eksende özetlemektedir:

- Yanlış sınıflandırma türü: Hedefli saldırılar belirli bir yanlış sınıfı üretmeyi amaçlarken, hedefsiz saldırılar herhengi bir yanlış sınıflandırmayı yeterli görür.
- Pertürbasyonun kapsamı: Bireysel pertürbasyonlar tek bir girdi için optimize edilirken (Carlini ve Wagner, 2017; Goodfellow, Shlens ve Szegedy, 2015; Madry, Makelov, Schmidt, Tsipras ve Vladu, 2018), evrensel pertürbasyonlar girdiden bağımsız olarak etkili olacak biçimde tasarlanır (Moosavi-Dezfooli, Fawzi, Fawzi ve Frossard, 2017; Mopuri, Garg ve Babu, 2017).
- Saldırgana sunulan kaynaklar: Beyaz kutu (white-box) senaryolarda saldırgan modelin mimarisi, ağırlıkları ve eğitim detayları hakkında tam bilgiye sahiptir ve genellikle gradyan tabanlı yöntemler kullanır (Carlini ve Wagner, 2017; Goodfellow, Shlens ve Szegedy, 2015; Madry vd., 2018); kara kutu (black-box) senaryolarda ise böyle bir bilgi bulunmaz ve evrimsel algoritmalar (Alzantot, Sharma, Chakraborty, Zhang, Hsieh ve Srivastava, 2019), gradyan kestirimi (Chen, Zhang, Sharma, Yi ve Hsieh,

2017; Ilyas, Engstrom, Athalye ve Lin, 2018) veya vekil (surrogate) modeller (Liu, Chen, Liu ve Song, 2017; Papernot vd., 2017) gibi daha maliyetli stratejiler kullanılır.

- Konuşlandırma türü: Dijital dünya saldırıları doğrudan sayısal dosya üzerinde değişiklik yaparken, fiziksel dünya saldırıları "havadan" (over-the-air) yakalanan sinyaller üzerinde de etkili olacak biçimde tasarlanır; örneğin basılı trafik işaretleri veya konuşma komutları (Eykholt vd., 2018; Sharif vd., 2016; Xu, Zhang, Liu vd., 2020).

Bu saldırılar başlangıçta neredeyse yalnızca sınıflandırma problemlerine odaklanmış olsa da, zamanla regresyon (Balda, Behboodi ve Mathar, 2019; Kos, Fischer ve Song, 2018), pekiştirmeli öğrenme (Hussenot, Geist ve Pietquin, 2020) ve görüntü bölütleme (Cisse, Adi, Neverova ve Keshet, 2017; Xie, Wang, Zhang, Zhou, Xie ve Yuille, 2017) gibi farklı problem türlerine de genişletilmiştir (Yuan, He, Zhu ve Li, 2019).

4. Açıklamaların Karşıt Senaryolardaki Güvenilirliği

Ghorbani, Abid ve Zou (2019), Dombrowski vd. (2019), Alvarez-Melis ve Jaakkola (2018b), Zhang, Wang, Shen, Ji, Luo ve Wang (2020) ve Kuppa ve Le-Khac (2020) gibi çalışmalar, girdide yapılan küçük değişikliklerin, çıktı sınıflandırması aynı kalsa dahi özellik önemi (feature-importance) açıklamalarında dramatik değişikliklere yol açabildiğini göstermiştir. Ghorbani, Abid ve Zou (2019), önerdikleri saldırıları ayrıca Koh ve Liang (2017)'ın etki fonksiyonlarına (influence functions) dayalı örnek tabanlı açıklamaları üzerinde de değerlendirmiştir. Zheng, Fernandes ve Prakash (2019) ise doğuştan açıklanabilir (prototip tabanlı) sınıflandırıcılar için, çıktıyı değiştirmeden açıklamaları değiştirebilen saldırılar geliştirmiştir.

Bu kırılganlığın kaynağına ilişkin iki temel açıklama öne sürülmüştür. Ghorbani, Abid ve Zou (2019) ile Dombrowski vd. (2019), karmaşık modellerin karar sınırlarının düzgün olmayan (non-smooth) geometrisini sorumlu tutmaktadır; buna göre girdide yapılan küçük değişiklikler gradyan yönünü (karar sınırına dik yönü) önemli ölçüde değiştirebilmekte, bu da çoğu açıklama yönteminin dayandığı gradyan bilgisini bozmaktadır. Zhang vd. (2020) ile Kuppa ve Le-Khac (2020) ise kırılganlığı tahminler ile açıklamalar arasındaki bir boşluğa (gap) bağlamaktadır; bu hipotezin doğuştan açıklanabilir modeller için de geçerli olup olmadığı ise açık bir soru olarak durmaktadır (Vadillo, Santana ve Lozano, 2025).

Öte yandan Aivodji vd. (2019), Aïvodji, Arai, Gambs ve Hara (2021) ile Lakkaraju ve Bastani (2020), önyargılı veya güvenilmez bir modele ait, kullanıcının güvenini yönlendirebilecek "sahte güvenilir" açıklamaların (fairwashing) üretilebildiğini göstermiştir; bu yaklaşımlar doğrudan karşıt saldırıya dayanmasa da

benzer bir manipülasyon amacı taşımaktadır. Benzer biçimde; Anders, Pasliev, Dombrowski, Müller ve Kessel (2020), Dimanov, Bhatt, Jamnik ve Weller (2020), Heo, Joo ve Moon (2019), Le Merrer ve Trédan (2020) ve Slack, Hilgard, Jia, Singh ve Lakkaraju (2020) tahmin performansına zarar vermeden yanlış veya yanıltıcı açıklamalar üretebilen "karşıt modeller" tasarlamıştır.

Savunma tarafında ise özellik tabanlı açıklama yöntemlerinin gürbüzlüğünü artırmaya yönelik düzenli hale getirme (regularization) stratejileri (Boopathy vd., 2020; Chen, Wu, Rastogi, Liang ve Jha, 2019; Tang, Liu, Yang, Zou ve Hu, 2022) ile gradyan tabanlı açıklamalar için açıklama-ortalama alma (explanation-averaging) yaklaşımları (Rieger ve Hansen, 2020) geliştirilmiştir. LIME (Ribeiro, Singh ve Guestrin, 2016) ve SHAP (Lundberg ve Lee, 2017) gibi model-agnostik yöntemler için de özel savunma stratejileri önerilmiştir (Carmichael ve Scheirer, 2023; Ghalebikesabi, Ter-Minassian, DiazOrdaz ve Holmes, 2021; Vreš ve Robnik-Šikonja, 2022). Bununla birlikte, çok sayıda açıklama türü ve yöntemini kapsayan bütüncül bir gürbüzlük değerlendirmesi hâlâ açık bir problem olarak durmaktadır (Huang, Zhao, Jin ve Huang, 2023).

5. Açıklanabilir Makine Öğrenmesi İçin Genişletilmiş Karşıt Örnek Çerçevesi

Vadillo, Santana ve Lozano (2025)'nun bu derlemedeki asıl katkısı, karşıt örnek kavramını insanın yalnızca girdiyi değil, aynı zamanda modelin çıktısını ve açıklamasını da değerlendirdiği senaryolara genişletmesidir.

5.1. İnsanın Girdi ve Çıktıyı Gözlemlediği Senaryolar

Klasik adversarial örnekler, insan algısının girdi üzerinde değişmediği ancak modelin tahmininin değiştiği bir çerçeveye dayanır (Yuan vd., 2019). Ancak Vadillo, Santana ve Lozano (2025), insanın çıktıyı da gözlemlediği durumlarda dört olası senaryo tanımlamaktadır. Burada $f(x)$ modelin tahminini, $h(x)$ insan sınıflandırmasını ve y_x gerçek (ground-truth) sınıfı temsil etmektedir:

$h(x) = y_x$	A.0.1 $f(x) = y_x$ Hem model hem de insan doğru sınıflandırır.
	A.0.2 $f(x) \neq y_x$ Model yanlış, insan doğru sınıflandırır. (Klasik karşıt saldırının hedeflediği senaryo)
$h(x) \neq y_x$	A.0.3 $f(x) = y_x$ İnsan yanlışlıkla modelin doğru tahminine katılmaz.
	A.0.4 $f(x) \neq y_x$ Hem model hem insan aynı yanlış sınıfa karar verir.

Klasik karşıt saldırılar, doğru sınıflandırılmış bir girdiden (A.0.1) yola çıkarak A.0.2 senaryosunu üretmeyi hedefler; ancak kullanıcı çıktıyı da gözlemlediğinde saldırının fiilî etkinliği, insan öznelerin tespit edilen tutarsızlığı düzeltip düzeltemeyeceğine bağlıdır. Yazarlar ayrıca, tıbbi teşhis gibi yüksek karmaşıklıktaki görevlerde uzman bir öznenin de yanılabileceğini (Pillai, Oza ve Sharma, 2019) ve bu durumda saldırganın "insan hatasını onaylama" (human error confirmation) saldırısıyla A.0.3'ten A.0.4'e geçişi zorlayabileceğini belirtmektedir; örneğin bir model, uzmanın yanlış teşhisini onaylayacak biçimde manipüle edilebilir (Bortsova vd., 2021; Finlayson vd., 2019; Goddard, Roudsari ve Wyatt, 2012).

5.2. İnsanın Açıklamayı da Gözlemlediği Senaryolar

Açıklanabilir modeller söz konusu olduğunda, $A_f(x)$ modelin $f(x)$ kararı için sunduğu açıklamayı, $A_h(x)$ ise insanın kendi kriterlerine göre üreteceği açıklamayı temsil eder. Vadillo, Santana ve Lozano (2025), literatür açısından ilgi çekici üç ana senaryo tanımlamaktadır:

A.1 $(f(x) = y_x \wedge A_f(x) \not\approx A_h(x))$: Model doğru tahmin yapar, ancak açıklama yanlış veya insanın vereceğinden farklıdır. Bu tür saldırılar post-hoc özellik-önemi açıklamaları (Alvarez-Melis ve Jaakkola, 2018b; Anders vd., 2020; Dombrowski vd., 2019; Ghorbani, Abid ve Zou, 2019; Kuppa ve Le-Khac, 2020) ve prototip tabanlı sınıflandırıcılar (Zheng, Fernandes ve Prakash, 2019) için incelenmiştir. Alt-senaryo A.1.1, açıklamanın hâlâ tutarlı görünen ama yanlış gerekçeler sunması durumudur; örneğin bir kredi başvurusu doğru biçimde reddedilirken açıklama "başvuru sahibi çok genç" gibi ilgisiz ama makul görünen bir gerekçe sunabilir; bu, başvurudaki asıl sorunun (örn. "maaş çok düşük") düzeltilmesini engelleyebilir (Ustun, Spangher ve Liu, 2019). Benzer biçimde yanlış bir tıbbi açıklama hatalı tedavi kararlarına yol açabilir (Bortsova vd., 2021; Ghassemi, Oakden-Rayner ve Beam, 2021). Bu şema, cinsiyet, ırk veya din gibi hassas özelliklere haksız yere atıf yapan önyargılı açıklamalar üretmek için de kullanılabilir (Slack, Hilgard, Lakkaraju ve Singh, 2021); ya da tam tersine, gerçek önyargıları saklayacak güvenilir görünümlü açıklamalar üretmek için kullanılabilir (Aivodji vd., 2019; Lakkaraju ve Bastani, 2020).

A.2 $(f(x) \neq y_x \wedge A_f(x) \not\approx A_h(x))$: Hem sınıflandırma hem açıklama yanlıştır (Kuppa ve Le-Khac, 2020; Sun, Song, Cai, Du ve Guizani, 2022; Zhang vd., 2020). Alt-senaryo A.2.1, açıklamanın modelin (yanlış) tahminiyle tutarlı olduğu; dolayısıyla insanın yanlış tahmine olan güvenini artırabileceği durumdur. Bu senaryo, yazarlar tarafından açıklanabilir modeller için klasik karşıt örneklerin en doğrudan genişletmesi olarak nitelendirilmektedir. A.2.2 ise

açıklamanın ne modelin çıktısıyla ne de gerçek sınıfla tutarlı olduğu, tüm faktörler arasında toptan bir uyumsuzluğun bulunduğu daha aşırı bir durumdur.

A.3 $(f(x) \neq y_x \wedge A_f(x) \approx A_h(x) \wedge A_f(x) \sim y_x)$: Modelin sınıflandırması yanlıştır, ancak sağlanan açıklama insan bakış açısından gerçek sınıfla tutarlıdır (Huang vd., 2023; Subramanya, Pillai ve Pirsiavash, 2019; Zhan, Zheng, Wang vd., 2022). Bu durum açıklamaya olan güveni artırırken çıktı ile açıklama arasında bir tutarsızlık da barındırır; ancak saldırgan, hem gerçek sınıfa hem de modelin yanlış çıktısına uyan belirsiz bir açıklama bularak bu tutarsızlığı gizleyebilir. Bu tür saldırılar, kullanıcıyı yanlış bir sınıfı doğru veya haklı olarak kabul etmeye ikna etmek ya da birden fazla makul çıktı sınıfının bulunduğu durumlarda kullanıcının dikkatini tercih edilen bir sınıfa yönlendirmek için kullanılabilir.

6. Açıklama Türüne Göre Saldırı Tasarımı

Vadillo, Santana ve Lozano (2025), her açıklama türünün kendine özgü zafiyetleri olduğunu vurgulamaktadır.

- Özellik tabanlı açıklamalar için vurgulanan bölgelerin insan tarafından algılanabilir, anlamlı ve ilgili olması gerekir. Saldırgan, hedefli saldırılarda yanlış sınıfı destekleyecek bir bölgeyi öne çıkarabilir; hedefsiz saldırılarda ise birden fazla bölgeyi vurgulayarak veya tekdüze bir harita üreterek belirsizlik yaratabilir (Noppel, Peter ve Wressnegger, 2023).
- Prototip tabanlı açıklamalar için en yakın prototiplerin temel özelliklerinin girdide algılanabilir olması ve çıktı sınıfıyla ilişkilendirilebilmesi gerekir; prototipler ne kadar genel (örn. semantik kavramlar) olursa, belirsiz açıklamalar üretme olasılığı o kadar artar.
- Kural tabanlı açıklamalar, modelin çıktısıyla hizalı ama güvenilir/etik görünen özellikleri kullanan açıklamalar hedeflenerek kandırılabilir; örneğin bir suç tekrarı (recidivism) tahmin modeli etik olmayan gerekçelere dayanırken açıklama etik görünebilir (Aivodji vd., 2019; Lakkaraju ve Bastani, 2020).
- Karşı olgusal açıklamalar için saldırgan, ilgisiz özelliklerde değişiklik öneren bir karşı olgusal açıklama zorlayabilir (asıl ilgili eksiklikleri gizleyerek) ya da adil olmayan/önyargılı öneriler üretebilir (Slack, Hilgard, Lakkaraju ve Singh, 2021).

7. Senaryo, Kullanıcı ve Göreve Göre Saldırı Tasarımı

Gerçekçi tehdit modellemesi amacıyla Vadillo, Santana ve Lozano (2025), sekiz farklı senaryo (S1–S8) tanımlamaktadır:

S1 →Çıktının zamanında düzeltilememesi: Otonom araçlar veya kitlesel içerik filtreleme gibi hızlı karar alma süreçlerinde insan denetimi mümkün

değildir; açıklamalar pratikte önemsizdir ve saldırganın tek amacı çıktıyı değiştirmektir (Fujiyoshi, Hirakawa ve Yamashita, 2019; Kuchipudi, Nannapaneni ve Liao, 2020).

S2 →Model hata ayıklama, geliştirme, doğrulama: Herhangi bir göreve uygulanabilir; geliştirici modelin neden yanlış tahmin yaptığını anlamaya çalışırken A.2.1 ve A.3 tipi saldırılarla yanlış sınıflandırmalar meşrulaştırılabilir, ya da A.1.1 ile modelin uygunsuz davranışları maskelenebilir (Doshi-Velez ve Kim, 2018).

S3 →Model kararlarının uzman yargısından daha bağlayıcı olması: Suç tekrarı riski veya kredi risk yönetimi gibi alanlarda S2'ye benzer stratejiler geçerlidir; ancak burada açıklamalar model son-kullanıcı tarafından işletilse dahi anlamlıdır.

S4 → Uzmanlığı olmayan kullanıcı: Karar kriterlerinin gizli veya bilinmeyen olduğu bankacılık ya da kötü amaçlı yazılım sınıflandırma gibi senaryolarda, kullanıcının deneyimsizliğinden yararlanılarak herhangi bir saldırı şeması şüphe uyandırmadan uygulanabilir; bu nedenle bu tür senaryolarda kullanılan modellerin daha güçlü güvenlik önlemleriyle donatılması gerekmektedir.

S5 →Orta düzey uzmanlık: Karmaşık tıbbi teşhis gibi zorlu senaryolarda açıklamanın, girdideki temel semantik özelliklerle ve/veya çıktı sınıfıyla yeterince tutarlı olması beklenir (A.1.1, A.2.1, A.3).

S6 →Kısmi uzmanlık: Büyük ölçekli görsel tanıma gibi hiyerarşik sınıflandırma görevlerinde kullanıcı bazı faktörlerde uzmanken bazılarında değildir; saldırganın çıktı ve açıklamaları yalnızca kullanıcının aşina olduğu unsurlarla tutarlı kılması gerekir.

S7 →Yüksek uzmanlık: Kullanıcı, tanımı gereği yanlış bir çıktıyı veya açıklamayı fark edecektir; ancak görüntüde birden fazla nesnenin bulunduğu belirsiz durumlarda (Stock ve Cisse, 2018) veya doğal dil işlemede anlam belirsizliği taşıyan görevlerde (Agirre ve Edmonds, 2006) belirsizlik yoluyla kullanıcı yine de yanıltılabilir.

S8 → Açıklamanın çıktının kendisinden daha kritik olması: Kestirimci bakım (Serradilla, Zugasti, Rodriguez ve Zurutuza, 2022) veya tıbbi teşhis (Stiglic vd., 2020) gibi alanlarda, bir sistemin neden arızalanacağını ya da bir hastanın neden riskli görüldüğünü bilmek, yalnızca sonucu bilmekten daha önemlidir; bu modeller A.1 ve A.2.1 tipi saldırılara özellikle duyarlıdır.

8. Örnek Uygulamalar: Bağlam Duyarlı Karşıt Saldırılar

Vadillo, Santana ve Lozano (2025), önerdikleri çerçeveyi somutlaştırmak amacıyla iki temsili görüntü sınıflandırma görevi üzerinde örnekler sunmaktadır: (i) COVIDx veri setiyle eğitilmiş, %92,6 doğruluğa sahip önceden eğitilmiş bir COVID-Net modeli (Wang, Lin ve Wong, 2020) kullanılarak yapılan göğüs

röntgeni (CXR) sınıflandırması ve (ii) %74,9 Top-1 doğruluğa sahip bir ResNet-50 modeliyle ImageNet veri setinde büyük ölçekli görsel tanıma. Açıklama yöntemi olarak Grad-CAM tabanlı belirginlik (salience) haritaları (Selvaraju vd., 2017) ile en yakın k=3 eğitim görüntüsüne dayalı prototip tabanlı açıklamalar kullanılmıştır.

Göğüs röntgeni görevinde, hedefli bir Projeksiyonlu Gradyan İnişi(Projected Gradient Descent, PGD) saldırısı (Madry vd., 2018) kullanılarak sınıflandırma kaybı ile açıklama kaybını (hedef belirginlik haritası ile modelin ürettiği açıklama arasındaki Öklid mesafesi) birleştiren genel bir kayıp fonksiyonu tanımlanmıştır (benzer bir yaklaşım için ayrıca bkz. Zhang vd., 2020). Elde edilen örnekler arasında şunlar yer almaktadır: yalnızca çıktıyı değiştiren fakat açıklamayı kontrolsüz bırakan klasik bir PGD saldırısı; çıktıyı "normal" olarak değiştirirken orijinal belirginlik haritasını koruyan ve böylece A.3 paradigmasını örnekleyen bir saldırı; doğru sınıflandırmayı korurken açıklamayı seçici biçimde yalnızca görüntünün bir yarısını vurgulayacak şekilde değiştiren A.1.1 tipi bir saldırı; hem çıktıyı hem açıklamayı (tutarlı fakat bilgilendirici olmayan biçimde) değiştiren A.2.1 saldırısı; ve tüm faktörler arasında toptan bir uyumsuzluk üreten A.2.2 saldırısı.

ImageNet görevinde ise sınıf belirsizliğinden yararlanılmıştır: "Great Pyrenees" köpek ırkına ait bir girdi, orijinal belirginlik haritası korunurken "Kuvasz", "White wolf" ve "Labrador Retriever" gibi görsel olarak benzer ırklara yönlendirilmiştir; bu benzerlik, her sınıfa ait en yakın prototip görüntüler üzerinden de doğrulanmıştır. Ayrıca görüntüde birden fazla kavramın (örn. bir köpek ve bir takım elbise) bulunduğu belirsiz bir girdide, açıklamanın seçilen nesneyi vurgulayacak biçimde manipüle edilmesiyle A.2.1 saldırısı örneklenmiş; aynı girdi üzerinde hem yanlış sınıf hem de belirsizliği daha da artıran bir açıklama üretilerek A.3 saldırısı gösterilmiştir. Son olarak, prototip tabanlı açıklama için, hedef sınıfa ("Doberman") ait olup aynı zamanda kaynak girdiyle görsel benzerlik taşıyan eğitim görüntülerine olan uzaklığı en aza indiren bir karşıt örnek üretilerek A.3 paradigması bir kez daha örneklenmiştir.

Bu örnekler, açıklamaların kontrol altına alınmadığı klasik saldırıların açıklamalarda tutarsızlık yaratarak insan gözlemcisini uyarabileceğini; ancak açıklamayı da hesaba katan saldırıların, modelin hem çıktısını hem de gerekçesini insan için makul kılabildiğini göstermektedir.

9. Tartışma: Yorumlanabilirlik ile gürbüzlük Arasındaki Bağlantı

Açıklamaların kullanılması, paradoksal biçimde yeni güvenlik açıkları da doğurabilir; çünkü bir modelin hangi özellikleri en çok dikkate aldığını ifşa eden bir açıklama, saldırganın daha etkili saldırılar tasarlamasına yardımcı olabilir (Sokol ve Flach, 2019; Viganò ve Magazzeni, 2020). Öte yandan literatür, karşıt olarak eğitilmiş gürbüz (robust) modellerin genellikle daha yorumlanabilir olduğunu da göstermektedir (Etmann, Lunz, Maass ve Schoenlieb, 2019; Ros ve Doshi-Velez,

2018; Tsipras, Santurkar, Engstrom, Turner ve Madry, 2019; Zhang ve Zhu, 2019). Etmann vd. (2019), girdilerin karar sınırlarına olan uzaklığı arttıkça belirginlik haritalarıyla daha iyi hizalandığını, dolayısıyla daha yorumlanabilir hale geldiğini göstermiştir. Benzer biçimde, açıklama yöntemleri de belirli savunma stratejilerine ilham kaynağı olmuş (Liu, Yang ve Hu, 2018; Tao, Ma, Liu ve Zhang, 2018; Zhang, Ye, Wang ve Yang, 2018); tersine, karşıt saldırı yöntemleri de açıklama üretmek veya analiz etmek için bir araç olarak kullanılmıştır (Elliott, Law ve Russell, 2021; Moore, Hammerla ve Watkins, 2019). Liu, Du, Guo, Liu ve Hu (2021), yorumlama yöntemleri ile karşıt saldırı ve savunmalar arasındaki benzerlikleri ayrıntılı biçimde tartışmıştır.

10. Sonuç ve Gelecek Araştırma Yöntemleri

Vadillo, Santana ve Lozano (2025)'nun ortaya koyduğu çerçeve, açıklanabilir makine öğrenmesi modellerine yönelik karşıt saldırıların, klasik "girdi-çıktı" ikilisinin ötesinde, insanın açıklamayı da değerlendirdiği çok boyutlu bir uzayda incelenmesi gerektiğini göstermektedir. Bu çerçeve, hem saldırgan açısından (hangi senaryoda hangi saldırı paradigmasının gerçekçi bir tehdit oluşturduğu) hem de savunmacı ya da geliştirici açısından (bir modelin güvenilir olması için hangi gereksinimleri karşılaması gerektiği) bir yol haritası sunmaktadır.

Yazarlar, gelecekteki araştırmalar için çeşitli yönler önermektedir. İlk olarak, tahminin güven skoru gibi ek faktörlerin çerçeveye dahil edilmesi, insanın model tahminini kabul etme eğilimini daha ince taneli biçimde modelleyebilir (Nguyen, Kim ve Nguyen, 2021). İkinci olarak, insan geri bildiriminin modelin aktif biçimde geliştirilmesinde kullanıldığı açıklama güdümlü etkileşimli öğrenme senaryoları (human-in-the-loop) daha fazla araştırılmayı hak etmektedir (Ghai, Liao, Zhang, Bellamy ve Mueller, 2021; Guo vd., 2022; Schramowski vd., 2020; Teso vd., 2023; Teso ve Kersting, 2019). Üçüncü olarak, çerçevede tanımlanan tüm saldırı paradigmalarını otomatik biçimde üretebilen, genel ve birleştirici bir saldırı algoritmasının geliştirilmesi önemli bir açık problem olarak durmaktadır. Son olarak, özellikle prototip tabanlı yaklaşımlar başta olmak üzere farklı açıklama yöntemlerinin karşıt gürbüzlüğü konusunda hâlâ sınırlı araştırma yapılmış olması, açıklama yöntemlerinin güvenilirliğini artırmak için daha derin analizlere ihtiyaç duyulduğunu göstermektedir (Vadillo, Santana ve Lozano, 2025).

Sonuç olarak, açıklanabilirlik ve karşıt gürbüzlük araştırmalarının artık birbirinden bağımsız ele alınamayacağı açıktır: bir modelin yalnızca doğru tahmin üretmesi değil, bu tahminin insan tarafından denetlenebilir, tutarlı ve manipülasyona dirençli bir gerekçeyle desteklenmesi de güvenilir yapay zekâ sistemlerinin temel bir gereksinimidir.

Kaynakça

Agirre, E., & Edmonds, P. (2006). Word sense disambiguation: Algorithms and applications (Vol. 33). Springer.

Aivodji, U., Arai, H., Fortineau, O., Gambs, S., Hara, S., & Tapp, A. (2019). Fairwashing: The risk of rationalization. Proceedings of the 36th International Conference on Machine Learning (ICML), 97, 161–170.

Aïvodji, U., Arai, H., Gambs, S., & Hara, S. (2021). Characterizing the risk of fairwashing. Advances in Neural Information Processing Systems, 34, 14822–14834.

Alvarez-Melis, D., & Jaakkola, T. (2018a). Towards robust interpretability with self-explaining neural networks. Advances in Neural Information Processing Systems, 31, 7775–7784.

Alvarez-Melis, D., & Jaakkola, T. S. (2018b). On the robustness of interpretability methods. Proceedings of the 2018 ICML Workshop on Human Interpretability in Machine Learning (WHI 2018), 66–71.

Alzantot, M., Sharma, Y., Chakraborty, S., Zhang, H., Hsieh, C.-J., & Srivastava, M. B. (2019). GenAttack: Practical black-box attacks with gradient-free optimization. Proceedings of the Genetic and Evolutionary Computation Conference (GECCO '19), 1111–1119.

Anders, C., Pasliev, P., Dombrowski, A.-K., Müller, K.-R., & Kessel, P. (2020). Fairwashing explanations with off-manifold detergent. Proceedings of the 37th International Conference on Machine Learning (ICML), 119, 314–323.

Baehrens, D., Schroeter, T., Harmeling, S., Kawanabe, M., Hansen, K., & Müller, K.-R. (2010). How to explain individual classification decisions. Journal of Machine Learning Research, 11(61), 1803–1831.

Balda, E. R., Behboodi, A., & Mathar, R. (2019). Perturbation analysis of learning algorithms: Generation of adversarial examples from classification to regression. IEEE Transactions on Signal Processing, 67(23), 6078–6091.

Ballet, V., Renard, X., Aigrain, J., Laugel, T., Frossard, P., & Detyniecki, M. (2019). Imperceptible adversarial attacks on tabular data. NeurIPS 2019 Workshop on Robust AI in Financial Services.

Belinkov, Y., & Bisk, Y. (2018). Synthetic and natural noise both break neural machine translation. International Conference on Learning Representations (ICLR).

Boopathy, A., Liu, S., Zhang, G., Liu, C., Chen, P.-Y., Chang, S., & Daniel, L. (2020). Proper network interpretability helps adversarial robustness in classification. Proceedings of the 37th International Conference on Machine Learning (ICML), 119, 1014–1023.

Bortsova, G., González-Gonzalo, C., Wetstein, S. C., Dubost, F., Katramados, I., Hogeweg, L., Liefers, B., van Ginneken, B., Pluim, J. P. W., Veta, M., Sánchez, C. I., & de Bruijne, M. (2021). Adversarial attack vulnerability of medical image analysis systems: Unexplored factors. Medical Image Analysis, 73, 102141.

Carlini, N., & Wagner, D. (2017). Towards evaluating the robustness of neural networks. Proceedings of the 2017 IEEE Symposium on Security and Privacy (SP), 39–57.

Carmichael, Z., & Scheirer, W. J. (2023). Unfooling perturbation-based post hoc explainers. Proceedings of the 37th AAAI Conference on Artificial Intelligence, 37, 6925–6934.

Cartella, F., Anunciação, O., Funabiki, Y., Yamaguchi, D., Akishita, T., & Elshocht, O. (2021). Adversarial attacks for tabular data: Application to fraud detection and imbalanced data. Proceedings of the 2021 AAAI Workshop on Artificial Intelligence Safety (SafeAI).

Chen, C., Li, O., Tao, D., Barnett, A., Rudin, C., & Su, J. K. (2019). This looks like that: Deep learning for interpretable image recognition. Advances in Neural Information Processing Systems, 32, 8930–8941.

Chen, J., Wu, X., Rastogi, V., Liang, Y., & Jha, S. (2019). Robust attribution regularization. Advances in Neural Information Processing Systems, 32, 14300–14310.

Chen, P.-Y., Zhang, H., Sharma, Y., Yi, J., & Hsieh, C.-J. (2017). ZOO: Zeroth order optimization based black-box attacks to deep neural networks without training substitute models. Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security (AISec), 15–26.

Cisse, M. M., Adi, Y., Neverova, N., & Keshet, J. (2017). Houdini: Fooling deep structured visual and speech recognition models with adversarial examples. Advances in Neural Information Processing Systems, 30, 6977–6987.

Dimanov, B., Bhatt, U., Jamnik, M., & Weller, A. (2020). You shouldn't trust me: Learning models which conceal unfairness from multiple explanation methods. Proceedings of the 24th European Conference on Artificial Intelligence (ECAI), 97, 2473–2480.

Dombrowski, A.-K., Alber, M., Anders, C., Ackermann, M., Müller, K.-R., & Kessel, P. (2019). Explanations can be manipulated and geometry is to blame. Advances in Neural Information Processing Systems, 32, 13589–13600.

Doshi-Velez, F., & Kim, B. (2018). Considerations for evaluation and generalization in interpretable machine learning. In Explainable and interpretable models in computer vision and machine learning (pp. 3–17). Springer.

Elliott, A., Law, S., & Russell, C. (2021). Explaining classifiers using adversarial perturbations on the perceptual ball. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 10693–10702.

Etmann, C., Lunz, S., Maass, P., & Schoenlieb, C. (2019). On the connection between adversarial robustness and saliency map interpretability. Proceedings of the 36th International Conference on Machine Learning (ICML), 97, 1823–1832.

Eykholt, K., Evtimov, I., Fernandes, E., Li, B., Rahmati, A., Xiao, C., Prakash, A., Kohno, T., & Song, D. (2018). Robust physical-world attacks on deep learning visual classification. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1625–1634.

Finlayson, S. G., Bowers, J. D., Ito, J., Zittrain, J. L., Beam, A. L., & Kohane, I. S. (2019). Adversarial attacks on medical machine learning. Science, 363(6433), 1287–1289.

Fujiyoshi, H., Hirakawa, T., & Yamashita, T. (2019). Deep learning-based image recognition for autonomous driving. IATSS Research, 43(4), 244–252.

Gautam, S., Boubekki, A., Hansen, S., Salahuddin, S., Jenssen, R., Höhne, M., & Kampffmeyer, M. (2022). ProtoVAE: A trustworthy self-explainable prototypical variational model. Advances in Neural Information Processing Systems, 35, 17940–17952.

Ghai, B., Liao, Q. V., Zhang, Y., Bellamy, R., & Mueller, K. (2021). Explainable active learning (XAL): Toward AI explanations as interfaces for machine teachers. Proceedings of the ACM on Human-Computer Interaction, 4(CSCW3), 235:1–235:28.

Ghalebikesabi, S., Ter-Minassian, L., DiazOrdaz, K., & Holmes, C. C. (2021). On locality of local explanation models. Advances in Neural Information Processing Systems, 34, 18395–18407.

Ghassemi, M., Oakden-Rayner, L., & Beam, A. L. (2021). The false hope of current approaches to explainable artificial intelligence in health care. The Lancet Digital Health, 3(11), e745–e750.

Ghorbani, A., Abid, A., & Zou, J. (2019). Interpretation of neural networks is fragile. Proceedings of the AAAI Conference on Artificial Intelligence, 33, 3681–3688.

Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). Automation bias: A systematic review of frequency, effect mediators, and mitigators. Journal of the American Medical Informatics Association, 19(1), 121–127.

Goodfellow, I., Shlens, J., & Szegedy, C. (2015). Explaining and harnessing adversarial examples. International Conference on Learning Representations (ICLR).

Guidotti, R., Monreale, A., Giannotti, F., Pedreschi, D., Ruggieri, S., & Turini, F. (2019). Factual and counterfactual explanations for black box decision making. IEEE Intelligent Systems, 34(6), 14–23.

Guo, L., Daly, E. M., Alkan, O., Mattetti, M., Cornec, O., & Knijnenburg, B. (2022). Building trust in interactive machine learning via user contributed interpretable rules. Proceedings of the 27th International Conference on Intelligent User Interfaces (IUI), 537–548.

Hase, P., Chen, C., Li, O., & Rudin, C. (2019). Interpretable image recognition with hierarchical prototypes. Proceedings of the AAAI Conference on Human Computation and Crowdsourcing, 7, 32–40.

Heo, J., Joo, S., & Moon, T. (2019). Fooling neural network interpretations via adversarial model manipulation. Advances in Neural Information Processing Systems, 32, 2925–2936.

Huang, W., Zhao, X., Jin, G., & Huang, X. (2023). SAFARI: Versatile and efficient evaluations for robustness of interpretability. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 1988–1998.

Hussenot, L., Geist, M., & Pietquin, O. (2020). CopyCAT: Taking control of neural policies with constant attacks. Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS '20), 548–556.

Ilyas, A., Engstrom, L., Athalye, A., & Lin, J. (2018). Black-box adversarial attacks with limited queries and information. Proceedings of the 35th International Conference on Machine Learning (ICML), 80, 2137–2146.

Kim, B., Wattenberg, M., Gilmer, J., Cai, C., Wexler, J., Viegas, F., & Sayres, R. (2018). Interpretability beyond feature attribution: Quantitative testing with concept activation vectors (TCAV). Proceedings of the 35th International Conference on Machine Learning (ICML), 80, 2668–2677.

Koh, P. W., & Liang, P. (2017). Understanding black-box predictions via influence functions. Proceedings of the 34th International Conference on Machine Learning (ICML), 70, 1885–1894.

Kos, J., Fischer, I., & Song, D. (2018). Adversarial examples for generative models. Proceedings of the 2018 IEEE Security and Privacy Workshops (SPW), 36–42.

Kuchipudi, B., Nannapaneni, R. T., & Liao, Q. (2020). Adversarial machine learning for spam filters. Proceedings of the 15th International Conference on Availability, Reliability and Security (ARES '20), 1–6.

Kuppa, A., & Le-Khac, N.-A. (2020). Black box attacks on explainable artificial intelligence (XAI) methods in cyber security. 2020 International Joint Conference on Neural Networks (IJCNN), 1–8.

Lakkaraju, H., & Bastani, O. (2020). "How do I fool you?": Manipulating user trust via misleading black box explanations. Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES '20), 79–85.

Lakkaraju, H., Kamar, E., Caruana, R., & Leskovec, J. (2019). Faithful and customizable explanations of black box models. Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (AIES '19), 131–138.

Le Merrer, E., & Trédan, G. (2020). Remote explainability faces the bouncer problem. Nature Machine Intelligence, 2(9), 529–539.

Li, O., Liu, H., Chen, C., & Rudin, C. (2018). Deep learning for case-based reasoning through prototypes: A neural network that explains its predictions. Proceedings of the AAAI Conference on Artificial Intelligence, 32(1), 3530–3537.

Lipton, Z. C. (2018). The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. Queue, 16(3), 31–57.

Liu, N., Du, M., Guo, R., Liu, H., & Hu, X. (2021). Adversarial attacks and defenses: An interpretation perspective. ACM SIGKDD Explorations Newsletter, 23(1), 86–99.

Liu, N., Yang, H., & Hu, X. (2018). Adversarial detection with model interpretation. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '18), 1803–1811.

Liu, Y., Chen, X., Liu, C., & Song, D. (2017). Delving into transferable adversarial examples and black-box attacks. International Conference on Learning Representations (ICLR).

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765–4774.

Madry, A., Makelov, A., Schmidt, L., Tsipras, D., & Vladu, A. (2018). Towards deep learning models resistant to adversarial attacks. International Conference on Learning Representations (ICLR).

Moore, J., Hammerla, N., & Watkins, C. (2019). Explaining deep learning models with constrained adversarial examples. In PRICAI 2019: Trends in artificial intelligence (pp. 43–56). Springer.

Moosavi-Dezfooli, S.-M., Fawzi, A., Fawzi, O., & Frossard, P. (2017). Universal adversarial perturbations. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 86–94.

Mopuri, K. R., Garg, U., & Babu, R. V. (2017). Fast feature fool: A data independent approach to universal adversarial perturbations. Proceedings of the British Machine Vision Conference 2017 (BMVC), 30.1–30.12.

Nauta, M., Van Bree, R., & Seifert, C. (2021). Neural prototype trees for interpretable fine-grained image recognition. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 14933–14943.

Nguyen, G., Kim, D., & Nguyen, A. (2021). The effectiveness of feature attribution methods and its correlation with automatic evaluation scores. Advances in Neural Information Processing Systems, 34, 26422–26436.

Noppel, M., Peter, L., & Wressnegger, C. (2023). Disguising attacks with explanation-aware backdoors. Proceedings of the 2023 IEEE Symposium on Security and Privacy (SP), 664–681.

Papernot, N., McDaniel, P., Goodfellow, I., Jha, S., Celik, Z. B., & Swami, A. (2017). Practical black-box attacks against machine learning. Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security (ASIA CCS '17), 506–519.

Pillai, R., Oza, P., & Sharma, P. (2019). Review of machine learning techniques in health care. Proceedings of the 2019 International Conference on Recent Innovations in Computing (ICRIC), 103–111.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16), 1135–1144.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2018). Anchors: High-precision model-agnostic explanations. Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI '18/IAAI '18/EAAI '18), 1527–1535.

Rieger, L., & Hansen, L. K. (2020). A simple defense against adversarial attacks on heatmap explanations. ICML Workshop on Human Interpretability in Machine Learning (WHI).

Robnik-Šikonja, M., & Kononenko, I. (2008). Explaining classifications for individual instances. IEEE Transactions on Knowledge and Data Engineering, 20(5), 589–600.

Ros, A. S., & Doshi-Velez, F. (2018). Improving the adversarial robustness and interpretability of deep neural networks by regularizing their input gradients. Proceedings of the AAAI Conference on Artificial Intelligence, 32(1), 1660–1669.

Ross, A. S., Hughes, M. C., & Doshi-Velez, F. (2017). Right for the right reasons: Training differentiable models by constraining their explanations. Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI), 2662–2670.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206–215.

Saralajew, S., Holdijk, L., Rees, M., Asan, E., & Villmann, T. (2019). Classification-by-components: Probabilistic modeling of reasoning over a set of components. Advances in Neural Information Processing Systems, 32, 2792–2803.

Schramowski, P., Stammer, W., Teso, S., Brugger, A., Herbert, F., Shao, X., Luigs, H.-G., Mahlein, A.-K., & Kersting, K. (2020). Making deep neural networks right for the right scientific reasons by interacting with their explanations. Nature Machine Intelligence, 2(8), 476–486.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), 618–626.

Serradilla, O., Zugasti, E., Rodriguez, J., & Zurutuza, U. (2022). Deep learning models for predictive maintenance: A survey, comparison, challenges and prospects. Applied Intelligence, 52, 10934–10964.

Sharif, M., Bhagavatula, S., Bauer, L., & Reiter, M. K. (2016). Accessorize to a crime: Real and stealthy attacks on state-of-the-art face recognition. Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS '16), 1528–1540.

Simonyan, K., Vedaldi, A., & Zisserman, A. (2014). Deep inside convolutional networks: Visualising image classification models and saliency maps. Workshop of the 2014 International Conference on Learning Representations (ICLR).

Slack, D., Hilgard, A., Lakkaraju, H., & Singh, S. (2021). Counterfactual explanations can be manipulated. Advances in Neural Information Processing Systems, 34, 62–75.

Slack, D., Hilgard, S., Jia, E., Singh, S., & Lakkaraju, H. (2020). Fooling LIME and SHAP: Adversarial attacks on post hoc explanation methods. Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, 180–186.

Sokol, K., & Flach, P. (2019). Counterfactual explanations of machine learning predictions: Opportunities and challenges for AI safety. Proceedings of the 2019 AAAI Workshop on Artificial Intelligence Safety (SafeAI), 95–99.

Stiglic, G., Kocbek, P., Fijacko, N., Zitnik, M., Verbert, K., & Cilar, L. (2020). Interpretability of machine learning-based prediction models in healthcare. WIREs Data Mining and Knowledge Discovery, 10(5), e1379.

Stock, P., & Cisse, M. (2018). ConvNets and ImageNet beyond accuracy: Understanding mistakes and uncovering biases. Computer Vision – ECCV 2018, 11210, 504–519.

Štrumbelj, E., & Kononenko, I. (2010). An efficient explanation of individual classifications using game theory. Journal of Machine Learning Research, 11(1), 1–18.

Subramanya, A., Pillai, V., & Pirsiavash, H. (2019). Fooling network interpretation in image classification. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2020–2029.

Sun, S., Song, B., Cai, X., Du, X., & Guizani, M. (2022). CAMA: Class activation mapping disruptive attack for deep neural networks. Neurocomputing, 500, 989–1002.

Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., & Fergus, R. (2014). Intriguing properties of neural networks. International Conference on Learning Representations (ICLR).

Tang, R., Liu, N., Yang, F., Zou, N., & Hu, X. (2022). Defense against explanation manipulation. Frontiers in Big Data, 5, 704203.

Tao, G., Ma, S., Liu, Y., & Zhang, X. (2018). Attacks meet interpretability: Attribute-steered detection of adversarial samples. Advances in Neural Information Processing Systems, 31, 7717–7728.

Teso, S., Alkan, Ö., Stammer, W., & Daly, E. (2023). Leveraging explanations in interactive machine learning: An overview. Frontiers in Artificial Intelligence, 6, 1066049.

Teso, S., & Kersting, K. (2019). Explanatory interactive machine learning. Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (AIES), 239–245.

Thys, S., Ranst, W. V., & Goedemé, T. (2019). Fooling automated surveillance cameras: Adversarial patches to attack person detection. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 49–55.

Tsipras, D., Santurkar, S., Engstrom, L., Turner, A., & Madry, A. (2019). Robustness may be at odds with accuracy. International Conference on Learning Representations (ICLR).

Ustun, B., Spangher, A., & Liu, Y. (2019). Actionable recourse in linear classification. Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19), 10–19.

Vadillo, J., Santana, R., & Lozano, J. A. (2025). Adversarial attacks in explainable machine learning: A survey of threats against models and humans. WIREs Data Mining and Knowledge Discovery, 15(1), e1567. https://doi.org/10.1002/widm.1567

Viganò, L., & Magazzeni, D. (2020). Explainable security. Proceedings of the 2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), 293–300.

Vreš, D., & Robnik-Šikonja, M. (2022). Preventing deception with explanation methods using focused sampling. Data Mining and Knowledge Discovery.

Wachter, S., Mittelstadt, B., & Russell, C. (2017). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. Harvard Journal of Law & Technology, 31(2), 842–887.

Wang, L., Lin, Z. Q., & Wong, A. (2020). COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. Scientific Reports, 10(1), 19549.

Xie, C., Wang, J., Zhang, Z., Zhou, Y., Xie, L., & Yuille, A. (2017). Adversarial examples for semantic segmentation and object detection. Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), 1378–1387.

Xu, K., Zhang, G., Liu, S., Wang, Q., Sun, J., Yang, F., Yu, B., Kailkhura, B., Lin, X., & Chen, X. (2020). Adversarial T-shirt! Evading person detectors in a physical world. Proceedings of the 2020 European Conference on Computer Vision (ECCV), 665–681.

Yosinski, J., Clune, J., Nguyen, A., Fuchs, T., & Lipson, H. (2015). Understanding neural networks through deep visualization. 2015 ICML Workshop on Deep Learning.

Yuan, X., He, P., Zhu, Q., & Li, X. (2019). Adversarial examples: Attacks and defenses for deep learning. IEEE Transactions on Neural Networks and Learning Systems, 30(9), 2805–2824.

Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. Computer Vision—ECCV 2014, 8689, 818–833.

Zhan, Y., Zheng, B., Wang, Q., Mou, N., Guo, B., Li, Q., Shen, C., & Wang, C. (2022). Towards black-box adversarial attacks on interpretable deep learning systems. Proceedings of the 2022 IEEE International Conference on Multimedia and Expo (ICME), 1–6.

Zhang, C., Ye, Z., Wang, Y., & Yang, Z. (2018). Detecting adversarial perturbations with saliency. Proceedings of the IEEE 3rd International Conference on Signal and Image Processing (ICSIP), 271–275.

Zhang, T., & Zhu, Z. (2019). Interpreting adversarially trained convolutional neural networks. Proceedings of the 36th International Conference on Machine Learning (ICML), 97, 7502–7511.

Zhang, X., Wang, N., Shen, H., Ji, S., Luo, X., & Wang, T. (2020). Interpretable deep learning under fire. Proceedings of the 29th USENIX Security Symposium (USENIX Security 20), 1659–1676.

Zhang, Y., Tiňo, P., Leonardis, A., & Tang, K. (2021). A survey on neural network interpretability. IEEE Transactions on Emerging Topics in Computational Intelligence, 5(5), 726–742.

Zheng, H., Fernandes, E., & Prakash, A. (2019). Analyzing the interpretability robustness of self-explaining models. ICML 2019 Security and Privacy of Machine Learning Workshop.

5. Bölüm

Metasezgisel Optimizasyon Algoritmaları Kullanarak Çoklu Odak Görüntü Füzyonu

Eyüp SIRAMKAYA[1]

Özet

Kamera lenslerinin sahip olduğu doğal alan derinliği sınırlaması nedeniyle, farklı derinliklerde bulunan nesnelerin tek bir pozlama ile eşzamanlı olarak keskin biçimde görüntülenmesi mümkün değildir. Bu sınırlamanın giderilmesine yönelik olarak, farklı odak mesafelerinde elde edilen birden fazla görüntünün keskin bölgelerinin birleştirildiği çoklu odak görüntü füzyonu (multi-focus image fusion) yaklaşımından yararlanılmaktadır. Bu bölümde, blok tabanlı görüntü füzyonunda kritik bir parametre olan optimal blok boyutunun belirlenmesi problemi ele alınmaktadır. Blok boyutunun füzyon kalitesi üzerindeki belirleyici etkisi ve sabit bir blok boyutunun farklı görüntüler için her durumda uygun olmayabileceği dikkate alınarak, metasezgisel optimizasyon algoritmalarının söz konusu problemdeki kullanımı incelenmektedir. Özellikle Yapay Arı Kolonisi (ABC), Diferansiyel Evrim (DE), Parçacık Sürü Optimizasyonu (PSO) ve Akbulut (2025) tarafından karşılaştırılan diğer metasezgisel algoritmaların performansları; dokuz nesnel kalite metriği ve öznel değerlendirmeler çerçevesinde değerlendirilmektedir.

1. Giriş

Görüntü füzyonu (image fusion), aynı sahneye ilişkin birden fazla görüntüden elde edilen bilgilerin bütünleştirilerek tek, kapsamlı ve yüksek kaliteli bir görüntünün oluşturulması işlemidir. Bu teknik; tıbbi görüntüleme, uzaktan algılama, robotik görme sistemleri, dijital fotoğrafçılık, güvenlik ve gözetleme gibi çok sayıda uygulama alanında önemli bir yere sahiptir (Aslantas ve Kurban, 2010).

Gerçek optik sistemlerde kullanılan mercekler, yalnızca belirli bir mesafede bulunan nesneleri keskin biçimde odaklayabilmektedir. Alan derinliği (depth of field) olarak ifade edilen bu sınırlama, farklı uzaklıklardaki nesneleri içeren sahnelerde tüm bölgelerin tek bir görüntü içerisinde eşzamanlı olarak

[1] Bilgisayar Mühendisliği, Mühendislik-Mimarlık Fakültesi, Nevşehi Hacı Bektaş Veli Üniversitesi, TÜRKİYE, ORCID: 0000-0002-6011-7302

odaklanmasını engellemektedir (Aslantas ve Kurban, 2010; Banharnsakun, 2019). Örneğin, ön plandaki bir nesneye odaklanıldığında arka plan bulanıklaşırken, arka plana odaklanılması durumunda ön planın netliği azalmaktadır. Bu durum, özellikle mikroskopi, fotoğrafçılık ve makine görüşü uygulamalarında önemli kısıtlamalara neden olmaktadır.

Çoklu odak görüntü füzyonu, söz konusu sınırlamanın giderilmesine yönelik etkili bir yaklaşım olarak geliştirilmiştir. Temel yaklaşım, aynı sahnenin farklı odak mesafelerinde elde edilen en az iki görüntüsündeki keskin bölgelerin belirlenerek birleştirilmesi ve böylece her bölgede net (everywhere-in-focus) tek bir görüntünün oluşturulmasıdır. Bu sayede sahnenin tamamı, bulanık bölgeler en aza indirilecek biçimde, yüksek çözünürlüklü ve bilgi açısından zengin tek bir görüntü içerisinde temsil edilebilmektedir.

Görüntü füzyon yöntemleri genel olarak iki temel kategori altında incelenmektedir: dönüşüm alanı tabanlı yöntemler ve uzamsal alan tabanlı yöntemler. Dalgacık dönüşümü (wavelet transform), Laplace piramidi ve Fourier dönüşümü gibi teknikler dönüşüm alanı tabanlı yöntemlerin başlıca örnekleri arasında yer almaktadır. Bu yöntemler çoklu çözünürlüklü analiz olanağı sağlamakla birlikte, yüksek hesaplama maliyeti ve konum değişmezliğinin sağlanamaması (shift-variance) gibi çeşitli dezavantajlara sahiptir (Aslantas ve Kurban, 2010).

Uzamsal alan tabanlı yöntemler ise işlemleri doğrudan piksel düzeyinde gerçekleştirmeleri nedeniyle daha yalın ve hesaplama açısından daha verimli bir yapı sunmaktadır. Bu yöntemler içerisinde blok tabanlı füzyon yaklaşımları öne çıkmaktadır. Blok tabanlı yöntemlerde kaynak görüntüler m×n boyutlu bloklara ayrılmakta, ardından her bloğun keskinlik değeri hesaplanarak daha keskin olan blok birleşik görüntüye aktarılmaktadır (Aslantas ve Kurban, 2010; Banharnsakun, 2019). Bu yaklaşımda en kritik parametrelerden biri blok boyutudur. Çok küçük blok boyutları keskinlik ölçümünün güvenilirliğini azaltabilirken, çok büyük blok boyutları hem odaklanmış hem de odaklanmamış bölgeleri aynı blok içerisinde barındırarak füzyon kalitesinin düşmesine neden olabilmektedir (Akbulut, 2025).

Bu çerçevede, optimal blok boyutunun belirlenmesi bir optimizasyon problemi olarak ele alınmakta ve metasezgisel algoritmalar bu probleme yönelik etkili çözüm olanakları sunmaktadır. Bu bölümde, blok tabanlı çoklu odak görüntü füzyonunun matematiksel temelleri, kullanılan kalite metrikleri ve çeşitli metasezgisel optimizasyon algoritmalarının söz konusu problemdeki performansları kapsamlı biçimde ele alınmaktadır.

2. Blok Tabanlı Çoklu Odak Görüntü Füzyonu

2.1. Temel İlkeler ve Yöntem

Blok tabanlı çoklu odak görüntü füzyonunun temel amacı, farklı odak mesafelerinde elde edilen iki kaynak görüntüden (A ve B) daha keskin bölgeleri seçerek her yerde net bir birleşik görüntü (F) oluşturmaktır. Yöntem temelde üç aşamadan oluşmaktadır: (1) kaynak görüntülerin m×n boyutlu bloklara ayrılması, (2) her bloğun keskinlik değerinin bir kriter fonksiyonu aracılığıyla belirlenmesi ve (3) daha keskin olan bloğun birleşik görüntüye aktarılması (Aslantas ve Kurban, 2010).

Füzyon kuralı matematiksel olarak şu şekilde ifade edilmektedir (Aslantas ve Kurban, 2010):

$$F_i = A_i \;\; eğer\ KriterA(i) \geq KriterB(i), \;\; aksi\ takdirde\ F_i = B_i$$

Burada F_i, birleşik görüntünün i'inci bloğunu; A_i ve B_i ise sırasıyla A ve B kaynak görüntülerinin i'inci bloklarını temsil etmektedir. KriterA(i) ve KriterB(i), ilgili blokların keskinlik değerlerini ifade etmektedir. Keskinlik değerlerinin eşit olması durumunda blokların ortalamasının alınması mümkündür (Çakıroğlu ve ark., 2023).

Blok tabanlı yöntemin başlıca avantajlarından biri, uzamsal alanda doğrudan piksel değerleri üzerinde çalışması ve bu nedenle dönüşüm tabanlı yöntemlere kıyasla kayıt (registration) sorunlarından daha az etkilenmesidir. Bunun yanı sıra, uzamsal alanda çalışması ve blokların bağımsız olarak işlenebilmesi, yöntemin hesaplama bakımından görece yalın olmasını ve paralel işlemeye uygunluk göstermesini sağlamaktadır (Akbulut, 2025).

2.2. Keskinlik Kriter Fonksiyonları

Blok tabanlı füzyonun başarısı, büyük ölçüde kullanılan keskinlik ölçüm fonksiyonunun etkinliğine bağlıdır. Literatürde farklı amaç ve özelliklere sahip çok sayıda kriter fonksiyonu önerilmiş olup, bunların yaygın olarak kullanılan başlıca örnekleri aşağıda sunulmaktadır:

2.2.1. Uzamsal Frekans

Uzamsal frekans (Spatial Frequency, SF), bir görüntüdeki genel aktivite düzeyini karakterize eden ve yatay ile dikey doğrultulardaki piksel farklarının karelerine dayanan bir metriktir (Eskicioglu ve Fisher, 1995). M×N boyutlu bir F görüntüsü için uzamsal frekans aşağıdaki biçimde hesaplanmaktadır (Aslantas ve Kurban, 2010):

$$SF = \sqrt{(RF^2 + CF^2)}$$
$$RF = [1/(M \times N) \times \Sigma_i \Sigma_j (F(i,j) - F(i,j-1))^2]^{(1/2)}$$
$$CF = [1/(M \times N) \times \Sigma_i \Sigma_j (F(i,j) - F(i-1,j))^2]^{(1/2)}$$

Burada RF satır frekansını, CF ise sütun frekansını temsil etmektedir. Görüntünün keskinliği arttıkça yüksek frekanslı içerik miktarı da arttığından SF değeri yükselmekte; bulanıklığın artmasıyla birlikte SF değeri azalmaktadır. Bu özellik, SF metriğinin keskin blokların belirlenmesinde kullanılmasını uygun hâle getirmektedir (Banharnsakun, 2019; Çakıroğlu ve ark., 2023).

2.2.2. Varyans

Varyans (Variance, VAR), bir görüntü bloğundaki piksel yoğunluk değerlerinin ortalama değer etrafındaki dağılımını ölçen bir metriktir. Odaklanmış bölgeler çoğunlukla daha yüksek kontrast ve ayrıntı içerdiğinden, bu bölgelerde varyans değerlerinin daha yüksek olması beklenmektedir (Aslantas ve Kurban, 2010):

$$VAR = 1/(M \times N) \times \Sigma_i \Sigma_j (f(i,j) - \bar{f})^2$$

Akbulut (2025) tarafından gerçekleştirilen karşılaştırmalı çalışmada varyans (QMVar), yalnızca bir kalite metriği olarak değil, aynı zamanda optimizasyon sürecinde oluşturulan birleşik görüntünün fitness değerini belirleyen temel ölçüt olarak kullanılmıştır.

2.2.3. Değiştirilmiş Laplace Toplamı

Değiştirilmiş Laplace toplamı (Sum-Modified Laplacian, SML), kenar bilgisini yüksek frekanslı bir tahmin aracı olarak kullanan bir keskinlik metriğidir. Standart Laplace operatöründen farklı olarak yatay ve dikey ikinci türevlerin mutlak değerlerini toplaması, işaretlerin birbirini götürmesinden kaynaklanan iptal sorununu ortadan kaldırır (Aslantas ve Kurban, 2010; Banharnsakun, 2019):

$$\nabla^2 f(i,j) = |2f(i,j) - f(i-1,j) - f(i+1,j)| + |2f(i,j) - f(i,j-1) - f(i,j+1)|$$
$$SML = \Sigma_i \Sigma_j \nabla^2 f(i,j)$$

Banharnsakun (2019), SML metriğinin çoklu odak görüntü füzyonundaki kullanımını incelemiş ve bu metriği farklı keskinlik ölçütleriyle karşılaştırmıştır.

2.3. Optimal Blok Boyutunun Önemi

Blok boyutu (m×n), füzyon kalitesini doğrudan etkileyen kritik parametrelerden biridir. Çok küçük bloklar kullanıldığında kriter fonksiyonları keskin ve bulanık bölgeler arasında hatalı ayrımlar gerçekleştirebilmektedir; bunun temel nedeni, küçük blokların yeterli düzeyde uzamsal bağlam bilgisi sağlayamamasıdır. Buna karşılık, çok büyük blokların tercih edilmesi durumunda bir bloğun hem odaklanmış hem de bulanık bölgeleri birlikte içermesi olası hâle gelmekte ve bu durum birleşik görüntüde bulanık alanların oluşmasına neden olabilmektedir (Aslantas ve Kurban, 2010).

Akbulut (2025) çalışmasında, blok boyutunun 8 pikselin altına düşmesi veya 24 pikselin üzerine çıkması durumunda keskin blokların belirlenmesine ilişkin kalite metriklerinin olumsuz etkilenebileceği ve büyük blokların bulanık bölgeleri de içerebileceği belirtilmektedir. Bu doğrultuda, deneysel çalışmada blok boyutu için alt sınır 8, üst sınır ise 24 olarak belirlenmiştir. Ayrıca sabit bir blok boyutunun farklı görüntülerdeki keskin ve bulanık bölgelerin geometrik özelliklerine her zaman uyum sağlayamaması, blok boyutunun optimizasyon yoluyla belirlenmesini gerekli kılmaktadır (Akbulut, 2025).

Aslantas ve Kurban (2010), Diferansiyel Evrim algoritmasını blok boyutu optimizasyonunda kullanarak bu yaklaşımın literatürdeki önemli örneklerinden birini ortaya koymuştur. Söz konusu çalışmada, blok boyutunun optimizasyon yoluyla belirlenmesinin füzyon performansını iyileştirebileceği gösterilmiş; sonraki çalışmalarda ise farklı metasezgisel algoritmaların aynı problem üzerindeki performansları karşılaştırılmıştır (Banharnsakun, 2019; Çakıroğlu ve ark., 2023; Akbulut, 2025).

3. Metasezgisel Optimizasyon Algoritmaları

Metasezgisel algoritmalar (metaheuristic algorithms), doğadan veya fiziksel olgulardan esinlenilerek karmaşık, çok boyutlu ve doğrusal olmayan problemlerin çözümünde kullanılmak üzere geliştirilen üst düzey sezgisel yöntemlerdir. Bu algoritmalar, gradyan bilgisine ihtiyaç duymaksızın geniş çözüm uzaylarını etkin biçimde araştırabilmekte ve yerel minimumlara takılma olasılığını azaltmaya yönelik mekanizmalar içermektedir (Akbulut, 2025). Çoklu odak görüntü füzyonunda blok boyutu optimizasyonu, yüksek boyutlu ve doğrusal olmayan bir problem sınıfında yer aldığından, metasezgisel yaklaşımlar bu problem için uygun bir optimizasyon çerçevesi sunmaktadır.

3.1. Diferansiyel Evrim Algoritması

Diferansiyel Evrim (Differential Evolution, DE) algoritması, Storn ve Price (1997) tarafından geliştirilen, popülasyon tabanlı ve stokastik bir doğrudan arama

algoritmasıdır. DE, sürekli çözüm uzaylarında doğrusal olmayan, türevlenemeyen ve çok modlu fonksiyonların optimizasyonunda etkili biçimde kullanılabilen bir yöntemdir.

Algoritmanın temel aşamaları şu şekilde sıralanabilir: (1) başlangıç popülasyonunun rastgele oluşturulması, (2) mutasyon operatörü aracılığıyla aday vektörlerin üretilmesi, (3) çaprazlama operatörü ile çözüm çeşitliliğinin desteklenmesi ve (4) daha başarılı bireylerin bir sonraki nesle aktarılmasını sağlayan seçim işleminin gerçekleştirilmesi. Mutasyon operatörü aşağıdaki formülle ifade edilmektedir (Aslantas ve Kurban, 2010):

$$v_i,G+1 = xr1,G + F \times (xr2,G - xr3,G)$$

Burada r1, r2, r3 rastgele seçilen farklı bireyler, F ∈ [0, 2] ölçekleme faktörü ve G mevcut nesli göstermektedir. Aslantas ve Kurban (2010), Diferansiyel Evrim algoritmasını çoklu odak görüntü füzyonunda blok boyutu optimizasyonuna uygulamıştır.

3.2. Yapay Arı Kolonisi Algoritması

Yapay Arı Kolonisi (Artificial Bee Colony, ABC) algoritması, Karaboga ve Basturk (2007) tarafından bal arılarının besin arama davranışından esinlenilerek geliştirilmiştir. Algoritma, işçi arılar (employed bees), gözetleyici arılar (onlooker bees) ve keşifçi arılar (scout bees) olmak üzere üç arı grubunun etkileşimini modellemektedir.

İşçi arılar mevcut çözümleri yerel olarak iyileştirirken, gözetleyici arılar işçi arılardan elde edilen bilgileri değerlendirerek daha umut verici çözüm bölgelerine yönelmektedir. Keşifçi arılar ise tükenen çözümlerin yerine rastgele yeni çözümler oluşturmaktadır. Bu yapı, algoritmanın yerel arama (sömürü) ile küresel arama (keşif) arasında bir denge kurmasına olanak sağlamaktadır. ABC algoritmasının önemli avantajlarından biri, stokastik seçim mekanizmasının da katkısıyla yerel minimumlara takılma riskinin görece düşük olmasıdır (Akbulut, 2025).

Banharnsakun (2019), standart ABC algoritmasına üç temel geliştirme ekleyerek En-İyi-Şimdiye-Kadar ABC (Best-So-Far ABC, BSF-ABC) varyantını önermiştir. Bu geliştirmeler; (1) gözetleyici arıların popülasyondaki en iyi konuma yönlendirilmesini sağlayan En-İyi-Şimdiye-Kadar yöntemi (BSF), (2) çözümün yerel minimumda sıkışması durumunda arama davranışını dinamik olarak değiştiren Ayarlanabilir Arama Yarıçapı (ASR) ve (3) fitness hesabındaki sayısal hassasiyet sorunlarını azaltmayı amaçlayan Nesnel Değer Tabanlı Karşılaştırma (OBC) mekanizmasından oluşmaktadır. Deneysel sonuçlar, BSF-

ABC'nin standart ABC, PSO, DE ve GA algoritmalarına kıyasla daha düşük RMSE ve daha yüksek uzamsal frekans değerleri ürettiğini ve daha hızlı yakınsama sağladığını göstermiştir (Banharnsakun, 2019).

ABC algoritmasının blok tabanlı füzyondaki güçlü performansı, farklı araştırma grupları tarafından elde edilen bulgularla da desteklenmektedir. Çakıroğlu ve ark. (2023), ABC ve JSA gibi sürü tabanlı algoritmaların fizik tabanlı algoritmalara kıyasla genel olarak daha yüksek füzyon kalitesi sağladığını göstermiştir. Akbulut (2025) ise 16 farklı metasezgisel algoritmayı karşılaştırdığı çalışmasında ABC'nin nesnel metrikler ile öznel değerlendirmeleri birlikte dikkate alan genel performans bakımından en yüksek toplam skoru elde ettiğini ortaya koymuştur.

3.3. Parçacık Sürü Optimizasyonu

Parçacık Sürü Optimizasyonu (Particle Swarm Optimization, PSO), Kennedy ve Eberhart (1995) tarafından kuş sürülerinin ve balık sürülerinin kolektif davranışlarından esinlenilerek geliştirilen popülasyon tabanlı bir optimizasyon algoritmasıdır. Her parçacık, kendi en iyi deneyimini ve sürünün ulaştığı küresel en iyi konumu dikkate alarak konumunu güncellemektedir.

Blok boyutu optimizasyonu açısından PSO'nun bazı pratik sınırlılıkları bulunmaktadır. Banharnsakun (2019), PSO'nun yeterli düzeyde keşif sağlayamaması durumunda erken yakınsama (premature convergence) problemiyle karşılaşabileceğini belirtmiştir. Çakıroğlu ve ark. (2023) da PSO'nun bazı görüntü çiftlerinde bulanık bölgeler içeren füzyon sonuçları üretebildiğini gözlemlemiştir. Akbulut (2025) çalışmasında ise PSO'nun bazı görüntü çiftlerinde üst sıralarda yer alırken diğerlerinde daha düşük performans gösterdiği ve genel değerlendirmede orta sıralarda konumlandığı görülmektedir.

3.4. Diğer Metasezgisel Algoritmalar

Literatürde blok tabanlı çoklu odak görüntü füzyonuna uygulanmış çok sayıda farklı metasezgisel algoritma bulunmaktadır. Akbulut (2025), aşağıda belirtilen 16 algoritmayı sistematik bir karşılaştırma kapsamında incelemiştir:

Biyocoğrafya Tabanlı Optimizasyon (BBO), Simon (2008) tarafından önerilmiş ve türlerin coğrafi dağılımından esinlenilmiştir. Güve-Alev Optimizasyonu (MFO), güvelerin ışık kaynaklarına doğru spiral yörüngeler boyunca hareket etme davranışını modellemekte ve Mirjalili (2015) tarafından geliştirilmiştir. Öğretim-Öğrenme Tabanlı Optimizasyon (TLBO), algoritmaya özgü parametrelere ihtiyaç duymadan sınıf ortamındaki öğretmen-öğrenci etkileşimini temel almaktadır (Rao ve ark., 2011). Çoklu Evren Optimize Edici (MVO), kara delik, beyaz delik ve solucan deliği kavramlarını matematiksel

operatörler aracılığıyla modellemektedir (Mirjalili ve ark., 2016). Guguk Kuşu Araması (CS), guguk kuşlarının üreme parazitliği davranışını ve Lévy uçuşlarını modellemektedir (Yang ve Deb, 2009). Ateşböceği Algoritması (FA), ateşböceklerinin parlaklık tabanlı çekim mekanizmasından yararlanmaktadır (Łukasik ve Zak, 2009). Yusufçuk Algoritması (DA) ise yusufçukların sürü davranışlarını ve avlanma stratejilerini modellemektedir (Mirjalili, 2016).

Çakıroğlu ve ark. (2023) ise Archimedes Optimizasyon Algoritması (AOA), Atomik Orbital Arama (AOS) ve Denge Optimize Edici (EO) gibi fizik tabanlı algoritmaları; ABC, PSO ve JSA gibi sürü tabanlı algoritmalarla karşılaştırmıştır. Elde edilen sonuçlar, sürü tabanlı algoritmaların genel olarak daha yüksek füzyon performansı sergilediğini ortaya koymuştur. Fizik tabanlı algoritmalar arasında EO bazı metriklerde daha başarılı sonuçlar verirken, AOS hesaplama süresi bakımından en hızlı yakınsama davranışını göstermiştir.

4. Optimizasyon Tabanlı Füzyon Sürecinin Adım Adım Açıklaması

Metasezgisel optimizasyon algoritmalarının blok tabanlı çoklu odak görüntü füzyonuna entegrasyonu belirli bir işlem prosedürü çerçevesinde gerçekleştirilmektedir. Aslantas ve Kurban (2010) tarafından geliştirilen blok tabanlı yaklaşım ile Akbulut (2025) tarafından uygulanan optimizasyon çerçevesi birlikte değerlendirildiğinde, süreç aşağıdaki temel adımlar üzerinden açıklanabilir:

Adım 1 — Parametre Yapılandırması: Optimizasyon algoritmasının kontrol parametreleri ve yakınsama kriterleri belirlenmektedir. Akbulut (2025) tarafından gerçekleştirilen karşılaştırmalı analizde, algoritmalar arasında adil bir karşılaştırma yapılabilmesi amacıyla tüm algoritmalar için maksimum iterasyon sayısı 100 ve popülasyon büyüklüğü 20 olarak belirlenmiştir.

Adım 2 — İlk Popülasyonun Oluşturulması: Her birey, m ve n boyutlarından oluşan bir blok boyutu çözümünü temsil etmektedir. Akbulut (2025) çalışmasında blok boyutları için alt sınır 8 ve üst sınır 24 olarak belirlenmiş ve başlangıç çözümleri bu aralık içerisinde oluşturulmuştur.

Adım 3 — Fitness Değerlendirmesi: Her bireyin temsil ettiği blok boyutu kullanılarak kaynak görüntüler m×n boyutlu bloklara ayrılmaktadır. Her karşılık gelen blok için uzamsal frekans (QMSF) değeri hesaplanmakta ve QMSF değeri daha yüksek olan blok birleşik görüntüye aktarılmaktadır. Daha sonra oluşturulan birleşik görüntünün varyans (QMVar) değeri hesaplanmakta ve bu değer optimizasyon algoritmasının fitness ölçütü olarak kullanılmaktadır (Çakıroğlu ve ark., 2023; Akbulut, 2025).

Adım 4 — Optimizasyon Operatörleri: Seçilen algoritmaya özgü operatörler uygulanarak yeni çözümler üretilmekte ve mevcut popülasyon

güncellenmektedir. Bu aşamanın gerçekleştirilme biçimi kullanılan algoritmaya göre değişmektedir. DE'de mutasyon, çaprazlama ve seçim; ABC'de işçi, gözetleyici ve keşifçi arı aşamaları; PSO'da ise hız ve konum güncellemeleri uygulanmaktadır.

Adım 5 — Yakınsama Kontrolü: Maksimum iterasyon sayısına ulaşılıp ulaşılmadığı veya önceden belirlenen iyileştirme ölçütünün sağlanıp sağlanmadığı kontrol edilmektedir. Koşul sağlanmamışsa süreç Adım 3'e döndürülmekte; koşul sağlanmışsa optimizasyon sonlandırılmakta ve elde edilen en iyi blok boyutu kullanılarak oluşturulan birleşik görüntü çıktı olarak sunulmaktadır.

Bu prosedürün temelinde, QMSF kullanılarak oluşturulan birleşik görüntünün QMVar değeri üzerinden değerlendirilmesi bulunmaktadır. Akbulut (2025) çalışmasında kullanılan fitness fonksiyonu, oluşturulan birleşik görüntünün QMVar değerini maksimize etmeyi amaçlamaktadır. Böylece her aday blok boyutu için oluşturulan birleşik görüntünün kalite düzeyi sayısal olarak değerlendirilmekte ve optimizasyon algoritması daha yüksek QMVar değerlerine karşılık gelen blok boyutlarına yönlendirilmektedir.

$$Qvar(F) = 1/(M \times N) \times \Sigma_i \Sigma_j (F(i,j) - \mu)^2$$

Burada μ, birleşik görüntünün ortalama piksel değerini; M ve N ise görüntünün sırasıyla boyutlarını ifade etmektedir. Buna göre, her aday blok boyutu için oluşturulan birleşik görüntünün QMVar değeri hesaplanmakta ve optimizasyon sürecinde daha yüksek QMVar değerine sahip çözümler tercih edilmektedir.

5. Kalite Metrikleri

Çoklu odak görüntü füzyonunun başarısının değerlendirilmesinde çeşitli nesnel kalite metriklerinden yararlanılmaktadır. Akbulut (2025), tek bir kalite metriğinin farklı görüntülerin ve farklı füzyon senaryolarının tüm özelliklerini yeterli biçimde temsil edemeyeceğini belirttiğinden, daha kapsamlı bir değerlendirme için birden fazla metriğin birlikte kullanılmasını benimsemiştir.

5.1. Referans Tabanlı Metrikler

Referans görüntünün bulunduğu yapay test senaryolarında, füzyon kalitesi doğrudan referans görüntü ile karşılaştırılabilmektedir. Bu tür değerlendirmelerde MSE, PSNR ve MI gibi referans tabanlı ölçütlerden yararlanılmaktadır (Aslantas ve Kurban, 2010; Banharnsakun, 2019).

Ortalama Kare Hatası (MSE): Füzyon görüntüsü ile referans görüntü arasındaki piksel düzeyindeki farkların karelerinin ortalamasını ifade etmektedir. MSE

değerinin düşük olması, füzyon görüntüsünün referans görüntüye daha yakın olduğunu ve dolayısıyla daha iyi bir füzyon kalitesine işaret ettiğini göstermektedir.

Tepe Sinyal-Gürültü Oranı (PSNR): Füzyon görüntüsündeki sinyal gücünün gürültüye oranını logaritmik ölçekte ifade etmektedir. Daha yüksek PSNR değerleri, füzyon görüntüsünün referans görüntü ile daha yüksek düzeyde benzerlik gösterdiğine işaret etmektedir.

Karşılıklı Bilgi (MI): Kaynak görüntüler ile füzyon görüntüsü arasındaki ortak bilgi miktarını ölçen bir metriktir. Daha yüksek MI değerleri, kaynak görüntülerde bulunan bilginin daha büyük bir bölümünün füzyon görüntüsüne aktarılabildiğini göstermektedir.

5.2. Referanssız Metrikler

Gerçek çoklu odak görüntülerinde ideal bir referans görüntünün bulunmaması durumunda, füzyon kalitesinin değerlendirilmesi için referans gerektirmeyen metriklerden yararlanılmaktadır. Akbulut (2025) çalışmasında bu amaçla dokuz farklı nesnel kalite metriği kullanılmıştır.

Kenar Tabanlı Metrik (QMQAB/F): Kaynak görüntülerde bulunan kenar bilgisinin füzyon görüntüsüne ne ölçüde aktarıldığını değerlendiren ve Xydeas ve Petrovic (2000) tarafından tanımlanan bir metriktir.

Chen-Blum Metriği (QMCB): İnsan görsel algısından esinlenilerek geliştirilen ve beş temel aşamadan oluşan bu metrik, kontrast duyarlılığı filtrelemesi ile yerel kontrast hesaplamalarını içermektedir (Chen ve Blum, 2009).

Farkların Korelasyonlarının Toplamı (QMSCD): Kaynak görüntüler ile füzyon görüntüsü arasındaki fark görüntülerinin korelasyonlarını toplamakta ve füzyon görüntüsünün kaynak görüntülerden ne ölçüde bilgi aktardığını değerlendirmektedir (Aslantas ve Bendes, 2015).

Doğrusal Olmayan Korelasyon Bilgi Entropisi (QMW): Girdi ve çıktı görüntüleri arasındaki doğrusal olmayan bağımlılık ilişkilerini entropi temelli bir yaklaşım aracılığıyla ölçmektedir.

Varyans (QMVar), Entropi (QME), Uzamsal Frekans (QMSF) ve Standart Sapma (QMSD): Bu metrikler, birleşik görüntünün piksel dağılımı, bilgi içeriği ve ayrıntı düzeyi gibi farklı özelliklerini karakterize etmektedir. Akbulut (2025), bu ölçütleri QMQAB/F, QMCB, QMW, QMMI ve QMSCD ile birlikte kullanarak dokuz boyutlu nesnel bir değerlendirme gerçekleştirmiştir.

5.3. Skor Tabanlı Karşılaştırma Yöntemi

Birden fazla metriğin birlikte kullanıldığı karşılaştırmalı çalışmalarda sonuçların yorumlanması güçleşebilmektedir; çünkü bir algoritma belirli bir metrikte üstün performans gösterirken başka bir metrikte daha düşük sonuçlar verebilmektedir. Bu

güçlüğün azaltılması amacıyla Akbulut (2025), her algoritmanın her kalite metriğindeki sıralamasını puana dönüştüren skor tabanlı bir karşılaştırma yöntemi kullanmıştır. En iyi sonuca 16, en düşük sonuca ise 1 puan verilmekte; metriklerden elde edilen puanlar toplanarak algoritmaların genel sıralaması oluşturulmaktadır. Böylece farklı kalite metriklerinden elde edilen sonuçların tek bir toplam skor üzerinden daha anlaşılır ve dengeli biçimde karşılaştırılması mümkün hâle gelmektedir.

6. Deneysel Sonuçlar ve Karşılaştırmalar

6.1. Veri Setleri

Akbulut (2025) çalışmasında, Lytro, MFI-WHU ve Grayscale veri setlerinden seçilen beş çoklu odak görüntü çifti üzerinde deneyler gerçekleştirilmiştir. Görüntüler 256×256 çözünürlüğe sahiptir. Veri setleri farklı görüntü özelliklerine sahip sahneler içermekte; bazı görüntüler blok tabanlı füzyon için daha uygun özellikler gösterirken, bazıları keskin ve bulanık bölgeler arasındaki doğrusal olmayan geçişler nedeniyle daha zorlayıcı örnekler sunmaktadır.

Banharnsakun (2019) çalışmasında ise hem yapay olarak oluşturulmuş hem de gerçek merceklerle elde edilmiş doğal görüntüler kullanılmıştır. Yapay görüntüler, her yerde net bir referans görüntüye Gauss bulanıklaştırması uygulanması yoluyla oluşturulduğundan, füzyon kalitesi söz konusu referans görüntü ile doğrudan karşılaştırılabilmektedir.

6.2. Algoritmaların Genel Performans Karşılaştırması

Akbulut (2025), 16 farklı metasezgisel algoritmayı beş çoklu odak görüntü çifti üzerinde değerlendirmiş ve her algoritmayı 30 bağımsız kez çalıştırarak ortalama ve standart sapma değerlerini hesaplamıştır. Bunun yanında algoritmaların CPU çalışma süreleri karşılaştırılmış ve elde edilen füzyon görüntüleri öznel olarak değerlendirilmiştir. Genel bulgular, ABC algoritmasının toplam performans bakımından öne çıktığını göstermektedir.

Yapay Arı Kolonisi (ABC) algoritması, çalışmanın genel değerlendirmesinde hem nesnel hem de öznel sonuçlar bakımından en güçlü algoritma olarak öne çıkmıştır. Bununla birlikte, algoritmaların performansının görüntü çiftine ve kullanılan kalite metriğine bağlı olarak değişebildiği görülmektedir. Nitekim referans çalışmanın tablolarında bazı görüntü çiftlerinde PSO, GOA, MVO, SCA, SSA, TSO veya DA gibi farklı algoritmaların da üst sıralarda yer aldığı görülmektedir. Dolayısıyla ABC'nin üstünlüğü, her görüntü çiftinde mutlak olarak birinci olmasından ziyade, genel değerlendirmede dengeli ve yüksek performans göstermesi bağlamında ele alınmalıdır (Akbulut, 2025).

Hesaplama süresi bakımından Cuckoo Search (CS) algoritması ortalama 0,39 saniye ile en hızlı algoritma olurken, Selfish Herd Optimizer (SHO) ortalama 92,04 saniye ile en yavaş algoritma olmuştur. ABC ise füzyon kalitesi açısından en başarılı sonuçları vermesine rağmen CPU süresi bakımından 15. sırada yer almıştır. Bu bulgular, füzyon kalitesi ile hesaplama maliyeti arasında belirgin bir ödünleşim bulunduğuna işaret etmektedir (Akbulut, 2025).

Algoritmaların 30 bağımsız çalıştırmasından elde edilen standart sapma değerlerinin genel olarak düşük olması, sonuçların kararlılığı ve yakınsama davranışının tutarlılığı açısından olumlu bir bulgu olarak değerlendirilmektedir. Bununla birlikte, bu durum algoritmaların farklı başlangıç noktalarından tamamen bağımsız olduğunu göstermemekte; deneysel koşullar altında sonuçların görece istikrarlı ve tekrarlanabilir olduğunu göstermektedir (Akbulut, 2025).

6.3. Sürü Tabanlı ve Fizik Tabanlı Algoritmaların Karşılaştırılması

Çakıroğlu ve ark. (2023), blok tabanlı çoklu odak görüntü füzyonu kapsamında sürü tabanlı ve fizik tabanlı altı algoritmayı Lytro veri setindeki görüntü çiftleri üzerinde karşılaştırmıştır. Elde edilen sonuçlar, sürü tabanlı ABC ve JSA algoritmalarının, fizik tabanlı AOA, AOS ve EO algoritmalarına kıyasla genel olarak daha yüksek füzyon performansı sergilediğini göstermiştir. Bununla birlikte AOS algoritması, hesaplama süresi bakımından en verimli yöntem olarak belirlenmiştir.

Bu bulgular, sürü tabanlı algoritmaların görüntü füzyonu gibi çok modlu ve doğrusal olmayan optimizasyon problemlerinde fizik tabanlı algoritmalara kıyasla daha uygun bir adaptasyon mekanizması sergileyebileceğine işaret etmektedir. Bununla birlikte, araştırmacılar hesaplama bütçesinin sınırlı olduğu uygulamalarda AOS algoritmasının cazip bir alternatif oluşturduğunu da vurgulamıştır (Çakıroğlu ve ark., 2023).

6.4. Geleneksel Yöntemlerle Karşılaştırma

Çakıroğlu ve ark. (2023), metasezgisel algoritmalar ile geleneksel dönüşüm tabanlı yöntemleri (DWT, DCT, SWT, PCA ve bunların kombinasyonları) kapsamlı biçimde karşılaştırmıştır. Elde edilen sonuçlara göre DCT, QNMI, QPAB/F, Qy ve QCB metriklerinde en başarılı geleneksel yöntem olurken, DWT QSCD ve QSTD metriklerinde öne çıkmıştır. Metasezgisel yöntemler ise QNMI, QSCD, Qy ve QSTD metriklerinde DCT'den daha yüksek sonuçlar elde etmiştir. Bu karşılaştırma, optimizasyon tabanlı yaklaşımların geleneksel yöntemlere kıyasla ilave bir iyileştirme potansiyeli taşıdığını göstermektedir.

Akbulut (2025), blok tabanlı optimizasyon yöntemi ile Ouyang ve ark. tarafından önerilen yeni bir yöntem arasında karşılaştırma gerçekleştirmiştir. Sonuçlar, blok tabanlı optimizasyon yönteminin çeşitli kalite metriklerinde Ouyang yöntemiyle

karşılaştırılabilir sonuçlar üretebildiğini göstermektedir. Bu bulgu, blok tabanlı yaklaşımın daha yeni yöntemlerin ortaya çıkmasına rağmen rekabetçi ve uygulamaya değer bir alternatif olma niteliğini sürdürdüğüne işaret etmektedir.

7. Tartışma ve Değerlendirme

Bu bölüm kapsamında incelenen çalışmalar, metasezgisel optimizasyon algoritmalarının blok tabanlı çoklu odak görüntü füzyonuna sağladığı katkıları farklı boyutlarıyla ortaya koymaktadır. Elde edilen bulgular çerçevesinde özellikle aşağıdaki hususlar dikkat çekmektedir:

İlk olarak, algoritma seçiminin füzyon kalitesi üzerinde anlamlı bir etkisinin bulunduğu görülmektedir. Akbulut (2025), tek bir kalite metriğinde yüksek performans gösterilmesinin genel başarıyı garanti etmediğini ortaya koymuştur. Metasezgisel algoritmaların stokastik yapısı nedeniyle performansın görüntü çiftine ve kullanılan değerlendirme ölçütüne göre değişebildiği görülmektedir. Bu nedenle, algoritmaların performansının çoklu kalite metrikleri birlikte dikkate alınarak değerlendirilmesi önem taşımaktadır.

İkinci olarak, hesaplama maliyeti ile füzyon kalitesi arasında belirgin bir ödünleşim bulunmaktadır. ABC kalite bakımından öne çıkarken CPU süresi açısından 15. sırada yer almakta; buna karşılık CS en düşük ortalama CPU süresine sahip algoritma olurken SHO en yüksek ortalama CPU süresini göstermektedir (Akbulut, 2025). Dolayısıyla uygulama gereksinimleri doğrultusunda yalnızca füzyon kalitesinin değil, hesaplama maliyetinin de değerlendirmeye dâhil edilmesi gerekmektedir.

Üçüncü olarak, görüntünün yapısal özelliklerinin hem blok tabanlı füzyonun hem de kullanılan algoritmaların performansını etkileyebildiği görülmektedir. Blok tabanlı yönteme uygun özellikler taşıyan görüntülerde algoritmaların sonuçları birbirine daha yakın seyrederken, yönteme daha az uygun olan veya keskin ve bulanık bölgeler arasında doğrusal olmayan geçişler içeren görüntülerde performans farklılıkları daha belirgin hâle gelebilmektedir. Bu nedenle algoritma performansının farklı görüntü karakteristikleri üzerinde değerlendirilmesi önem taşımaktadır (Akbulut, 2025).

Dördüncü olarak, nesnel kalite metrikleri ile öznel görsel değerlendirmenin birlikte kullanılması önem taşımaktadır. Akbulut (2025), hiçbir tek kalite metriğinin görüntü kalitesinin tüm boyutlarını yeterli biçimde temsil edemeyeceğini ve bu nedenle dokuz nesnel metrikle birlikte görsel değerlendirmelerin de gerçekleştirilmesi gerektiğini göstermiştir. Özellikle büyütülmüş görüntü bölgelerinin incelenmesi, metriklerden elde edilen sonuçların görsel açıdan yorumlanmasına katkı sağlamaktadır.

Gelecekteki araştırmalar açısından çeşitli yönelimler öngörülebilir. Çok amaçlı metasezgisel optimizasyon yaklaşımlarının benimsenmesi, birden fazla kalite metriğinin eşzamanlı olarak optimize edilmesine ve daha dengeli sonuçların elde edilmesine katkı sağlayabilir (Çakıroğlu ve ark., 2023). Bunun yanı sıra, derin öğrenme tabanlı yöntemler ile blok tabanlı optimizasyon yaklaşımlarının hibrit yapılarda bir araya getirilmesi umut vadeden bir araştırma alanı olarak değerlendirilebilir. Farklı görüntü içeriklerine uyarlanabilen ve blok boyutunu dinamik olarak güncelleyebilen yaklaşımların geliştirilmesi de gelecekteki çalışmalar açısından önemli bir araştırma yönü oluşturmaktadır.

8. Sonuç

Bu bölümde, metasezgisel optimizasyon algoritmalarının çoklu odak görüntü füzyonunda optimal blok boyutunun belirlenmesi amacıyla kullanımına odaklanılmıştır. Alan derinliği sınırlamasından kaynaklanan odak probleminin çözümünde blok tabanlı füzyon yaklaşımından yararlanılabilmekte; bununla birlikte elde edilen füzyon performansı, blok boyutunun uygun biçimde belirlenmesine önemli ölçüde bağlı bulunmaktadır.

Aslantas ve Kurban (2010) tarafından blok boyutunun optimizasyonunda Diferansiyel Evrim yaklaşımının kullanılmasının ardından, farklı metasezgisel algoritmaların bu problem üzerindeki performansları çeşitli çalışmalar kapsamında araştırılmıştır. Akbulut (2025) tarafından gerçekleştirilen çalışma ise 16 algoritmayı aynı problem çerçevesinde karşılaştırarak ABC'nin genel olarak hem nesnel hem de öznel değerlendirmelerde üstün performans sergilediğini ortaya koymaktadır.

Hesaplama maliyeti, özellikle çalışma süreleri yüksek olan algoritmalar bakımından gerçek uygulamalarda dikkate alınması gereken önemli bir unsurdur. Akbulut (2025) çalışmasında CS en hızlı, SHO ise en yavaş algoritma olarak belirlenirken, ABC kalite bakımından üstün olmasına rağmen hesaplama süresi açısından daha maliyetli bir yöntem olarak ortaya çıkmıştır. Ayrıca blok tabanlı optimizasyon yönteminin Ouyang yöntemiyle çeşitli metriklerde karşılaştırılabilir sonuçlar üretmesi, söz konusu yaklaşımın güncel yöntemler karşısındaki rekabetçiliğini desteklemektedir.

Sonuç olarak, metasezgisel optimizasyon algoritmaları kullanılarak gerçekleştirilen blok tabanlı çoklu odak görüntü füzyonu, optimal blok boyutunun görüntüye özgü olarak belirlenmesine ve farklı kalite ölçütleri altında değerlendirilmesine olanak sağlayan bir çözüm yaklaşımı sunmaktadır. Akbulut (2025) tarafından gerçekleştirilen kapsamlı karşılaştırma, özellikle ABC algoritmasının genel performans bakımından güçlü bir seçenek olduğunu; bununla birlikte algoritma seçiminde görüntü karakteristikleri ile hesaplama maliyetinin de dikkate alınması gerektiğini göstermektedir.

Kaynakça

Akbulut, H. (2025). Meta-heuristic optimization for optimal block size in multi-focus image fusion: a comprehensive comparative study. *The Visual Computer, 41,* 11025–11051. https://doi.org/10.1007/s00371-025-04084-4

Aslantas, V. ve Bendes, E. (2015). A new image quality metric for image fusion: the sum of the correlations of differences. *AEU – International Journal of Electronics and Communications, 69,* 1890–1896. https://doi.org/10.1016/j.aeue.2015.09.004

Aslantas, V. ve Kurban, R. (2009). A comparison of criterion functions for fusion of multi-focus noisy images. *Optics Communications, 282,* 3231–3242. https://doi.org/10.1016/j.optcom.2009.05.021

Aslantas, V. ve Kurban, R. (2010). Fusion of multi-focus images using differential evolution algorithm. *Expert Systems with Applications, 37,* 8861–8870. https://doi.org/10.1016/j.eswa.2010.06.011

Banharnsakun, A. (2019). Multi-focus image fusion using best-so-far ABC strategies. *Neural Computing and Applications, 31,* 2025–2040. https://doi.org/10.1007/s00521-015-2061-2

Banharnsakun, A., Achalakul, T. ve Sirinaovakul, B. (2011). The best-so-far selection in artificial bee colony algorithm. *Applied Soft Computing, 11,* 2888–2901.

Chen, Y. ve Blum, R. S. (2009). A new automated quality assessment algorithm for image fusion. *Image and Vision Computing, 27,* 1421–1432. https://doi.org/10.1016/j.imavis.2007.12.002

Çakıroğlu, F., Kurban, R., Durmuş, A. ve Karaköse, E. (2023). Multi-focus image fusion by using swarm and physics based metaheuristic algorithms: a comparative study with Archimedes, Atomic Orbital Search, Equilibrium, Particle Swarm, Artificial Bee Colony and Jellyfish Search optimizers. *Multimedia Tools and Applications, 82,* 44859–44883. https://doi.org/10.1007/s11042-023-16651-9

Eskicioglu, A. M. ve Fisher, P. S. (1995). Image quality measures and their performance. *IEEE Transactions on Communications, 43,* 2959–2965. https://doi.org/10.1109/26.477498

Karaboga, D. ve Basturk, B. (2007). A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. *Journal of Global Optimization, 39,* 459–471. https://doi.org/10.1007/s10898-007-9149-x

Kennedy, J. ve Eberhart, R. (1995). Particle swarm optimization. *Proceedings of ICNN'95 – International Conference on Neural Networks, 4,* 1942–1948. https://doi.org/10.1109/ICNN.1995.488968

Łukasik, S. ve Zak, S. (2009). Firefly algorithm for continuous constrained optimization tasks. *Lecture Notes in Computer Science, 5796,* 169–178. https://doi.org/10.1007/978-3-642-04441-0_8

Mirjalili, S. (2015). Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm. *Knowledge-Based Systems, 89,* 228–249. https://doi.org/10.1016/j.knosys.2015.07.006

Mirjalili, S. (2016). Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems. *Neural Computing and Applications, 27,* 1053–1073. https://doi.org/10.1007/s00521-015-1920-1

Mirjalili, S., Mirjalili, S. M. ve Hatamlou, A. (2016). Multi-verse optimizer: a nature-inspired algorithm for global optimization. *Neural Computing and Applications, 27,* 495–513. https://doi.org/10.1007/s00521-015-1870-7

Rao, R. V., Savsani, V. J. ve Vakharia, D. P. (2011). Teaching–learning-based optimization: a novel method for constrained mechanical design optimization problems. *Computer-Aided Design, 43,* 303–315. https://doi.org/10.1016/j.cad.2010.12.015

Simon, D. (2008). Biogeography-based optimization. *IEEE Transactions on Evolutionary Computation, 12,* 702–713. https://doi.org/10.1109/TEVC.2008.919004

Storn, R. ve Price, K. (1997). Differential evolution – A simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization, 11,* 341–359.

Xydeas, C. S. ve Petrovic, V. (2000). Objective image fusion performance measure. *Electronics Letters, 36,* 308–309. https://doi.org/10.1049/el:20000267

Yang, X. S. ve Deb, S. (2009). Cuckoo search via Lévy flights. *2009 World Congress on Nature and Biologically Inspired Computing (NaBIC),* 210–214. https://doi.org/10.1109/NABIC.2009.5393690

6. Bölüm

Parçalı Metinden İlişkisel Bilgiye: Büyük Dil Modellerinde Grafik Destekli Bilgi Erişimli Üretim (GraphRAG)

Hayri İNCEKARA[1]

Özet

Büyük dil modelleri (BDM'ler), bilgiyi milyarlarca parametre içinde örtük olarak depolar; bu temsil biçimi, gerçek dışı içerik üretimi (halüsinasyon), bilgi güncelliği ve alan uzmanlığı eksikliği gibi yapısal sorunların kaynağıdır. Bilgi erişimli üretim (Retrieval-Augmented Generation, RAG), üretimi çıkarım anında erişilen harici kanıtlara dayandırarak bu sorunları hafifletir; ancak klasik RAG'in bilgi temsili birbirinden kopuk metin parçalarına dayanır. Bu parçalı temsil, cevabı birden fazla bilgi parçasının zincirlenmesini gerektiren çok adımlı sorularda ve derlemin bütününe yönelen küresel sorularda ilkesel bir darboğaz oluşturur. Grafik destekli bilgi erişimli üretim (GraphRAG), bilgiyi varlıklar ve ilişkiler üzerinden açıkça yapılandırarak bu darboğaza cevap veren güncel araştırma alanıdır. Bu bölüm, GraphRAG'i iki tamamlayıcı mercekle inceler: Grafiklerin RAG hattındaki işlevsel rolleri (bilgi tabanı inşası, erişim algoritmaları ve hat denetimi) ile grafik tabanlı indeksleme, grafik güdümlü erişim ve grafik destekli üretim aşamalarından oluşan iş akışı. Klasik RAG ile GraphRAG'in sistematik karşılaştırması sunulmakta; Microsoft GraphRAG, LightRAG, HippoRAG, Think-on-Graph, RoG, GNN-RAG, G-Retriever, HybridRAG ve PathRAG sistemleri grafik kaynağı, erişim mekanizması ve hedef sorgu profili eksenlerinde karşılaştırmalı bir tabloyla tanıtılmaktadır. Bölüm, değerlendirme pratiği, uygulama alanları ve ajan tabanlı GraphRAG ile çok dillilik dâhil açık araştırma sorunlarının tartışılmasıyla sonlanmaktadır.

Anahtar Kelimeler: Büyük dil modelleri, bilgi erişimli üretim, GraphRAG, bilgi grafikleri, halüsinasyon, çok adımlı akıl yürütme, ajan tabanlı RAG

[1] Department of Artificial Intelligence and Machine Learning, Konya Technical University, Konya, Türkiye. hincekara@ktun.edu.tr, ORCID: 0000-0002-0898-3422

1. Giriş

Bir dil modeline derleminizdeki iki belge arasında hiç açıkça yazılmamış bir bağlantıyı sorduğunuzda ne olur? Klasik bilgi erişimli üretim sistemleri bu soruya çoğu zaman iki yoldan biriyle cevap verir: ya bağlantının halkalarından yalnızca birini bulup eksik bir yanıt üretir ya da eksik kalan halkayı modelin parametrik bilgisinden, yani doğrulanamaz bir kaynaktan tamamlar. İkinci durum, tam da RAG'in ortadan kaldırmayı vadettiği halüsinasyon sorununun geri dönüşüdür. Sorunun kökeninde, harici bilginin birbirinden bağımsız metin parçaları (chunk) hâlinde temsil edilmesi yatar: parçalar arasındaki ilişkiler indeksleme sırasında kaybolur ve erişim, bu ilişkileri yeniden kuramaz (Peng vd., 2025; Zhang vd., 2025).

Grafik destekli bilgi erişimli üretim (GraphRAG), bu kayıp ilişkilerin peşine düşen araştırma alanıdır. Bilgi, düğümlerin varlıkları ve kenarların ilişkileri temsil ettiği grafik yapılarında örgütlenir; erişim, grafik üzerinde gezinme, alt-graf çıkarımı ya da topluluk özetleri aracılığıyla ilişkisel bağlamı koruyacak biçimde yapılır. Alan, kısa sürede dikkat çekici bir kurumsallaşma yaşamıştır: Peng vd. (2025) tarafından hazırlanan ve yazarlarının alanın ilk kapsamlı derlemesi olarak nitelediği çalışma (ön baskı 2024) ACM Transactions on Information Systems dergisinde, Zhu vd. (2026) tarafından hazırlanan ve grafiklerin RAG içindeki işlevlerini veri yönetimi perspektifinden ele alan derleme ise ACM Computing Surveys dergisinde yayımlanmıştır. Erken bir tarama Procko ve Ochoa (2024) tarafından sunulmuş; Han vd. (2025) ve Zhang vd. (2025) alanı farklı sınıflandırmalarla genişletmiştir. Bu yoğunluk, GraphRAG'in geçici bir mühendislik hilesi değil, bilgi temsili ile dil üretimi arasındaki köklü bir sorunun cevabı olduğuna işaret etmektedir.

Bu bölümün Türkçe alanyazına iki katkısı hedeflenmektedir. Birincisi, GraphRAG'i tek bir derlemenin izinden değil, birbirini tamamlayan iki mercekle sunmaktır: grafiklerin RAG hattında üstlendiği işlevsel roller (Zhu vd., 2026) ve uçtan uca iş akışı (Peng vd., 2025). İkincisi, alandaki temsili sistemleri ortak eksenler (grafik kaynağı, erişim mekanizması, hedef sorgu profili) üzerinde karşılaştırmalı tablolarla konumlandırmak ve Türkçe terminoloji önerileriyle birlikte sunmaktır. Bölümün dayandığı hakemli derlemeler şunlardır: GraphRAG için Peng vd. (2025) ve Zhu vd. (2026); genel RAG için Huang ve Huang (2026) ile Zhao vd. (2026); BDM–bilgi grafiği birleşimi için Pan vd. (2024) ve Wang vd. (2025); halüsinasyon için Ji vd. (2023) ve Lavrinovics vd. (2025); bilgi grafikleri için Hogan vd. (2021) ve Ji vd. (2022). Bu derlemelerin yanında Lewis vd. (2020) ile başlayan birincil kaynak zinciri de anlatıya dâhil edilmiştir.

Bölümün geri kalanı şöyle örgütlenmiştir: İkinci kısım kavramsal arka planı (parametrik bilgi, halüsinasyon, bilgi grafikleri) kurar. Üçüncü kısım RAG'in

evrimini naif hatlardan ajan tabanlı sistemlere uzanan bir çizgide izler. Dördüncü kısım GraphRAG'ı tanımlar ve klasik RAG ile sistematik biçimde karşılaştırır. Beşinci kısım grafiklerin işlevsel rollerini, altıncı kısım üç aşamalı iş akışını inceler. Yedinci kısım temsili sistemleri karşılaştırmalı olarak sunar. Sekizinci, dokuzuncu ve onuncu kısımlar sırasıyla değerlendirme pratiğini, uygulama alanlarını ve açık sorunları ele alır.

2. Kavramsal Arka Plan

2.1. Parametrik Bilginin Doğası ve Sınırları

BDM'ler, eğitim derleminden öğrendiklerini ağırlık matrisleri içinde dağıtık ve örtük biçimde saklar. Bu temsilin üç pratik sonucu vardır. Birincisi, bilgi eğitim kesim tarihinde donmuştur, modelin bu tarihten sonraki gelişmeleri bilmesi mümkün değildir ve bilginin güncellenmesi maliyetli bir yeniden eğitim gerektirir (Huang ve Huang, 2026). İkincisi, bilgi denetlenemez: modelin belirli bir olguyu bilip bilmediği, ancak ona sorularak ve çıktısı doğrulanarak anlaşılabilir. Üçüncüsü, kurumsal senaryolardaki özel belgeler eğitim verisinde hiç yer almamış olabilir; model bu durumda, sahip olmadığı bilgiyi üretmesi beklenen bir konuma itilir (Fan vd., 2024). Pan vd. (2024), bu tabloyu bilgi grafikleriyle karşılaştırmalı biçimde özetler: BDM'ler genel bilgi, dil anlama ve genelleme bakımından güçlü; örtük temsil, halüsinasyon eğilimi, kara kutu yapı ve güncel bilgi eksikliği bakımından zayıftır. Bilgi grafikleri ise yapısal, doğru ve yorumlanabilir bilgi sunar; buna karşılık eksik kapsam ve doğal dili anlayamama gibi kısıtlar taşır. İki teknolojinin tamamlayıcılığı, GraphRAG dâhil tüm birleştirme araştırmalarının hareket noktasıdır.

2.2. Halüsinasyon: Taksonomi ve Grafiklerle Azaltım

Ji vd. (2023), doğal dil üretiminde halüsinasyonu kaynak içerikle tutarsız ya da kaynaktan doğrulanamayan içerik üretimi olarak tanımlar ve iki türe ayırır. İçsel halüsinasyonda üretilen içerik kaynakla doğrudan çelişir, dışsal halüsinasyonda ise içerik kaynaktan doğrulanamaz. Aynı derleme, halüsinasyonun veri kaynaklı (gürültülü ya da çelişik eğitim verisi) ve model kaynaklı (kodlama-çözme kusurları, parametrik bilginin bağlama baskın çıkması) nedenlerini ayrıştırır. Lavrinovics vd. (2025) ise soruna bilgi grafikleri penceresinden bakar: grafiklerin sunduğu yapılandırılmış ve birbirine bağlı olgu koleksiyonu, modelin belirli konulardaki bilgi boşluklarını dolduran bir bağlam kaynağı olarak halüsinasyon azaltımında umut verici bir yol açar. Ancak yazarlar, halüsinasyon tespitinin, değerlendirmesinin ve çok dilli ortamlara genellenmesinin hâlâ çözülmemiş sorunlar barındırdığını vurgular. GraphRAG

bu iki hattın kesişimindedir: erişilen ilişkisel kanıt, hem üretimi olgulara bağlar hem de yanıtın hangi grafik yolundan türetildiğinin izlenebilmesini sağlar.

2.3. Bilgi Grafikleri

Hogan vd. (2021), bilgi grafiğini gerçek dünyaya ilişkin bilgiyi biriktirmek ve aktarmak amacıyla düzenlenmiş bir veri grafiği olarak tanımlar: düğümler ilgi alanındaki varlıkları, kenarlar varlıklar arasındaki ilişkileri temsil eder. Derleme, yönlü kenar etiketli grafikler ve özellik grafikleri gibi veri modellerini, sorgulama ve doğrulama dillerini, bilginin tümdengelimli (ontolojiler, kurallar) ve tümevarımlı (grafik gömmeleri, grafik sinir ağları) tekniklerle işlenmesini kapsamlı biçimde ele alır. Ji vd. (2022) ise alanı dört eksende derler: bilgi temsili öğrenimi, bilgi edinimi, zamansal bilgi grafikleri ve bilgi kaynaklı uygulamalar. GraphRAG açısından kritik nokta şudur: kullanılan grafik her zaman hazır bir kaynak (Wikidata, Freebase, DBpedia gibi) olmak zorunda değildir; birçok sistem grafiği doğrudan hedef derlemden, çoğu zaman bizzat BDM'ler aracılığıyla inşa eder (Edge vd., 2024; Zhu vd., 2026). Dolayısıyla bilgi edinimi alanının kalite ve eksiklik sorunları, GraphRAG'in indeksleme aşamasına olduğu gibi taşınır. Wang vd. (2025), madalyonun öbür yüzünü, yani BDM'lerin grafik öğrenmesini nasıl dönüştürdüğünü derlemekte ve iki alanın karşılıklı beslenme ilişkisini belgelemektedir.

3. Bilgi Erişimli Üretimin Evrimi

3.1. Özgün Mimari: Parametrik ve Parametrik Olmayan Bellek

RAG teriminin kaynağı, Lewis vd. (2020) tarafından bilgi yoğun doğal dil işleme görevleri için önerilen mimaridir. Bu mimaride üretici modelin parametrik belleği, Wikipedia'nın yoğun vektör indeksinden erişilen parametrik olmayan bellekle birleştirilir; erişim, soru ve geçitleri iki ayrı kodlayıcıyla ortak vektör uzayına gömen yoğun geçit erişimiyle (DPR) yapılır. Karpukhin vd. (2020), açık alan soru yanıtlamada yoğun temsillerin geleneksel seyrek yöntemlere kıyasla daha isabetli erişim sağladığını göstermiş; Lewis vd. (2020) ise erişimle koşullandırılan üretimin salt parametrik modellere göre daha özgül, çeşitli ve olgusal dil ürettiğini raporlamıştır. Bu iki çalışma, izleyen tüm RAG araştırmasının referans noktasıdır.

3.2. Naif Hattan Modüler Sistemlere

Gao vd. (2023), RAG araştırmalarının gelişimini üç paradigmada toplar. Naif RAG, belgelerin parçalanıp gömüldüğü ve sorguya en benzer parçaların istem içine eklendiği temel hattır. Gelişmiş RAG, erişim öncesi (sorgu yeniden yazımı ve genişletme) ve erişim sonrası (yeniden sıralama, bağlam sıkıştırma)

iyileştirmelerle isabeti artırır. Modüler RAG, erişim, bellek ve yönlendirme işlevlerini değiştirilebilir modüller olarak ele alır. Huang ve Huang (2026) aynı süreci bilgi erişimi disiplini gözünden dört evrede inceler: erişim öncesi hazırlık, erişim, erişim sonrası işleme ve üretim; bu ayrıştırma, performans darboğazlarının hangi evreden kaynaklandığını teşhis etmeye yarar. Zhao vd. (2026) ise çerçeveyi metnin ötesine taşır: erişimin üretime hangi noktada ve nasıl entegre edildiğine göre bir sınıflandırma önerir ve RAG'in kod, görüntü ve ses dâhil yapay zekâ ile içerik üretiminin genelinde uygulanabilir bir teknik olduğunu gösterir.

3.3. Ajan Tabanlı RAG'e Doğru

Sabit iş akışlı RAG hatları, sorgunun karmaşıklığına uyum sağlayamaz: basit bir olgu sorusu ile çok kaynaklı bir araştırma sorusu aynı boru hattından geçer. Ajan tabanlı RAG (agentic RAG), bu katılığı özerk ajan tasarım örüntüleriyle aşmayı önerir: yansıtma, planlama, araç kullanımı ve çoklu ajan iş birliği örüntüleri, erişim stratejisinin sorguya göre dinamik biçimde yönetilmesine ve bağlam anlayışının yinelemeli olarak rafine edilmesine olanak tanır (Singh vd., 2025). Bu eğilim GraphRAG ile doğal biçimde kesişir: grafik üzerinde hangi düğümden hangi kenara ilerleneceği kararı, özünde bir ajan kararıdır ve yedinci kısımda görüleceği üzere Think-on-Graph gibi sistemler bu kesişimin erken örnekleridir (Sun vd., 2024).

3.4. Parçalı Temsilin Yapısal Sınırları

Klasik RAG'in dayandığı düz metin erişimi dört yapısal sınırlılık taşır. Birincisi, parçalama işlemi parçalar arasındaki anlamsal ve ilişkisel bağları koparır; cevabı iki parçadaki bilgilerin birleşimi olan bir soruda, benzerlik temelli erişim her iki parçayı birden getirmeyi garanti edemez. İkincisi, vektör benzerliği yüzeysel anlamsal yakınlığı ölçer; uzmanlık sorgularının ardındaki mantıksal yapıyı çözümleyemez. Üçüncüsü, derlem büyüdükçe indeksleme ve erişim verimliliği düşer (Zhang vd., 2025). Dördüncüsü ve en az fark edileni, Edge vd. (2024) tarafından adlandırılan küresel duyu-verme sorunudur: "Bu derlemin ana temaları nelerdir?" türünden sorular hiçbir yerel erişim hedefi tanımlamaz; getirilebilecek hiçbir parça kümesi cevabı içermez, çünkü cevap derlemin bütününün bir özetidir. GraphRAG, dördü de bilginin ilişkisiz parçalar hâlinde temsilinden doğan bu sınırlılıklara, temsilin kendisini değiştirerek cevap verir.

4. GraphRAG: Tanım ve Klasik RAG ile Karşılaştırmalı Çerçeve

Peng vd. (2025), GraphRAG'i varlıklar arasındaki yapısal bilgiden yararlanarak daha isabetli ve kapsamlı erişim gerçekleştiren, ilişkisel bilgiyi

yakalayan ve böylece daha doğru, bağlama duyarlı yanıtlar üreten bir çerçeve olarak tanımlar. Zhang vd. (2025) bu üstünlüğü üç yenilikle açıklar: varlık ilişkilerini ve alan hiyerarşilerini açıkça yakalayan grafik yapılı bilgi temsili, çok adımlı akıl yürütmeye olanak veren ve bağlamı koruyan grafik erişim teknikleri ve erişilen bilgiyi tutarlı üretime dönüştüren yapı-farkındalıklı bütünleştirme. Peng vd. (2025) ayrıca GraphRAG ile bilgi tabanlı soru yanıtlama (KBQA) arasındaki sürekliliğe dikkat çeker: bilgi erişimine dayalı KBQA yöntemleri, uygulama odaklı bir GraphRAG alt kümesi olarak okunabilir. Tablo 1, iki paradigmanın temel eksenlerdeki farklarını özetlemektedir.

Tablo 1. Klasik (vektör tabanlı) RAG ile GraphRAG'in karşılaştırması

Eksen	**Klasik RAG**	**GraphRAG**
Bilgi temsili	Bağımsız metin parçaları ve vektör gömmeleri	Varlık ve ilişkilerden oluşan grafik; parçalar grafiğe bağlanabilir
Erişim birimi	Metin parçası (chunk)	Düğüm, üçlü, yol, alt-graf veya topluluk özeti
Erişim mekanizması	Vektör benzerliği (ve seyrek eşleştirme)	Grafik gezinme, alt-graf çıkarımı, GNN puanlama, benzerlikle melez
Çok adımlı sorular	Parçaların rastlantısal birlikteliğine bağımlı	İlişki zinciri grafikte açık bir yol olarak izlenir
Küresel sorular	İlkesel olarak yetersiz (yerel erişim hedefi yok)	Topluluk özetleriyle derlem geneli yanıt üretilebilir
İndeksleme maliyeti	Görece düşük (parçalama ve gömme)	Yüksek (grafik çıkarımı, çoğu zaman BDM çağrılarıyla)
Güncellenebilirlik	Yeni parçaların eklenmesi kolay	Grafik tutarlılığı gerektirir; artımlı güncelleme ayrı bir sorundur
İzlenebilirlik	Kaynak parça gösterilebilir	Yanıtın türetildiği ilişki yolu açıkça gösterilebilir

Kaynak: Peng vd. (2025), Edge vd. (2024), Guo vd. (2024) ve Zhang vd. (2025) temel alınarak yazarlar tarafından derlenmiştir.

5. Grafiklerin RAG Hattındaki İşlevsel Rolleri

GraphRAG üzerine yazılan ilk derlemeler alanı ağırlıklı olarak iş akışı üzerinden anlatmıştır. Zhu vd. (2026) ise farklı ve tamamlayıcı bir soru sorar: grafik, RAG hattının neresinde, hangi işlevi görür? Yazarlara göre mevcut derlemeler grafiğin rolünü büyük ölçüde bilgi grafiği gezinmesine indirger; oysa grafik yapısının erişim, istemleme ve hat denetimi üzerindeki daha geniş etkileri

yeterince incelenmemiştir. Bu işlevsel bakış, grafiği tek bir aşamanın aracı olmaktan çıkarıp hattın tamamına yayılan bir tasarım ilkesi olarak görür ve alanı grafik öğrenmesi, veritabanı sistemleri ve doğal dil işleme topluluklarının kesişimine yerleştirir (Zhu vd., 2026).

5.1. Bilgi Tabanı İnşasında Grafikler

İlk işlev, harici bilgi kaynağının kendisinin grafik olarak inşasıdır. DBpedia, Wikidata ve Freebase gibi açık bilgi grafikleri, varlık ve ilişkilerin yapılandırılmış örgütlenmesinin başlıca örnekleridir; ancak BDM tabanlı yaklaşımlar bu ilişkisel örüntülerden, özellikle karmaşık ve çok adımlı akıl yürütme görevlerinde, çoğu zaman tam olarak yararlanamamaktadır (Zhu vd., 2026). Grafik veri yönetimi yöntemleri burada devreye girer: varlık çözümleme, şema tasarımı ve indeks yapıları, ham metinden ya da hazır kaynaklardan sorgulanabilir ve tutarlı bir bilgi tabanı üretilmesini sağlar. Metinden grafik inşasında BDM'lerin kendisinin çıkarıcı olarak kullanılması (Edge vd., 2024), bu işlevin güncel ve maliyet belirleyici biçimidir.

5.2. Erişim Algoritmalarında Grafikler

İkinci işlev, erişimin kendisinin grafik algoritmalarıyla yapılmasıdır. Komşuluk genişletme, en kısa yol, rastgele yürüyüş ve kişiselleştirilmiş PageRank gibi klasik algoritmalar, sorgu varlıklarından başlayarak ilgili bilgiye topolojik yakınlık üzerinden ulaşır; grafik sinir ağları ise yapıyı öğrenilmiş temsillere dönüştürerek aday ögeleri puanlar (Peng vd., 2025; Zhu vd., 2026). Vektör benzerliğinden temel farkı şudur: benzerlik iki metnin ne kadar aynı şeyi söylediğini ölçerken, topolojik yakınlık iki varlığın bilgi yapısında birbirine ne kadar bağlı olduğunu ölçer. Çok adımlı sorularda aranan tam da ikincisidir.

5.3. Hat Denetimi ve İstemlemede Grafikler

Üçüncü ve en az görünür işlev, grafiğin RAG hattının denetim katmanında kullanılmasıdır. Erişilen alt-grafın hangi biçimde doğrusallaştırılıp isteme yerleştirileceği, akıl yürütme sürecinin hangi grafik yolunu izleyeceği ve çok turlu erişimde bir sonraki adımın nereden başlayacağı gibi kararlar, grafiği bir veri deposu değil bir denetim yapısı olarak kullanır (Zhu vd., 2026). Düşünce grafiği türü istemleme yaklaşımları ile ajan tabanlı gezinme stratejileri bu işlevin örnekleridir; yedinci kısımda ele alınan ToG ve RoG sistemleri, denetim işlevinin somutlaşmış hâlleridir.

6. GraphRAG İş Akışı: Üç Aşamalı Çerçeve

Peng vd. (2025), GraphRAG iş akışını üç aşamada biçimselleştirir: grafik tabanlı indeksleme (G-Indexing), grafik güdümlü erişim (G-Retrieval) ve grafik destekli üretim (G-Generation). Beşinci kısımdaki işlevsel roller bu aşamalara şöyle eşlenir: bilgi tabanı inşası indekslemenin, erişim algoritmaları erişimin, hat denetimi ise erişim ile üretim arasındaki geçişin sorusudur.

6.1. Grafik Tabanlı İndeksleme

İndeksleme, sorgulara temel oluşturacak grafik verisinin seçilmesini ya da inşasını ve erişime uygun biçimde indekslenmesini kapsar. Grafik verisi iki kaynaktan gelir: Wikidata ve Freebase gibi hazır açık bilgi grafikleri ile hedef derlemden inşa edilen grafikler; ikinci senaryoda varlık ve ilişkiler çoğu zaman bizzat BDM'lerce çıkarılır (Edge vd., 2024; Peng vd., 2025). İndeks türleri birbirini tamamlar: grafik indeksleri yapısal gezinmeyi, metin indeksleri grafik ögelerine iliştirilmiş açıklamalar üzerinden anahtar kelime erişimini, vektör indeksleri anlamsal benzerlik erişimini destekler; pratikte melez kullanım yaygındır (Peng vd., 2025). İndeks tasarımı, erişimin hangi ayrıntı düzeylerinde çalışabileceğini belirlediğinden sistemin karakterini büyük ölçüde bu aşama tayin eder.

6.2. Grafik Güdümlü Erişim

Erişim aşaması, sorguya karşılık grafikten hangi ögelerin getirileceğine karar verir. Peng vd. (2025) erişimcileri üç sınıfta toplar. Parametrik olmayan erişimciler klasik grafik algoritmalarına ve kural tabanlı eşleştirmeye dayanır; hızlıdır ancak sorgu anlama yetenekleri sınırlıdır. Dil modeli tabanlı erişimciler sorgu anlamlandırmayı BDM'ye devreder; model gezinme kararlarını doğal dil üzerinden verebilir. GNN tabanlı erişimciler grafiği kodlayıp aday ögeleri sorguyla benzerliklerine göre puanlar; örneğin GNN-RAG grafiği kodladıktan sonra her varlığa skor atar ve eşik üzerindeki varlıkları getirir (Mavromatis ve Karypis, 2024; Peng vd., 2025). Grafik ve metin gömmelerinin farklı anlamsal uzaylarda kalması durumunda karşıtsal öğrenmeyle hizalama yapılır (Peng vd., 2025). Erişim ayrıntı düzeyi ikinci tasarım eksenidir: düğüm, üçlü, yol, alt-graf ve topluluk özeti düzeyleri, gürültü ile bağlam genişliği arasındaki ödünleşimin farklı noktalarıdır; yinelemeli ve koşullu erişim paradigmaları bu ödünleşimi sorgu karmaşıklığına göre dinamik yönetmeye çalışır.

6.3. Grafik Destekli Üretim

Üretim aşamasının temel sorusu, grafik yapısındaki bilginin dizi tabanlı bir dil modeline nasıl sunulacağıdır. En yaygın yol, erişilen üçlü, yol ya da alt-grafların

doğal dile veya yapılandırılmış metin biçimlerine doğrusallaştırılarak isteme yerleştirilmesidir. Alternatif yol, grafik kodlayıcıların (örneğin GNN'lerin) grafiği sürekli temsillere dönüştürüp modelin girdi uzayına yansıtmasıdır; bu, yapısal bilgiyi daha sadık korur ancak ek eğitim gerektirir (Peng vd., 2025; He vd., 2024). Doğrusallaştırma biçiminin üretim kalitesine etkisi bulunduğundan biçim seçimi de başlı başına bir tasarım kararıdır.

7. Temsili GraphRAG Sistemleri

Bu kısım, alandaki tasarım uzayını dokuz temsili sistem üzerinden somutlaştırır. Tablo 2, sistemleri grafik kaynağı, erişim mekanizması ve hedef sorgu profili eksenlerinde karşılaştırmaktadır; izleyen alt bölümler her sistemi ayrıntılandırır.

Tablo 2. Temsili GraphRAG sistemlerinin karşılaştırması

Sistem	Grafik kaynağı	Erişim mekanizması	Öne çıkan özellik	Kaynak
Graph RAG (Microsoft)	Derlemden BDM ile inşa	Topluluk özetleri + harita-indirgeme	Derlem geneli küresel sorular	Edge vd. (2024)
LightRAG	Derlemden BDM ile inşa	Çift düzeyli (varlık / tema) erişim	Artımlı indeks güncelleme	Guo vd. (2024)
HippoRAG	Derlemden BDM ile inşa	Kişiselleştirilmiş PageRank	Tek adımda çok adımlı erişim	Gutiérrez vd. (2024)
Think-on-Graph	Hazır bilgi grafiği	BDM ajanıyla ışın araması	Eğitim gerektirmeyen gezinme	Sun vd. (2024)
RoG	Hazır bilgi grafiği	İlişki yolu planlama + erişim	Sadık ve yorumlanabilir çıkarım	Luo vd. (2024)
GNN-RAG	Hazır bilgi grafiği	GNN puanlama + BDM akıl yürütme	GNN–BDM iş bölümü	Mavromatis ve Karypis (2024)
G-Retriever	Metinsel grafikler	Ödül toplayan Steiner ağacı	Kompakt alt-graf çıkarımı	He vd. (2024)
HybridRAG	Derlemden inşa + vektör indeksi	Graf ve vektör erişiminin birleşimi	Finansal belge analizi	Sarmah vd. (2024)
PathRAG	Derlemden inşa	İlişkisel yollarla budama	Erişim sonrası gürültü azaltımı	Chen vd. (2025)

7.1. Microsoft GraphRAG: Küresel Sorular İçin Topluluk Özetleme

Edge vd. (2024) tarafından önerilen ve alanın adını popülerleştiren Graph RAG, derlem geneli küresel duyu-verme sorularını hedefler. Sistem, indeksleme aşamasında bir BDM ile kaynak belgelerden varlık merkezli bir bilgi grafiği çıkarır; grafik üzerinde topluluk tespiti uygulayarak yakın ilişkili varlık kümelerini hiyerarşik topluluklara ayırır ve her topluluk için önceden özet üretir. Sorgu zamanında topluluk özetleri harita-indirgeme tarzında kısmi yanıtlara dönüştürülür ve nihai küresel yanıtta birleştirilir. Yazarlar, yaklaşık bir milyon belirteçlik veri kümelerinde bu yaklaşımın yanıt kapsayıcılığı ve çeşitliliği bakımından vektör tabanlı RAG'e göre belirgin iyileşme sağladığını raporlar (Edge vd., 2024). Maliyet indekslemeye yüklenmiştir: grafiğin ve özetlerin BDM ile üretilmesi, derlem büyüdükçe önemli bir gider oluşturur.

7.2. LightRAG: Çift Düzeyli Erişim ve Artımlı Güncelleme

Guo vd. (2024) tarafından önerilen LightRAG, grafik yapılarını metin indeksleme sürecine gömerek maliyet sorununa cevap arar. Çift düzeyli erişim paradigmasında düşük düzey erişim belirli varlıklara ve yakın ilişkilerine, yüksek düzey erişim geniş temalara ve kavramlar arası bağlantılara yönelir; grafik yapıları vektör temsilleriyle bütünleştirildiğinden erişim hem ilişkisel bağlamı korur hem de hızlıdır. Sistemin ayırt edici özelliği artımlı güncelleme algoritmasıdır: yeni belgeler geldiğinde tüm indeksin yeniden inşası yerine grafiğin ilgili kısmı güncellenir; bu, LightRAG'i dinamik veri ortamlarına uygun kılar (Guo vd., 2024).

7.3. HippoRAG: Bellek Esinli Tek Adımlı Çok Adımlılık

Gutiérrez vd. (2024) tarafından önerilen HippoRAG, insan uzun süreli belleğine ilişkin hipokampal indeksleme kuramından esinlenir. Derlemden BDM ile bilgi grafiği çıkarılır; sorgu zamanında grafik üzerinde kişiselleştirilmiş PageRank çalıştırılarak sorgu varlıklarından yayılan ilişkisel etkinlik hesaplanır. Sonuç dikkat çekicidir: çok adımlı sorular, yinelemeli erişim turlarına gerek kalmadan tek erişim adımında yanıtlanabilir; adımlar arası bütünleştirme işi erişimin kendisine devredilmiştir (Gutiérrez vd., 2024). HippoRAG bu yönüyle, çok adımlılığı sorgu-zamanı akıl yürütmesinden indeks yapısına taşıyan yaklaşımların temsilcisidir.

7.4. Bilgi Grafiği Üzerinde Akıl Yürütme: ToG, RoG, GNN-RAG ve G-Retriever

Hazır bilgi grafikleriyle çalışan bir yaklaşım ailesi, BDM'yi grafikle etkileşen bir akıl yürütücü olarak konumlandırır. Think-on-Graph, BDM'yi grafik üzerinde adım adım gezinen bir ajan olarak kullanır: model, ışın aramasıyla umut vadeden akıl yürütme yollarını yinelemeli biçimde keşfeder ve genişletir; süreç ek eğitim gerektirmez ve keşfedilen yollar yanıtın izlenebilir gerekçesini oluşturur (Sun vd.,

2024). Reasoning on Graphs, planlama-erişim-akıl yürütme çerçevesi benimser: model önce bilgi grafiğine dayanan ilişki yolları biçiminde planlar üretir, planlarla grafikten geçerli yollar getirilir ve sadık, yorumlanabilir çıkarım bu yollar üzerinden yapılır (Luo vd., 2024). GNN-RAG, grafik sinir ağlarının alt-graf akıl yürütme gücünü BDM'lerin dil yeteneğiyle birleştirir: yoğun bir alt-graf üzerinde GNN ile aday yanıtlar ve ilgili yollar çıkarılır, yollar doğal dile dönüştürülüp BDM'ye sunulur (Mavromatis ve Karypis, 2024). G-Retriever, sahne grafiği anlama, sağduyu akıl yürütme ve bilgi grafiği akıl yürütme gibi gerçek dünya metinsel grafiği uygulamalarında soru yanıtlamayı hedefler; erişimi ödül toplayan Steiner ağacı eniyilemesi olarak formüle ederek sorguyla ilgili, bağlantılı ve kompakt bir alt-graf çıkarır ve bunu grafik kodlayıcı üzerinden modele bağlar; yazarlar yapının halüsinasyonu azalttığını raporlar (He vd., 2024). Tablo 2'ye alınmayan GRAG da metinsel grafiklerde alt-graf düzeyinde erişime odaklanan yakın bir yaklaşımdır (Hu vd., 2024).

7.5. Melez ve Yol Tabanlı Yaklaşımlar: HybridRAG ve PathRAG

Grafik erişimi ile vektör erişimi birbirinin alternatifi olmak zorunda değildir. Sarmah vd. (2024) tarafından finansal belge analizi için önerilen HybridRAG, bilgi grafiği tabanlı erişimle vektör tabanlı erişimi tek hatta birleştirir; finansal kazanç çağrısı dökümleri üzerindeki deneyler, birleşimin iki tekniğin tek başına kullanımına kıyasla hem erişim hem yanıt kalitesinde iyileşme sağladığını gösterir. PathRAG ise grafik erişiminin getirdiği fazlalık ve gürültüye yönelir: erişilen grafik bilgisi ilişkisel yollar temel alınarak budanır ve modele yalnızca sorguyla ilgili yol yapıları sunulur (Chen vd., 2025). İki sistem, alandaki iki pratik eğilimi örnekler: melezleşme ve erişim sonrası yapısal sadeleştirme.

8. Değerlendirme Görevleri, Veri Kümeleri ve Ölçütler

GraphRAG sistemleri en yoğun biçimde bilgi tabanlı soru yanıtlama ve çok adımlı soru yanıtlama görevlerinde değerlendirilir (Peng vd., 2025). Çok adımlı değerlendirmenin yaygın veri kümelerinden HotpotQA, yanıtın birden fazla belgeden gelen bilgilerin birleştirilmesini gerektirmesi ve destekleyici cümlelerin açıklanabilirlik amacıyla etiketlenmiş olması bakımından ayırt edicidir (Yang vd., 2018). Bilgi tabanlı soru yanıtlamada Freebase ve Wikidata türevi grafikler üzerinde tanımlanmış görevler kullanılır; ToG, RoG ve GNN-RAG'in başlıca değerlendirme zemini bunlardır (Sun vd., 2024; Luo vd., 2024; Mavromatis ve Karypis, 2024). Alana özgü değerlendirme de gelişmektedir: Xiong vd. (2024), tıp alanındaki RAG sistemlerinin sistematik kıyaslanması için MIRAGE kıyaslama ortamını önermiş ve erişimin tıbbi soru yanıtlama başarımına etkisini birden çok model ve derlem üzerinde incelemiştir.

Ölçüt tarafında, kesin eşleşme ve F1 gibi klasik ölçütlerin yanında, tek doğru yanıtı olmayan küresel sorular için BDM'nin hakem olarak kullanıldığı karşılaştırmalı değerlendirmeler öne çıkar; Edge vd. (2024) yanıtları kapsayıcılık ve çeşitlilik gibi ölçütler üzerinden bu yöntemle karşılaştırmıştır. Görev, veri kümesi ve ölçütlerdeki bu dağınıklık, sistemlerin adil karşılaştırılmasını güçleştirmektedir; ortak kıyaslama ortamlarının geliştirilmesi güncel derlemelerde açık bir ihtiyaç olarak dile getirilir (Peng vd., 2025; Zhang vd., 2025). Değerlendirmede sıkça gözden kaçan bir boyut da maliyettir: iki sistemin yalnızca doğruluk üzerinden karşılaştırılması, indeksleme ve sorgu başına BDM çağrısı sayısındaki büyük farkları görünmez kılar; adil bir kıyas, doğruluk-maliyet düzleminde yapılmalıdır.

9. Uygulama Alanları

GraphRAG'in katma değeri, bilginin doğal olarak ilişkisel örgütlendiği uzmanlık alanlarında belirginleşir. Zhang vd. (2025), uygulamaları sağlık, hukuk, finans ve bilimsel araştırma gibi profesyonel bağlamlar üzerinden inceler; bu bağlamların ortak özelliği derin alan uzmanlığı ve dağıtık kaynaklardan bilgi bütünleştirme gereksinimidir. Biyomedikal alanda hastalık, ilaç, gen ve semptom varlıkları arasındaki ilişkiler doğal bir grafik oluşturur; erişimin tıbbi soru yanıtlamaya katkısı kıyaslama düzeyinde de belgelenmiştir (Xiong vd., 2024; Pan vd., 2024). Finans alanında HybridRAG, kazanç çağrısı dökümlerinden bilgi çıkarımında graf-vektör birleşiminin etkisini göstermiştir (Sarmah vd., 2024). Kurumsal bilgi yönetiminde, büyük ve heterojen belge havuzları üzerinde bütüncül kavrayış gerektiren analizler, topluluk özetlemeli yaklaşımların birincil senaryosudur (Edge vd., 2024). RAG'in metin dışı kipliklere (kod, görüntü, ses) genişlemesi (Zhao vd., 2026) ile BDM–grafik etkileşiminin grafik öğrenmesi tarafındaki karşılığı (Wang vd., 2025) birlikte düşünüldüğünde, çok kipli varlık ve ilişkiler içeren grafiklerle çalışan GraphRAG sistemleri öngörülebilir bir sonraki adımdır.

10. Açık Sorunlar ve Gelecek Araştırma Yönelimleri

Alandaki derlemeler, GraphRAG'in önündeki sorunları büyük ölçüde örtüşen başlıklarda toplar. Birincisi bilgi kalitesi ve grafik inşa maliyetidir: BDM'lerle otomatik grafik çıkarımı ölçeklenebilirlik sağlar, ancak çıkarım hataları grafiğe taşınır ve hatalı bir kenar, üzerinden geçen tüm akıl yürütme yollarını zehirleyebilir. Grafiklerin eksikliği ve bakım maliyeti, Pan vd. (2024) tarafından bilgi grafiklerinin yapısal zayıflıkları arasında zaten sayılmaktadır ve GraphRAG'e miras kalır. İkincisi dinamikliktir: gerçek dünya bilgisi sürekli değişirken grafik indekslerinin bu değişimi düşük maliyetle izlemesi gerekir; LightRAG'in artımlı güncellemesi bir adımdır ancak genel çözüm yoktur (Guo vd., 2024; Peng vd., 2025). Üçüncüsü verimlilik ve ölçeklenebilirliktir: grafik gezinme ve alt-graf çıkarımı vektör aramaya kıyasla pahalı

olabilir; indekslemedeki BDM çağrıları derlem boyutuyla orantılı gider üretir (Edge vd., 2024; Zhang vd., 2025). Dördüncüsü değerlendirme standardizasyonudur (Peng vd., 2025).

Beşinci başlık, ajan tabanlı GraphRAG'dir. Yansıtma, planlama ve araç kullanımı örüntülerinin (Singh vd., 2025) grafik gezinmesiyle birleşmesi, erişim stratejisini sorguya göre planlayan ve gerektiğinde grafiği sorgu sırasında genişleten sistemlerin önünü açmaktadır; ToG bu birleşimin erken örneğidir (Sun vd., 2024). Altıncı başlık güvenlik ve gizliliktir: kurumsal grafiklerin hassas bilgi içerdiği senaryolarda erişim denetimi ve bilgi sızıntısı ayrıca incelenmelidir (Zhang vd., 2025). Yedinci başlık ise çok dilliliktir: Lavrinovics vd. (2025), halüsinasyon tespiti, değerlendirmesi ve bilgi bütünleştirmenin çok dilli ortamlara genellenmesini açık bir araştırma sorunu olarak tanımlar. Bu sorun Türkçe için doğrudan bir fırsata dönüşmektedir: Türkçe alan derlemlerinden bilgi grafiği inşası, Türkçenin eklemeli yapısına uygun varlık çözümleme yöntemleri ve bu grafikler üzerinde çalışan GraphRAG sistemleri, ulusal alanyazın için hem özgün hem uygulama değeri yüksek bir araştırma programı sunmaktadır.

11. Sonuç

Bu bölümde GraphRAG, iki tamamlayıcı mercekle incelenmiştir: grafiklerin RAG hattındaki işlevsel rolleri ile üç aşamalı iş akışı. Alanyazın, klasik vektör tabanlı RAG'in parçalı bilgi temsilinin çok adımlı akıl yürütme ve derlem geneli kavrayış gerektiren sorularda yapısal bir darboğaz oluşturduğu konusunda uzlaşmaktadır (Peng vd., 2025; Edge vd., 2024; Zhu vd., 2026). GraphRAG bu darboğaza temsilin kendisini değiştirerek cevap verir: bilgi, varlıklar ve ilişkiler üzerinden açıkça yapılandırılır; erişim, benzerliğin yanında topolojik yakınlığı da kullanır; üretim, izlenebilir ilişki yollarına dayanır.

Temsili sistemlerin karşılaştırması, tasarım uzayının tek bir en iyi noktası olmadığını göstermektedir. Topluluk özetlemeli küresel yanıtlama (Edge vd., 2024), çift düzeyli erişim ve artımlı güncelleme (Guo vd., 2024), bellek esinli tek adımlı çok adımlılık (Gutiérrez vd., 2024), grafik üzerinde ajan tabanlı gezinme (Sun vd., 2024; Luo vd., 2024), GNN–BDM iş bölümü (Mavromatis ve Karypis, 2024; He vd., 2024) ve melez erişim (Sarmah vd., 2024) farklı sorgu profillerine ve maliyet bütçelerine hitap eden, birbirini dışlamayan seçeneklerdir. Uygulayıcı için doğru soru "hangi sistem en iyisidir" değil, "benim sorgu profilim, derlem ölçeğim ve güncelleme sıklığım hangi tasarım noktasını gerektirir" sorusudur. Grafik inşa maliyeti, bilgi kalitesi, dinamik güncelleme, standart değerlendirme, ajan tabanlı gezinme ve çok dillilik, alanın olgunlaşması için önümüzde duran sorunlardır ve son ikisi, Türkçe doğal dil işleme topluluğuna somut bir araştırma gündemi önermektedir.

Kaynakça

Chen, B., Guo, Z., Yang, Z., Chen, Y., Chen, J., Liu, Z., Shi, C. ve Yang, C. (2025). PathRAG: Pruning graph-based retrieval augmented generation with relational paths. *arXiv preprint* arXiv:2502.14902. https://doi.org/10.48550/arXiv.2502.14902

Edge, D., Trinh, H., Cheng, N., Bradley, J., Chao, A., Mody, A., Truitt, S., Metropolitansky, D., Ness, R. O. ve Larson, J. (2024). From local to global: A Graph RAG approach to query-focused summarization. *arXiv preprint* arXiv:2404.16130. https://doi.org/10.48550/arXiv.2404.16130

Fan, W., Ding, Y., Ning, L., Wang, S., Li, H., Yin, D., Chua, T.-S. ve Li, Q. (2024). A survey on RAG meeting LLMs: Towards retrieval-augmented large language models. *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* içinde (ss. 6491–6501). ACM. https://doi.org/10.1145/3637528.3671470

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, M. ve Wang, H. (2023). Retrieval-augmented generation for large language models: A survey. *arXiv preprint* arXiv:2312.10997. https://doi.org/10.48550/arXiv.2312.10997

Guo, Z., Xia, L., Yu, Y., Ao, T. ve Huang, C. (2024). LightRAG: Simple and fast retrieval-augmented generation. *arXiv preprint* arXiv:2410.05779. https://doi.org/10.48550/arXiv.2410.05779

Gutiérrez, B. J., Shu, Y., Gu, Y., Yasunaga, M. ve Su, Y. (2024). HippoRAG: Neurobiologically inspired long-term memory for large language models. *Advances in Neural Information Processing Systems, 37*. arXiv:2405.14831.

Han, H., Wang, Y., Shomer, H., Guo, K., Ding, J., Lei, Y., Halappanavar, M., Rossi, R. A., Mukherjee, S., Tang, X., He, Q., Hua, Z., Long, B., Zhao, T., Shah, N., Javari, A., Xia, Y. ve Tang, J. (2025). Retrieval-augmented generation with graphs (GraphRAG). *arXiv preprint* arXiv:2501.00309. https://doi.org/10.48550/arXiv.2501.00309

He, X., Tian, Y., Sun, Y., Chawla, N. V., Laurent, T., LeCun, Y., Bresson, X. ve Hooi, B. (2024). G-Retriever: Retrieval-augmented generation for textual graph understanding and question answering. *Advances in Neural Information Processing Systems, 37*, 132876–132907. arXiv:2402.07630.

Hogan, A., Blomqvist, E., Cochez, M., D'Amato, C., de Melo, G., Gutierrez, C., Kirrane, S., Labra Gayo, J. E., Navigli, R., Neumaier, S., Ngonga Ngomo, A.-C., Polleres, A., Rashid, S. M., Rula, A., Schmelzeisen, L., Sequeda, J., Staab, S. ve Zimmermann, A. (2021). Knowledge graphs. *ACM Computing Surveys, 54*(4), Makale 71, 1–37. https://doi.org/10.1145/3447772

Hu, Y., Lei, Z., Zhang, Z., Pan, B., Ling, C. ve Zhao, L. (2024). GRAG: Graph retrieval-augmented generation. *arXiv preprint* arXiv:2405.16506. https://doi.org/10.48550/arXiv.2405.16506

Huang, Y. ve Huang, J. X. (2026). A survey on retrieval-augmented text generation for large language models. *ACM Computing Surveys, 58*(12), Makale 300, 1–38. https://doi.org/10.1145/3805774

Ji, S., Pan, S., Cambria, E., Marttinen, P. ve Yu, P. S. (2022). A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Transactions on Neural Networks and Learning Systems, 33*(2), 494–514. https://doi.org/10.1109/TNNLS.2021.3070843

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A. ve Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys, 55*(12), Makale 248, 1–38. https://doi.org/10.1145/3571730

Karpukhin, V., Oğuz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D. ve Yih, W. (2020). Dense passage retrieval for open-domain question answering. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* içinde (ss. 6769–6781). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.emnlp-main.550

Lavrinovics, E., Biswas, R., Bjerva, J. ve Hose, K. (2025). Knowledge graphs, large language models, and hallucinations: An NLP perspective. *Journal of Web Semantics, 85*, Makale 100844. https://doi.org/10.1016/j.websem.2024.100844

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S. ve Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems, 33*, 9459–9474.

Luo, L., Li, Y.-F., Haffari, G. ve Pan, S. (2024). Reasoning on graphs: Faithful and interpretable large language model reasoning. *Proceedings of the Twelfth International Conference on Learning Representations (ICLR)*. arXiv:2310.01061.

Mavromatis, C. ve Karypis, G. (2024). GNN-RAG: Graph neural retrieval for large language model reasoning. *arXiv preprint* arXiv:2405.20139. https://doi.org/10.48550/arXiv.2405.20139

Pan, S., Luo, L., Wang, Y., Chen, C., Wang, J. ve Wu, X. (2024). Unifying large language models and knowledge graphs: A roadmap. *IEEE Transactions on Knowledge and Data Engineering, 36*(7), 3580–3599. https://doi.org/10.1109/TKDE.2024.3352100

Peng, B., Zhu, Y., Liu, Y., Bo, X., Shi, H., Hong, C., Zhang, Y. ve Tang, S. (2025). Graph retrieval-augmented generation: A survey. *ACM Transactions on Information Systems, 44*(2), 1–52. https://doi.org/10.1145/3777378

Procko, T. T. ve Ochoa, O. (2024). Graph retrieval-augmented generation for large language models: A survey. *2024 Conference on AI, Science, Engineering, and Technology (AIxSET)* içinde (ss. 166–169). IEEE.

Sarmah, B., Mehta, D., Hall, B., Rao, R., Patel, S. ve Pasquali, S. (2024). HybridRAG: Integrating knowledge graphs and vector retrieval augmented generation for efficient information extraction. *Proceedings of the 5th ACM International Conference on AI in Finance (ICAIF)* içinde (ss. 608–616). ACM. https://doi.org/10.1145/3677052.3698671

Singh, A., Ehtesham, A., Kumar, S. ve Talaei Khoei, T. (2025). Agentic retrieval-augmented generation: A survey on agentic RAG. *arXiv preprint* arXiv:2501.09136. https://doi.org/10.48550/arXiv.2501.09136

Sun, J., Xu, C., Tang, L., Wang, S., Lin, C., Gong, Y., Shum, H.-Y. ve Guo, J. (2024). Think-on-Graph: Deep and responsible reasoning of large language model on knowledge graph. *Proceedings of the Twelfth International Conference on Learning Representations (ICLR)*. arXiv:2307.07697.

Wang, S., Huang, J., Chen, Z., Song, Y., Tang, W., Mao, H., Fan, W., Liu, H., Liu, X., Yin, D. ve Li, Q. (2025). Graph machine learning in the era of large language models (LLMs). *ACM Transactions on Intelligent Systems and Technology, 16*(5), Makale 104, 1–40. https://doi.org/10.1145/3732786

Xiong, G., Jin, Q., Lu, Z. ve Zhang, A. (2024). Benchmarking retrieval-augmented generation for medicine. *Findings of the Association for Computational Linguistics: ACL 2024* içinde (ss. 6233–6251). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.findings-acl.372

Yang, Z., Qi, P., Zhang, S., Bengio, Y., Cohen, W. W., Salakhutdinov, R. ve Manning, C. D. (2018). HotpotQA: A dataset for diverse, explainable multi-hop question answering. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)* içinde (ss. 2369–2380). Association for Computational Linguistics. https://doi.org/10.18653/v1/D18-1259

Zhang, Q., Chen, S., Bei, Y., Yuan, Z., Zhou, H., Hong, Z., Chen, H., Xiao, Y., Zhou, C., Dong, J., Chang, Y. ve Huang, X. (2025). A survey of graph retrieval-augmented generation for customized large language models.

arXiv preprint arXiv:2501.13958. https://doi.org/10.48550/arXiv.2501.13958

Zhao, P., Zhang, H., Yu, Q., Wang, Z., Geng, Y., Fu, F., Yang, L., Zhang, W., Jiang, J. ve Cui, B. (2026). Retrieval-augmented generation for AI-generated content: A survey. *Data Science and Engineering*. https://doi.org/10.1007/s41019-025-00335-5

Zhu, Z., Huang, T., Wang, K., Ye, J., Chen, X. ve Luo, S. (2026). Graph-based approaches and functionalities in retrieval-augmented generation: A comprehensive survey. *ACM Computing Surveys*. https://doi.org/10.1145/3795880

7. Bölüm

Büyük Dil Modellerinin (LLM) Temelleri, Mimarileri, Uygulamaları ve Gelecek Araştırmaları

Mahmut SAYAR[1]

1. Giriş

Küresel yapay zekâ teknolojilerindeki gelişim incelendiğinde; insan dilini anlama ve insani özellik ve yetenekleri üretme fiillerinin bu teknolojileri geliştiren tüm paydaşlar için en zorlayıcı görevler olduğu gözlenmektedir. Doğal dil işleme (NLP) çalışmaları, bu zorluğu aşabilmek için uzun yıllar kural tabanlı sistemlerden istatistiksel dil modellerine ve sonrasında derin öğrenme tabanlı mimarilere uzanan bir dönüşüm geçirmiştir (Zhao et al., 2023). Özellikle Yinelemeli Sinir Ağları ve Uzun Kısa Süreli Bellek gibi geleneksel derin öğrenme mimarileri, sıralı verileri işlemedeki başarılarına rağmen, uzun vadeli bağımlılıkları yönetmede yetersiz kalmış ve paralel hesaplama kısıtları nedeniyle büyük veri kümeleri üzerinde ölçeklenememiştir (Vaswani et al., 2017). Doğal dil işleme araştırmalarındaki bu kısıt, Vaswani et al. (2017) tarafından geliştirilen ve yinelemeli modelleri devre dışı bırakarak öz dikkat mekanizmasını temel alan transformer mimarisinin geliştirilmesiyle kökten bir dönüşüme uğramıştır. Transformer mimarisinin sunduğu paralel işlem yeteneği, dil modellerinin daha önce hayal edilemeyen büyüklükte veri kümeleriyle eğitilmesinin önünü açmıştır. Bu mimari zemin üzerine inşa edilen Bidirectional Encoder Representations from Transformers-BERT ve Generative Pre-trained Transformer-GPT gibi önceden eğitilmiş dil modelleri, ön eğitim ve ince ayar modelini literatüre kazandırmıştır (Devlin et al., 2018), (Radford et al., 2018).

Model parametrelerinin, eğitim veri setlerinin ve hesaplama güçlerinin milyarlarca parametre seviyesine çıkarılması sayesinde literatürde Büyük Dil Modelleri (LLM) olarak adlandırılan yeni bir aşamaya geçilmiştir (Bommasani et al., 2021). LLM' ler, geleneksel NLP görevlerindeki başarıları yanında, model ölçeği belirli kritik eşikleri aştığında aniden ortaya çıkan beliren yetenekler ile de karakterize edilmektedir. Bu beliren yetenekler arasında birkaç örnekle öğrenme, sıfır örnekle akıl yürütme ve karmaşık çok adımlı problem çözme yer almaktadır (Brown et al., 2020; Wei et al., 2022). Bu çalışmada, büyük dil modellerinin

[1] Dr. Öğretim Üyesi, İzmir Bakırçay Üniversitesi, İzmir Bakırçay Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, İşletme Bölümü, İzmir, Türkiye,
ORCID: 0000-0002-1852-6276

temelleri, mimari yapıları, eğitim metodolojileri, pratik uygulama alanları, etik ve teknik sınırlılıkları bütüncül bir yapıyla ele alınmaktadır. Bu kapsamda, çalışmada istatistiksel modellerden nöral ve büyük dil modellerine geçişteki temel teorik gelişmelerin neler olduğu; transformer mimarisi ve öz dikkat mekanizması dil temsilini nasıl dönüştürdüğü; hizalama, pekiştirmeli öğrenme ve parametre verimli ince ayar tekniklerinin modellerin yönlendirilebilirliğini nasıl artırdığı; LLM' lerin farklı disiplinlerdeki katkıları, sınırlılıkları ve taşıdıkları etik risklerin neler olduğu ve gelişmiş artırılmış ve ajan tabanlı sistemlerin gelecekteki araştırma yönelimlerinin neler olacağı sorularına yanıt aranmaktadır.

2. Büyük Dil Modellerinin Temelleri

2.1. Dil Modellerinin Gelişimi

Dil modelleme genel anlamda bir kelime dizisinin veya token dizisinin olasılık dağılımını belirleme süreci olarak tanımlanabilmektedir. Geleneksel istatistiksel dil modelleri, zincir kuralına dayanarak n-gram olarak adlandırılan, bir sonraki kelimenin olasılığını, kendisinden önce gelen sınırlı sayıdaki kelimeye bağlı olarak tahmin etmeye çalışmaktadır (Zhao et al., 2023). Ancak bu yaklaşımlar, eğitim verisinde görünmeyen kelime dizileri için veri seyrekliği nedeniyle olasılığın sıfıra düşmesi problemiyle karşılaşmış ve bu sorunu aşmak için Kneser-Ney gibi karmaşık yumuşatma yöntemlerine ihtiyaç duymuştur. Yapay sinir ağları çalışmalarının katkılarıyla nöral dil modelleri geliştirilmiş ve kelimeler arasındaki anlamsal ilişkiler sürekli vektör uzaylarında temsil edilmeye başlanmıştır. Nöral yaklaşımlar, dil modellemeyi bir sonraki kelimeyi tahmin etme veya maskelenmiş kelimeleri doldurma görevi olarak formüle etmiştir (Devlin et al., 2018; Radford et al., 2018). Bu yaklaşım, LLM' lerin temelini oluşturan denetimsiz ön eğitimin de yöntemsel hareket noktasını belirlemiştir (Bommasani et al., 2021).

2.2. Derin Öğrenme ve Temsil Öğrenme

Dilin bilgisayarlar tarafından işlenebilmesi için sembolik kelimelerin sayısal vektörlere dönüştürülmesi gerekmektedir. Geleneksel tek sıcaklık vektörü gösterimleri, kelimeler arasındaki anlamsal benzerlikleri ve eş anlamlılık ilişkilerini yakalayamamaktadır. Derin öğrenme ile birlikte gelişen temsil öğrenme teorisi, kelimeleri düşük boyutlu ve yoğun vektörlerle temsil eden word embeddings diye adlandırılan kelime gömmeleri kavramını ortaya çıkarmıştır (Zhao et al., 2023). Büyük dil modellerinde gömme katmanları, her tokenı yüksek boyutlu sürekli bir vektör uzayına (*dmodel*) izdüşürür (Vaswani et al., 2017). Temsil öğrenmenin asıl başarısı, Word2Vec gibi durağan gömmelerin ötesine geçerek; bir kelimenin bağlama göre değişen anlamını yakalayabilen bağlamsal

temsil yeteneğidir.. Örneğin, BERT ve GPT modelleri, aynı kelimenin farklı cümlelerdeki anlamsal nüanslarını, çevreleyen diğer tokenların vektör alanlarıyla etkileşime sokarak dinamik bir şekilde oluşturmaktadır (Devlin et al., 2018; Radford et al., 2018).

2.3. Attention ve Transformer Yaklaşımı

Transformer öncesi RNN ve LSTM gibi NLP mimarileri, dizideki her bir kelimeyi sırayla işlemek durumunda olduğundan bu durum sıralı hesaplama darboğazı ve bilgi kaybı ve gradyan kaybolması olarak iki temel sınırlılık oluşturmaktaydı. Sıralı hesaplama darboğazı, t anındaki hesaplamanın yapılabilmesi için $t-1$ anındaki gizli durum vektörünün beklenmesi gerekiyordu, bu da donanımların paralel işlem gücünden tam olarak yararlanılmasını engelliyordu şeklinde açıklanabilir. Bilgi kaybı ve gradyan kaybolması ise uzun dizilerde, dizinin başındaki bir bilginin dizinin sonuna taşınırken RNN hücreleri içinde sönümlenmesinin veya bozulmasının kaçınılmaz olduğu şeklinde açıklanabilmektedir (Vaswani et al., 2017). Vaswani et al. (2017) tarafından önerilen dikkat mekanizması, dizideki elemanlar arasındaki mesafeden bağımsız olarak, her bir tokenın diğer tüm tokenlarla doğrudan bağlantı kurmasını sağlayarak bu darboğazları ortadan kaldırmaktadır. Formülsel olarak öz dikkat şu şekilde ifade edilmektedir:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Burada Q sorgu, K anahtar ve V değer matrisleri, girdi vektörlerinin doğrusal izdüşümleri; sqrt(dk) ise iç çarpım değerlerinin çok büyümesini ve softmax fonksiyonunun gradyan sönümlenmesi bölgesine girmesini engelleyen ölçekleme katsayısıdır. Öz dikkat mekanizmasının sıraya duyarsız yapısı nedeniyle, dizideki kelimelerin sırasını modele bildirmek amacıyla konumsal kodlama yöntemleri geliştirilmiştir. Orijinal transformer da sinüsoidal konumsal kodlamalar kullanılırken (Vaswani et al., 2017); modern LLM' lerde göreli konumsal kodlamalar tercih edilmektedir. Bunlar arasında en dikkat çekenleri, attention logitlerine göreli mesafeye dayalı bir ceza matrisi ekleyen ALiBi (Press et al., 2022) ve döner matrisler kullanarak konumsal bilgiyi doğrudan iç çarpım işlemine entegre eden Rotary Position Embedding-RoPE teknolojileridir (Su et al., 2021).

2.4. Büyük Dil Modellerinin Temel Çalışma Mantığı

LLM' lerin temel çalışma mantığı iki aşamalı bir transfer öğrenme sürecine dayanmaktadır. Bunlar, devasa veri kümelerinde gerçekleştirilen öz denetimli ön eğitim ve ardından hedeflenen görevlere veya insan tercihlerine göre uyarlama yapılan fine tuning aşamalarıdır (Bommasani et al., 2021; Zhao et al., 2023). Ön eğitim aşamasında model, Common Crawl ve C4 veri kümeleri gibi internetten çekilen milyarlarca kelimelik yapılandırılmamış metin üzerinde bir sonraki kelimeyi tahmin etmeye çalışmaktadır (Raffel et al., 2020; Zhao et al., 2023). Bu basit amaç fonksiyonu, modelin metinlerdeki syntax, semantik, dil bilgisi yapılarını ve gerçek hatta dünya bilgisini örtük bir şekilde öğrenmesini sağlamaktadır (Brown et al., 2020). Modelin parametre sayısı, veri boyutu ve hesaplama gücü ölçeği arttıkça bu basit sonraki kelime tahmini görevi, karmaşık mantıksal çıkarım ve bağlam içi öğrenme gibi olağanüstü yeteneklerin belirmesine yol açmaktadır (Wei et al., 2022).

3. LLM Mimarileri ve Eğitim Yaklaşımları

3.1. Transformer Mimarisi

Transformer mimarisi, çok başlı dikkat katmanları ve konum bazlı ileri beslemeli ağların ardışık şekilde istiflenmesinden oluşmaktadır. Çok başlı dikkat mekanizması, modelin aynı anda farklı anlamsal alt uzaylardan gelen bilgileri işlemesini sağlamaktadır.

$$\mathrm{MultiHead}(Q, K, V) = \mathrm{Concat}(\mathrm{head}_1, \ldots, \mathrm{head}_h)W^O$$

$$\mathrm{head}_i = \mathrm{Attention}(QW_i^Q, KW_i^K, VW_i^V)$$

Katmanlar arasında eğitim kararlılığını artırmak için artık bağlantılar ve katman normalizasyonu kullanılmaktadır (Vaswani et al., 2017). Modern LLM' lerde katman normalizasyonunun blok girişlerine uygulandığı Pre-LN yapısı, eğitim kararlılığını artırdığı için standart hale gelmektedir (Raffel et al., 2020; Zhao et al., 2023).

3.2. Encoder ve Decoder Yaklaşımları

Transformer mimarisi, dikkat maskeleme yöntemlerine göre yalnızca encoder, yalnızca decoder ve encoder-decoder olarak üç temel varyasyona ayrılmaktadır (Raffel et al., 2020; Zhao et al., 2023): Yalnızca encoder tamamen görünür dikkat maskesi kullanmaktadır. Dizideki her bir token, sağındaki ve solundaki tüm tokenlara erişebilmektedir (Devlin et al., 2018). BERT bu sınıfın en bilinen temsilcisidir. Metin sınıflandırma ve soru cevaplama gibi anlama görevlerinde

etkilidir. Yalnızca decoder nedensel dikkat maskesi kullanmaktadır Bir token yalnızca kendisinden önce gelen tokenlara dikkat yöneltebilmektedir (Radford et al., 2018; Brown et al., 2020). Üretken görevlerde üstün olan GPT serisi ve LLaMA, Mistral ve PaLM gibi modern LLM' lerin çoğunluğu bu mimariye dayanmaktadır. Encoder-decoder girdiyi çift yönlü işleyen bir encoder ve çıktıyı nedensel olarak üreten bir decoder dan oluşmaktadır İki stack arasında çapraz dikkat bağı kurulmaktadır (Vaswani et al., 2017; Raffel et al., 2020). T5 ve BART bu yaklaşımla metinden metine dönüşüm görevlerini başarıyla yürütmektedir. Prefix decoder ise causal decoder varyasyonudur. Ancak girdi kısmında çift yönlü dikkate izin verirken; üretim kısmında nedensel maskeleme uygulamaktadır (Zhao et al., 2023).

Tablo 1: Encoder ve decoder yaklaşımları

Mimari Türü	Dikkat Maskeleme Şekli	Temsilci Modeller	Temel Kullanım Amacı
Encoder-only	Çift yönlü (sınırsız)	BERT	Sınıflandırma, bilgi çıkarımı
Decoder-only	Nedensel (causal)	GPT serisi, LLaMA, Mistral	Üretken görevler, sohbet, akıl yürütme
Encoder-Decoder	Çapraz dikkatli karışık	T5, BART	Çeviri, özetleme, görev dönüşümleri
Prefix Decoder	Girdide çift yönlü, çıktıda nedensel	PaLM, GLM	Çok görevli üretim ve analiz

3.3. Pre-training ve Fine-tuning

Ön eğitim aşamasında milyarlarca parametreye sahip model dondurulmadan tüm ağırlıkları güncellenmektedir. Ancak belirli bir downstream göreve uyarlamak için modelin tüm parametrelerinin güncellenmesi hesaplama ve depolama açısından son derece maliyetlidir. Bu sorunu çözmek amacıyla parameter efficient fine-tuning-PEFT/delta tuning teknikleri geliştirilmiştir (Ding et al., 2023; Zhao et al., 2023). Adapter tuning' de transformer katmanları arasına küçük doğrusal darboğaz katmanları eklenmektedir ve yalnızca bu katmanlar eğitilmektedir. Prefix tuning' de her katmandaki anahtar ve değer matrislerinin önüne eğitilebilir sürekli vektörler yerleştirilmektedir. Prompt tuning' de modelin girdi gömme katmanına eğitilebilir sanal token vektörleri eklenmektedir. Low-rank adaptation-LoRA' da ise ağırlık güncellemelerini ΔW düşük rütbeli iki matrisin çarpımı $A \times B$ olarak yeniden parametrelendirir:

$$W' = W + \Delta W = W + \frac{\alpha}{r}(A \times B)$$

Burada $r \ll d$ (rütbe/rank) olup eğitilen parametre sayısını %99'un üzerinde azaltırken orijinal model performansını korumaktadır (Hu et al., 2021; Ding et al., 2023).

3.4. Instruction Tuning ve RLHF

Önceden eğitilmiş bir LLM, kullanıcıların niyetlerini anlamak yerine yalnızca metni devam ettirmeye meyillidir. Modelleri yönlendirilebilir ve yararlı kılmak için talimat uyumlama ve insan geri bildirimiyle pekiştirmeli öğrenme adımları uygulanmaktadır (Ouyang et al., 2022; Zhao et al., 2023). Talimat uyumlama ile modelin, doğal dilde yazılmış talimat formatındaki veri setleriyle eğitilerek görünmeyen görevlere sıfır örnekle genelleme yapabilmesi sağlanmaktadır (Chung et al., 2022). Peng et al. (2023) çalışmalarında GPT-4 gibi güçlü kapalı kaynaklı modelleri eğitmen olarak kullanarak açık kaynaklı modeller için yüksek kaliteli talimat verileri üretilebileceğini göstermiştir. İnsan geri bildirimiyle pekiştirmeli öğrenmede model çıktıları insan etiketleyiciler tarafından sıralanarak bir ödül modeli eğitilmektedir. Ardından, hedef model Proximal Policy Optimization-PPO algoritması kullanılarak bu ödül modeline göre optimize edilmektedir. Bu süreçte, modelin genel yeteneklerinin gerilemesi durumuna hizalama vergisi denilmektedir. Bu vergiyi azaltmak için pekiştirmeli öğrenme gradyanlarına ön eğitim veri kaybı karıştırılmaktadır (Ouyang et al., 2022). Son dönemlerde, insan yerine doğrudan bir yapay zekâ modelinin kurallara göre değerlendirme yaptığı Reinforcement Learning from AI Feedback- RLAIF yöntemleri de sıklıkla kullanılmaktadır (Bai et al., 2022; Lee et al., 2023).

3.5. Parametre Ölçeklendirme ve Büyük Model Yaklaşımı

Dil modellerinde ölçeklendirmenin performansı nasıl etkilediği KM ölçekleme yasası (Kaplan et al., 2020) ve Chinchilla ölçekleme yasası (Hoffmann et al., 2022) olarak iki temel yasayla açıklanmaktadır. KM ölçekleme yasası OpenAI ekibi, model performansının model boyutu (*N)* veri seti boyutu (*D)* ve toplam hesaplama gücü (*C)* ile bir güç yasası ilişkisi içinde olduğunu savunmaktadır. Bu yasaya göre, artan bütçelerde parametre sayısının (*N)* veri miktarından (*D)* çok daha hızlı ölçeklendirilmesi gerektiği öne sürülmüştür. Chinchilla ölçekleme yasası Google DeepMind ekibi, Kaplan et al. (2020) çalışmalarının veri miktarını sabit tutarak hata yaptığını göstermiş ve daha geniş bir yelpazede gerçekleştirdikleri deneysel analizlerle, hesaplama-optimal bir eğitim için model boyutu ile eğitim token sayısının eşit oranda ölçeklendirilmesi gerektiğini kanıtlamıştır. Bu yasaya göre, 70 milyar parametreye sahip Chinchilla modelinin 1.4 trilyon token ile eğitilmesi durumunda, kendisinden çok daha büyük olan Gopher (280B) ve GPT-3 (175B) modellerinden daha iyi performans

sergilediği gösterilmiştir. Bu yasanın oluşturduğu en büyük dönüşüm, çıkarım maliyetlerini düşürmek adına küçük modellerin token sayısını artırarak çok daha uzun süre eğitilmesidir. LLaMA (Touvron et al., 2023) ve Mistral (Jiang et al., 2023) gibi modeller bu yaklaşımla geliştirilmiştir. Mistral 7B, çıkarım hızını ve bellek verimliliğini artırmak amacıyla Grouped-Query Attention- GQA ve uzun dizileri verimli şekilde işleyen Sliding Window Attention-SWA mekanizmalarını kullanarak 13B'lik LLaMA modellerini geride bırakmıştır.

3.6. RAG ve Gelişmiş LLM Yaklaşımları

Büyük dil modelleri, statik parametrik hafızaları nedeniyle güncel bilgilere erişememekte ve halüsinasyon adı verilen uydurma bilgi üretme problemine karşı savunmasız kalmaktadır (Zhao et al., 2023). Bu sorunu aşmak amacıyla Retrieval-Augmented Generation-RAG teknikleri geliştirilmiştir (Lewis et al., 2020). RAG sistemi, Wikipedia veya kurumsal veri tabanı gibi parametrik olmayan bir harici indeks ile BART, T5 veya herhangi bir LLM gibi parametrik bir üreticiyi bir araya getirmektedir. Lewis et al. (2020) RAG-sequence ve RAG-token olarak iki temel RAG kurgusu önermiştir. RAG-sequence harici kaynaktan alınan en ilgili k adet belgenin her birini kullanarak tüm dizinin üretim olasılığını hesaplamakta ve bu olasılıkları arındırıp özelleştirmektedir. RAG-token her bir tokenın üretimi sırasında farklı bir belge kümesine dikkat yönelterek daha dinamik bir içerik üretimi yapmaktadır. RETRO (Borgeaud et al., 2022) ve Atlas (Izacard et al., 2022) gibi modern yapılar, milyarlarca tokenlık veritabanlarını ön eğitim aşamasından itibaren modele entegre ederek parametrik boyutu küçük tutarken yüksek performans elde etmektedir.

3.7. Multimodal ve Agent Tabanlı Yaklaşımlar

Dil modellerinin yalnızca metin formatıyla sınırlı kalmayıp görüntü, ses ve video gibi farklı duyulara ait girdileri de işlemesiyle Multimodal Large Language Models-MLLM ortaya çıkmıştır (Gemini Team, 2023; Zhao et al., 2023). Bu süreç, görsel-dilsel hizalama ön eğitimi ve ardından görsel talimat uyumlama aşamalarından oluşmaktadır. Ayrıca LLM' leri otonom karar alıcılar haline getiren ajan tabanlı ve araç kullanan yaklaşımlar literatürde hızla artmaktadır: Reasoning and Acting-ReAct (Yao et al., 2022) modelin adım adım akıl yürütme izleri üretmesini ve Wikipedia API sorgusu gibi harici araçları tetikleyen eylemler gerçekleştirmesini sağlayan entegre bir prompt metodolojisidir. Akıl yürütme eylemleri yönlendirirken; eylemden gelen gözlemler de akıl yürütmeyi güncellemektedir. Toolformer (Schick et al., 2023) modelin hesap makinesi, takvim veya çeviri API' leri gibi dış araçları ne zaman ve nasıl çağıracağını tamamen kendi kendine öğrenmesini sağlayan bir mimaridir. Multi-agent

sistemler MetaGPT (Hong et al., 2023) ve Generative Agents (Park et al., 2023) gibi yapılar, birden fazla LLM ajanının yazılım mimarı ve test mühendisi gibi farklı sosyal rolleri üstlenerek karmaşık yazılım veya simülasyon projelerini otonom olarak yürütmesini sağlamaktadır. GPT-4 tabanlı olan Coscientist (Boiko et al., 2023) sistemi; internet araması yapma, API dokümantasyonunu okuma, python kodları çalıştırma ve bulut laboratuvarlarındaki sıvı işleme robotlarını otonom olarak kontrol etme yeteneklerini birleştirerek, paladyum katalizörlü çapraz kenetlenme reaksiyonları gibi karmaşık kimyasal deneyleri tamamen kendi kendine tasarlayıp optimize edebilmiştir.

4. LLM' lerin Yetenekleri ve Kullanım Alanları

Büyük dil modelleri; metin üretimi, dil anlama, bağlam işleme, soru sorma, cevap verme, özetleme, çeviri, bilgi çıkarımı, sınıflandırma ve kod üretimi gibi geniş yelpazede yeteneklere sahiptir (Zhao et al., 2023). Akıl yürütme yeteneği, modelin problemi alt problemlere bölerek çözdüğü düşünce zinciri (Wei et al., 2022) ve birden fazla çözüm yolu arasından oylama yaptığı öz tutarlılık (Wang et al., 2022) gibi tekniklerle pratik olarak gerçekleşmektedir. Ancak, dil bazlı çıktı üretme başarısı ile insani anlama, bilinç veya bilişsel süreçler arasında kesin bir teorik ayrım yapılması akademik güvenilirlik açısından zorunludur. LLM' lerin akıl yürütme performansı, insan zihnindeki mantıksal çıkarım süreçlerinden yapısal olarak farklılık göstermektedir. Modeller, semboller arasındaki olasılıksal korelasyonları ve metinsel kalıpları devasa parametre uzaylarında eşleştiren gelişmiş istatistiksel örüntü eşleştiricileri olarak görev yapmaktadır (Bommasani et al., 2021). Bu nedenle, LLM’ nin düşündüğü, bildiği veya anladığı şeklindeki ifadeler bilimsel açıdan hatalı olup; modellerin dil üretim başarısının ardındaki deterministik matematiksel ve istatistiksel modeller göz ardı edilmemelidir.

5. LLM Uygulamaları

5.1. Eğitim

Büyük dil modellerinin eğitim süreçlerine entegrasyonu, geleneksel öğrenme yöntemlerini kökten değiştirebilecek bir potansiyele sahiptir. Bu alanda LLM' ler; bireyselleştirilmiş sanal öğretmenlik, soru-cevap asistanlığı ve diyalog bazlı dinamik öğrenme süreçlerinin kurgulanmasında aktif olarak kullanılmaktadır (Kasneci et al., 2023; Thirunavukarasu et al., 2023). Modellerin sağladığı en önemli avantaj, her öğrencinin bireysel öğrenme hızına, bilgi düzeyine ve bilişsel gereksinimlerine uygun kişiselleştirilmiş eğitim materyalleri sunabilmesi ve küresel ölçekte dil engellerini ortadan kaldırarak eğitimde fırsat eşitliğine katkı sağlayabilmesidir. Eğitimde yapay zekâ kullanımının ciddi pedagojik sınırlılıkları ve riskleri de bulunmaktadır. Modellerin ürettiği eğitsel içeriklerin

doğruluğunun ve bilimsel geçerliliğinin gerçek zamanlı olarak denetlenmesindeki zorluklar, bu teknolojinin doğrudan eğitsel bir araç olarak kabulünü zorlaştırmaktadır. Öğrencilerin hazır bilgiye kontrolsüz erişimi nedeniyle eleştirel düşünme, problem çözme ve kavramlar arasında bağ kurabilme gibi üst düzey bilişsel yeteneklerinin gerilemesi riski de bulunmaktadır. Akademik dürüstlük ihlalleri, intihal ve uydurulmuş bilgilerin doğruymuş gibi kabul edilmesi gibi etik riskler, eğitim süreçlerinde insan denetiminin korunmasını ve yeni ölçme değerlendirme yöntemlerinin geliştirilmesini gerekli kılmaktadır (Kasneci et al., 2023; Sallam, 2023).

5.2. Sağlık

Klinik tıp ve sağlık hizmetleri, büyük dil modellerinin en yüksek katma değer oluşturduğu; aynı zamanda hata payının sıfır olması gerektiği nedeniyle en hassas etik sınırlılıkları barındırdığı alanların başında gelmektedir. Sağlık alanında LLM' ler; klinik karar destek sistemlerinin desteklenmesi, karmaşık tıbbi raporların ve radyoloji sonuçlarının hastaların anlayabileceği sade bir dile dönüştürülmesi ve epikriz özetleri ile hasta mektupları gibi idari evrak işlerinin otomasyonu gibi kritik işlevlerde kullanılmaktadır (Jeblick et al., 2022; Thirunavukarasu et al., 2023). Bu uygulamalar, hekimlerin klinik dışı bürokratik işlere harcadığı zamanı ciddi ölçüde azaltarak doğrudan hasta bakımına odaklanmalarına olanak tanımakta ve hekim-hasta arasındaki bilgi seviyesi farklılıklarını azaltmaktadır. Sağlık sektöründe insan sağlığının doğrudan söz konusu olması, LLM' lerin klinik entegrasyonunun önündeki en büyük engeli oluşturmaktadır. Güncel genel amaçlı modellerin klinik akıl yürütme seviyesi henüz uzman tıp profesyonellerinin bütünsel teşhis yeteneğine ulaşamamıştır (Singhal et al., 2023; Thirunavukarasu et al., 2023). Modellerin olasılıksal doğasından kaynaklanan halüsinasyon üretimi, hastalara yanlış veya hayati tehlike arz eden tedavi önerilerinin sunulması riskini beraberinde getirmektedir (Sallam, 2023). Buna ek olarak, hasta verilerinin korunması ve gizlilik standartlarının ihlal edilmesi endişesi, klinik ortamlarda kapalı döngü ve yerel olarak barındırılan özel modellerin kullanımını zorunlu kılmaktadır. GPT-4'ün USMLE tıp sınavlarındaki yüksek başarısı (Nori et al., 2023) ve tıp alanına özel olarak eğitilen Med-PaLM gibi modellerin klinik bilgi seviyesi (Singhal et al., 2023), bu teknolojinin potansiyelini kanıtlarken; hekim denetimi olmadan otonom teşhis süreçlerinde kullanımının henüz ciddi riskler barındırdığını göstermektedir.

5.3. Bilimsel Araştırma

Akademik araştırma ve bilimsel keşif süreçlerinde büyük dil modelleri, geleneksel yöntemleri hızlandıran ve araştırmacılara yeni perspektifler sunan hızlandırıcı roller üstlenmektedir. Bilimsel süreçlerde LLM' ler; devasa literatür yığınlarının taranması, araştırma makalelerinin dilsel olarak yapılandırılması, veri analizine yönelik kod bloklarının üretilmesi ve karmaşık deney tasarımlarının optimize edilmesi amacıyla kullanılmaktadır (Sallam, 2023). Özellikle araştırmacıların bilimsel çalışmalarını küresel standartlarda ifade etmelerine yardımcı olarak akademik dünyada dil temelli eşitsizlikleri azaltmaktadır. LLM' lerin mevcut bilimsel yöntemin dışına çıkarak tamamen özgün ve çığır açıcı bilimsel yenilikler üretme kapasitesi bulunmamaktadır. Çünkü bu modeller yalnızca geçmiş veriler ve örüntüler üzerinden ilişkiler ve çıkarımlar sunabilmektedir (Thirunavukarasu et al., 2023). Bilimsel çalışmalarda yapay zekâ kullanımının en riskli boyutu, modellerin gerçekte var olmayan akademik yayınları, DOI numaralarını ve araştırmacı adlarını uydurması ve bu sahte kaynakların bilimsel literatüre sızmasıdır (Sallam, 2023). Buna rağmen, otonom kimya ajanı Coscientist gibi ileri düzey sistemlerin, literatür taramasından sıvı işleme robotlarının kontrolüne kadar tüm deneysel aşamaları otonom olarak yöneterek başarılı kimyasal reaksiyonlar gerçekleştirebilmesi (Boiko et al., 2023), modellerin bilimsel araştırmalardaki rolünün otonom deney tasarımcılığına doğru geliştiğini göstermektedir.

5.4. Hukuk

Hukuk alanı, yoğun metin analizi ve belgelendirme süreçleri içermesi nedeniyle büyük dil modellerinin operasyonel verimlilik sağladığı kritik bir uygulama alanıdır. Hukuki süreçlerde LLM' ler; uzun sözleşmelerin risk analizlerinin yapılması, geniş içtihat arşivlerinde arama yapılması, hukuki belge taslaklarının hazırlanması ve dava dosyalarının özetlenmesi gibi görevlerde kullanılabilmektedir (Bommasani et al., 2021). Bu entegrasyon, avukatların ve hukuk danışmanlarının binlerce sayfalık dokümanı dakikalar içinde analiz etmesini sağlayarak operasyonel süreçlerde büyük ölçekli zaman ve maliyet tasarrufu sağlamaktadır. Ancak hukuki terim ve kavramların yerel mevzuatlara, yargı yetkilerine ve sosyo-politik ilişkilere göre gösterdiği farklılıklar, modellerin her alanda yorumlar yapmasını sınırlamaktadır. En büyük risklerden biri, modelin gerçekte var olmayan mahkeme kararlarını veya yasal maddeleri emsal olarak uydurması ve bu durumun savunma veya mütalaalarda kullanılarak yargıyı yanıltmasıdır. Ayrıca, müvekkillere ait hassas bilgilerin ticari LLM platformlarına yüklenmesi durumunda ortaya çıkan gizlilik ihlalleri ve modellerin eğitim verilerindeki önyargılar nedeniyle adalete erişimde makine taraflılığı oluşturma riski hukuki etik tartışmalarının merkezinde yer almaktadır.

GPT-4 modelinin ABD Baro Sınavını en başarılı %10'luk dilimde olması (Bubeck et al., 2023), modellerin karmaşık hukuki mantık yürütme kapasitesini ortaya koyarken; her çıktının son adımda yetkin bir hukukçu tarafından doğrulanması gerekmektedir.

5.5. Yazılım Geliştirme

Yazılım mühendisliği, büyük dil modellerinin doğrudan kod üretim yetenekleri nedeniyle en hızlı ve köklü dönüşüm geçiren sektörlerden biridir. LLM' ler yazılım geliştirme döngüsünde; otomatik kod tamamlama, kodlardaki mantıksal hataların tespiti ve giderilmesi, API veya kütüphanelere yönelik teknik dokümantasyonların yazılması ve algoritma optimizasyonları gibi süreçlerde aktif rol almaktadır (Chen et al., 2021; Bubeck et al., 2023). Bu araçlar, yazılım geliştiricilerin rutin ve tekrarlayan kod şablonlarını yazma yükünü hafifleterek onların üst düzey mimari tasarımlara odaklanmalarına imkân tanımakta ve yazılım geliştirme maliyetlerini düşürmektedir. Ancak modellerin karmaşık, dağıtık ve çok katmanlı yazılım sistemlerinin mimari kararlarını vermedeki yetersizliği en büyük teknik sınırlılıktır. Modellerin eğitim veri setlerinde yer alan hatalı veya güncelliğini yitirmiş kod blokları nedeniyle, üretilen kodların ciddi güvenlik açıkları barındırabilmesi sektörel açıdan büyük bir tehdittir. Ayrıca, telif hakkı koruması altındaki lisanslı kodların model çıktılarında doğrudan kullanılmasıyla ortaya çıkan fikri mülkiyet ihlalleri ve güvenli olmayan kod bloklarının kurumsal sistemlere sızması riski de bulunmaktadır (Chen et al., 2021). Codex ve benzeri kod modellerinin sunduğu yüksek kod üretim başarısı, yazılımcı iş gücü piyasasında rollerin değiştiğini ve yazılımcıların kod yazan kişilerden ziyade kod denetleyen ve sistem tasarlayan mühendislere dönüştüğünü göstermektedir.

5.6. İşletme, Finans ve Diğer Alanlar

Kurumsal yönetim, finans ve müşteri ilişkileri, büyük dil modellerinin veri analitiği ve doğal dil işleme yeteneklerinden en üst düzeyde yararlanan diğer önemli sektörlerdir. Bu alanlarda LLM' ler; finansal piyasa trendlerinin analizi, karmaşık kurumsal raporların ve bilançoların işlenmesi, müşteri hizmetlerinde üretken yapay zekâ tabanlı akıllı sohbet robotlarının kullanılması ve pazar araştırmalarının otomatikleştirilmesi gibi süreçlerde tercih edilmektedir. Bu entegrasyonlar, kurumlara veri odaklı karar alma yapılarını hızlandırma, müşteri memnuniyetini kesintisiz destekle artırma ve operasyonel maliyetleri en aza indirme gibi stratejik avantajlar sağlamaktadır. Bununla birlikte, finansal piyasaların yüksek volalite ve gürültülü yapısı, genel amaçlı modellerin tahmin başarılarını sınırlamaktadır. Modellerin olasılıksal doğası nedeniyle yapabileceği hatalı finansal öngörüler, kurumsal düzeyde büyük maddi kayıplara yol açma potansiyeli taşımaktadır. Ayrıca, otomatik üretilen manipülatif finansal analizlerin piyasaya sürülmesiyle oluşabilecek piyasa manipülasyonu riskleri ve manipülatif botların finansal istikrarı sarsma

potansiyeli ciddi birer tehdittir (Bommasani et al., 2021). Bu sınırlılıkları aşmak adına, Wu et al. (2023) tarafından 50 milyar parametre ile finans alanına özel veriler kullanılarak eğitilen BloombergGPT gibi dikey uzmanlık modelleri geliştirilmektedir Böylelikle genel modellerin finansal açıdan yetersizlikleri ve riskleri en aza indirilmeye çalışılmaktadır.

6. LLM'lerin Sınırlılıkları, Riskleri ve Etik Boyutları

6.1. Halüsinasyon ve Güvenilirlik

Büyük dil modellerinin temel teknik sorunu, etkileyici ve ikna edici bir üslupla tamamen yanlış veya uydurma bilgiler üretmeleridir. Halüsinasyon olarak adlandırılan bu durum girdiyle, bağlamla ve gerçeklikle çelişen halüsinasyon olmak üzere genel olarak üç sınıfa ayrılmaktadır (Sallam, 2023; Zhao et al., 2023). Girdiyle çelişen halüsinasyon, modelin, kullanıcının prompt içinde verdiği doğrudan girdideki bilgilerle çelişen çıktılar üretmesi olarak tanımlanmaktadır. Bağlamla çelişen halüsinasyon, modelin uzun metin üretim süreçlerinde, kendi ürettiği önceki cümlelerle çelişen iddialar ortaya koymasına verilen isimdir. Gerçeklikle ve bilgiyle çelişen halüsinasyon ise sahte akademik kaynak üretimi gibi modelin dünya bilgisi, bilimsel gerçekler veya nesnel gerçeklerle uyuşmayan durumları uydurması olarak tanımlanmaktadır. Bu sorun, modellerin hakikat kavramına sahip olmamasından ve yalnızca token dizilim olasılıklarını optimize etmelerinden kaynaklanmaktadır (Bommasani et al., 2021).

6.2. Önyargı ve Adalet

LLM' ler, ön eğitim verilerinde yer alan toplumsal önyargıları, cinsiyetçi, ırkçı ve ayrımcı ifadeleri öğrenmekte ve ürettikleri çıktılarda öğrendikleri bu kalıpları yeniden üretmektedir. Bu duruma, belirli meslek gruplarının erkeklerle eşleştirildiği ve daha düşük gelirli veya hizmet sektörü mesleklerinin kadınlarla eşleştirildiği örnek olarak verilebilmektedir. Böylece, modeller kurumsal karar alma mekanizmalarına entegre edildiğinde sistematik adaletsizliklerin artmasına yol açma riskleri oluşmaktadır (Bommasani et al., 2021; Sallam, 2023).

6.3. Gizlilik ve Güvenlik

Ön eğitimde kullanılan verilerin genişliği, modellerin kişisel verileri, gizli yazışmaları veya telifli bilgileri ezberlemesine yol açmaktadır (Bommasani et al., 2021). Kötü niyetli kişi veya kişiler, sorgu enjeksiyonu veya üyelik çıkarımı saldırılarıyla model parametrelerinin gerisinde kalan hassas kişisel sağlık veya finans verilerini sızdırabilmektedir (Zhao et al., 2023). Ayrıca, kötü niyetli aktörlerin eğitim verilerini manipüle ederek data poisoning olarak adlandırılan modele belirli açık

kapılar bırakması da ciddi siber güvenlik tehditleri arasında yer almaktadır (Bommasani et al., 2021).

6.4. Telif Hakları ve Fikri Mülkiyet

Milyarlarca internet sayfasının telif hakları gözetilmeden eğitim verisi olarak kullanılması, fikri mülkiyet hukuku açısından büyük tartışmalar oluşturmaktadır (Bommasani et al., 2021). Sanatçıların, yazarların ve yazılımcıların ürettiği telifli içeriklerin LLM çıktıları içinde neredeyse doğrudan kopyalanarak sunulması, mevcut telif hakları yasalarının sınırlarını zorlamaktadır (Sallam, 2023).

6.5. Akademik Etik

Akademik alanda LLM kullanımı; ödevlerin yapay zekâya yazdırılması, sınav güvenliğinin sarsılması ve araştırma dürüstlüğünün kaybolması gibi riskleri beraberinde getirmektedir. Bilimsel yayıncılıkta, LLM' lerin katkısı ICMJE/COPE yönergeleri kapsamında bir yazar sorumluluğu taşıyamayacağı için, modellerin doğrudan yazar olarak listelenmesi akademik çevre tarafından kesin olarak reddedilmektedir (Sallam, 2023).

6.6. Şeffaflık ve Açıklanabilirlik

Derin sinir ağlarının milyarlarca parametre içeren karmaşık yapısı, dil modellerini birer kara kutu haline dönüştürmektedir (Bommasani et al., 2021). Modelin belirli bir klinik veya hukuki kararı verirken hangi anlamsal yolları izlediği, hangi verilere dayandığı deterministik olarak açıklanamamaktadır. Bu durum, kararların denetlenmesini imkânsız hale getirmekte ve kritik sektörlerde LLM' lere duyulan güveni sarsmaktadır (Sallam, 2023).

6.7. Çevresel ve Hesaplama Maliyetleri

Büyük modellerin eğitimi ve çıkarım süreçleri, devasa hesaplama kaynaklarına ve yüksek elektrik tüketimlerine sebep olmaktadır (Bommasani et al., 2021). Tek bir büyük model pre-training sürecinin tonlarca karbondioksit salınımına yol açması; yapay zekâ araştırmalarının çevresel sürdürülebilirliği ve küresel iklim krizi üzerindeki etkileri konusunda ciddi endişeler oluşturmaktadır. Bu durum, literatürü daha yeşil ve enerji verimli mimarilere yönlendirmektedir (Zhao et al., 2023).

7. LLM' lerin Bilimsel Araştırmalardaki Rolü

Büyük dil modelleri, araştırmacıların literatür tarama, taslak oluşturma, kod yazma ve özetleme süreçlerini hızlandırarak bilimsel üretkenliği artırmaktadır. Özellikle ana dili İngilizce olmayan araştırmacıların makalelerinin dilini kolaylıkla çevirebilmesi açısından küresel ölçekte araştırma adaletini destekleyen bir araç

olarak yerini almaktadır. Ancak bu araçların en ciddi risklerden biri, modelin uydurduğu sahte atıf ve kaynakların literatüre girmesi ve bu sahte kaynakların peer review süreçlerinden gözden kaçarak yayımlanmasıdır. Ayrıca, araştırmacıların literatürü derinlemesine okumak yerine LLM özetlerine bağımlı kalması, bilim insanlarının eleştirel analiz, derin çıkarım ve bağımsız düşünme yeteneklerinin zamanla azalmasına yol açabilecektir. Akademik dürüstlüğü korumak amacıyla araştırmalarda kullanılan tüm yapay zekâ araçlarının metodoloji bölümlerinde açık şekilde beyan edilmesi ve üretilen her çıktının insan denetiminden geçirilmesi oldukça önemlidir (Sallam, 2023).

8. Gelecek Araştırmaları ve Perspektifleri

Büyük dil modelleri araştırmalarında güncel çalışmalar incelendiğinde; gelecek araştırmaların, parametre sayısını büyütmenin yanında hesaplama optimal ve küçük ve alana özgü uzman modellerin geliştirilmesine doğru yöneldiği görülmektedir. Hoffmann et al. (2022) tarafından atıfta bulunulan Chinchilla ölçekleme yasaları, bütçe sınırları dahilinde en yüksek model performansını elde etmek için parametrik boyutu artırmak yerine modellerin çok daha geniş ve yüksek kaliteli veri kümeleriyle eğitilmesi gerektiğini deneysel olarak kanıtlamıştır. Bu dönüşüm, çıkarım maliyetlerini ve donanım gereksinimlerini en aza indirirken; yüksek doğruluk sunabilen, yerel olarak barındırılmaya uygun küçük ölçekli modellerin önünü açmaktadır. Gelecekte tıp, finans veya hukuk gibi sektörlerin hassas veri yapılarıyla ön eğitime tabi tutulmuş ve ince ayarı yapılmış özel amaçlı uzman sistemlerin geliştirilmesine yönelik eğilimlerin güçlenmesi öngörülmektedir.

Modellerin yönlendirilebilirliğini ve güvenliğini artırmak amacıyla yapılan alignment çalışmalarında, maliyetli ve ölçeklenmesi zor olan insan geri bildirimlerine dayalı süreçlerin yerini daha gelişmiş ve otonom yaklaşımlar almaktadır. İnsan etiketçilerin sağladığı verilerin subjektiflik, tutarsızlık barındırması ve yüksek maliyeti olması araştırmacıları yapay zekâ ajanlarının geri bildirimlerinden yararlanan yapay zekâ geri bildirimli pekiştirmeli öğrenme (RLAIF) yapılarına yönlendirmektedir (Bai et al., 2022; Lee et al., 2023). Ayrıca, geleneksel insan geri bildiriminden pekiştirmeli öğrenme (RLHF) süreçlerinin karmaşık ödül modelleme ve istikrarsız pekiştirmeli öğrenme adımlarına duyduğu ihtiyacı ortadan kaldıran doğrudan tercih optimizasyonu gibi yenilikçi algoritmalar, modellerin toplumsal değerlerle, güvenlik protokolleriyle ve kullanıcı niyetleriyle çok daha verimli ve kararlı bir şekilde hizalanmasını sağlamaktadır (Zhao et al., 2023).

Büyük dil modellerinin statik yapısından kaynaklanan güncellik yitimi ve halüsinasyon gibi yapısal sınırlılıkların aşılması için dinamik bilgi entegrasyonu ve sürekli öğrenme teknikleri güncel araştırmalar arasında yer almaktadır. Lewis et al. (2020) tarafından temelleri atılan Arama-Artırımlı Üretim (RAG) mimarileri,

modellerin parametrik belleğindeki statik bilgiyi harici dinamik arama motorları, yapılandırılmış kurumsal veri tabanları ve gerçek zamanlı bilgi depolarıyla bir araya getirerek üretilen içeriklerin doğrulanabilirliğini artırmaktadır (Zhao et al., 2023). Gelecekteki sistemlerde, RAG entegrasyonunun sorgu-cevap mekanizması yanında modelin zamanla değişen dünya bilgisini parametre ağırlıklarını tamamen güncellemeden, bellek içi dinamik indeksleme ve sürekli öğrenme protokolleri aracılığıyla kendi bünyesine adapte edebileceği hibrit mimariler şeklinde kurgulanması yönünde güçlü bir eğilim bulunmaktadır.

Yapay zekâ sistemlerinin tekil ve pasif sohbet robotu rolünü aşarak otonom karar alıcılar haline gelmesi, gelecek vizyonunun en çarpıcı boyutlarından birini oluşturmaktadır. ReAct (Yao et al., 2022) ve Toolformer (Schick et al., 2023) gibi tekniklerle başlayan araç kullanma ve akıl yürütme işlemi, günümüzde birden fazla yapay zekâ ajanının ortak bir amaç doğrultusunda iş birliği yaptığı otonom çoklu ajan ekosistemlerine dönüşmektedir. Bu kapsamda, yazılım mühendisliği süreçlerini otonom olarak simüle edebilen MetaGPT (Hong et al., 2023) veya insan davranışlarını yapay topluluklar içinde taklit eden sistemler (Park et al., 2023) dikkat çekmektedir. Daha da önemlisi, GPT-4 tabanlı Coscientist gibi sistemlerin, internet araması, API okuma, kod çalıştırma ve fiziksel laboratuvar donanımlarını kontrol etme yeteneklerini birleştirerek karmaşık kimyasal reaksiyonları tamamen otonom olarak tasarlayıp gerçekleştirebilmesi (Boiko et al., 2023), yapay zekânın deneysel bilimlerde bağımsız bir araştırma ortağı haline gelebileceği güçlü bir gelecek öngörüsüdür.

Büyük dil modellerinin yaygınlaşmasıyla birlikte ortaya çıkan etik, yasal ve ekolojik riskler, sorumlu yapay zekâ ve yapay zekâ yönetişimi kavramlarını gelecek araştırmaların merkezine taşımaktadır. Modellerin tarafsızlık, adalet, şeffaflık ve açıklanabilirlik standartlarını yasal düzenlemelerle uyumlu hale getirecek teknik denetim araçlarının ve metodolojilerinin geliştirilmesi zorunlu bir araştırma yönelimi olarak kabul edilmektedir (Bommasani et al., 2021). Ayrıca, büyük modellerin eğitim ve çıkarım süreçlerinde harcanan devasa enerji bütçelerinin oluşturduğu çevresel tahribat, araştırmacıları daha yeşil, karbon-nötr ve enerji verimli mimarilere yöneltmektedir (Zhao et al., 2023). Gelecekteki araştırmalar, yapay zekânın teknik yeteneklerini daha ileriye taşırken; bu araştırmaların aynı zamanda veri gizliliğini, telif haklarını, fikri mülkiyet haklarını ve toplumsal refahı güvence altına alacak çok disiplinli yönetişim yapılarının inşasına odaklanacağı öngörülmektedir.

9. Sonuç

Büyük dil modelleri, doğal dil işleme alanında ortaya çıkan istatistiksel ve kural tabanlı yaklaşımlardan derin temsil öğrenimine ve sonrasında üretken yapay zekâ dönemine uzanan devrimsel bir teknolojik ilerlemenin ürünüdür. Vaswani et al.

(2017) tarafından geliştirilen ve yinelemeli mekanizmaların paralel hesaplama kısıtlarını ortadan kaldıran öz dikkat mekanizmalı transformer mimarisi, dil modellerini parametre ölçeği bazında milyarlarca seviyeye taşımaktadır. Bu devasa parametre ölçeklemesi, modellerin belirli eşikleri aştığında bağlam içi öğrenme ve çok adımlı mantıksal çıkarım yapabilme gibi beliren yetenekleri aniden sergilemelerinin yolunu açmaktadır (Wei et al., 2022). Dil üretimindeki bu önemli başarı, modellerin gerçek anlamda düşündüğü, bildiği veya anladığı yanılsamasına yol açmamalıdır. LLM' ler, semboller arasındaki olasılıksal korelasyonları ve metinsel kalıpları devasa parametre uzaylarında eşleştiren gelişmiş istatistiksel örüntü eşleştiricileridir (Bommasani et al., 2021).

Yapılan çalışmalar, dil modellerinin eğitim (Kasneci et al., 2023), tıp ve sağlık bilimleri (Thirunavukarasu et al., 2023), yazılım mühendisliği (Chen et al., 2021) ve hukuk (Bommasani et al., 2021) gibi kritik disiplinlerde devasa operasyonel verimlilik artışı sağladığını göstermektedir. Modellerin bu kritik alanlardaki otonom işlevleri hekim ve hukukçu gibi insan karar alıcıların bütünsel akıl yürütme seviyesinin henüz gerisindedir (Singhal et al., 2023). Coscientist gibi otonom çoklu araç entegrasyonu sağlayan yapılar aracılığıyla yapay zekânın bilimsel keşiflerde bağımsız bir araştırma ortağı haline gelmesi (Boiko et al., 2023) büyük bir bilimsel devrimin geleceğini gösterse de; modellerin olasılıksal doğası gereği ürettiği halüsinasyonlar, önyargılar ve gizlilik açıkları (Sallam, 2023; Zhao et al., 2023), sistemlerin her aşamada insan denetimi altında tutulmasını etik ve yasal zorunluluk altında tutmaktadır.

Gelecekteki araştırma yönelimleri, yapay zekânın dönüştürücü gücünü korurken; onu daha güvenli, açıklanabilir ve enerji verimli bir yapıya kavuşturmayı amaçlamaktadır. Chinchilla ölçekleme yasaları doğrultusunda, sadece parametre boyutunu büyütmek yerine veri miktarı yüksek ve çıkarım maliyeti düşük küçük modellerin (Touvron et al., 2023) ve alana özgü özelleştirilmiş dikey sistemlerin önemi giderek artmaktadır. RAG gibi harici bilgi kaynaklarıyla entegre çalışan sistemler (Lewis et al., 2020) ve insan hizalamasında daha kararlı otonom algoritmaların kullanılması, halüsinasyonları ve önyargıları en aza indirmede önemli rol oynamaktadır. Sonuç olarak, yapay zekanın sunduğu üst düzey dönüştürücü güçten güvenle yararlanabilmek; teknolojik faydaların abartılmadığı, risklerin deneysel ve bilimsel kanıtlarla ele alındığı, güçlü denetim ve hizalama mekanizmalarının işletildiği çok disiplinli bir ortak sorumluluk ve yönetişim anlayışını zorunlu kılmaktadır.

Kaynakça

Boiko, D. A., MacKnight, R., Kline, B., & Gomes, G. (2023). Autonomous chemical research with large language models. Nature, 624(7992), 570-558. https://doi.org/10.1038/s41586-023-06792-0

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., et al. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.

Borgeaud, S., Mensch, A., Hoffmann, J., Buchatskaya, E., Cai, T., Rutherford, E., de Las Casas, D., Hendricks, L. A., Welbl, J., Clark, A., et al. (2022). Improving language models by retrieving from trillions of tokens. International Conference on Machine Learning, 2206-2240.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877-1901.

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Liang, S., et al. (2023). Sparks of artificial general intelligence: Early experiments with GPT-4. arXiv preprint arXiv:2303.12712.

Chung, H. W., Hou, L., Longpre, S., Zoph, B., Tay, Y., Fedus, W., Li, Y., Wang, X., Mostajabi, M., Alikhani, S., et al. (2022). Scaling instruction-finetuned language models. arXiv preprint arXiv:2210.11416.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Ding, N., Qin, Y., Yang, G., Wei, F., Yang, Z., Su, Y., Hu, S., Chen, Y., Chan, C. M., Chen, W., et al. (2023). Parameter-efficient fine-tuning of large-scale pre-trained language models. Nature Machine Intelligence, 5(3), 220-235. https://doi.org/10.1038/s42256-023-00626-4

Guu, K., Lee, K., Zhao, Z., Luan, Y., & Chang, M. W. (2020). Retrieval augmented language model pre-training. International Conference on Machine Learning, 3929-3938.

Hoffmann, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., de Las Casas, D., Hendricks, L. A., Welbl, J., Clark, A., et al. (2022). Training compute-optimal large language models. arXiv preprint arXiv:2203.15556.

Hong, S., Zheng, X., Chen, J., Cheng, Y., Wang, J., Zhang, C., Wang, Z., Yau, S. K. S., Lin, Z., Zhou, L., et al. (2023). MetaGPT: Meta programming for multi-agent collaborative framework. arXiv preprint arXiv:2308.00352.

Hu, J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, We. (2021). LoRA: Low-rank adaptation of large language models. arXiv preprint arXiv:2106.09685.

Jiang, A. Q., Sablayrolles, A., Mensch, A., Bamford, C., Chaplot, D. S., de las Casas, D., Bressand, F., Lengyel, G., Lample, G., Saulnier, L., et al. (2023). Mistral 7B. arXiv preprint arXiv:2310.06825.

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., Maningo, J., & Tseng, V. (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLOS Digital Health, 2(2), e0000198. https://doi.org/10.1371/journal.pdig.0000198

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W. T., Rocktäschel, T., et al. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems, 33, 9459-9474.

Nori, H., King, N., McKinney, S. M., Carignan, D., & Horvitz, E. (2023). Capabilities of GPT-4 on medical challenge problems. arXiv preprint arXiv:2303.13375.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., et al. (2022). Training language models to follow instructions with human feedback. Advances in Neural Information Processing Systems, 35, 27730-27744.

Park, J. S., O'Brien, J. C., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, 1-22.

Peng, B., Li, C., He, P., Galley, M., & Gao, J. (2023). Instruction tuning with GPT-4. arXiv preprint arXiv:2304.03277.

Press, O., Smith, N. A., & Lewis, M. (2022). Train short, test long: Attention with linear biases enables input length extrapolation. International Conference on Learning Representations.

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

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. Journal of Machine Learning Research, 21(140), 1-67.

Sallam, M. (2023). ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns. Healthcare, 11(6), 887. https://doi.org/10.3390/healthcare11060887

Schick, T., Dwivedi-Yu, J., Dessì, R., Raileanu, R., Lomeli, M., Zettlemoyer, L., Cancedda, N., & Scialom, T. (2023). Toolformer: Language models can teach themselves to use tools. Advances in Neural Information Processing Systems, 36.

Singhal, K., Azizi, S., Tu, T., Mahdavi, S. S., Wei, J., Chung, H. W., Scales, N., Tanwani, A., Cole-Lewis, H., Pfohl, S., et al. (2023). Large language models encode clinical knowledge. Nature, 620(7972), 172-180. https://doi.org/10.1038/s41586-023-06291-2

Su, J., Lu, Y., Pan, S., Murtadha, A., Wen, B., & Liu, Y. (2021). Roformer: Enhanced transformer with rotary position embedding. arXiv preprint arXiv:2104.09864.

Thirunavukarasu, A. J., Ting, D. S. J., Elangovan, K., Gutierrez, L., Tan, T. F., & Ting, D. S. W. (2023). Large language models in medicine. Nature Medicine, 29(10), 2430-2440. https://doi.org/10.1038/s41591-023-02448-8

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M. A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., et al. (2023). LLaMA: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971.

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

Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2022). ReAct: Synergizing reasoning and acting in language models. International Conference on Learning Representations.

Zhao, K., Zhou, K., Zhao, W. X., Wan, J., Zhang, F., Zhang, D., Gai, K., & Wen, J. R. (2023). A survey of large language models. arXiv preprint arXiv:2303.18223.