

# Bilgisayar Bilimleri ve Mühendisliđi Alanında Güncel Arařtırmalar

Editör: Doç. Dr. Mürsel Ozan İNCETAŐ



ÖZGÜR  
YAYINLARI

# Bilgisayar Bilimleri ve Mühendisliđi Alanında Güncel Arařtırmalar

**Editör:**

Doç. Dr. Mürsel Ozan İNCETAŞ



Published by

**Özgür Yayın-Dağıtım Co. Ltd.**

Certificate Number: 45503

📍 15 Temmuz Mah. 148136. Sk. No: 9 Şehitkamil/Gaziantep

☎ +90.850 260 09 97

📞 +90.532 289 82 15

🌐 www.ozgurayinlari.com

✉ info@ozgurayinlari.com

---

## Bilgisayar Bilimleri ve Mühendisliği Alanında Güncel Araştırmalar

Editör: Doç. Dr. Mürsel Ozan İNCETAŞ

---

Language: Turkish

Publication Date: 2026

Cover design by Mehmet Çakır

Cover design and image licensed under CC BY-NC 4.0

Print and digital versions typeset by Çizgi Medya Co. Ltd.

**ISBN (PDF):** 978-625-8562-58-3

**DOI:** <https://doi.org/10.58830/ozgur.pub1230>

---



This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0). To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc/4.0/>  
This license allows for copying any part of the work for personal use, not commercial use, providing author attribution is clearly stated.

---

Suggested citation:

İncetaş, M. O. (ed) (2026). *Bilgisayar Bilimleri ve Mühendisliği Alanında Güncel Araştırmalar*. Özgür Publications.

DOI: <https://doi.org/10.58830/ozgur.pub1230>. License: CC-BY-NC 4.0

---

*The full text of this book has been peer-reviewed to ensure high academic standards. For full review policies, see <https://www.ozgurayinlari.com/>*

---



## Ön Söz

Bilgisayar bilimleri ve mühendisliđi, günümüz dünyasında bilimsel arařtırmaların ve teknolojik gelişmelerin en hızlı ilerlediđi alanlardan biri haline gelmiştir. Yapay zekâ, veri bilimi, görüntü işleme, bilgisayar ağları, siber güvenlik ve yazılım mühendisliđi gibi birçok alt disiplin; hem akademik arařtırmaların hem de endüstriyel uygulamaların merkezinde yer almaktadır. Bu hızlı gelişim, söz konusu alanlarda yürütölen güncel çalışmaların paylaşılmasını ve farklı arařtırma toplulukları arasında bilgi alışverişinin güçlendirilmesini daha da önemli kılmaktadır.

Bilgisayar Bilimleri ve Mühendisliđi Alanında Güncel Arařtırmalar başlıklı bu kitap, bilgisayar bilimleri ve mühendisliđinin farklı alt alanlarında yürütölen güncel akademik çalışmaların bir araya getirilmesi amacıyla hazırlanmıştır. Kitapta yer alan bölümler; teorik yaklaşımlardan uygulamalı çalışmalara kadar geniş bir yelpazede arařtırma sonuçlarını içermekte olup, alanın güncel eğilimlerini ve arařtırma yönelimlerini yansıtmayı hedeflemektedir.

“Bilgisayar Bilimleri ve Mühendisliđi Alanında Güncel Arařtırmalar” kitabının arařtırmacılar, akademisyenler, lisansüstü öğrenciler ve bilgisayar bilimleri ile bilgisayar mühendisliđi alanına ilgi duyan tüm okuyucular için yararlı bir kaynak olması dileđiyle...



# İçindekiler

Ön Söz

iii

## Bölüm 1

---

- Yapay Zekâ Destekli Modern Web Uygulamaları: Mimari Yaklaşımlar ve Dağıtım Stratejileri 1  
*Mert Yağcıoğlu*

## Bölüm 2

---

- Perakendede Yapay Zeka Destekli Stok Yönetimi: YOLOv4-tiny ile Ürün Tespiti ve Envanter Otomasyonu 21  
*İrem Şenyer Yapıcı*

## Bölüm 3

---

- Tarımsal İzlemede İHA Tabanlı RGB Görüntülerin Değerlendirilmesi: Farklı Filtreleme Stratejilerinin Nesne Ayrıştırma Başarımı Üzerine Karşılaştırmalı Bir Analiz 35  
*Nazan Kemaloğlu Alagöz*

## Bölüm 4

---

- İğnecikli Sinir Ağları ile Görüntü İşleme: Temel Kavramlar, Uygulamalar ve Gelecek Perspektifleri 59  
*Rukiye Uzun Arslan*

## Bölüm 5

---

- Steganografi Alanındaki Araştırma Eğilimleri: Ulusal Literatür Üzerine Tematik Bir İnceleme 69  
*Mürsel Ozan İncetaş*  
*Murat Meriçelli*

## Bölüm 6

---

Çok Modlu Yapay Zekâda Görsel Dil Modelleri: Mimari Temeller ve Sektörel Uygulamalar 91

*İrem Cakcak*

*Burhan Duman*

## Bölüm 7

---

Yapay Zekâ Tabanlı Sentetik Hasta Modellemesi ile Karar Destek Sistemlerinin Geliştirilmesi 165

*Utku Köse*

*Emine Betül Sürücü*

## Bölüm 8

---

Bilgisayar Bilimleri ve Mühendisliğinde Dönüşüm: Üretken Yapay Zekâ, Güvenli Sistemler ve Ölçeklenebilir Hesaplama 183

*Furkan Atlan*

## Bölüm 9

---

Histogram Tabanlı Çok Seviyeli Eşikleme Tekniklerinin Renk Nicemleme Üzerindeki Etkileri 205

*Mahmut Kılıçaslan*

# Yapay Zekâ Destekli Modern Web Uygulamaları: Mimari Yaklaşımlar ve Dağıtım Stratejileri

Mert Yağcıoğlu<sup>1</sup>

## Özet

Yapay zekanın (YZ) web geliştirmeye entegrasyonu, modern web uygulamalarının tasarımını, uygulamasını ve dağıtımını yeniden şekillendiriyor. Yalnızca izole kodlama araçlarıyla sınırlı kalmak yerine, YZ giderek kullanıcı arayüzü oluşturma, arka uç servis oluşturma, otomatik test, güvenlik analizi ve sürekli dağıtım dahil olmak üzere tüm geliştirme yaşam döngüsünü destekliyor. Buna paralel olarak, modern web sistemleri bulut platformlarını, uç ortamları ve tarayıcılar içindeki istemci tarafı yürütmeyi kapsayan dağıtılmış mimarilere doğru evriliyor. Bu bölüm, YZ teknolojilerinin geliştirme aşamaları ve hesaplama katmanları boyunca nasıl konumlandırıldığını inceleyerek, YZ destekli modern web uygulamalarına kapsamlı bir mimari bakış açısı sunmaktadır. Yaşam döngüsü aşamasına, mimari katmana ve entegrasyon stratejisine dayalı olarak YZ destekli web geliştirme yaklaşımlarını sistematik olarak sınıflandırmak için kavramsal bir sınıflandırma çerçevesi tanıtılmaktadır. Bu çerçeveye dayanarak, insan-YZ iş birliğinin yinelemeli ve ölçeklenebilir geliştirme süreçlerini nasıl destekleyebileceğini gösteren uçtan uca YZ odaklı bir web geliştirme hattı önerilmektedir. Ek olarak, geleneksel web geliştirme uygulamaları ile YZ destekli yaklaşımlar arasında niteliksel bir karşılaştırma sunularak, otomasyon potansiyeli, geliştirme verimliliği, güvenlik entegrasyonu ve mimari esneklik açısından farklılıklar vurgulanmaktadır. Analiz, yapay zekanın geliştiricilerin yerini almak yerine, onları destekleyen bir katman görevi gördüğünü ve kritik tasarım kararlarında insan gözetimini korurken verimliliği artırdığını göstermektedir. Genel olarak, bu bölüm, ortaya çıkan yapay zeka tabanlı web sistemlerini ve bunların mimari etkilerini anlamak için yapılandırılmış ve teknolojiden bağımsız bir temel sunmaktadır.

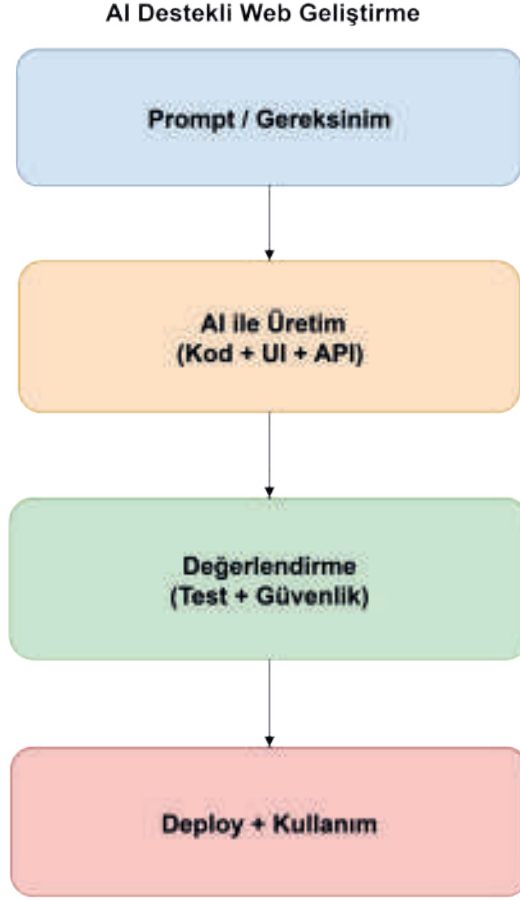
1 Dr. Öğr. Üyesi, İstanbul Arel Üniversitesi Mühendislik Fakültesi Bilgisayar Mühendisliği Bölümü, mertyaagcioglu@arel.edu.tr, ORCID ID: 0000-0001-6493-6447

## 1. Giriş

Son yıllarda web uygulamaları, yalnızca statik içerik sunan yapılardan kullanıcı davranışlarına uyum sağlayabilen, ölçeklenebilir ve yüksek etkileşimli sistemlere dönüşmüştür. Bu dönüşüm süreci, frontend ve backend bileşenlerinin karmaşıklığını artırırken, yazılım geliştirme süreçlerinin daha kısa sürede ve daha yüksek kaliteyle tamamlanmasını zorunlu hâle getirmiştir. Bu bağlamda, yapay zekâ (YZ) tabanlı araçlar, web geliştirme süreçlerinin farklı aşamalarında giderek daha etkin bir şekilde kullanılmaya başlanmıştır.

Günümüzde yapay zekâ destekli web geliştirme araçları; kod üretimi, kullanıcı arayüzü tasarımı, hata ayıklama, otomatik test oluşturma ve güvenlik açıklarının tespiti gibi pek çok alanda geliştiricilere destek sunmaktadır. Özellikle büyük dil modelleri (Large Language Models – LLM) tabanlı sistemler, geliştiricilerin doğal dil girdileriyle fonksiyonel kod parçaları, kullanıcı arayüzü bileşenleri ve API tanımları oluşturabilmesine imkân tanımaktadır (Zou et al., 2025). Bu durum, yazılım geliştirme yaşam döngüsünde insan–makine etkileşimini yeni bir boyuta taşımaktadır.

Bununla birlikte, mevcut çalışmaların büyük bir kısmı yapay zekâ destekli araçları belirli bir geliştirme aşaması veya tekil bir teknoloji üzerinden ele almakta; bulut altyapıları, uç (edge) sistemler ve istemci tarafı (client-side) uygulamalar arasındaki bütüncül mimari entegrasyonu yeterince irdelememektedir. Oysa modern web uygulamaları, yalnızca sunucu tarafında çalışan yapay zekâ servislerinden değil, aynı zamanda tarayıcı içinde çalışan istemci tarafı modellerden ve gecikmeyi azaltmaya yönelik uç bilişim çözümlerinden de faydalanmaktadır.



*řekil 1. AI destekli web geliřtirme*

Bulut tabanlı yapay zekâ servisleri, yüksek hesaplama gücü ve ölçeklenebilirlik avantajları sunarken; uç biliřim ve istemci tarafı yapay zekâ yaklaşımları, düşük gecikme, veri gizliliđi ve çevrimdışı çalışma gibi kritik kazanımlar sağlamaktadır (Bulut, 2025). Bu çok katmanlı yapı, web geliřtirme süreçlerinde kullanılan yapay zekâ araçlarının yalnızca fonksiyonel deđil, aynı zamanda mimari açıdan da deđerlendirilmesini gerekli kılmaktadır. Yapay zekâ destekli web geliřtirme yaklaşımlarının, bulut, uç ve istemci katmanları arasında nasıl konumlandığını ve bu katmanlar arasındaki etkileşimi özetleyen genel görünüm řekil 1’de sunulmaktadır. řekil, yapay zekânın yalnızca sunucu tarafında çalışan bir bileřen olmadığını; tarayıcı içi modeller ve uç biliřim çözümleriyle birlikte çok katmanlı bir yapı içerisinde ele alınması gerektiğini vurgulamaktadır.

Bu çalışmada, web geliştirmede kullanılan yapay zekâ araçları ve yaklaşımları; bulut, uç ve istemci katmanlarını kapsayan bütüncül bir bakış açısıyla ele alınmaktadır. Çalışmanın temel amacı, yapay zekâ destekli web geliştirme süreçlerinde kullanılan mimarileri, frameworkleri ve entegrasyon modellerini sistematik olarak incelemek ve bu yaklaşımların modern web uygulamalarına olan katkılarını ortaya koymaktır. Ayrıca, geliştiriciler için buluttan mobil tarayıcıya kadar uzanan yapay zekâ destekli bir web geliştirme hattı (pipeline) önerilmekte ve bu hattın potansiyel avantajları tartışılmaktadır.

## 2. Türkiye’de Yapay Zekâ Destekli Web Geliştirme: Mevcut Eğilimler ve Odak Noktaları

Yapay zekâ (YZ) destekli araçların yazılım geliştirme süreçlerine entegrasyonu, son yıllarda hem akademiye hem de endüstride hızla artan bir ilgi alanı hâline gelmiştir. Türkiye’deki çalışmalar incelendiğinde, bu ilginin özellikle kodsuz/az kodlu (no-code/low-code) platformlar, üretken yapay zekâ ile kod üretimi, geliştirme maliyeti ve verimlilik etkileri ile YZ destekli test ve kalite değerlendirme başlıklarında yoğunlaştığı görülmektedir (Koç et al., 2025). Bu bölümün amacı, Türkiye’de yapay zekâ destekli web geliştirme alanında öne çıkan akademik eğilimleri ve çalışma odaklarını özetleyerek, çalışmanın konumlandığı bağlamı ortaya koymaktır.

### 2.1. Kodsuz/Az Kodlu ve Yapay Zekâ Destekli Web Geliştirme

Türkiye’de yapay zekâ destekli web geliştirme çalışmalarının önemli bir bölümü, kodsuz ve az kodlu platformlar etrafında şekillenmektedir. Bu çalışmalar, yapay zekâ ile güçlendirilmiş geliştirme ortamlarının yazılım üretim süresini kısaltma, maliyetleri düşürme ve teknik uzmanlık gereksinimini azaltma potansiyeline odaklanmaktadır. Özellikle kurumsal dijital dönüşüm süreçleri bağlamında ele alınan bu yaklaşımlar, web uygulamalarının daha hızlı hayata geçirilmesini mümkün kılan araç merkezli çözümler sunmaktadır.

Bu eksenindeki çalışmalar, yapay zekânın web geliştirmede bir hızlandırıcı ve kolaylaştırıcı rol üstlenebileceğini göstermektedir (Koç et al., 2025). Ancak bu literatürde, çoğunlukla araçların sunduğu kazanımlar ön plana çıkarılmakta; bu araçların bulut, uç veya istemci katmanlarında nasıl konumlandırıldığına ya da farklı mimari bileşenlerle nasıl bütünleştiğine dair tartışmalar sınırlı kalmaktadır.

### 2.2. Üretken Yapay Zekâ ve Kod Üretimi Perspektifi

Üretken yapay zekâ sistemlerinin kod üretimi yetenekleri, Türkiye’deki akademik çalışmalarda da geniş yer bulmaktadır. Büyük dil modellerinin

farklı programlama dillerinde kod üretme, mevcut kodu açıklama ve hata ayıklama gibi yetkinlikleri, yazılım geliştirme süreçlerini dönüştüren önemli bir unsur olarak ele alınmaktadır (Tatlısu et al., 2025). Bu çalışmalar, genellikle üretken yapay zekâ araçlarının çıktılarının doğruluğu, kod kalitesi ve geliştirici verimliliği üzerindeki etkilerini incelemektedir.

Web geliştirme bağlamında değerlendirildiğinde, bu yaklaşımlar özellikle frontend bileşen üretimi, backend servis tanımları ve API geliştirme gibi alanlarda potansiyel faydalar sunmaktadır (Akbulut, 2025). Bununla birlikte, mevcut çalışmaların büyük bir kısmı üretken yapay zekânın genel kod üretim kabiliyetlerine odaklanmakta; web uygulamalarına özgü mimari yapıların (örneğin bileşen tabanlı frontend mimarileri veya dağıtık servis yapıları) bu araçlarla nasıl bütünleştiğini sınırlı düzeyde ele almaktadır.

### 2.3. İnsan–Yapay Zeka Etkileşimi ve Prompt Tabanlı Yaklaşımlar

Türkiye’de yürütülen bazı çalışmalar, yapay zekâ sistemleriyle etkileşim biçimlerini inceleyerek “prompt” kavramının çıktı kalitesi üzerindeki belirleyici rolünü vurgulamaktadır. Her ne kadar bu çalışmalar doğrudan web geliştirme odağında olmasa da elde edilen bulgular yazılım geliştirme süreçleri açısından da önemli çıkarımlar sunmaktadır. Üretken yapay zekâdan elde edilen çıktının niteliği, geliştiricinin sisteme sunduğu girdilerin açıklığına, bağlamına ve yinelemeli etkileşim döngülerine güçlü biçimde bağlıdır (Aktaş, 2025).

Web geliştirme süreçleri açısından bu durum, kullanıcı arayüzü üretimi, kod üretimi, test senaryosu oluşturma ve güvenlik analizi gibi adımların, insan denetimi altında iteratif bir biçimde yürütülmesi gerektiğini göstermektedir. Bu bağlamda, Türkiye’deki literatür insan–yapay zekâ iş birliğinin önemini dolaylı biçimde ortaya koymakta; tamamen otonom sistemlerden ziyade geliştiriciyi merkeze alan yaklaşımların öne çıktığını göstermektedir.

### 2.4. Test, Kalite Güvencesi ve DevOps Boyutu

Yapay zekânın yazılım test süreçlerinde kullanımı, Türkiye’deki çalışmalarda daha çok test otomasyonu ve kalite güvencesi çerçevesinde ele alınmaktadır. Bu çalışmalar, yapay zekâ tabanlı test yaklaşımlarının hata tespiti, test kapsamının genişletilmesi ve ölçeklenebilirlik açısından sunduğu avantajlara odaklanmaktadır (Ağdaş, 2025). Web geliştirme özelinde değerlendirildiğinde, bu yaklaşım özellikle üretken yapay zekâ ile üretilen kodların doğrulanması açısından kritik bir tamamlayıcı rol üstlenmektedir.

Bununla birlikte, mevcut literatürde yapay zekâ destekli test ve DevOps yaklaşımlarının, web geliştirmeye özgü uçtan uca CI/CD hatları içerisinde nasıl konumlandırılacağına dair bütüncül modeller sınırlı düzeyde ele alınmaktadır.

(Kaya, 2025). Bu durum, test, güvenlik ve dağıtım süreçlerinin yapay zekâ ile ele alındığı entegre geliştirme hatlarına yönelik ihtiyacı ortaya koymaktadır.

## 2.5. Genel Değerlendirme ve Çalışmanın Konumu

Türkiye'deki çalışmalar incelendiğinde, yapay zekâ destekli web geliştirme alanında ağırlıklı olarak araç odaklı, verimlilik merkezli ve belirli geliştirme aşamalarına yoğunlaşan yaklaşımların benimsendiği görülmektedir. Buna karşın, web geliştirme sürecinin tüm yaşam döngüsünü kapsayan; bulut, uç ve istemci katmanlarını birlikte ele alan bütüncül mimari çerçeveler sınırlı düzeyde ele alınmaktadır.

Bu kitap bölümünde sunulan yaklaşım, Türkiye'deki eğilimleri dikkate alarak; yapay zekâ destekli web geliştirmeyi yalnızca kullanılan araçlar üzerinden değil, mimari konumlandırma, entegrasyon biçimleri ve uçtan uca geliştirme süreçleri perspektifinden ele almayı amaçlamaktadır (Yavuz, 2025). Bu yönüyle bölüm, mevcut çalışmalarını tamamlayıcı bir bağlam sunmakta ve okuyucuya alanın genel görünümünü kavramsal bir çerçeve içerisinde aktarmayı hedeflemektedir.

## 3. Yapay Zekâ Destekli Web Geliştirme İçin Kavramsal Sınıflandırma Çerçevesi

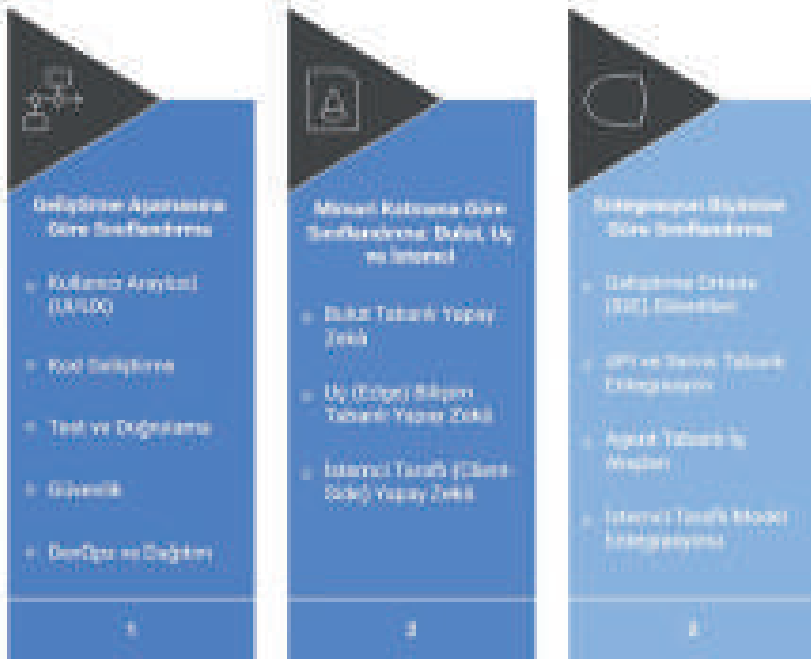
Yapay zekâ destekli web geliştirme, günümüzde çok sayıda araç, yaklaşım ve mimari çözümü kapsayan geniş ve heterojen bir alan hâline gelmiştir. Kod üretiminden kullanıcı arayüzü tasarımına, test otomasyonundan dağıtım süreçlerine kadar uzanan bu ekosistem, farklı geliştirme aşamalarında ve farklı hesaplama katmanlarında konumlanan yapay zekâ bileşenlerini içermektedir. Bu çeşitlilik, alandaki yaklaşımların yalnızca araç isimleri üzerinden ele alınmasını güçleştirmekte; daha soyut ve genellenebilir bir bakış açısını gerekli kılmaktadır.

Bu bölümde, yapay zekâ destekli web geliştirme yaklaşımlarını sistematik biçimde ele alabilmek amacıyla kavramsal bir sınıflandırma çerçevesi sunulmaktadır. Sunulan çerçeve, belirli bir araca ya da teknolojiye odaklanmak yerine, yapay zekânın web geliştirme sürecindeki rolünü, konumunu ve entegrasyon biçimini esas almaktadır. Böylece okuyucuya, hızla değişen araç ekosisteminden bağımsız olarak uzun vadede geçerliliğini koruyabilecek bir değerlendirme zemini sunulması hedeflenmektedir.

Önerilen kavramsal çerçeve üç temel boyut etrafında yapılandırılmıştır:

1. Web geliştirme yaşam döngüsündeki geliştirme aşamasına göre sınıflandırma

2. Yapay zekâ bileşenlerinin konumlandığı mimari katmana göre sınıflandırma
3. Yapay zekâ araçlarının geliştirme sürecine entegrasyon biçimine göre sınıflandırma



*Şekil 2. Yapay zekâ destekli web geliştirme için sınıflandırma çerçevesi*

Yapay zekâ destekli web geliştirme yaklaşımlarının sistematik biçimde ele alınabilmesi amacıyla oluşturulan kavramsal sınıflandırma çerçevesi Şekil 2’de görsel olarak özetlenmektedir. Bu şekil, yapay zekânın web geliştirme süreçlerindeki rolünü; geliştirme aşaması, mimari katman ve entegrasyon biçimi olmak üzere üç tamamlayıcı boyutta ele alarak, farklı yaklaşımların ortak bir referans çerçevesi içerisinde değerlendirilmesini mümkün kılmaktadır.

Şekil 2’de sunulan bu çok boyutlu yapı, yapay zekâ destekli web geliştirme yaklaşımlarının yalnızca ne yaptığı üzerinden değil, nerede ve nasıl konumlandığı üzerinden de analiz edilmesini sağlayan bütüncül bir bakış açısı sunmaktadır.

### **3.1. Geliştirme Aşamasına Göre Sınıflandırma**

Web uygulama geliştirme süreci, kullanıcı arayüzü tasarımından başlayarak kodlama, test, güvenlik ve dağıtım gibi birbirini izleyen ancak çoğu zaman iç içe geçen aşamalardan oluşmaktadır (Hancı & Gülseçen, 2023). Yapay zekâ

destekli araçlar, bu aşamaların her birinde farklı sorumluluklar üstlenerek geliştirme sürecinin çeşitli noktalarında değer üretmektedir.

### **3.1.1. Kullanıcı Arayüzü (UI/UX)**

Bu aşamada yapay zekâ, doğal dil girdilerinden kullanıcı arayüzü bileşenleri üretme, taslak arayüzler oluşturma ve erişilebilirlik iyileştirmeleri önerme gibi görevlerde kullanılmaktadır (Jayaram et al., 2024). Özellikle bileşen tabanlı modern web frameworkleri bağlamında, tekrar eden arayüz yapılarını otomatik olarak oluşturabilen yapay zekâ sistemleri, geliştirme süresini önemli ölçüde azaltabilmektedir.

### **3.1.2. Kod Geliştirme**

Kodlama aşamasında yapay zekâ; frontend ve backend bileşenlerinin üretilmesi, mevcut kodun açıklanması, yeniden düzenlenmesi ve hata ayıklanması gibi görevleri üstlenmektedir (Kalkan & Arslan, 2025). Bu bağlamda yapay zekâ, geliştiricinin yerini alan bir unsurdan ziyade, yazılım üretim sürecini hızlandıran ve destekleyen bir yardımcı olarak konumlanmaktadır.

### **3.1.2. Test ve Doğrulama**

Yapay zekâ destekli test araçları, birim testlerinden uçtan uca test senaryolarına kadar geniş bir yelpazede otomatik test üretimi sağlayabilmektedir (Aslandođdu et al., 2022). Özellikle yapay zekâ tarafından üretilen kodların doğrulanması açısından bu aşama kritik bir tamamlayıcı rol oynamaktadır.

### **3.1.3. Güvenlik**

Web güvenliği bağlamında yapay zekâ; kod içerisindeki potansiyel güvenlik açıklarını tespit etme, zafiyet analizi ve güvenli kodlama önerileri sunma amacıyla kullanılmaktadır (Çiçek, 2025). Bu tür araçlar, güvenliği geliştirme sürecinin erken safhalarına entegre etmeyi mümkün kılmaktadır.

### **3.1.4. DevOps ve Dağıtım**

DevOps aşamasında yapay zekâ; CI/CD süreçlerinin otomasyonu, yapılandırma hatalarının tespiti ve dağıtım optimizasyonu gibi görevlerde kullanılmaktadır (Karagöz & Türhal, 2025). Bu durum, web uygulamalarının daha hızlı ve daha güvenilir şekilde üretim ortamına alınmasına katkı sağlamaktadır.

## 3.2. Mimari Katmana Göre Sınıflandırma: Bulut, Uç ve İstemci

Yapay zekâ destekli web geliştirme araçları, çalıştıkları hesaplama ortamına bağlı olarak farklı mimari özellikler sergilemektedir. Bu bağlamda araçlar; bulut, uç (edge) ve istemci (client) tabanlı yaklaşımlar altında sınıflandırılabilir (Ali & Zeebaree, 2025).

### 3.2.1. Bulut Tabanlı Yapay Zekâ

Bulut katmanı, yüksek hesaplama gücü ve ölçeklenebilirlik gerektiren yapay zekâ işlemleri için doğal bir ortam sunmaktadır. Büyük modellerin eğitimi ve karmaşık çıkarım işlemleri çoğunlukla bu katmanda gerçekleştirilmektedir. Ancak gecikme ve veri gizliliği gibi konular, bulut tabanlı çözümlerin dikkatle değerlendirilmesini gerektirmektedir.

### 3.2.2. Uç (Edge) Bilişim Tabanlı Yapay Zekâ

Uç bilişim yaklaşımı, yapay zekâ işlemlerinin kullanıcıya daha yakın noktalarda gerçekleştirilmesini sağlayarak gecikmeyi azaltmayı hedeflemektedir. Etkileşimli web uygulamaları açısından bu yaklaşım, performans ve kullanıcı deneyimi bakımından önemli avantajlar sunmaktadır.

### 3.2.3. İstemci Taraflı (Client-Side) Yapay Zekâ

İstemci taraflı yapay zekâ, doğrudan tarayıcı içerisinde çalışan modelleri kapsamaktadır. Veri gizliliği ve çevrimdışı çalışma gibi avantajlar sunmasına karşın, donanım kısıtları nedeniyle model karmaşıklığı sınırlı kalmaktadır. Buna rağmen, kişiselleştirme ve anlık etkileşim gerektiren senaryolarda giderek daha fazla önem kazanmaktadır.

## 3.3. Entegrasyon Biçimine Göre Sınıflandırma

Yapay zekâ araçlarının web geliştirme süreçlerine nasıl entegre edildiği, geliştirici deneyimini ve sistem esnekliğini doğrudan etkilemektedir. Bu bağlamda farklı entegrasyon biçimleri öne çıkmaktadır.

### 3.3.1. Geliştirme Ortamı (IDE) Eklentileri

Bu yaklaşımda yapay zekâ araçları, geliştiricinin kullandığı IDE veya kod editörü içerisine eklenti olarak entegre edilir. Anlık kod önerileri ve hata bildirimleri sunarak geliştirme sürecini destekler.

### 3.3.2. API ve Servis Tabanlı Entegrasyon

YZ yetenekleri, harici servisler aracılığıyla API üzerinden sunulur. Bu yapı, farklı web uygulamaları ve araç zincirleriyle kolay entegrasyon imkânı sağlar.

### 3.3.3. Agent Tabanlı İş Akışları

Agent tabanlı yaklaşımlar, birden fazla yapay zekâ bileşeninin belirli görevleri otonom veya yarı-otonom biçimde üstlendiği iş akışlarını ifade etmektedir. Bu yaklaşım, uçtan uca web geliştirme hatlarının oluşturulmasına olanak tanımaktadır.

### 3.3.4. İstemci Taraflı Model Entegrasyonu

Bu entegrasyon biçimi, yapay zekâ modellerinin doğrudan web uygulamasının istemci tarafına gömülmesini kapsamaktadır ve özellikle kişiselleştirme odaklı uygulamalarda tercih edilmektedir.

## 3.4. Genel Değerlendirme

Bu bölümde sunulan kavramsal sınıflandırma çerçevesi, yapay zekâ destekli web geliştirme yaklaşımlarını araç isimlerinden bağımsız, zamanla geçerliliğini koruyabilecek bir bakış açısıyla ele almaktadır. Geliştirme aşamaları, mimari katmanlar ve entegrasyon biçimleri birlikte değerlendirildiğinde, yapay zekânın web geliştirme süreçlerindeki rolü daha net ve sistematik bir biçimde anlaşılabilir.

Bu çerçeve, sonraki bölümlerde ele alınacak olan yapay zekâ destekli web geliştirme pipeline'ı için kavramsal bir temel oluşturmakta; okuyucuya farklı yaklaşımları karşılaştırma ve bağlamlandırma imkânı sunmaktadır.

## 4. Yapay Zekâ Destekli Web Geliştirme Pipeline'ı

Yapay zekâ destekli web geliştirme yaklaşımlarının pratikte değer üretebilmesi, bu yaklaşımların dağıtım araçlar ve tekil çözümler olmaktan çıkarılarak uçtan uca bir geliştirme hattı (pipeline) içerisinde ele alınmasını gerektirmektedir. Aksi takdirde, yapay zekâ yalnızca belirli aşamalarda kullanılan yardımcı bir araç olarak kalmakta; geliştirme sürecinin bütününe nüfuz edememektedir.

Bu bölümde, yapay zekânın web geliştirme yaşam döngüsünün farklı aşamalarına sistematik biçimde entegre edildiği kavramsal bir web geliştirme pipeline'ı ele alınmaktadır. Sunulan yapı, belirli bir teknolojiye ya da araca bağımlı olmaksızın; farklı ölçeklerdeki web projelerine uyarlanabilecek genel bir yol haritası sunmayı amaçlamaktadır. Pipeline, hem insan denetimini merkeze alan yarı-otonom senaryoları hem de ileri düzey otomasyon yaklaşımlarını destekleyecek şekilde kurgulanmıştır.

#### 4.1. Pipeline'ın Genel Yapısı ve Tasarım İlkeleri

Yapay zekâ destekli web geliştirme pipeline'ı, klasik yazılım geliştirme süreçleriyle benzer aşamalar içerirse de bu aşamaların yapay zekâ tabanlı geri besleme döngüleri ile birbirine bağlanması bakımından farklılaşmaktadır. Pipeline, doğrusal bir akıştan ziyade, gerektiğinde önceki aşamalara geri dönülebilen iteratif ve esnek bir yapı sunmaktadır.

Bu yaklaşımın temel tasarım ilkeleri şu şekilde özetlenebilir:

- Geliştiriciyi merkeze alan insan–makine iş birliği
- Otomasyon ile denetim arasında dengeli bir yapı
- Bulut, uç ve istemci katmanlarını kapsayan dağıtık mimari uyum
- Tekrarlanabilir ve genişletilebilir geliştirme adımları

Bu ilkeler doğrultusunda ele alınan pipeline, web geliştirme sürecini yedi temel aşamada incelemektedir. Yapay zekâ destekli web geliştirme pipeline'ının temel aşamaları ve bu aşamalar arasındaki geri besleme ilişkileri Şekil 3'te gösterilmektedir. Şekil, geliştirme sürecinin doğrusal bir akıştan ziyade, yapay zekâ destekli geri besleme döngüleriyle zenginleştirilmiş iteratif bir yapı sunduğunu görsel olarak ortaya koymaktadır.



Şekil 3. Yapay zekâ destekli web geliştirme hattı

## 4.2. Gereksinim Analizi ve Doğal Dil Tabanlı Başlatma

Pipeline'in ilk aşaması, sistem gereksinimlerinin doğal dil kullanılarak ifade edilmesine dayanmaktadır. Bu aşamada geliştiriciler veya paydaşlar, teknik ayrıntılara girmeden uygulamanın temel işlevlerini, kullanıcı senaryolarını ve beklentilerini tanımlamaktadır. Yapay zekâ, bu girdileri analiz ederek işlevsel gereksinimler, temel bileşenler ve olası sistem mimarisi hakkında ön çıkarımlar üretmektedir.

Bu yaklaşım, teknik olmayan paydaşların da geliştirme sürecine erken aşamada katkı sağlamasına olanak tanırken; geliştirici açısından da soyut gereksinimlerin daha hızlı biçimde somut yapılarla dönüştürülmesini mümkün kılmaktadır.

## 4.3. Yapay Zekâ Destekli UI ve Frontend Üretimi

İkinci aşamada, tanımlanan gereksinimlere bağlı olarak kullanıcı arayüzü bileşenleri ve frontend yapıları ele alınmaktadır. Yapay zekâ, doğal dilde ifade edilen ihtiyaçlardan yola çıkarak sayfa düzenleri, arayüz bileşenleri ve temel etkileşim senaryoları üretebilmektedir.

Bu aşamada üretilen çıktılar, nihai ürün olarak değil; geliştiricinin üzerinde düzenleme yapabileceği taslak ve başlangıç noktaları olarak değerlendirilmelidir. Böylece tekrar eden arayüz tasarımlarında önemli ölçüde zaman kazanımı sağlanırken, tasarım üzerindeki insan denetimi korunmaktadır.

## 4.4. Backend ve API Katmanının Oluşturulması

Frontend bileşenleriyle uyumlu olacak şekilde, pipeline'in bir sonraki aşamasında backend servisleri ve API yapıları ele alınmaktadır. Yapay zekâ destekli araçlar, veri modellerinin tanımlanması, doğrulama kurallarının oluşturulması ve temel iş mantığının yapılandırılması gibi görevlerde geliştiriciye destek sunmaktadır.

Bu yaklaşım, frontend ve backend arasında sıkça karşılaşılan uyumsuzlukların azaltılmasına katkı sağlamakta; servis odaklı ve modüler mimarilerin daha hızlı biçimde hayata geçirilmesini mümkün kılmaktadır.

## 4.5. Otomatik Test ve Doğrulama Süreçleri

Yapay zekâ destekli web geliştirme pipeline'ında test ve doğrulama, sonradan eklenen bir adım değil; geliştirme sürecinin ayrılmaz bir parçası olarak ele alınmaktadır. Bu aşamada yapay zekâ, üretilen kod yapısına ve tanımlanan kullanıcı senaryolarına bağlı olarak birim testleri, entegrasyon testleri ve uçtan uca test senaryoları oluşturabilmektedir.

Özellikle yapay zekâ tarafından üretilen kodların güvenilirliğinin sağlanması açısından bu aşama kritik öneme sahiptir. Otomatik test üretimi, geliştiricinin test kapsamını genişletmesine yardımcı olurken, insan denetimi altında yapılan doğrulamalar sistem kalitesini desteklemektedir.

#### 4.6. Güvenlik Analizi ve Kalite Kontrol

Pipeline'in bu aşamasında, kod ve yapılandırmalar güvenlik ve kalite açısından ele alınmaktadır. Yapay zekâ tabanlı analiz araçları, potansiyel güvenlik açıklarını, hatalı yapılandırmaları ve iyi uygulamalardan sapmaları tespit etmeye yönelik öneriler sunmaktadır.

Bu yaklaşım, güvenliğin geliştirme sürecinin geç aşamalarına bırakılmasını önleyerek, daha sürdürülebilir ve güvenli web uygulamalarının geliştirilmesine katkı sağlamaktadır.

#### 4.7. CI/CD Süreçleri ve Dağıtım

Dağıtım aşamasında yapay zekâ, sürekli entegrasyon ve sürekli dağıtım süreçlerinin otomasyonunu desteklemektedir. Yapılandırma dosyalarının oluşturulması, dağıtım stratejilerinin seçilmesi ve kaynak kullanımının optimize edilmesi gibi görevler bu aşamada ele alınmaktadır.

Bu sayede web uygulamaları, farklı ortamlar arasında daha tutarlı ve hataya daha az açık biçimde taşınabilmektedir. Pipeline'in bu aşaması, özellikle ölçeklenebilir ve sürekli güncellenen web uygulamaları açısından kritik bir rol oynamaktadır.

#### 4.8. İstemci Tarafı Yapay Zekâ ve Geri Besleme Döngüsü

Pipeline'in son aşamasında, istemci taraflı yapay zekâ bileşenleri web uygulamasına entegre edilmektedir. Kişiselleştirme, öneri sistemleri veya etkileşimli özellikler bu aşamada devreye alınmaktadır. Kullanıcı etkileşimlerinden elde edilen geri besleme, pipeline'in önceki aşamalarına aktarılmakta ve sistemin zaman içinde iyileştirilmesine katkı sağlamaktadır.

Özellikle istemci taraflı yapay zekâ bileşenlerinden elde edilen kullanıcı geri bildirimlerinin, pipeline'in önceki aşamalarına aktarılması, Şekil 3'te gösterildiği üzere sistemin zaman içinde kendini iyileştirebilen bir yapıya verilmesini mümkün kılmaktadır.

#### 4.9. Bölümün Genel Değerlendirmesi

Bu bölümde ele alınan yapay zekâ destekli web geliştirme pipeline'ı, klasik yazılım geliştirme süreçlerini ikame etmeyi değil; bu süreçleri daha esnek,

daha hızlı ve daha tutarlı hâle getirmeyi amaçlayan kavramsal bir yol haritası sunmaktadır. Pipeline'ın temel gücü, yapay zekâ ile otomasyonu artırırken geliştirici denetimini koruyan dengeli yapısında yatmaktadır.

Sunulan yaklaşım, önceki bölümde tanımlanan kavramsal sınıflandırma çerçevesini somut bir geliştirme hattı ile ilişkilendirmekte ve sonraki bölümlerde ele alınacak karşılaştırmalı değerlendirmeler için doğal bir zemin oluşturmaktadır.

## **5. Klasik Web Geliştirme ve Yapay Zeka Destekli Web Geliştiriminin Karşılaştırmalı Değerlendirilmesi**

Web uygulamalarının geliştirilme biçimi, son yıllarda yapay zekâ tabanlı araçların yaygınlaşmasıyla birlikte önemli bir dönüşüm sürecine girmiştir. Geleneksel web geliştirme yaklaşımı, uzun yıllar boyunca geliştiricinin bilgi birikimi ve manuel üretim süreçleri üzerine inşa edilmiş; tasarım, kodlama, test ve dağıtım aşamaları büyük ölçüde insan emeğiyle yürütülmüştür. Yapay zekâ destekli yaklaşımlar ise bu sürece otomasyon, öngörü ve karar destek mekanizmaları ekleyerek geliştirme pratiklerini yeniden şekillendirmektedir.

Bu bölümde, klasik web geliştirme yaklaşımı ile yapay zekâ destekli web geliştirme yaklaşımı üstünlük iddiası olmaksızın, kavramsal ve niteliksel bir bakış açısıyla karşılaştırılmaktadır. Amaç, her iki yaklaşımın güçlü ve sınırlı yönlerini ortaya koyarak okuyucuya bağlamsal bir değerlendirme sunmaktır.

### **5.1. Geliştirme Süreci ve Zaman Yönetimi**

Klasik web geliştirme sürecinde, gereksinimlerin analiz edilmesi, kullanıcı arayüzünün tasarlanması ve kodun üretilmesi aşamaları genellikle ardışık ve manuel bir biçimde ilerlemektedir. Bu yapı, özellikle proje kapsamı genişledikçe geliştirme süresinin uzamasına neden olabilmektedir.

Yapay zekâ destekli web geliştirme yaklaşımı ise erken prototipleme ve tekrar eden görevlerin otomasyonu sayesinde geliştirme sürecini daha esnek hâle getirmektedir. Doğal dil girdilerinden taslak arayüzler veya kod bileşenleri üretilebilmesi, sürecin başlangıç aşamalarında önemli bir hızlanma sağlamaktadır. Bununla birlikte, bu hız kazanımı insan denetimi ile dengelenmediğinde, ilerleyen aşamalarda yeniden düzenleme gereksinimleri ortaya çıkabilmektedir.

### **5.2. Kod Üretimi ve Yeniden Kullanılabilirlik**

Geleneksel yaklaşımda kod üretimi, geliştiricinin deneyimine ve bilgi birikimine doğrudan bağlıdır. Bu durum, yüksek kontrol ve özelleştirilebilirlik

sağlarken, tekrar eden kod parçalarının manuel olarak üretilmesi nedeniyle verimlilik kaybına yol açabilmektedir.

Yapay zekâ destekli yaklaşımlar, kod üretiminde yarı-otonom bir yapı sunarak tekrar eden kalıpların hızlı biçimde oluşturulmasını mümkün kılmaktadır. Bu durum, özellikle standartlaştırılmış web bileşenlerinde ve servis yapılarında yeniden kullanılabilirliği artırmaktadır. Ancak üretilen kodun kalitesinin değerlendirilmesi ve bağlamsal uygunluğunun sağlanması, geliştirici denetimini zorunlu kılmaktadır.

### 5.3. Test, Doğrulama ve Hata Tespiti

Klasik web geliştirme süreçlerinde test faaliyetleri çoğu zaman geliştirme sürecinin ileri aşamalarında ele alınmakta; manuel testler ve sınırlı otomasyon çözümleri yaygın olarak kullanılmaktadır. Bu yaklaşım, hataların geç tespit edilmesine ve düzeltme maliyetlerinin artmasına neden olabilmektedir.

Yapay zekâ destekli web geliştirme yaklaşımı, test üretiminin ve doğrulama süreçlerinin geliştirme hattının erken aşamalarına entegre edilmesini teşvik etmektedir. Yapay zekâ tabanlı test senaryoları, kod yapısına ve kullanıcı akışlarına bağlı olarak daha geniş bir test kapsamı sunabilmektedir. Bununla birlikte, otomatik testlerin de bağlamdan bağımsız olarak mutlak doğruluk sağladığı varsayımı, dikkatle ele alınması gereken bir konudur.

### 5.4. Güvenlik ve Kalite Farkındalığı

Geleneksel yaklaşımlarda web güvenliği çoğu zaman ayrı bir uzmanlık alanı olarak ele alınmakta ve geliştirme sürecinin geç safhalarında değerlendirilmekteydi. Bu durum, güvenlik açıklarının sistem mimarisine yerleşmesine neden olabilmektedir.

Yapay zekâ destekli web geliştirme, güvenlik analizlerinin geliştirme sürecine daha erken entegre edilmesine imkân tanımaktadır. Kod inceleme, zafiyet analizi ve güvenli kodlama önerileri sunan yapay zekâ araçları, geliştirici farkındalığını artırmaktadır. Ancak bu araçların sunduğu çıktılar, kesin karar mekanizmaları olarak değil; destekleyici öneriler olarak değerlendirilmelidir.

### 5.5. Mimari Esneklik ve Dağıtık Sistem Uyumu

Klasik web geliştirme yaklaşımları, tarihsel olarak sunucu merkezli mimariler üzerine inşa edilmiştir. Bu yapı, belirli ölçeklere kadar yeterli olmakla birlikte, modern web uygulamalarının dağıtık ve düşük gecikme gereksinimlerini her zaman karşılayamamaktadır.

Yapay zekâ destekli web geliştirme yaklaşımı, bulut, uç ve istemci katmanlarını kapsayan dağıtık mimarilere daha doğal bir uyum sağlamaktadır. Yapay zekâ bileşenlerinin farklı katmanlarda konumlandırılabilmesi, performans ve kullanıcı deneyimi açısından yeni olanaklar sunmaktadır. Bununla birlikte, bu esneklik mimari karmaşıklığı da beraberinde getirmekte ve dikkatli bir tasarım gerektirmektedir.

### 5.6. Geliştirici Deneyimi ve İnsan–Makine Etkileşimi

Klasik web geliştirme yaklaşımında geliştirici, sürecin merkezinde yer almakta ve tüm karar mekanizmalarını doğrudan yönetmektedir. Bu durum yüksek kontrol sağlarken, bilişsel yükü artırabilmektedir.

Yapay zekâ destekli yaklaşımlar, geliştiricinin rolünü tamamen ortadan kaldırmak yerine, karar destek sistemleri aracılığıyla yeniden tanımlamaktadır. İnsan–makine iş birliğine dayalı bu yapı, geliştiricinin yaratıcılığa ve mimari tasarıma daha fazla odaklanmasına imkân tanımaktadır. Ancak bu dönüşüm, geliştiricinin yapay zekâ çıktılarını eleştirel bir bakış açısıyla değerlendirebilme yetkinliğini de gerekli kılmaktadır.

### 5.7. Genel Değerlendirme

Klasik web geliştirme ve yapay zekâ destekli web geliştirme yaklaşımları, birbirini dışlayan değil; farklı bağlamlarda tamamlayıcı roller üstlenen iki paradigma olarak değerlendirilebilir. Yapay zekâ destekli yaklaşımlar, hız, otomasyon ve ölçeklenebilirlik açısından önemli avantajlar sunarken; insan denetimi, bağlamsal değerlendirme ve mimari kararlar hâlen geliştiricinin sorumluluğundadır.

Bu karşılaştırmalı değerlendirme, yapay zekâ destekli web geliştirme yaklaşımlarının mutlak bir çözüm değil; doğru bağlamda ve dengeli biçimde kullanıldığında değer üreten bir araç seti sunduğunu göstermektedir. Bir sonraki bölümde, bu yaklaşımların gelecekte nasıl evrilebileceği ve hangi araştırma alanlarının öne çıktığı ele alınacaktır.

## 6. Sonuç ve Gelecek Perspektifleri

Bu kitap bölümünde, yapay zekâ destekli web geliştirme alanı; kavramsal sınıflandırmalar, mimari katmanlar ve uçtan uca geliştirme süreçleri bağlamında bütüncül bir bakış açısıyla ele alınmıştır. Sunulan yaklaşım, yapay zekâyı yalnızca belirli araçlar veya kısa vadeli teknolojik eğilimler üzerinden değerlendirmek yerine, web geliştirme yaşam döngüsünün tamamına yayılan yapısal bir dönüşüm unsuru olarak konumlandırmaktadır.

Bölüm boyunca ele alınan kavramsal sınıflandırma çerçevesi, yapay zekâ destekli web geliştirme yaklaşımlarını; geliştirme aşamaları, mimari konumlandırma ve entegrasyon biçimleri açısından sistematik biçimde incelemeyi mümkün kılmıştır. Bu çerçeve, hızla değişen araç ekosistemine rağmen geçerliliğini koruyabilecek bir referans noktası sunmakta ve okuyucunun farklı yaklaşımları ortak bir dil üzerinden değerlendirmesine imkân tanımaktadır. Benzer şekilde, yapay zekâ destekli web geliştirme pipeline'ı, bu kavramsal yapıyı pratik bir yol haritası ile ilişkilendirerek, yapay zekânın geliştirme sürecine nasıl entegre edilebileceğine dair açıklayıcı bir perspektif sunmuştur.

Karşılaştırmalı değerlendirme, klasik web geliştirme ile yapay zekâ destekli yaklaşımlar arasındaki farkların mutlak üstünlükler üzerinden değil, bağlamsal uygunluk çerçevesinde ele alınması gerektiğini göstermektedir. Yapay zekâ destekli yaklaşımlar; hız, otomasyon ve ölçeklenebilirlik açısından önemli kazanımlar sağlarken, mimari kararlar, güvenlik sorumluluğu ve bağlamsal değerlendirme hâlen geliştiricinin denetiminde kalmaktadır. Bu durum, yapay zekânın web geliştirme süreçlerinde insan emeğinin yerini alan bir unsurdan ziyade, insan-makine iş birliğini güçlendiren bir destekleyici olarak ele alınmasının daha sürdürülebilir bir yaklaşım olduğunu ortaya koymaktadır.

Gelecek perspektifi açısından değerlendirildiğinde, web uygulamalarının giderek daha fazla "AI-native" bir yapıya evrileceği öngörülebilir. Yapay zekânın yalnızca geliştirme sürecinde değil, doğrudan web uygulamasının çalışma zamanında (runtime) da aktif rol üstlenmesi; kişiselleştirme, bağlamsal farkındalık ve etkileşim düzeyi yüksek uygulamaların yaygınlaşmasını beraberinde getirecektir. Bu dönüşüm, istemci tarafı yapay zekâ modelleri, uç bilişim çözümleri ve tarayıcı içi çalıştırma teknolojilerinin daha fazla önem kazanmasına yol açacaktır.

Bununla birlikte, bu evrimin beraberinde getirdiği bazı temel zorluklar da bulunmaktadır. Yapay zekâ tarafından üretilen kodun doğrulanması, güvenlik açıklarının erken tespiti, veri gizliliğinin korunması ve model bağımlılığı gibi konular, gelecekte web geliştirme alanında daha fazla tartışılacak başlıklar arasında yer alacaktır. Özellikle agent tabanlı ve yarı-otonom geliştirme iş akışlarının yaygınlaşması, etik sorumluluklar ve insan denetiminin sınırları gibi yeni araştırma alanlarını da gündeme getirmektedir.

Akademik ve endüstriyel perspektiften bakıldığında, gelecekteki çalışmaların; yapay zekâ destekli web geliştirme yaklaşımlarını gerçek dünya projeleri üzerinde uzun vadeli ve karşılaştırmalı biçimde incelemesi önem taşımaktadır. Geliştirme süresi, bakım maliyeti, güvenlik dayanıklılığı ve geliştirici deneyimi gibi ölçütlerin birlikte ele alındığı çalışmalar, bu alandaki kavramsal çerçevelerin somut etkilerini daha net biçimde ortaya koyacaktır. Ayrıca, farklı ölçeklerdeki

web uygulamaları için bulut-uç-istemci katmanlarının nasıl dengeleneceğine dair tasarım ilkeleri, önümüzdeki dönemde öne çıkan araştırma konuları arasında yer alacaktır.

Sonuç olarak, yapay zekâ destekli web geliştirme, geçici bir araç trendi olmanın ötesinde, web uygulamalarının nasıl tasarlandığını, geliştirildiğini ve sürdürüldüğünü yeniden tanımlayan yapısal bir dönüşüm sürecini temsil etmektedir. Bu kitap bölümünde sunulan kavramsal yaklaşımlar ve değerlendirmeler, okuyucuya bu dönüşümü anlamlandırma ve kendi bağlamına uyarlama konusunda rehberlik etmeyi amaçlamaktadır. Yapay zekânın olanakları ile insan uzmanlığının dengeli biçimde bir araya getirildiği yaklaşımlar, geleceğin web geliştirme pratiklerinin temelini oluşturacaktır.

## Kaynakça

- Ağdaş, M. T. (2025). İnovasyonun Görünmeyen Ekibi: Yapay Zekâ Ajanları ile Akıllı Ar-Ge. *Disiplinlerarası Yaklaşımla Ar-Ge, İnovasyon ve Sürdürülebilir Kalkınma*, 167.
- Akbulut, M. (2025). Yapay Zeka Uygulamalarının Python Kod Üretimindeki Performans Karşılaştırması. *Sinop Üniversitesi Fen Bilimleri Dergisi*, 10(1), 259-288.
- Aktaş, S. G. (2025). YAPAY ZEKÂ VE GAZETECİLİK ETKİLEŞİMİNDE PROMPT MÜHENDİSLİĞİ: CHATGPT İLE KEŞİFSEL BİR ÇALIŞMA. *Gümüşhane Üniversitesi İletişim Fakültesi Elektronik Dergisi*, 13(1), 611-644.
- Ali, C. S. M., & Zeebaree, S. R. (2025). Cloud-based web applications for enterprise systems: A review of AI and marketing innovations. *Asian Journal of Research in Computer Science*, 18(4), 427-451.
- Aslandođdu, Ö. F., Aydın, H., & Çetinkaya, A. (2022). Bulut üzerinde Tensorflow Javascript ile geliştirilen Öğretilebilir Makine (ÖM) Modelinin Doğruluk Tahmini ve Analizi: Yüz Tanıma Sistemi Uygulaması. *International Journal of Management Information Systems and Computer Science*, 6(1), 66-77.
- Bulut, C. (2025). SAĞLIK KURUMLARINDA YAPAY ZEKÂ DESTEKLİ KARAR DESTEK SİSTEMLERİNİN KULLANIMI. *Uluslararası Sağlık Yönetimi ve Stratejileri Araştırma Dergisi*, 11(1), 27-37.
- Çiçek, A. E. (2025). TÜRKİYE'NİN YAPAY ZEKÂ TABANLI SİBER GÜVENLİK STRATEJİSİ: ULUSAL GÜVENLİĞİ GÜÇLENDİRMEK VE KÜRESEL SİBER YÖNETİŞİME YÖN VERMEK. *Yönetim Bilimleri Dergisi*, 23(56), 993-1012.
- HANCI, A. K., & GÜLSEÇEN, S. (2023). DevOps Test Süreç Geliştirmede Yeni Bir Model Önerisi. *Duzce University Journal of Science & Technology*, 11(4).
- Jayaram, K., Abhishek, K. A., Krishna, A., & Rayas, K. G. (2024). Rubreak: An AI-assisted UI/UX workflow implementation. In *Computer Science Engineering* (pp. 178-184). CRC Press.
- Kalkan, B., & Arslan, R. B. (2025, September). A P11Y Based Approach To Web Accessibility Evaluation. In *2025 10th International Conference on Computer Science and Engineering (UBMK)* (pp. 1281-1286). IEEE.
- KARAGÖZ, B., & TURHAL, Ü. Ç. KURUMSAL DÜZEYDE DEVOPS KÜLTÜRÜNÜN ENTEGRASYONU VE ALTYAPI OTOMASYONUNUN MODERNİZASYONU. *yaz*, 1.
- Kaya, H. (2025). Dijital Dönüşüm Çağında Yönetim Bilişim Sistemlerinin Stratejik Evrimi ve Yenilikçi Uygulamaları. *Yönetim Bilişim Sistemleri Alanında Yenilikçi Çözümler ve*, 49.

- Koç, O., Yücedağ, İ., & Şentürk, Ü. (2025). The Impact of Artificial Intelligence Enhanced No-Code Software Development Platforms on Software Processes: A Literature Review. *Duzce University Journal of Science and Technology*, 13(1), 383-401.
- Tatlisu, E., Bağlı, H. H., Turan, A. Z., Hamurcu, A., Eser, A., & Avcı, S. K. (2025). Yapay zekâ çağında endüstriyel tasarım: Uygulamalar ve eğitim üzerine etkileri. *Öneri Dergisi*, 20(MX Yaratıcı Endüstriler Çalıştayı 2024: Yapay Zeka Çağında Yaratıcı Endüstriler Özel Sayısı), 154-188.
- Yavuz, U. (2025). Üretken Yapay Zekâ Teknolojilerinin Yazılım Geliştirici İstihdamı Üzerindeki Etkileri. *Niksar Akademik Çalışmalar Dergisi*, 2(1), 26-48.
- Zou, G., Li, P., Yang, S., Hu, S., Pang, S., & Gan, Y. (2025). LLM-enhanced service Semantic Representation and Category co-occurrence feature Augmentation for Web API recommendation. *Information Processing & Management*, 62(6), 104219.

## Perakendede Yapay Zeka Destekli Stok Yönetimi: YOLOv4-tiny ile Ürün Tespiti ve Envanter Otomasyonu

İrem Şenyer Yapıcı<sup>1</sup>

### Özet

Perakende sektörü, yüksek ürün çeşitliliği, kısa ürün yaşam döngüsü ve talep dalgalanmalarının belirginliği nedeniyle etkin stok yönetimini zorunlu kılan dinamik operasyonel alanlardan biridir. Geleneksel manuel stok takip yöntemleri; insan hatasına açıklık, düşük zaman çözünürlüğü, yüksek personel maliyeti ve kayıt güncellemelerinde gecikme gibi yapısal problemler içermektedir. Bu çalışmada, market raflarındaki ürünlerin görüntüler üzerinden gerçek zamanlı olarak tespit edilmesi ve stok bilgilerinin otomatik güncellenmesi için YOLOv4-tiny algoritmasına dayalı uçtan uca bir sistem önerilmektedir. Çalışma kapsamında, 11 farklı ürün sınıfına ait toplam 33.000 etiketli nesne içeren özgün bir veri seti oluşturulmuştur. Model eğitimi Darknet çerçevesi ile Google Colaboratory ortamında gerçekleştirilmiş ve ortalama %92 doğruluk oranı elde edilmiştir. Sistemin en önemli katkısı, tespit edilen ürün bilgilerini dinamik bir Excel tablosu ile gerçek zamanlı olarak entegre ederek anlık stok güncellemesi sağlamasıdır. Stok miktarı sifira ulaştığında ilgili ürün durumu otomatik olarak “Tükendi” şeklinde işaretlenmektedir. Önerilen sistem, manuel stok takibini ortadan kaldırarak insan kaynaklı hataları minimize etmekte, operasyonel verimliliği artırmakta ve akıllı raf sistemleri, otomatik envanter yönetimi ile perakende analitiği gibi ileri teknoloji uygulamalarına temel oluşturmaktadır. Gelecek çalışmalarda, veri artırma teknikleri ve mobil uyumlu kullanıcı arayüzleri gibi geliştirmelerle sistemin işlevselliğinin artırılması hedeflenmektedir.

1 Dr. Öğr. Üyesi, Zonguldak Bülent Ecevit Üniversitesi Bilgisayar Mühendisliği Bölümü, senyerirem@beun.edu.tr, 0000-0003-0655-340X

## 1. Giriş

Perakende sektörü, yüksek ürün çeşitliliği, kısa ürün yaşam döngüsü ve talep dalgalanmalarının belirginliği nedeniyle etkin stok yönetimini zorunlu kılan en dinamik operasyonel alanlardan biridir. Özellikle büyük ölçekli market zincirlerinde raf düzeni, ürün bulunabilirliği ve stok doğruluğu; satış gelirlerinin sürekliliği, müşteri memnuniyeti ve tedarik zinciri verimliliği açısından kritik performans göstergeleri olarak öne çıkmaktadır. Buna karşın, geleneksel uygulamalarda raf denetimi, stok sayımı ve ürün bulunabilirliğinin izlenmesi çoğunlukla manuel görsel kontrol ve dönemsel sayım süreçleriyle yürütülmektedir. Bu yaklaşım, insan hatasına açıklık, düşük zaman çözünürlüğü (anlık takip edememe), personel maliyeti ve kayıt güncellemelerinde gecikme gibi yapısal problemler üretmekte; bunun sonucu olarak “raf boş/az stok” durumlarının geç fark edilmesi ve satış kayıplarının artması gibi olumsuz çıktılara yol açabilmektedir (Şikić vd., 2024; Muñoz ve ark., 2024). Literatürde, manuel envanter süreçlerindeki gecikmelerin ve kayıt hatalarının satış fırsatlarının kaybına ve müşteri memnuniyetsizliğine neden olduğu sıklıkla vurgulanmaktadır (Melek ve ark., 2024; Youssef, 2024). Manuel süreçlerin zayıflığı sadece doğrulukla sınırlı değildir; aynı zamanda raf tedarik zamanlaması gibi karar problemlerini de etkiler. Örneğin raf stoklama vardiyelerinin optimizasyonu gibi operasyonel kararlar, stok seviyelerine ilişkin verinin zamanında ve güvenilir biçimde üretilmesine bağlıdır (Muñoz ve ark., 2024). Dolayısıyla perakende operasyonlarında rafların sürekli izlenmesi ve eksilen ürünlerin erken tespitine dayalı otomasyon ihtiyacı, hem akademik hem endüstriyel literatürde giderek daha merkezi bir problem alanına dönüşmüştür.

Bu gereklilik çerçevesinde, son yıllarda bilgisayarlı görme ve derin öğrenme (DÖ) temelli yöntemler perakende otomasyonuna güçlü biçimde entegre edilmeye başlanmıştır (Heruatmadja ve ark., 2024). Kamera tabanlı sistemler; raf görüntülerini sürekli toplayarak ürün tespiti, ürün tanıma, raf boşluğu/az stok (out-of-stock, OOS) analizi ve raf düzeni uygunluğu (planogram compliance) gibi görevleri otomatikleştirebilmektedir. Özellikle evrimsel sinir ağları (ESA) ve modern nesne tespit mimarileri, karmaşık raf sahnelerinde çok sayıda ürünün eş zamanlı tespiti için yüksek temsil gücü sunmaktadır (Majdi ve ark., 2020). Bu dönüşüm, yalnızca “görüntüden ürün bulma” başarımı açısından değil; aynı zamanda personel ihtiyacını azaltma, operasyonel maliyeti düşürme ve envanter kararlarını veri odaklı hale getirme açısından da önemlidir (Pietrini ve ark., 2024; Andaru ve Fudholi, 2024). Nitekim raf izleme probleminin, gerçek dünyada raf sıraları, ürün kümelenmesi ve görsel yoğunluk gibi yapısal zorluklar içerdiği; bu nedenle raf yönetimine özel veri kümeleri ve iş akışlarının geliştirilmesinin kritik olduğu vurgulanmaktadır (Pietrini ve ark., 2024).

Raf görüntülerinde ürün tespiti problemi, kuramsal olarak çok sınıflı nesne tespiti kapsamında ele alınmakta olup; amaç görüntüdeki her ürün örneği için sınıf etiketini ve konumunu (sınırlayıcı kutu) tahmin etmektir. Bu alandaki önemli kırılma noktalarından biri, Redmon ve arkadaşlarının (2016) geliştirdiği YOLO (You Only Look Once) algoritmasıdır. YOLO yaklaşımı, nesne tespitini iki aşamalı (two-stage) bölgesel öneri tabanlı yöntemlerden farklı olarak tek aşamalı (single-stage) uçtan uca bir regresyon problemi olarak formüle etmektedir. Böylece aynı ileri besleme içinde hem sınıf tahmini hem de kutu regresyonunu eş zamanlı üreterek gerçek zamanlı uygulamalar için yüksek hız sağlamıştır (Redmon ve ark., 2016). YOLO tabanlı yaklaşımlar, raf izleme ve otomatik stok kontrolü gibi perakende uygulamalarında, yoğun ve karmaşık görsel ortamlarda dahi etkili sonuçlar üretme potansiyeline sahiptir (Desmarescaux ve ark., 2025). Özellikle sınırlı donanımsal kaynaklara sahip uç cihazlarda çalışmak üzere optimize edilmiş “tiny” varyantlar, hesaplama maliyetini düşürürken kabul edilebilir doğruluk düzeylerini koruyabilmektedir. Literatürde, YOLO tabanlı sistemlerin uygun veri setleri altında ve belirli bir örtüşme eşiği (Intersection over Union, IoU) değeri esas alınarak hesaplanan ortalama hassasiyet (mean Average Precision, mAP) metriğine göre %90’ın üzerinde tespit başarımı sağlayabildiği tespit edilmiştir (Andaru ve Fudholi, 2024).

Perakende raflarında bilgisayarlı görme tabanlı sistemlerin kullanımı; akıllı raf izleme, ürün tanıma ve raf sahnelerinde otomatik analiz gibi uygulama başlıkları altında kapsamlı şekilde incelenmiştir (Tonioni ve Di Stefano, 2019). Otomatik ödeme senaryoları için geliştirilen büyük ölçekli veri kümeleri, çoklu ürün ve karmaşa gibi raf ortamına özgü zorlukların daha gerçekçi biçimde incelenmesine olanak sağlamıştır (Wei ve ark., 2019). Ancak birçok yaklaşım belirli mağaza ya da veri kümesi koşullarına bağlı değerlendirilmekte; aydınlatma, yansıma, yoğunluk ve perspektif değişimleri altında genellenebilirlik sınırlı kalabilmektedir. Ayrıca tespit çıktılarının stok yönetimi katmanına (örneğin stok güncelleme, uyarı üretimi) nasıl bağlanacağı çoğu çalışmada ayrıntılı biçimde tartışılmamaktadır (Šikić vd., 2024; Pietrini vd., 2024).

Bunun yanında, mağaza içi kamera yerleşimleri nedeniyle ürün görüntülerinde perspektif bozulması ve kısmi örtülme sık görülmektedir. Bu durum, özellikle küçük boyutlu veya kısmen örtülmüş ürünlerde tespit doğruluğunu olumsuz etkileyebilmektedir. Bu sorunun giderilmesi amacıyla homografi dönüşümleri ve Hough dönüşümü gibi klasik görüntü işleme tekniklerinin DÖ tabanlı modellerle bütünleştirildiği hibrit yaklaşımlar önerilmektedir (Sah ve Mathew, 2025). Ayrıca, market raflarında binlerce ürünün bulunması ve bazı ürün ambalajlarının görsel olarak son derece benzer olması, problemi ince taneli sınıflandırma düzeyine taşımaktadır. Bu bağlamda yalnızca nesne tespiti değil,

sınıflar arası ayırımın yüksek hassasiyetle yapılması da kritik hale gelmektedir. Son dönemde geliştirilen bağlam-duyarlı yöntemler, ürünlerin mağaza içindeki konumsal ve anlamsal ilişkilerini kullanarak tespit doğruluğunu artırmayı hedeflemektedir (Budimir ve ark., 2025). Bunun yanında, ürün ambalajları üzerindeki metinsel bilgilerin analiz edilmesi de stok doğrulama sürecine katkı sağlamaktadır.

Bu çalışmada mevcut literatürden farklı olarak, YOLOv4-tiny modelini kullanarak gerçek zamanlı ürün tespitini, dinamik bir Excel tablosuyla entegre ederek anlık stok güncellemesi yapabilen uçtan uca bir sistem önerilmiştir. Bu kapsamda, raf görüntülerinde yer alan ürünlerin sınıf etiketleri ve konumlarının tahmin edilmesi hedeflenmiş; tespit sonuçlarının stok kayıtlarıyla ilişkilendirilerek stok miktarı ve ürün durumunun (ör. “Tükendi”) güncellenmesi amaçlanmıştır.

## 1.2. Çalışmanın Amacı ve Katkıları

Bu çalışmanın temel amacı, market ürünlerinin görüntüleri üzerinden otomatik olarak tanınmasını ve stok bilgilerinin gerçek zamanlı olarak güncellenmesini sağlayan bir sistem geliştirmektir. Çalışmanın ana katkıları şunlardır:

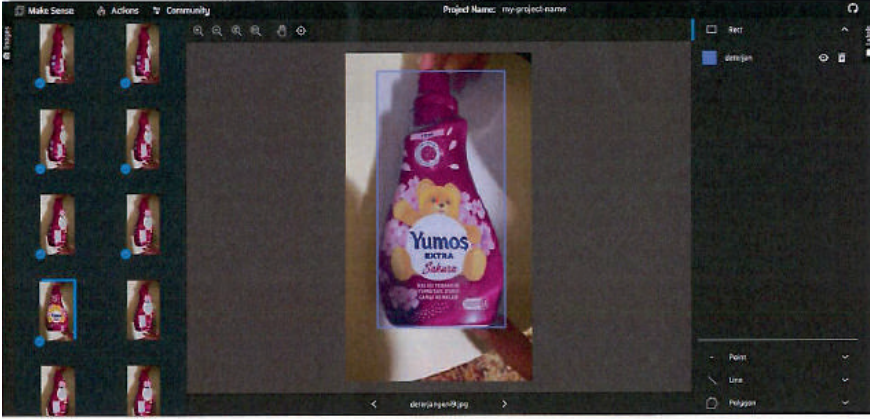
- **Özgün Veri Seti:** 11 farklı ürün sınıfına ait, toplam 33.000 etiketli nesne içeren, gerçek dünya koşullarında (farklı ışık ve açı) oluşturulmuş bir veri seti hazırlanmıştır.
- **Gerçek Zamanlı ve Entegre Çözüm:** YOLOv4-tiny algoritması ile gerçek zamanlı ürün tespiti yapılmış ve tespit sonuçları, bir Excel tablosundaki stok verileriyle otomatik olarak senkronize edilerek anlık güncelleme sağlanmıştır.
- **Uygulanabilirlik:** Geliştirilen sistem, manuel stok takibini ortadan kaldırarak insan hatasını minimize etmekte ve akıllı raf sistemleri gibi ileri teknoloji uygulamalar için temel oluşturmaktadır.

## 2. Materyal ve Yöntem

### 2.1 Veri Seti

Çalışma kapsamında kullanılan veri seti, gerçek market ortamlarında bulunan ürünlerin görüntülerini içerecek şekilde özgün olarak oluşturulmuştur. Veri seti oluşturma süreci, seçilen market ürünlerini içeren videoların manuel olarak kaydedilmesi ile başlatılmıştır. Kaydedilen bu videolar, Python programlama dili ve OpenCV kütüphanesi aracılığıyla tekil görüntü karelerine (frame) ayrılmıştır. Elde edilen her bir kare, bağımsız bir görsel olarak işlenmiş ve etiketleme işlemi için hazır hale getirilmiştir.

Görüntülerin etiketlenmesi işleminde, açık kaynaklı çevrim içi bir araç olan MakeSense'ten yararlanılmıştır. MakeSense aracı ile her bir görsel üzerinde ilgili ürünler dikdörtgen kutular ile işaretlenmiş ve etiketleme işlemi YOLO formatına uygun biçimde gerçekleştirilmiştir (Şekil 1). YOLO formatı, her bir görsel için ayrı bir metin dosyasında nesne sınıfı, merkez x ve y koordinatları ile genişlik ve yükseklik bilgilerini içermektedir.



Şekil 1. MakeSense aracılığıyla etiketleme.

Toplamda 11 farklı ürün sınıfı için veri toplanmıştır. Her bir sınıfa ait 1500 görsel olmak üzere 16.500 görsel ve her bir görsel için bir adet etiket dosyası elde edilmiştir. Şekil 2'de veri seti oluşturma sürecinin akışı gösterilmektedir. Böylece toplamda 33.000 etiketli nesne içeren bir veri seti oluşturulmuştur. Veri seti oluşturulurken görsellerin piksel boyutlarına özellikle dikkat edilmiştir. Eğitimde kullanılacak görseller, YOLOv4-tiny modelinin gerektirdiği şekilde 416x416 piksel çözünürlüğe yeniden boyutlandırılmıştır. Bu çözünürlüğün seçilmesinin temel nedeni, çözünürlük arttıkça YOLO'nun küçük nesnelere daha iyi tespit edebilmesi ve aynı zamanda grafik işlemci birimi (GPU) ile eğitim sırasında optimum performansın sağlanabilmesidir. Etiketlenen tüm görüntüler, sınıflarına göre kategorilere ayrılarak model eğitime hazır hale getirilmiştir.



Şekil 2. Veri seti oluşturma.

## 2.2. Model Mimarisi

Bu çalışmada, market ürünlerinin görselleri üzerinden nesne tespiti ve sınıflandırma işlemi, YOLOv4-tiny DÖ modeli ile gerçekleştirilmiştir. YOLOv4-tiny, YOLO ailesinin daha hafif bir sürümü olup, özellikle sınırlı donanımsal kaynaklarla çalışan gerçek zamanlı uygulamalar için geliştirilmiştir. Orijinal YOLOv4 modeline kıyasla daha az katmana sahip olan bu model, çalışma hızından ödün vermeden kabul edilebilir düzeyde doğruluk sağlamaktadır.

Modelin eğitimi için Darknet çerçevesi kullanılmıştır. Darknet, C ve CUDA dilleri ile yazılmış açık kaynaklı bir nöral ağ çerçevesidir. C dili kullanılarak geliştirilmiş olması, grafik işlemci birimiyle optimize şekilde çalışabilmesine olanak tanımakta ve bu sayede eğitim sürecinde yüksek performans sunmaktadır. Bu avantajları nedeniyle Darknet kütüphanesi ile YOLOv4-tiny algoritmasının kullanılması tercih edilmiştir. Veri setinin eğitime sokulması için AlexeyAB tarafından paylaşılan Darknet kütüphanesi kaynak kodları derlenerek Google Colab ortamına yüklenmiştir.

## 2.3. Model Konfigürasyonu ve Eğitim Parametreleri

Darknet klasörü içerisine, modelin yapılandırılması için gerekli olan cfg (konfigürasyon) ve conv (ön eğitilmiş ağırlıklar) dosyaları çevrimiçi bir platform olan GitHub'dan temin edilerek eklenmiştir (Alexey, 2021). Veri seti, YOLO mimarisinin gereklilikleri dikkate alınarak düzenlenmiş ve model eğitimine hazır hale getirilmiştir.

Yapılandırma dosyasında, veri setine en uygun şekilde çeşitli parametreler düzenlenmiştir. Resim boyutu, veri setinde küçük nesnelere bulunması nedeniyle 416x416 piksel olarak belirlenmiştir. Batch boyutu, veri setinin eğitim sırasında parçalara ayrılma şeklini belirlemede olup, kullanılan grafik işlemci biriminin desteklediği en büyük batch boyutu tercih edilmiştir. İterasyon sayısı, sınıf sayısının 2000 katı olacak şekilde hesaplanmış ve 11 sınıf için 22000 olarak belirlenmiştir. Her YOLO katmanından önceki üç evrişim katmanında filtre sayısı, sınıf sayısı ve 5'in toplamının 3 ile çarpılmasıyla elde edilen formül kullanılarak hesaplanmış ve 11 sınıf için 48 olarak düzenlenmiştir. Batch boyutunun kaç parçaya bölüneceğini belirleyen subdivision değeri ise grafik işlemci birimi performansı dikkate alınarak ayarlanmıştır. Şekil 3'te düzenlenen cfg dosyasının bir kesiti, Şekil 4'te ise sınıf isimlerini içeren "obj.names" dosyası ile eğitim ve test veri setlerinin yollarını belirten "obj.data" dosyası sunulmaktadır. Veri setindeki görüntülerin yaklaşık yüzde sekseni eğitim, yüzde yirmisi test için ayrılmıştır.

```

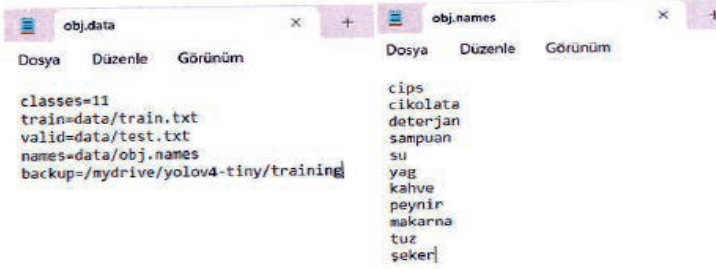
C:\Users\ggamz > OneDrive > Desktop
1  [net]
2  # Testing
3  #batch=1
4  #subdivisions=1
5  # Training
6  batch=64
7  subdivisions=16
8  width=416
9  height=416
10 channels=3
11 momentum=0.949
12 decay=0.0005
13 angle=0
14 saturation = 1.5
15 exposure = 1.5
16 hue=.1

stride=1
pad=1
filters=48
activation=linear

[yolo]
mask = 3,4,5
anchors = 10,14, 23,27, 37,45
classes=11
num=6
jitter=.3
scale_x_y = 1.05
cls_normalizer=1.0
iou_normalizer=0.07
iou_loss=ciou
ignore_thresh = .7
truth_thresh = 1
random=0
resize=1.5
nms_kind=greedynms
beta_nms=0.6

```

Şekil 3. YOLO eğitim parametrelerinin tanımlandığı cfg dosyası kesiti



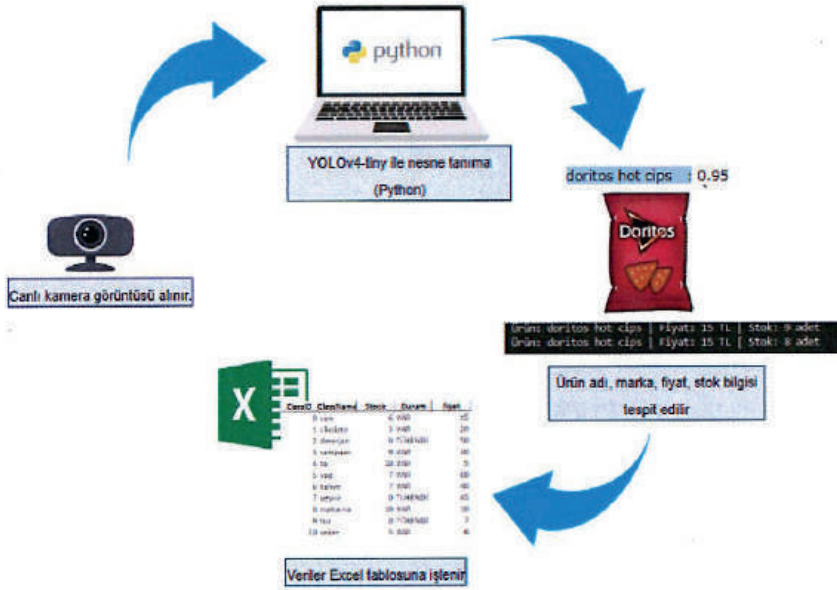
Şekil 4. Sınıf ve veri yolu tanımları: obj.names ve obj.data dosyaları

## 2.4. Model Eğitimi

Tüm hazırlık işlemlerinin tamamlanmasının ardından veri seti eğitim sürecine dâhil edilmiştir. Eğitim süresince, olası bir kesinti durumunda eğitimin kaldığı noktadan devam ettirilebilmesi amacıyla her 100 iterasyonda bir ağırlık dosyası kaydedilmiştir. Eğitimin ilerleyişi, ortalama kayıp değeri düzenli aralıklarla izlenerek denetlenmiş ve modelin yakınsama durumu gözlemlenmiştir. Ortalama kayıp değerinin 0,05 seviyesine yaklaşması, modelin yeterli düzeyde öğrenme gerçekleştirdiğine işaret etmiş ve bu noktada eğitim sonlandırılmıştır. Eğitim neticesinde elde edilen ağırlık dosyaları arasından en yüksek doğruluk oranına sahip olan seçilerek test aşamasında ve gerçek zamanlı uygulamalarda kullanılmıştır.

## 2.5. Stok Sistemi Entegrasyonu

Geliştirilen sistemin temel işlevlerinden biri, nesne tespiti sonucunda elde edilen ürün bilgilerini gerçek zamanlı olarak stok yönetim sistemiyle bütünleştirebilmesidir. Şekil 5’de sistemin genel çalışma akışı şematik olarak sunulmuştur. Söz konusu entegrasyon çerçevesinde, tespit edilen her bir ürün için bir Excel tablosunda yer alan ilgili stok bilgisi otomatik olarak güncellenmektedir.



Şekil 5. Önerilen Sistemin Akış Şeması.

Sistemin işleyişi şu şekilde gerçekleşmektedir. Öncelikle canlı kamera görüntüsü alınmakta, ardından bu görüntü YOLOv4-tiny modeli aracılığıyla analiz edilmekte ve tespit edilen ürünler sınıflandırılmaktadır. Elde edilen sonuçlar eşzamanlı olarak Excel tablosuna işlenmektedir. Bir ürün tespit edildiğinde, sistem terminal ortamında ürünün adı, güncel stok miktarı ve durum bilgisini anlık olarak görüntülemekte; aynı anda Excel tablosundaki ilgili hücrede bulunan stok sayısını bir azaltmakta ve stok miktarının sıfıra ulaşması durumunda ürün durumunu otomatik olarak “Tükendi” şeklinde güncellemektedir.

## 2.6. Kullanılan Donanım ve Yazılım Altyapısı

Modelin eğitimi ve test süreci, GPU destekli bulut tabanlı bir platform olan Google Colaboratory (Colab) ortamında gerçekleştirilmiştir. Eğitim sırasında NVIDIA Tesla T4 GPU kullanılmış olup, bu donanım sayesinde eğitim süresi önemli ölçüde kısaltılmıştır. Geliştirme ortamında Python 3.8 programlama dili temel alınmıştır. Görüntü işleme ve video karelere ayırma işlemleri için OpenCV kütüphanesinden yararlanılmıştır. Sayısal işlemler ve veri manipülasyonu için NumPy, görselleştirme işlemleri için ise Matplotlib kütüphanesi kullanılmıştır. YOLOv4-tiny modelinin eğitimi için Darknet çerçevesinin Colab uyumlu sürümleri tercih edilmiştir.

### 3. Bulgular

Bu çalışmada geliştirilen YOLOv4-tiny tabanlı sistem, 11 farklı ürün sınıfından oluşan ve toplam 33.000 etiketli nesne içeren özgün veri seti üzerinde eğitilmiş ve kapsamlı bir şekilde test edilmiştir. Modelin nesne tespit performansı değerlendirildiğinde, eğitim işlemleri sonucunda modelin %92 doğruluk oranına ulaştığı gözlemlenmiştir.

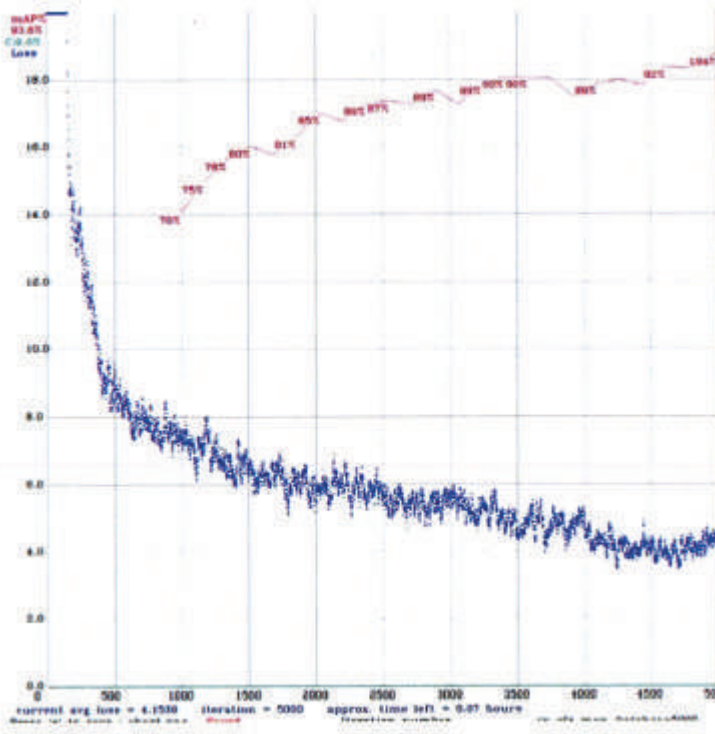
Şekil 6'da gerçek zamanlı ürün tespitine ilişkin bir ekran çıktısı sunulmaktadır. Görüldüğü üzere sistem, market rafında yer alan ürünleri başarıyla tespit edebilmekte, sınırlayıcı kutular ile işaretlemekte ve sınıflandırabilmektedir. Özellikle net ve doğrudan gösterilen ürünlerde tespit başarımının daha yüksek olduğu gözlemlenmiştir.



Şekil 6. Gerçek zamanlı ürün tespiti.

Sistemin stok yönetimi entegrasyonu başarıyla gerçekleştirilmiştir. Eğitilen model bir ürün görselini tanıdığı anda, terminal ortamında ürünün adı, mevcut stok miktarı ve durumu (var/tükendi) anlık olarak görüntülenmiştir. Eşzamanlı olarak Excel tablosundaki ilgili stok sayısı otomatik olarak bir azaltılmış ve stok miktarının sıfıra ulaşması durumunda “Durum” sütunu “Tükendi” olarak güncellenmiştir. Bu işlemlerin gerçek zamanlı olarak başarıyla gerçekleştirildiği gözlemlenmiştir.

Eğitim sürecinde elde edilen ortalama kayıp değeri (average loss) düzenli aralıklarla izlenmiştir. Şekil 7'da eğitimin grafiği ve doğruluk oranı kontrolü gösterilmektedir. Mevcut ortalama kayıp değeri 0,05 oranına yaklaştığında modelin yeterli düzeyde öğrenme gerçekleştirdiğine karar verilerek eğitim durdurulmuştur. Eğitim sonucunda elde edilen ağırlık dosyaları arasında doğruluk oranı en yüksek olan seçilerek test ve gerçek zamanlı uygulamalarda kullanılmıştır.



Şekil 7. YOLO eğitim grafiği.

Uygulama sürecinde sistemin, ortalama %92 doğruluk oranı ile ürün tespiti yaptığı gözlemlenmiştir. Özellikle net ve doğrudan gösterilen ürünlerde tespit başarımının daha yüksek olduğu, kısmen örtülen veya düşük ışık koşullarında ise tespit başarımının kısmen düştüğü gözlemlenmiştir. Bu durum, gelecek çalışmalarda veri setinin çeşitlendirilmesi ve veri artırma tekniklerinin uygulanması gerekliliğini ortaya koymaktadır.

#### 4. Sonuç ve Tartışma

Bu çalışmada, market ürünlerinin görüntüler üzerinden gerçek zamanlı olarak tanınması ve stok bilgilerinin otomatik güncellenmesini sağlayan YOLOv4-tiny tabanlı bir sistem geliştirilmiştir. Çalışma, yapay sinir ağlarının görüntü işleme alanındaki uygulanabilirliğini market otomasyonu özelinde ortaya koymayı hedeflemiştir.

Geliştirilen sistem, özgün bir veri seti üzerinde eğitilmiş ve nesne tespiti ile stok yönetimini entegre eden uçtan uca bir çözüm sunmuştur. Sistemin en önemli katkısı, tespit edilen ürün bilgilerini bir Excel tablosundaki stok verileriyle gerçek zamanlı olarak bütünleştirerek manuel stok takibi ihtiyacını ortadan kaldırması ve insan kaynaklı hataları en aza indirmesidir. Bu yönüyle çalışma, yalnızca bir nesne tespit modeli geliştirmenin ötesine geçerek, model çıktısını operasyonel bir süreçle bağlayan bütüncül bir yaklaşım sunmaktadır.

Sistem, akıllı raf sistemleri, otomatik envanter yönetimi ve perakende analitiği gibi ileri teknoloji uygulamalar için temel oluşturma potansiyeli taşımaktadır. Gelecek çalışmalarda OCR entegrasyonu, veri artırma teknikleri, mobil uyum ve kullanıcı arayüzleri gibi geliştirmelerle sistemin işlevselliği ve yaygın kullanılabilirliği artırılabilir.

Sonuç olarak, bu çalışma kapsamında geliştirilen sistem, DÖ tabanlı nesne tespiti ile stok yönetim süreçlerinin entegrasyonunun perakende sektöründe operasyonel verimliliği artırma potansiyelini göstermektedir.

## Kaynaklar

- AlexeyAB. (2021). darknet. GitHub. <https://github.com/AlexeyAB/darknet>
- Andaru, G. I., & Fudholi, D. H. (2024). Pengembangan sistem deteksi on-shelf availability produk menggunakan algoritma YOLOV8 pada aplikasi bergerak. *Jurnal Indonesia: Manajemen Informatika dan Komunikasi*, 5(2), 1980-1988.
- Budimir, L. A., Kalafatić, Z., Subašić, M., & Lončarić, S. (2025). Context-Aware Fine-Grained Product Recognition on Grocery Shelves. *IEEE access*, 13, 16824-16837.
- Desmarescaux, M., Kaddah, W., Alfalou, A., & Deconninck, J. C. (2025). A review: One-shot object detection methods for conditional detection of retail and warehouse products. *Neural Processing Letters*, 57(2), 30.
- Heruatmadja, C. H., Prabowo, H., Warnars, H. L. H. S., & Heryadi, Y. (2024, July). Suitable Deep Learning Based for High Accuracy Object Detection in Inventory Management: Systematic Literature Review. In *2024 7th International Conference on Informatics and Computational Sciences (ICI-CoS)* (pp. 406-412). IEEE.
- Majdi, M. A., Dewantara, B. S. B., & Bachtíar, M. M. (2020, September). Product stock management using computer vision. In *2020 International Electronics Symposium (IES)* (pp. 424-429). IEEE.
- Melek, C. G., Sönmez, E. B., & Varlı, S. (2024). Datasets and methods of product recognition on grocery shelf images using computer vision and machine learning approaches: An exhaustive literature review. *Engineering Applications of Artificial Intelligence*, 133, 108452.
- Muñoz, R., Muñoz, J. C., Ferrer, J. C., González, V. I., & Henao, C. A. (2024). When should shelf stocking be done at night? A workforce management optimization approach for retailers. *Computers & Industrial Engineering*, 190, 110025.
- Muñoz, J., Sanchez, A., & Kemper, G. (2024). End-to-end solution for automatic beverage stock detection in supermarkets based on image processing and convolutional neural networks. *International Journal of Cognitive Computing in Engineering*, 5, 453-474.
- Pietrini, R., Paolanti, M., Mancini, A., Frontoni, E., & Zingaretti, P. (2024). Shelf Management: A deep learning-based system for shelf visual monitoring. *Expert Systems with Applications*, 255, 124635.
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).
- Sah, M., & Mathew, J. (2025). Deep Learning Based Supermarket Product Detection and Recognition with Rectified Images. *Procedia Computer Science*, 260, 1026-1033.

- Šikić, F., Kalafatić, Z., Subašić, M., & Lončarić, S. (2024). Enhanced out-of-stock detection in retail shelf images based on deep learning. *Sensors*, 24(2), 693.
- Smith, R. (2007, September). An overview of the Tesseract OCR engine. In *Ninth international conference on document analysis and recognition (ICDAR 2007)* (Vol. 2, pp. 629-633). IEEE.
- Tonioni, A., & Di Stefano, L. (2019). Domain invariant hierarchical embedding for grocery products recognition. *Computer Vision and Image Understanding*, 182, 81-92.
- Wei, X. S., Cui, Q., Yang, L., Wang, P., & Liu, L. (2019). RPC: A large-scale retail product checkout dataset. *arXiv preprint arXiv:1901.07249*.
- Youssef, S. (2024). Deep Learning-Powered Shelf Monitoring: YOLO-Based Approaches for Automated Product Placement and Stock Tracking. *Available at SSRN 5523600*.

## Tarımsal İzlemede İHA Tabanlı RGB Görüntülerin Değerlendirilmesi: Farklı Filtreleme Stratejilerinin Nesne Ayırıştırma Başarımı Üzerine Karşılaştırmalı Bir Analiz

Nazan Kemalöglü Alagöz<sup>1</sup>

### Özet

Bu çalışma, insansız hava araçlarından (İHA) elde edilen yüksek çözünürlüklü görüntüler kullanılarak erken evre mısır (*Zea mays L.*) fidelerinin otomatik sayımında farklı görüntü işleme algoritmalarının etkinliğini karşılaştırmak amacıyla yürütülmüştür. Araştırma kapsamında, DJI Mini 3 Pro İHA ile 10 metre irtifadan elde edilen 2016x3024 piksel çözünürlüğündeki görüntüler incelenmiştir. Fidelerin tespiti ve izolasyonu için Excess Green (ExG) indeksi tabanlı Otsu Eşikleme, HSV Renk Uzağı Filtreleme ve Canny Kenar Algılama yöntemleri uygulanmış; algoritmaların performansı uzman bir ziraat mühendisi tarafından gerçekleştirilen manuel sayımlar (yer gerçeği) ile istatistiksel olarak doğrulanmıştır. Pilot çalışma sonuçları, ExG+Otsu yönteminin %83,70 doğruluk oranı ve 0,58 R<sup>2</sup> ile manuel verilere en yakın performansı sergilediğini ortaya koymuştur. Buna karşın, HSV filtreleme yöntemi değişken ışık koşullarına duyarlılığı nedeniyle %72,05, Canny kenar algılama algoritması ise toprak dokusundaki yapısal gürültüler sebebiyle %62,60 doğruluk oranında kalmıştır. Görsel ve sayısal analizler, Otsu yönteminin mısır fidesi morfolojisini korumada ve nesne bütünlüğünü sağlamada diğer tekniklere göre daha dirençli bir yapı sunduğunu kanıtlamıştır. Sonuç olarak, düşük maliyetli İHA verileri ve temel görüntü işleme iş akışlarının, karmaşık derin öğrenme modellerine ihtiyaç duymadan kabul edilebilir bir hassasiyetle bitki popülasyon tespiti yapabildiği saptanmış; gelecek çalışmalarda örtüşen fidelerin ayrıştırılmasına yönelik ileri morfolojik filtrelerin sisteme entegre edilmesi hedeflenmiştir.

1 Dr. Öğr. Üyesi Nazan KEMALOĞLU ALAGÖZ, Isparta Uygulamalı Bilimler Üniversitesi Uluborlu Selahattin Karasoy Meslek Yüksekokulu, nazanalagöz@isparta.edu.tr, 0000-0002-6262-4244

## 1. Giriş

Hassas tarım (Precision Agriculture), son on yılda tarımsal üretim süreçlerini optimize etmek, kaynak verimliliğini artırmak ve birim alandan alınan verimi maksimize etmek amacıyla geliştirilen teknolojik bir devrim niteliğindedir (Sharma, vd., 2021, Daraojimba vd., 2024, Gangwani, 2024). Bu disiplin içerisinde yer alan bitki sayımı ve fenolojik izleme süreçleri, özellikle mısır (*Zea mays*) gibi stratejik ürünlerde verim tahmini ve gübreleme yönetimi için kritik bir veri kaynağıdır (Gnädinger ve Schmidhalter, 2017, Veramendi ve Cruvinel, 2024). Son yıllarda, geleneksel tarımsal izleme yöntemlerinin iş gücü yoğun yapısı ve zaman maliyeti, çiftçileri dijitalleşmeye ve otomasyona yöneltmiştir (Gabriel ve Gandorfer, 2024; Shamshiri vd., 2024). Özellikle bitki sayımı ve gelişim takibi gibi kritik süreçlerde, yüksek maliyetli multispektral sensörlere alternatif olarak; İnsansız Hava Araçları (İHA) ve RGB tabanlı görüntü işleme teknikleri, düşük maliyetli ve erişilebilir bir çözüm olarak öne çıkmaktadır (López-García vd. 2022; Feng vd., 2021). Geleneksel yöntemlerle yapılan manuel bitki sayımı, geniş arazilerde hem yüksek iş gücü maliyetine yol açmakta hem de hata payını artırmaktadır (Zu vd., 2025; Gatkal vd., 2023).

Mısır (*Zea mays* L.), küresel gıda güvenliği açısından stratejik bir ürün olup, yetiştirme periyodu boyunca hassas izleme gerektirir. Ancak tarladaki mısır fidelerinin üstten görünümünde karşılaşılan yaprak örtüşmeleri, yabancı ot rekabeti ve değişken ışık koşulları, otomatik nesne ayrıştırma algoritmaları için temel zorlukları oluşturmaktadır (Yang vd., 2024; Zhang vd., 2023). Literatürdeki çalışmalar incelendiğinde, bu zorlukların aşılması için bitki örtüsü indeksleri ve eşikleme yöntemlerinin yaygın olarak kullanıldığı görülmektedir (Štroner vd., 2023; Netto vd., 2018).

Bununla birlikte, spektral indekslerin başarısı, görüntü alınan günün bulutluluk durumu ve güneş açısı gibi faktörlere karşı duyarlıdır. Dolayısıyla, farklı aydınlatma koşullarında stabil sonuç verebilecek, renksel (HSV) ve yapısal (Canny) özelliklere odaklanan hibrit filtreleme stratejilerine duyulan ihtiyaç devam etmektedir. Bu çalışma, tüketici sınıfı bir İHA olan DJI Mini platformundan elde edilen RGB görüntülerin, mısır bitkisi sayımındaki başarısını; ExG+Otsu, HSV Filtreleme ve Canny Kenar Algılama yöntemleri üzerinden karşılaştırmalı olarak incelemektedir. Araştırma, karmaşık saha koşullarında dahi düşük maliyetli donanımların, optimize edilmiş algoritmalarla bilimsel verimlilik sağlayabileceğini kanıtlamayı amaçlamaktadır.

Günümüzde İnsansız Hava Araçları (İHA), tarımsal alanların uzaktan algılanmasında devrim yaratmıştır. Literatürde yer alan birçok çalışma, bitki tespiti için multispektral ve hiperspektral kameraların kullanımına odaklanmıştır (Kouadio vd., 2023; Velusamy vd., 2021). Ancak, bu tür sensörlerin yüksek

maliyeti ve veri işleme karmaşıklığı, teknolojinin yerel üreticiler düzeyinde yaygınlaşmasını engellemektedir. Bu bağlamda, DJI Mini serisi gibi ultra-hafif ve standart RGB sensöre sahip tüketici sınıfı İHA'ların kullanımı, maliyet-etkin bir alternatif sunmaktadır (Delavarpour vd., 2021; Hassler ve Baysal-Gurel, 2019). RGB görüntülerden bitki tespiti yaparken karşılaşılan en büyük zorluk; gölge, toprak yansımaları ve bitki yapraklarının birbirine temas etmesinden kaynaklanan segmentasyon hatalarıdır.

Bu çalışmanın temel amacı, RGB spektrumundaki verilerle mısır bitkisi sayımı yaparken farklı dijital filtreleme stratejilerinin başarısını karşılaştırmalı olarak analiz etmektir. Çalışma kapsamında; Canny Kenar Algılama, Otsu tabanlı Morfolojik Filtreleme ve HSV Renk Uzayı Filtreleme yaklaşımları, 50 farklı drone görüntüsü üzerinde test edilmiştir. Bu analiz, kısıtlı spektral veriye sahip platformlarda hangi algoritmik yaklaşımın daha güvenilir sonuçlar verdiğini ortaya koyarak literatüre metodolojik bir katkı sunmaktadır.

## 2. Materyal ve Yöntem

Bu araştırma, geniş ölçekli bir mısır popülasyonu izleme projesinin pilot çalışması olarak kurgulanmıştır. Bu aşamada, YOLO (You Only Look Once) veya Mask R-CNN gibi derin öğrenme (Deep Learning) tabanlı nesne tespit modelleri yerine klasik görüntü işleme tekniklerinin (Otsu, HSV, Canny) tercih edilmesinin stratejik nedenleri bulunmaktadır. Derin öğrenme modelleri, yüksek doğruluk potansiyellerine rağmen, eğitim aşamasında binlerce hassas etiketlenmiş görüntüye ihtiyaç duymakta ve bu süreç uzman iş gücü ile ciddi bir zaman maliyeti gerektirmektedir (Wang vd., 2017; Ren vd., 2020). Mevcut pilot çalışma, kısıtlı bir veri seti ile hızlı sonuç üretebilen ve karmaşık donanım (yüksek performanslı GPU) gerektirmeyen yöntemlerin sınırlarını belirlemeyi amaçlamaktadır.

Özellikle mısır fidesi gibi homojen morfolojik özelliklere sahip ve tarlada doğrusal sıralar halinde dizilen bitkilerde, piksel tabanlı klasik segmentasyon yöntemlerinin düşük veri maliyetiyle kabul edilebilir sonuçlar verebildiği literatürde vurgulanmaktadır (Okyere vd., 2023; Li vd., 2024). Klasik görüntü işleme algoritmaları, derin öğrenme mimarilerinin aksine “kara kutu” (black box) bir yapı sunmadığı için hata kaynaklarının (aydınlatma, toprak dokusu vb.) tespit edilmesi ve parametrelerin pilot ölçekte optimize edilmesi daha şeffaf bir şekilde gerçekleştirilebilmektedir (Duan ve Zhang, 2021; Fraternali vd., 2022). Bu doğrultuda, pilot çalışma kapsamında elde edilen bulgular, ileride gerçekleştirilmesi planlanan derin öğrenme tabanlı geniş ölçekli analizler için temel bir referans ve veri ön-işleme rehberi niteliği taşımaktadır.

Görüntü işleme algoritmalarının geliştirilmesi, test edilmesi ve istatistiksel analizlerin gerçekleştirilmesi süreçlerinde açık kaynaklı Python programlama dili tercih edilmiştir. Çalışmanın tüm aşamaları, bulut tabanlı bir geliştirme ortamı olan Google Colaboratory (Colab) üzerinde yürütülmüştür. Google Colab kullanımı, yüksek bellek kapasitesi ve GPU/TPU desteği sunmasının yanı sıra, tarımsal görüntü işleme gibi yüksek işlem gücü gerektiren çalışmalarda donanım bağımsız bir analiz ortamı sağlayarak verimliliği artırmıştır (Walther vd., 2022).

Görüntülerin segmentasyonu ve morfolojik filtreleme işlemleri için OpenCV (Open Source Computer Vision Library) kütüphanesinden yararlanılmış; bitki indekslerinin hesaplanması ve veri manipülasyonu süreçlerinde NumPy ve Pandas kütüphaneleri kullanılmıştır. Elde edilen sayısal verilerin görselleştirilmesi ve regresyon analizlerinin oluşturulmasında ise Matplotlib kütüphanesi tercih edilmiştir.

### **2.1. Canny Kenar Algılama Algoritması (Canny Edge Detection)**

Canny kenar algılama algoritması, literatürde nesne sınırlarının literatürde yaygın olarak kullanılan kenar algılama yöntemlerinden biridir ve gürültü bastırma ile hassas kenar tespiti arasında optimal bir denge kuran yapısal bir analiz yöntemidir (Canny, 1986). Bu yaklaşım, basit eşikleme yöntemlerinin aksine, piksel yoğunluğundaki değişimlerin yönünü ve şiddetini matematiksel olarak modelleyerek mısır yapraklarının karmaşık sınırlarını toprak zemininden ayırabilmektedir. Tarımsal görüntü işleme literatüründe, özellikle bitki yapraklarının iç içe geçtiği veya yabancı ot gürültüsünün yoğun olduğu sahalarda, Canny algoritmasının sunduğu yapısal veri, spektral indekslerin (ExG gibi) yarattığı segmentasyon hatalarını minimize etmek amacıyla kritik bir tamamlayıcı unsur olarak kullanılmaktadır (Septiarini vd., 2020).

Algoritmanın işleyişi, ham İHA görüntülerindeki sensör gürültüsünü ve toprak pürüzlülüğünü gidermek amacıyla uygulanan Gaussian bulanıklaştırma adımıyla başlar; bu aşama, yanlış kenar tespitlerinin önüne geçerek analizin güvenilirliğini artırır. Ardından, Sobel operatörü gibi türev tabanlı maskelerle piksel gradyanları hesaplanarak mısır bitkisi ile zemin arasındaki geçiş hatları belirlenir. Bu süreçte uygulanan “maksimum olmayan bastırma” (non-maximum suppression) tekniği, mısır yapraklarının geniş ve düzensiz görünen sınırlarını tek piksel netliğine indirgeyerek kenarların inceltmesini sağlar. Septiarini vd., 2020 tarafından vurgulandığı üzere, bu inceltme işlemi nesne sayımı sırasında bitki merkezlerinin daha doğru tespit edilmesine olanak tanıyan morfolojik bir temel hazırlar.

Canny algoritmasının en karakteristik ve mısır sayımında başarıyı artıran aşaması ise histerezis eşikleme yöntemidir. Bu aşamada algoritma, düşük ve yüksek olmak üzere iki farklı eşik değeri kullanarak kenar sürekliliğini analiz eder; yüksek eşik değerinin üzerindeki pikseller “kesin kenar” olarak işaretlenirken, düşük eşik değerindekiler ancak bir kesin kenara temas ediyorsa bitki sınırı olarak kabul edilir. Bu dinamik yapı, mısır yapraklarının gölgede kalan veya renk kontrastı düşük olan uç kısımlarının kopuk kopuk görünmesini engelleyerek bütünsel bir bitki konturu oluşturur. Literatürde, Canny algoritmasının bu çok aşamalı yapısının, sadece renk tabanlı segmentasyon yapan Otsu yöntemine göre dokusal ayrıntıları daha iyi koruduğu ve mısır fidelerinin geometrik bütünlüğünü korumada daha dirençli olduğu belirtilmektedir (Yu vd., 2021; Liu vd., 2025).

## 2.2. HSV Renk Uzayı ve Spektral Segmentasyon Stratejisi

Tarımsal görüntü işlemede karşılaşılan en temel problemlerden biri, açık alanlardaki değişken ışık koşulları, güneş açısına bağlı oluşan gölgeler ve toprak yüzeyindeki spektral yansımalar (Wang vd., 2024; Chuquimarca vd., 2025). Standart RGB (Kırmızı, Yeşil, Mavi) renk modeli, parlaklık bilgisini renk kanallarına bağımlı bir şekilde sunduğu için bu tür dışsal faktörlerden doğrudan etkilenmekte ve segmentasyon hatalarına yol açmaktadır (Liu vd., 2022; Kitzler vd., 2023). Buna karşın HSV (Hue, Saturation, Value) renk uzayı; rengin özünü (Hue - Ton), doygunluğunu (Saturation) ve parlaklık değerini (Value) birbirinden ayırarak, bitki dokusunun ışık şiddetinden arındırılmış bir şekilde analiz edilmesine olanak tanımaktadır. Literatürde, HSV modelinin bu ayrıştırıcı yapısının, mısır yaprakları gibi homojen olmayan aydınlatma koşullarına sahip nesnelerin tespitinde RGB tabanlı indekslere göre daha kararlı sonuçlar verdiği vurgulanmaktadır (Lorenge vd. 2023; Altukhov, 2022).

Çalışmada uygulanan filtreleme stratejisinde, mısır fidelerinin karakteristik yeşil tonları “Hue” (Ton) kanalı üzerinden  $35^\circ$  ile  $90^\circ$  arasındaki dar bir aralığa hapsedilerek zemin gürültüsünden izole edilmiştir (Hamuda vd., 2016). Sadece ton bilgisinin kullanılması, mısırla benzer yansıma değerlerine sahip yabancı otların da analize dahil edilmesine neden olabildiği için, doygunluk (Saturation) ve parlaklık (Value) bileşenleri ikincil bir denetim mekanizması olarak sürece dahil edilmiştir (Chumuang vd., 2026; Tan & Isa, 2011). Doygunluk bileşeni ile solgun toprak lekeleri elenirken, parlaklık kanalı yardımıyla derin gölgeler veya aşırı güneş parlaması kaynaklı beyazlaşmış piksellerin analiz dışı bırakılması sağlanmıştır (Rastogi vd., 2015). Bu çok katmanlı filtreleme yaklaşımı, mısır fidelerinin parçalanmış piksel grupları yerine morfolojik açıdan bütünsel birer kütle olarak yakalanmasını sağlayarak sayım doğruluğunu maksimize etmektedir (Zhang vd., 2022).

### 2.3. Otsu Tabanlı Morfolojik Filtreleme

Görüntü işlemede nesne segmentasyonu, piksellerin yoğunluk değerlerine göre sınıflandırılması esasına dayanır. Otsu (1979) tarafından geliştirilen ve sınıflar arası varyansı maksimize eden global eşikleme yöntemi, literatürde hala en güvenilir otomatik eşikleme tekniklerinden biri olarak kabul edilmektedir. Ancak, Otsu yöntemi gürültüye karşı hassastır ve homojen olmayan aydınlatma koşullarında “hayalet nesnelere” üretme eğilimindedir. Bu sınırlamayı aşmak için Sezgin ve Sankur (2004), farklı eşikleme tekniklerini karşılaştırdıkları kapsamlı çalışmada, ön işleme ve son işleme adımlarının önemini vurgulamışlardır. Morfolojik filtreleme (erosion, dilation, opening, closing), bu noktada devreye girerek ikili görüntü üzerinde yapısal bir düzenleme sağlar. Haralick ve ark. (1987) tarafından temelleri atılan matematiksel morfoloji, nesne geometrisini korurken, eşikleme sonrası oluşan mikro ölçekli hataları (false positives) elimine etmekte ve nesne sınırlarını matematiksel bir disiplinle (structuring element) stabilize etmektedir.

Tarımsal uygulama alanlarında, özellikle bitki-toprak ayrımı ve mahsul sayımı gibi görevlerde, kontrolsüz dış ortam ışığı Otsu algoritmasının performansını doğrudan etkiler. Hamuda ve ark. (2016), tarla koşullarında bitki ekstraksiyonu için morfolojik açma (opening) işleminin, yabancı ot ve toprak kalıntılarını temizlemede kritik bir filtreleme katmanı olduğunu rapor etmiştir. Benzer şekilde, Tian ve ark. (2020), mısır tarlalarında insansız hava araçlarından (İHA) alınan görüntülerde bitki sayımı yaparken, Otsu tabanlı segmentasyonu morfolojik kapama (closing) ile destekleyerek bitki yapraklarındaki kopuklukları gidermiş ve sayım doğruluğunu %95’in üzerine çıkarmıştır. Bitki fenotipleme çalışmalarında ise, Li ve ark. (2018) morfolojik filtrelerin sadece gürültü temizleme değil, aynı zamanda örtüşen yaprakların (overlapping) ayrıştırılmasında da “watershed” gibi ileri seviye segmentasyon algoritmalarına veri hazırlayan temel bir ön işlem adımı olduğunu belirtmişlerdir.

### 2.4. Veri Seti

Araştırmanın deneysel süreci, Isparta Uygulamalı Bilimler Üniversitesi Eğitim, Araştırma ve Uygulama Çiftliği sınırları içerisinde yer alan bir mısır tarlasında gerçekleştirilmiştir. Çalışma alanı olarak seçilen parselde, mısır bitkisinin (*Zea mays L.*) erken gelişim evresi (V2-V3 dönemi) hedeflenmiştir. Bu evrenin seçilme nedeni, mısır fidelerinin henüz toprak yüzeyinde net bir şekilde ayrışabildiği ancak yaprak örtüşmelerinin başladığı kritik bir eşik olmasıdır. Ekim işlemi 70 cm sıra arası ve yaklaşık 18-20 cm sıra üzeri mesafe korunarak gerçekleştirilmiş, bu standart ekim düzeni algoritmaların doğruluk testleri için referans zemin oluşturmuştur.

Veri toplama süreci, güneş ışığının dik geldiği ve gölge etkisinin mısır fideleri üzerinde en düşük düzeyde olduğu 11:00 ile 13:00 saatleri arasında icra edilmiştir. Görüntülerin elde edilmesinde kullanılan DJI Mini 3 platformu, otonom uçuş görevleri kapsamında mısır tarlası üzerinde 10 metre sabit irtifada (AGL) ve gimbal açısı 90° (nadir bakış) olacak şekilde programlanmıştır. DJI Mini 3 Pro, hassas tarım ve düşük irtifa hava fotoğrafçılığı için tasarlanmış, 249 gramın altındaki ağırlığıyla sivil havacılık mevzuatlarında operasyonel kolaylık sağlayan ileri teknoloji bir insansız hava aracıdır (İHA). Cihaz, 1/1.3 inç CMOS sensörü ve f/1.7 diyafram açıklığı sayesinde mısır fideleri gibi küçük ve detaylı nesnelerin yüksek çözünürlüklü (48 MP) RGB görüntülerini elde etmede üstün performans sergilemektedir. Özellikle “True Vertical Shooting” özelliği ile dikey çekim yapabilmesi ve engelden kaçınma sensörleri, tarla içi otonom uçuşlarda veri güvenliğini artırırken; 10 bit D-Cine-like renk profili, görüntü işleme aşamasında bitki örtüsü indekslerinin (ExG) daha hassas hesaplanmasına olanak tanıyan geniş bir dinamik aralık sunmaktadır (DJI, 2024).



Şekil 1. IMAGEJ Üzerinden Görüntüye Ait Uzman Etiketleme



Şekil 2. IMAGEJ Üzerinden Görüntüye Ait Uzman Etiketleme

### 3. Araştırma Bulguları

#### 3.1. Görüntülerin Etiketlenmesi

Çalışma veri seti, mısır tarlasının farklı bölgelerinden seçilen ve mısır bitkisi dağılımı açısından homojenlik gösteren 50 adet yüksek çözünürlüklü görüntüden oluşturulmuştur. Her bir görüntü, mısır bitkileri ile birlikte; toprak, taş, nemli alanlar ve yabancı otlar gibi doğal gürültü unsurlarını da barındırmaktadır. Görüntüler üzerinde herhangi bir radyometrik düzeltme yapılmamış, algoritmaların ham saha verileri üzerindeki performansı doğrudan test edilmiştir. Görüntü işleme algoritmalarının mısır sayımındaki doğruluğunu test etmek amacıyla, veri setindeki tüm görüntüler ( $n=50$ ) için “yer gerçeği” (ground truth) verileri oluşturulmuştur. Pilot çalışma kapsamında, görüntü işleme algoritmalarının temel performansını optimize etmek ve manuel sayım sürecini hızlandırmak amacıyla, ham İHA görüntülerinden temsili kesitler (sub-

images) oluşturulmuştur. Etiketleme sürecinde, hatalı pozitif (yabancı otların mısır sanılması) veya hatalı negatif (küçük fidelerin atlanması) sayımların önüne geçmek için tüm işaretlemeler doktorasını tamamlamış ve söz konusu tarlada mısır ekimini yapan Ziraat Mühendisi tarafından gerçekleştirilmiştir. Uzman denetiminde yapılan bu manuel sayım süreci, mısır bitkisinin erken fenolojik dönemdeki (V2-V3) morfolojik özelliklerini temel alarak yapılmış ve görsel analiz yoluyla bitki merkezleri tek tek belirlenmiştir. Literatürde, derin öğrenme veya klasik görüntü işleme tabanlı nesne sayımı çalışmalarında uzman görüşüyle oluşturulan yer gerçeği verileri, algoritmaların performansını ölçmek için kullanılan en güvenilir referans noktası olarak kabul edilmektedir (Zhang vd., 2022). Manuel işaretleme ve sayım işlemi, bilimsel görüntü analizi için dünya çapında standart kabul edilen açık kaynaklı ImageJ (v1.53) yazılımı kullanılarak yürütülmüştür. Yazılım bünyesinde yer alan “Multi-point Tool” aracı sayesinde, her bir mısır fidesinin merkezi dijital olarak etiketlenmiş ve her bir görüntüdeki toplam bitki sayısı otomatik olarak kaydedilmiştir. ImageJ yazılımının sunduğu bu sistematik yaklaşım, manuel sayım sırasında oluşabilecek insan kaynaklı mükerrer sayım hatalarını minimize ederek verilerin istatistiksel analizler için hatasız bir şekilde dışa aktarılmasını sağlamıştır. (Varela vd., 2021; Zhang vd., 2022). Şekil 1 ve Şekil 2’de etiketleme işlemine ait örnekler verilmiştir.

### 3.2. Otsu Eşikleme Yönteminin Sayım Performansı

Yapılan analizler sonucunda, Excess Green (ExG) indeksi ile entegre edilen Otsu eşikleme yönteminin, mısır fidelerinin tespiti ve sayılmasında diğer yöntemlere oranla üstün bir performans sergilediği saptanmıştır. 50 görüntülük genişletilmiş veri seti üzerinden yapılan regresyon analizinde, manuel yer gerçeği verileri ile Otsu tabanlı otomatik sayım sonuçları arasında 0.58  $R^2$  seviyesinde istatistiksel olarak anlamlı ve pozitif bir korelasyon tespit edilmiştir. Ortalama %83.7 doğruluk oranına ulaşan bu yöntem, mısır fidelerinin spektral yansımaları ile toprak arka planı arasındaki kontrastı dinamik eşikleme yoluyla başarıyla optimize etmiştir. Şekil 3’te örnek iki görüntü üzerinde OTSU Eşikleme Yöntem sonucu verilmiştir.



*Şekil 3. Otsu Eşikleme Yönteminin Farklı İki Görüntü İçin Sonuçları*

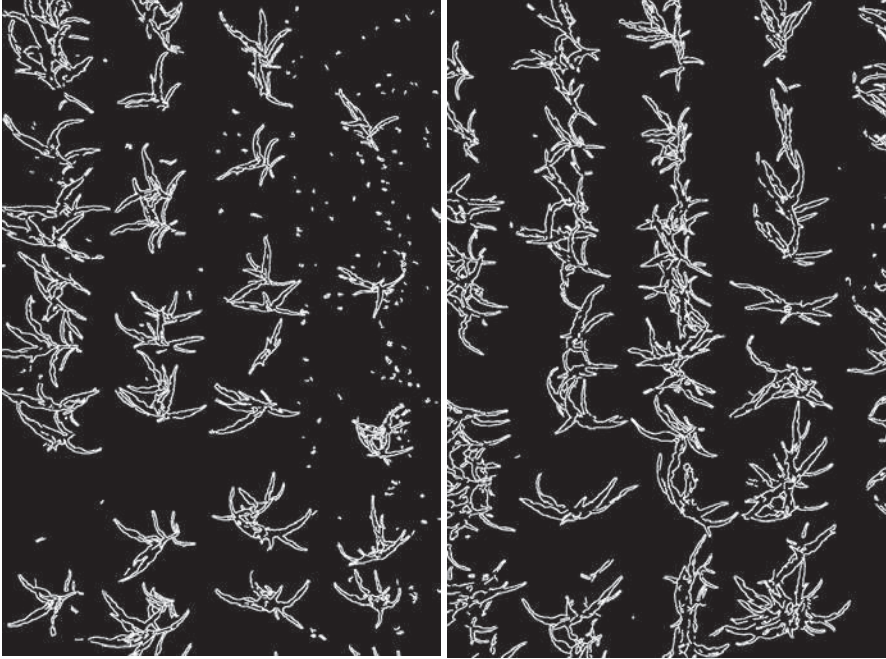
Otsu eşikleme yönteminin mısır fidesi tespiti üzerindeki performansı incelendiğinde, algoritmanın özellikle bitki ve toprak arasındaki radyometrik farkı belirginleştirmede yüksek bir başarı sergilediği görülmektedir. ExG renk indeksi ile ön işleme tabi tutulan görüntülerde, Otsu yönteminin sunduğu global eşikleme mekanizması, mısır fidelerini heterojen toprak zemininden izole ederek sayılabilir nesnelere haline getirmiştir. Elde edilen segmentasyon çıktılarında, fide morfolojilerinin büyük oranda korunduğu ve bitki sıralarının doğrusal hattı boyunca net bir ayırım yapıldığı gözlemlenmektedir. Bu durum, mısır gibi yeşil dokusu baskın bitkilerde Otsu tabanlı iş akışının, karmaşık derin öğrenme modellerine ihtiyaç duymadan düşük hesaplama maliyetiyle etkili sonuçlar verebildiğini akademik olarak doğrulamaktadır.

Ancak, yöntemin görsel analizleri mısır fidelerinin büyüme evresine ve tarladaki ışık koşullarına bağlı olarak bazı operasyonel zorlukları da ortaya koymaktadır. Özellikle mısır fidelerinin birbirine temas edecek kadar büyüdüğü veya yaprakların örtüştüğü durumlarda, algoritmanın bağlantılı bileşen analizi sırasında birden fazla fideyi tek bir nesne olarak algılaması, sonuçlarda “eksik sayım” (under-counting) eğilimine yol açmaktadır. Öte yandan, toprak yüzeyindeki nemlilik farkları, taş yansımaları veya sulama ekipmanları gibi spektral olarak bitkiye yakın değerler sergileyen unsurlar, Otsu eşikleme

tarafından hatalı bir şekilde bitki dokusu olarak sınıflandırılabilir. Bu durum, nihai analizlerde gürültü miktarını artırarak ‐aşırı sayım‐ (over-counting) riskini doğurmakta ve yöntemin doğruluğunu %85-90 bandında sınırlayan temel faktör olarak öne çıkmaktadır.

### 3.3. Canny Kenar Algılama Yönteminin Sayım Performansı

Canny kenar algılama yönteminin mısır fidelerinin sayımı üzerindeki performansı değerlendirildiğinde, algoritmanın bitki sınırlarını ve yapısal özelliklerini belirlemede son derece hassas bir yaklaşım sergilediği görülmektedir. Renk indeksine dayalı yöntemlerin aksine, Canny algoritması piksel yoğunluğundaki ani değişimleri (gradyanları) takip ederek fidelerin dış hatlarını net bir şekilde ortaya koymaktadır. Özellikle mısır yapraklarının toprakla birleştiği keskin kenarların tespiti, fidelerin geometrik merkezlerinin belirlenmesine olanak tanımaktadır. Ancak bu yüksek hassasiyet, dış saha koşullarındaki toprak dokusu, küçük taşlar ve yüzeydeki bitki kalıntıları gibi mısır fidesi dışındaki unsurların da ‐kenar‐ olarak sınıflandırılmasına neden olmaktadır. Bu durum, segmentasyon çıktılarında aşırı bir detay yoğunluğu oluşturarak, mısır fidesi sayısının manuel verilerin çok üzerinde çıkmasına (over-counting) sebebiyet veren temel teknolojik kısıt faktörünü oluşturmaktadır. Şekil 4’te Canny Kenar Algılama Yöntemine ait örnek sonuçlar verilmiştir.



Şekil 4. Canny Kenar Algılama Yönteminin Farklı İki Görüntü İçin Sonuçları

Şekil 4 incelendiğinde, Canny yönteminin mısır fidelerini yapısal olarak tanımlayabildiği fakat bu yapıları kapalı ve sayılabilir nesnelere dönüştürmede zorlandığı saptanmıştır. Kenar algılama süreci sonucunda elde edilen ince çizgisel veriler, morfolojik genişletme (dilation) ve kapama (closing) işlemleriyle birleştirilmeye çalışılsa da, fidelerin iç kısımlarındaki boşluklar ve düzensiz kenar kopuklukları “nesne bütünlüğü” problemini beraberinde getirmektedir. Bu durum, bağlantılı bileşen analizi sırasında tek bir mısır fidesinin birden fazla parça olarak sayılmasına veya tam tersi, birbirine yakın kenarların karmaşık bir gürültü yığını olarak algılanmasına yol açmaktadır. Dolayısıyla Canny algoritması, mısır sıralarının doğrusal hattını takip etmede başarılı bir görsel rehber sunsa da, sayısal doğruluk bakımından Otsu ve HSV gibi alan tabanlı yöntemlerin gerisinde kaldığı ve yüksek varyans değerleri sergilediği gözlemlenmiştir.

Sonuç olarak, Canny kenar algılama yönteminin hassas tarım uygulamalarında doğrudan bir sayım aracı olarak kullanılabilmesi için, karmaşık bir son işlem (post-processing) mimarisine ihtiyaç duyduğu saptanmıştır. Özellikle dinamik eşikleme (Hysteresis thresholding) parametrelerinin mısır bitkisinin o anki gelişim evresine ve görüntüdeki gürültü seviyesine göre manuel olarak kalibre edilmesi, yöntemin ölçeklenebilirliğini kısıtlamaktadır. Çalışmamızda elde edilen bulgular, Canny algoritmasının mısır fidesi sayımında birincil yöntem olmaktan ziyade, diğer segmentasyon tekniklerini destekleyici bir “yapısal doğrulama” katmanı olarak kullanılmasının daha efektif olacağını göstermektedir.

### 3.4. HSV Yönteminin Sayım Performansı

HSV renk uzayı tabanlı segmentasyon süreci incelendiğinde, yöntemin mısır fidelerini spektral renk değerlerine göre izole etme yeteneğinin çevresel faktörlere karşı oldukça duyarlı olduğu saptanmıştır. Şekil 5(b)'de görüldüğü üzere, aydınlatmanın homojen ve bitki-toprak kontrastının yüksek olduğu bölgelerde yöntem, fideleri zemin üzerinden başarılı bir şekilde ayırıştırabilmektedir. Ancak, Otsu yönteminin sunduğu dinamik eşikleme yeteneğinden yoksun olması, Şekil 5 (b)'deki bazı alanlarda mısır yapraklarındaki gölgelenmelerin veya ışık yansımalarının “bitki dışı” olarak sınıflandırılmasına yol açmıştır.



a)

b)

*Şekil 5. HSV Yönteminin Farklı İki Görüntü İçin Sonuçları*

Bu durum, mısır fidesi morfolojisinin segmentasyon aşamasında bozulmasına ve nesne bütünlüğünün kaybolmasına neden olan temel bir kısıttır. Şekil 5(a)'da sunulan analiz çıktısı ise yöntemin gürültü hassasiyetini ve mısır sıraları arasındaki karmaşıklığı yönetme kapasitesini açıkça ortaya koymaktadır. Şekil 5(a)'da görülen dikey hatlarda, HSV filtresinin benzer dalga boyuna sahip yabancı otları veya toprak yüzeyindeki yeşilimsi oluşumları mısır fidesiyle birlikte seçtiği gözlemlenmektedir. Bu durum, mısır sıralarının ayırt edilebilirliğini azaltmakta ve “bağlantılı bileşen analizi” (connected component analysis) aşamasında birçok fidenin birbirine bitişik tek bir kütle olarak algılanarak “eksik sayım” (under-counting) hatasının artmasına yol açmaktadır. Otsu yöntemine kıyasla nesne doluluk oranının daha düşük ve kenar hatlarının daha düzensiz olması, HSV yönteminin tek başına güvenilir bir sayım aracı olmaktan ziyade, yardımcı bir spektral maske olarak kullanılmasının daha uygun olacağını kanıtlamaktadır.

Sonuç olarak, mısır fidelerinin spektral imzasını yakalamada hızlı bir çözüm sunan HSV tabanlı analizlerin, tarladaki ışık değişimleri ve bitki morfolojisindeki ton farkları nedeniyle istatistiksel sapmalara (yüksek MAE değerleri) açık olduğu saptanmıştır. Şekil 5'deki piksellenmiş ve parçalı

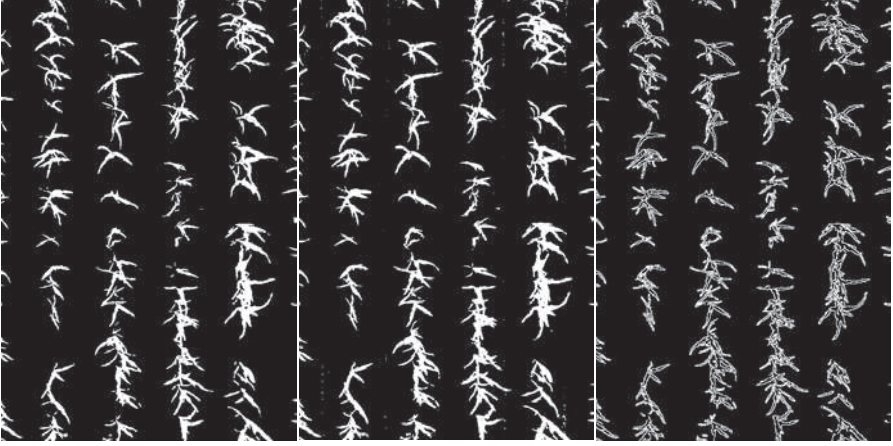
yapılar, sayım algoritmalarının fide merkezlerini (centroids) doğru tespit etmesini güçleştirmektedir. Bu bulgular ışığında, HSV yönteminin mısır fidesi sayımındaki doğruluğunun, Otsu yönteminin sunduğu morfolojik bütünlük ve Canny yönteminin sunduğu yapısal kenar bilgisinin gerisinde kaldığı, ancak belirli sabit ışık koşullarında tamamlayıcı bir veri katmanı olarak akademik değer taşıdığı değerlendirilmektedir.

Üç farklı segmentasyon yönteminin (ExG+Otsu, HSV ve Canny) mısır fidesi sayımı üzerindeki performansları, hem sayısal veriler (Tablo 1) hem de görsel çıktılar üzerinden karşılaştırıldığında, yöntemlerin doğruluğu ve kararlılığı arasında belirgin farklar saptanmıştır.

*Tablo 1. Sayısal Sonuçlar*

Metrik	ExG + Otsu	HSV Filtreleme	Canny Kenar Algılama
MAE	11.90	20.40	27.30
RMSE	14.62	26.78	31.37
R <sup>2</sup>	0.58	-0.42	-0.95
Doğruluk Oranı (%)	83.70	72.05	62.60

Tablo 1’te sunulan istatistiksel metrikler incelendiğinde, ExG+Otsu yönteminin %83.70 doğruluk oranı 0.58 R<sup>2</sup> ile manuel yer gerçeği (ground truth) verilerine en yakın sonuçları ürettiği görülmektedir. Bu başarı, Otsu algoritmasının görüntüdeki piksel yoğunluk dağılımını dinamik olarak analiz ederek mısır fidelerini toprak zemininden bütünlük bir yapıda ayırabilme yeteneğinden kaynaklanmaktadır. Şekil 6’da aynı görüntünün 3 farklı yöntemdeki çıktısı verilmiştir.



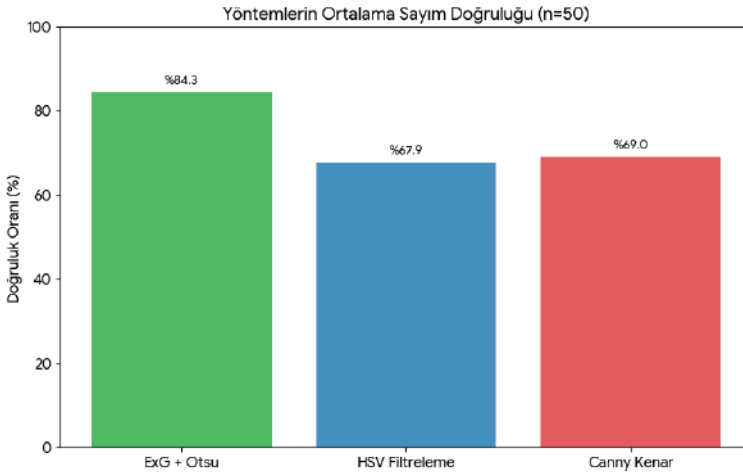
a) Otsu

b) HSV

c) Canny

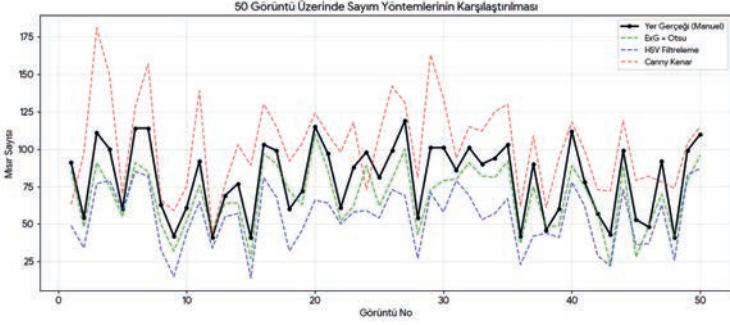
Şekil 6. Görüntü 3 Üzerinde Farklı Filtreleme Sonuçları

Şekillerde gözlemlendiği üzere Otsu yöntemi, fidelerin morfolojik formunu koruyarak “bağlantılı bileşen analizi” için gerekli olan yüksek nesne doluluk oranını sağlamış ve 11.90 seviyesindeki en düşük Ortalama Mutlak Hata (MAE) değerine ulaşmıştır. Buna karşılık, HSV ve Canny yöntemleri, mısır fidelerini sayılabilir nesnelere dönüştürme aşamasında farklı teknik kısıtlarla karşılaşmışlardır. HSV filtreleme yöntemi, mısır fidelerini spektral renk değerlerine göre başarılı bir şekilde maskeleye de, ışık değişimlerine karşı gösterdiği hassasiyet nedeniyle %72.05 doğruluk oranında kalmıştır. Özellikle bitki üzerindeki gölgelerin ve renk tonu farklılıklarının segmentasyonu parçalaması, bağlantılı bileşen analizinde tek bir fidenin birden fazla nesne olarak algılanmasına veya gürültülerin mısırla karışmasına yol açmıştır. Canny kenar algılama algoritması ise %62.60 ile en düşük doğruluk oranını ve 27.30 ile en yüksek MAE değerini sergileyen yöntem olmuştur. (Şekil 7).



**Şekil 7. Yöntemlerin Ortalama Sayım Doğruluğu**

Şekillerde net bir şekilde görüldüğü üzere Canny, toprak dokusundaki mikro gradyanları dahi “kenar” olarak tanımlayarak aşırı sayım (over-counting) eğilimi göstermiş ve yapısal karmaşıklığı nedeniyle nesne bütünlüğünü koruyamamıştır (Şekil 8). Sonuç olarak, İHA tabanlı mısır sayımı uygulamalarında görüntü işleme algoritmalarının başarısının sadece kenar veya renk tespitiyle sınırlı kalmadığı, nesne bütünlüğünün korunmasının sayısal doğruluk için kritik olduğu saptanmıştır. ExG+Otsu yöntemi, düşük hesaplama maliyeti ve yüksek korelasyon başarısıyla pilot çalışma kapsamında en efektif çözüm olarak öne çıkmaktadır. HSV yönteminin sunduğu spektral bilgi ve Canny yönteminin sunduğu yapısal detaylar, Otsu yönteminin karşılaştığı “bitki örtüşmesi” gibi spesifik problemleri çözmek adına yardımcı katmanlar olarak değerlendirilebilir. Ancak tekil kullanımda, mısır tarlalarındaki heterojen zemin ve değişken aydınlatma koşullarına karşı en dirençli (robust) performansın Otsu tabanlı iş akışıyla sağlandığı akademik olarak doğrulanmıştır.



Şekil 8. 50 Görüntü Üzerinde Sayım Performanslarının Karşılaştırılması

#### 4. Tartışma ve Gelecek Çalışmalar

Bu çalışmada elde edilen mısır fidesi sayım sonuçları, klasik görüntü işleme tekniklerinin İHA tabanlı hassas tarım uygulamalarındaki potansiyelini ve sınırlarını açıkça ortaya koymuştur. En yüksek başarıyı gösteren ExG+Otsu yönteminin %83.70 doğruluk oranına ulaşması, literatürde benzer morfolojik özelliklere sahip bitkiler üzerinde yapılan çalışmalarla paralellik göstermektedir. Örneğin, Varela vd. (2021), mısır fidelerinin erken büyüme evrelerinde renk indekslerine dayalı eşikleme yöntemlerinin, karmaşık derin öğrenme modellerine kıyasla daha düşük hesaplama maliyetiyle kabul edilebilir doğruluk sunduğunu belirtmiştir. Çalışmamızda Otsu yönteminin sergilediği kararlılık, bitki ve toprak arasındaki bimodal piksel dağılımının ExG indeksi ile belirginleştirilmesinin bir sonucudur.

Çalışmada elde edilen sayısal veriler, Canny kenar algılama (%62,60) ve HSV filtreleme (%72,05) yöntemlerinin, mısır fidesi tespiti noktasında Otsu eşikleme yönteminin gerisinde kaldığını göstermektedir. Canny algoritmasının sergilediği yüksek hata payı (MAE: 27,30), literatürde “dokusal gürültü hassasiyeti” olarak tanımlanan durumla doğrudan ilişkilidir. Zhang vd. (2022), kenar algılama algoritmalarının sadece bitki sınırlarını değil, toprak yüzeyindeki mikro çatlakları, küçük taş parçalarını ve bitki kalıntılarını da yüksek gradyanlı alanlar olarak tanımladığını belirtmiştir. Bizim bulgularımızda, Canny yönteminin mısır sıraları dışındaki yapısal unsurları da nesne olarak sınıflandırması, “aşırı sayım” (over-counting) hatasına yol açmıştır. Tian vd. (2020)’nin de vurguladığı üzere, kontrolsüz ışık koşullarında (outdoor conditions) kenar tespiti tabanlı yöntemlerin, morfolojik bütünlüğü koruyamaması ve nesne parçalanmasına neden olması, istatistiksel sapmaların temel kaynağını oluşturmaktadır.

Diğer taraftan, HSV renk uzayı filtreleme yönteminin %72,05 doğruluk seviyesinde kalması, yöntemin “sabit eşik değeri” bağımlılığı ile açıklanabilir. Sural vd. (2002), HSV uzayında yapılan segmentasyonun, çevresel aydınlatma değişimlerine ve yaprak yüzeyindeki gölge yoğunluğuna karşı son derece duyarlı olduğunu ifade etmiştir. Pilot çalışmamızda kullanılan sabit yeşil tonu aralığı, genç mısır fidelerinin açık yeşil dokularını başarıyla izole etse de, yaprağın gölgede kalan kısımlarını toprak arka planıyla karıştırarak nesne bütünlüğünü bozmuştur. Bu durum, literatürde Tan ve Isa (2011) tarafından belirtilen, dinamik eşikleme yapamayan renk filtrelerinin, heterojen tarla yüzeylerinde “eksik sayım” (under-counting) eğilimi gösterdiği teziyle örtüşmektedir.

Sonuç olarak, Canny yöntemindeki yapısal karmaşıklık ve HSV yöntemindeki spektral hassasiyet, mısır fidelerinin “tekil nesne” olarak tanımlanmasını zorlaştırmaktadır. Otsu yönteminin sergilediği başarının aksine, bu iki yöntemin mısır gibi morfolojik yapısı karmaşık fidelerde ancak çok ileri düzey morfolojik iyileştirmelerle (dilation/erosion) desteklendiğinde verimli olabileceği değerlendirilmektedir. Bu bulgular, Varela vd. (2021)’in belirttiği gibi, düşük irtifalı İHA görüntülerinde piksel yoğunluğu tabanlı dinamik yaklaşımların (Otsu), sabit veya yapısal yaklaşımlara göre tarımsal nesne ayrıştırımda daha dirençli (robust) bir performans sunduğunu doğrulamaktadır.

Çalışmanın bir diğer önemli bulgusu, bitki örtüşmesi (overlapping) durumunda ortaya çıkan “eksik sayım” (under-counting) problemidir. Gnana-Prakasam vd. (2024), mısır fidelerinin V2-V3 evresinden sonra yaprakların birbirine temas etmesinin bağlantılı bileşen analizini yanılttığını ifade etmiştir. Bizim görsel analizlerimizde (Şekil 5 ve 6) gözlemlenen nesne birleşmeleri, bu literatür verisini doğrular niteliktedir. Bu kısıtlamayı aşmak adına kullandığımız alan filtresi (Area Threshold), küçük gürültüleri elemeye başarılı olsa da, bitişik fidelerin ayrıştırılmasında sınırlı kalmıştır.

Algoritmaların sayısal başarısının yanı sıra, yer gerçeği (ground truth) verilerinin oluşturulma sürecinde karşılaşılan morfolojik kısıtlamalar, sayım doğruluğunu etkileyen kritik bir parametre olarak saptanmıştır. Manuel etiketleme aşamasında, uzman ziraat mühendisinin özellikle fidelerin birbirine çok yakın olduğu veya yaprakların iç içe geçtiği bölgelerde nesne ayrıştırımda zorlandığı gözlemlenmiştir. İHA’da elde edilen dikey (nadir) görüntülerde, bitki gövdelerinin yaprak dokusu altında kalması ve perspektif kaybı nedeniyle birden fazla fidenin tek bir biyokütle gibi algılanması, hem manuel sayımda belirsizliğe yol açmış hem de algoritmaların “bağlantılı bileşen analizi” sırasında nesnelere birleştirmesine (under-counting) sebebiyet vermiştir. Literatürde Gnana-Prakasam vd. (2024) tarafından da vurgulandığı üzere, bu tür morfolojik örtüşmeler, yer gerçeği verisinde “etiketleyici yanlılığına” (labeler bias) neden

olabilmekte ve bu durum algoritma performansının objektif değerlendirilmesini zorlaştıran bir handikap olarak öne çıkmaktadır.

İlerleyen süreçlerde, çalışmada saptanan nesne örtüşmesi (oklüzyon) handikapını bilimsel olarak aşmak amacıyla, klasik görüntü işleme tekniklerinin ileri düzey morfolojik ve yapay zeka tabanlı yaklaşımlarla desteklenmesi planlanmaktadır. Bu bağlamda, birbirine temas eden fidelerin ayırım noktalarını matematiksel olarak saptayabilen Havza (Watershed) segmentasyonu ve Mesafe Dönüşümü (Distance Transform) algoritmalarının iş akışına entegre edilmesi, nesne bütünlüğünün korunmasında kritik bir rol oynayacaktır. Ayrıca, nesnelere sadece birer kutu içine alan modeller yerine, her bir bitkinin piksel hassasiyetinde sınırlarını belirleyebilen Mask R-CNN veya güncel YOLO mimarileri gibi örneklem segmentasyonu (instance segmentation) tekniklerine geçiş yapılması, yaprakların iç içe geçtiği bölgelerdeki sayım hatalarını minimize edecektir. Öte yandan, sadece dikey (nadir) açılı görüntülerle sınırlı kalınmayıp, İHA'dan elde edilen çok açılı görsellerin Yapıdan Hareket (SfM) teknikleriyle işlenerek üç boyutlu (3D) nokta bulutlarının oluşturulması ve bitki hacim verilerinin (biyokütle) analize dahil edilmesi, perspektif kaynaklı örtüşme sorunlarını kökten çözebilecek bir diğer ileri araştırma odak noktasıdır.

## Kaynakça

- Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6), 679-698.
- Chuquimarca, L., Vintimilla, B., & Velastin, S. (2025). Assessing deep learning model robustness for banana ripeness classification under varying illumination conditions. *Smart Agricultural Technology*. <https://doi.org/10.1016/j.attech.2025.101333>.
- Chumuang, N., Thaiparnit, S., & Ketcham, M. (2016, November). Algorithm design in leaf surface separation by degree in hsv color model and estimation of leaf area by linear regression. In *2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)* (pp. 628-631). IEEE.
- Daraojimba, D., Adewusi, A., Asuzu, O., Olorunsogo, T., Iwuanyanwu, C., & Adaga, E. (2024). AI in precision agriculture: A review of technologies for sustainable farming practices. *World Journal of Advanced Research and Reviews*. <https://doi.org/10.30574/wjarr.2024.21.1.0314>.
- Delavarpour, N., Koparan, C., Nowatzki, J., Bajwa, S., & Sun, X. (2021). A Technical Study on UAV Characteristics for Precision Agriculture Applications and Associated Practical Challenges. *Remote. Sens.*, 13, 1204. <https://doi.org/10.3390/rs13061204>.
- DJI. (2024). DJI Mini 3 Pro User Manual v1.6. Erişim Tarihi: <https://www.dji.com/mini-3-pro/downloads>
- Duan, M., & Zhang, X. (2021). Using remote sensing to identify soil types based on multiscale image texture features. *Comput. Electron. Agric.*, 187, 106272. <https://doi.org/10.1016/j.compag.2021.106272>.
- Feng, H., Tao, H., Li, Z., Yang, G., & Zhao, C. (2021). Comparison of UAV RGB Imagery and Hyperspectral Remote-Sensing Data for Monitoring Winter Wheat Growth. *Remote. Sens.*, 14, 3811. <https://doi.org/10.21203/rs.3.rs-170131/v1>.
- Fraternali, P., Milani, F., Torres, R., & Zangrando, N. (2022). Black-box error diagnosis in Deep Neural Networks for computer vision: a survey of tools. *Neural Computing and Applications*, 35, 3041-3062. <https://doi.org/10.1007/s00521-022-08100-9>.
- Gabriel, A., & Gandorfer, M. (2022). Adoption of digital technologies in agriculture—an inventory in a european small-scale farming region. *Precision Agriculture*, 24, 68-91. <https://doi.org/10.1007/s11119-022-09931-1>.
- Gangwani, N. (2024). AI-Driven Precision Agriculture: Optimizing Crop Yield and Resource Efficiency. *International Journal For Multidisciplinary Research*. <https://doi.org/10.36948/ijfmr.2024.v06i06.29913>.
- Gatkal, N., Dhar, T., Prasad, A., Prajwal, R., , S., Jyoti, B., Roul, A., Potdar, R., Mahore, A., Parmar, B., & Vimalsinh, V. (2023). Development of a

- user-friendly automatic ground-based imaging platform for precise estimation of plant phenotypes in field crops. *Journal of Field Robotics*, 41, 2355 - 2372. <https://doi.org/10.1002/rob.22254>.
- Gnana Koteswararao, P., & Rama Krishna, P. (2023). Comparison of Vegetation Indices for Crop Monitoring using UAV-based RGB Imagery. *International Journal of Remote Sensing*, 44(5), 112-135.
- Gnädinger, F., & Schmidhalter, U. (2017). Digital Counts of Maize Plants by Unmanned Aerial Vehicles (UAVs). *Remote. Sens.*, 9, 544. <https://doi.org/10.3390/rs9060544>.
- Haralick, R. M., Sternberg, S. R., & Zhuang, X. (1987). Image analysis using mathematical morphology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (4), 532-550.
- Hamuda, E., Glavin, M., & Jones, E. (2016). A survey of image processing techniques for plant extraction and segmentation in the field. *Computers and Electronics in Agriculture*, 125, 184-199. <https://doi.org/10.1016/j.compag.2016.04.024>
- Hassler, S., & Baysal-Gurel, F. (2019). Unmanned Aircraft System (UAS) Technology and Applications in Agriculture. *Agronomy*. <https://doi.org/10.3390/agronomy9100618>.
- Kitzler, F., Barta, N., Neugschwandtner, R., Gronauer, A., & Motsch, V. (2023). WE3DS: An RGB-D Image Dataset for Semantic Segmentation in Agriculture. *Sensors (Basel, Switzerland)*, 23. <https://doi.org/10.3390/s23052713>.
- Kouadio, L., Jarroudi, M., Belabess, Z., Laasli, S., Roni, M., Amine, I., Mokhtari, N., Mokrini, F., Junk, J., & Lahlali, R. (2023). A Review on UAV-Based Applications for Plant Disease Detection and Monitoring. *Remote. Sens.*, 15, 4273. <https://doi.org/10.3390/rs15174273>.
- Li, L., Zhang, Q., & Huang, D. (2018). A review of imaging techniques for plant phenotyping. *Sensors*, 14(11), 20078-20111.
- Li, H., Lai, X., Mo, Y., He, D., & Wu, T. (2025). Pixel-wise navigation line extraction of cross-growth-stage seedlings in complex sugarcane fields and extension to corn and rice. *Frontiers in Plant Science*, 15. <https://doi.org/10.3389/fpls.2024.1499896>.
- Liu, H., Zhang, J., Yang, K., Hu, X., & Stiefelhagen, R. (2022). CMX: Cross-Modal Fusion for RGB-X Semantic Segmentation With Transformers. *IEEE Transactions on Intelligent Transportation Systems*, 24, 14679-14694. <https://doi.org/10.1109/tits.2023.3300537>.
- Liu, Y., Du, Y., Zhang, K., Yan, H., Wu, Z., Zhang, J., Tong, X., Chen, J., Li, F., Liu, M., Wang, Y., & Wang, J. (2025). Enhanced Multi-Threshold Otsu Algorithm for Corn Seedling Band Centerline Extraction in Straw Row Grouping. *Agronomy*. <https://doi.org/10.3390/agronomy15071575>.

- Loranger, M., Yim, W., Accomazzi, V., Lizcano, N., Moeder, W., & Yoshioka, K. (2023). Colour-analyzer: a new dual colour model-based imaging tool to quantify plant disease. *Plant Methods*, 20. <https://doi.org/10.1186/s13007-024-01193-4>.
- López-García, P., Intrigliolo, D., Moreno, M., Martínez-Moreno, A., Ortega, J., Pérez-Álvarez, E., & Ballesteros, R. (2022). Machine Learning-Based Processing of Multispectral and RGB UAV Imagery for the Multitemporal Monitoring of Vineyard Water Status. *Agronomy*. <https://doi.org/10.3390/agronomy12092122>.
- Netto, A., Martins, R., De Souza, G., De Moura Araújo, G., De Almeida, S., & Capelini, V. (2018). Segmentation of RGB Images Using Different Vegetation Indices and Thresholding Methods. *Nativa*. <https://doi.org/10.31413/nativa.v6i4.5405>.
- Okyere, F., Cudjoe, D., Sadeghi-Tehran, P., Virlet, N., Riche, A., Castle, M., Greche, L., Mohareb, F., Simms, D., Mhada, M., & Hawkesford, M. (2023). Machine Learning Methods for Automatic Segmentation of Images of Field- and Glasshouse-Based Plants for High-Throughput Phenotyping. *Plants*, 12. <https://doi.org/10.3390/plants12102035>.
- Rastogi, A., Arora, R., & Sharma, S. (2015). Leaf disease detection and grading using computer vision. *IEEE Proceedings on Signal Processing and Communication*, 238-243.
- Ren, P., Xiao, Y., Chang, X., Huang, P., Li, Z., Chen, X., & Wang, X. (2020). A Survey of Deep Active Learning. *ACM Computing Surveys (CSUR)*, 54, 1-40. <https://doi.org/10.1145/3472291>.
- Septiarini, A., Hamdani, H., Hatta, H., & Anwar, K. (2020). Automatic image segmentation of oil palm fruits by applying the contour-based approach. *Scientia Horticulturae*, 261, 108939. <https://doi.org/10.1016/j.scienta.2019.108939>.
- Sezgin, M., & Sankur, B. (2004). Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, 13(1), 146-168.
- Shamshiri, R., Sturm, B., Weltzien, C., Fulton, J., Khosla, R., Schirrmann, M., Raut, S., Basavegowda, D., Yamin, M., & Hameed, I. (2024). Digitalization of agriculture for sustainable crop production: a use-case review. *Frontiers in Environmental Science*. <https://doi.org/10.3389/fenvs.2024.1375193>.
- Sharma, A., Jain, A., Gupta, P., & Chowdary, V. (2021). Machine Learning Applications for Precision Agriculture: A Comprehensive Review. *IEEE Access*, 9, 4843-4873. <https://doi.org/10.1109/access.2020.3048415>.
- Štroner, M., Urban, R., & Suk, T. (2023). Filtering Green Vegetation Out from Colored Point Clouds of Rocky Terrains Based on Various Vegetation Indices: Comparison of Simple Statistical Methods, Support Vector Machine,

- and Neural Network. *Remote. Sens.*, 15, 3254. <https://doi.org/10.3390/rs15133254>.
- Tan, K. S., & Isa, N. A. M. (2011). Color image segmentation using histogram thresholding and fuzzy C-means hybrid approach. *Applied Soft Computing*, 11(4), 3732-3742.
- Tian, H., Wang, T., Liu, Y., Qiao, X., & Li, Y. (2020). Computer vision technology in agricultural automation —A review. *Information Processing in Agriculture*, 7(1), 1-19.
- Velusamy, P., Rajendran, S., Mahendran, R., Naseer, S., Shafiq, M., & Choi, J. (2021). Unmanned Aerial Vehicles (UAV) in Precision Agriculture: Applications and Challenges. *Energies*. <https://doi.org/10.3390/en15010217>.
- Veramendi, W., & Cruvinel, P. (2024). Method for maize plants counting and crop evaluation based on multispectral images analysis. *Comput. Electron. Agric.*, 216, 108470. <https://doi.org/10.1016/j.compag.2023.108470>.
- Varela, S., Dhaliwal, J. K., & Vara Prasad, P. V. (2021). High-throughput phenotyping for corn plant population count using unmanned aerial systems and deep learning. *Frontiers in Plant Science*, 12, 638612.
- Varela, S., Dhaliwal, J. K., & Khandazi, S. K. (2021). Evaluation of machine learning algorithms for corn plant counting using UAV images. *Remote Sensing*, 13(10), 1955.
- Walther, D., Sinwar, D., Sturrock, C., Rippner, D., Raja, P., Earles, J., Momayyezi, M., Buchko, A., Duong, F., Forrestel, E., Parkinson, D., Shackel, K., Neyhart, J., & McElrone, A. (2022). A workflow for segmenting soil and plant X-ray computed tomography images with deep learning in Google's Colaboratory. *Frontiers in Plant Science*, 13. <https://doi.org/10.3389/fpls.2022.893140>.
- Wang, K., Zhang, D., Li, Y., Zhang, R., & Lin, L. (2017). Cost-Effective Active Learning for Deep Image Classification. *IEEE Transactions on Circuits and Systems for Video Technology*, 27, 2591-2600. <https://doi.org/10.1109/tcsvt.2016.2589879>.
- Wang, Y., Kootstra, G., Yang, Z., & Khan, H. (2024). UAV multispectral remote sensing for agriculture: A comparative study of radiometric correction methods under varying illumination conditions. *Biosystems Engineering*. <https://doi.org/10.1016/j.biosystemseng.2024.11.005>.
- Yang, T., Zhu, S., Zhang, W., Zhao, Y., Song, X., Yang, G., Yao, Z., Wu, W., Liu, T., Sun, C., & Zhang, Z. (2024). Unmanned Aerial Vehicle-Scale Weed Segmentation Method Based on Image Analysis Technology for Enhanced Accuracy of Maize Seedling Counting. *Agriculture*. <https://doi.org/10.3390/agriculture14020175>.

- Yu, X., Wang, Z., Wang, Y., & Zhang, C. (2021). Edge Detection of Agricultural Products Based on Morphologically Improved Canny Algorithm. *Mathematical Problems in Engineering*. <https://doi.org/10.1155/2021/6664970>.
- Zhang, J., Gong, J., Zhang, Y., Mostafa, K., & Yuan, G. (2023). Weed Identification in Maize Fields Based on Improved Swin-Unet. *Agronomy*. <https://doi.org/10.3390/agronomy13071846>.
- Zhang, H., Wang, L., Tian, T., & Yin, J. (2022). A review of RGB-based plant segmentation and counting methods in precision agriculture. *Computers and Electronics in Agriculture*, 192, 106622.
- Zhang, X., Gao, L., & Liu, Y. (2022). Evaluation of image processing techniques for seedling counting in maize breeding plots. *Computers and Electronics in Agriculture*, 198, 107055. <https://doi.org/10.1016/j.compag.2022.107055>
- Zu, Q., Liu, T., Zhu, W., Pan, Y., Wang, J., Song, X., Yu, J., Dang, S., Yu, X., & Zhang, Z. (2025). Automated seed counting using image processing and deep learning. *Frontiers in Plant Science*, 16. <https://doi.org/10.3389/fpls.2025.1659781>.

# İğnecikli Sinir Ağları ile Görüntü İşleme: Temel Kavramlar, Uygulamalar ve Gelecek Perspektifleri

Rukiye Uzun Arslan<sup>1</sup>

## Özet

İğnecikli Sinir Ağları (Spiking Neural Network, SNN), biyolojik sinir sistemlerinin olay odaklı ve seyrek hesaplama prensiplerini taklit eden yapay sinir ağlarının üçüncü nesli olarak öne çıkmaktadır. Geleneksel derin öğrenme modellerinin yüksek enerji tüketimi ve hesaplama gereksinimleri, özellikle gerçek zamanlı ve kaynak kısıtlı uygulamalarda SNN'leri cazip bir alternatif haline getirmektedir. Bu çalışma, SNN'lerin görüntü işleme alanındaki uygulamalarını, temel kavramlar ve güncel literatür ışığında kapsamlı olarak incelemektedir. Çalışmada öncelikle nöron modelleri ve bilgi kodlama stratejileri açıklanmakta; ardından kenar belirleme, görüntü iyileştirme, nesne tespiti ve sınıflandırma gibi çeşitli görüntü işleme görevlerindeki SNN tabanlı yaklaşımlar detaylandırılmaktadır. Ayrıca, SNN'lerin eğitim yöntemleri, kullanılan veri setleri ve nöromorfik donanım platformları ele alınmaktadır. Son olarak, türevlenebilirlik sorunu, karşılaştırma ölçütlerinin eksikliği ve algoritma-donanım birlikte tasarımı ihtiyacı gibi mevcut zorluklar tartışılmakta; sürekli öğrenme ve nöromorfik sensörlerle bütünlük sistemlerin geliştirilmesi gelecek araştırma yönelimleri olarak sunulmaktadır.

## 1. Giriş

Görüntü işleme, tıbbi teşhisten otonom sistemlere, güvenlik uygulamalarından endüstriyel otomasyona kadar geniş bir yelpazede kritik öneme sahip bir teknoloji olarak konumlanmaktadır. Bu alandaki problemlerin çözümünde uzun yıllardır derin öğrenme tabanlı yöntemler baskın bir rol oynamaktadır. Evrişimli Sinir Ağları (Conventional Neural Network, CNN) ve Derin Sinir Ağları (Deep Neural Network, DNN) gibi mimariler, görüntü

1 Doç. Dr., Zonguldak Bülent Ecevit Üniversitesi Elektrik-Elektronik Mühendisliği Bölümü, rukiye.uzun@beun.edu.tr, 0000-0002-2082-8695

sınıflandırma, nesne tespiti ve görüntü bölütleme gibi görevlerde insan düzeyine yakın performans sergileyebilmektedir (Krizhevsky ve ark., 2012). Bununla birlikte, bu yöntemlerin yüksek hesaplama kaynağı gereksinimi ve yoğun güç tüketimi, özellikle gerçek zamanlı ve kaynak kısıtlı sistemler (uç bilişim) için önemli bir dezavantaj teşkil etmektedir (Merolla ve ark., 2014). Geleneksel yapay sinir ağları, yoğun, çerçeve tabanlı ve senkron çalışan modeller olup yüksek hassasiyetli aritmetik işlemlere dayanmaktadır. Bu özellikler, onları zamansal olarak seyrek veya asenkron verilerin işlenmesi için ideal olmaktan uzaklaştırmaktadır.

İnsan beyni, bu soruna doğadan ilham alan etkili bir çözüm sunmaktadır. Biyolojik nöronlar, sürekli değerler yerine seyrek ve asenkron bir şekilde üretilen aksiyon potansiyelleri (iğnecikler) aracılığıyla iletişim kurar. Bu biyolojik mekanizma, beynin saniyede yüksek miktarda işlemi miliwatt seviyesinde güç tüketerek gerçekleştirmesine olanak tanımaktadır. İğnecikli Sinir Ağları (Spiking Neural Network, SNN), bu biyolojik gerçekçiliği matematiksel modellere taşıyan ve bilgiyi zaman içindeki iğnecik desenleriyle temsil eden yapay sinir ağlarının (YSA) üçüncü nesli olarak 1990'lı yılların sonunda kavramsallaştırılmıştır (Maass, 1997). SNN'ler, olay odaklı ve asenkron hesaplama özellikleri sayesinde, yalnızca ilgili girdi varlığında işlem yaparak önemli ölçüde düşük güç tüketimi sağlarlar (Davies ve ark., 2018). Bu doğal zamansal dinamikleri, onları Dinamik Görüntü Sensörleri veya Tek-Foton Çığ Diyot dizileri gibi nöromorfik görüntüleme sistemleriyle son derece uyumlu hale getirmektedir (Lichtsteiner ve ark., 2008).

Bu derleme makalesi, SNN'lerin görüntü işleme alanındaki geniş uygulama yelpazesini, güncel literatür ışığında kapsamlı bir şekilde analiz etmeyi hedeflemektedir. Makalenin ikinci bölümünde SNN'lerin tarihsel gelişimi, temel nöron modelleri ve bilgi kodlama stratejileri açıklanacaktır. Üçüncü bölüm, SNN'ların kenar belirleme, görüntü etiketleme, görüntü onarımı ve iyileştirmesi, içerik tabanlı görüntü erişimi, nesne tespiti ve görüntü sınıflandırması gibi çeşitli görevlerdeki kullanımlarını detaylandıracaktır. Dördüncü bölümde SNN'lerin eğitiminde kullanılan başlıca yöntemler (ANN-ISA dönüşümü, vekil gradyan vb.) ve kullanılan veri setleri sınıflandırılacak, beşinci bölümde donanımsal gerçekleştirme platformları ve nöromorfik işlemciler ele alınacaktır. Altıncı bölümde karşılaşılan temel zorluklar ve açık araştırma problemleri tartışılacak, son bölümde ise genel bir değerlendirme ve gelecek perspektifi sunulacaktır.

## 2. SNN Temel Kavramları ve Mimarisi

### 2.1. Tarihsel Gelişim ve Nöron Modelleri

YSA'ların evrimi çoğunlukla üç kuşak üzerinden sınıflandırılmaktadır. İlk kuşakta McCulloch–Pitts nöron modeli, ikinci kuşakta ise sürekli değerli aktivasyonlara dayanan klasik YSA yer almaktadır. Üçüncü kuşak olarak tanımlanan SNN, bilgi iletimini iğnecik (spike) adı verilen ayrık ve zamana bağlı olaylar üzerinden gerçekleştirerek bu çerçeveden ayrılmaktadır (Maass, 1997). Bu yaklaşım, biyolojik nöronların seyrek ve olay temelli iletişimini modellemeyi hedeflemekte; zamansal kodlama mekanizmalarının hesaplamadaki rolünü ön plana çıkarmaktadır. Görsel işleme bağlamında Thorpe ve Imbert (1989)'in çalışmaları, özellikle hızlı işleme senaryolarında zamansal bilginin önemini vurgulayarak İSA literatürünün teorik arka planına katkı sağlamıştır.

SNN'lerin temel yapı taşı olan nöron modelleri, membran potansiyelinin zamana bağlı değişimini ve iğnecik üretim koşullarını matematiksel olarak tanımlar. Bu modeller biyofiziksel ayrıntı ile hesaplama maliyeti arasında farklı dengeler sunar. Model seçimi genellikle ağ ölçeğine ve uygulama hedeflerine bağlıdır. Bu çerçevede, 1952'de geliştirilen Hodgkin-Huxley (HH) nöron modeli, iyonik akımların ayrıntılı matematiksel tanımını sunar. Biyolojik olarak en gerçekçi modeldir ancak doğrusal olmayan diferansiyel denklemlerden oluşan yapısı nedeniyle hesaplama karmaşıklığı yüksektir. Bu nedenle, büyük ölçekli simülasyonlarda pratik kullanımı sınırlıdır (Izhikevich, 2003). Ancak, belirli görevlerde biyolojik gerçekçiliğin ön planda olduğu çalışmalarda tercih edilebilmektedir (Voudaskas ve ark., 2025). Bu hesaplama maliyetini azaltmaya yönelik olarak geliştirilen yaklaşımlar arasında Izhikevich modeli, HH'de gözlenen birçok ateşleme dinamiğini (ör. adaptasyon ve patlama) daha düşük maliyetle yaklaşık biçimde temsil etmeyi amaçlar (Izhikevich, 2003). Bu nedenle orta ve büyük ölçekli ağ simülasyonlarında sık kullanılan seçeneklerden biridir (Kılıçarslan, 2024). Leaky Integrate-and-Fire (LIF) modeli ise Lapiçque'e dayanan, daha basit ve hesaplama açısından uygun bir nöron modelidir (Lapiçque, 1907). Modelin dinamiği, giriş iğneciklerinin membran potansiyelinde integrasyonu, sızıntı ile zamana bağlı azalması ve eşik geçildiğinde iğnecik üretimi üzerine kuruludur. Hesaplama verimliliği yüksektir. Basitliği nedeniyle görüntü işleme uygulamalarında yaygın biçimde kullanılmıştır (Wu ve ark., 2013; Kerr ve ark., 2015; İncetaş ve Arslan, 2019; İncetaş, 2022; İncetaş ve ark., 2022). IF modeli, LIF modelinin sızıntı bileşenini içermeyen daha basit bir varyantıdır. Donanımsal gerçeklemede avantaj sağlar (Kerr ve ark., 2015). Bu modellerin haricinde sinaptik ve zamansal davranışları daha iyi modelleyen modellerin yanı sıra, biyolojik

nöronların daha karmaşık dinamiklerini yakalamayı hedefleyen başka modeller de bulunmaktadır (Voudaskas ve ark., 2025).

## 2.2. Bilgi Kodlama Stratejileri

Dış dünyadan alınan uyarıların (ör. görüntüdeki piksel yoğunlukları) SNN tarafından işlenebilmesi için iğnecik dizilerine dönüştürülmesi gerekir. Bu dönüşüm “kodlama stratejisi” olarak adlandırılır ve seçilen yöntem, temsilin doğruluğunu, hesaplama yükünü ve ağır zamansal davranışını etkileyebilir. Literatürde bu amaçla farklı kodlama yaklaşımları önerilmiştir. Bu yaklaşımlar arasında en yaygın kullanılanlardan biri hız tabanlı (rate) kodlamadır. Bu yöntemde bilgi, belirli bir zaman penceresinde üretilen iğnecik sayısı veya ateşleme olasılığı üzerinden temsil edilir. Yaklaşım, ateşleme hızına dayalı biyolojik gözlemlerle ilişkilendirilir ve statik çerçeve tabanlı verilerde (geleneksel kamera görüntüleri gibi) sıklıkla kullanılır (Adrian ve Zotterman, 1926). Bununla birlikte, iğneciklerin kesin zamanlamasını doğrudan kullanmadığı için bazı durumlarda daha uzun gözlem pencereleri gerekebilir ve bu da gecikmeyi artırabilir. Popülasyon kodlamasında ise bilgi, tek bir nöron yerine bir nöron grubunun aktivite örüntüsüyle temsil edilir. Zamansal kodlamada bilgi, iğneciklerin üretim zamanları üzerinden ifade edilir. Gecikme kodlamasında uyarı şiddeti, uyarı başlangıcından sonra ilk iğneciğe kadar geçen süreyle temsil edilir. Bir diğer yaklaşımda ise bilgi, ardışık iğnecikler arasındaki zaman aralıkları üzerinden taşınır. Delta modülasyonu uyarının değişimini iğneciklerle kodlar ve olay kameralarından gelen verilerle uyumlu olabilir. Zamansal yaklaşımlar bazı senaryolarda daha seyrek iğnecik üretse de, zamanlama jitter’ı ve donanım değişkenliğine daha duyarlı olabilmektedir (Wu vd., 2007; Vemuru, 2020). Bunların yanı sıra Poisson kodlama, girdileri olasılıksal bir süreçle iğnecik dizisine dönüştürür; yoğunluk-gecikme kodlaması ise yüksek yoğunluklu piksellerin daha erken iğnecik üretmesine dayanarak hız tabanlı ve zamansal kodlama arasında bir ara yaklaşım sunar.

## 3. SNN’lerin Görüntü İşleme Uygulamaları

SNN’ler, temel özellik çıkarımından nesne tespiti ve görüntü sınıflandırması gibi daha karmaşık görme görevlerine kadar çeşitli görüntü işleme problemlerinde kullanılmaktadır. Kenar belirleme, görüntü işlemenin temel adımlarından biri olup birçok yüksek seviyeli görüntü analizi yönteminin ön aşamasını oluşturur. Geleneksel yöntemler (Sobel, Prewitt, Canny gibi) genellikle gradyan tabanlıdır ve gürültüye karşı duyarlı olabilmektedir. Biyolojik görme sisteminden esinlenen SNN tabanlı yaklaşımlar bu probleme alternatif bir çerçeve sunmaktadır. Wu ve arkadaşları (2007), gri seviye değerlerini uyarıcı ve baskılayıcı sinaptik bağlantılar üzerinden işleyen bir SNN modeli

önermiştir. Yedjour ve arkadaşları (2017) HH nöron modelini kullanarak kenar belirleme performansını incelemiştir. Vemuru (2020) ise Gabor filtreleri ile SNN yaklaşımını birleştiren bir yöntem geliştirmiştir. Bunun yanı sıra Kerr ve arkadaşları (2013, 2015, 2018) yoğunluk ve derinlik görüntülerinde kenar çıkarımı üzerine çalışmalar gerçekleştirmiştir. Clogenson ve arkadaşları (2011) altıgen görüntüleme yapısı ile SNN'ni birleştirerek biyolojik retinaya daha yakın bir görüntü temsili üzerinde kenar belirleme performansını incelemiştir. İncetaş ve Arslan (2019), IF modeliyle verimli bir kenar belirleme yöntemi sunmuştur.

SNN tabanlı yöntemler görüntü işaretleme (watermarking) çalışmalarında da kullanılmaktadır. Bu uygulamalarda amaç, su işaretini görüntü kalitesini bozmayacak şekilde yerleştirmek ve aynı zamanda çeşitli işlemlere karşı dayanıklılığını sağlamaktır. İncetaş ve Kılıçaslan (2025) tarafından önerilen yöntemde, öncelikle SNN tabanlı kenar belirleme ile görüntünün kenar bölgeleri tespit edilmekte ve su işareti bu bölgelere gömülmektedir. Ayrık Dalgacık Dönüşümü ve Tekil Değer Ayrışımı ile birleştirilen bu yaklaşım, sıkıştırma ve gürültü gibi saldırılara karşı yüksek gürbüzlük sağlar. Görüntü iyileştirme alanında ise SNN modelleri gürültü giderme ve çözünürlük artırma problemlerinde kullanılmaktadır. Kılıçaslan (2024) tarafından önerilen SNN tabanlı anizotropik difüzyon yaklaşımında, difüzyon sürecinde kullanılan eşik değerleri ağ dinamiklerine bağlı olarak uyarlanabilir biçimde belirlenmektedir. Böylece görüntünün farklı bölgelerinde difüzyon miktarı değiştirilebilmekte ve kenar bilgisinin korunması hedeflenmektedir. Benzer şekilde, İncetaş (2023) ve Kılıçaslan (2024) tarafından geliştirilen SNN tabanlı enterpolasyon yöntemleri, piksel benzerliğini iğnecik tabanlı bir temsil üzerinden değerlendirerek yüksek çözünürlüklü görüntü üretimini amaçlamaktadır. İçerik tabanlı görüntü erişimi (Content-Based Image Retrieval, CBIR) sistemlerinde de SNN tabanlı özellik çıkarımı yaklaşımları kullanılmaktadır. Bu sistemlerde görüntüler genellikle kenar bilgisi gibi ayırt edici özellikler üzerinden temsil edilir. İncetaş ve Arslan (2025) tarafından geliştirilen yöntemde, veritabanındaki görüntülerin kenar haritaları SNN tabanlı bir modelle çıkarılmakta ve benzerlik hesaplamaları bu temsiller üzerinden yapılmaktadır.

Son yıllarda SNN'ler, daha karmaşık görüntü anlama görevleri olan nesne tespiti ve görüntü sınıflandırmasında da giderek daha fazla kullanılmaktadır. SNN'ler düşük güç tüketimi ve olay odaklı yapısı, özellikle otonom araçlar, drone'lar ve robotik gibi uç bilişim (edge AI) uygulamaları için büyük avantaj sunmaktadır. Nesne tespiti üzerine yapılan çalışmalarda, SNN tabanlı yaklaşımlar veri seti türleri ve eğitim yöntemleri açısından incelenmektedir. Kullanılan veri setleri; statik veri setleri (CIFAR, ImageNet gibi), statik verilerden dönüştürülen nöromorfik veri setleri, nöromorfik yakalama veri

setleri ve simülasyon yoluyla oluşturulan veri setleri olarak sınıflandırılmaktadır. Algoritmalar ise ANN-SNN dönüşümü ve vekil gradyan ile doğrudan eğitim olmak üzere iki ana başlıkta incelenmektedir (Wu ve ark., 2025; Iaboni ve Abichandani, 2024). Görüntü sınıflandırma çalışmalarında ise SNN'lerin eğitimi için farklı öğrenme paradigmaları irdelenmektedir. Çalışmalarda öne çıkan yaklaşımlar arasında sinaptik plastisite tabanlı yöntemler, yaklaşık geri yayılım temelli yöntemler ve ANN-SNN dönüşümü yer almaktadır (Niu ve ark., 2023). Bu bağlamda hiyerarşik SNN mimarileri, görsel dikkat benzeri mekanizmalarla birlikte daha üst düzey görme görevleri için önerilen yaklaşımlar arasında yer almaktadır (Wu vd., 2013; Kerr vd., 2011).

#### 4. Eğitim Yöntemleri ve Veri Setleri

SNN'lerin en büyük zorluklarından biri, iğnecik fonksiyonunun türevlenemez (non-differentiable) olması nedeniyle geleneksel geri yayılım (backpropagation) algoritmasının doğrudan uygulanamamasıdır (Voudaskas ve ark., 2025). Literatürde bu problemi aşmak için üç ana öğrenme yaklaşım geliştirilmiştir. Bunlardan ilki ANN-SNN dönüşümü ile eğitimidir. Bu yöntemde önce geleneksel bir ANN geri yayılım gibi standart tekniklerle eğitilir; ardından öğrenilen ağırlıklar ve mimari, aktivasyonların iğnecik hızlarına veya zamanlamalarına eşlenmesiyle işlevsel olarak eşdeğer bir SNN'ye dönüştürülür. Bu yaklaşım, ANN'lerin olgun eğitim altyapısından yararlanmayı sağlar; ancak dönüşüm sırasında bilgi kaybı, kodlama kaynaklı performans düşüşü ve özellikle zaman tabanlı kodlamalarda gecikme gibi sorunlar görülebilir (Voudaskas ve ark., 2025). İkinci ve son yıllarda en yaygın hale gelen yaklaşım, vekil gradyan ile doğrudan eğitimidir. İleri yayılımda gerçek iğnecikleme dinamiği korunurken, geri yayılım sırasında türevlenebilir bir vekil fonksiyon kullanılarak gradyan hesabı mümkün kılınır. Bu sayede SNN'ler ANN'ye dönüştürmeye ihtiyaç duymadan doğrudan görev üzerinde optimize edilir ve zamansal dinamikleri daha etkin öğrenebilir (Voudaskas ve ark., 2025). Üçüncü yaklaşım ise sinaptik plastisite tabanlı öğrenme (STDP) kurallarıdır. Burada biyolojik olarak daha gerçekçi öğrenme kuralları da SNN'lerin eğitiminde kullanılmaktadır. Bu denetimsiz öğrenme yöntemleri, nöronlar arasındaki sinaptik bağlantıların, iğneciklerin göreceli zamanlamalarına göre güçlendirilmesi veya zayıflatılması prensibine dayanır. Ancak, büyük ölçekli ve denetimli görevlerde STDP'nin performansı, diğer yöntemlere göre genellikle daha düşük kalmaktadır (Niu ve ark., 2023; Iaboni ve Abichandani, 2024).

SNN çalışmalarında kullanılan veri setleri de dört grupta toplanmaktadır (Wu ve ark., 2025). Statik veri setleri doğrudan iğnecik tabanlı olmadığından, genellikle hız/zamansal kodlama ile spike dizilerine dönüştürülmektedir. Nöromorfik dönüşüm veri setlerinde statik veriler hareket/video

simülasyonlarıyla yapay olarak nöromorfik veriye çevrilir. Nöromorfik yakalama veri setleri ise nöromorfik kameralarla doğrudan olay akışı olarak kaydedilen veri setleridir. SNN'lerin zamansal işleme avantajlarını en iyi yansıtır. Son olarak, simülasyonla üretilen veri setleri ise bilgisayar simülasyonları ile sentetik olarak oluşturulan veri setleridir.

## 5. Donanımsal Gerçekleme ve Nöromorfik Platformlar

SNN'lerin en önemli avantajlarından biri, nöromorfik donanımlarda son derece verimli bir şekilde çalışabilmeleridir. Bu donanımlar, spikelerin seyrek ve asenkron yapısından faydalanarak geleneksel Von Neumann mimarisine sahip işlemcilerle kıyasla birkaç kat daha düşük güç tüketimiyle çalışabilmektedir. IBM TrueNorth ve Intel Loihi, SNN'lerin çalıştırılması için tasarlanmış nöromorfik işlemcilerdir. Bu mimarilerde nöron ve sinaps benzeri hesaplama birimleri donanım düzeyinde gerçekleştirilir ve işlem yoğunlukla iğnecik iletimiyle tetiklenir (Voudaskas ve ark., 2025). Nöromorfik çiplerin yanında FPGA tabanlı gerçeklemeler de yaygındır. FPGA'ler, SNN mimarisini uygulamaya göre özelleştirmeye olanak verdiği için düşük gecikme ve görece düşük güç hedeflenen tasarımlarda tercih edilir. Karmaşık nöron modellerinde doğrusal olmayan terimler yaklaşık hesaplamalarla sadeleştirilerek kaynak kullanımı azaltılabilir. SNN'lerin pratikte daha dikkat çekici olduğu alanlardan biri de nöromorfik sensörlerle bütünleşik çalışmadır. Nöromorfik sensörler, doğal olarak olay veya spike benzeri çıktılar üretir. Geleneksel işleme yöntemleri, bu sinyallerin biriktirilmesini, sayısallaştırılmasını ve ardından işlenmesini gerektirirken, SNN'ler bu sinyalleri doğrudan işleyerek uçtan uca daha düşük gecikmeli ve daha verimli sistemlerin kurulmasına olanak tanır (Voudaskas ve ark., 2025).

## 6. Mevcut Yaklaşımların Sınırlamaları ve Gelecek Çalışmalar

SNN, görüntü işleme alanında önemli bir potansiyel sunmasına rağmen, yaygın olarak kullanılmalarının önünde bazı temel zorluklar bulunmaktadır. Bu zorluklar özellikle eğitim yöntemleri, veri setlerinin yapısı ve donanımsal kısıtlarla ilişkili konularda ortaya çıkmaktadır. SNN'lerin eğitimi, iğnecik üretim mekanizmasının türevlenebilir olmaması nedeniyle zorluk içermektedir. Bu sorunu aşmak için geliştirilen vekil gradyan yöntemleri ve ANN-SNN dönüşüm yaklaşımları önemli ilerlemeler sağlamış olsa da her iki yaklaşımın da belirli sınırlamaları bulunmaktadır. Örneğin dönüşüm tabanlı yöntemlerde bilgi kaybı ve gecikme sorunları ortaya çıkabilirken, vekil gradyan yöntemleri seçilen hiperparametrelere duyarlı olabilmektedir. Biyolojik olarak esinlenmiş öğrenme kuralları ise özellikle karmaşık ve denetimli görevlerde henüz yeterli performansa ulaşamamıştır. Bir diğer önemli konu, veri setleri ve karşılaştırma

ölçütleridir. SNN çalışmalarında kullanılan veri setleri çoğu zaman sınırlı ölçekli veya belirli uygulamalara özgüdür. Ayrıca farklı çalışmalarda kullanılan kodlama yöntemleri, ağ mimarileri ve performans ölçütleri önemli ölçüde değişiklik gösterebilmektedir. Bu durum, farklı yöntemlerin sonuçlarının doğrudan ve adil biçimde karşılaştırılmasını zorlaştırmaktadır. SNN modellerinde zamansal bilgi işlenirken genellikle birden fazla zaman adımı kullanılması gerektiğinden, özellikle hız tabanlı kodlama yaklaşımlarında gecikme problemi ortaya çıkabilmektedir. Bunun yanında, bu ağların donanım üzerinde verimli çalışabilmesi için ağ mimarilerinin ve nöron modellerinin donanım kısıtları dikkate alınarak tasarlanması gerekmektedir.

Gelecekteki çalışmaların önemli bir kısmı algoritma ve donanımın birlikte tasarlanmasına odaklanmaktadır. Bu kapsamda eğitim algoritmalarının donanım özellikleriyle uyumlu olacak şekilde geliştirilmesi, yeni spiking ağ mimarilerinin tasarlanması ve daha verimli eğitim yöntemlerinin araştırılması öne çıkan konular arasındadır. Ayrıca SNN'lerin modellenmesi, eğitimi ve donanıma aktarımı için kapsamlı yazılım araçlarının geliştirilmesi ile sürekli öğrenme, pekiştirmeli öğrenme ve çoklu veri türlerinin birlikte işlendiği uygulama alanlarında yeni çalışmalar yapılması beklenmektedir.

## 7. Sonuç

Bu çalışma, SNN'lerin görüntü işleme alanında araştırılan yaklaşımlar arasında yer aldığını ve farklı uygulamalarda değerlendirildiğini göstermektedir. Biyolojik sinir sistemlerinden esinlenen olay tabanlı ve enerji verimli hesaplama yapıları sayesinde SNN'ler, kenar belirleme, görüntü iyileştirme ve görüntü sınıflandırma gibi çeşitli görevlerde başarılı sonuçlar sunabilmektedir. Özellikle düşük güç tüketimi gerektiren uç bilişim uygulamaları açısından dikkat çekici bir alternatif oluşturmaktadır. SNN'lerin en önemli avantajları arasında seyrek ve asenkron hesaplama yapısı ile nöromorfik sensörlerle uyumlu çalışabilme yeteneği yer almaktadır. Bu özellikler, otonom sistemler, robotik ve nesnelerin interneti gibi kaynak kısıtlı ortamlarda kullanımını desteklemektedir. Bununla birlikte, eğitim yöntemlerindeki zorluklar, standart veri setleri ve karşılaştırma ölçütlerinin sınırlı olması gibi bazı sorunlar hâlen araştırma konusudur. Gelecekte algoritma–donanım birlikte tasarımı, yeni spiking ağ mimarileri ve daha gelişmiş eğitim yöntemleri üzerine yapılacak çalışmaların SNN'lerin kullanım alanlarını genişletmesi beklenmektedir. Ayrıca nöromorfik sensörlerle bütünleşik çalışan sistemlerin geliştirilmesi, enerji verimli yapay zekâ uygulamalarının ilerlemesine katkı sağlayabilir. Bu yönleriyle SNN'ler geleceğin görüntü işleme sistemleri için önemli bir araştırma alanı olarak değerlendirilmektedir.

## Kaynaklar

- Clogenson, M., Kerr, D., McGinnity, T. M., Coleman, S. A., & Wu, Q. (2011). Biologically inspired edge detection using spiking neural networks and hexagonal images. In *International Conference on Neural Computation Theory and Applications* (pp. 381-384). SciTePress.
- Davies, M., Srinivasa, N., Lin, T. H., Chinya, G., Cao, Y., Choday, S. H., ... & Wang, H. (2018). Loihi: A neuromorphic manycore processor with on-chip learning. *Ieee micro*, 38(1), 82-99.
- Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology*, 117(4), 500.
- Iaboni, C., & Abichandani, P. (2024). Event-based spiking neural networks for object detection: a review of datasets, architectures, learning rules, and implementation. *IEEE Access*, 12, 180532-180596.
- Izhikevich, E. M. (2003). Simple model of spiking neurons. *IEEE Transactions on neural networks*, 14(6), 1569-1572.
- İncetaş, M. O. (2022). Anisotropic diffusion filter based on spiking neural network model. *Arabian Journal for Science and Engineering*, 47(8), 9849-9860.
- İncetaş, M. O., Kılıçaslan, M., & Farshi, T. R. (2022). Image retrieval with SNN-based multi-level thresholding. *Gümüşhane Üniversitesi Fen Bilimleri Dergisi*, Özel Sayı, 98-108.
- İncetaş, M. O. (2023). Image interpolation based on spiking neural network model. *Applied Sciences*, 13(4), 2438.
- İncetaş, M. O., & Arslan, R. U. (2019). Edge Detection Using Integrate and Fire Neuron. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 23(2), 611-616.
- İncetaş, M. O., & Arslan, R. U. (2025). Spiking neural network-based edge detection model for content-based image retrieval. *Signal, Image and Video Processing*, 19(2), 169.
- İncetaş, M. O., & Kılıçaslan, M. (2025). Image watermarking based on spiking neural networks. *Cluster Computing*, 28(11), 736.
- Kerr, D., Coleman, S., & McGinnity, T. M. (2018). Biologically inspired intensity and depth image edge extraction. *IEEE Transactions on Neural Networks and Learning Systems*, 29(11), 5356-5365.
- Kerr, D., Coleman, S. A., McGinnity, T. M., & Clogenson, M. (2013, August). Biologically inspired intensity and range image feature extraction. In *The 2013 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE.
- Kerr, D., McGinnity, T. M., Coleman, S., & Clogenson, M. (2015). A biologically inspired spiking model of visual processing for image feature detection. *Neurocomputing*, 158, 268-280.

- Kılıçaslan, M. (2024). Image interpolation with spiking neural network based pixel similarity. *Signal, Image and Video Processing*, 18(10), 6925-6936.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.
- Lapicque, L. (1907). Recherches quantitatives sur l'excitation électrique des nerfs traitée comme une polarisation. *Journal de physiologie et de pathologie générale*, 9, 620-635.
- Lichtsteiner, P., Posch, C., & Delbruck, T. (2008). A 128 $\times$ 128 120 dB 15 $\mu$ s latency asynchronous temporal contrast vision sensor. *IEEE journal of solid-state circuits*, 43(2), 566-576.
- Maass, W. (1997). Networks of spiking neurons: the third generation of neural network models. *Neural networks*, 10(9), 1659-1671.
- Merolla, P. A., Arthur, J. V., Alvarez-Icaza, R., Cassidy, A. S., Sawada, J., Akopyan, E., ... & Modha, D. S. (2014). A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science*, 345(6197), 668-673.
- Niu, L. Y., Wei, Y., Liu, W. B., Long, J. Y., & Xue, T. H. (2023). Research progress of spiking neural network in image classification: A review. *Applied intelligence*, 53(16), 19466-19490.
- Saifei, W. U., Yuan, Z., Di, X. I. E., Hai, Y. U., & Jiang, Z. H. U. (2025). Comprehensive survey of spiking neural networks for visual object detection: from biological mechanisms to state-of-the-art applications. *Journal of Functional Materials and Devices*, 31(5), 364-378.
- Thorpe, S. J., & Imbert, M. (1989). Biological constraints on connectionist modelling. *Connectionism in perspective*, 63-92.
- Voudaskas, M., MacLean, J. I., Dutton, N. A., Stewart, B. D., & Gyongy, I. (2025). Spiking neural networks in imaging: A review and case study. *Sensors*, 25(21), 6747.
- Wu, Q., McGinnity, M., Maguire, L., Belatreche, A., & Glackin, B. (2007, August). Edge detection based on spiking neural network model. In *International Conference on Intelligent Computing* (pp. 26-34). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Yedjour, H., Meftah, B., Lézoray, O., & Benyettou, A. (2017). Edge detection based on Hodgkin–Huxley neuron model simulation. *Cognitive processing*, 18(3), 315-323.
- Zhang, A., Cao, H., Shan, N., Wang, J., Pu, M., & Song, Y. (2025). Spiking neural networks for object detection and semantic segmentation across event-driven and frame-based modalities: A review. *Intelligent Opto-Electronics*, 1(2), 250007-1.

## Steganografi Alanındaki Araştırma Eğilimleri: Ulusal Literatür Üzerine Tematik Bir İnceleme

Mürsel Ozan İncetas<sup>1</sup>

Murat Meriçelli<sup>2</sup>

### Özet

Steganografi, dijital ortamlarda gizli verilerin metin, görüntü, ses, video gibi taşıyıcı nesnelere içerisine gömülmesini temel alan bir bilgi gizleme yöntemidir. Kriptografiden farklı olarak, yalnızca mesajın içeriğini değil, iletişimin varlığını da gizlemeyi hedefler. Antik Yunan'dan günümüze uzanan köklü bir geçmişe sahip olan steganografi, dijital çağla birlikte siber güvenlik, adli bilişim ve telif hakkı koruma gibi alanlarda stratejik önem kazanmıştır. Bu çalışma, Türkiye'de steganografi alanında hazırlanmış lisansüstü tezlerin genel görünümünü, metodolojik eğilimlerini ve araştırma boşluklarını ortaya koymayı amaçlamaktadır. YÖK Ulusal Tez Merkezi'nde "steganografi" anahtar kelimesiyle taranan tezler PRISMA ilkeleri doğrultusunda değerlendirilmiş; tam metnine erişilen 111 tez (92 yüksek lisans, 19 doktora) betimsel analiz yöntemiyle incelenmiştir. Tezler; tür, yıl, üniversite, enstitü, anabilim dalı, danışman unvanı, sayfa sayısı, tablo-şekil kullanımı, taşıyıcı ortam, kullanılan yöntemler ve performans ölçütleri açısından analiz edilmiştir. Bulgular, tezlerin büyük çoğunluğunun Fen Bilimleri Enstitüleri bünyesinde ve bilgisayar mühendisliği ağırlıklı programlarda yürütüldüğünü göstermektedir. En çok tercih edilen yöntem LSB (45 tez) iken, kriptografik algoritmalarından AES (10 tez) öne çıkmaktadır. Performans değerlendirmesinde PSNR (35 tez) en sık kullanılan ölçüttür. Görüntü steganografisi baskın olmakla birlikte metin, ses, video ve ağ ortamlarında da çalışmalar mevcuttur. Steganaliz içeren tezlerin sınırlı sayıda kalması, alanın gizleme odaklı bir perspektifle şekillendiğini göstermektedir. Yapay zekâ tabanlı yöntemlerin (CNN, GAN) ise henüz yeterli düzeyde temsil edilmediği tespit edilmiştir. Sonuç olarak,

- 1 Doç. Dr., Alanya Alaaddin Keykubat Üniversitesi, ALTSO MYO, Bilgisayar Teknolojileri Bölümü, ozan.incetas@alanya.edu.tr, ORCID: 0000-0002-1016-1655
- 2 Dr. Öğr. Üyesi, Alanya Alaaddin Keykubat Üniversitesi, ALTSO MYO, Bilgisayar Teknolojileri Bölümü, murat.mericelli@alanya.edu.tr, ORCID: 0000-0003-0168-3221

Türkiye’de steganografi alanındaki lisansüstü tezler teknik ve uygulama odaklı bir çizgide ilerlemekte; ancak frekans alanı teknikleri, yapay zekâ uygulamaları ve steganaliz konularında metodolojik çeşitliliğin artırılması gerekmektedir. Bu çalışmanın, alanda çalışacak araştırmacılara yol gösterici bir kaynak olması beklenmektedir.

## 1. Giriş

Dijitalleşen dünyada bilginin stratejik bir öneminin olması, verinin güvenli bir şekilde iletilmesini ve depolanmasını zorunlu kılmıştır. Bilgi güvenliğinin temellerinden biri olan gizlilik kavramı, verinin yetkisiz kişilerin erişimine karşı korunmasını ifade eder (Stamp, 2011; Whitman & Mattord, 2009). Gizlilik, yalnızca verinin içeriğinin okunmasını engellemekle sınırlı değildir; bazı durumlarda iletişimin varlığının dahi gizli tutulması kritik bir gereksinim hâline gelmektedir.

Bilgi gizleme kavramı içerisinde yer alan kriptografi ve steganografi, benzer amaçlara hizmet etseler de yaklaşım ve yöntem bakımından birbirlerinden farklıdır. Kriptografi, mesajı matematiksel işlemlerle şifreleyerek üçüncü taraflar için anlamsız duruma getirirken, mesajın varlığını gizlememektedir (Buchmann & Buchamann, 2004). Steganografi ise mesajı bir taşıyıcı nesnenin içerisine gömerek, bizzat bilginin varlığını dahi gizlemeyi amaçlamaktadır (Kessler & Hosmer, 2011). Bu yönüyle steganografinin temelini oluşturan fark edilemezlik ilkesi, yöntemin tercih edilmesinde belirleyici olmaktadır (Mishra & Bhanodiya, 2015; Wang & Wang, 2004). Kriptografik bir veri potansiyel bir saldırgan için doğrudan bir hedef oluştururken, steganografik veri sıradan bir dijital içerik gibi algılandığından şüphe uyandırmamaktadır. Günümüzde siber güvenlik, istihbarat, dijital adli bilişim ve telif hakkı koruma uygulamaları gibi birçok alanda steganografinin önemi giderek artmaktadır (Dalal & Juneja, 2021).

Steganografinin kökenleri dijital çağdan çok önceye dayanmaktadır (Alabdali & Alzahrani, 2021; Kahn, 1996; Kessler & Hosmer, 2011). Antik Yunan’da Herodot’un aktardığı, mesajın bir kölenin kazınmış kafa derisine yazılarak gizlenmesi ya da balmumu tabletlerin altına kazınan metinler, bilginin fiziksel ortamlar aracılığıyla gizlenmesine yönelik erken dönem uygulamalar olarak kabul edilmektedir. Orta Çağ ve Rönesans dönemlerinde görünmez mürekkepler ve sanatsal eserler içerisine gizlenen semboller, steganografinin hem askeri hem de kültürel bağlamda kullanıldığını göstermektedir. II. Dünya Savaşı sırasında geliştirilen mikro nokta teknolojisi ise, bilginin fark edilmeden taşınması açısından steganografinin stratejik önemini ortaya koyan önemli bir dönüm noktası olmuştur.

Dijital çağın başlamasıyla birlikte steganografi, fiziksel ortamlardan dijital ortamlara taşınmış; taşıyıcı nesnelere olarak metin, görüntü, ses, video ve ağ paketleri kullanılmaya başlanmıştır (Bhattacharyya, 2011; Kishor vd., 2016). Dijital görüntülerde piksellerin en az anlamlı bitleri, ses dosyalarındaki frekans bileşenleri veya ağ protokollerinin başlık alanları, veri gizleme amacıyla yaygın biçimde kullanılan ortamlar hâline gelmiştir. Bu çeşitlilik, steganografi çalışmalarında kullanılan yöntemlerin de uzamsal alan, frekans alanı ve adaptif yaklaşımlar gibi farklı teknik sınıflar altında ele alınmasını beraberinde getirmiştir (Dhawan & Gupta, 2021; Majeed vd., 2021).

Steganografik yöntemlerin değerlendirilmesinde yalnızca gizleme işleminin başarıyla gerçekleştirilmesi yeterli görülmemekte; gizlenen verinin algılanabilirliği, taşıyıcı nesne üzerindeki bozulma düzeyi ve sistemin saldırılara karşı dayanıklılığı gibi ölçütler de önem kazanmaktadır. Bu bağlamda, literatürde PSNR, MSE, SSIM ve BER gibi performans ölçütleri yaygın olarak kullanılmaktadır (Beram, 2014; Malik vd., 2025; Setiadi, 2021). Bununla birlikte, steganografi sistemlerinin güvenlik düzeyinin bütüncül biçimde değerlendirilebilmesi için, gizleme yöntemlerinin steganaliz karşısındaki dirençlerinin de incelenmesi gerekmektedir. Steganaliz, gizli verinin varlığını tespit etmeyi amaçlayan yaklaşımları kapsamakta olup, steganografi ve steganaliz arasındaki ilişki alanının metodolojik olgunluğunu belirleyen temel unsurlardan biri olarak değerlendirilmektedir (Dalal & Juneja, 2021; Michaylov & Sarmah, 2025).

Bu çalışmada, steganografi alanında hazırlanmış lisansüstü tezler; kullanılan taşıyıcı ortamlar, tercih edilen yöntem ve teknikler, performans değerlendirme ölçütleri ve steganaliz uygulamalarının varlığı açısından sistematik biçimde incelenmiştir. Böylece alanın Türkiye özelindeki metodolojik eğilimlerinin ortaya konulması ve steganografi çalışmalarının hangi eksenler etrafında şekillendiğinin belirlenmesi amaçlanmıştır.

## 2. Yöntem

### 2.1 Araştırma Deseni

Bu çalışma, Türkiye’de steganografi alanında hazırlanmış lisansüstü tezlerin genel görünümünü ortaya koymayı amaçlayan, betimsel nitelikte bir doküman incelemesi olarak tasarlanmıştır. Araştırma kapsamında, farklı üniversite ve enstitülerde yürütülen yüksek lisans ve doktora tezleri sistematik bir yaklaşımla incelenmiş; alanın zamansal gelişimi, kurumsal dağılımı ve yapısal özellikleri bütüncül bir çerçevede ele alınmıştır. Çalışmada, incelenen tezlerden elde edilen nicel veriler betimsel istatistikler kullanılarak analiz edilmiş; tez türü, yayın yılı, üniversite, anabilim dalı ve danışman unvanı gibi değişkenler üzerinden

alanın genel görünümünü ortaya konmuştur. Elde edilen bulgular, Türkiye’de steganografi alanındaki akademik üretimin mevcut durumunu betimlemenin yanı sıra, alandaki araştırma yoğunluklarını ve gelişim eğilimlerini görünür kılmayı amaçlamaktadır. Bu doğrultuda çalışma, hem alan yazına bütüncül bir bakış sunmayı hem de gelecekte yapılacak lisansüstü araştırmalar için yol gösterici bir kaynak oluşturmayı hedeflemektedir.

## **2.2 Veri Kaynağı**

Araştırmanın veri kaynağını, Yükseköğretim Kurulu Ulusal Tez Merkezi’nde yer alan lisansüstü tezler oluşturmaktadır (YÖK, 2025). Tezlerin belirlenmesi sürecinde, özet bölümünde “steganografi” ifadesine yer verilen çalışmalar esas alınmış ve bu ölçüt doğrultusunda kapsamlı bir tarama gerçekleştirilmiştir. Bu yaklaşım, çalışmanın doğrudan steganografi alanına odaklanmasını sağlamak ve konu dışı tezlerin araştırma kapsamı dışında bırakılmasını temin etmek amacıyla tercih edilmiştir.

Gerçekleştirilen tarama sonucunda toplam 115 lisansüstü tez tespit edilmiştir. Tezlerin dâhil edilme ve hariç tutulma süreci, sistematik derleme çalışmalarında yaygın olarak kabul gören raporlama standartlarını belirleyen PRISMA ilkeleri çerçevesinde yürütülmüştür. Bu kapsamda, tam metnine erişim sağlanamayan 4 tez çalışma dışı bırakılmış; nihai analizler 111 tez üzerinden gerçekleştirilmiştir. Çalışmaya dâhil edilen tezler; tez türü, yayın yılı, üniversite, enstitü, anabilim dalı, danışman akademik unvanı, tezlerin toplam sayfa sayısı ile içerik yapısını yansıtan tablo ve şekil sayıları gibi ölçütler açısından ayrıntılı biçimde incelenmiştir. Elde edilen nicel veriler, alanın zamansal gelişimini, kurumsal dağılımını ve yapısal özelliklerini ortaya koymak amacıyla betimsel istatistikler kullanılarak analiz edilmiştir.

## **2.3. Veri Toplama Süreci**

Araştırma kapsamında incelenen tezlere ilişkin bibliyografik ve yapısal veriler, Yükseköğretim Kurulu Ulusal Tez Merkezi üzerinden erişilen tezlerin tam metinleri ve özet bölümleri incelenerek manuel olarak toplanmıştır (YÖK, 2025). Veri toplama sürecinde, her bir tezin sistematik ve tutarlı biçimde değerlendirilmesini sağlamak amacıyla önceden belirlenmiş değişkenler esas alınmıştır. Bu doğrultuda; tez türü, yayın yılı, üniversite, enstitü, anabilim dalı, danışman akademik unvanı, toplam sayfa sayısı ile tezlerde yer alan tablo ve şekil sayıları gibi ölçütleri içeren bir veri kayıt formu hazırlanmıştır. İlgili bilgiler, her tez için ayrı ayrı olmak üzere doğrudan tez metinlerinden kontrol edilerek forma işlenmiştir. Veri girişleri sırasında olası hata ve tutarsızlıkların önüne geçmek amacıyla, eksik ya da belirsiz bilgiler tekrar gözden geçirilmiş ve gerekli durumlarda tezler çapraz olarak yeniden incelenmiştir. Tezlere

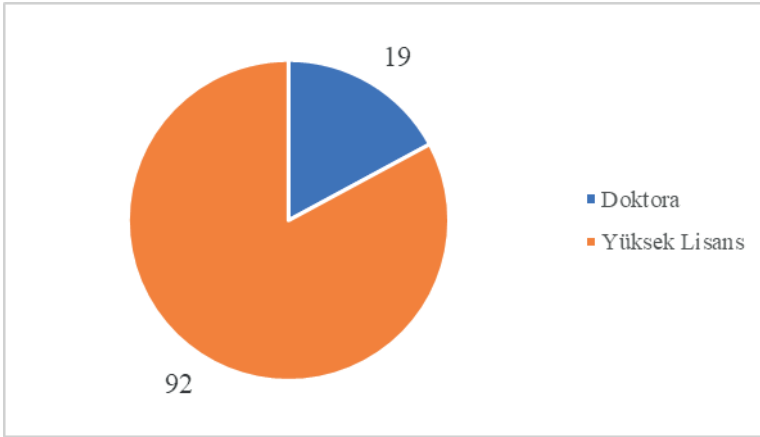
ait anabilim dalı ve bölüm bilgileri, orijinal haliyle kaydedilmiş ve analiz aşamasında karşılaştırılabilirlik sağlamak amacıyla yeniden sınıflandırılmak üzere korunmuştur. Toplanan veriler, analiz sürecine uygun biçimde tablollaştırılmış ve betimsel istatistiksel işlemlere hazır hâle getirilmiştir. Bu süreçte, verilerin araştırmanın amaçlarıyla uyumlu, karşılaştırılabilir ve tekrarlanabilir olmasına özen gösterilmiştir.

#### **2.4. Veri Analizi**

Araştırma kapsamında elde edilen nicel veriler, betimsel istatistik yöntemleri kullanılarak analiz edilmiştir. Bu doğrultuda; tez türü, yayın yılı, üniversite, enstitü, anabilim dalı ve danışman unvanı gibi değişkenlere ilişkin veriler frekans ve yüzde değerleri üzerinden değerlendirilmiştir. Nicel bulguların daha anlaşılır ve karşılaştırılabilir biçimde sunulabilmesi amacıyla, elde edilen sonuçlar tablolar ve grafikler aracılığıyla görselleştirilmiştir. Tezlerin anabilim dalı ve bölüm adlarının üniversiteler arasında farklılık göstermesi nedeniyle, bu değişkene ilişkin veriler analiz aşamasında yeniden sınıflandırılmıştır. Bu kapsamda, içerik benzerliği ve disiplinler yakınlıktan yola çıkarak farklı adlandırmalara sahip ancak benzer akademik odağı paylaşan programlar ortak gruplar altında birleştirilmiştir.

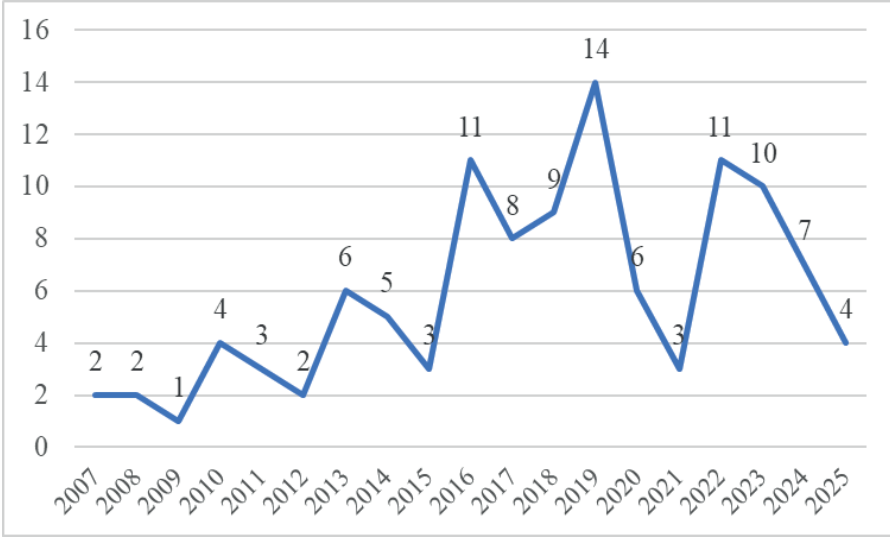
Nitel verilerin analizinde ise betimsel analiz yaklaşımı benimsenmiştir. Bu amaçla, tezlerin özet ve anahtar kelime bölümlerinde yer alan ifadeler incelenmiş; sık tekrar eden kavramlar belirlenerek kavramsal yoğunluklar ortaya konmuştur. Elde edilen kavramlar doğrultusunda oluşturulan anahtar kelime bulutu, steganografi alanındaki baskın temaların, yöntemsel eğilimlerin ve araştırma odaklarının görsel olarak yorumlanmasına imkân sağlamıştır. Analiz sürecinin son aşamasında, nicel ve nitel bulgular birlikte ele alınarak toplam bir değerlendirme yapılarak tartışılmıştır. Bu yaklaşım sayesinde, Türkiye’de steganografi alanında yürütülen lisansüstü çalışmaların mevcut durumu ve gelişim yönelimleri kapsamlı biçimde ortaya konmuştur.

### 3. Bulgular



Şekil 1 Tez türüne göre dağılımı

Şekil 1’de tez türlerine göre dağılım incelendiğinde, çalışmaya dâhil edilen lisansüstü tezlerin büyük çoğunluğunu yüksek lisans tezlerinin oluşturduğu görülmektedir. Toplam 111 tezin %82,9’u (n=92) yüksek lisans, %17,1’i (n=19) ise doktora tezlerinden oluşmaktadır. Bu dağılım, steganografi alanındaki akademik çalışmaların ağırlıklı olarak yüksek lisans düzeyinde yürütüldüğünü, doktora düzeyindeki çalışmaların ise daha sınırlı kaldığını göstermektedir. Doktora tezlerinin sayıca daha az olması, alanın henüz derinlemesine ve uzun soluklu araştırmalar açısından gelişim sürecinde olduğuna işaret edebileceği gibi, konunun çoğunlukla uygulama ve yöntem geliştirme odaklı ele alındığını da akla getirir. Şekil 2’ de tezlerin yıllara göre dağılımı görülmektedir.



Şekil 2 Yıllara göre tez sayısı

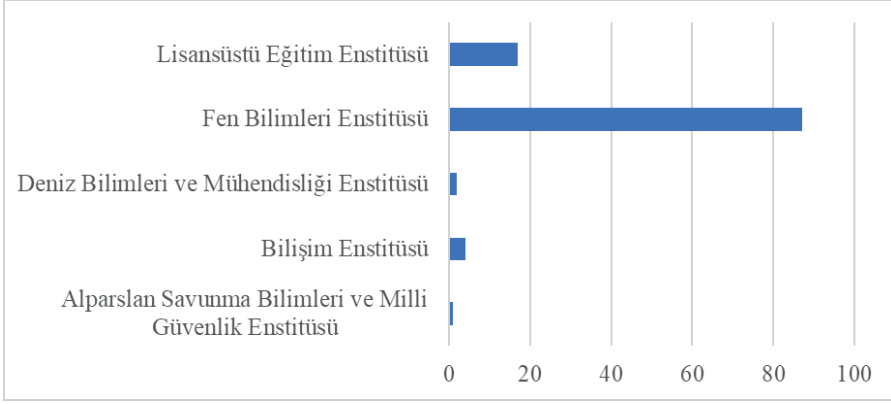
Şekil 2’de yer alan çizgi grafikten elde edilen bulgular, steganografi alanındaki lisansüstü çalışmaların zamansal gelişiminin üç temel evrede değerlendirilebileceğini göstermektedir. İlk evreyi oluşturan 2007–2012 yılları arasında tez sayılarının düşük seviyede seyretmesi, alanın Türkiye’de sınırlı sayıda araştırmacı tarafından ele alındığını ve henüz yaygınlaşmadığını düşündürmektedir. İkinci evre olarak değerlendirilebilecek 2013–2019 döneminde ise tez sayılarında belirgin bir artış eğilimi dikkat çekmektedir. Özellikle 2016 ve 2019 yıllarında gözlenen yükselişler, bilgi güvenliği, veri gizleme ve dijital medya teknolojilerindeki gelişmelerin steganografiye olan akademik ilgiyi artırmış olabileceğine işaret etmektedir. Üçüncü evrede (2020 ve sonrasında) tez sayılarında dalgalı bir yapı görülmekle birlikte, yıllık üretimin belirli bir seviyenin altına düşmemesi, alanın akademik gündemde kalmaya devam ettiğini göstermektedir. Bu dalgalanmanın, küresel ölçekte yaşanan pandemi süreci, araştırma önceliklerindeki değişimler, yapay zeka çalışmalarının ortaya çıkışı ve lisansüstü eğitim dinamikleriyle ilişkili olabileceği değerlendirilmektedir. Tablo 1’de üniversitelere göre tezlerin dağılımı yer almaktadır.

**Tablo 1 Üniversitelere Göre Tezlerin Dağılımı**

Üniversite	Program sayısı
Altınbaş Üniversitesi	11
Gazi Üniversitesi	10
Fırat Üniversitesi	8
Selçuk Üniversitesi	8
İstanbul Aydın Üniversitesi	6
Düzce Üniversitesi	5
Ankara Üniversitesi	4
İstanbul Ticaret Üniversitesi	4
Kocaeli Üniversitesi	4
Sakarya Üniversitesi	4
Trakya Üniversitesi	4
Çankaya Üniversitesi	3
Ankara Yıldırım Beyazıt Üniversitesi	2
Atılım Üniversitesi	2
Başkent Üniversitesi	2
Deniz Harp Okulu Komutanlığı	2
Gaziantep Üniversitesi	2
İstanbul Teknik Üniversitesi	2
Kadir Has Üniversitesi	2
Karadeniz Teknik Üniversitesi	2
Üsküdar Üniversitesi	2

Tablo 1’de, çalışmaların belirli üniversitelerde yoğunlaştığı görülmektedir. En yüksek tez sayısına sahip olan Altınbaş Üniversitesi (n=11) ve Gazi Üniversitesi (n=10), alanın kurumsal düzeyde en aktif üretim merkezleri olarak öne çıkmaktadır. Bu üniversiteleri Fırat Üniversitesi ve Selçuk Üniversitesi (n=8) takip etmektedir. Bu dağılım, steganografi alanındaki akademik üretimin hem devlet hem de vakıf üniversitelerinde var olduğunu göstermektedir. Ayrıca, tezlerin farklı üniversitelere dağılmış olması, alanın belirli merkezlerle sınırlı kalmadığını; Türkiye genelinde farklı akademik birimlerde çalışıldığını ortaya koymaktadır. Çok sayıda tezin ise belirli üniversitelerde kümelenmesi, bu kurumlarda konuya ilişkin akademik kadro birikimi, danışman uzmanlaşması veya lisansüstü programların teknik altyapı yeterlilikleriyle ilişkili olabilir. Ek olarak, dağılımın homojen olmaması, steganografi alanında kurumsal uzmanlaşmanın henüz sınırlı sayıda üniversitede yoğunlaştığını göstermektedir.

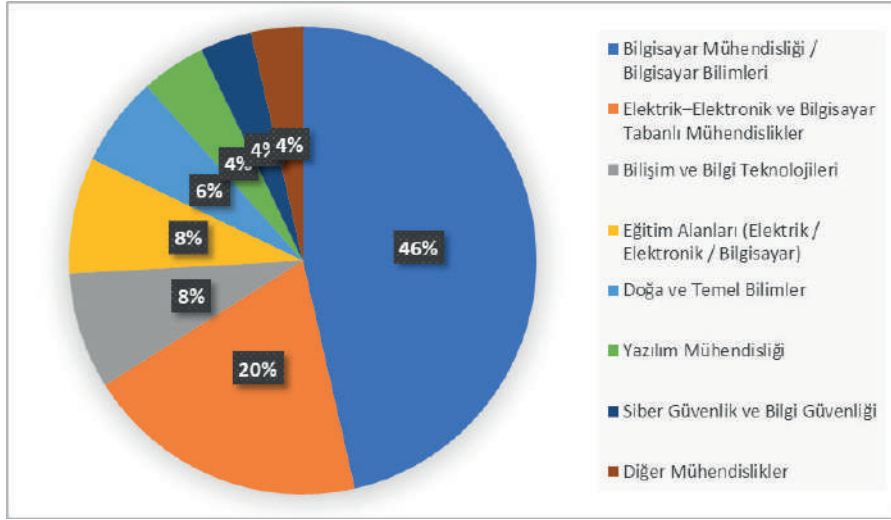
Bu durum, alanın Türkiye’de gelişmekte olan ancak belirli araştırma merkezlerinde derinleşen bir yapıya sahip olduğunu düşündürmektedir. Şekil 3’te üniversitelere göre tezlerin dağılımı gösterilmektedir.



Şekil 3 Enstitülere göre tezlerin dağılımı

Şekil 3 incelendiğinde, steganografi alanındaki lisansüstü tezlerin çok büyük bir bölümünün fen bilimleri enstitüsü bünyesinde hazırlandığı görülmektedir (n=87). Bu sayı, toplam tezlerin açık ara çoğunluğunu oluşturmaktadır. Alanın teknik ve mühendislik temelli bir disiplin olarak konumlandığını açık biçimde ortaya koymaktadır. Bunu, 17 tez ile lisansüstü eğitim enstitüsü izlemektedir. Lisansüstü Eğitim Enstitülerinin son yıllarda birçok üniversitede farklı enstitülerin birleştirilmesiyle oluşturulduğu dikkate alındığında, bu sayı yapısal dönüşümün bir yansıması olarak değerlendirilebilir.

Bilişim Enstitüsü (n=4) ve Deniz Bilimleri ve Mühendisliği Enstitüsü (n=2) bünyesinde hazırlanan tezler ise alanın belirli alt uzmanlık alanlarında da çalışıldığını göstermektedir. En düşük tez sayısının Alparslan Savunma Bilimleri ve Milli Güvenlik Enstitüsü bünyesinde görülmesi (n=1) ise daha sınırlı kaldığını göstermektedir. Bu durum, gelecekte disiplinler arası çalışmalar için potansiyel bir gelişim alanı olarak değerlendirilebilir. Şekil 4’te alanlara göre tezlerin dağılımı yer almaktadır.

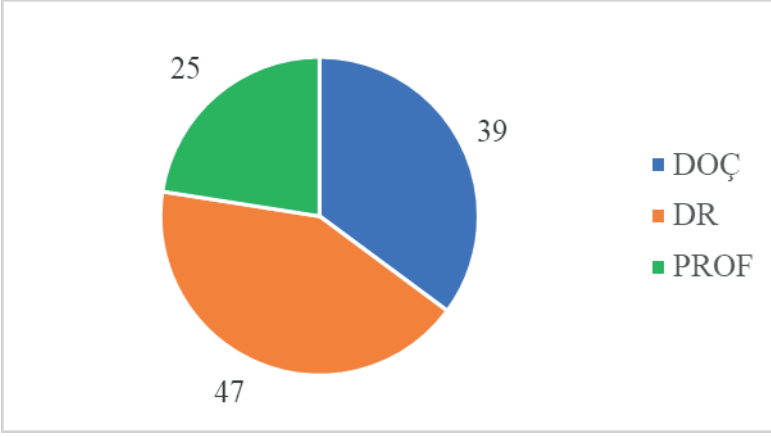


Şekil 4 Alanlara göre tezlerin dağılımı

Şekil 4 ele alındığında, tezlerin büyük ölçüde bilgisayar temelli mühendislik disiplinlerinde yoğunlaştığı anlaşılmaktadır. Toplam 111 tezin 52'si (%46,8) Bilgisayar Mühendisliği / Bilgisayar Bilimleri grubunda yer almaktadır. Bu oran, alanın neredeyse yarısının doğrudan bilgisayar bilimi paradigması içinde üretildiğini göstermektedir. Bunu 22 tez (%19,8) ile Elektrik-Elektronik ve Bilgisayar Tabanlı Mühendislikler grubu takip etmektedir. Bu bulgu, steganografinin yalnızca yazılım veya algoritma düzeyinde değil; sinyal işleme, gömülü sistemler ve haberleşme temelli mühendislik alanlarıyla da güçlü bir ilişki içinde olduğunu göstermektedir. Bu iki grup birlikte değerlendirildiğinde, toplam tezlerin yaklaşık %66'sının doğrudan mühendislik ve teknik altyapı odaklı disiplinlerde üretildiği anlaşılmaktadır. Bu durum, steganografinin Türkiye'de ağırlıklı olarak teknik, algoritmik ve uygulama temelli bir araştırma alanı olarak konumlandığını ortaya koymaktadır.

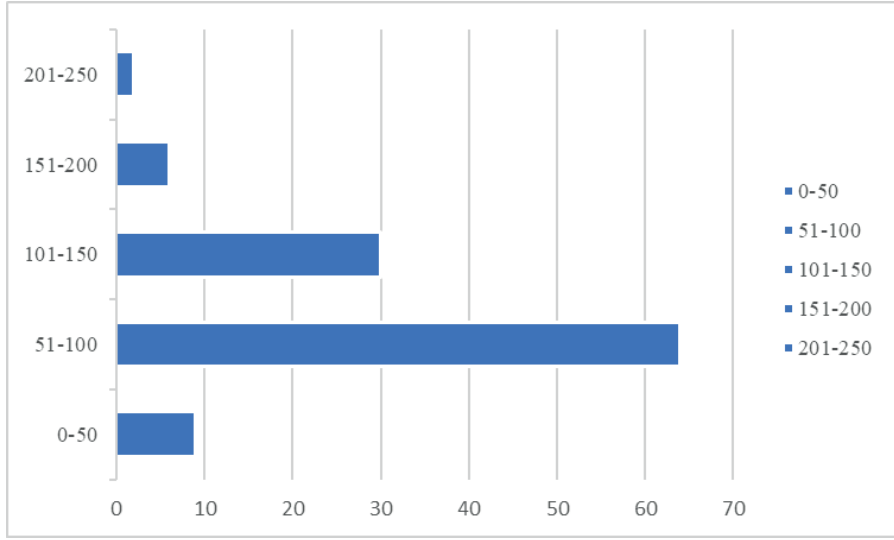
Bilişim ve Bilgi Teknolojileri (%8,1) ile Eğitim Alanları (Elektrik / Elektronik / Bilgisayar) (%8,1) gruplarının benzer oranlara sahip olması dikkat çekicidir. Özellikle eğitim temelli programlarda üretilen tezler, alanın pedagojik veya uygulamalı öğretim boyutunun da ele alındığını göstermektedir. Buna karşılık, doğrudan Siber Güvenlik ve Bilgi Güvenliği grubunda yer alan tezlerin oranının %3,6 gibi görece düşük bir seviyede kalması önemli bir bulgudur. Steganografi doğası gereği bilgi gizleme ve güvenli iletişimle doğrudan ilişkili olmasına rağmen, çalışmaların çoğunun "siber güvenlik" etiketi altında değil, daha çok bilgisayar mühendisliği şemsiyesi altında yürütülmesi dikkat çekmektedir. Alanın kurumsal olarak güvenlik disiplininden ziyade

teknik mühendislik disiplinleri içinde konumlandığı düşünülebilir. Doğa ve Temel Bilimler (%6,3) ile Diğer Mühendislikler (%3,6) gruplarındaki tez sayılarının sınırlı olması, steganografinin matematiksel kuramsal çerçevede veya disiplinlerarası modelleme bağlamında henüz güçlü bir temsil alanı bulamadığını işaret eder. Özellikle matematik ve hesaplamalı bilimler temelli çalışmaların düşük oranı, alanın teorik derinleşmeden ziyade uygulama geliştirme odaklı ilerlediğini ortaya koyar. Şekil 5'te danışman unvanına göre tezlerin dağılımı gösterilmektedir.



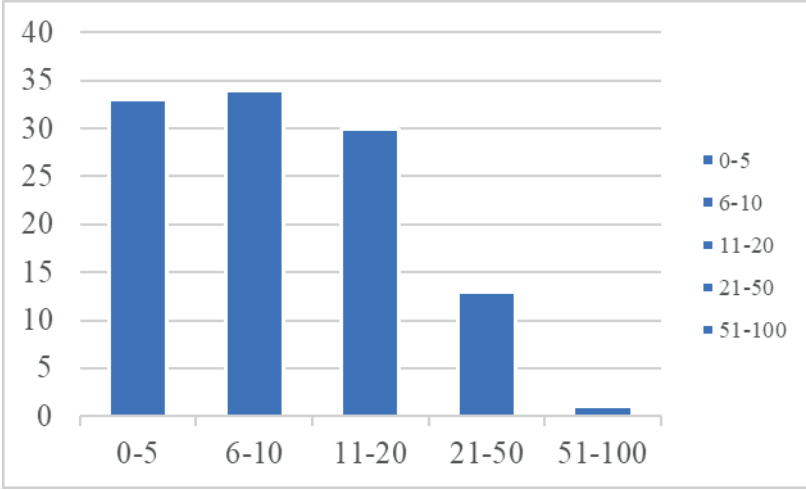
Şekil 5 Danışman unvanına göre tezlerin dağılımı

Şekil 5'e bakıldığında, 47 tezin doktor, 39 tezin doçent ve 25 tezin profesör unvanına sahip akademisyenler tarafından yürütüldüğü görülmektedir. Bilişim ve mühendislik temelli araştırma alanlarında, aktif proje üretimi ve uygulama geliştirme süreçlerinde daha yoğun rol alan Dr. ve Doç. unvanlarının tez danışmanlığında daha yüksek sayılara ulaşması yapısal bir durum olarak değerlendirilebilir. En yüksek sayının 47 tez ile Dr. grubunda yer alması, alanın uygulama ve teknik geliştirme odaklı niteliğiyle uyumludur. Profesör danışman sayısının 25 olması ise steganografinin görece yeni ve dinamik bir araştırma alanı olması nedeniyle orta kuşak akademisyenler tarafından daha yoğun sahiplenildiği şeklinde yorumlanabilir. Genel dağılım, alanın yalnızca belirli bir akademik unvan grubunun tekelinde olmadığını; farklı kıdem düzeylerinden akademisyenlerin katkısıyla sürdürüldüğünü göstermektedir. Şekil 6'da sayfa sayılarına göre tezlerin dağılımı yer almaktadır.



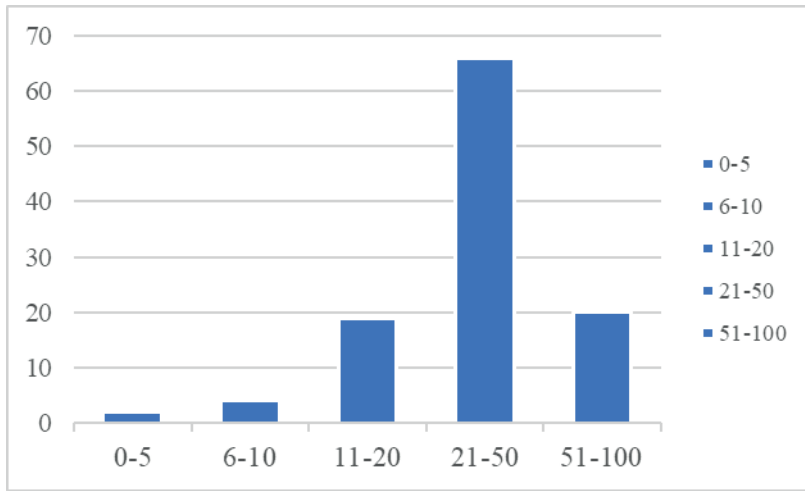
Şekil 6 Sayfa sayılarına göre tezlerin dağılımı

Şekil 6 incelendiğinde, çalışmaların büyük bölümünün 51–100 sayfa aralığında yoğunlaştığı görülmektedir ( $n=64$ ). Bu aralığı 101–150 sayfa aralığında yer alan 30 tez izlemektedir. Buna karşılık 0–50 sayfa aralığında 9, 151–200 sayfa aralığında 6 ve 201–250 sayfa aralığında yalnızca 2 tez bulunmaktadır. Bu dağılım, tezlerin ağırlıklı olarak orta hacimli çalışmalar şeklinde hazırlandığını göstermektedir. 51–100 sayfa aralığındaki belirgin yoğunluk, çalışmaların büyük bölümünün belirli bir yöntem, veri seti ve performans analizi çerçevesinde yapılandırıldığını düşündürmektedir. 101–150 sayfa aralığında yer alan tezler ise görece daha kapsamlı literatür taraması ve deneysel karşılaştırmalar içeren araştırmalara işaret etmektedir. 151 sayfa ve üzerindeki tezlerin sınırlı sayıda olması, alan çalışmalarının çoğunlukla belirli bir teknik yaklaşım veya uygulama problemi etrafında kurgulandığını göstermektedir. 201–250 sayfa aralığındaki düşük frekans ise oldukça geniş kapsamlı ve derinlemesine yapılandırılmış çalışmaların nispeten daha az üretildiğini ortaya koymaktadır. Şekil 7’de Tablo sayısına göre tezlerin dağılımı gösterilmektedir.



Şekil 7 Tablo sayısına göre tezlerin dağılımı

Şekil 7'ye bakıldığında, çalışmaların büyük bölümünün 0–10 tablo aralığında yoğunlaştığı görülmektedir. 6–10 tablo aralığında 34 tez, 0–5 tablo aralığında ise 33 tez bulunmaktadır. Bu durum, araştırmaların önemli bir kısmında tablo kullanımının sınırlı düzeyde tutulduğunu göstermektedir. 11–20 tablo aralığında yer alan 30 tez, orta yoğunlukta tablo kullanımına sahip çalışmaların da önemli bir yer tuttuğunu ortaya koymaktadır. Buna karşılık 21–50 tablo aralığında yalnızca 13 tez bulunması, yüksek yoğunluklu tablo kullanımının daha sınırlı olduğunu göstermektedir. 51–100 tablo aralığında sadece 1 tez yer alması ise aşırı yoğun tablo temelli raporlamanın istisnai olduğunu ortaya koymaktadır. Buna göre, steganografi alanındaki lisansüstü tezlerin çoğunlukla görsel veri sunumunu dengeli ve sınırlı bir çerçevede kullandığı; aşırı tablo yoğunluğuna dayalı büyük ölçekli veri analizlerinin ise oldukça nadir olduğu söylenebilir. Şekil 8'de şekil sayısına göre tezlerin dağılımı yer almaktadır.



*Şekil 8 Şekil sayısına göre tezlerin dağılımı*

Şekil 8 ele alındığında, çalışmaların büyük bölümünün 21–50 şekil aralığında yoğunlaştığı görülmektedir ( $n=66$ ). Bu aralık açık ara en yüksek frekansa sahiptir ve toplam tezlerin yarısından fazlasını oluşturmaktadır. 51–100 şekil aralığında yer alan 20 tez ise görsel yoğunluğu yüksek çalışmaların da önemli bir yer tuttuğunu göstermektedir. 11–20 şekil aralığında 19 tez bulunması, orta düzeyde görsel kullanımının da yaygın olduğunu ortaya koymaktadır. Buna karşılık 0–5 ( $n=2$ ) ve 6–10 ( $n=4$ ) aralıklarında yer alan tezlerin oldukça sınırlı sayıda olması, düşük görsel içerikli çalışmaların istisnai kaldığını göstermektedir. Bu dağılım, steganografi alanındaki lisansüstü tezlerin tablo kullanımındaki sınırlılığa karşın büyük ölçüde görsel temelli raporlamaya dayandığını ortaya koymaktadır. Tablo 2’de görüntü dışı steganografi ortamlarına göre tezlerin dağılımı yer almaktadır.

*Tablo 2 Görüntü Dışı Steganografi Ortamlarına Göre Tez Dağılımı*

Steganografi Ortamı	Tez Sayısı
Metin	8
Ses	8
Video	7
Ağ	8

Tablo 2 incelendiğinde, görüntü dışı steganografi ortamları arasında tez dağılımının oldukça dengeli olduğu görülmektedir. Metin ( $n=8$ ), ses ( $n=8$ ) ve ağ ( $n=8$ ) ortamları eşit sayıda çalışmaya konu olurken, video

ortamında 7 tez bulunmaktadır. Buradan anlaşılacağı üzere, araştırmacıların görüntü temelli steganografi dışında farklı veri taşıyıcı ortamlarına da benzer düzeyde ilgi gösterdiği görülmektedir. Özellikle metin, ses ve ağ tabanlı steganografi çalışmalarının eşit sayıda olması, bu alanların metodolojik ve uygulama açısından birbirine yakın araştırma potansiyeline sahip olduğunu düşündürmektedir. Video ortamındaki tez sayısının bir miktar daha düşük olması (n=7) ise bu alandaki teknik karmaşıklık, yüksek veri boyutu ve işlem maliyetleri gibi faktörlerle ilişkili olabilir; ancak dağılım genel olarak değerlendirildiğinde “ortamlar arasında belirgin bir fark yoktur. Tablo 3’te steganografi çalışmalarında kullanılan yöntemler ifade edilmektedir.

*Tablo 3 Steganografi Çalışmalarında Kullanılan Yöntemler*

Kategori	Yöntem	Tez Sayısı
Kriptografi	AES	10
	DES	4
	RSA	5
Steganografi Tekniği	LSB	45
	DCT	5
Yapay Zeka	CNN	2

Tablo 3 değerlendirildiğinde, steganografi çalışmalarının ağırlıklı olarak geleneksel gizleme teknikleri üzerine yoğunlaştığı görülmektedir. Özellikle LSB (Least Significant Bit) yönteminin açık ara en sık tercih edilen teknik olması, alanın önemli ölçüde uzamsal alan (spatial domain) temelli yaklaşımlar etrafında şekillendiğini göstermektedir. LSB’nin uygulama kolaylığı, düşük hesaplama maliyeti ve hızlı sonuç üretmesi, lisansüstü tez çalışmalarında tercih edilmesinde belirleyici faktörler olarak değerlendirilebilir. Bununla birlikte, bu durum aynı zamanda literatürde yöntemsel çeşitliliğin sınırlı kaldığına da işaret etmektedir. Frekans alanına dayalı tekniklerden DCT’nin daha sınırlı sayıda tezde yer alması, transform domain yaklaşımlarının görece daha teknik ve karmaşık yapısı ile ilişkilendirilebilir. Ancak literatürde DCT tabanlı yöntemlerin dayanıklılık açısından avantajlı olduğu bilinmesine rağmen tezlerde sınırlı temsil edilmesi, araştırmaların daha çok uygulama odaklı ve temel düzeyde kaldığını düşündürmektedir.

Kriptografik algoritmaların (AES, DES, RSA) belirli sayıda tezde kullanılmış olması, steganografinin tek başına yeterli görülmediği ve çoğu çalışmada veri gizleme öncesinde ek bir güvenlik katmanı oluşturma eğiliminin bulunduğunu göstermektedir. Özellikle AES’in diğer kriptografik yöntemlere kıyasla daha fazla tercih edilmesi, güncel güvenlik standartlarıyla uyumlu ve

daha güçlü kabul edilen algoritmaların akademik çalışmalara yansıdığını ortaya koymaktadır. DES'in daha sınırlı kullanımını ise, algoritmanın güncel güvenlik gereksinimlerini karşılamada zayıf kalmasıyla açıklanabilir.

Yapay zekâ tabanlı yöntemlerin, özellikle CNN temelli yaklaşımların, sınırlı sayıda tezde yer alması ise alanın henüz dönüşüm sürecinde olduğunu göstermektedir. Derin öğrenme tekniklerinin hem gizleme hem de steganaliz süreçlerinde artan önemi dikkate alındığında, mevcut dağılım geleneksel yöntemlerin hâlen baskın olduğunu, ancak yeni nesil yaklaşımların akademik literatüre giriş yapmaya başladığını ortaya koymaktadır. Tablo 4'te steganografi çalışmalarında kullanılan performans ölçütleri yer almaktadır.

*Tablo 4 Steganografi Çalışmalarında Kullanılan Performans Ölçütleri*

Performans Ölçütü	Tez Sayısı (n)
PSNR	35
MSE	18
SSIM	15
BER	11

Tablo 4 incelendiğinde, tezlerde en sık kullanılan performans ölçütünün PSNR olduğu görülmektedir (n=35). Bu metriği MSE (n=18) ve SSIM (n=15) izlemektedir. BER ise 11 tezde kullanılmıştır. Bu durum, çalışmaların ağırlıklı olarak görüntü kalitesini ölçmeye odaklandığını göstermektedir. PSNR'nin açık ara önde olması, özellikle görüntü steganografisi çalışmalarında gizleme sonrasında oluşan kalite kaybını sayısal olarak ifade etme ihtiyacının belirleyici olduğunu düşündürmektedir. MSE'nin de görece yüksek bir kullanım oranına sahip olması, piksel düzeyindeki hata analizinin hâlâ temel değerlendirme araçlarından biri olduğunu göstermektedir. SSIM'in daha sınırlı sayıda çalışmada yer alması ise algısal benzerliği dikkate alan ölçütlerin klasik hata temelli metriklerle kıyasla daha az tercih edildiğini ortaya koymaktadır. Bu durum, performans değerlendirmesinde geleneksel kalite metriklerinin hâlen baskın olduğunu göstermektedir. BER'in 11 tezde kullanılmış olması, dayanıklılık ve hata oranı analizlerinin tamamen ihmal edilmediğini; ancak kalite temelli ölçütlere kıyasla daha geri planda kaldığını düşündürmektedir. Özellikle saldırı, sıkıştırma veya iletim hataları gibi senaryolara karşı sistem performansını ölçen çalışmaların sayısının görece sınırlı olduğu söylenebilir. Şekil 8'de anahtar kelime bulutu görünümü yer almaktadır. Şekil 9'da anahtar kelime bulutu görülmektedir.



PSNR ve MSE gibi metriklerin görünürlüğü ise çalışmaların algılanamazlık ve bozulma düzeyini nicel olarak kanıtama kaygısını yansıttığını göstermektedir.

Üçüncü önemli katman, “Kriptografi”, “Şifreleme”, “AES”, “DES” ve “RSA” gibi kavramlarla temsil edilen güvenlik merkezli yaklaşımlardır. Bu küme, steganografinin çoğu zaman kriptografik yöntemlerle birlikte veya hibrit yapılar içinde ele alındığını göstermektedir. Gizli bilginin yalnızca saklanması değil, aynı zamanda içeriğinin de korunması hedeflenmekte; bu durum steganografi–kriptografi bütünleşmesinin güçlü bir araştırma yönelimi olduğunu ortaya koymaktadır.

Ayrıca “Ses Steganografisi”, “Görüntü Steganografisi” ve “Video Steganografisi” kavramlarının birlikte görünmesi, çalışmaların tek bir medya türüyle sınırlı kalmadığını; farklı taşıyıcı ortamlar üzerinden gizleme tekniklerinin karşılaştırmalı veya paralel biçimde ele alındığını göstermektedir. Ancak bu kümelenme içinde görüntü steganografisinin daha baskın olması, pratikte görsel verilerin hâlen en yaygın ve tercih edilen kapak ortamı olduğunu düşündürmektedir.

#### 4. Tartışma ve Sonuç

Bu çalışmada, Türkiye’de steganografi alanında hazırlanmış lisansüstü tezler; tür, yıl, üniversite, enstitü, çalışma alanı, danışman unvanı, sayfa sayısı, tablo-şekil kullanımı, kullanılan yöntemler, taşıyıcı ortamlar ve performans ölçütleri açısından sistematik olarak incelenmiştir. Sonuçlar, alanın genel görünümünü ortaya koymanın yanı sıra metodolojik eğilimler ve araştırma boşlukları hakkında da önemli ipuçları sunmaktadır.

Ulaşılan sonuçlar, alanın metodolojik bir olgunluğa eriştiğini ancak belirli teknik kalıpların dışına çıkmakta henüz direnç gösterdiğini ortaya koymaktadır. Tezlerin büyük çoğunluğunun yüksek lisans düzeyinde olması (%82,9), steganografinin Türkiye’de doktora düzeyinde henüz yeterli derinliğe ulaşmadığını göstermektedir. Doktora düzeyindeki çalışmaların %17,1 gibi sınırlı bir oranda kalması, alanın kuramsal derinleşme ve uzun soluklu araştırma projeleri üretme kapasitesinin henüz gelişim aşamasında olduğuna işaret etmektedir. Yıllara göre dağılım, 2013 sonrasında belirgin bir artış olduğunu, ancak 2020’den itibaren dalgalı bir seyir izlendiğini ortaya koymaktadır. Bu durum, küresel ölçekte yaşanan pandemi süreci, araştırma önceliklerindeki değişimler ve yapay zekâ çalışmalarının yükselişiyle ilişkili olabilir (Invernici vd., 2024; Singh, 2025).

Üniversite bazlı dağılım, Altınbaş ve Gazi üniversitelerinin öne çıktığını, ancak üretimin tekelleşmediğini göstermektedir. Fen Bilimleri Enstitüleri’nin açık ara önde olması, steganografinin Türkiye’de ağırlıklı olarak mühendislik

temelli bir disiplin olarak konumlandığını doğrulamaktadır. Buna karşın, sosyal bilimler veya disiplinler arası programlarda yürütülen tezlerin yok denecek kadar az olması, alanın teknik yönünün baskınlığını pekiştirmektedir.

Steganografinin kurumsal olarak “Siber Güvenlik” programlarından ziyade (%3,6), doğrudan “Bilgisayar Mühendisliği” (%46,8) çalışma alanı altında ele alınmaktadır. Bu durum, steganografinin Türkiye’deki akademik algıda bir savunma ya da güvenlik disiplininin çok, algoritmik bir veri işleme problemi olarak konumlandırıldığını ortaya koymaktadır. Doğa ve Temel Bilimler programlarındaki düşük temsil oranı ise, alanın matematiksel ve teorik modelleme boyutunun uygulama odaklı yaklaşımların gölgesinde kaldığını desteklemektedir.

Yöntemsel eğilimler açısından LSB tekniğinin (%40,5 ile 45 tez) ve PSNR (%31,5) ölçütünün baskınlığı, ulusal literatürün uygulama kolaylığı ve düşük maliyetli çözümlere yöneldiğini kanıtlamaktadır. LSB’nin kullanım avantajına rağmen, modern güvenlik ihtiyaçları karşısındaki zayıflığı bilinmektedir (Abdulhameed Alher vd., 2024; Tran vd., 2022). Buna karşın, uluslararası literatürde derin öğrenme tabanlı (CNN, GAN) steganografi ve steganaliz çalışmaları hızla artarken (Malik vd., 2025; Michaylov & Sarmah, 2025) Türkiye’deki tezlerde bu yöntemlerin oldukça sınırlı kalması, küresel akıllı steganografi eğilimlerinin ulusal literatüre entegrasyonunda kısmi bir gecikme yaşandığını göstermektedir. Öte yandan, steganografinin AES gibi algoritmalarla hibritlenmesi, verinin hem varlığını hem içeriğini korumaya yönelik güçlü bir güvenlik farkındalığının geliştiğine işaret etmektedir (Badhan & Malhi, 2024; Banoori vd., 2025). Taşıyıcı ortam çeşitliliği açısından metin, ses, video ve ağ steganografisi üzerine yapılan çalışmaların sayıca dengeli olması olumludur. Buna karşın, görüntü steganografisinin hâlâ baskın olması, araştırmacıların geleneksel medya türlerine yöneldiğini, güncel taşıyıcılar konusunda ise henüz yeterli çalışma yapılmadığını ortaya koymaktadır.

Çalışmanın en dikkat çekici sonuçlarından biri, gizleme tekniklerine odaklanan (Das vd., 2011), yoğun literatüre kıyasla steganaliz çalışmalarının geri planda kalmış olmasıdır. Literatürün sağlıklı bir dönüşüm sürecine girebilmesi için gizleme yöntemleri kadar, bu yöntemleri deşifre edecek savunma mekanizmalarının da akademik gündeme taşınması kritik bir gerekliliktir.

Sonuç olarak, Türkiye’de steganografi alanındaki lisansüstü tezler; teknik, deneysel ve performans odaklı bir çizgide ilerlemekte klasik yöntemler hâlâ baskınlığını korurken kriptografi ile bütünleşik yaklaşımlar da önemli bir yer tutmaktadır. Bununla birlikte, yapay zekâ tabanlı yöntemler, frekans alanı teknikleri, steganaliz odaklı çalışmalar ve disiplinler arası yaklaşımlar açısından önemli boşluklar bulunmaktadır (Alhomoud, 2021; Laishram & Tuithung,

2018). Gelecek arařtırmaların bu alanlara yönelmesi hem ulusal alanyazının zenginleşmesine hem de uluslararası düzeyde rekabet edebilirliđin artmasına katkı sağlayacaktır.

## Kaynaklar

- Abdulhameed Alher, Z., M Al Imran, B., & Al Ali, I. (2024). LSB as a steganography tool in information security. *5th International Conference on Communication Engineering and Computer Science (CIC-COCOS'24)*, 348-355.
- Alabdali, N., & Alzahrani, S. (2021). An overview of steganography through history. *Int. J. Sci. Eng. Sci*, 5, 41-44.
- Alhomoud, A. M. (2021). Image steganography in spatial domain: Current status, techniques, and trends. *Intelligent Automation & Soft Computing*, 27(1).
- Badhan, A., & Malhi, S. S. (2024). A Review on Hybrid Cryptography approach with Steganography. *2024 12th International Conference on Internet of Everything, Microwave, Embedded, Communication and Networks (IEMECON)*, 1-7.
- Banoori, S. Z., Khan, W., Rahman, S., Masood, F., Salam, A., Amin, F., De La Torre, I., Villar, M. G., Garay, H., & Choi, G. S. (2025). An improved hybrid image steganography method using AES algorithm. *Scientific Reports*.
- Beram, F. G. (2014). Effective parameters of image steganography techniques. *International Journal of Computer Applications Technology and Research*, 3(6), 361-363.
- Bhattacharyya, S. (2011). A survey of steganography and steganalysis technique in image, text, audio and video as cover carrier. *Journal of global research in computer science*, 2(4).
- Buchmann, J., & Buchamann, J. (2004). *Introduction to cryptography* (C. 335). Springer.
- Dalal, M., & Juneja, M. (2021). Steganography and Steganalysis (in digital forensics): a Cybersecurity guide. *Multimedia Tools and Applications*, 80(4), 5723-5771.
- Das, S., Das, S., Bandyopadhyay, B., & Sanyal, S. (2011). Steganography and Steganalysis: different approaches. *arXiv preprint arXiv:1111.3758*.
- Dhawan, S., & Gupta, R. (2021). Analysis of various data security techniques of steganography: A survey. *Information Security Journal: A Global Perspective*, 30(2), 63-87.
- Invernici, F., Bernasconi, A., & Ceri, S. (2024). Exploring the evolution of research topics during the COVID-19 pandemic. *Expert Systems with Applications*, 252, 124028.
- Kahn, D. (1996). The history of steganography. *International workshop on information hiding*, 1-5.
- Kessler, G. C., & Hosmer, C. (2011). An overview of steganography. *Advances in Computers*, 83, 51-107.

- Kishor, S. N., Ramaiah, G. N. K., & Jilani, S. A. K. (2016). A review on steganography through multimedia. *2016 International conference on research advances in integrated navigation systems (RAINS)*, 1-6.
- Laishram, D., & Tuithung, T. (2018). A survey on digital image steganography: current trends and challenges. *proceedings of 3rd international conference on internet of things and connected technologies (ICIOTCT)*, 26-27.
- Majeed, M. A., Sulaiman, R., Shukur, Z., & Hasan, M. K. (2021). A review on text steganography techniques. *Mathematics*, 9(21), 2829.
- Malik, K. R., Sajid, M., Almogren, A., Malik, T. S., Khan, A. H., Altameem, A., Rehman, A. U., & Hussien, S. (2025). A hybrid steganography framework using DCT and GAN for secure data communication in the big data era. *Scientific Reports*, 15(1), 19630.
- Michaylov, K. D., & Sarmah, D. K. (2025). Steganography and steganalysis for digital image enhanced Forensic analysis and recommendations. *Journal of Cyber Security Technology*, 9(1), 1-27.
- Mishra, R., & Bhanodiya, P. (2015). A review on steganography and cryptography. *2015 International Conference on Advances in Computer Engineering and Applications*, 119-122.
- Setiadi, D. R. I. M. (2021). PSNR vs SSIM: imperceptibility quality assessment for image steganography. *Multimedia tools and applications*, 80(6), 8423-8444.
- Singh, A. (2025). From Algorithms to AI: A Comprehensive Review of Core Concepts in Computer Science. *Global Research Repo*, 1(2), 129-153.
- Stamp, M. (2011). *Information security: principles and practice*. John Wiley & Sons.
- Tran, D. N., Zepernick, H.-J., & Chu, T. M. C. (2022). LSB data hiding in digital media: A survey. *EAI Endorsed Transactions on Industrial Networks and Intelligent Systems*, 1-50.
- Wang, H., & Wang, S. (2004). Cyber warfare: steganography vs. steganalysis. *Communications of the ACM*, 47(10), 76-82.
- Whitman, M. E., & Mattord, H. J. (2009). *Principles of information security*. Thomson Course Technology Boston, MA.
- YÖK. (2025). *YÖK Ulusal Tez Merkezi*. <https://tez.yok.gov.tr/UlusalTezMerkezi/>

## Çok Modlu Yapay Zekâda Görsel Dil Modelleri: Mimari Temeller ve Sektörel Uygulamalar

İrem Cakcak<sup>1</sup>

Burhan Duman<sup>2</sup>

### Özet

Yapay zekâ alanındaki son gelişmeler, metin ve görsel veriyi ortak bir temsil uzayında birleştirebilen Görsel-Dil Modellerinin (VLM) farklı disiplinlerde etkin biçimde kullanılmasını mümkün kılmıştır. Geleneksel tek-modlu görsel tanıma sistemlerinin ötesine geçen bu yenilikçi modeller; bağlamsal yorumlama, çok modlu muhakeme ve görev odaklı üretim süreçlerini destekleyen bütüncül yaklaşımlar sunmaktadır. Bu çalışma, Büyük Dil Modelleri (LLM) ve VLM tabanlı yaklaşımların temel mimari bileşenlerini detaylı bir şekilde ele alırken, uygulama boyutunda ağırlıklı olarak görsel dil modellerinin sektörel entegrasyonuna odaklanmayı amaçlamaktadır. Çalışmada öncelikle, her iki model ailesinin temelini oluşturan Transformer mimarisi; öz-dikkat (self-attention) mekanizmaları, tokenizasyon ve konumsal kodlama gibi süreçler bağlamında teknik olarak incelenerek LLM'lerin çalışma prensipleri açıklanmıştır. Devamında ise odak noktası tamamen çok modlu yapılara kaydırılarak, VLM ve eylem boyutuyla genişletilmiş Vision-Language-Action (VLA) mimarilerinin spesifik kullanım alanları detaylandırılmıştır. Bu kapsamda; otonom sürüş sistemlerinde uçtan uca planlama, robotik sistemlerde mekânsal temellendirme, sağlık alanında yapılandırılmış bilgi çıkarımı, tarımda uzaktan algılama ve alan-öзgü tanı sistemleri ile insan odaklı görsel analiz görevlerindeki güncel yaklaşımlar incelenmiştir. Yapılan incelemeler, VLM tabanlı sistemlerin pasif bilgi işleyiciler olmaktan çıkarak, yüksek seviyeli semantik rehberlik sağlayan karar destek bileşenlerine evrildiğini göstermektedir. Bununla birlikte, modellerin gerçek dünya sistemlerine aktarımında karşılaşılan hesaplama maliyetleri, gerçek zamanlılık kısıtları ve halüsinasyon riskleri gibi mühendislik darboğazları tartışılmış; gelecekteki eğilimin parametre-verimli ve alan-öзgü hibrit mimarilere yöneleceği sonucuna varılmıştır.

- 1 Yüksek Lisans Öğrencisi, Isparta Uygulamalı Bilimler Üniversitesi Lisansüstü Eğitim Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı , iremcakcak1012@gmail.com, 0009-0004-8929-3950
- 2 Dr. Öğr. Üyesi, Isparta Uygulamalı Bilimler Üniversitesi Teknoloji Fakültesi Bilgisayar Mühendisliği, burhanduman@isparta.edu.tr, 0000-0001-5614-1556

## 1. Giriş

Yapay zekâ sistemleri uzun yıllar boyunca belirli görevler için tasarlanan, çoğunlukla tek bir veri türüyle çalışan ve dar kapsamlı çıktılar üreten yapılar olarak geliştirilmiştir. Metin işleyen doğal dil sistemleri ile görüntüye odaklanan bilgisayarlı görü modelleri büyük ölçüde birbirinden bağımsız ilerlemiştir. Ancak son dönemde büyük ölçekli öğrenme modellerinin yaygınlaşması, hesaplama altyapısındaki ilerlemeler ve özellikle Transformer tabanlı mimarilerin olgunlaşması, bu ayrılmış yapıyı önemli ölçüde dönüştürmüştür. Bilgi işleme ve karar verme süreçleri artık daha bütüncül, daha bağlama duyarlı ve daha esnek sistemler üzerinden ele alınmaktadır.

Bu dönüşümün merkezinde yer alan Büyük Dil Modelleri (Large Language Models – LLM), yalnızca metni analiz eden sistemler olmaktan çıkarak; doğal dilin istatistiksel örüntülerini ve anlamsal yapısını öğrenebilen, çok adımlı muhakeme kurabilen, yönerge izleyebilen ve farklı görev türlerine uyarlanabilen genel amaçlı bileşenlere dönüşmüştür. LLM'ler, geniş ölçekli ön-egitim süreçleri sayesinde yalnızca cevap üretmekle kalmamakta; bağlamı takip edebilmekte, tutarlı metinler oluşturabilmekte ve karmaşık problem yapılarını dil üzerinden temsil edebilmektedir.

Bununla birlikte, gerçek dünyadaki bilginin yalnızca metinden ibaret olmadığı açıktır. İnsan karar verme süreçleri çoğu zaman görsel, mekânsal ve dilsel bilgilerin birlikte değerlendirilmesine dayanır. Bu gereksinim, metin, görüntü ve video gibi farklı veri kaynaklarının ortak bir temsil uzayında birleştirilmesini zorunlu hâle getirmiştir. Bu noktada öne çıkan Görsel–Dil Modelleri (Vision–Language Models – VLM), görsel içerik ile dilsel ifadeler arasında semantik bir bağ kurarak sahne anlama, nesne–ilişki çıkarımı ve görsel bağlama dayalı akıl yürütme gibi görevlerde geleneksel tek-modlu yaklaşımların sınırlarını genişletmektedir.

Geleneksel bilgisayarlı görü sistemleri çoğunlukla “görselde ne var?” sorusuna sınıflandırma ya da nesne tespiti gibi çıktılar üretirken, VLM tabanlı yaklaşımlar bu içeriği dilsel olarak gerekçelendirebilmekte, görsel öğeler arasındaki ilişkileri açıklayabilmekte ve kullanıcı niyetine göre görev odaklı yanıtlar oluşturabilmektedir. Böylece kullanıcı etkileşimi de değişmektedir. Önceden tanımlı komut setleri ve etiketleme temelli arayüzler yerine; doğal dilde soru sorma, açıklama talep etme, görev tanımlama ve adım adım yönlendirme gibi daha esnek iletişim biçimleri ön plana çıkmaktadır. Bu yönüyle görsel–dil modelleri, yalnızca algısal doğruluğu artıran sistemler değil; insan–makine etkileşimini yeniden yapılandıran çok modlu karar destek bileşenleri olarak değerlendirilebilir.

Ancak çok modlu sistemlerin pratikte güvenilir biçimde kullanılabilmesi, yalnızca akıcı metin üretimi ya da yüksek doğruluklu görsel temsil öğrenimi ile sınırlı değildir. Gerçek dünya uygulamalarında model çıktısının görsel kanıtı dayanması, bağlama uygun terminoloji kullanması ve tutarlı bir gerekçelendirme sunması kritik önem taşır. Özellikle otonom sürüş, sağlık, endüstriyel denetim ve tarım gibi karar maliyetinin yüksek olduğu alanlarda; halüsinasyon (görüntüde bulunmayan içerik üretimi), yanlış temellendirme, alan dışı genelleme zayıflığı, yüksek hesaplama maliyeti ve gerçek zamanlılık kısıtları temel mühendislik zorlukları olarak ortaya çıkmaktadır. Bu nedenle değerlendirme yalnızca model doğruluğu üzerinden değil; gecikme bütçesi, donanım sınırlamaları, veri dağılımı kayması ve sistem entegrasyon gereksinimleri üzerinden yapılmalıdır.

Son yıllarda literatürde gözlenen eğilim, görsel–dil modellerinin eylem boyutuyla bütünleşmesine yöneliktir. Görsel algı ve dilsel muhakemenin robotik ya da otonom sistemler gibi fiziksel dünyayla etkileşen yapılara aktarılması; planlama, görev ayrıştırma ve çok adımlı karar süreçlerinin yönetilmesini gerektirmektedir. Vision-Language-Action (VLA) mimarileri bu doğrultuda geliştirilmiş olup, dil tabanlı komutları görsel bağlamla ilişkilendirerek doğrudan eylem temsillerine dönüştürebilen bütüncül bir sistem yaklaşımı sunmaktadır. Bu evrim, sistemlerin denetimlenebilirlik, güvenlik ve açıklanabilirlik gereksinimlerini daha kritik bir noktaya taşımaktadır; çünkü üretilen çıktılar artık yalnızca metinsel öneriler değil, fiziksel sonuçlar doğurabilecek kararlar hâline gelmektedir.

Bu çalışma, söz konusu teknolojik ve mimari dönüşümü kavramsal temelleri ve uygulama alanları çerçevesinde sistematik biçimde ele almayı amaçlamaktadır. İlk eksende, LLM'lerin dayandığı Transformer mimarisi; öz-dikkat mekanizmaları, tokenizasyon ve konumsal kodlama süreçleri üzerinden incelenerek dil modellerinin teknik altyapısı açıklanmaktadır. İkinci eksende ise odak çok modlu yapılara kaydırılarak; VLM'lerin mimari türleri, eylem boyutuyla genişletilmiş VLA yaklaşımları, temel görev sınıfları ve otonom sürüş, sağlık, tarım ve robotik gibi alanlardaki uygulama örnekleri detaylandırılmaktadır. Bölümün temel katkısı, LLM ve VLM ekosistemini bütüncül bir çerçevede değerlendirmek; gerçek dünya entegrasyonunda karşılaşılan sınırlılıkları ortaya koymak ve gelecekteki eğilimin parametre-verimli, alan-özü hibrit mimarilere doğru evrildiğini tartışmaktır.

## 2. Büyük Dil Modelleri (LLM)

Büyük dil modelleri geniş ölçekli metin veri kümeleri üzerinde eğitilen ve doğal dil işleme görevlerini yerine getirebilen derin öğrenme tabanlı modellerdir. Bu modeller; metin anlama, özetleme, çeviri, dil modelleme ve içerik üretimi

gibi görevlerde kullanılmaktadır. Büyük dil modellerin büyük çoğunluğu, dilsel bağlamı ve uzun menzilli bağımlılıkları etkili biçimde modelleyebilen Transformer mimarisine dayanmaktadır.

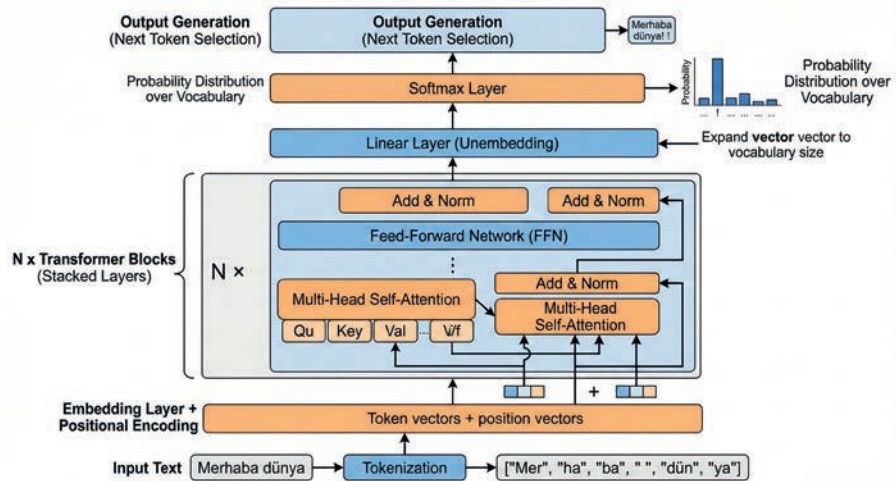
“Büyük” ifadesi yalnızca kavramsal bir nitelendirme değildir; modelin eğitildiği veri miktarının hacmine ve içerdiği parametre sayısının büyüklüğüne doğrudan işaret eder. Literatürde tanımlanan ölçeklenme yasaları (scaling laws), veri miktarı ve parametre sayısı arttıkça model performansının belirli bir doğrultuda ve öngörülebilir biçimde iyileştiğini göstermektedir.

## 2.1. Büyük Dil Modellerin Temel Mimarisini

Büyük dil modellerin temel mimarisini, Transformer tabanlı derin öğrenme yaklaşımlarına dayanmaktadır. Transformer mimarisini, önceki yinelemeli ağ yapılarının (RNN, LSTM) aksine, sıralı hesaplama zorunluluğunu ortadan kaldırarak dikkat (self-attention) mekanizması üzerinden paralel işlemeye olanak tanımaktadır. Bu özellik, modelin daha büyük veri kümeleri ve parametre sayılarıyla verimli biçimde eğitilmesini mümkün kılmıştır.

Günümüzde yaygın olarak kullanılan büyük dil modelleri, Transformer mimarisinin encoder, decoder veya encoder–decoder temelli farklı yapılandırmaları üzerine inşa edilmektedir.

### 2.1.1. Büyük Dil Modellerin Temel Yapısal Bileşenleri



Şekil 1: Transformer Tabanlı LLM'in Genel Çalışma Akışı

Yüksek soyutlama düzeyinde incelendiğinde, Transformer tabanlı bir büyük dil modeli (LLM) üç ana bileşenden oluşmaktadır (Şekil 1). İlk aşamada yer alan sözcükleyici (tokenizer), girdi metnini modelin işleyebileceği daha küçük alt birimlere (tokenlara) ayırır ve her birini sözlükte tanımlı benzersiz bir sayısal kimlikle temsil eder. Bu dijitalleştirme işlemi, metinsel verinin sayısal formata dönüştürülerek modelin dil üzerinde matematiksel işlemler yapabilmesine olanak tanır. Sayısallaştırılan bu girdiler, daha sonra modelin temel hesaplama merkezini oluşturan Transformer blokları yığına aktarılır. Ardışık olarak düzenlenmiş onlarca katmandan oluşabilen bu bloklar, girdiyi derin bağlamsal temsillere dönüştürür. Modelin “katman derinliği” olarak da bilinen bu mimari büyüklüğü, karmaşık dilsel bağımlılıkları öğrenme kapasitesini doğrudan etkileyen en kritik tasarım parametrelerinden biridir. Sürecin son aşamasında ise dil modelleme kafası (language modeling head) devreye girer. Bu bileşen, Transformer bloklarından elde edilen bağlamsal temsilleri kullanarak modelin sözlüğündeki her bir token için bir olasılık dağılımı üretir ve metin üretim sürecinin bir sonraki adımını belirler.

### 2.1.2. Transformer Bloğunun İç Yapısı

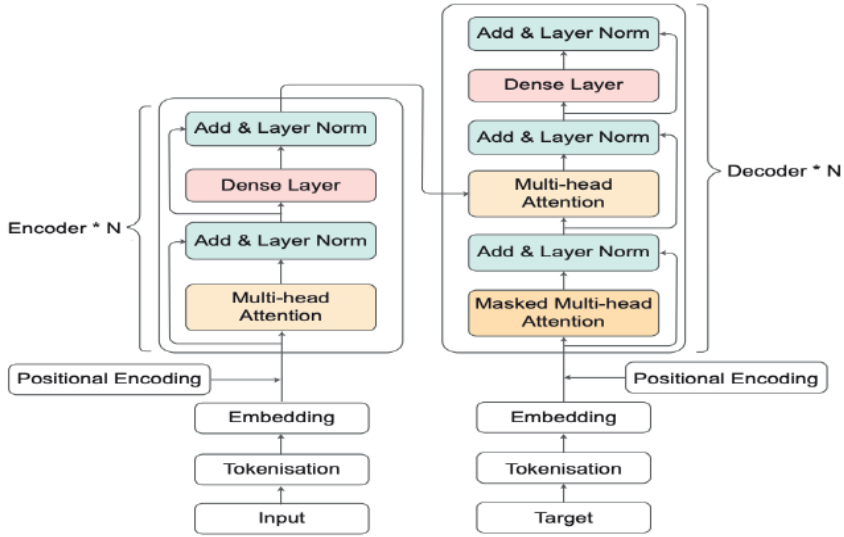
Transformer tabanlı bir LLM’de, her bir transformer bloğunun iç yapısı ve temel bileşenleri Şekil 2’de gösterilmektedir. Bir transformer bloğu, self-attention (öz-dikkat) mekanizması ve Feed-Forward Network (ileri beslemeli sinir ağı) bileşenlerinden oluşur.

#### 2.1.2.1. Öz-Dikkat Mekanizması (Self-Attention) Katmanı

Self-attention mekanizması, bir tokenın temsili oluşturulurken dizideki diğer tokenlarla olan ilişkilerinin dikkate alınmasını sağlar. Bu sayede model, yerel bağlamın ötesinde uzun menzilli bağımlılıkları modelleyebilir ve bağlama duyarlı temsiller üretebilir. Öz-dikkat yapısı, özellikle sıralı verilerde bağlamsal ilişkilerin öğrenilmesinde kritik bir rol oynamaktadır.

#### 2.1.2.2. İleri Beslemeli Sinir Ağı (Feed-Forward Network)

Self-attention katmanından elde edilen çıktılar, her token için bağımsız olarak uygulanan iki katmanlı bir Feed-Forward Network’e aktarılır. Bu bileşen, doğrusal olmayan dönüşümler aracılığıyla temsillerin zenginleştirilmesini sağlar. Transformer bloğu ayrıca artık (residual) bağlantılar ve katman normlama (layer normalization) bileşenlerini içerir. Artık bağlantılar derin ağlarda gradyan akışını desteklerken, katman normlama eğitim sürecinin kararlılığını artırmaya yardımcı olur.



Şekil 2: Transformer kodlayıcı-çözücü (encoder-decoder) mimarisi.

Kaynak: Vaswani vd. (2017).

### 2.1.3. Transformer Mimarisi Türleri

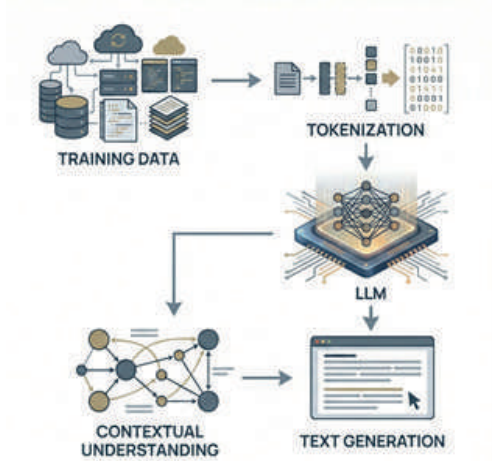
Transformer mimarileri, kullanım amaçlarına ve üstlendikleri doğal dil işleme görevlerinin niteliğine bağlı olarak üç temel varyasyonda tasarlanmaktadır. Kodlayıcı odaklı (encoder-only) modeller, metin üretiminden ziyade girdiyi bağlamsal olarak anlamaya odaklanır. BERT gibi mimariler bu gruba dahildir ve özellikle metin sınıflandırma, duygu analizi, öznelik çıkarımı ve cümle temsili üretimi gibi analitik görevlerde yaygın biçimde kullanılmaktadır.

Buna karşılık, GPT serisi ve Phi gibi modellerin temsil ettiği kod çözücü odaklı (decoder-only) yapılar üretim merkezli bir yaklaşım benimser. Bu modeller oto-regresif (autoregressive) bir prensiple çalışarak her adımda bir sonraki tokeni tahmin eder ve metni ardışık biçimde üretir. Bu özellikleri sayesinde metin üretimi, soru yanıtlama, metin tamamlama ve diyalog sistemleri gibi görevlerde yüksek performans göstermektedir.

Kodlayıcı-kod çözücü (encoder-decoder) mimarileri ise hem girdiyi temsil eden bir kodlayıcı hem de bu temsile dayalı çıktı üreten bir kod çözücü bileşenini bir arada barındırır. T5 ve BART gibi modeller bu yapıyı kullanmaktadır. Bu mimari, özellikle bir girdi dizisinin farklı bir çıktı dizisine dönüştürülmesini gerektiren makine çevirisi, metin özetleme ve çeşitli sequence-to-sequence görevlerinde tercih edilmektedir.

## 2.2. Büyük Dil Modellerinin İşleyişi

Şekil 3'te gösterildiği üzere büyük dil modelleri metinsel dokümanları işlemek için çok aşamalı bir hesaplama süreci kullanır. Bu süreç; metnin model tarafından işlenebilir bir forma dönüştürülmesini, bağlamsal temsillerin oluşturulmasını ve kullanıcı girdisine uygun bir çıktının üretilmesini kapsar.



Şekil 3 : LLM çalışma prensibi

### 2.2.1. Tokenizasyon (Sözcükleme)

Tokenizasyon, büyük dil modelleri için temel bir ön işleme adıdır ve metinsel girdilerin sayısal temsillere dönüştürülmesini sağlar. Bu süreçte, girdi metni model tarafından işlenebilecek tokenlara ayrıştırılır. Tokenlar; kelime, sub-word (alt-kelime) veya karakter düzeyinde tanımlanabilir ve bu temsil biçimi, modelin dil verisi üzerinde işlem yapabilmesi için zorunlu bir ara katman oluşturur.

Tokenizasyon sürecinde alt-kelime tabanlı yaklaşımlar, dildeki çeşitliliği daha dengeli biçimde ele alarak nadir kelimelerin ve biçimbilimsel varyasyonların modele dahil edilmesine olanak tanır. Bu nedenle, büyük ve heterojen metin kümeleriyle eğitilen modern büyük dil modelleride alt-kelime temsilleri yaygın olarak tercih edilmektedir.

Modern tokenizasyon yöntemleri, bilinmeyen kelime (Out-of-Vocabulary, OOV) problemini önemli ölçüde azaltarak modelin farklı dil yapıları ve yazım biçimleriyle karşılaşması durumunda daha kararlı sonuçlar üretmesine katkı sağlar. Özellikle boşluklardan bağımsız çalışan yöntemler, sosyal medya verileri

veya gürültülü metinler gibi düzensiz veri kaynaklarında daha esnek bir yapı sunmaktadır.

Yaygın olarak kullanılan tokenizasyon yöntemleri ve temel özellikleri Tablo 1'de özetlenmektedir.

*Tablo 1:Tokenizasyon Yöntemleri*

Yöntem	Temel Mekanizma	Seçim Kriteri	Kullanım Alanı / Özellikler
<b>Byte-Pair Encoding (BPE)</b>	Metindeki en sık görülen ardışık sembol çiftlerini birleştirerek sözlüğü genişletir.	Frekansı en yüksek çiftlerin seçimi.	Dengeli sözlük boyutu sağlar; alt-kelime temsilleriyle bilinmeyen kelime sorununu azaltır.
<b>WordPiece</b>	Alt birimleri, eğitim verisinin olasılığını en çok artıracak biçimde birleştirir.	Maksimum likelihood artışı sağlayan çiftlerin seçimi.	Daha kontrollü sözlük büyütme sunar; bağlam duyarlılığı daha yüksektir.
<b>Unigram / SentencePiece</b>	Geniş bir başlangıç token kümesi oluşturur; düşük olasılıklı tokenları iteratif şekilde eler.	Minimum kayıp veren token kümesinin belirlenmesi.	Boşluk bağımsız çalışır; gürültülü veya çok dilli veri üzerinde daha stabil performans gösterir.

### 2.2.2. Gömme (Embedding)

Şekil 4'te gösterildiği üzere, tokenizasyonun ardından gelen gömme (embedding) aşaması, ayrı token kimliklerini model tarafından işlenebilecek sürekli ve düşük boyutlu vektör temsillerine dönüştürür. Tokenlar yalnızca ayrı kimlikler (ID) olarak temsil edildiklerinde semantik bilgi taşımaz; gömme katmanı bu ayrı temsilleri ortak bir anlam uzayında konumlandırarak modele uygun bir giriş temsili oluşturur. Böylece kelimeler arasındaki anlamsal benzerlikler ve bağlamsal ilişkiler vektör uzayında ifade edilebilir hâle gelir ve Transformer tabanlı modellerin öz-dikkat mekanizması aracılığıyla bağlama duyarlı çıktılar üretmesinin temeli atılmış olur.

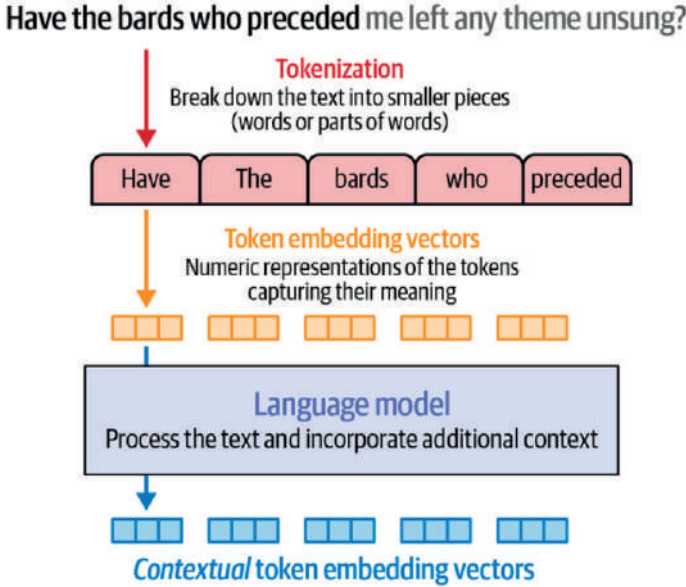
Matematiksel olarak bir tokenın gömme vektörü

$$e_t = W_e [t]$$

şeklinde tanımlanır. Burada  $W_e \in \mathbb{R}^{V \times d}$  öğrenilen embedding matrisini,  $d$  modelin gizli boyutunu (model dimension),  $e_t \in \mathbb{R}^d$  ise ilgili tokenın sürekli vektör temsili ifade eder.

Embedding yaklaşımları tarihsel gelişim sürecinde statik ve dinamik olmak üzere ikiye ayrılmaktadır. Word2Vec ve GloVe gibi statik yöntemlerde her kelime bağlamdan bağımsız tek bir sabit vektörle temsil edilir; bu durum özellikle çok anlamlı kelimelerde sınırlılıklar doğurur. Transformer tabanlı modern modellerde ise token temsilleri bağlama duyarlı biçimde üretilir. BERT ve GPT gibi yapılarda bir tokenın temsili, bulunduğu bağlam ve çevresindeki diğer tokenlar dikkate alınarak dinamik biçimde güncellenir. Bu bağlamsal temsiller büyük ölçekli ön-eğitim (pre-training) süreçlerinde öğrenilmekte ve görev odaklı ince ayar (fine-tuning) aşamalarında optimize edilmektedir.

Ancak embedding vektörleri tek başına dizisel sıra bilgisini içermez. Bu nedenle Transformer mimarisinde gömme temsilleri konumsal kodlama (positional encoding) ile birlikte kullanılarak nihai giriş vektörü oluşturulur.



Şekil 4. Jetonların sözcükleme (tokenizasyon) sürecinden geçirilerek gömme (embedding) vektörlerine dönüştürülmesi ve bağlama duyarlı temsillerin elde edilmesi. Kaynak: Jay Alammar (2024)

### 2.2.3. Konumsal Kodlama (Positional Encoding)

Transformer tabanlı büyük dil modelleri metin dizilerinin sıralı yapısının doğru biçimde modellenmesi, dilin sözdizimsel ve anlamsal özelliklerinin yakalanabilmesi açısından önemli bir gerekliliktir. Ancak Transformer mimarisi tokenları paralel olarak işlediğinden, bu sıralı yapı bilgisini doğrudan temsil

etmez. Yalnızca gömme (embedding) temsillerine dayalı bir modelleme yaklaşımı, kelimelerin diziliminden kaynaklanan anlam farklılıklarını ayırt etmekte yetersiz kalabilir. Bu sınırlılığı gidermek amacıyla geliştirilen konumsal kodlama (positional encoding) yöntemleri, her bir tokenin dizideki konum bilgisini matematiksel olarak modele entegre etmektedir.

Konumsal bilgi genellikle başlangıçtaki gömme vektörüne eklenen bir konum vektörü aracılığıyla modele dahil edilir:

$$z_i = e_i + p_i$$

Burada  $e_i \in \mathbb{R}^d$ , dizinin  $i$ . pozisyonundaki jetona ait embedding vektörünü;  $p_i \in \mathbb{R}^d$ , aynı pozisyon için tanımlanan mutlak konumsal gömme vektörünü;  $z_i \in \mathbb{R}^d$  ise konumsal bilgi ile zenginleştirilmiş nihai giriş temsiliyi ifade etmektedir.

Literatürde konumsal bilginin modele entegrasyonu iki temel yaklaşım altında incelenmektedir:

- *Mutlak Konumsal Gömme (Absolute Positional Embeddings – APE)*: Her pozisyon için sabit (sinüzoidal) veya veriden öğrenilebilir özgün bir vektör tanımlar. Ancak bu yöntemler genellikle önceden belirlenmiş bir maksimum dizi uzunluğuyla sınırlı kalır ve tokenlar arasındaki göreceli mesafeyi doğrudan modelleyemez.
- *Göreceli Konumsal Gömme (Relative Positional Embeddings – RPE)*: Konumsal bilgiyi doğrudan dikkat (attention) mekanizmasına entegre ederek token çiftleri arasındaki mesafeyi baz alır. Uzun bağlamlı (long-context) modern dil modellerinde esneklik ve ölçeklenebilirlik sağlayan bu yaklaşımın öne çıkan iki türü bulunmaktadır:
  - o *RoPE (Rotary Position Embeddings)*: Sorgu (query) ve anahtar (key) vektörlerine sinüzoidal rotasyon uygulayarak mesafe bilgisini hesaplamaya dâhil eder. Bu yapı, eğitimde görülmeyen uzun dizilerde bile modele güçlü bir genelleme yeteneği kazandırır.
  - o *ALiBi (Attention with Linear Biases)*: Konum bilgisini vektörlere eklemek yerine, dikkat skorlarına doğrusal bir sapma (ceza) olarak uygular. Tokenlar arası mesafe arttıkça dikkat ağırlığı sistematik olarak düşürülür, bu da özellikle uzun metinlerde oldukça kararlı bir çıkarım (ekstrapolasyon) davranışı sunar.

#### 2.2.4. Dönüştürücü (Transformer) Blokları ve Öz-Dikkat (Self-Attention) Mekanizması

Transformer mimarisinin temel yapı taşı, birden fazla kez yinelenen dönüştürücü (transformer) bloklardır. Şekil 5'te gösterildiği üzere her bir blok; öz-dikkat (self-attention) mekanizması, iki katmanlı ileri beslemeli sinir ağı (feed-forward network), artık bağlantılar (residual connections) ve katman normlama (layer normalization) bileşenlerinden oluşur. Giriş temsilleri, embedding ve konumsal kodlama bileşenlerinin birleştirilmesiyle elde edilir ve blok içerisinde sırasıyla işlenir.

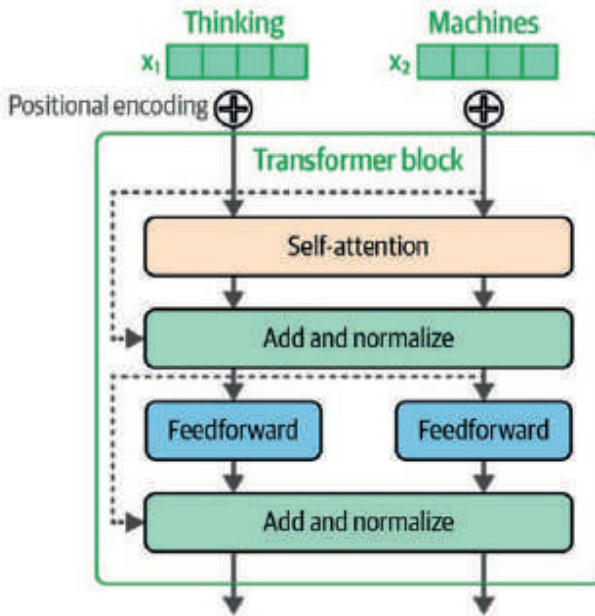


Figure 3-29. A Transformer block from the original Transformer paper.

Şekil 5: Transformer bloğunun genel yapısı

Blok yapısının ilk aşaması öz-dikkat mekanizmasıdır. Self-attention, her bir tokenın temsilini oluştururken dizideki diğer tüm tokenlarla olan ilişkisini dikkate alır. Bu sayede model yalnızca yerel komşuluk bilgisine değil, tüm dizi boyunca oluşan uzun menzilli bağımlılıklara erişebilir. Şekil 6'da şematik olarak gösterildiği üzere, bu mekanizma sorgu (Query, Q), anahtar (Key, K) ve değer (Value, V) temsilleri üzerinden çalışır.

Matematiksel olarak, giriş matrisi  $X \in \mathbb{R}^{n \times d}$  olmak üzere, öz-dikkat için gerekli doğrusal projeksiyonlar şu şekilde tanımlanır:

$$Q = XW_Q, K = XW_K, V = XW_V$$

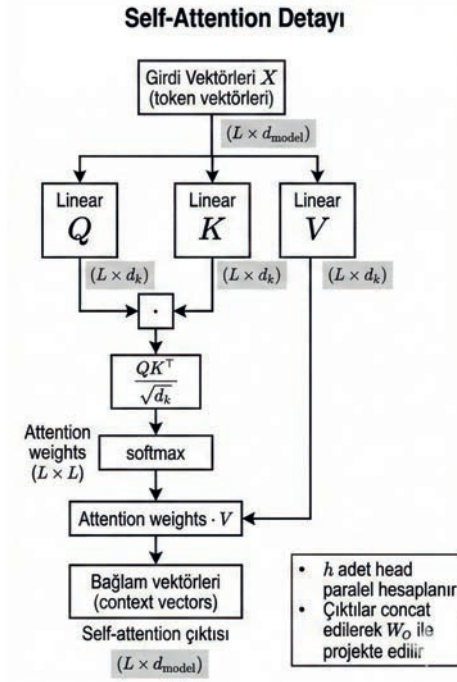
Burada  $W_Q, W_K, W_V$  öğrenilebilir ağırlık matrisleridir. Dikkat skorları, sorgu ve anahtar matrisleri arasındaki benzerlik üzerinden hesaplanır ve ölçeklenmiş noktasal çarpım (scaled dot-product) formunda ifade edilir:

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

Bu ifadeye  $QK^T \in \mathbb{R}^{n \times n}$  matrisi, tüm token çiftleri arasındaki benzerlik skorlarını içerir.  $\sqrt{d_k}$  ile ölçekleme işlemi, yüksek boyutlu temsillerde sayısal kararlılığı artırmak amacıyla uygulanır. Softmax işlemi her token için dikkat ağırlıklarını olasılık dağılımına dönüştürür ve önemli görülen tokenlara daha yüksek ağırlık verilmesini sağlar. Elde edilen dikkat ağırlıkları değer (V) matrisi ile çarpılarak her token için bağlama duyarlı, zenginleştirilmiş yeni bir temsil üretilir.

Öz-dikkat katmanının çıktısı doğrudan iletilmez; giriş temsili ile toplanarak artık bağlantı oluşturulur ve ardından katman normlama uygulanır. Bu “add & normalize” adımı, derin mimarilerde gradyan akışını stabilize ederek eğitimin daha kararlı ilerlemesine katkıda bulunur. Sonraki aşamada her tokena bağımsız olarak uygulanan iki katmanlı ileri beslemeli sinir ağı, doğrusal olmayan dönüşümler aracılığıyla temsilleri daha soyut özelliklere dönüştürür. Feed-forward çıktısına da tekrar artık bağlantı ve normlama uygulanarak blok tamamlanır.

Bu iki aşamalı yapı (self-attention + feed-forward), model derinliği boyunca yinelenerek daha üst düzey bağlamsal temsillerin öğrenilmesini sağlar. Şekil 5'te görüldüğü gibi, blok içerisindeki bu düzenli yapı Transformer mimarisinin hem paralel hesaplama verimliliğini hem de yüksek temsil kapasitesini mümkün kılan temel tasarım unsurudur.



Şekil 6: Self-attention mekanizması şematik gösterimi

### 2.2.5. İleri Beslemeli Ağ (Feed-Forward Network, FFN)

İleri beslemeli ağ ya da MLP (Multi-Layer Perceptron) alt bloğu, her transformer bloğunda çok başlı öz-dikkat katmanını izleyen temel bileşenlerden biridir ve modelin hesaplama kapasitesinin büyük kısmını üzerinde toplar. Öz-dikkat katmanı tokenlar arasındaki bağlamsal ilişkileri çözümlerken, FFN bu zenginleştirilmiş temsilleri her bir token için bağımsız olarak (position-wise) dönüştürür. Modelin ezberleme (memorization) ve genelleme (interpolasyon) yetenekleri ağırlıklı olarak bu katmanda şekillenir ve öğrenilen dilsel bilgilerin önemli bir bölümü burada depolanır.

Yapısal olarak FFN, iki doğrusal dönüşüm ve aralarındaki doğrusal olmayan bir aktivasyon fonksiyonundan oluşur. Öz-dikkat çıktısı olan  $\in \mathbb{R}^{d_{model}}$  vektörü, ilk doğrusal katman ( $W_1$ ) ile boyutu genellikle  $4 \times d_{model}$  olacak şekilde genişletilir. Ardından (klasik yapıda ReLU, modern dil modellerinde ise çoğunlukla GELU veya SwiGLU gibi) bir aktivasyon fonksiyonundan geçirilen veri, ikinci doğrusal katman ( $W_2$ ) ile tekrar  $d_{model}$  boyutuna sıkıştırılarak projeksiyon tamamlanır. Bu genişletme-sıkıştırma yapısı, modelin daha yüksek boyutlu bir temsil uzayında karmaşık, doğrusal olmayan örüntüler öğrenmesine

ve sonrasında bunları tekrar self-attention ile uyumlu boyuta projekte etmesine imkân tanır.

Eğitim sürecinde gradyan akışını ve yakınsamayı iyileştirmek için FFN bloğu; artık bağlantılar (residual connections) ve normalizasyon katmanlarıyla birlikte kullanılır. Modern tasarımlarda genellikle blok girişinde uygulanan ön-normalizasyon (LayerNorm veya RMSNorm) tercih edilmektedir. Bazı gelişmiş ve devasa ölçekli mimarilerde ise FFN, her token için yalnızca ilgili “uzman” ağların aktifleştiği Mixture-of-Experts (MoE) yapılarıyla genişletilmektedir. Bu bütünleşik yapı sayesinde modelin ifade gücü, yüksek kapasiteli ve seyrek (sparse) bir hesaplama modülüyle maksimize edilir.

### 2.2.6. Çıktı Katmanı ve Token Üretimi (Logits, Softmax, Sampling)

Büyük dil modellerinde metin üretimi, son Transformer bloğundan elde edilen gizli temsillerin sözlük uzayına projeksiyonu ile gerçekleştirilir. Her zaman adımında elde edilen  $h_t \in \mathbb{R}^{d_{\text{model}}}$  boyutlu gizli vektör, doğrusal bir çıktı katmanı aracılığıyla model sözlüğündeki her bir token için bir uygunluk skoru (logit) üretir:

$$z_t = W_{\text{out}} h_t + b_{\text{out}}$$

Burada  $W_{\text{out}} \in \mathbb{R}^{|V| \times d_{\text{model}}}$  ağırlık matrisi,  $b_{\text{out}} \in \mathbb{R}^{|V|}$  sapma vektörü ve  $|V|$  sözlük boyutudur. Elde edilen  $z_t$  vektörü ham skorları temsil eder ve henüz bir olasılık dağılımı değildir.

Bu skorlar Softmax fonksiyonu ile normalize edilerek olasılık dağılımına dönüştürülür:

$$p_t(i) = \frac{\exp(z_t(i))}{\sum_{j=1}^{|V|} \exp(z_t(j))}$$

Bu işlem sonucunda her aday token için 0 ile 1 arasında bir olasılık değeri elde edilir ve toplamları 1 olacak şekilde normalize edilir. Eğitim aşamasında bu dağılım, gerçek hedef token ile karşılaştırılarak çapraz entropi (cross-entropy) kaybı hesaplanır ve model parametreleri geriye yayılım ile güncellenir.

Çıkarım (inference) sürecinde ise model, hesaplanan olasılık dağılımından bir sonraki tokeni seçerek metni adım adım üretir. Bu seçim süreci deterministik (greedy) veya olasılık temelli örnekleme stratejileri (top-k, top-p gibi) ile gerçekleştirilebilir. Ayrıca sıcaklık (temperature) parametresi kullanılarak dağılımın keskinliği ayarlanabilir; düşük sıcaklık değerleri daha tutarlı ve

öngörülebilir çıktılar üretirken, yüksek değerler daha çeşitli ve yaratıcı metinlere olanak tanır.

Metin üretimi oto-regresif bir mekanizma ile ilerler. Model her adımda bir sonraki tokenın koşullu olasılığını  $P(x_t | x_{<t})$  biçiminde hesaplar ve seçilen token giriş dizisine eklenir. Bu süreç, belirlenen bir sonlandırma koşuluna ulaşılan kadar tekrarlanır. Böylece büyük dil modelleri, bağlama duyarlı ve tutarlı metinleri ardışık token üretimi yoluyla oluşturur.

### 3. Görsel–Dil Modelleri (Vision–Language Models, VLM)

Görsel dil modelleri, görüntü ve metin bilgisini ortak bir temsil çerçevesinde birleştirerek görsel içerik hakkında dil tabanlı çıkarım yapabilen çok modlu (multimodal) modellerdir. Bu modeller, yalnızca “görselde ne var?” sorusunu yanıtlayan geleneksel görsel tanıma yaklaşımlarının ötesine geçerek; nesnelere ve aralarındaki ilişkileri doğal dilde tanımlama, bağlamsal yorumlama ve soruya ya da komuta göre gerekçelendirilmiş yanıt üretme gibi daha üst düzey etkileşimleri mümkün kılar. Böylece kullanıcı, görsel içerikle etiketleme veya arama odaklı bir süreç yerine, soru sorma, açıklama isteme ya da yönlendirme verme gibi daha doğal bir iletişim biçimiyle etkileşime girebilir.

Görsel dil modellerinin pratik değeri, görsel ipuçlarını metinsel bağlamla birleştirerek tek bir görüntü üzerinden farklı türde görevleri destekleyebilmesinden kaynaklanır. Görüntü açıklama (image captioning), görsel soru–cevaplama (visual question answering, VQA) ve çapraz-modalite erişim (örneğin metinden görsel bulma ya da görselden metin bulma) gibi görevler, aynı model ailesi altında ele alınabilmekte; model, kullanıcının niyetine göre görseldeki ilgili kanıtları seçerek tutarlı bir dil çıktısına dönüştürebilmektedir. Bu bağlamda VLM yetenekleri literatürde belirli görev sınıfları altında incelenmektedir.

Görsel dil modellerinin başarısı büyük ölçüde web ölçekli görüntü–metin çiftleri üzerinde gerçekleştirilen ön-eğitim (pre-training) süreci sayesinde, her iki modaliteden zengin ve genellenebilir temsiller öğrenilmesine dayanır. Bu sayede modeller, etiketli veriye sıkı biçimde bağımlı kalmadan yeni görüntüler üzerinde sıfır-atış (zero-shot) çıkarım yapabilir ve daha önce karşılaşmadıkları kavramları uygun metinsel ifadelerle ilişkilendirebilir. Bununla birlikte, görsel dil modellere yönelik ilgi hızla artmasına rağmen alan-öзgöl senaryolarda güvenilir kullanım için önemli güçlükler devam etmektedir. Üretilen çıktının görsel kanıta dayalı biçimde temellendirilmesi (grounding) her zaman yeterince açık olmayabilir; teknik veya klinik bağlamlarda gerekli olan terim doğruluğu ve ayrıntı düzeyi korunamayabilir; model eğitimde görmediği alanlara taşındığında genelleme performansı düşebilir.

Bu nedenle görsel dil modellerinin değerlendirilmesi, yalnızca genel amaçlı ölçütlerle sınırlı tutulmamalı; hedef uygulamanın gerektirdiği doğruluk, tutarlılık ve açıklanabilirlik kriterleri doğrultusunda alan verisi üzerinde test edilmesi ve gerektiğinde uyarlanması gerekmektedir. Bu çerçevede görsel dil modellerinin yetenekleri, literatürde yaygın biçimde kullanılan görev sınıfları üzerinden ele alınmaktadır. Görev temelli bu ayırım, görsel ve metinsel temsillerin etkileşim biçimini ve model çıktısının niteliğini görünür kılarak karşılaştırmalı değerlendirmeyi kolaylaştırır. Bu nedenle aşağıda, en sık kullanılan temel görev alanları kısaca özetlenmektedir.

### 3.1. Görsel Dil Modellerinin Temel Görev Alanları

Görsel dil modellerinin (VLM) görsel ve dilsel temsilleri ortak bir semantik uzayda birleştirme yeteneği, literatürde çeşitli temel görev alanları üzerinden değerlendirilmektedir. Bu görevler, modellerin çok modlu anlama, eşleştirme, muhakeme ve üretim kapasitesini ölçmeyi amaçlar.

*Görsel Soru–Cevaplama (VQA)*, bir görüntü ve bu görüntüye ilişkin doğal dildeki sorunun birlikte değerlendirilerek uygun yanıtın üretilmesini hedefler. Bu görev; nesne tanıma, sayma ve uzamsal ilişkileri çözümleme gibi alt becerileri bir araya getirerek modelin çok modlu kavrama yeteneğini ortaya koyar.

*Görsel Betimleme (Image Captioning)*, görüntüdeki sahnenin doğal dilde ifade edilmesidir. Modelin yalnızca görsel içeriği tanıması değil, aynı zamanda bunu akıcı ve bağlama uygun bir metne dönüştürmesi beklenir. Bu nedenle üretim (generation) odaklı VLM yeteneklerini yansıtır.

*Görüntü–Metin Eşleştirme ve Getirme (Image–Text Retrieval)*, metin ve görüntü çiftlerinin ortak temsil uzayında hizalanmasına dayanır. Amaç, bir metne en uygun görüntüyü ya da bir görüntüye en uygun metni benzerlik skorları üzerinden belirlemektir. Bu yaklaşım, arama ve bilgi erişim sistemlerinde yaygın olarak kullanılmaktadır.

*Görsel Muhakeme (Visual Reasoning)*, görüntüdeki öğeler arasındaki ilişkileri (neden–sonuç, karşılaştırma, olay akışı vb.) modelleyerek mantıksal çıkarım yapılmasını kapsar. Bu görev, basit tanımanın ötesine geçerek çok-adımlı akıl yürütme kapasitesini değerlendirir.

*Görsel Referanslama (Visual Grounding)*, metindeki belirli bir ifadenin görüntüdeki ilgili nesne veya bölgeyle eşleştirilmesini hedefler. Bu yetenek, model çıktılarının görsel kanıtla dayandırılmasını sağladığı için açıklanabilirlik açısından önemlidir.

*Belge Zekâsı ve Düzen Anlama (Document Intelligence & OCR)* ise yapılandırılmış dokümanlarda yalnızca metnin tanınmasını değil, aynı zamanda

sayfa düzeni ve görsel bileşenler arasındaki yapısal ilişkilerin yorumlanmasını içerir. Bu alan, kurumsal bilgi çıkarımı ve otomasyon uygulamalarında kritik bir rol oynamaktadır.

### 3.2. Görsel Dil Modellerinin Temel Yapısı

VLM'in mimari yapısı, görsel ve metinsel bilgiyi birlikte işleyebilmek amacıyla tasarlanmış üç temel bileşen etrafında şekillenir (Şekil 7). Bu bileşenler, farklı mimari düzenlemelere sahip modellerde değişen biçimlerde konumlandırılrsa da, görsel dil modellerinin çalışma prensibinin temelini oluşturmaktadır.

#### 3.2.1. Görüntü Kodlayıcı (Image Encoder)

Görüntü kodlayıcı, ham piksel değerlerinden oluşan görüntü veya video girdilerini, anlamsal açıdan zengin ve yoğun vektör temsillerine dönüştürmekle sorumludur. Bu bileşenin temel amacı, görsel içeriği dil bileşeninin işleyebileceği bir temsil uzayına taşımaktır. Bu süreçte yalnızca düşük seviyeli görsel özellikler (renk, kenar, doku) değil; aynı zamanda nesnelere, sahne yapısı ve uzamsal ilişkiler de soyutlanmış biçimde kodlanır.

Literatürde erken dönem VLM çalışmalarında sıklıkla evrişimli sinir ağları (CNN) kullanılmıştır. Ancak güncel yaklaşımlarda, küresel bağlamı daha etkin biçimde modelleyebilmesi nedeniyle Vision Transformer (ViT) tabanlı kodlayıcılar yaygınlaşmıştır. ViT mimarilerinde görüntü, yamalara ayrılarak görsel belirteçler (visual tokens) hâline getirilir ve bu belirteçler dikkat mekanizmaları aracılığıyla tüm görüntü genelinde etkileşime girer. Bu yapı, özellikle dil bileşeniyle kurulacak etkileşim açısından esnek ve ayrıntılı bir temsil sunar.

Görüntü kodlayıcının çıktısı, mimariye bağlı olarak tek bir global görüntü vektörü (global embedding) veya görüntünün yamalarına karşılık gelen bir görsel token dizisi (patch-level tokens) biçiminde elde edilebilir. Token-temelli temsil, görüntünün farklı bölgelerine ait bilgiyi ayrı ayrı taşıdığı için nesne/bölge düzeyinde daha ayrıntılı temsil sağlarken; global temsil daha kompakt olup özellikle eşleştirme ve arama (retrieval) gibi görevlerde verimlidir.

Bu bileşende omurga olarak ResNet gibi CNN tabanlı yapılar veya ViT/DeiT/Swin Transformer gibi Transformer tabanlı kodlayıcılar kullanılmakta; ayrıca pek çok güncel VLM'de CLIP-ViT görsel kodlayıcısı yaygın bir tercih olarak öne çıkmaktadır.

#### 3.2.2. Dil Bileşeni (Language Encoder / Decoder)

Dil bileşeni, metinsel girdilerin işlenmesinden ve gerekli durumlarda metin çıktısının üretilmesinden sorumludur. Bu bileşen, mimariye bağlı olarak yalnızca

bir dil kodlayıcıdan (encoder), yalnızca bir dil kod çözücünden (decoder) veya her ikisinin birlikte kullanıldığı bir yapıdan oluşabilir.

Dil kodlayıcılar, verilen metni anlamsal ve bağlamsal bir temsile dönüştürerek özellikle anlama ve eşleştirme temelli görevlerde kullanılır. Dil kod çözücüler ise görsel ve metinsel bağlamı birlikte değerlendirerek ardışık kelime üretimi gerçekleştirir ve görüntü açıklama ya da görsel soru-cevaplama gibi üretim odaklı görevlerin merkezinde yer alır. Güncel görsel dil modellerde bu bileşen çoğunlukla büyük dil modelleri müzerine inşa edilmekte; Transformer tabanlı mimariler, dilsel akıl yürütme ve tutarlılık açısından temel yapı taşı olarak kullanılmaktadır.

Encoder-only dil bileşenleri daha çok görüntü–metin eşleştirme/retrieval senaryolarında kullanılırken, decoder-only yapılar ve encoder–decoder modeller özellikle captioning ve VQA gibi üretim odaklı görevlerde öne çıkar. Güncel sistemlerde büyük dil modelleri çoğunlukla instruction-tuning ile çok modlu yönerge takibi (multimodal instruction following) yapacak şekilde uyarlanarak daha doğal ve görev yönelimli çıktılar üretir.

Bu kapsamda, görüntü–metin eşleştirme/retrieval odaklı yapılar CLIP, üretim ve yönerge takip eden multimodal çıktılara BLIP-2 / InstructBLIP ve LLM tabanlı görsel-diyalog yaklaşımlarına LLaVA örnek olarak verilebilir.

### 3.2.3. Modaliteler Arası Birleştirme (Cross-Modal Fusion)

Modaliteler arası birleştirme mekanizması, görüntü ve metin temsillerinin aynı bağlam içerisinde ilişkilendirilmesini sağlayarak görsel dil modellerinin ayırt edici yeteneklerini ortaya çıkaran bileşendir. Bu mekanizmanın temel işlevi, görsel ve dilsel bilgiler arasında hizalama kurmak ve iki modalite arasındaki karşılıklı bağımlılıkları öğrenmektir.

Bu süreç literatürde genellikle erken birleşim (early fusion), geç birleşim (late fusion) ve ara düzey / çapraz dikkat tabanlı birleşim (cross-attention fusion) olmak üzere üç ana yaklaşım altında incelenmektedir.

Geç birleşim (Late Fusion) yaklaşımında, görsel ve metinsel girdiler ayrı kodlayıcılarda işlenir ve yalnızca son aşamada benzerlik ölçümleri üzerinden ilişkilendirilir. İki akışlı (dual-encoder) mimariler bu yaklaşımın tipik örneklerini oluşturur. Bu yöntem hesaplama açısından verimli ve ölçeklenebilir olmakla birlikte, modaliteler arasında ayrıntılı etkileşim kurulmasına sınırlı imkân tanır.

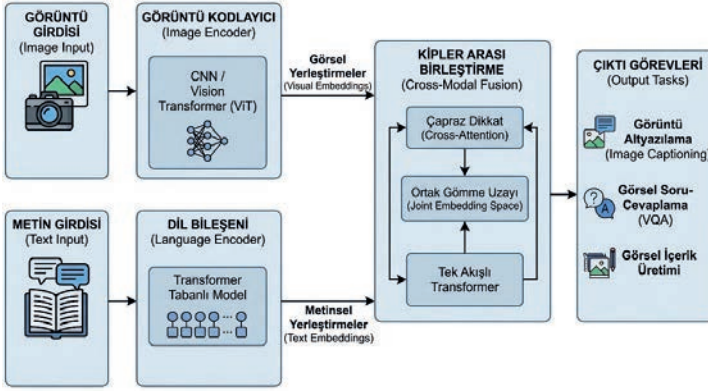
Erken birleşim (Early Fusion) yaklaşımında ise görsel ve metinsel temsiller, modelin giriş seviyesinde veya ilk katmanlarında birleştirilerek birlikte işlenir. Bu sayede her iki modalite baştan itibaren ortak bir temsil uzayında değerlendirilir.

Ancak bu yaklaşım, model karmaşıklığını ve hesaplama maliyetini artırabilmesi nedeniyle pratikte daha sınırlı kullanıma sahiptir.

Ara düzey birleşim ve çapraz dikkat (Cross-Attention Fusion), güncel VLM mimarilerinde en yaygın tercih edilen yaklaşımdır. Bu yöntemde görsel ve metinsel temsiller ayrı olarak elde edilir, ancak Transformer tabanlı çapraz dikkat katmanları aracılığıyla derinlemesine etkileşime sokulur. Tek akışlı (single-stream) ve iki akışlı (two-stream) mimariler, bu yaklaşımı farklı biçimlerde uygulayarak görsel ve dilsel bilgiler arasında çift yönlü bilgi akışı sağlar.

Çapraz dikkat mekanizmalarında, bir modalitenin temsilleri (ör. metin token'ları) diğer modaliteden (ör. görsel token'lar) bilgi “çekerek” hizalama kurar; böylece belirli kelimeler ile görüntünün ilgili bölgeleri arasında daha ince taneli ilişkiler öğrenilebilir.

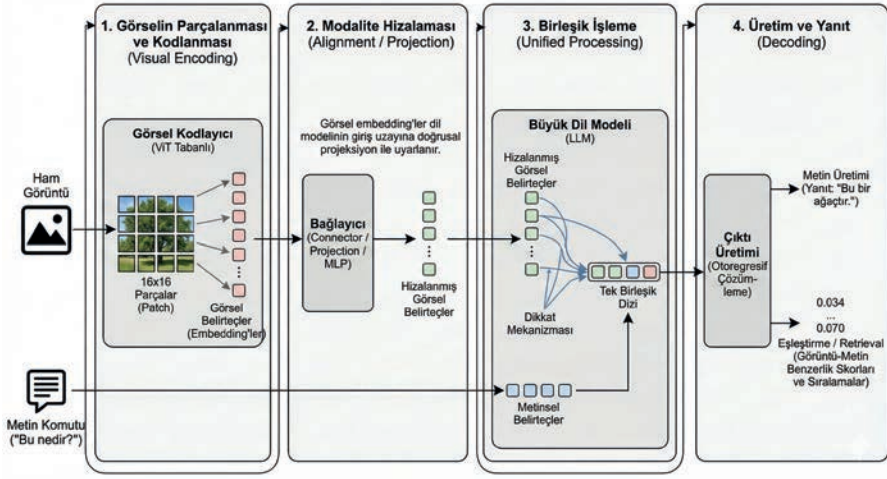
LLM tabanlı güncel görsel dil modellerde ise görsel kodlayıcı çıktısının dil modeline aktarılabilmesi için çoğunlukla projeksiyon/adapter benzeri bir köprü katmanı kullanılır; bu sayede görsel temsiller dil modelinin işleyebileceği biçime dönüştürülerek çapraz-modal etkileşim güçlendirilir.



Şekil 7: Görsel Dil Modelinin (VLM) genel mimarisi ve bileşenleri

### 3.3. Görsel–Dil Modellerinin İşleyişi

Görsel dil modelleri (VLM), görsel ve metinsel girdileri ortak bir anlamsal uzayda temsil ederek çok modlu çıkarım yapmayı hedefler. Bu modellerin çalışma mantığı, farklı modalitelerden gelen bilgilerin sayısal temsillere dönüştürülmesi, hizalanması ve birleşik bir bağlamda işlenmesi üzerine kuruludur. Genel iş akışı dört temel aşamada ele alınmaktadır (Şekil 8).



Şekil 8: Görsel dil modellerde uçtan uca iş akışı: (1) görsel kodlama (patch görsel belirteçler), (2) modalite hizalaması (bağlayıcı/projeksiyon), (3) LLM içinde birleşik işleme ve dikkat temelli çok modlu muhakeme, (4) görev türüne bağlı çıktı (otoregresif metin üretimi veya görüntü-metin eşleştirme/retrieval).

### 3.3.1. Girdi Temsili: Görüntü ve Metin

Görsel dil modeller tipik olarak bir görüntü (tek/çoklu) ve buna eşlik eden bir metin girdisi (soru, komut veya prompt) alır. Metin girdisi tokenizer aracılığıyla belirteçlere (token) ayrılarak modele beslenir. Görüntü tarafında ise modelin doğrudan piksel alanında çalışması yerine, görsel içerik önce öğrenilebilir temsillere dönüştürülür. Bu dönüşüm, özellikle Transformer tabanlı görsel kodlayıcılarda, görüntünün sabit boyutlu parçalara (patch) ayrılması ve her parçanın bir “görsel belirteç” gibi temsil edilmesiyle gerçekleştirilir.

### 3.3.2. Görsel Kodlama: Görsel Temsillerin Üretilmesi

Görsel kodlayıcı (çoğunlukla CNN veya ViT tabanlı), görüntüden ya tek bir global embedding ya da daha sık olarak görsel token dizisi üretir. Token dizisi yaklaşımı, görüntünün farklı bölgelerine ait bilgiyi ayrı ayrı taşıdığı için özellikle ayrıntı gerektiren görevlerde (ör. VQA, doküman yorumlama) daha zengin bir temsil sağlar. Görsel kodlayıcının çıktısı, modelin sonraki aşamalarda metinle ilişkilendireceği “görsel kanıt” (visual evidence) niteliğindedir (bkz. Şekil 8, 1. aşama).

### 3.3.3. Modalite Hizalaması: Bağlayıcı/Projeksiyon Katmanı

Görsel kodlayıcıdan gelen temsiller ile dil modelinin giriş uzayı çoğu zaman aynı boyutta değildir; ayrıca bu temsillerin dağılım özellikleri de farklı olabilir. Bu nedenle görsel dil modeller, görsel temsilleri dil modelinin beklediği uzaya taşıyan bir bağlayıcı (connector) / projeksiyon bileşeni kullanır. En basit formda bu bileşen doğrusal bir projeksiyon veya MLP iken, bazı tasarımlarda daha güçlü bir ara modül (ör. sorgu tabanlı dönüştürücüler) kullanılabilir. Amaç, görsel temsillerin dil modelinin bağlamına “yerleştirilebilir” hâle getirilmesi ve metinsel belirteçlerle aynı dizide birlikte işlenebilmesidir (bkz. Şekil 8, 2. aşama).

### 3.3.4. Birleşik İşleme: LLM İçinde Çok Modlu Muhakeme

Hizalanmış görsel belirteçler ve metinsel belirteçler, dil modeli gövdesine tek bir birleşik dizi olarak verilir. Bu aşamada dikkat (attention) mekanizması, metindeki sorgu ile görüntüdeki ilgili bölgeleri temsil eden görsel belirteçler arasında bağ kurar. Böylece model, metinsel bağlamı görsel kanıtla koşullandırarak çok modlu muhakeme yapabilir. Pratikte bu birleşim, modelin “hangi nesne”, “nerede”, “kaç tane”, “hangi özelliğe” gibi sorulara yanıt üretebilmesini veya görüntü hakkında tutarlı açıklamalar oluşturabilmesini sağlar (bkz. Şekil 8, 3. aşama).

### 3.3.5. Çıktı Üretimi: Retrieval ve Üretim Modları

Görsel dil modellerinin çıktısı, mimarinin hedeflediği görev türüne göre iki ana biçimde ele alınabilir:

- *Eşleştirme/Retrieval Odaklı Çıktı:* Dual-encoder benzeri tasarımlarda model, görüntü ve metin temsillerinin benzerliğini hesaplayarak bir skor üretir ve adaylar arasında sıralama yapar. Bu yaklaşım, metinden görsel bulma veya görselden metin bulma gibi erişim senaryolarında tercih edilir.
- *Üretim Odaklı Çıktı:* LLM-tabanlı görsel dil modellerde model, birleşik bağlam üzerinden otoregresif biçimde metin üretir; her adımda bir sonraki belirteci tahmin ederek yanıtı oluşturur. Bu mod, görüntü açıklama, VQA ve görsel diyalog gibi görevlerde kullanılır (bkz. Şekil 8, 4. aşama).

## 3.4. Görsel–Dil Modellerinin Mimari Türleri

VLM mimarileri genellikle görsel ve metinsel temsillerin hangi aşamada ve ne ölçüde etkileştiği, modelin benzerlik skoru mu yoksa metin mi ürettiği

ve buna karşılık gelen hesaplama maliyeti–performans dengesi açısından sınıflandırılmaktadır.

### 3.4.1. Çift Kodlayıcı (Dual-Encoder) Yaklaşımlar

Dual-encoder mimarilerde görüntü ve metin, iki ayrı kodlayıcı tarafından bağımsız olarak gömlemelere (embeddings) dönüştürülür ve ilişkilendirme genellikle son aşamada bir benzerlik ölçümü (ör. kosinüs benzerliği) üzerinden yapılır. Bu yaklaşım, özellikle büyük ölçekli veri tabanlarında görüntü–metin eşleştirme ve arama (retrieval) senaryolarında pratik avantaj sağlar; çünkü aday öğelerin temsil vektörleri önceden hesaplanabilir ve hızlı biçimde karşılaştırılabilir. Bununla birlikte, modaliteler arası etkileşim son aşamada sınırlı kaldığından, ayrıntılı hizalama (ör. kelime–bölge ilişkisi) veya çok adımlı görsel akıl yürütme gerektiren görevlerde tek başına yetersiz kalabilir. Temsilci örnekler arasında CLIP ailesi yer almaktadır.

### 3.4.2. Kodlayıcı–Çözücü (Encoder–Decoder) Yaklaşımlar

Encoder–decoder mimariler, görsel içeriği bir kodlayıcı ile temsile dönüştürdükten sonra bu temsili kullanan bir çözücü ile metin üretmeyi hedefler. Bu çerçeve, görüntü açıklama (captioning), açık uçlu görsel soru–cevaplama (VQA) ve görsel içerikten rapor/özet üretimi gibi üretim odaklı görevler için doğrudan bir çözüm sunar. Temsillerin çözücü tarafında otoresif biçimde üretilmesi, daha zengin doğal dil çıktıları sağlarken; çıkarım süresini uzatabilen bir maliyet de doğurur. Derleme literatüründe BLIP/BLIP-2 benzeri tasarımlar, hem görsel temsil kalitesi hem de üretim performansı açısından sık atıf alan örneklerdendir.

### 3.4.3. Unified / Single-Stream (Tek Akışlı Birleşik Transformer) Yaklaşımlar

Tek akışlı modellerde görsel ve metinsel belirteçler (tokens), tek bir Transformer içinde ortak bir dizi olarak işlenir. Bu tasarım, modalitelerin “başından itibaren” birbirini görmesine izin verdiği için, ince taneli çapraz-modal bağıntıların öğrenilmesini kolaylaştırır. Buna karşılık, özellikle retrieval gibi çok sayıda aday çiftin değerlendirilmesini gerektiren senaryolarda her çift için ortak işleme maliyeti ortaya çıktığından, hesaplama verimliliği dual-encoder yaklaşımlara kıyasla düşebilir. Tek akışlı modellerin güçlü olduğu alanlar çoğunlukla VQA, görsel çıkarım ve karmaşık hizalama gerektiren görevlerdir.

### 3.4.4. Donmuş LLM Tabanlı (Frozen Backbone + Adapter/Mapping) Yaklaşımlar

Son dönemde yaygınlaşan bir hat, güçlü bir dil modelini (çoğunlukla decoder-only bir LLM) büyük ölçüde donmuş halde tutup, görsel kodlayıcıdan çıkan temsilleri dil modeline taşıyan hafif köprü katmanları (projeksiyon/adapter) üzerinden çok modlu yetenek kazandırmaktır. Bu yaklaşımda amaç, LLM'in dilsel akıcılığı ve talimat takip kabiliyetini korurken, görsel bağlamı modele düşük parametre maliyetiyle entegre etmektir. Böylece çok modlu diyalog, görsel talimat izleme ve uzun biçimli açıklama üretimi gibi senaryolarda güçlü sonuçlar elde edilebilir. LLaVA sınıfı modeller bu hattın temsilci örnekleri arasında gösterilebilir. Bununla birlikte, model boyutu ve çıkarım maliyeti pratik uygulamalarda belirleyici olabilir; ayrıca görsel bilginin “ne kadar derin” işlendiği, kullanılan köprü tasarımının kapasitesine bağlıdır.

### 3.4.5. Türler Arası Karşılaştırma ve Görev Uyumları

Derleme çalışmalarında genel eğilim; retrieval/eşleştirme için dual-encoder mimarilerin ölçeklenebilirlik avantajı sunduğu, üretim ve talimat takip senaryolarında encoder-decoder veya LLM tabanlı (adapter'lı) yaklaşımların daha uygun olduğu, ayrıntılı hizalama ve çapraz-modal muhakeme gerektiren görevlerde ise single-stream tasarımların öne çıktığı yönündedir. Bu nedenle mimari seçimi; hedef görev, veri ölçeği, gecikme kısıtı ve donanım kaynaklarına göre birlikte değerlendirilmelidir (Tablo 2).

*Tablo 2. Görsel-Dil Modellerinde (VLM) Kullanılan Temel Mimari Yaklaşımların Karşılaştırılması*

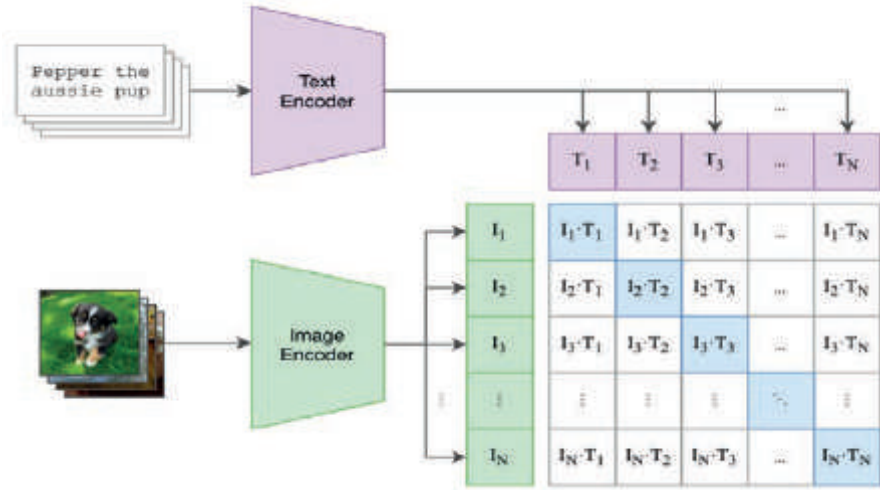
Mimari	Güçlü yön	Sınırlılık	Uygun görev
Çift Kodlayıcı	Hızlı, indekslenebilir, ölçeklenebilir	Derin görsel dil etkileşimi sınırlı	Retrieval / eşleştirme
Kodlayıcı-Çözücü	Metin üretiminde güçlü, görsele koşullu çıktı	Görev/etiketli veri bağımlılığı artabilir	Captioning, VQA
Tek Akışlı Birleşik Transformer	Erken ve yoğun etkileşim, muhakemede esneklik	Hesaplama maliyeti yüksek olabilir	Çok yönlü anlama-üretim
Donmuş LLM Tabanlı	Maliyet-etkin, hızlı uyarlama	Omurga kapasitesiyle sınırlanır	LLM'ye görsel yetenek ekleme

### 3.5. Görsel Dil Modellerinde Yaygın Olarak Kullanılan Modeller

Görsel–dil modelleri, mimari tasarımları ve hedefledikleri görevler açısından farklılaşmaktadır. Bu bölümde literatürde yaygın olarak kullanılan CLIP, BLIP, LLaVA, Kosmos ve Gemini modelleri kısaca incelenmektedir.

#### 3.5.1. CLIP

CLIP, VLM alanında dual-encoder (çift kodlayıcı) yaklaşımının en bilinen örneklerinden biridir (Şekil 9). Model, görüntü ve metni iki ayrı kodlayıcıyla bağımsız biçimde temsil eder ve bu temsilleri ortak bir gömme uzayında hizalamayı hedefler. Eğitim sürecinde kontrastif öğrenme kullanılarak doğru görüntü–metin çiftlerinin yakınlştırılması, yanlış eşleşmelerin ise uzaklaştırılması amaçlanır. Bu tasarım, özellikle görsel–metin eşleştirme ve çapraz erişim (retrieval) görevlerinde pratik avantaj sağlar; çünkü hem görseller hem de metinler ölçeklenebilir biçimde indekslenebilir. CLIP’in etkisi, sınıf etiketlerini doğal dil açıklamaları gibi ele alarak sıfır-atış (zero-shot) kullanım senaryolarında güçlü sonuçlar verebilmesiyle daha da belirginleşmiştir.



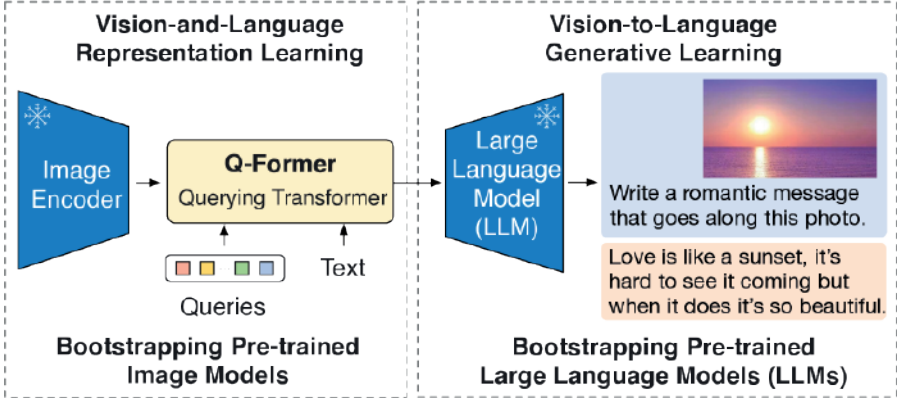
Şekil 9: CLIP modelinde metin ve görüntü girdilerinin ayrı kodlayıcılar tarafından ortak gömme uzayına projekte edilmesi ve tüm görüntü–metin çiftleri arasında benzerlik skorlarının hesaplanması.

Kaynak: OpenAI (2021), CLIP.

### 3.5.2. BLIP ve BLIP-2

BLIP ailesi, görsel dil modellerde hem anlama (ör. eşleştirme, VQA) hem de üretim (captioning) görevlerini destekleyen bir yaklaşım ortaya koyar (Şekil 10). BLIP’te görsel ve dil temsillerinin etkileşimi, görev ihtiyacına göre farklı düzenlemelerle ele alınabilir ve model, gürültülü web verisi gibi gerçek dünya kaynaklarından öğrenmeye uygun bir çerçeve sunmayı hedefler.

BLIP-2 ise daha verimli bir tasarım çizgisine yönelerek, güçlü bir görsel kodlayıcı ve güçlü bir dil modelini çoğunlukla dondurulmuş (frozen) biçimde kullanır; iki bileşeni bağlamak için hafif bir ara modül ile görsel bilginin LLM’ye aktarılmasını sağlar. Bu sayede tüm modeli baştan eğitmek yerine, sınırlı sayıda parametre ile etkili uyarlama yapılabilir. BLIP-2’nin bu yönü, “frozen backbone + bağlayıcı modül” yaklaşımının görsel dil modellerde neden popülerleştiğini gösteren iyi bir örnektir.

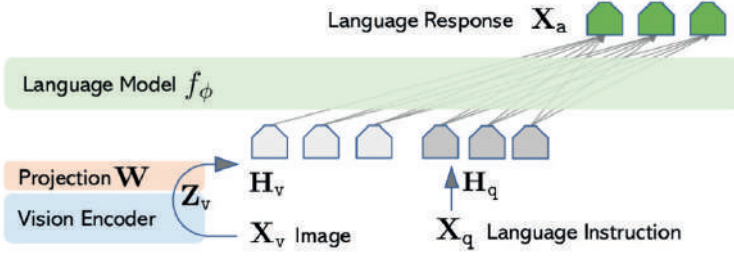


Şekil 10: BLIP-2 mimarisinde görsel kodlayıcıdan elde edilen temsillerin Q-Former (Querying Transformer) aracılığıyla büyük dil modeli (LLM) ile hizalanması ve görselden dile üretim süreci.

Kaynak: Salesforce BLIP-2, *Weights & Biases* raporu (2023).

### 3.5.3. LLAVA

LLaVA, CLIP tabanlı bir görsel kodlayıcıyı büyük bir dil modeliyle basit bir projeksiyon katmanı üzerinden bağlayan LLM-merkezli bir yaklaşımdır (Şekil 11). Görsel sohbet, talimat takibi ve açıklama üretimi gibi görevlerde başarılıdır. Bununla birlikte, küçük nesnelere ve hassas görsel ayrıntılar konusunda performansı sınırlı olabilir.



Şekil 11: LLaVA mimarisinde görsel kodlayıcıdan elde edilen görsel temsillerin bir projeksiyon katmanı aracılığıyla büyük dil modeliyle hizalanması ve dil-görsel talimatlara dayalı yanıt üretim süreci.

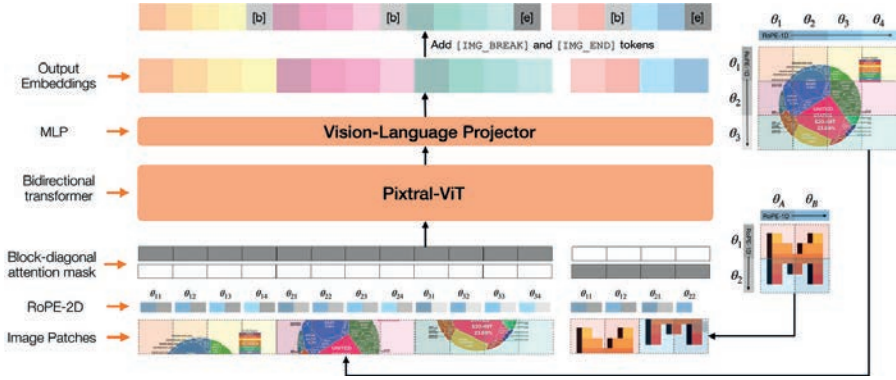
Kaynak: Encord (2023).

### 3.5.4. KOSMOS

Kosmos modelleri, görüntü ve metni tek bir Transformer mimarisi içinde işleyen birleşik (unified) bir yapıya sahiptir. Özellikle Kosmos-2 ile birlikte nesne konumlarını dilsel referanslarla ilişkilendiren grounding yeteneği öne çıkmıştır. OCR ve nesneye bağlı açıklamalarda güçlü olmakla birlikte, ölçek ve erişim açısından sınırlıdır.

### 3.5.5. PIXTRAL

Pixtral, Mistral AI tarafından geliştirilen ve görüntüyü dil modelinin bağlamına taşıyarak metin üreten (üretici) bir görsel-dil modelidir (Şekil 12). Görsel kodlayıcıdan elde edilen temsiller, büyük dil modeli (LLM) ile bütünleştirilerek görsel soru-cevaplama, görüntü açıklama ve özellikle doküman/şema yorumlama gibi görevlerde kullanılmaktadır. Uzun bağlam desteği ve çoklu görsel girdi senaryolarına uygun tasarımıyla dikkat çeken Pixtral, CLIP benzeri saf eşleştirme modellerine kıyasla daha yüksek hesaplama maliyetine sahiptir; performansı ise eğitildiği veri dağılımına (doküman ağırlıklı vs. doğal görüntüler) bağlı olarak değişkenlik gösterebilmektedir.



Şekil 12: Pixtral-Large mimarisinde görüntü yamalarının (image patches) Pixtral-ViT aracılığıyla işlenmesi, RoPE-2D ve blok-diyagonal attention maskesi ile konumsal bilginin modellenmesi ve görsel temsillerin görsel-dil projeksiyon katmanını üzerinden dil modeliyle hizalanması süreci.

Uyarlanmıştır: Encord (2024)

### 3.6. Görsel Dil Modellerinde Kullanılan Metrikler

Görsel dil modellerinin performansı tek bir metrikle temsil edilemez; çünkü modeller farklı görev sınıflarında (ör. eşleştirme/retrieval, açıklama üretimi, soru-cevaplama, grounding) farklı çıktı biçimleri üretir. Bu nedenle değerlendirme, görevin doğasına uygun ölçütlerle çok boyutlu biçimde ele alınmalıdır. Bu bölümde görsel dil modeller için literatürde yaygın kullanılan değerlendirme ölçütleri, görev ailelerine göre sınıflandırılarak özetlenmekte ve her bir metrik grubunun hangi tür çıktıları ölçmekte daha işlevsel olduğu kısaca açıklanmaktadır.

#### 3.6.1. Görüntü Açıklama Görevleri için Metrikler (Image Captioning)

Bu metrikler, modelin bir resim için ürettiği metnin, insan tarafından yazılmış referans metne ne kadar benzediğini ölçer.

##### 3.6.1.1. BLEU (Bilingual Evaluation Understudy)

Başlangıçta makine çevirisi için önerilmiş bir metriktir ve üretilen cümlelerin referans cümlelerle paylaştığı  $n$ -gram sayısını ölçer. Aday cümledeki kaç kelimenin referans metinde (gerçek metin) geçtiğini dikkate alarak, hedef metnin referans metne göre doğruluğunu hesaplar.

$$BLEUScore = BP * \exp\left(\sum_{i=1}^N W_i * \ln(p_i)\right)$$

- *BP (Brevity Penalty)*: Aday cümlelerin referans cümleye göre aşırı kısa olmasını cezalandıran terimdir.
- $w_i$ :  $i$ . dereceden n-gram hassasiyetinin ağırlığıdır. Genellikle tüm n-gramlar için eşit seçilir:

$$w_i = \frac{1}{N}$$

- $p_i$ :  $i$ . dereceden değiştirilmiş n-gram hassasiyetini (modified precision) ifade eder. Aday metindeki n-gram'ların referans metin(ler)deki karşılıklarıyla örtüşme oranını ölçer.
- $N$ : Dikkate alınan maksimum n-gram derecesidir (çoğunlukla  $N = 4$ ).

### 3.6.1.2. METEOR

Hassasiyet (precision) ve geri çağırma (recall) değerlerinin harmonik ortalamasını temel alan, ancak geri çağırma parametresine daha fazla ağırlık vererek bu bileşenleri bir ceza terimiyle çarpan bütüncül bir değerlendirme metriğidir. Tek bir parametreye odaklanan geleneksel metriklerin aksine, her iki veri kaynağından gelen bilgileri sentezleyerek daha kapsamlı bir analiz sunar. Eşanlımlı kelimeleri ve morfolojik varyasyonları dikkate alma yeteneği sayesinde cümle düzeyinde BLEU metriğine kıyasla daha esnek ve korelasyonu yüksek sonuçlar üretmekle birlikte; anlamsal eşdeğerliği tam olarak garanti edememesi, metriğin temel sınırlılığı olarak değerlendirilmektedir.

### 3.6.1.3. ROUGE-L

Otomatik metin özetlerini değerlendirmek için geliştirilmiştir. Bu metriğin bir varyasyonu olan ROUGE-L, en uzun ortak alt dizi (Longest Common Subsequence - LCS) yöntemine dayanarak metinler arasındaki yapısal örtüşmeyi hesaplamaktadır. Kapsam yönü oldukça güçlü olan bu yaklaşım, uzun cümleleri lehine değerlendirme eğilimi göstermekte ve serbest anlam değişimlerini (paraphrase) yakalamada belirli sınırlılıklar barındırmaktadır.

$$ROUGE-L = F\beta = \frac{(1 + \beta^2) * P * R}{\beta^2 * P + R}$$

- *P (Precision)*: En uzun ortak alt diziye (LCS) dayalı hassasiyet ölçüsüdür.
- *R (Recall)*: En uzun ortak alt diziye (LCS) dayalı geri çağırma ölçüsüdür.

- $\beta$ : Precision ve recall arasındaki dengeyi kontrol eden parametredir. Uygulamada genellikle  $\beta = 1$  olarak seçilir ve bu durumda precision ile recall eşit ağırlıklandırılır

#### 3.6.1.4. CIDEr (Consensus-based Image Description Evaluation)

Görüntü altyazılama (image captioning) görevlerine özgü olarak geliştirilmiş, konsensüs temelli bir değerlendirme metriğidir. Bu yöntem, referans altyazılardaki n-gram dizilerini TF-IDF ağırlıklandırma tekniğiyle analiz ederek, aday cümle ile referans kümesi arasındaki benzerliği kosinüs benzerliği (cosine similarity) üzerinden hesaplamaktadır. İnsan değerlendirmesiyle yüksek korelasyon sergilemesi ve ayırt edici terimleri ön plana çıkarması bakımından oldukça etkili bir ölçüt olan CIDEr; buna karşın, referans setinde yer almayan ancak görüntü içeriğiyle anlamsal olarak örtüşen alternatif ifadeleri düşük puanlandırma eğilimi göstermektedir.

#### 3.6.1.5. SPICE (Semantic Propositional Image Caption Evaluation)

Metni nesne, öznitelik ve ilişkilerden oluşan sahne grafiklerine dönüştürerek anlamsal doğruluğu ölçen bir metriktir. İnsan yargılarıyla yüksek korelasyon (0,88) sergilemesiyle CIDEr ve METEOR gibi metriklerden ayrılan bu yöntem, içerik zenginliğini başarılı bir şekilde ödüllendirmektedir. Ancak dilbilgisel akıcılığı göz ardı etmesi ve yüksek hesaplama maliyeti sebebiyle, kapsamlı bir değerlendirme için genellikle BLEU gibi n-gram temelli metriklerle birlikte kullanılması tercih edilmektedir.

Değerlendirme metrikleri objektif bir ölçüt sunsa da BLEU, ROUGE ve CIDEr gibi n-gram temelli yöntemler, anlamsal eşdeğerliği ve görsel sadakati ölçmede sınırlı kalabilmektedir. Özellikle modellerin görüntüde bulunmayan nesnelere üretmesi (halüsinasyon) durumunda bu metrikler yetersiz kalabilmektedir. Bu nedenle, içerik tutarlılığını daha hassas analiz edebilmek için SPICE gibi anlamsal metriklerin veya doğrudan görüntü içeriğine dayalı doğruluk kontrollerinin kullanımı kritik önem arz etmektedir.

### 3.6.2. Görsel Soru-Cevaplama (VQA) için Metrikler

VQA görevlerinde model performansını çok boyutlu analiz edebilmek amacıyla, standart doğruluk ölçütlerinin yanı sıra görevin niteliğine (örn. metin okuma, anlamsal yakınlık) özgü farklı metrikler de kullanılmaktadır:

### 3.6.2.1. Accuracy (Doğruluk)

VQA literatüründeki en temel ölçüttür. Özellikle çoktan seçmeli veya kapalı uçlu sorularda, model çıktısının yer gerçekliği (ground truth) etiketiyle eşleşme durumunu ikili (binary) bir sistem üzerinden raporlar.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Burada TP (True Positive) ve TN (True Negative) sırasıyla doğru pozitif ve doğru negatif tahminleri; FP (False Positive) ve FN (False Negative) ise yanlış pozitif ve yanlış negatif tahminleri ifade etmektedir. Accuracy metriği, modelin tüm örnekler üzerindeki doğru sınıflandırma oranını ölçerek genel performansını özetler.

### 3.6.2.2. VQA Score (Consensus Accuracy)

Veri setinde tek bir “altın cevap” yerine birden fazla geçerli insan yanıtının bulunduğu durumlarda kullanılır. Modelin başarısı, ürettiği cevabın insan annotatörler arasındaki uzlaşa (konsensüs) kümesine olan yakınlığına göre değerlendirilir.

### 3.6.2.3. Exact Match (EM)

Model tarafından üretilen yanıtın referans yanıt ile karakter dizilimi bakımından birebir örtüşmesini şart koşar. Kısa ve kesin cevaplı görevlerde etkili bir ölçüt olmakla birlikte, morfolojik varyasyonlara veya eşanlamlı kullanımlara karşı toleransı düşüktür.

### 3.6.2.4. F1-Score

Yanıtın birden fazla kelimededen oluştuğu durumlarda, tahmin edilen ve referans metin arasındaki sözcük örtüşmesini hassasiyet (precision) ve geri çağırma (recall) metriklerinin harmonik ortalaması üzerinden dengeli bir şekilde ölçer.

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Burada Precision, model tarafından üretilen sözcüklerin ne kadarının referans yanıtta yer aldığını; Recall ise referans yanıtta sözcüklerin ne kadarının model tarafından doğru şekilde üretildiğini ifade eder.

### 3.6.2.5. ANLS (*Average Normalized Levenshtein Similarity*)

Özellikle görseldeki metni okumayı gerektiren (OCR-VQA) görevlerde tercih edilir. Tahmin ile referans metin arasındaki yazım benzerliğini Levenshtein düzenleme mesafesi üzerinden normalize ederek hesaplar; böylece küçük OCR hatalarına veya yazım yanlışlarına karşı esnek bir değerlendirme sağlar.

### 3.6.2.6. WUPS (*Wu-Palmer Similarity*)

Kelime düzeyindeki anlamsal eşdeğerliği ölçmek için WordNet taksonomisini kullanır. Model yanıtı ile referans arasındaki hiyerarşik anlamsal yakınlığı hesaplayarak, tam eşleşme olmasa dahi anlamsal açıdan doğru (eş anlamlı) yanıtları kısmi puanla ödüllendirir.

### 3.6.2.7. Top-k Accuracy

Geniş yanıt uzayına sahip veya olasılıksal dağılım üreten modellerde, doğru yanıtın modelin en yüksek olasılık atadığı ilk k tahmin içerisinde bulunma oranını ifade eder.

### 3.6.2.8. Human Evaluation (*İnsan Değerlendirmesi*)

Otomatik metriklerin anlamsal nüansları veya gerekçelendirmeyi (reasoning) yakalamada yetersiz kaldığı açık uçlu sorularda uygulanır. Yanıtlar; görsel dayanaklandırma (grounding), tutarlılık ve bağlamsal doğruluk gibi nitel kriterler üzerinden insan hakemler tarafından puanlanır.

## 3.6.3. Görsel-Metin Eşleştirme ve Retrieval Metrikleri

### 3.6.3.1. Recall@K ( $R@1$ , $R@5$ , $R@10$ )

Her sorgu için doğru eşleşmenin ilk K sonuç içinde yer alma oranını ölçer; retrieval çalışmalarında en sık raporlanan metriktir ve “ilk sonuçlar ne kadar isabetli?” sorusuna yanıt verir.

### 3.6.3.2. Median Rank ( $MedR$ )

Doğru eşleşmenin sıralamadaki konumlarının medyanıdır; “tipik olarak doğru eşleşme kaçınıcı sırada geliyor?” bilgisini verir ve düşük olması daha iyidir.

### 3.6.3.3. Mean Rank ( $MR$ )

Doğru eşleşmenin sıralamadaki konumlarının ortalamasıdır; düşük değer daha iyidir ancak aykırı (çok geride kalan) örneklerden daha fazla etkilenir, bu yüzden çoğunlukla MedR ile birlikte raporlanır.

#### 3.6.3.4. Mean Reciprocal Rank (MRR)

Doğru eşleşme üst sıralarda geldikçe daha yüksek puan verir ( $1/\text{rank}$ ); özellikle her sorgu için tek doğru eşleşmenin olduğu senaryolarda sıralama kalitesini iyi özetler.

#### 3.6.3.5. $n\text{DCG}@K$ (Normalized Discounted Cumulative Gain)

İlgili sonuçları üst sıralarda bulmayı daha fazla ödüllendirir; bir sorgu için birden fazla ilgili örnek olduğunda  $R@K$ 'ya göre daha anlamlı bir sıralama metriğidir.

#### 3.6.3.6. $mAP$ (mean Average Precision)

Sıralama boyunca ilgili sonuçları yakalama başarısını tek skorla özetler; çoklu ilgili örnek bulunan retrieval senaryolarında hem kapsama hem de sıralama kalitesini birlikte yansıtır.

#### 3.6.3.7. $\text{Precision}@K$ ( $P@K$ )

İlk  $K$  sonucun ne kadarının ilgili olduğunu ölçer; çoklu ilgili etiketlemesi varsa değerlidir, tek doğru eşleşme senaryosunda genellikle  $R@K$  daha yaygındır.

#### 3.6.3.8. $R\text{-Precision}$

İlgili sonuç sayısı  $R$  ise, ilk  $R$  sonuçtaki precision değeridir; ilgili örnek sayısının sorguya göre değiştiği durumlarda sıralama kalitesini dengeli biçimde değerlendirmek için kullanılır.

### 3.6.4. Çok Modlu Üretim Görevlerinde Bütüncül Değerlendirme Metrikleri

Multimodal üretim modellerinin başarısı, günümüzde  $n$ -gram tabanlı yüzeysel ölçümlerin ötesine geçerek; anlamsal, yapısal ve olgusal tutarlılığı hedefleyen çok katmanlı bir değerlendirme rejimine dayanmaktadır. Literatürdeki temel yaklaşımlar şunlardır:

#### 3.6.4.1. Gömme (Embedding) Tabanlı Metrikler

Üretilen içerik ile referans veri arasındaki ilişkiyi kelime örtüşmesi yerine, vektör uzayındaki anlamsal izdüşümler üzerinden analiz eder. BERTScore ve CLIPScore gibi yöntemler, sözdizimsel farklılıklara rağmen bağlamsal uyumu yakalayarak, modelin “anlama” kapasitesini ölçümler.

### 3.6.4.2. Yapısal ve İlişkisel Doğruluk

Dilbilgisi akıcılığından ziyade sahne içeriğine ve nesne ilişkilerine odaklanır. SPICE gibi metrikler, görseldeki varlıkların ve niteleyicilerin metne doğru aktarılıp aktarılmadığını bir “sahne grafiği” üzerinden denetleyerek, içerik bütünlüğünü ve olgusal doğruluğu sağlar.

### 3.6.4.3. Halüsinasyon ve Görsel Sadakat

Modelin görüntüde bulunmayan nesnelere üretme (halüsinasyon) sorununu ele alır. Klasik benzerlik testlerinin tespit edemediği bu fabrikasyon hataları, görsel sadakat (faithfulness) ilkesini temel alan özelleşmiş metriklerle (örn. CHAIR, POPE) saptanarak modelin güvenilirliği sınanır.

### 3.6.4.4. Model Tabanlı Değerlendirme (LLM-as-a-Judge)

Karmaşık ve açık uçlu görevlerde, gelişmiş dil modelleri (LLM) tanımlı rubrikler ışığında birer “hakem” olarak kullanılır. İnsan muhakemesine en yakın nüanslı sonuçları sunan bu yöntem, maliyet etkinliği ve ölçeklenebilirliği sayesinde modern araştırmalarda yaygınlaşmaktadır.

### 3.6.4.5. İnsan Değerlendirmesi

Otomatik metriklerin validasyonu için “altın standart” kabul edilir. Maliyet ve süre kısıtlarına rağmen; doğruluk, akıcılık ve bağlamsal nüansların en kesin ölçümü için hibrit değerlendirme süreçlerinin vazgeçilmez bir parçasıdır.

## 3.6.5. Görsel Dil Modelleri İçin Yeni Metrik Önerileri ve Benchmark Setleri

Görsel dil modellerinin değerlendirilmesinde klasik metrikler (ör. captioning’de BLEU/CIDEr, retrieval’da R@K) tek başına yeterli görülmemektedir; çünkü bu ölçütler yüksek skor üretse bile modelin görselde olmayan ayrıntıları eklemesi veya görsel kanıtla zayıf bağ kurması her zaman yakalanamayabilir. Bu nedenle literatürde eğilim, görsel sadakat/grounding ve muhakeme boyutlarını daha doğrudan ölçen değerlendirme kurulumlarına yönelmiştir.

Bu doğrultuda MMMU, MMBench ve MathVista gibi yeni nesil benchmark’lar modelin yalnızca görsel içerik algısını değil, çok-disiplinli bilgi kullanımı, sağlamlık ve matematiksel–mantıksal görsel muhakeme kapasitesini de sınamayı hedefler; HallusionBench ise halüsinasyon ve yanıltıcı görsel ipuçlarına karşı dayanıklılığı görünür kılar. Metrik tarafında POPE/H-POPE benzeri yaklaşımlar nesne/özellik düzeyinde halüsinasyonu daha kararlı biçimde ölçmeye odaklanırken, açık uçlu çıktılarda tek bir doğru tanımının zorlaşması

nedeniyle rubrik tabanlı model-hakemli değerlendirme (LLM/VLM-as-a-judge) kullanımını da yaygınlaştırmaktadır. Genel olarak değerlendirme pratikleri, tek bir skor yerine doğruluk, sadakat ve muhakemeyi birlikte raporlayan çok boyutlu çerçevelere doğru evrilmektedir.

### 3.7. Görsel Dil Modellerinin Kullanım Alanları

#### 3.7.1. Görsel Dil Modellerinin Otonom Sürüş Sistemlerindeki Kullanımı

##### 3.7.1.1. Otonom Algı, Sahne Anlama ve Çevresel Farkındalık

Otonom sistemlerde güvenli ve bağlama duyarlı karar verme, çevrenin yalnızca geometrik olarak algılanmasını değil; sahnedeki nesnelere, insanların ve çevresel koşulların anlamsal olarak yorumlanmasını gerektirmektedir. Bu gereksinim, görsel temsiller ile dil tabanlı muhakemeyi birleştiren Vision-Language Model (VLM) yaklaşımlarını otonom algı literatüründe önemli bir konuma taşımıştır. Son yıllarda yapılan çalışmalar, VLM'lerin 3B nesne tespiti, insan odaklı çevresel algı ve geniş ölçekli sokak görünümünü analizlerinde açık sözcük dağılımına dayalı, sıfır atışlı ve komutla yönlendirilebilir yetenekler sunduğunu göstermektedir (Sapkota et al., 2025; Greco et al., 2025; Peng et al., 2025).

3B nesne tespiti literatürünü VLM perspektifiyle sistematik biçimde inceleyen Sapkota ve arkadaşları (2025), geleneksel nokta bulutu ve voksel tabanlı yöntemlerin (PointNet++, VoteNet, PV-RCNN vb.) yoğun veri etiketleme gereksinimi, kapalı sınıf öğrenme yapısı ve semantik genelleme eksikliği gibi temel sınırlılıklarına dikkat çekmektedir. Derleme çalışması, bu sınırlamaları aşmaya yönelik geliştirilen yaklaşımları kapsamlı biçimde ele almaktadır. Dil-görsel hizalama temelli yöntemler (CLIP, PointCLIP v2), çok modlu büyük modeller (PaLM-E, BLIP-2, LLaVA, CogVLM) ve 3B sahnede dil temellendirme odaklı sistemler (3D-LLM, SpatialVLM, Cube R-CNN) bu çerçevede incelenmektedir. Bu modellerin ortak özelliği, görsel özellikleri dil uzayına projekte ederek referans ifadelerini çözümleyebilmesi ve açık sınıf tanıma yoluyla semantik temellendirmeyi güçlendirmesidir. OMNI3D (234K görüntü, 3M nesne etiketi), ScanNet, ScanRefer ve Nr3D/Sr3D veri setlerinde raporlanan sonuçlar; daha önce görülmemiş nesnelere tespiti ve referans tabanlı 3B algıda anlamlı performans artışları göstermektedir. Örneğin CoDA çerçevesinde novel sınıflarda %80'e varan mAP artışı, 3DVLP modelinde ScanRefer üzerinde IoU@0.25 doğruluğunun %51.70'e ulaşması bu semantik genellemenin nicel göstergeleridir. Bununla birlikte çalışma, VLM tabanlı 3B sistemlerin yüksek hesaplama maliyeti, düşük kare hızları (örn. Instruct3D ≈

8 FPS) ve mekânsal halüsinasyon riski nedeniyle gerçek zamanlı otonom sürüş uygulamalarında doğrudan kullanımının sınırlı kaldığını vurgulamaktadır.

İnsan odaklı çevresel farkındalık alanında Greco ve arkadaşları (2025), yaya nitelik tanıma (Pedestrian Attribute Recognition, PAR) problemini VLM tabanlı bir çerçevede deneysel olarak değerlendirmiştir. Çalışmada çoklu nitelik çıkarımı klasik çok-etiketli sınıflandırma yaklaşımı yerine VQA formatında ele alınmıştır. Model, doğal dil soruları üzerinden aynı anda birden fazla niteliği çıkarabilmektedir. PAR 2023 değerlendirmelerinde MIVIA PAR, UPAR ve SYNTH-PEDES veri setleri üzerinde BLIP-2 tabanlı yaklaşım 0.921 ortalama doğruluk (mA) ile en yüksek performansı göstermiştir. Buna karşılık CNN tabanlı bir model 0.709 mA doğruluk üretmiş ancak 150 FPS hızına ulaşarak gerçek zamanlılık avantajı sağlamıştır. VLM tabanlı yaklaşımın <0.5 FPS işlem hızında kalması, doğruluk üstünlüğüne rağmen gerçek zamanlı video analizi açısından önemli bir sınırlılık oluşturmaktadır. Bu bulgular, görsel–dil ön-eğitiminin semantik nitelik çıkarımında avantaj sağladığını ancak hesaplama maliyetinin mühendislik açısından kritik olduğunu göstermektedir.

Daha geniş ölçekli çevresel farkındalık bağlamında Peng ve arkadaşları (2025), Street View Analytics literatürünü sistematik olarak incelemiş ve sokak görünümü görüntülerinin çok modlu analizine odaklanmıştır. Google Street View, Baidu Street View, Tencent Maps ve Mapillary gibi platformlardan elde edilen görüntüler; CLIP, BLIP-2, LLaVA, GPT-4V ve GPT-3.5 + görsel kodlayıcı kombinasyonları ile analiz edilerek kentsel güvenlik algısı, yürünebilirlik, ulaşım yoğunluğu ve sosyoekonomik göstergeler gibi üst düzey kavramların tahmini gerçekleştirilmiştir. Bu yaklaşımlar, görüntüye metin istemleri eşliğinde açık alan bilgisini entegre ederek yüksek seviyeli semantik çıkarım üretmektedir. Manuel etiketleme ihtiyacının azalması ve zengin semantik yorum üretimi önemli avantajlar sağlarken; verinin zamansal güncelliği, bağlamsal hassasiyet ve model kararlarının açıklanabilirliği temel sınırlılıklar olarak belirtilmektedir.

Genel olarak değerlendirildiğinde, Vision–Language Modelleri otonom algı, sahne anlama ve çevresel farkındalık alanlarında açık sözcük dağarcıklı genelleme, referans çözümleme ve insan merkezli semantik yorum üretimi açısından geleneksel yöntemlere kıyasla belirgin kazanımlar sağlamaktadır. Bununla birlikte yüksek hesaplama maliyeti, düşük kare hızları ve gecikme gibi mühendislik kısıtları nedeniyle bu modellerin kısa vadede bağımsız gerçek zamanlı algı modülleri olarak değil, üst seviye semantik muhakeme ve karar destek bileşenleri olarak konumlandırılması daha uygulanabilir bir yaklaşım olarak görünmektedir (Sapkota et al., 2025; Greco et al., 2025; Peng et al., 2025). Tablo 3’de Vision–Language Modellerinin otonom algı ve çevresel farkındalık görevlerinde kullanımının karşılaştırmalı analizi verilmiştir.

Tablo 3. Vision-Language Modellerinin otonom algı ve çevresel farkındalık görevlerinde kullanımı

Özellik	3D Nesne Tespiti ve Görsel Dil Modeller	Yaya Nitelik Tanıma (PAR)	Sokak Görünümü Analitiği (SVA)
İlgili Çalışma	<i>A Review of 3D Object Detection with Vision-Language Models</i> (Sapkota et al., 2025)	<i>An Experimental Evaluation of Smart Sensors for Pedestrian Attribute Recognition Using Multi-Task Learning and Vision Language Models</i> (Greco et al., 2025)	<i>VLM-enabled Street View Analytics: A Systematic Literature Review</i> (Peng et al., 2024)
Çalışmanın Amacı	Geleneksel nokta bulutu ve voksel tabanlı 3D tespit yöntemlerinin sınırlamalarını aşarak, açık sözcük dağarcığı (open-vocabulary), sıfır atışlı ve dil komutlarına duyarlı 3D nesne algısı sağlamak	Düşük çözünürlüklü güvenlik kamerası görüntülerinden yaya niteliklerini (cinsiyet, kıyafet rengi, aksesuar vb.) çoklu görevli ve VLM tabanlı yaklaşımlarla doğru biçimde tanımak	Sokak görünümü görüntülerini kullanarak kentsel güvenlik, yürünebilirlik, çevresel algı ve sosyoekonomik göstergeleri çok modlu biçimde analiz etmek
Kullanılan Model / Mimari	Geleneksel: PointNet++, VoteNet, PV-RCNN VLM tabanlı: PaLM-E, BLIP-2, LLaVA, CogVLM, CLIP, PointCLIP v2, 3D-LLM, SpatialVLM, Cube R-CNN	Kazanan yaklaşım: <b>BLIP-2 (VQA olarak kullanılmış)</b> Diğerleri: ResNet50 + PLIP, Swin Transformer varyantları, BeitV2, EfficientNetV2	GPT-4V, GPT-3.5 + görsel kodlayıcılar, CLIP, BLIP-2, LLaVA, LLaMA
Temel Teknik Yaklaşım	Dil-görsel hizalama ile 3D sahnelerde semantik temellendirme; sıfır atışlı tespit; metin tabanlı referans çözümleme (ör. “kırmızı sandalyenin yanındaki çanta”)	Problemin görsel sınıflandırma yerine <b>VQA olarak formüle edilmesi</b> ; çoklu niteliklerin tek modelle çıkarılması	Görüntü + metin istemleri ile yüksek seviyeli çevresel çıkarım; açık alan bilgisi (open-domain knowledge) ile kentsel yorumlama
Kullanılan Veri Setleri	KITTI, SUN RGB-D, Waymo Open Dataset, ScanNet, ScanRefer, Nr3D/Sr3D, <b>OMNI3D</b> (234K görüntü, 3M nesne etiketi)	Eğitim: MIVIA PAR (105.244 görüntü), UPAR, SYNTH-PEDES Test: 20.000 görüntülük gizli test seti	Google Street View (GSV), Baidu Street View, Tencent Maps, Mapillary + OpenStreetMap (OSM)
Değerlendirme Metrikleri	mAP, IoU@0.25 / IoU@0.5, doğruluk, FPS	Ortalama doğruluk (mA), standart sapma, FPS	Uygulamaya bağlı metrikler (sınıflandırma doğruluğu, korelasyon, nitel analiz)
Öne Çıkan Nicel Sonuçlar	CoDA: Novel object discovery’de %80’e varan mAP artışı 3DVLP: ScanRefer’da %51.70 doğruluk (IoU@0.25)	BLIP-2 kullanan iROC-ULPGC: <b>0.921 mA</b> (en yüksek doğruluk) Baseline CNN: 0.709 mA fakat 150 FPS	Görsel dil modeller, kentsel güvenlik ve yürünebilirlik gibi kavramları insan algısına yakın biçimde modelleyebilmiştir

<b>Avantajlar</b>	Sıfır atışlı öğrenme, dil komutlarına duyarlılık, anlamsal genelleme	Çok yüksek doğruluk, tek modelle çoklu nitelik çıkarımı	Manuel etiket ihtiyacını azaltma, zengin semantik açıklamalar
<b>Sınırlılıklar</b>	Yüksek hesaplama maliyeti, düşük FPS (ör. Instruct3D $\approx$ 8 FPS), mekansal halüsinasyonlar	Gerçek zamanlı kullanım için çok yavaş ( $<0.5$ FPS)	Zamansal güncellik, bağlamsal hassasiyet ve karar sürecinde şeffaflık eksikliği

### 3.7.1.2. Otonom Karar Verme, Planlama ve Navigasyon

Otonom sürüş sistemlerinde karar verme, planlama ve navigasyon bileşenleri; algılanan çevresel bilginin güvenli, tutarlı ve açıklanabilir sürüş davranışlarına dönüştürülmesinden sorumludur. Geleneksel modüler yaklaşımlar, karmaşık ve nadir görülen (long-tail) senaryolarda insan benzeri muhakeme gerektiren durumlarda sınırlı kalmaktadır. Son yıllarda Vision–Language Model (VLM) ve Large Language Model (LLM) tabanlı yaklaşımlar, sürüş problemini yalnızca geometrik bir optimizasyon olarak değil; dil destekli akıl yürütme ve bağlamsal karar verme çerçevesinde ele alarak bu sınırlamaları aşmayı hedeflemektedir (Liu et al., 2023; Wang et al., 2024).

Bu doğrultuda öne çıkan VLM-AD yaklaşımının temel amacı, uçtan uca otonom sürüş modellerindeki muhakeme eksikliğini gidermektir (Chen et al., 2024). Önerilen mimaride, güçlü bir VLM (GPT-4o) doğrudan bir kontrolcü olarak değil, yalnızca eğitim aşamasında bir öğretmen model olarak konumlandırılır. Sürüş sahnesi, planlanan yörünge ve araç durumu üzerinden üretilen doğal dil açıklamaları ile yapılandırılmış sürüş eylemleri; bilgi damıtma (knowledge distillation) yöntemiyle UniAD ve VAD gibi uçtan uca sürüş modellerine aktarılır. Test aşamasında VLM'nin tamamen devre dışı bırakıldığı bu tasarım, sistemin gerçek zamanlı çalışmasını sağlarken dil tabanlı muhakeme bilgisinin sürüş davranışlarına yansıtılmasına olanak tanır.

Elde edilen deneysel sonuçlar, VLM-AD mimarisinin planlama doğruluğu ve güvenlik açısından anlamlı kazanımlar sağladığını göstermektedir. nuScenes veri setinde UniAD için raporlanan ortalama L2 planlama hatası 1.03 m iken, VLM-AD ile bu değer yaklaşık 0.88–0.89 m seviyesine düşmüş; çarpışma oranı ise %0.31'den %0.19–0.24 aralığına gerilemiştir (Chen et al., 2024). CARLA simülasyonlarında yapılan kapalı döngü deneyler de dil destekli damıtmanın daha kararlı sürüş davranışları ürettiğini ortaya koymuştur.

Bu tür görev-odaklı çalışmaların metodolojik arka planı literatürdeki kapsamlı derlemelerde de vurgulanmaktadır. *A Survey on Large Language Model-Powered Autonomous Driving* çalışması, büyük dil modellerinin süreçlere entegrasyonunu inceleyerek modelleri muhakeme yeteneklerine göre sınıflandırırken (Liu

et al., 2023); *Vision Language Models in Autonomous Driving: A Survey and Outlook* çalışması görsel dil modellerinin uçtan uca sürüşteki rollerini ve dil zenginleştirilmiş veri setlerinin önemini ele almaktadır (Wang et al., 2024). Sonuç olarak; yüksek hesaplama maliyeti, gecikme ve uzamsal halüsinasyonlar gibi donanımsal ve sistemsel darboğazlar göz önüne alındığında, VLM'lerin doğrudan kontrolcü yerine eğitim aşamasında rehberlik eden bileşenler olarak kullanılması kısa vadede en uygulanabilir mimari yaklaşım olarak öne çıkmaktadır (Chen et al., 2024; Liu et al., 2023; Wang et al., 2024). Tablo 4' de otonom karar verme, planlama ve navigasyon için VLM/LLM tabanlı çalışmaların teknik karşılaştırması sunulmuştur.

**Tablo 4. Otonom Karar Verme, Planlama ve Navigasyon için VLM/LLM Tabanlı Çalışmaların Teknik Karşılaştırması**

Özellik	VLM-AD	LLM-Powered AD Survey	VLM in AD Survey
Çalışmanın Amacı	Dil tabanlı muhakemeyi uçtan uca sürüşe aktarmak	Büyük dil modellerinin otonom sürüşteki rolünü sınıflandırmak	Görsel dil modellerinin AD'deki kullanım alanlarını özetlemek
Kullanılan Modeller	GPT-4o (öğretmen) + UniAD / VAD	GPT-Driver, DriveVLM, DriveGPT4	CLIP, LLaVA, GPT-4V
Dilin Rolü	Eğitimde açıklama + eylem damıtma	Karar verme ve açıklanabilirlik	Açık sözcük dağarcıklı algı
Veri Setleri	nuScenes, CARLA	nuScenes, BDD100K, CARLA	nuScenes-QA, Talk2Car, DriveLM
Performans Metrikleri	L2 hata, çarpışma oranı	Çeşitli (planlama, güvenlik)	Görev-odaklı doğruluk
Temel Sonuç	L2 hata ↓, çarpışma ↓	Long-tail senaryolarda güçlü	Algı-planlama entegrasyonu
Kaynak	Chen et al. (2024)	Liu et al. (2023)	Wang et al. (2024)

### 3.7.1.3. Otonom Sistemler için Veri Üretimi, Etiketleme ve Edge (Uç) Uygulamaları

Otonom sistemlerin gerçek dünya koşullarında güvenilir biçimde çalışabilmesi, yalnızca gelişmiş algı ve karar verme modellerine değil; bu modellerin eğitimi için gerekli yüksek kaliteli verinin ölçeklenebilir biçimde üretilmesine ve kaynak kısıtlı uç (edge) cihazlarda çalıştırılabilmesine de bağlıdır. Vision-Language Modelleri (VLM), bu iki gereksinimi birlikte ele alabilen yaklaşımlar sunarak literatürde giderek daha merkezi bir rol üstlenmektedir.

Otonom sürüş alanında en yüksek maliyetli süreçlerden biri olan 3D nesne etiketleme problemine çözüm getirmek amacıyla Ma ve arkadaşları, VLM

destekli yarı-otomatik bir veri üretim çerçevesi önermektedir (Ma et al., 2026). Önerilen sistemin teknik mimarisinde, stereo görüntülerden derinlik tahmini yapan PSMNet, görüntü kenar bilgisi ile seyrek LiDAR ölçümlerini birleştiren Frustum Graph Depth Correction (FGDC) modülü ve Transformer tabanlı TransFusion 3D dedektöründen oluşan çok aşamalı bir boru hattı (pipeline) kurgulanmıştır. Bu yapıda, seyrek LiDAR verileri kamera derinlik bilgisiyle zenginleştirilerek yoğun bir pseudo-LiDAR temsili oluşturulur. Elde edilen 3D sınırlayıcı kutular (bounding box), 2D görüntü düzlemine yansıtılarak GLM-4 tabanlı bir VLM'e ardışık soru-cevap istemleriyle (örneğin “kutu bir yaya içeriyor mu?”) sunulur ve hatalı etiketler semantik olarak filtrelenir.

Bu sistemin deneysel sonuçları, VLM tabanlı doğrulama mekanizmasının manuel etiketlemeye kıyasla yaklaşık yirmi kat hızlanma sağladığını ve %99'un üzerinde filtreleme hassasiyetine ulaştığını kanıtlamaktadır. Ayrıca, otomatik üretilmiş bu etiketlerle eğitilen algı modellerinin özellikle araç tespitinde belirgin performans artışları sergilediği rapor edilmiştir (Ma et al., 2026).

Veri üretimindeki bu başarıların pratik sistemlere entegrasyonu, VLM'lerin uç cihazlarda çalıştırılabilmesine yönelik mühendislik çözümlerini zorunlu kılmaktadır. Sharshar ve arkadaşları tarafından sunulan kapsamlı derleme çalışması, büyük ölçekli görsel dil modellerinin IoT, mobil ve gömülü sistemler gibi kaynak kısıtlı platformlara nasıl uyarlanabileceğini sistematik olarak incelemeyi hedeflemektedir (Sharshar et al., 2024). Analiz sonuçlarına göre, büyük VLM'lerin doğrudan uç cihazlara taşınması gecikme, bellek ve enerji tüketimi açısından sürdürülebilir değildir. Bu darboğazı aşmak için kurgulanan uç cihaz mimarilerinde, bilgi damıtma, niceleme ve token azaltma gibi model sıkıştırma teknikleri merkeze alınmaktadır. Örneğin, MobileVLM-V2 gibi yapılar, görsel belirteç sayısını azaltan hafif projeksiyon katmanları sayesinde mobil donanımlarda kabul edilebilir gecikmelerle çok modlu çıkarım yapılmasına olanak tanır.

Ayrıca bu çalışma, sürüş verilerinin merkezi sunuculara aktarılmadan işlenmesini sağlayan federe öğrenme (federated learning) yaklaşımlarının gizlilik risklerini ve ağ gecikmelerini ortadan kaldırdığını; ancak uç cihaz mimarilerindeki bu yerel modellerin fiziksel saldırılara ve model zehirlenmelerine karşı hala savunmasız olduğunu vurgulamaktadır (Sharshar et al., 2024). Genel bir değerlendirmeye; görsel dil modelleri bir yandan etiketleme süreçlerinde insan maliyetini düşüren semantik denetleyiciler olarak işlev görürken, diğer yandan kısa vadede tam ölçekli genel modeller yerine, sisteme entegre edilmiş, damıtılmış ve görev-öзgü mimariler olarak uç cihazlarda gerçek zamanlı algı yetenekleri sunmaktadır (Ma et al., 2026; Sharshar et al., 2024). Otonom sistemlerde veri üretimi ve uç uygulamalar için Vision-Language Model yaklaşımlarının karşılaştırılması Tablo 5' de verilmiştir.

**Tablo 5. Otonom Sistemlerde Veri Üretimi ve Uç Uygulamalar için Vision-Language Model Yaklaşımlarının Karşılaştırılması**

Özellik	VLM Destekli Otomatik 3D Nesne Etiketleme	Uç Ağlar için Vision-Language Modelleri (VLM)
<b>Kaynak</b>	Ma et al. (2026)	Sharshar et al. (2024)
<b>Çalışmanın Amacı</b>	Otonom sürüş için yüksek maliyetli manuel 3D etiketleme sürecini azaltmak; kamera ve LiDAR verilerini kullanarak yarı-otomatik ve güvenilir 3D etiket üretmek	Büyük ölçekli görsel dil modellerinin kaynak kısıtlı uç (edge) cihazlarda düşük gecikme, düşük enerji tüketimi ve gizlilik korunarak çalıştırılabilmesini sağlamak
<b>Temel Yaklaşım</b>	Pseudo-LiDAR üretimi + Transformer tabanlı 3D tespit + VLM tabanlı semantik doğrulama ile kapalı döngü etiketleme	Model sıkıştırma, hafif mimariler ve federated learning kullanarak görsel dil modelleri uç cihazlara uyarlama
<b>Kullanılan VLM / LLM</b>	GLM-4 (etiket doğrulama ve semantik filtreleme için)	CLIP, MobileVLM-V2, Moondream2, MiniVLM, EdgeVL
<b>Görsel Algı Modelleri</b>	PSMNet (stereo derinlik), FGDC (derinlik düzeltme), TransFusion (3D nesne tespiti)	CLIP-tabanlı görsel kodlayıcılar, tek-akışlı ve token-azaltmalı VLM mimarileri
<b>Teknik Yenilik</b>	<ul style="list-style-type: none"> <li>Stereo görüntülerden pseudo-LiDAR üretimi</li> <li>3D kutuların 2D projeksiyonu</li> <li>VLM ile ardışık soru-cevap tabanlı etiket doğrulama</li> </ul>	<ul style="list-style-type: none"> <li>Knowledge distillation (bilgi damıtma)</li> <li>Quantization (8-bit ve altı)</li> <li>Lightweight Downsampling Projection (LDP)</li> <li>Federated learning</li> </ul>
<b>Kullanılan Veri Setleri</b>	GA Dataset, TJ Dataset (gerçek dünya); KITTI (kıyaslama); nuScenes (farklı hava/gece koşulları)	Otonom sürüş, sağlık ve uzaktan algılama alanlarından çeşitli veri setleri (AerialVLN, MedBLIP, ChangeCLIP vb.)
<b>Değerlendirme Metrikleri</b>	Etiket doğrulama hassasiyeti (Precision), Ortalama Doğruluk (AP), Etiketleme süresi	Gecikme (latency), bellek kullanımı, enerji tüketimi, doğruluk, cihaz uyumluluğu
<b>Nicel Sonuçlar</b>	<ul style="list-style-type: none"> <li>Manuel etiketlemeye göre <math>\sim 20\times</math> hızlanma (300 sn <math>\rightarrow</math> 15.2 sn)</li> <li>VLM doğrulama hassasiyeti <math>&gt; 99\%</math></li> <li>BEVDet performansı: büyük araçlarda <math>+38\%</math> AP, küçük araçlarda <math>+15.6\%</math> AP</li> </ul>	<ul style="list-style-type: none"> <li>MobileVLM ve Moondream2 gibi modellerle uç cihazlarda gerçek zamanlı çıkarım mümkün</li> <li>Bulut bağımlılığı ve ağ gecikmesi belirgin biçimde azaltılmış</li> </ul>
<b>Avantajlar</b>	Yüksek kaliteli etiket üretimi, düşük insan maliyeti, ölçeklenebilir veri üretimi	Düşük gecikme, gizlilik korunumu, enerji verimliliği
<b>Sınırlılıklar</b>	Büyük görsel dil modeller sunucu tarafında çalışmak zorunda	Sıkıştırma sonrası anlamsal kapasite kaybı, güvenlik (model poisoning) riskleri
<b>Otonom Sistemlere Katkı</b>	Algı modelleri için ölçeklenebilir ve güvenilir eğitim verisi üretimi	Gerçek zamanlı, gizlilik dostu ve sahaya uyarlanabilir VLM tabanlı algı

### 3.7.2. Görsel Dil Modellerinin Robotik Sistemlerindeki Kullanımı

#### 3.7.2.1. VLM/VLA Tabanlı Robot Kontrolü ve Gerçek Dünya Uygulamaları

Robotik alanda Vision–Language–Action (VLA) modelleri, görsel algı (vision), dil temsilleri (language) ve eylem üretimini (action) tek bir uçtan uca öğrenme çerçevesinde birleştirilerek, robotların doğal dil ile tanımlanan görevleri yerine getirebilmesini hedeflemektedir. Bu yaklaşım, klasik modüler robotik mimarilerde ayrı ayrı ele alınan algı, planlama ve kontrol bileşenlerinin, yüksek seviyeli anlamsal temsiller aracılığıyla doğrudan ilişkilendirilmesini amaçlamaktadır.

Bu doğrultuda geliştirilen RT-2 (Robotics Transformer-2) modeli, büyük ölçekli görsel–dil modellerinin web verilerinden edindiği semantik bilginin robotik kontrol politikalarına aktarılabilirliğini gösteren önemli bir örnek olarak rapor edilmiştir (Brohan et al., 2023). RT-2’de robot eylemleri, ayrıık metin belirteçleri (tokens) olarak temsil edilmekte; böylece robot kontrol problemi, doğal dil üretimine benzer bir ardışık tahmin problemi şeklinde ele alınmaktadır. Bu kapsamda robotun altı serbestlik dereceli (6-DoF) konumu, yönelimi ve tutucu (gripper) durumu ayrıklaştırılarak modelin çıktısı kelime dağarcığına dâhil edilmektedir (Brohan et al., 2023). Bu yaklaşım literatürde action-as-text tokenization olarak adlandırılmaktadır.

Mimari açıdan RT-2, PaLI-X (5B ve 55B parametre) ve PaLM-E (12B parametre) gibi büyük ölçekli Vision–Language Model (VLM) omurgaları üzerine inşa edilmiştir (Brohan et al., 2023). Eğitim sürecinde robotik trajektori verileri ile web ölçekli görsel–dil görevlerinin (ör. görsel soru-cevaplama ve görüntü altyazılama) birlikte kullanıldığı bir co-fine-tuning stratejisi uygulanmaktadır. Bu strateji, modelin robotik görevlerde uzmanlaşırken, web verilerinden öğrenilen genel semantik bilgiyi korumasını amaçlamaktadır.

Deneyisel değerlendirmelerde RT-2’nin masüstü manipülasyon görevlerinde (ör. pick-and-place, nesne ayırma ve hedefe taşıma) görülmemiş nesnelere, arka planlar ve dilsel komutlar altında önceki modellere kıyasla daha yüksek başarı oranları elde ettiği rapor edilmiştir (Brohan et al., 2023). Çalışmada, bazı zero-shot senaryolarda başarı oranının önceki yaklaşımlara göre 2 ila 6 kat arttığı belirtilmektedir. Ayrıca modelin, robotik eğitim verisinde açıkça yer almayan bağlamsal ve ilişkisel talimatları yorumlayabildiği gözlemlenmiştir (Brohan et al., 2023).

VLA literatürünü daha geniş bir perspektiften ele alan Sapkota ve arkadaşları, bu modelleri tekil uygulama başarılarından ziyade genel amaçlı robotik ajanlara doğru evrilen bir mimari paradigma olarak konumlandırmaktadır (Sapkota et al., 2025). İlgili survey çalışmasında 80’den fazla VLA modeli incelenmiş

ve bu modeller erken füzyon (early fusion), çift sistemli (dual-system) ve kendi kendini düzelten (self-correcting) mimariler olmak üzere üç ana grupta sınıflandırılmıştır. Çift sistemli mimarilerde, hızlı tepki veren düşük seviyeli kontrol modülleri ile daha yavaş ancak muhakeme odaklı planlama modüllerinin ayrıştırıldığı ifade edilmektedir (Sapkota et al., 2025).

Aynı çalışmada, milyarlarca parametreye sahip VLA modellerinin pratik robotik sistemlerde çalıştırılabilmesi için LoRA, niceleme (quantization) ve düşük dereceli adaptasyon gibi parametre-verimli öğrenme tekniklerinin kritik öneme sahip olduğu vurgulanmaktadır (Sapkota et al., 2025). Bununla birlikte, otoregresif belirteç üretimine dayalı VLA modellerinin genellikle 3–5 Hz gibi düşük kontrol frekanslarında çalışabildiği ve bunun gerçek zamanlı ve güvenli robot kontrolü açısından önemli bir sınırlama oluşturduğu belirtilmektedir.

Gerçek dünya uygulamalarına odaklanan Kawaharazuka ve arkadaşları ise VLA modellerini veri toplama stratejileri, robot donanımı ve eğitim paradigmaları ile birlikte ele almaktadır (Kawaharazuka et al., 2023). Bu çalışmada sensörimotor modeller, dünya modelleri ve sađlar (affordance) temelli yaklaşımlar karşılaştırılmış; özellikle sürekli eylem uzaylarında difüzyon tabanlı politika başlıklarının daha pürüzsüz ve kararlı kontrol sağladığı rapor edilmiştir (Kawaharazuka et al., 2023). Ayrıca, farklı robot platformları arasında öğrenilen temsillerin doğrudan aktarılmasının performans kaybına yol açabildiği ve embodiment transfer probleminin hâlen açık bir araştırma konusu olduğu vurgulanmaktadır.

Genel olarak değerlendirildiğinde, VLA modellerinin robotik sistemlere dil temelli muhakeme, görev genelmesi ve açıklanabilirlik kazandırdığı görülmektedir. Ancak bu kazanımların; hesaplama maliyeti, düşük kontrol frekansı, güvenlik garantileri ve donanım bağımlılığı gibi mühendislik kısıtlarıyla dengelenmesi gerektiği literatürde açıkça ifade edilmektedir (Brohan et al., 2023; Sapkota et al., 2025; Kawaharazuka et al., 2023). VLA modellerinin robotik alandaki karşılaştırmalı analizi Tablo 6' da verilmiştir.

**Tablo 6. Vision-Language-Action (VLA) Modellerinin Robotik Alandaki Karşılaştırmalı Özeti**

Özellik	VLM Destekli Otomatik 3D Nesne Etiketleme	Uç Ağlar için Vision-Language Modelleri (VLM)
<b>Kaynak</b>	Ma et al. (2026)	Sharshar et al. (2024)
<b>Çalışmanın Amacı</b>	Otonom sürüş için yüksek maliyetli manuel 3D etiketleme sürecini azaltmak; kamera ve LiDAR verilerini kullanarak yarı-otomatik ve güvenilir 3D etiket üretmek	Büyük ölçekli Görsel dil modellerinin kaynak kısıtlı uç (edge) cihazlarda düşük gecikme, düşük enerji tüketimi ve gizlilik korunarak çalıştırılabilmesini sağlamak
<b>Temel Yaklaşım</b>	Pseudo-LiDAR üretimi + Transformer tabanlı 3D tespit + VLM tabanlı semantik doğrulama ile kapalı döngü etiketleme	Model sıkıştırma, hafif mimariler ve federated learning kullanarak görsel dil modelleri uç cihazlara uyarlama
<b>Kullanılan VLM / LLM</b>	GLM-4 (etiket doğrulama ve semantik filtreleme için)	CLIP, MobileVLM-V2, Moondream2, MiniVLM, EdgeVL
<b>Görsel Algı Modelleri</b>	PSMNet (stereo derinlik), FGDC (derinlik düzeltme), TransFusion (3D nesne tespiti)	CLIP-tabanlı görsel kodlayıcılar, tek-akışlı ve token-azaltmalı VLM mimarileri
<b>Teknik Yenilik</b>	<ul style="list-style-type: none"> <li>Stereo görüntülerden pseudo-LiDAR üretimi</li> <li>3D kutuların 2D projeksiyonu</li> <li>VLM ile ardışık soru-cevap tabanlı etiket doğrulama</li> </ul>	<ul style="list-style-type: none"> <li>Knowledge distillation (bilgi damıtma)</li> <li>Quantization (8-bit ve altı)</li> <li>Lightweight Downsampling Projection (LDP)</li> <li>Federated learning</li> </ul>
<b>Kullanılan Veri Setleri</b>	GA Dataset, TJ Dataset (gerçek dünya); KITTI (kıyaslama); nuScenes (farklı hava/gece koşulları)	Otonom sürüş, sağlık ve uzaktan algılama alanlarından çeşitli veri setleri (AerialVLN, MedBLIP, ChangeCLIP vb.)
<b>Değerlendirme Metrikleri</b>	Etiket doğrulama hassasiyeti (Precision), Ortalama Doğruluk (AP), Etiketleme süresi	Gecikme (latency), bellek kullanımı, enerji tüketimi, doğruluk, cihaz uyumluluğu
<b>Nicel Sonuçlar</b>	<ul style="list-style-type: none"> <li>Manuel etiketlemeye göre <math>\sim 20\times</math> hızlanma (300 sn <math>\rightarrow</math> 15.2 sn)</li> <li>VLM doğrulama hassasiyeti <math>&gt; \%99</math></li> <li>BEVDet performansı: büyük araçlarda <math>+\%38</math> AP, küçük araçlarda <math>+\%15.6</math> AP</li> </ul>	<ul style="list-style-type: none"> <li>MobileVLM ve Moondream2 gibi modellerle uç cihazlarda gerçek zamanlı çıkarım mümkün</li> <li>Bulut bağımlılığı ve ağ gecikmesi belirgin biçimde azaltılmış</li> </ul>
<b>Avantajlar</b>	Yüksek kaliteli etiket üretimi, düşük insan maliyeti, ölçeklenebilir veri üretimi	Düşük gecikme, gizlilik korunumu, enerji verimliliği
<b>Sınırlılıklar</b>	Büyük görsel dil modeller sunucu tarafında çalışmak zorunda	Sıkıştırma sonrası anlamsal kapasite kaybı, güvenlik (model poisoning) riskleri
<b>Otonom Sistemlere Katkı</b>	Algı modelleri için ölçeklenebilir ve güvenilir eğitim verisi üretimi	Gerçek zamanlı, gizlilik dostu ve sahaya uyarlanabilir VLM tabanlı algı

### 3.7.2.2. Görsel Dil Modellerinin Robotik Algı, Anlama ve Görev Yönlendirmede Kullanımı

Robotik alanda Vision–Language–Action (VLA) modelleri, görsel algı, dil temsilleri ve eylem üretimini tek bir uçtan uca öğrenme çerçevesinde birleştirerek doğal dil ile tanımlanan görevlerin doğrudan robot kontrolüne aktarılmasını hedeflemektedir. Bu yaklaşım, klasik modüler robotik mimarilerde ayrı ele alınan algı, planlama ve kontrol bileşenlerini yüksek seviyeli semantik temsiller üzerinden bütünleştirmeyi amaçlamaktadır (Tablo 7).

Brohan ve arkadaşları (2023) tarafından geliştirilen RT-2 (Robotics Transformer-2), büyük ölçekli görsel–dil modellerinden elde edilen semantik bilginin robotik kontrol politikalarına aktarılabilirliğini göstermektedir. RT-2’de robot eylemleri ayrık metin belirteçleri olarak temsil edilmekte ve kontrol problemi, doğal dil üretimine benzer otoregresif bir tahmin süreci şeklinde modellenmektedir. Altı serbestlik dereceli (6-DoF) konum, yönelim ve tutucu durumu ayrıklaştırılarak modelin çıktı kelime dağarcığına dâhil edilmekte; bu yaklaşım literatürde action-as-text tokenization olarak adlandırılmaktadır. Mimari olarak RT-2, PaLI-X (5B ve 55B parametre) ve PaLM-E (12B parametre) omurgaları üzerine inşa edilmiştir. Eğitim sürecinde robotik trajektori verileri ile web ölçekli görsel–dil görevleri birlikte kullanılarak co-fine-tuning uygulanmış; böylece robotik uzmanlaşma sağlanırken genel semantik bilginin korunması amaçlanmıştır. Deneylerde modelin masaüstü manipülasyon görevlerinde görülmemiş nesnelere ve komutlar altında daha yüksek başarı oranları elde ettiği, bazı zero-shot senaryolarda başarının önceki yöntemlere göre 2–6 kat arttığı rapor edilmiştir (Brohan et al., 2023).

VLA literatürünü kapsamlı biçimde inceleyen Sapkota ve arkadaşları (2025), bu modelleri genel amaçlı robotik ajanlara doğru evrilen bir mimari paradigma olarak konumlandırmaktadır. Seksenin üzerinde VLA modeli; erken füzyon, çift sistemli ve kendi kendini düzelten mimariler olarak sınıflandırılmıştır. Çift sistemli yapılarda hızlı düşük seviyeli kontrol modülleri ile daha yavaş ancak muhakeme odaklı planlama modüllerinin ayrıştırıldığı belirtilmektedir. Ayrıca milyarlarca parametrelili VLA modellerinin pratik sistemlerde uygulanabilmesi için LoRA, niceleme ve düşük dereceli adaptasyon gibi parametre-verimli öğrenme tekniklerinin kritik olduğu vurgulanmaktadır. Bununla birlikte otoregresif belirteç üretimine dayalı modellerin genellikle 3–5 Hz gibi düşük kontrol frekanslarında çalışabildiği ve bunun gerçek zamanlı, güvenli robot kontrolü açısından önemli bir sınırlama oluşturduğu ifade edilmektedir (Sapkota et al., 2025).

Gerçek dünya uygulamalarına odaklanan Kawaharazuka ve arkadaşları (2023), VLA modellerini veri toplama stratejileri, robot donanımı ve eğitim

paradigmaları bağlamında değerlendirmiştir. Çalışmada sensörimotor modeller, dünya modelleri ve sağlar (affordance) temelli yaklaşımlar karşılaştırılmış; özellikle sürekli eylem uzaylarında difüzyon tabanlı politika başlıklarının daha pürüzsüz ve kararlı kontrol sağladığı rapor edilmiştir. Ayrıca farklı robot platformları arasında öğrenilen temsillerin doğrudan aktarımında performans kaybı yaşanabildiği ve embodiment transfer probleminin hâlen açık bir araştırma alanı olduğu belirtilmiştir (Kawaharazuka et al., 2023).

Genel olarak VLA modelleri, robotik sistemlere dil temelli muhakeme, görev genellemesi ve açıklanabilirlik kazandırmaktadır. Ancak bu kazanımlar; yüksek hesaplama maliyeti, düşük kontrol frekansı, güvenlik garantileri ve donanım bağımlılığı gibi mühendislik kısıtları ile dengelenmek zorundadır (Brohan et al., 2023; Sapkota et al., 2025; Kawaharazuka et al., 2023).

*Tablo 7. Görsel Dil Modellerinin Robotik Algı, Anlama ve Görev Yönlendirmede Kullanımı*

Özellik	VLM-Social-Nav	NaVid	Human-Guided Mobile Robot Navigation (HSM)	Planning with Vision-Language Models
<b>Kaynak</b>	Song et al., 2024	Zhang et al., 2024	Wang et al., 2025	Hu et al., 2023
<b>Çalışma Alanı</b>	Sosyal farkındalıklı robot navigasyonu	Haritasız görsel-dil navigasyonu (VLN-CE)	İnsan odaklı akıllı üretim ortamları	Robot destekli öğretim ve görev planlama
<b>Temel Amaç</b>	İnsan-robot etkileşimlerini dikkate alarak güvenli ve sosyal açıdan uygun rota seçimi	Harita, derinlik veya odometri olmadan yalnızca RGB video ve dil ile navigasyon	Gürültülü ve yapılandırılmamış üretim ortamlarında insan komutlarıyla güvenilir hareket	Doğal dil talimatlarını sıralı robot eylemlerine dönüştürmek
<b>Temel Yaklaşım</b>	VLM ile aday navigasyon yollarını skorlayarak seçim	Video tabanlı uçtan uca VLM; bir sonraki eylemi doğrudan üretme	Modüler mimari: 3B rekonstrüksiyon + VLM algı + LLM planlama	VLM destekli görev ayrıştırma ve planlama
<b>Kullanılan VLM / LLM</b>	CLIP-türevi VLM + skorlayıcı yapı	LLaMA-VID tabanı, Vicuna-7B	LSeg (CLIP tabanlı) + GPT-3.5	Genel amaçlı VLM + LLM
<b>Görsel Kodlayıcı</b>	CLIP tabanlı encoder	EVA-CLIP	ViT-L/16 (LSeg)	CLIP / ViT tabanlı
<b>Dil Modeli</b>	VLM içi dil bileşeni	Vicuna-7B	GPT-3.5	GPT-türevi LLM
<b>Mimari Özellik</b>	Sosyal bağlam farkındalığı	Instruction-queried & instruction-agnostic tokenlar	3 aşamalı modüler yapı (rekonstrüksiyon- algı-planlama)	Dil → görev → eylem zinciri
<b>Girdi Türü</b>	RGB görüntü + doğal dil	Monoküler RGB video + doğal dil	RGB-D video + doğal dil	Görsel gözlem + doğal dil

<b>Çıktı Türü</b>	En uygun navigasyon rotası	Düşük seviyeli yürütülebilir eylemler	Python tabanlı robot komutları	Sıralı görev ve eylem planı
<b>Veri Setleri</b>	Sosyal navigasyon senaryoları	R2R, RxR, VLN-CE	Özel fabrika verisi + AI Habitat	Eğitim ve öğretim senaryoları
<b>Değerlendirme Metrikleri</b>	Başarı oranı, güvenlik ölçütleri	SR, SPL	Piksel doğruluğu, navigasyon başarı oranı	Görev tamamlama oranı
<b>Öne Çıkan Sonuçlar</b>	Sosyal açıdan daha güvenli navigasyon	Gerçek dünyada %66 başarı; çapraz veri setinde %41.7 SR artışı	Simülasyonda %92.5 navigasyon başarısı	Eğitim senaryolarında güvenilir planlama
<b>Güçlü Yan</b>	İnsan merkezli navigasyon	Sensörsüz ve haritasız genelleme	Endüstriyel ortamlara uygunluk	Dil tabanlı esnek görev planlama
<b>Sınırlılıklar</b>	Hesaplama maliyeti	Gerçek zamanlılık sınırlı	Sistem karmaşıklığı	Fiziksel alrıdan ziyade planlama odaklı

### 3.7.2.3. Görsel Dil Modelleri ile Algı, Grounding, Affordance ve Düşük-Seviyeli Robotik Biliş

Robotik sistemlerde algıdan eyleme uzanan bilişsel süreçler, yalnızca nesne tanıma veya sahne sınıflandırma çıktılarıyla sınırlı değildir. Robotların çevreyle güvenli ve etkili biçimde etkileşime girebilmesi; mekânsal temellendirme (grounding), eylem uygunluğu (affordance), çevresel durum algısı ve düşük-seviyeli fiziksel özelliklerin doğru biçimde yorumlanmasını gerektirmektedir. Bu bağlamda görsel dil modelleri, görsel gözlemleri dilsel ve kavramsal temsillerle ilişkilendirerek, klasik denetimli algı yaklaşımlarının ötesinde daha esnek ve genellenebilir robotik biliş çerçeveleri sunmaktadır (Tablo 8).

Düşük-seviyeli robotik algıya odaklanan Osada ve arkadaşları (2024), yakınlık sensörleri açısından kritik olan kızılötesi yansıtma (reflectance) değerinin, yalnızca fiziksel ölçümle değil dağılımsal semantik bilgi üzerinden de tahmin edilebileceğini göstermiştir. Metin tabanlı tahmin için GPT-3.5 ve GPT-4, görüntü tabanlı tahmin için CLIP temelli bir VLM kullanılmıştır. Sonuçlar, GPT-4'ün yalnızca nesne adından %14.7 ortalama hata oranı ile görüntü tabanlı ResNet modellerini geride bıraktığını; CLIP tabanlı yaklaşımın ise %11.8 hata oranı ile en düşük hatayı elde ettiğini göstermektedir. Bu bulgular, dil modellerinde gömülü semantik bilginin düşük-seviyeli robotik görevlerde kullanılabilirliğini ortaya koymaktadır.

Algısal çıktının doğrudan eyleme dönüştürülmesini amaçlayan bir diğer yaklaşım olan RoboPoint (Zhu ve ark., 2024), robotik manipülasyon ve navigasyon görevlerinde gerekli eylem noktalarını, sınırlayıcı kutular yerine

doğrudan görüntü düzleminde  $(x, y)$  koordinatları olarak tahmin etmeyi hedeflemektedir. Model mimarisi, Vicuna-v1.5-13B tabanlı bir dil modeli ile CLIP ön-egitimli görsel kodlayıcı üzerine inşa edilmiş olup, komut ince ayarı (instruction fine-tuning) ve ölçeklenebilir sentetik veriyle eğitilmiştir. Sonuçlar, RoboPoint modelinin gerçek dünya robot görevlerindeki başarı oranını %30'un üzerinde artırdığını ve diğer VLM tabanlı yaklaşımlara kıyasla daha yüksek doğruluk sağladığını göstermektedir.

Çevresel durum farkındalığına yönelik Sharshar ve arkadaşları (2023), ön-egitimli VLM'leri kara kutu optimizasyon yöntemleriyle birleştirmiştir. Kapı durumu, musluk akışı veya ortam temizliği gibi durumlar VQA ve görüntü–metin eşleştirme görevleri üzerinden tanınmıştır. BLIP-2, OFA, CLIP ve ImageBind gibi modeller için oluşturulan çoklu metin istemleri, genetik algoritma tabanlı optimizasyonla ayarlanmıştır. Optimize edilmiş istemlerle tek bir VLM kullanılarak farklı çevresel durumların %90'ın üzerinde doğrulukla tanımladığı ve özellikle şeffaf yüzeyler gibi zorlayıcı senaryolarda performansın arttığı rapor edilmiştir.

Mekânsal temellendirmeyi üç boyuta taşıyan VLM-Grounder yaklaşımı (Xu et al., 2024), sıfır atış 3B görsel temellendirme problemine odaklanmaktadır. Sistem, 3B nokta bulutuna ihtiyaç duymadan yalnızca 2B görüntüler ve doğal dil sorguları üzerinden hedef nesnenin 3B sınırlayıcı kutusunu tahmin etmektedir. GPT-4o ajan olarak kullanılmış; Grounding DINO, SAM ve YOLOv8-World gibi 2B algı bileşenleri dinamik birleştirme, geri bildirimli temellendirme ve çok görünümlü projeksiyon mekanizmalarıyla entegre edilmiştir. ScanRefer ve Nr3D veri setlerinde önerilen yöntemin önceki zero-shot yaklaşımlara kıyasla anlamlı doğruluk artışı sağladığı ve bazı gözetimli yöntemlerle karşılaştırılabilir performans sunduğu bildirilmiştir.

Bu çalışmalar, görsel dil modellerinin robotikte yalnızca yüksek-seviyeli planlama değil; düşük-seviyeli fiziksel özellik tahmini, eylem noktası çıkarımı ve çok boyutlu mekânsal temellendirme gibi görevlerde de etkili olduğunu göstermektedir. Ortaya çıkan eğilim, robotik bilişin tekil algı modülleri yerine görsel, dilsel ve fiziksel bilgiyi ortak bir temsil uzayında bütünleştiren çok modlu sistemler üzerinden ele alınması yönündedir.

**Tablo 8. Görsel–Dil Modellerinin Robotik Algı, Mekânsal Temellendirme ve Eylem Uygunluğu Görevlerindeki Kullanımı**

Özellik	Yansıtma Tahmini ile Yakınlık Algılama	RoboPoint: Mekânsal Eylem Noktası Tahmini	Çevresel Durum Tanıma (VLM + Optimizasyon)	VLM-Grounder: Sıfır Atış 3B Temellendirme
Kaynak	Osada et al., 2024	Zhu et al., 2024	Sharshar et al., 2023	Xu et al., 2024
Temel Amaç	Nesnelerin IR yansıtma değerlerini tahmin ederek yakınlık algısını iyileştirmek	Robot eylemleri için doğrudan (x, y) eylem noktaları tahmin etmek	Çevresel durumları doğal dil üzerinden tanımak	2B görüntülerden 3B hedef nesne temellendirmek
Robotik Problem Türü	Düşük-seviyeli algı ve kavrama hassasiyeti	Mekânsal uygunluk ve hassas manipülasyon	Durum farkındalığı	3B mekânsal temellendirme
Kullanılan Modeller	GPT-3.5, GPT-4, CLIP; VGG16, ResNet101 (kıyas)	Vicuna-13B + CLIP	BLIP-2, OFA, CLIP, ImageBind	GPT-4o, Grounding DINO, SAM
Temel Teknikler	Few-shot prompting, CoT, CLIP embedding regresyonu	Instruction fine-tuning, sentetik veri üretimi	VQA, ITR, genetik algoritma ile istem optimizasyonu	Dinamik stitching, geri bildirim, çok görünümlü projeksiyon
Veri Setleri	54 nesne, 324 görüntü-yansıtma çifti	RoboRefIt, WHERE2PLACE	Özel çevresel durum veri setleri	ScanRefer, Nr3D
Öne Çıkan Sonuçlar	CLIP %11.8 hata, GPT4 %14.7 hata	%30+ başarı artışı	%90+ doğruluk	Acc@0.25'te %51.6

### 3.7.3. Görsel Dil Modellerinin Sağlık Alanındaki Kullanımı

#### 3.7.3.1. Radyoloji ve Klinik Görüntülemelerde Görsel Dil Modelleri

Radyoloji alanında Görsel Dil Modellerinin (VLM) kullanımı, klasik serbest metin rapor üretiminin ötesine geçerek; klinik bilginin yapılandırılması, halüsinasyonların azaltılması ve karar destek süreçlerine entegrasyonu hedefleyen bütüncül sistemlere evrilmektedir. Son dönemdeki çalışmaların temel amacı, radyoloji görüntüleri ve raporlarının birlikte ele alındığı çok modlu çerçeveler aracılığıyla metin tabanlı yaklaşımların bağlam sınırlamalarını ve klinik tutarsızlık sorunlarını aşmaktır (Zhang et al., 2024; Li et al., 2024; Zhong et al., 2025; Wang et al., 2023).

Bu doğrultuda öne çıkan VLM-KG, radyoloji görüntüleri ile serbest metin raporlarını birlikte kullanarak yapılandırılmış bilgi grafikleri üretmeyi hedeflemektedir (Zhang et al., 2024). Önerilen mimaride, uzun raporları işleyebilmek için 32K bağlam uzunluğunu destekleyen Qwen1.5-0.5B dil modeli ve görsel tarafta kontrastif biçimde eğitilmiş MedCLIP kodlayıcısı kullanılmıştır. Görsel ve dilsel temsiller arasındaki boyut uyumsuzluğu çok

katmanlı transformatör tabanlı bir projektörle giderilmiş ve RadGraph şemasına dayalı varlık–ilişki–varlık üçlüleri üzerinden denetimli öğrenme uygulanmıştır. Deneysel sonuçlar, görsel bağlamın entegrasyonu ile bilgi grafiği üretimindeki halüsinasyonların anlamlı biçimde azaldığını ve performans metriklerinde belirgin iyileşmeler sağlandığını göstermektedir.

Radyoloji rapor üretiminde güvenilirliği artırmayı amaçlayan SERPENT-VLM (Li et al., 2024), görüntüde bulunmayan bulguların rapora eklenmesi sorununu çözmek için kendi kendini iyileştiren (self-refining) bir eğitim stratejisi hedeflemektedir. Sistemin mimarisinde, dondurulmuş Swin-Transformer V2 görsel kodlayıcısı ile LLaMA2-7B dil modeli entegre edilmiş; standart dil modeli kaybına ek olarak, üretilen rapor ile görsel temsil arasındaki tutarlılığı zorlayan bir “self-refining loss” tanımlanmıştır. IU-Xray ve ROCO veri setleri üzerindeki sonuçlar, bu yaklaşımın rapor doğruluğunu artırırken halüsinasyonları sistematik olarak azalttığını kanıtlamaktadır.

Görsel dil modellerinin klinik iş akışına derin entegrasyonunu hedefleyen ve raporlamayı tanı/prognoz tahminiyle birleştirmeyi amaçlayan bir diğer sistem, CT Pulmonary Angiogram (CTPA) odaklı ajan tabanlı çerçevedir (Zhong et al., 2025). Bu kapsamlı mimaride, pulmoner emboli ile ilişkili 32 anormallığı tespit eden çok etiketli bir sınıflandırıcının ardından; CT-CHAT, RadFM ve M3D gibi modeller “okuma ajanı” olarak konumlandırılmıştır. Bölgesel bulgular LLaMA 3 tabanlı bir yazma ajanı tarafından yapılandırılmış raporlara dönüştürülmekte ve son aşamada görüntü, klinik değişkenler ve rapor çıktıları Cox Proportional Hazards modeli kullanılarak sağkalım tahmini yapılmaktadır. Çok merkezli veri setlerinden elde edilen sonuçlar, bu çok ajanlı yaklaşımın hem rapor kalitesini hem de klinik sonuç tahminini anlamlı ölçüde iyileştirdiğini ortaya koymaktadır.

Önceden eğitilmiş bir VLM mimarisinin klinik alana çok aşamalı ince ayar stratejileriyle uyarlanmasını amaçlayan ClinicalBLIP modelinde ise (Wang et al., 2023); InstructBLIP tabanında CLIP görsel kodlayıcısı, Flan-T5-XL dil modeli ve Q-Former yapısı bir araya getirilmiş, CheXBERT ile çıkarılan tıbbi etiketler modele istem (prompt) olarak entegre edilmiştir. IU-Xray ve MIMIC-CXR veri setlerindeki sonuçlar, klinik bağlama uyarlanan bu çok modlu modellerin rapor doğruluğu ve tutarlılığında önemli kazanımlar sağladığını göstermektedir. Genel bir değerlendirmeye; bu çalışmalar görsel dil modellerinin radyolojide salt rapor üretimi sınırlarını aşarak yapılandırılmış bilgi çıkarımı ve klinik karar destek mekanizmaları için vazgeçilmez hâle geldiğini doğrulamaktadır. Tablo 9’da radyoloji ve klinik görüntüleme VLM tabanlı yaklaşımların karşılaştırılması verilmiştir.

**Tablo 9. Radyoloji ve Klinik Görüntülemede VLM Tabanlı Yaklaşımların Karşılaştırılması**

Özellikler	VLM-KG	SERPENT-VLM	ClinicalBLIP	CTPA-VLM
<b>Kaynak</b>	Multimodal Radiology Knowledge Graph Generation	Self-Refining Radiology Report Generation Using VLM	Vision-Language Model for Generating Textual Descriptions From Clinical Images	VLM for Report Generation and Outcome Prediction in CT Pulmonary Angiogram
<b>Çalışma Alanı</b>	Radyoloji bilgi çıkarımı ve klinik bilgi yapılandırma	Güvenilir radyoloji rapor üretimi	Klinik görüntüden rapor üretimi	Toraks BT (CTPA) tabanlı tanı ve prognoz
<b>Temel Amaç</b>	Görüntü + rapor kullanarak bilgi grafiği üretmek ve halüsinasyonları azaltmak	Görüntüyle uyumsuz rapor üretimini azaltmak	Klinik bağlama uyumlu rapor üretimini güçlendirmek	Tanı, yapılandırılmış rapor ve klinik sonuç tahminini entegre etmek
<b>Temel Yaklaşım</b>	Görsel dil temsillerini bilgi grafiği triplet'lerine dönüştürme	Self-refining kayıp ile görsel-metin hizalaması	Çok aşamalı klinik ince ayar ve ön bilgi entegrasyonu	Anormallik-rehberli ajan tabanlı raporlama
<b>Kullanılan VLM / LLM</b>	Qwen1.5-0.5B	LLaMA2-7B	Flan-T5-XL	LLaMA-3
<b>Görsel Kodlayıcı</b>	MedCLIP	Swin Transformer V2 (frozen)	CLIP	3D I3D
<b>Dil Modeli</b>	Qwen1.5-0.5B	LLaMA2-7B	Flan-T5-XL	LLaMA-3
<b>Mimari Özellik</b>	Transformer tabanlı projektör + visual instruction tuning	LoRA destekli eşleme + Self-Refining Loss	Q-Former + InstructBLIP mimarisi	Reading Agent + Writing Agent + CoxPH
<b>Girdi Türü</b>	Radyoloji görüntüsü + serbest metin rapor	Radyoloji görüntüsü	Klinik görüntü	CTPA hacimleri + klinik değişkenler
<b>Çıktı Türü</b>	Bilgi grafiği triplet'leri	Serbest metin radyoloji raporu	Klinik rapor / açıklama	Tanı, yapılandırılmış rapor, prognoz
<b>Veri Setleri</b>	MIMIC-CXR, IU-Xray	IU-Xray, ROCO	IU-Xray, MIMIC-CXR	BUH, JHU, INSPECT
<b>Değerlendirme Metrikleri</b>	BLEU-1/2/3/4, ROUGE-L	BLEU-4, ROUGE-L, BERTScore	BLEU-A, METEOR, ROUGE-L	AUROC, F1, BERT-F1, BLEU-4
<b>Öne Çıkan Sonuçlar</b>	Daha düşük halüsinasyon ve daha yüksek yapısal doğruluk	Halüsinasyonların belirgin biçimde azalması	Klinik tutarlılığı yüksek raporlar	Tanı ve prognoz başarısında anlamlı artış
<b>Güçlü Yan</b>	Görsel bağlamla güvenilir bilgi çıkarımı	Görsel-metin uyumunun doğrudan optimize edilmesi	Klinik bağlama güçlü adaptasyon	Uçtan uca klinik iş akışına entegrasyon
<b>Sınırlılıklar</b>	Bilgi grafiği şemasına bağımlılık	Hesaplama maliyeti ve eğitim karmaşıklığı	Gerçek zamanlı kullanım için ağır mimari	Yüksek hesaplama ve veri gereksinimi

### 3.7.3.2. Uzmanlık Alanına Özgü Klinik Görsel Dil Model Uygulamaları: VQA, Tam Destekli Yorumlama ve Görev Odaklı Modeller

Uzmanlık alanına özgü klinik görsel dil modeli uygulamaları, genel amaçlı rapor üretiminden farklı olarak belirli klinik görevlerin doğrudan desteklenmesine odaklanmaktadır. Bu yaklaşımlar, özellikle uzman eksikliğinin belirgin olduğu alanlarda tanı destek süreçlerinin güvenilirliğini ve erişilebilirliğini artırmayı amaçlamaktadır.

Dermatoloji alanında çok dilli Görsel Soru-Cevaplama (VQA) görevine odaklanan ve IKIM ekibi tarafından MEDIQA-M3G 2024 yarışmasında sunulan çalışma, bu vizyonun başarılı bir örneğidir (Liu et al., 2024). Çalışmanın amacı, dermatolog eksikliği olan senaryolarda VLM tabanlı sistemlerin tanı destek potansiyelini değerlendirmek ve özellikle Çince veriler üzerinde en ideal modelleme stratejisini belirlemektir. Önerilen sistem mimarisinde, biyomedikal alan için özelleştirilmiş LLaVA-med modeli temel alınmış ve yalnızca Çince verilerle minimal düzeyde bir ince ayar (fine-tuning) sürecinden geçirilmiştir. Modelden alınan Çince çıktılar ise Mixtral-8×7B-instruct büyük dil modeli kullanılarak İngilizce ve İspanyolcaya aktarılmıştır. Eğitim mimarisine dair en dikkat çekici strateji; modelin yalnızca tek bir epoch (epok) boyunca (öğrenme oranı =  $2e^{-5}$ , yığın boyutu = 4, gradyan birikimi = 16) eğitilmesi ve Fitzpatrick17k ile DermNet gibi harici veri setlerinin aşırı uyum (overfitting) riski nedeniyle eğitim dışı bırakılmasıdır. Deneysel sonuçlar, her dil için ayrı bir model eğitmek yerine, tek dilde optimize edilmiş bir modelin çıktılarını LLM tabanlı çeviri ile çoğaltmanın çok daha başarılı olduğunu kanıtlamıştır. Bu stratejinin bir sonucu olarak IKIM ekibi, yarışmanın Çince ve İspanyolca kategorilerinde birincilik, İngilizce kategorisinde ise üçüncülük elde etmiştir (Liu et al., 2024).

Benzer şekilde görev odaklı bir VLM yaklaşımıyla önkol ultrason görüntülerinden el hareketi çözümüleme problemini ele alan bir diğer çalışma, klasik yoğun etiketli sınıflandırma yöntemlerini Vision-Language çerçevesiyle değiştirmeyi amaçlamaktadır (Bimbraw et al., 2024). Kurgulanan mimaride, görsel dil modellerinin bağlam içi öğrenme (in-context learning - ICL) yetenekleri ve dikkatle tasarlanmış istem (prompt) yapıları kullanılarak; tam bir ince ayara gerek kalmadan, yalnızca kullanıcıya özgü birkaç örnekle (few-shot) sistemin uyarlanabilmesi hedeflenmiştir. Özellikle klinik açıdan gerçeğe en yakın olan oturumlar arası (cross-session) ve kullanıcılar arası (cross-subject) senaryolarda gerçekleştirilen deneylerin sonuçları, few-shot örnek sayısı arttıkça modelin genelleme performansının belirgin biçimde yükseldiğini göstermektedir. Bu bulgular, ultrason gibi operatöre ve bireysel anatomiye son derece duyarlı olan, veri kıtlığının yaşandığı durumlarda dahi

VLM tabanlı mimarilerin rekabetçi doğruluk oranlarına ulaşabildiğini ortaya koymaktadır (Bimbraw et al., 2024).

Genel bir çerçeveden değerlendirildiğinde, uzmanlık alanına özgü klinik VLM uygulamalarının başarısının karmaşık çoklu model mimarilerinden ziyade; doğru görev tanımlamasına, hedef odaklı ve sınırlı ince ayar stratejilerine, ayrıca model yetenekleriyle örtüşen doğru veri kullanımına bağlı olduğu sonucuna varılmaktadır (Bimbraw et al., 2024; Liu et al., 2024). Uzmanlık alanına özgü klinik VLM uygulamalarının karşılaştırılması Tablo 10’ da verilmiştir.

*Tablo 10. Uzmanlık Alanına Özgü Klinik Görsel Dil Model(VLM) Uygulamalarının Karşılaştırılması*

Özellik	Multilingual Visual Question Answering for Dermatology	Hand Gesture Decoding from Forearm Ultrasound Images
Kaynak	Liu et al. (2024)	Bimbraw et al. (2024)
Çalışma Alanı	Dermatoloji	Rehabilitasyon, protez ve insan-makine etkileşimi
Temel Amaç	Dermatoloji görüntülerinden klinik sorulara çok dilli ve doğru yanıt üretmek	Önkol ultrason görüntülerinden el hareketlerini veri ktlığı koşullarında çözümlmek
Temel Yaklaşım	VQA tabanlı klinik tanı destek sistemi + LLM tabanlı çeviri	Vision-Language çerçevesinde few-shot jest sınıflandırma
Kullanılan VLM / LLM	LLaVA-med + Mixtral-8×7B-instruct	GPT-Sonography (VLM tabanlı yaklaşım)
Görsel Kodlayıcı	LLaVA-med görsel omurgası	Ultrason görüntülerine uyarlanmış VLM görsel kodlayıcısı
Dil Modeli	LLaVA-med (temel) + Mixtral (çeviri)	VLM’nin entegre dil bileşeni
Mimari Özellik	Tek dilde fine-tuning + çok dilli çıktı üretimi	Fine-tuning olmadan in-context learning (ICL)
Girdi Türü	Dermatoloji görüntüsü + klinik soru	Ultrason görüntüsü + görev tanımlı prompt
Çıktı Türü	Klinik VQA yanıtı (çok dilli)	El jesti sınıf etiketi
Veri Setleri	MEDIQA-M3G (Çince, İngilizce, İspanyolca); Fitzpatrick17k, DermNet (deneysel)	Özel toplanmış ultrason jest veri seti
Değerlendirme Senaryosu	MEDIQA-M3G yarışması	Cross-session ve cross-subject testler
Değerlendirme Metrikleri	Yarışma sıralaması, doğruluk	Sınıflandırma doğruluğu
Öne Çıkan Sonuçlar	Çince & İspanyolca: 1.’lik; İngilizce: 3.’lük	Few-shot örnek sayısı arttıkça doğruluk belirgin artıyor
Güçlü Yan	Çok dilli klinik tanı desteği, düşük eğitim maliyeti	Hızlı kişiselleştirme, veri ktlığına dayanıklılık
Sınırlılıklar	Harici veri ve çoklu görüntü girdisi performansı düşürüyor	Jest çeşitliliği arttıkça performans sınırlanabiliyor

### 3.7.3.3. Temel ve Geniş Ölçekli Medikal VLM Yaklaşımları: Foundation Modeller, Etiket Kıtılığı ve Büyük Veri ile Öğrenme

Medikal Görsel Dil Modelleri (VLM) literatüründe son yılların temel eğilimi, etiketleme maliyetlerini minimize etmeyi ve sağlık sistemi ölçeğinde genellenebilirliği hedefleyen “foundation” (temel) model yaklaşımlarıdır. Patoloji, çok modlu klinik görevler ve nörogörüntüleme alanlarına odaklanan bu çalışmalar; etiket kıtlığı, görevler arası çatışma ve büyük ölçekli klinik verilerin entegrasyonu gibi temel sorunlara yenilikçi teknik çözümler sunmayı amaçlamaktadır.

Etiketleme sürecini tamamen ortadan kaldırmayı hedefleyen VLM-CPL yöntemi, patoloji görüntülerinin sınıflandırılması problemine odaklanmaktadır (Zhang et al., 2023). Önerilen mimaride, önceden eğitilmiş VLM’lerin sıfır-atış (zero-shot) çıkarım yetenekleri kullanılarak otomatik sözde etiketler (pseudo-labels) üretilir. Eğitim ve hedef veri dağılımları arasındaki uyumsuzluktan (domain gap) kaynaklanan etiket gürültüsünü filtrelemek amacıyla, PLIP ve BioMedCLIP modelleriyle üretilen etiketler çok aşamalı bir yarı denetimli öğrenme çerçevesinden geçirilir ve nihai sınıflandırma ResNet50 ile UNI kodlayıcıları üzerinden yapılır. Bu mimarinin merkezinde yer alan Çoklu Görünüm Uzlaşması (Multi-View Consensus - MVC) mekanizması, farklı veri artırma görünümüleri arasındaki tahmin tutarlılığını ölçerek güvenilir örnekleri seçer. Sınıf dengesizliğini hafifletmek için Sınıfa Duyarlı MVC (CMVC), etiket güvenilirliğini artırmak için ise Prompt-Feature Consensus (PFC) ve High-confidence Cross Supervision (HCS) modülleri sisteme entegre edilmiştir. Tüm slayt görüntüleri (WSI) düzeyinde ise ilgisiz yamaları eleyen Open-Set Prompting (OSP) mekanizması kullanılmıştır. Elde edilen deneysel sonuçlar, mekânsal ve renk dönüşümlerinin birlikte kullanımının MVC başarısını katladığını (örneğin HPH veri setinde doğruluğun 0.645’ten 0.904’e çıktığını) ve VLM-CPL yaklaşımının doğrudan sıfır-atış kullanımına kıyasla ortalama %18.8 doğruluk artışı sağladığını kanıtlamaktadır (Zhang et al., 2023).

Tıbbi görsel anlama (comprehension) ve görsel üretim (generation) görevlerini tek bir model çatısında birleştirmeyi amaçlayan bir diğer foundation yaklaşım HealthGPT’dir (Li et al., 2024). Bu çalışma, anlama görevlerinin gerektirdiği soyut/semantik temsiller ile üretim görevlerinin ihtiyaç duyduğu ayrıntılı/yüksek frekanslı görsel bilgi arasındaki görev çatışmasını çözmeyi hedeflemektedir. Geliştirilen Heterojen Düşük Dereceli Adaptasyon (Heterogeneous Low-Rank Adaptation - H-LoRA) mimarisi, görevlere özgü bilgileri izole edilmiş LoRA eklentilerinde saklayarak dinamik yönlendirme ile görevler arası negatif etkileşimi sınırlar. Sistem mimarisinde

görsel kodlayıcı olarak CLIP-L/14 kullanılmış (derin katmanlar anlama, sıç katmanlar üretim için); dil modeli tarafında Phi-3-mini ve Phi-4 tercih edilmiş ve VQGAN tabanlı ayırık jetonlaştırma uygulanmıştır. Üç Aşamalı Eğitim Stratejisi (TLS) ile optimize edilen modelin deneysel sonuçları, VL-Health veri setinde OmniMedVQA testi için 68.5 puan ve CT-MRI dönüşüm görevinde 79.38 SSIM skoru elde edildiğini göstermektedir. Ayrıca H-LoRA yapısının, MoELoRA yaklaşımına kıyasla eğitim süresini yaklaşık %33 oranında azalttığı rapor edilmiştir (Li et al., 2024).

Sağlık sistemi ölçeğinde nörogörüntüleme alanına odaklanan Prima modeli ise, gerçek dünya klinik MRI verilerini tüm sekanslarıyla işleyerek genellenebilir bir yapay zeka temeli oluşturmayı amaçlamaktadır (Peng et al., 2024). Prima mimarisi, MRI verisinin hiyerarşik doğasına tam uygun olarak tasarlanmış; hacim düzeyinde sıkıştırma için 3D VQ-VAE tabanlı bir volume tokenizer, ardından sekans ve çalışma düzeyinde temsil üretimi için iki aşamalı bir Vision Transformer (ViT) yapısı kurgulanmıştır. Eğitim sürecinde, MRI görüntüleri ile GPT-3.5 tarafından özetlenen radyoloji raporları (önyargıyı azaltmak amacıyla) CLIP hedefi üzerinden hizalanmıştır. UM-220K veri setiyle (220.000'den fazla MRI çalışması) eğitilen modelin prospektif sonuçları, 52 farklı radyolojik tanıda ortalama %92.0 AUROC değerine ulaştığını ve sıfır atış senaryolarında dahi CLIP, LLaVA, BioMedCLIP ve Med-Flamingo gibi modelleri geride bıraktığını ortaya koymaktadır. Klinik uygulama açısından modelin, radyolog iş listesi önceliklendirmesinde gerçek şiddet skorlarıyla yüksek korelasyon (Pearson  $r = 0.69$ ) ve sevk önerisi görevlerinde %85 üzeri AUROC sağladığı kanıtlanmıştır (Peng et al., 2024). Tablo 11' de temel ve geniş ölçekli medikal VLM yaklaşımlarının karşılaştırılması sunulmuştur.

*Tablo 11. Temel ve Geniş Ölçekli Medikal VLM Yaklaşımlarının Karşılaştırılması*

Özellikler	VLM-CPL	HealthGPT	Prima
Kaynak	Zhang et al. (2023)	Li et al. (2024)	Peng et al. (2024)
Çalışma alanı	Dijital patoloji	Çok modlu klinik görsel zekâ	Nörogörüntüleme (MRI)
Temel amaç	İnsan etiketlemesini ortadan kaldırarak etiketsiz patoloji verilerinden yüksek doğruluklu sınıflandırma yapmak	Tıbbi görsel anlama ve üretim görevlerini tek bir modelde görev çatışması olmadan birleştirmek	Sağlık sistemi ölçeğinde MRI verileriyle genellenebilir bir foundation model geliştirmek
Temel yaklaşım	Zero-shot VLM tabanlı pseudo-label üretimi + yarı denetimli öğrenme	Çok görevli birleşik öğrenme ve görev-özgü adaptasyon	Büyük ölçekli contrastive vision-language öğrenme

<b>Kullanılan VLM / LLM</b>	PLIP, BioMedCLIP (pseudo-label üretimi)	HealthGPT; LLM: Phi-3-mini (3.8B), Phi-4 (14B)	Prima (CLIP hedefli vision–language model)
<b>Görsel kodlayıcı</b>	PLIP / BioMedCLIP image encoder'ları	CLIP-L/14 (katman bazlı ayrıştırma)	Hiyerarşik Vision Transformer (Volume → Sequence → Study)
<b>Dil modeli</b>	prompt tabanlı VLM çıkarımı	Phi-3-mini, Phi-4	— (rapor–görüntü hizalaması için CLIP hedefi)
<b>Mimari özellikler</b>	MVC, CMVC, PFC, HCS, Open-Set Prompting (OSP)	H-LoRA, Hierarchical Visual Perception (HVP), Three-Stage Learning Strategy	3D VQ-VAE volume tokenizer, iki aşamalı ViT
<b>Girdi türü</b>	Patoloji yama görüntüleri (patch), WSI	Klinik görüntüler + doğal dil	Çok sekanslı MRI hacimleri + radyoloji raporları
<b>Çıktı türü</b>	Sınıf etiketleri (patoloji)	Metinsel cevaplar, üretilmiş tıbbi görüntüler	Tanı skorları, önceliklendirme puanları
<b>Veri setleri</b>	Patch: HPH, LC25K, NCT-CRC-HE-100K; WSI: DigestPath, TCGA-RCC	VL-Health: PubMedVision, LLaVA-Med, PathVQA, MIMIC-CXR-VQA, IXI, SynthRAD2023	UM-220K: 220K + MRI çalışma, 5.6M sekans, 362M kesit
<b>Değerlendirme metrikleri</b>	Accuracy	VQA score, SSIM	AUROC, Pearson korelasyonu
<b>Öne çıkan sonuçlar</b>	Ortalama %18.8 doğruluk artışı; HPH'de 0.645 → 0.883; MVC + ColorJitter ile 0.904	OmniMedVQA'da 68.5; CT→MRI dönüşümünde 79.38 SSIM; MoELoRA'ya kıyasla %33 daha hızlı eğitim	52 tamda %92.0 AUROC; sevk önerilerinde %85 + AUROC; iş listesi önceliklendirmede $r = 0.69$
<b>Güçlü yan</b>	Gürültüye dayanıklı pseudo-label temizleme ve etiketsiz öğrenme	Görev çatışmasını mimari düzeyde azaltan birleşik yapı	Gerçek dünya klinik verilerinde yüksek genellenebilirlik
<b>Sınırlılıklar</b>	Domain shift'e duyarlılık ve patch–WSI geçişinde karmaşıklık	Yüksek hesaplama maliyeti ve karmaşık eğitim süreci	Büyük bellek gereksinimi ve MRI odaklı sınırlı modalite

### 3.7.4. Görsel -Dil Modellerinin Tarım Alanındaki Kullanımı

#### 3.7.4.1. Uzaktan Algılama ve Mekânsal-Zamansal Tahmin ile Tarımsal Planlama

Tarımsal üretimin planlanması ve mahsul veriminin güvenilir biçimde tahmin edilmesi, yüksek boyutlu uzaktan algılama verilerinin çevresel ve iklimsel bağlarla bütünleştirilmesini gerektirmektedir. Bu doğrultuda Görsel Dil Modelleri (VLM), uydu görüntülerini meteorolojik zaman serileri ve kavramsal temsillerle harmanlayarak tarımsal karar destek süreçlerini optimize etmeyi amaçlamaktadır.

Bu bağlamda geliştirilen MMST-ViT modeli, iklim değişikliği etkilerini dikkate alan çok modlu ve mekânsal-zamansal bir mahsul verim tahmin çerçevesi sunmayı hedeflemektedir (Zhou et al., 2023). Sistemin mimarisi üç temel bileşenden oluşmaktadır. İlk bileşen olan çok modlu transformatör; Sentinel-2 uydu görüntülerindeki görsel girdileri (Pyramid Vision Transformer - PVT omurgasıyla) ve kısa vadeli meteorolojik değişkenleri, Çok Modlu Çoklu Başlık Dikkati (MM-MHA) mekanizması aracılığıyla ortak bir temsilde birleştirir. İkinci bileşen olan uzaysal transformatör, ilçe bazlı tarım alanları arasındaki mekânsal bağımlılıkları modellerken; üçüncü bileşen olan zamansal transformatör, uzun vadeli tarihsel hava verileri üzerinden iklim değişikliğinin mahsul üzerindeki kümülatif etkilerini yakalar (Zhou et al., 2023). Modelin eğitiminde ABD'deki 200'den fazla ilçeyi kapsayan 2017–2022 dönemine ait uydu, HRRR ve USDA verileri kullanılmıştır.

DeneySEL sonuçlar, MMST-ViT modelinin soya fasulyesi verim tahmininde  $RMSE = 3.9$ ,  $R^2 = 0.843$  ve Pearson korelasyonu = 0.918 değerlerine ulaşarak ConvLSTM, CNN-RNN ve GNN-RNN gibi önceki uzaysal-zamansal modelleri belirgin şekilde geride bıraktığını göstermektedir. Pamuk mahsulünde (birim farklılıklarından kaynaklanan yüksek varyans nedeniyle) daha yüksek RMSE değerleri gözlemlense de, yüksek  $R^2$  ve korelasyon oranları modelin güçlü açıklayıcılığını koruduğunu kanıtlamaktadır (Zhou et al., 2023).

Uzaktan algılama verilerinin tarıma özgü uyarlanması amaçlayan bir diğer çalışma olan AgriCLIP, genel amaçlı CLIP yapılarının tarımsal ince ayrıntıları yakalamadaki yetersizliğini aşmayı hedeflemektedir (Wang et al., 2024). AgriCLIP mimarisinde, kontrastif öğrenme yoluyla elde edilen küresel anlamsal temsiller ile DINO tabanlı kendi kendine denetimli görsel kodlayıcıdan alınan yerel ayrıntıya duyarlı özellikler, öğrenilebilir bir dönüşüm üzerinden (Ortalama Kare Hatası - MSE minimize edilerek) birbirine hizalanır. Modelin eğitimi, bitkisel üretim ve hayvancılık alanlarını kapsayan ve GPT-4

destekli metinlerle oluşturulmuş yaklaşık 600.000 görüntü-metin çiftinden oluşan ALive veri seti üzerinde gerçekleştirilmiştir.

Modelin daha önce hiç görülmemiş veriler üzerindeki sıfır-atış (zero-shot) sınıflandırma sonuçları, standart CLIP modeline kıyasla mutlak %9.07 oranında bir doğruluk artışı sağlandığını ve bu kazanımın özellikle ince taneli ayrımlar gerektiren görevlerde çok daha belirginleştiğini ortaya koymaktadır (Wang et al., 2024).

Uzaktan algılama alanındaki VLM literatürünü inceleyen kapsamlı derlemeler (Xie et al., 2024), bu modellerin güçlü görsel temsiller sunduğunu, ancak karmaşık akıl yürütme ve alana özgü bilgi entegrasyonu konularında hâlen geliştirilmeye açık olduğunu vurgulamaktadır. Genel bir değerlendirmeye; veri heterojenliği ve iklimsel belirsizlikler gibi zorluklara rağmen, VLM tabanlı mimariler makro ölçekli tarımsal tahmin ve planlama problemlerinde kritik bir potansiyel sunmaktadır. Uzaktan algılama ve mekânsal-zamansal tahmin ile tarımsal planlama alanındaki çalışmaların karşılaştırılması Tablo 12’ de verilmiştir.

*Tablo 12. Uzaktan Algılama ve Mekânsal-Zamansal Tahmin ile Tarımsal Planlama*

Özellikler	AgriCLIP (Wang et al., 2024)	MMST-ViT (Zhou et al., 2023)	Vision-Language Modeling Meets Remote Sensing (Xie et al., 2024)
<b>Kaynak / Makale</b>	<i>Adapting CLIP for Agriculture and Livestock via Domain-Specialized Cross-Model Alignment</i>	<i>Climate Change-Aware Crop Yield Prediction via Multi-Modal Spatial-Temporal Vision Transformer</i>	<i>Vision-Language Modeling Meets Remote Sensing: Models, Datasets and Perspectives</i>
<b>Çalışma Alanı</b>	Tarım ve hayvancılık	Uzaktan algılama tabanlı mahsul verim tahmini	Uzaktan algılama (genel inceleme)
<b>Temel Amaç</b>	Tarım ve hayvancılık alanına özgü ince taneli görsel ayrımları zero-shot olarak iyileştirmek	İklim değişikliği etkilerini dikkate alarak mahsul verimini mekânsal-zamansal bağlamda tahmin etmek	Uzaktan algılama alanında VLM yaklaşımlarını sistematik biçimde sınıflandırmak
<b>Temel Yaklaşım</b>	Alan-özü CLIP adaptasyonu ve cross-model alignment	Çok modlu mekânsal-zamansal Vision Transformer mimarisi	Kontrastif, instruction-based ve generation-based VLM yaklaşımlarının karşılaştırmalı analizi
<b>Kullanılan VLM / LLM</b>	CLIP / OpenCLIP (kontrastif); DINO (self-supervised)	VLM tabanlı çok modlu Transformer	CLIP türevleri, instruction-tuned görsel dil modeller

<b>Görsel Kodlayıcı</b>	CLIP image encoder + DINO	PVT-Tiny (PVT-T/4)	Çeşitli (CLIP, RemoteCLIP vb.)
<b>Dil Modeli</b>	CLIP text encoder (GPT-4 yalnızca prompt üretimi için)	—	LLM tabanlı instruction modeller (inceleme kapsamında)
<b>Mimari Özellikler</b>	Üç aşamalı öğrenme: kontrastif öğrenme, self-supervised öğrenme, cross-model alignment (MSE)	Multi-Modal Transformer (MM-MHA), Spatial Transformer (S-MHA), Temporal Transformer (T-MHA)	Literatür temelli mimari sınıflandırma
<b>Girdi Türü</b>	Tarım ve hayvancılığa ait görüntüler + metin istemleri	Sentinel-2 uydu görüntüleri + kısa ve uzun vadeli HRRR meteorolojik veriler	Çeşitli RS veri setleri (inceleme)
<b>Çıktı Türü</b>	Zero-shot sınıf tahminleri	Sürekli mahsul verim tahminleri	Performans karşılaştırmaları ve metodolojik çıkarımlar
<b>Veri Setleri</b>	ALive: ~600K image-text çifti (25 veri seti)	MS, LA, IA, IL eyaletleri; corn, cotton, soybean, winter wheat	AID, RSICD ve diğer RS veri setleri
<b>Değerlendirme Metrikleri</b>	Top-1 Accuracy (zero-shot)	RMSE, R <sup>2</sup> , Pearson korelasyonu	Top-1 Accuracy, Recall@K
<b>Öne Çıkan Sonuçlar</b>	20 veri setinde %48.27 zero-shot doğruluk; adapted CLIP'e göre +%9.07 mutlak artış	Soybean için RMSE = 3.9, R <sup>2</sup> = 0.843, Corr = 0.918	RemoteCLIP'in AID'de %87.9 doğruluk; PIR-CLIP'in RSICD'de SOTA
<b>Güçlü Yan</b>	İnce taneli tarımsal ayrımlarda güçlü zero-shot genelleme	İklimsel ve mekânsal bağımlılıkları birlikte modelleme	Alanın bütüncül ve sistematik çerçevesini sunma
<b>Sınırlılıklar</b>	Yüksek veri hazırlama maliyeti; alan dışı genellenebilirlik sınırlı	Bölgesel veri bağımlılığı ve hesaplama maliyeti	DeneySEL katkı yerine metodolojik özet sunması

### 3.7.4.2. Saha ve Bitki Düzeyi Tanılama ve Karar Destek

Tarımsal üretimde hastalık ve zararlı tespiti ile bitki gelişim evrelerinin belirlenmesi, salt görsel benzerliklere dayalı sınıflandırma problemlerinin ötesindedir. Bu görevler, ince taneli görsel ipuçlarının bitki türü, gelişim aşaması ve çevresel bağlamla birlikte yorumlanmasını gerektirir. Bu nedenle

Görsel Dil Modelleri (VLM), tarımsal tanı problemlerini bağlama duyarlı ve açıklanabilir karar destek sistemleri çerçevesinde ele almayı hedeflemektedir.

Bu doğrultuda geliştirilen AgriVLM, tarımsal görüntü ve metin verilerini çapraz modlu (cross-modal) biçimde işleyerek hastalık tanıma, gelişim evresi belirleme ve tarımsal soru-cevap görevlerini tek bir model çatısı altında birleştirmeyi amaçlamaktadır (Zhang et al., 2023). Modelin temel mimarisi; görsel özellik çıkarımı için Vision Transformer (ViT) tabanlı bir görüntü kodlayıcı, bu görsel temsilleri dil modelinin anlamsal uzayına aktaran bir Q-Former bileşeni ve temel dil modeli olarak ChatGLM'den oluşmaktadır. Tarımsal alana uyarılma sürecinde sınırlı veriyle aşırı uyum (overfitting) riskini azaltmak amacıyla, LoRA tabanlı parametre-verimli ince ayar (parameter-efficient fine-tuning) stratejisi uygulanmıştır. Elde edilen deneysel sonuçlar, beş farklı veri seti üzerinde değerlendirilen modelin (doğruluk, kesinlik, duyarlılık ve F1 skoru metriklerine göre) tüm görevlerde %90'ın üzerinde doğruluk oranına ulaştığını ve ince taneli ayırım gerektiren görevlerde genel amaçlı modellere kıyasla çok daha tutarlı bir performans sunduğunu göstermektedir (Zhang et al., 2023).

Tarımsal hastalık ve zararlı tanısına odaklanan bir diğer güncel çalışma olan LLMI-CDP ise, görsel tanı çıktıları ile alana özgü önleme önerilerini tek bir çerçevede birleştirmeyi hedeflemektedir (Liu et al., 2024). Önerilen sistem mimarisi, VisualGLM temeli üzerine inşa edilmiştir. Bu yapıda, modelin genel görsel-dil hizalama yeteneğini korumak ve tarımsal bilgi aktarımını sağlamak amacıyla görsel ve dil bileşenleri dondurulmuş, yalnızca LoRA katmanları eğitilmiştir. Toplam 15 mahsul türü ve 141 hastalık/zararlı kategorisini kapsayan 2.498 görüntülük veri seti üzerinde yapılan değerlendirme sonuçları, LLMI-CDP'nin hastalık tanımda %78.8 ve zararlı tanımda %86.7 doğruluk oranlarına ulaşarak temel VisualGLM modelini belirgin şekilde geride bıraktığını kanıtlamaktadır (Liu et al., 2024).

Karşılaştırma ölçütü olarak kullanılan genel amaçlı VisualGLM (Du et al., 2023), büyük ölçekli ön-eğitime sahip olmasına rağmen tarıma özgü ince ayar içermediği için bu tür spesifik görevlerde sınırlı bir performans sergilemektedir. Bu durum, tarımsal tanı problemlerinde alan-özel veri entegrasyonunun ve doğru uyarılma mimarilerinin kritik önemini açıkça ortaya koymaktadır. Tablo 13'te tarımsal saha ve bitki düzeyi tanılama için VLM tabanlı yaklaşımların karşılaştırılması sunulmuştur.

Tablo13. Tarımsal Saha ve Bitki Düzeyi Tanılama İçin VLM Tabanlı Yaklaşımların Karşılaştırılması

Özellikler	AgriVLM	LLMI-CDP	VisualGLM (Baseline)
Kaynak	Zhang et al. (2023)	Liu et al. (2024)	Du et al. (2023)
Çalışma Alanı	Tarımsal saha ve bitki düzeyi tanılama	Tarımsal hastalık ve zararlı tanısı	Genel amaçlı multimodal diyalog
Temel Amaç	Tarımsal görüntü ve metinlerden çoklu tanı ve karar destek üretmek	Hastalık/zararlı tanısı ve önleme önerilerini birleştirmek	Genel görsel dil etkileşimi sağlamak
Temel Yaklaşım	Çapraz modlu görsel dil temsil öğrenimi	Alan-özümlü LoRA uyarlaması	Genel amaçlı VLM ön eğitimi
Kullanılan VLM / LLM	ChatGLM tabanlı AgriVLM	VisualGLM + ChatGLM	ChatGLM-6B
Görsel Kodlayıcı	ViT tabanlı encoder	VisualGLM image encoder	VisualGLM image encoder
Dil Modeli	ChatGLM	ChatGLM	ChatGLM-6B
Mimari Özellik	Q-Former, LoRA ince ayar	Q-Former, dondurulmuş backbone + LoRA	Standart multimodal hizalama
Girdi Türü	Bitki ve hayvan görüntüleri + metin	Tarımsal görüntüler + metin	Genel görüntü + metin
Çıktı Türü	Tanı etiketleri, büyüme evresi, VQA yanıtları	Tanı etiketleri ve önleme önerileri	Görsel açıklama ve diyalog
Veri Setleri	5 görev veri seti + 5.593 tarımsal S-C örneği	2.498 görüntü, 141 sınıf	30M Çince + 300M İngilizce image-text
Değerlendirme Metrikleri	Accuracy, Precision, Recall, F1	Accuracy, insan & GPT-4 değerlendirmesi	Accuracy, VQA puanı
Öne Çıkan Sonuçlar	Tüm görevlerde >%90 doğruluk	Hastalık %78.8, zararlı %86.7	Hastalık %67.4, zararlı %72.4
Güçlü Yan	İnce taneli tarımsal görevlerde yüksek doğruluk	Tanı + öneri üretimini tek sistemde birleştirmesi	Güçlü genel dil yeteneği
Sınırlılıklar	Veri setleri görece küçük	Veri dili ağırlıklı olarak Çince	Tarımsal uzmanlık sınırlı

### 3.7.4.3. Tarım Odaklı Görsel Dil Model Tasarımı, Alan Uyarlaması ve Değerlendirme

Tarım alanında Görsel Dil Modellerinin (VLM) etkin biçimde kullanılabilmesi, genel amaçlı büyük modellerin ötesine geçerek; bitki hastalıkları, zararlı türleri ve yabancı otlar gibi birbirine son derece benzeyen sınıflar için ince taneli görsel ayrımların ve uzmanlık gerektiren kavramsal etiketlerin modele kazandırılmasını gerektirmektedir. Sadece görsel benzerliğe dayalı genellemenin yetersiz kaldığı bu alanda literatür, tarımsal görevler için güvenilir değerlendirme ortamlarının oluşturulması ve tarıma özgü bilgiyle eğitilmiş alan uyarlamalı modellerin geliştirilmesi olmak üzere iki tamamlayıcı eksen etrafında şekillenmeyi hedeflemektedir (Shinoda et al., 2024; Zhou et al., 2024).

Bu doğrultuda birinci eksen temsil eden AgroBench, tarım odaklı görsel dil modellerinin performansını gerçek dünya senaryoları üzerinden sistematik biçimde ölçmeyi amaçlayan kapsamlı bir değerlendirme çalışmasıdır (Shinoda et al., 2024). Sentetik ve sınırlı veri setlerinin ötesine geçmeyi hedefleyen bu kıyaslama mimarisi; 203 mahsul türü, 682 hastalık, 134 zararlı, 108 yabancı ot ve 98 tarımsal makine kategorisini kapsayan 4.218 görüntü ile 4.342 soru-cevap çiftinden oluşturulmuştur. Yedi temel görev altında gruplanan AgroBench üzerinde yapılan değerlendirme sonuçları, kapalı kaynaklı modellerde GPT-4o'nun, açık kaynaklılarda ise QwenVLM-72B'nin en yüksek performansı sergilediğini göstermektedir. Ancak hata analizleri, model başarısızlıklarının büyük ölçüde alan bilgisi eksikliğinden (%51.92) ve algısal hatalardan (%32.69) kaynaklandığını; özellikle görsel benzerliklerin çok yüksek olduğu yabancı ot tanımlama görevinin modeller için en zorlayıcı test olduğunu kanıtlamıştır (Shinoda et al., 2024).

AgroBench'in ortaya koyduğu bu alan bilgisi eksikliğini doğrudan çözmeyi amaçlayan AgroGPT ise, tarımsal uzmanlık kazandırılmış bir VLM geliştirme yaklaşımı sunmaktadır (Zhou et al., 2024). Zengin görüntü-metin çiftlerinin yokluğunu aşmak hedefiyle, genel amaçlı modeller ve uzman kaynaklar kullanılarak yaklaşık 70.000 diyalogluk talimat tabanlı bir eğitim veri seti (AgroInstruct) oluşturulmuştur. Modelin temel mimarisi, LLaVA yapısı üzerine inşa edilmiş olup görsel kodlayıcı olarak CLIP yerine SigLIP kullanılmaktadır. AgroGPT-3B ve AgroGPT-7B olarak iki ölçekte tasarlanan bu mimari; görsel-dil hizalama, genel görsel talimat ayarlama ve tarımsal uzmanlık kazandırmayı amaçlayan "expert tuning" (uzman ince ayarı) olmak üzere üç basamaklı bir eğitim sürecinden geçirilmiştir.

Elde edilen deneysel sonuçlar, AgroGPT'nin özellikle ince taneli hastalık ve zararlı tanımlama görevlerinde genel amaçlı büyük modellere kıyasla belirgin

bir üstünlük sağladığını ortaya koymaktadır. Hastalık tanımlama görevinde AgroGPT %51.37 doğruluğa ulaşırken, ChatGPT (%30.82) ve temel LLaVA modeli (%10.27) oldukça geride kalmıştır. Ayrıca insan uzmanlarla yapılan kör değerlendirme sonuçlarında, AgroGPT çıktılarının %86–%96 gibi ezici bir oranla daha fazla tercih edildiği rapor edilmiştir (Zhou et al., 2024).

Özetle; AgroBench tarımsal görevlerde mevcut modellerin sınırlarını sistematik biçimde ortaya koyarken, AgroGPT bu sınırların temel nedeni olan alan bilgisi eksikliğine yönelik uygulanabilir ve başarılı bir mimari çözüm sunmaktadır (Tablo 14).

*Tablo 14. Tarım Odaklı VLM Tasarımı ve Değerlendirmesi*

Özellikler	AgroBench	AgroGPT
<b>Kaynak</b>	Shinoda et al. (2024)	Zhou et al. (2024)
<b>Çalışma Türü</b>	Benchmark / değerlendirme	Model + veri üretimi
<b>Temel Amaç</b>	Tarımsal görsel dil modeller için uzman onaylı standart test ortamı oluşturmak	Tarımsal uzmanlık kazandırılmış verimli VLM geliştirmek
<b>Temel Yaklaşım</b>	Gerçek görüntüler ve uzman doğrulamalı QA ile kıyaslama	Instruction tuning + expert tuning
<b>Kullanılan VLM / LLM</b>	Mevcut görsel dil modeller (GPT-4o, QwenVLM vb.)	AgroGPT-3B, AgroGPT-7B
<b>Görsel Kodlayıcı</b>	—	SigLIP
<b>Dil Modeli</b>	—	LLaVA tabanlı LLM
<b>Mimari Özellik</b>	Model üretmez, performans analizi yapar	LLaVA tabanlı, üç aşamalı eğitim
<b>Girdi Türü</b>	Tarımsal görüntüler + görsel QA	Tarımsal görüntüler + talimatlar
<b>Çıktı Türü</b>	Doğruluk, hata analizi	Tam ve açıklamalı yanıtlar
<b>Veri Setleri</b>	4.218 görüntü, 4.342 QA, 7 görev	AgroInstruct (~70k), AgroEvals
<b>Değerlendirme Metrikleri</b>	Accuracy	Accuracy, insan tercihi
<b>Öne Çıkan Sonuçlar</b>	En büyük hata kaynağı bilgi eksikliği (%51.92)	Küçük modellerle yüksek tarımsal uzmanlık
<b>Güçlü Yan</b>	Güvenilir değerlendirme standardı	Alan-özgü bilgi entegrasyonu
<b>Sınırlılıklar</b>	Model üretmez	Sentetik talimatlara bağımlılık

### 3.7.5. Video Anlama ve Zaman-Bağımlı Görsel Dil Modelleri Uygulamaları

Görsel Dil Modellerinin (VLM) son yıllardaki gelişimi, bu yapıların tekil görüntülerden çıkararak zaman boyutu içeren video verileri üzerinde de kullanılabilirliğini gündeme getirmiştir. Video verisi; zaman içindeki değişimleri, hareket örüntülerini ve sahne sürekliliğini barındırması nedeniyle tek görüntüye dayalı algıdan çok daha karmaşık bir temsil ve akıl yürütme problemi sunmaktadır. Bu nedenle güncel literatür, doğrudan zamansal modelleme yapmak yerine, video verisinin dönüştürülerek uygun temsiller aracılığıyla VLM'lere sunulmasını hedeflemektedir (Chen et al., 2024; Wang et al., 2024).

Bu bağlamda geliştirilen IG-VLM (Image Grid-based Vision–Language Model) yaklaşımı, video anlama problemini özel bir video modeli eğitmeye gerek kalmadan çözmeyi amaçlamaktadır (Chen et al., 2024). Sistemin temel mimarisinde, videodan zamansal olarak örneklenen kareler tek bir birleşik görüntü ızgarası (image grid) hâlinde düzenlenerek önceden eğitilmiş (CogAgent, LLaVA v1.6 ve GPT-4V gibi) bir VLM'e tek girdi olarak sunulur. Böylece video boyunca gerçekleşen olayların zamansal özeti, modelin güçlü görüntü-metin hizalama ve muhakeme yetenekleri üzerinden dolaylı olarak modellenir. MSVD-QA, MSRVT-QA, ActivityNet-QA ve TGIF-QA veri setlerindeki deneysel sonuçlar; kare sayısının, sıralamasının ve ızgara yerleşiminin performansı doğrudan etkilediğini kanıtlamaktadır. Bu bulgular, video için özel olarak eğitilmemiş genel modellerin dahi uygun görsel temsil stratejileriyle zamansal akıl yürütme gerektiren görevleri başarıyla yerine getirebildiğini ortaya koymaktadır (Chen et al., 2024).

Video tabanlı VLM kullanımının gerçek dünya uygulamalarına yönelik bir diğer yenilikçi yaklaşımı, araç gözetimi ve trafik denetimi alanında plaka (ALPR) ile marka-model tanıma süreçlerini hedeflemektedir (Wang et al., 2024). Sabit trafik kameralarının yüksek kurulum maliyetlerine karşın, akıllı telefonlardan veya araç içi kameralardan elde edilen hareketli ve düşük kaliteli videoların kullanımını amaçlayan bu sistem; GPT-4o, Llama-3.2-Vision, LLaVA ve MiniCPM-V gibi modelleri hiçbir özel eğitim gerektirmeden sıfır-atış (zero-shot) biçimde kullanılmaktadır. Kurulan yenilikçi mimaride, video karelerinin kalitesini değerlendirmek ve en bilgilendirici olanları seçmek için CLIP-IQA ve BRISQUE tabanlı ölçütler kullanılmıştır. Ayrıca, araç marka ve model tanıma sürecinde modelin ilk tahminini web'den toplanan referans görüntülerle karşılaştıran ve gerekirse sorguyu düzelten bir öz-yansıtma (self-reflection) modülü geliştirilmiştir.

UFPR-ALPR ve gerçek dünya akıllı telefon video veri setleri üzerindeki değerlendirme sonuçları, plaka tanımadaki %90'ın üzerinde doğruluk elde edildiğini ve marka-model tespitinde geleneksel OCR tabanlı yöntemlere kıyasla belirgin bir üstünlük sağlandığını göstermektedir (Wang et al., 2024). Bu durum, görsel dil modellerinin düşük çözünürlük, hareket bulanıklığı ve açı değişimi gibi zorlu koşullarda oldukça dayanıklı çözümler üretebildiğini doğrulamaktadır.

Genel bir çerçeveden değerlendirildiğinde; görsel dil modellerinin video anlama görevlerinde doğrudan zamansal modelleme yerine temsil dönüştürme, kare seçimi ve test-zamanı muhakeme mimarileriyle son derece etkili biçimde kullanılabilirliği görülmektedir (Tablo 15). Bu yaklaşımlar, VLM'lerin yalnızca statik algı sistemleri olmaktan çıkarak, zaman-bağımlı karar destek bileşenlerine evrildiğini açıkça ortaya koymaktadır (Chen et al., 2024; Wang et al., 2024).

*Tablo 15. Video Anlama ve Zaman-Bağımlı VLM Uygulamaları*

Özellikler	IG-VLM	Self-Reflective VLM (Araç Gözetimi)
<b>Kaynak</b>	Chen et al. (2024)	Wang et al. (2024)
<b>Çalışmanın Amacı</b>	Video-QA görevlerini video için özel model eğitmeden VLM ile çözmek	Gerçek dünya videolarından ölçeklenebilir trafik denetimi sağlamak
<b>Problem Tanımı</b>	Uzun video dizilerinde zamansal bilginin görsel dil modellerde kaybolması	Düşük kalite ve hareketli videolarda ALPR başarısızlığı
<b>Temel Yaklaşım</b>	Videoyu image grid temsiline dönüştürerek VLM'e girdi vermek	Zero-shot VLM + öz-yansıtma ile doğrulama
<b>Kullanılan Görsel Dil Modeller</b>	CogAgent, LLaVA v1.6, GPT-4V	GPT-4o, Llama-3.2-Vision, LLaVA, MiniCPM-V
<b>Zamansal Bilgi Temsili</b>	Kare örnekleme + grid düzeni	Kare seçimi + çoklu sorgulama
<b>Kare Seçimi Stratejisi</b>	Sabit aralıklı örnekleme	CLIP-IQA, BRISQUE
<b>Ek Modüller</b>	Grid ve reasoning guidance promptları	Self-reflection, SAM tabanlı arka plan temizleme
<b>Veri Setleri</b>	MSVD-QA, MSRVT-QA, ActivityNet-QA, TGIF-QA	UFPR-ALPR, gerçek dünya akıllı telefon videoları
<b>Değerlendirme Görevleri</b>	Video Question Answering	ALPR, araç marka/model tanıma
<b>Değerlendirme Metrikleri</b>	Accuracy, VQA Score	Top-1 Accuracy

<b>Öne Çıkan Sonuçlar</b>	Zero-shot görsel dil modellerle etkili video-QA	ALPR'de %90+ doğruluk
<b>Güçlü Yanlar</b>	Video modeli gerektirmemesi	Zorlu koşullara dayanıklılık
<b>Sınırlılıklar</b>	Uzun olay zincirlerinde grid kapasitesi	Aşırı bulanıklık hâlen zorlayıcı

### 3.7.6. İnsan Odaklı Görsel Dil Modelleri Uygulamaları

İnsan odaklı görsel anlama (Human-Centric Vision), insan vücudu ve yüzüne ait görsel ipuçlarından poz (human pose), yüz ifadesi (facial expression) ve davranış (human behavior) gibi yüksek seviyeli semantik bilgilerin çıkarılmasını amaçlamaktadır. Geleneksel yöntemlerin yalnızca görsel özniteliklere dayalı sınırlılıklarını aşmayı hedefleyen güncel yaklaşımlar, Görsel Dil Modellerinin (VLM) anlamsal ve bağlamsal yorumlama yeteneklerini sürece entegre etmektedir.

İnsan pozunu tahmini alanında (Human Pose Estimation) önerilen “Vision-Language Model Guided Pose Knowledge Mining” yaklaşımı, poz bilgisinin otomatik olarak çıkarılmasını ve bu bilginin poz tahmin modellerine aktarılmasını hedeflemektedir (Zhang et al., 2023). Önerilen mimaride, CLIP tabanlı bir görsel-dil modeli kullanılarak poz içeren görüntüler; poz türü, vücut duruşu ve hareket bağlamını tanımlayan metinsel açıklamalarla hizalanır. Elde edilen bu yüksek seviyeli anlamsal bilgiler, klasik poz tahmin ağlarına yardımcı denetim (auxiliary supervision) sinyali olarak entegre edilir. Böylece sistemin yalnızca eklem noktaları arasındaki geometrik ilişkileri değil, pozun bütüncül anlamsal yapısını da öğrenmesi sağlanır. Deneysel sonuçlar, özellikle etiketli verinin sınırlı olduğu (low-label regime) ve karmaşık pozların yer aldığı senaryolarda bu yaklaşımın tahmin performansını anlamlı biçimde artırdığını göstermektedir (Zhang et al., 2023).

Yüz ifadesi analizi (Facial Expression Analysis) alanında öne çıkan SMILE-VLM yaklaşımı ise, 3B ve 4B yüz ifadesi tanıma problemini kendinden denetimli öğrenme (self-supervised learning) ve çoklu görünüm temsili (multi-view representation) perspektifinden çözmeyi amaçlamaktadır (Li et al., 2024). Bu sistemin mimarisinde VLM'ler, farklı kamera açıları ve zaman adımlarında elde edilen yüz görüntülerinin ortak ve tutarlı anlamsal temsillerini öğrenmek üzere konumlandırılmıştır. Görsel-dil hizalaması sayesinde, yüz kaslarının uzamsal ve zamansal değişimleri ile duygusal durumlar arasındaki karmaşık ilişkiler çok daha etkili biçimde modellenir. Elde edilen sonuçlar, etiketlenmiş 3B/4B verilerin kısıtlı olduğu durumlarda bile bu yaklaşımın tam denetimli

yöntemlerle rekabet edebilir bir performans ve yüksek genellenebilirlik sunduğunu kanıtlamaktadır (Li et al., 2024).

Bu çalışmalar birlikte değerlendirildiğinde, görsel–dil modellerinin insan odaklı görsel analiz görevlerinde doğrudan bir tahmin edici bileşen olmaktan ziyade, bilgi çıkarımı (knowledge extraction) ve temsil zenginleştirme (representation enrichment) aracı olarak konumlandığı görülmektedir. Poz ve yüz ifadesi gibi insan davranışını yansıtan karmaşık görsel yapıların VLM destekli yaklaşımlarla modellenmesi, yalnızca görsel bilgilere dayanan yöntemlerin sınırlamalarını tamamlayıcı nitelikte, çok modlu ve bağlama duyarlı bir analiz çerçevesi sunmaktadır. Tablo 16’ da insan odaklı Görsel–Dil Modelleme (Human-Centric Vision–Language Modeling) çalışmaları detaylı olarak analiz edilmiştir.

*Tablo 16. İnsan Odaklı Görsel–Dil Modelleme (Human-Centric Vision–Language Modeling) Çalışmaları*

Özellikler	VLM-Guided Pose Knowledge Mining	SMILE-VLM
<b>Kaynak</b>	Zhang et al. (2023)	Li et al. (2024)
<b>Problem Alanı</b>	İnsan pozu tahmini (Human Pose Estimation)	3B / 4B yüz ifadesi tanıma
<b>Temel Amaç</b>	Pozlara ait yüksek seviyeli anlamsal bilginin çıkarılması ve poz tahminine aktarılması	Etiket gerektirmeden çoklu görünüm ve zamansal yüz temsilleri öğrenmek
<b>Model Türü</b>	CLIP tabanlı VLM + klasik pose estimation ağı	CLIP tabanlı VLM + kendinden denetimli öğrenme
<b>Görsel Temsil</b>	Poz içeren görüntülerden çıkarılan görsel gömüler	Çoklu kamera ve zamansal yüz projeksiyonları
<b>Dil Bileşeni</b>	Poz, duruş ve hareketi tanımlayan metinsel açıklamalar	Yüz ifadelerine dair anlamsal dil temsilleri
<b>Temel Yaklaşım</b>	Poz bilgisi madenciliği ve yardımcı denetim	Self-supervised multi-view öğrenme + VLM hizalaması
<b>Öğrenme Stratejisi</b>	Anlamsal bilginin aşağı akış poz tahmin görevine transferi	Ortak görsel–dil anlamsal uzay öğrenimi
<b>Kullanılan Veri Setleri</b>	COCO, MPII	BU-3DFE, BU-4DFE, Bosphorus, BP4D
<b>Değerlendirme Senaryosu</b>	Sınırlı etiketli ve karmaşık poz içeren görüntüler	Etiketsiz 3B/4B yüz ifadesi tanıma
<b>Başlıca Bulgular</b>	Karmaşık ve az etiketli durumlarda doğruluk artışı	Denetimli yöntemlerle karşılaştırılabilir performans
<b>Uygulama Alanları</b>	İnsan hareketi ve davranış analizi	Affective computing, HCI, mental sağlık

#### 4. Değerlendirme ve Sonuç

İncelenen makaleler, büyük dil modelleri ve görsel dil modellerinin farklı disiplinlerde önemli ilerlemeler sağladığını ancak her alanın kendine özgü yapısal eksiklikler ve çözülmemiş sorunlar barındırdığını göstermektedir.

Görsel Dil Modelleri (VLM) ve bu modellerin eylem boyutuyla genişletilmiş versiyonları olan Vision–Language–Action (VLA) yaklaşımları, yapay zekâ literatüründe yalnızca görsel algıyı iyileştirmekle kalmamış; bilişsel muhakeme, bağlamsal yorumlama ve karar verme süreçlerini de kökten değiştirerek önemli bir paradigma dönüşümü yaratmıştır. İncelenen tüm çalışmalar geniş bir perspektiften değerlendirildiğinde; otonom sistemlerden robotiğe, sağlık bilimlerinden tarıma ve insan odaklı video analizlerine kadar pek çok alanda, salt geometrik ve modüler algı mimarilerinin yerini daha bütüncül, açıklanabilir, çok modlu ve bağlama duyarlı yapıların aldığı açıkça görülmektedir.

Otonom sistemler ve robotik alanında, klasik nokta bulutu veya sınırlayıcı kutu (bounding box) temelli algı sistemleri, VLM'lerin açık sözcük dağarcıklı (open-vocabulary) ve sıfır-atışlı (zero-shot) yetenekleri sayesinde yepyeni semantik bir derinlik kazanmıştır. Otonom sürüşte dil destekli bilgi damıtma stratejileri ve robotikte eylemlerin metinsel belirteçler (action-as-text) olarak modellenmesi, makinelerin görevleri sadece geometrik bir optimizasyonla değil, insan benzeri bir muhakemeyle yorumlamasını sağlamıştır. Bu sayede düşük seviyeli fiziksel özellik tahmininden sosyal navigasyona kadar uzanan geniş bir yelpazede, çevrenin çok daha bütüncül bir bilişsel çerçevede algılanması mümkün kılınmıştır.

Sağlık ve tarım gibi yüksek uzmanlık gerektiren alanlarda ise bu modellerin spesifik ihtiyaçlara göre evrildiği görülmektedir. Sağlık alanında VLM tabanlı sistemler, basit rapor üretiminin ötesine geçerek yapılandırılmış bilgi grafiği oluşturma, kendi kendini iyileştirme (self-refining) mekanizmalarıyla halüsinasyon azaltma ve çok ajanlı karar destek çerçevelerine dönüşmüştür. Etiket kıtlığı yaşanan biyomedikal alanlarda “foundation” (temel) modellerin sıfır-atışlı etiket üretimi büyük avantajlar sunarken; tarım ve çevresel planlamada genel amaçlı modellerin ince taneli sınıflarda (zararlı, hastalık vb.) yetersiz kaldığı tespit edilmiştir. Bu bağlamda, AgroGPT gibi alan-özgü ince ayarlı (expert-tuned) modellerin ve AgroBench gibi uzman onaylı değerlendirme standartlarının, sistem başarısında model ölçeğinden bile daha belirleyici olduğu kanıtlanmıştır. Benzer şekilde video analizi ve insan odaklı görevlerde, ağır zamansal modellemeler yerine görsel ızgara (image grid) gibi temsil dönüştürme stratejileri ve VLM destekli bilgi madenciliği, karmaşık davranışsal göstergeleri başarıyla semantik düzleme taşımaktadır.

Elde edilen tüm bu umut verici metodolojik kazanımlara rağmen literatür, bu modellerin doğrudan gerçek dünya sistemlerine aktarımında kritik mühendislik darboğazlarına işaret etmektedir. Milyarlarca parametrelili modellerin düşük frekanslarda çalışması ve yüksek hesaplama maliyetleri, otonom sürüş veya robotik kontrol gibi anlık tepki gerektiren güvenlik-kritik senaryolarda doğrudan kullanımı sınırlandırmaktadır. Ek olarak, mekânsal ve anlamsal halüsinasyon riskleri, medikal ve endüstriyel ortamlar gibi hataya sıfır toleranslı alanlarda VLM'lerin tek başına karar alıcı olarak konumlandırılmasını oldukça riskli hâle getirmektedir. Ayrıca bulut bağımlılığı ve donanım kısıtları, beraberinde veri gizliliği ve ağ gecikmesi problemlerini getirmektedir.

Bu tabloyu göz önünde bulundurduğumuzda, VLM ve VLA tabanlı sistemlerin geleceğinin yalnızca “daha büyük modeller” üretmekten ibaret olmadığı düşünülmektedir. Aksine geleceğin vizyonu; bilgi damıtma, niceleme (quantization) ve LoRA gibi parametre-verimli tekniklerle sıkıştırılmış, uç cihazlarda (edge) çalışabilen, göreve özgü uyarlanmış ve çok ajanlı (multi-agent) mimarilerle desteklenen hibrit sistemlere işaret etmektedir. Kısa ve orta vadede bu modeller, bağımsız kontrolcü sistemler olmaktan ziyade; üst düzey semantik rehberlik sağlayan, süreçlerin açıklanabilirliğini artıran ve insan operatörlerle uyum içinde çalışan “bilişsel karar destek bileşenleri” olarak konumlanacaktır. Görsel dil modellerinin başlattığı bu güçlü metodolojik dönüşümün güvenli, etik ve sürdürülebilir biçimde ilerlemesi; ancak alan-öзgü uyarlama, sıkı denetim mekanizmaları ve güvenilirlik odaklı tasarım ilkelerinin bütüncül bir yaklaşımla benimsenmesine bağlıdır.

## Kaynaklar

- Alammar, J. (2024). Hands-on large language models: Language understanding and generation. O'Reilly Media.
- Bimbraw, K., Singh, A., Patel, R., & Ghosh, S. (2024). Hand gesture decoding from forearm ultrasound images via vision–language models. *IEEE Transactions on Medical Imaging*. <https://arxiv.org/abs/2407.10870>
- Bordes, F., Pang, R. Y., Ajay, A., Li, A. C., Bardes, A., Petryk, S., & Chandra, V. (2024). An introduction to vision-language modeling. *arXiv*. <https://arxiv.org/abs/2405.17247>
- Brohan, A., Brown, N., Carbajal, J., et al. (2023). RT-2: Vision-language-action models transfer web knowledge to robotic control. *arXiv*. <https://arxiv.org/abs/2307.15818>
- Cai, H., et al. (2025). Unilaw-R1: A large language model for legal reasoning with reinforcement learning and iterative inference. *arXiv*.
- Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N., & Androutsopoulos, I. (2022). LexGLUE: A benchmark dataset for legal language understanding. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (ACL)*.
- Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y., Ye, W., Yu, P. S., Yang, Q., & Xie, X. (2024). A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology*, 15(3), Article 39. <https://doi.org/10.1145/3641289>
- Chen, Y., Ding, Z.-H., Wang, Z., Wang, Y., Zhang, L., & Liu, S. (2024). Asynchronous large language model enhanced planner for autonomous driving (AsyncDriver). *arXiv*. <https://arxiv.org/abs/2406.14556>
- Chen, X., Li, Y., Wang, H., & Zhang, Z. (2024). An image grid can be worth a video: Zero-shot video question answering using a vision–language model. *arXiv*.
- Du, Z., Qian, Y., Liu, X., Ding, M., Qiu, J., Yang, Z., & Tang, J. (2023). VisualGLM: Visual and language model for multimodal dialogue. *arXiv*. <https://arxiv.org/abs/2304.08857>
- Encord. (2023). LLaVA: Large language and vision assistant. <https://encord.com/blog/llava-large-language-vision-assistant/>
- Gkournelos, C., et al. (2024). An LLM-based approach for enabling seamless human–robot collaboration in assembly. *IEEE Robotics and Automation Letters*.
- Google DeepMind. (2024). AutoRT: Embodied foundation models for large-scale orchestration of robotic agents. <https://deepmind.google/discover/blog/>
- Greco, A., et al. (2025). An experimental evaluation of smart sensors for pedestrian attribute recognition using multi-task learning and vision–language models.

- Guha, N., Nyarko, J., Ho, D. E., Ré, C., & Chilton, L. (2023). LegalBench: A benchmark for measuring legal reasoning in large language models. arXiv. <https://arxiv.org/abs/2308.11462>
- He, K., et al. (2025). A survey of large language models for healthcare.
- Ho, H.-T., Nguyen, L. V., Pham, M.-T., Pham, Q.-H., Tran, Q.-D., Huy, D. N. M., & Nguyen, T.-H. (2025). A review on vision-language-based approaches: Challenges and applications. *Computers, Materials & Continua*, 82(2), 1733–1760. <https://doi.org/10.32604/cmc.2025.060363>
- Hu, Y., et al. (2023). Planning with vision-language models and a use case in robot-assisted teaching.
- Hugging Face. (n.d.). What are LLMs? In Agents Course. <https://huggingface.co/learn/agents-course/unit1/what-are-llms>
- Jadhav, A., & Mirza, V. (2025). Large language models in equity markets. *Frontiers in Artificial Intelligence*, 8, 1608365. <https://doi.org/10.3389/frai.2025.1608365>
- Jia, E., Mao, W., Liu, Y., Zhao, Y., Wen, Y., Zhang, C., Zhang, X., Wang, T., et al. (2023). ADriver-I: A general world model for autonomous driving. arXiv. <https://arxiv.org/abs/2311.13549>
- Jung, K. H. (2025). Large language models in medicine: Clinical applications, limitations, and ethics.
- Kawaharazuka, K., Iwase, M., & Sugita, N. (2023). Vision-language-action models for robotics: A review towards real-world applications. arXiv.
- Kong, Y., Nie, Y., Dong, X., Mulvey, J. M., Poor, H. V., Wen, Q., & Zohren, S. (2024). Large language models for financial and investment management: Applications and benchmarks.
- Li, B., et al. (2025). CoLE: A collaborative legal expert prompting framework for large language models in law.
- Li, J., et al. (2025). Guiding multiple remote users in physical tasks with language-driven robotic telepresence. *ACM Human-Computer Interaction*.
- Li, Q., Zhao, Y., Sun, Z., & Wang, X. (2024). SMILE-VLM: Self-supervised multi-view representation learning using vision-language models for 3D/4D facial expression recognition. arXiv.
- Li, T., Wang, H., Tan, J., Kong, L., Zhang, H., Pan, D., & Zhao, Z. (2025). Intelligent quality assessment of concrete vibration using computer vision and large language models. *Automation in Construction*, 180, 106507. <https://doi.org/10.1016/j.autcon.2025.106507>
- Lin, C., et al. (2025). Roles and potential of large language models in healthcare.
- Lin, T., Zhang, W., Li, S., Yuan, Y., Yu, B., Li, H., He, W., Jiang, H., Li, M., Song, X., Tang, S., Xiao, J., Lin, H., Zhuang, Y., & Ooi, B. C. (2025). HealthGPT: A medical large vision-language model for unifying comp-

- rehension and generation via heterogeneous knowledge adaptation. In Proceedings of the 42nd International Conference on Machine Learning (ICML 2025) (Proceedings of Machine Learning Research, Vol. 267, pp. 37975–37995). PMLR. <https://proceedings.mlr.press/v267/lin25a.html>
- Liu, H., Li, C., Li, Y., & Lee, Y. J. (2024). Multilingual visual question-answering for dermatology through vision–language model fine-tuning and large language model translations. In Proceedings of the MEDIQA-M3G 2024 Shared Task.
- Liu, J., Wang, Y., Zhang, Q., Zhao, S., & Li, H. (2024). A large language model for multimodal identification of crop diseases and pests. arXiv.
- Ma, Y., Yao, Z., Liu, X., Xiong, Z., He, X., & Wu, X. (2026). Efficient 3D object annotation via vision-derived pseudo-LiDAR and vision–language model validation. *Transportation Research Part C: Emerging Technologies*, 182, 105429. <https://doi.org/10.1016/j.trc.2025.105429>
- Macdonald, J., et al. (2023). Language, camera, autonomy! Prompt-engineered robot control for rapidly evolving deployment. arXiv.
- Mei, T., Zhang, W., & Yao, T. (2019). Vision and language: From visual perception to content creation. *Signal Processing*, 164, 22–35. <https://doi.org/10.1016/j.sigpro.2019.06.008>
- Minaee, S., Mikolov, T., Nikzad, N., Chenaghlu, M., Socher, R., Amatriain, X., & Gao, J. (2025). Large language models: A survey. arXiv. <https://arxiv.org/abs/2402.06196>
- Navarro, H. J., et al. (2025). Large language models in medicine: A systematic review.
- Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., & Mian, A. (2024). A comprehensive overview of large language models. arXiv. <https://arxiv.org/abs/2307.06435>
- Ong, J. C. L., et al. (2025). Ethical and regulatory challenges of large language models in healthcare.
- OpenAI. (2021). CLIP: Connecting vision and language. <https://openai.com/index/clip/>
- Osada, M., Garcia Ricardez, G. A., Suzuki, Y., & Taniguchi, T. (2024). Reflectance estimation for proximity sensing by vision-language models: Utilizing distributional semantics for low-level cognition in robotics. *Advanced Robotics*, 38(18), 1287–1306. <https://doi.org/10.1080/01691864.2024.2393408>
- Patil, R., & Gudivada, V. (2024). A review of current trends, techniques, and challenges in large language models (LLMs). *Applied Sciences*, 14(5), 2074. <https://doi.org/10.3390/app14052074>
- Peng, Y., et al. (2024). Vision-language model-enabled street view analytics: A systematic literature review.

- Peng, H., Zhang, L., Wu, J., & Patel, V. M. (2024). PRIMA: A foundation vision-language model for large-scale neuroimaging analysis. arXiv.
- Pollini, A., et al. (2024). Reducing latency in LLM-based robot control via ROS 2 integration.
- Polo Club. (n.d.). Transformer explainer: LLM transformer model visually explained. <https://poloclub.github.io/transformer-explainer>
- Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... Sutskever, I. (2021). Learning transferable visual models from natural language supervision. In Proceedings of the 38th International Conference on Machine Learning (ICML).
- SaM Solutions. (n.d.). LLM architecture: A comprehensive guide. <https://sam-solutions.com/blog/llm-architecture>
- Sapkota, M., et al. (2025). A review of 3D object detection with vision-language models.
- Sapkota, S., Zhang, Y., Wang, L., et al. (2025). Vision-language-action models: Concepts, progress, applications and challenges. arXiv.
- Shao, M., Basit, A., Karri, R., & Shafique, M. (2024). Survey of different large language model architectures: Trends, benchmarks, and challenges (arXiv:2412.03220). arXiv. <https://arxiv.org/abs/2412.03220>
- Sharshar, M., Kanaan, H., & Abou-Zeid, H. (2023). Robotic environmental state recognition with pre-trained vision-language models and black-box optimization. Robotics and Autonomous Systems, 168, 104504.
- Sharshar, A., Khan, L. U., Ullah, W., & Guizani, M. (2025). Vision-language models for edge networks: A comprehensive survey. IEEE Access. Advance online publication. <https://arxiv.org/abs/2502.07855>
- Shekhar, A. C., et al. (2025). Use of a large language model for ambulance dispatch and triage.
- Shinoda, R., Inoue, N., Kataoka, H., Onishi, M., & Ushiku, Y. (2024). Agro-Bench: Vision-language model benchmark in agriculture. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). <https://arxiv.org/abs/2403.08819>
- Shu, D., et al. (2024). LawLLM: Law large language model for the US legal system.
- Song, X., et al. (2024). VLM-Social-Nav: Socially aware robot navigation through scoring using vision-language models.
- Ullah, E., et al. (2024). Challenges and barriers of large language models in digital pathology.
- Wang, J. (2024). Hallucination reduction and optimization for large language model-based autonomous driving. Symmetry, 16(9), 1196.

- Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., Chen, Z., Tang, J., Chen, X., Lin, Y., Zhao, W. X., Wei, Z., & Wen, J. (2024). A survey on large language model based autonomous agents. *Frontiers of Computer Science*, 18(6), 186345.
- Wang, Y., Li, J., Xu, R., & Zhao, D. (2024). Vision language models in autonomous driving: A survey and outlook.
- Wang, Y., et al. (2025). Vision-language model-based human-guided mobile robot navigation in an unstructured environment for human-centric smart manufacturing.
- Xia. (2025, June 15). How vision language models are trained: A deep dive into the VLM training process. Medium. <https://medium.com/@xiaxiami/how-vision-language-models-are-trained-a-deep-dive-into-the-vlm-training-process-1ba1d8704bb0>
- Xu, Z., Wang, Y., Chen, X., Liu, J., & Shi, B. (2024). VLM-Grounder: A vision-language model agent for zero-shot 3D visual grounding. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 19,845–19,855).
- Yang, H., Liu, X.-Y., & Wang, C. D. (2023). FinGPT: Open-source financial large language models. arXiv. <https://arxiv.org/abs/2306.06031>
- Yang, R., et al. (2023). Large language models in health care.
- Zhang, C., et al. (2024). NaVid: Video-based vision-language model plans the next step for vision-and-language navigation.
- Zhang, J., Xu, C., & Li, B. (2024). ChatScene: Knowledge-enabled safety-critical scenario generation for autonomous vehicles. arXiv. <https://arxiv.org/abs/2405.14062>
- Zhang, J., Huang, J., Jin, S., & Lu, S. (2024). Vision-language models for vision tasks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. <https://arxiv.org/abs/2304.00685>
- Zhang, K., et al. (2025). Revolutionizing healthcare with large language models.
- Zhang, P., Lin, K., Li, D., Fu, Z., Cai, Y., Li, B., Yu, H., & Li, M. (2025). DAP-lanner: Dual-agent framework with multimodal large language model for autonomous driving motion planning. *Applied Soft Computing*.
- Zhang, Y., et al. (2024). Agentic LLM-based robotic systems for real-world applications: A survey.
- Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., Liu, P., Nie, J.-Y., & Wen, J.-R. (2025). A survey of large language models (Version 16). arXiv. <https://arxiv.org/abs/2303.18223>

- Zhong, L., Huang, Z., Liu, Y., Liao, W., Zhang, S., Wang, G., & Zhang, S. (2025). VLM-CPL: Consensus pseudo-labels from vision-language models for annotation-free pathological image classification. *IEEE Transactions on Medical Imaging*, 44(10), 4023–4036. <https://doi.org/10.1109/TMI.2025.3595111>
- Zhou, K., Yang, J., Loy, C. C., & Liu, Z. (2022). Learning to prompt for vision-language models. *International Journal of Computer Vision*, 130(9), 2337–2348. <https://arxiv.org/abs/2109.01134>
- Zhou, Y., Wang, H., Li, X., Zhang, Y., & Sun, J. (2024). AgroGPT: Efficient agricultural vision–language model with expert tuning. *arXiv*.
- Zhu, H., Zhang, Y., Li, Z., & Zhu, Y. (2024). RoboPoint: A vision-language model for spatial affordance prediction for robotics. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 14,732–14,741).

## Yapay Zekâ Tabanlı Sentetik Hasta Modellemesi ile Karar Destek Sistemlerinin Geliştirilmesi

Utku Köse<sup>1</sup>

Emine Betül Sürücü<sup>2</sup>

### Özet

Sağlık alanında karar destek sistemlerinin gelişmesi için yapay zeka çözümleri kullanılmaktadır. Geliştirilen yaklaşımlar, verilerin temel özelliklerini yansıtan yapay hasta grupları oluşturulmasına olanak sağlamaktadır. Sentetik hasta modelleri gerçek hasta verilerinin istatistiksel ve klinik özelliklerini taklit ederek yapay veri kümelerinin üretilmesinin mümkün kılmaktadır. Elde edilen sentetik veri kümeleri, özellikle hasta gizliliği ile ilgili kısıtlamalar veya veri erişimindeki sınırlılıklar nedeniyle büyük ölçekli tıbbi veri setlerine ulaşmanın zor olduğu durumlarda önemli bir alternatif sunmaktadır. Bu tür veri kümeleri, araştırmacıların farklı hasta senaryolarını incelemesine, tedavi süreçlerini simüle etmesine ve yapay zekâ tabanlı tahmin modelleri geliştirmesine olanak sağlamaktadır. Böylece sentetik hasta modelleme yaklaşımları, klinik karar destek sistemlerinin geliştirilmesi ve kişiselleştirilmiş tedavi stratejilerinin araştırılması açısından değerli bir araştırma altyapısı sunmaktadır. Bu çalışmada, yapay zekâ tabanlı sentetik hasta oluşturma ve klinik karar destek sistemlerindeki potansiyel uygulamalarına genel bir bakış sunmaktadır. Çalışma, veri entegrasyonu, tedavi simülasyonu, doz optimizasyonu ve tahmine dayalı modelleme yaklaşımları da dahil olmak üzere sentetik hasta modellemesinin temel bileşenlerini tartışmaktadır. Bunun yanında büyük ölçekli klinik verilerin modellenmesi ve hastaya özgü tedavi sonuçlarının tahmin edilmesi konusunda, özellikle grafik tabanlı derin öğrenme modelleri olmak üzere gelişmiş makine öğrenme tekniklerinin rolü incelenmektedir. Hesaplamalı hasta simülasyonlarını yapay zekâ algoritmalarıyla birleştiren hibrit modelleme stratejilerinin kişiselleştirilmiş tedavi stratejilerinin geliştirilmesini nasıl destekleyebileceğini vurgulamaktadır. Simülasyon tabanlı hasta modellemesini modern makine öğrenme yaklaşımlarıyla entegre ederek,

1 Prof.Dr., Süleyman Demirel Üniversitesi, utkukose@sdu.edu.tr, 0000-0002-9652-6415

2 Öğr. Gör., Süleyman Demirel Üniversitesi, betulsen@sdu.edu.tr, 0000-0001-8084-0015

karmaşık tedavi dinamiklerini analiz etmek ve sağlık uzmanlarının klinik karar verme süreçlerinde yardımcı olabilecek ön bilgiler üretmek mümkün hale gelmektedir. Genel olarak, yapay zekâ destekli sentetik hasta modellemesi, veri odaklı karar destek sistemlerini iyileştirmek ve kişiselleştirilmiş tıbbi geliştirmek için umut vadeden bir araştırma yönünü temsil etmektedir.

## 1. Giriş

Yapay zekâ teknolojilerindeki hızlı ilerleme, birden fazla sektörün gelişmesine önemli bir katkı sağlamaktadır. Gelişimin ve yenilenmenin en çok ihtiyaç olduğu sektörlerden biri olan sağlık sektöründe yapay zeka teknikleri oldukça yaygın bir şekilde kullanılmaktadır. Çok bilinmeyenli süreçlerin var olduğu ve çeşitli hasta senaryolarının olduğu onkoloji alanında yapay zeka önemli ölçüde gelişim sağlamaktadır. Klinik karar verme süreçleri modern sağlık hizmetlerinde önemli ölçüde gelişmektedir. Özellikle hastalığın sürecinin karmaşık ve hasta özelliklerinin büyük farklılıklar gösterdiği kanser tedavisinde klinik karar verme süreçlerinin geliştirilmesi daha önemli bir hale gelmektedir. Kanser tedavisinde birçok değişken bulunmaktadır. Hastaya özgü fizyolojik özellikler, tedavi farklılıkları, ilacın hastalardaki yanıt değişkenleri gibi farklılıklar sebebiyle risk ve fayda durumu değişmektedir (Topol, 2019).

Son zamanlara yapılan yapay zekâ destekli hesaplama modelleri, sağlık profesyonellerine büyük ölçekli tıbbi verileri analiz etmede ve kişiselleştirilmiş tedavi stratejileri için tahmine dayalı bilgiler üretmede yardımcı olmaktadır. Hastalığın teşhis aşamasında makine öğrenme algoritmaları, derin öğrenme mimarileri ve hesaplama modelleri, hastalığın gidişatı tahmininde, tedavi planlamasında ve ilaç keşfinde yaygın olarak uygulanmaktadır (Esteve vd., 2019). Yapay zekâ teknikleri büyük veri kümelerini analiz ederek tedavi sonuçları yüksek doğrulukta tahmin ederek klinik karar vermeyi amaçlamaktadır. Geliştirilen modellerin faydasının yanında zorluklar da bulunmaktadır. Büyük veri setlerine ulaşmak yapay zekâ temelli geliştirme yapmaktaki en temel zorluklardan birisi olmaktadır. Veri setlerine olan ulaşım sınırının sebebi hasta mahremiyeti ve etik düzenlemeler olmaktadır. Bunun yanında mevcut veri kümelerinin çoğu sınırlı hasta sayısı, gözlem konusundaki eksiklikler ve karışık türde veri yapısı içeriğinden kaynaklanmaktadır. Bu gibi durumlar da yapay zekâ modellerinin güvenli ve genel sonuçlar üretmesini zorlaştırmakta ve klinik uygulama entegrasyon sürecini karmaşık hale getirmektedir.

Hastalık için klinik karar verme süreçlerini iyileştirmek amacıyla faydalanılan yapay zeka tekniklerinin zorluklarından biri olan veri erişim sınırlandırmasının üstesinde gelmek için araştırmacılar, sentetik hasta verilerinin oluşturulmasına odaklanmaktadır (Bates vd., 2014). Sentetik veri kümeleri, hasta gizliliğini korumakta, büyük ölçekli hesaplama deneylerine

olanak sağlamakta, istatistiksel olarak gerçek hasta verilerine benzeyen yapay verilerden oluşmaktadır. Oluşturulan veri kümeleri, araştırmacıların hasta kayıtlarına doğrudan güvenmeden klinik senaryoları simüle etmeleri, tedavi stratejilerini değerlendirmeleri ve makine öğrenimi modellerini eğitmelerine olanak tanımaktadır (Chen vd., 2021). Araştırmalarda, kanser hücrelerinin karmaşıklığı ve bunun yanında kişiselleştirilmiş tedavi stratejilerine duyulan ihtiyaçlar sebebiyle sentetik hasta üretimi önemli hale gelmektedir. Karar destek sisteminin oluşması sürecinde ilk adım olan matematiksel modellemelerin yapılması hastalık ilerlemesini, ilaç yanıtını ve tedavi sonuçlarını simüle edebilen gerçekçi sanal hasta popülasyonları oluşturmaktır. Bu süreç tedavi desteği için çok güçlü bir çerçeve sağlamaktadır. Sanal hasta modelleri, tedavi yöntemlerini değerlendirmek, doz optimizasyonu yapmak ve karar verme süreçlerini desteklemek için klinik araştırmalarda giderek daha fazla kullanılmaktadır (Viceconti vd., 2021).

Bu çalışmada, yapay zekâ tabanlı sentetik hasta üretimi yaklaşımları ve bu yaklaşımların temel prensipleri ele alınmaktadır. Sentetik hasta üretiminin klinik karar destek sistemlerine olan katkısı belirtilmektedir. Aynı zamanda sentetik hasta modellerinin klinik veri eksikliğini azaltma, farklı tedavi senaryolarının simülasyonunu gerçekleştirme ve kişiselleştirilmiş tedavi kararlarının desteklenmesine yönelik potansiyeli de ele alınmaktadır. Çalışma veri odaklı üretken modelleri, matematiksel modelleme yaklaşımlarını ve yapay zekâyı farmakokinetik-farmakodinamik modellemeyle bütünleşmiş eden çeşitli sanal hasta simülasyonları için ele almaktadır (Dash, S., vd.,2019). Ek olarak, yapay zekâ tabanlı klinik karar destek sistemlerinin geliştirilmesinde sentetik hasta rolü, dijital ikiz teknolojileri perspektifleriyle birlikte incelenmektedir. Çalışmada kullanılan makine öğrenmesi algoritmaları ve derin öğrenme modelleri, sentetik hasta simülasyonları sonucu elde edilen veriler, analiz yapıldıktan sonra tedavi sonucu ile ilgili durumları tahmin edebilmektedir. Graph Neural Network (GNN) yapısı kullanılarak geliştirilen modeller, hasta verilerindeki ilişkileri analiz etmektedir. Böylelikle farklı hasta özellikleri ve tedavi parametreleri arasındaki ilişkiler tedavinin yanıtının doğru bir şekilde tahmin edilmesini mümkün kılmaktadır. Matematiksel modelleme, sentetik hasta üretimi ve yapay zekâ tabanlı tahmin mekanizmalarının birlikte kullanılması kurulmak istenen karar destek yapısı için bütüncül bir çözüm sunmaktadır.

## 2. Kanser Tedavisinde Klinik Karar Verme Süreçleri

Kanser tedavileri birçok klinik, biyolojik ve hastaya göre değişen parametreler içermektedir. Hastalıklı hücrelerin biyolojik özellikleri, genetik faktörü, hastalığın evresi, yaş faktörü, beslenme etkisi, egzersiz faktörü gibi birçok faktör tedavi planının belirlenmesinde önemli rol oynamaktadır. Bu sebeple kanser

hastalığında tedavinin şekli belirlenirken standartların yanında başka etmenlere de bakılarak tedavinin belirlenmesi gerekmektedir. Tedavi sürecinde cerrahi müdahale, kemoterapi ve hedefe yönelik tedavi yöntemleri kullanılmaktadır. Bu tedavi türlerinin seçimi ve yine tedavilerde kullanılacak ilaçların dozlarının belirlenmesi, hastalığın tipi ve evresi dışında hastanın bireysel özelliklerini de önemli ölçüde dikkate alarak yapılmaktadır. Farklı özellikteki hastalara aynı tip tedavi uygulandığında istenilen sonuçlar elde edilmeyebilmektedir. Bu durumda kişiye özel tedavinin önemini göstermektedir.

Tedavi protokollerinde doz ayarlanması oldukça kritik bir aşama olmaktadır. Yetersiz miktarda doz verilmesi tedavinin etkinliğini azaltabileceken, yüksek miktarda doz ise ciddi yan etkilere ve vücuttaki toksisite oranının artmasına sebep olmaktadır. Bu gibi sebeplerden ötürü tedavi planlamasının tedavinin etkinliğini arttırmasına yönelik ve yan etkileri minimize edecek ölçüde azaltılmasında çok önemli bir yer bulunmaktadır. Bununla birlikte kanser tedavisindeki karar verme süreçleri bazı belirsizlikler içermektedir. Bu belirsizliklerin en başında gelen hastanın tedaviye vereceği yanıtı tam olarak öngörememek, farklı tedavi seçeneklerinin etkilerinin kesin olarak bilinememesi ve tedavi sonucu oluşabilecek yan etkiler gibi faktörlerin bireyler arasında değişkenlik göstermesi tedavi sürecini çok daha karmaşık hale getirmektedir. Bu nedenlerden dolayı son yıllarda klinik karar verme süreçlerinin desteklemek amacıyla veri analizi, matematiksel modelleme ve yapay zekâ tabanlı yöntemlerin kullanımı tercih edilmektedir.

### **3. Klinik Karar Destek Sistemleri**

Klinik karar destek sistemleri, sağlık uzmanlarının tanı koymak, tedavi planlamak ve hastanın durumunu yönetmek gibi süreçlerinde daha bilinçli ve veri temelli kararlar almasına yardımcı olmak amacıyla geliştirilen sistemler olarak tanımlanabilmektedir. Hasta verilerini analiz ederek klinik karar verme sürecini destekleyen öneriler ve tahminler sunana bu sistemlerin temel amacı, sağlık uzmanlarının sahip olduğu klinik bilgi ve deneyimi veri analitiği ve hesaplamalı yöntemlerle destekleyerek daha doğru ve etkili kararlar alınmasını sağlamaktır. Her gün gelişen ve değişen teknoloji yöntemleriyle birlikte sağlık alanında klinik karar destek sistemlerinin kullanımı son yıllarda önemli ölçüde artmıştır. Elektronik sağlık kayıtlarının yaygınlaşması ve büyük veri analiz yöntemlerinin gelişmesi, bu sistemlerin daha geniş veri setlerini analiz edebilmesine olanak sağlamaktadır. Böylelikle hasta verileri, laboratuvar sonuçları, görüntüleme verileri ve tedavi geçmişi gibi birçok farklı veri kaynağına kolaylıkla tek bir sistemden ulaşılmasıyla birlikte klinik karar verme süreçleri desteklenebilmektedir.

Kanser tedavisinde klinik karar destek sistemleriyle birlikte özellikle tedavi planlaması, risk değerlendirmesi ve tedavi yanıtlarının tahmin edilmesi gibi konularda önemli bir gelişme sağlanmaktadır. Bu sistemlerle birlikte hastaların klinik özellikleri analiz edilerek farklı tedavi seçenekleri sunulabilmekte ve bu tedavilerin olası sonuçları değerlendirilebilmekte ve sonuç olarak daha uygun tedavi planları oluşturulabilmektedir. Bunların yanında tedaviye bağlı yan etkilerin ve toksisite risklerinin önceden tahmin edilmesi de klinik karar destek sistemleri aracılığıyla mümkün olabilmektedir. Son yıllarda yapay zekâ ve makine öğrenmesi yöntemlerinin gelişmesi klinik karar destek sistemlerinin de kapasitesi önemli ölçüde arttırmaktadır. Büyük ve karmaşık veri setlerinin analizi makine öğrenmesi algoritmalarıyla kolay ve hızlı bir şekilde yapılabilmekte, hasta özellikleri ile tedavi sonuçları arasındaki karmaşık ilişkileri ortaya çıkarabilmektedir. Bu sayede klinik karar destek sistemleri yalnızca mevcut durumdaki verileri analiz etmenin yanında aynı zamanda gelecekteki tedavi sonuçlarını da tahmin edebilen sistemler haline gelmektedir. Bununla birlikte bu sistemlerin geliştirilmesinde karşılaşılan en önemli zorluklardan biri yüksek kaliteli ve yeterli büyüklükte klinik veri kümelerine erişimin sınırlı olmasıdır. Klinik verilerin gizlilik ve etik nedenlerle paylaşımının kısıtlı olması, araştırmacıların geniş veri setlerine erişimini zorlaştırmaktadır. Bu durum, klinik karar destek sistemlerinin geliştirilmesinde alternatif veri üretim yöntemlerinin kullanılmasını gerekli hale getirmektedir.

#### 4. Yapay Zeka Tabanlı Sentetik Hasta Üretimi

##### 4.1 Sentetik Hasta Üretim Süreci

Sentetik hasta üretimi, veri kümelerinde gözlemlenen klinik değişkenliği yeniden üreten yapay hasta gruplarının oluşturulmasını ifade etmektedir. Sağlık alanındaki araştırmalarda, büyük ölçekli hasta verilerine erişim genellikle gizlilik düzenlemeleri, etik kısıtlamalar gibi nedenlerden dolayı genellikle sınırlı kalmaktadır. Bu sebepten ötürü sentetik veri üretimi, sağlık hizmetlerinde yapay zekâ modellerinin geliştirilmesi ve doğrulanması için önemli bir durum haline gelmektedir (Chen vd., 2021). Sentetik hasta üretim sürecinde hasta değişkenlerinin belirlenmesi ilk ve en önemli adım olmaktadır. Bu değişkenler hastanın bedensel ölçüleri, bireysel özellikleri, tedavi planları, laboratuvar bulguları ve yaşam tarzı gibi özelliklerden oluşabilmektedir. Bu gibi parametreler gerçek hastalarda gözlemlenen çeşitliliği yansıtan sanal hasta örnekleri oluşturmak için kullanılmaktadır (Tucker vd., 2020).

Son zamanlarda yapılan çalışmalar, sentetik hasta üretiminde genellikle vücut işleyişini ve tedavi yanıtını simüle eden hesaplamalı hasta modelleriyle desteklenmektedir. Modeller, vücut yüzey alanı, genetik özellikler, tedavi dozu ve çevresel faktörler gibi klinik değişkenlerden elde edilmiş bir simülasyon

çerçevesine entegre edilmektedir (Davenport, T., 2019). Bu yaklaşımlar, tahmine dayalı modelleme ve karar destek sistemi geliştirme için kullanılabilen büyük ölçekli sanal olarak aynı hastalığa sahip hasta grubunun oluşturulmasına olanak tanımaktadır (Barbiero & Lió, 2020). Lösemi tedavi modellemesinde, sentetik hasta verileri vücut ağırlığı, boy, tedavi dozu ve hematolojik parametre gibi değişkenleri içerebilmektedir. Bu değişkenler, hesaplamalı modellerin farklı hastaların kemoterapi tedavisine zaman içinde nasıl yanıt verebileceğini simüle etmesine olanak tanımaktadır. Elde edilen veri setleri daha sonra klinik karar verme süreçlerini desteklemek üzere tasarlanmış yapay zekâ algoritmalarının eğitilmesinde kullanılabilir.

#### **4.2 Sentetik Hastalar Kullanarak Tedavi Simülasyonu**

Sentetik hasta grupları oluşturulduktan sonra, tedavi simülasyon modelleri bu hastaların farklı tedavi müdahalelerine nasıl yanıt verdiğini değerlendirmek için kullanılmaktadır. Hesaplamalı tedavi simülasyon modelleri, tedavi yanıtını yansıtan klinik parametrelerin gelişimini izlemektedir. Lösemi tedavisinde, beyaz kan hücresi (WBC) sayısı ve mutlak nötrofil sayısı (ANC) gibi hematolojik göstergeler, tedavi ilerlemesinin izlenmesinde merkezi bir rol oynamaktadır (Jost vd., 2020). Biyolojik parametreler kemik iliğinin vücuttaki aktivesini yansıtmakta, kemoterapinin sonucunu ve toksisite riskini değerlendirmek için kullanılmaktadır. Tedavi simülasyonu, ilaç yanıtı ve ilacın vücuttaki toksisite miktarının bireyler arasında önemli ölçüde değişebildiği akut lenfoblastik lösemi (ALL) gibi hastalıklarda özellikle önemli olduğunu Köse ve arkadaşları (2025) gerçekleştirdikleri çalışmada bahsetmektedir. Çalışmalarında WBC ve ANC parametrelerin zamana bağlı değişimini modellemek amacıyla GNN tabanlı bir dijital ikiz yaklaşımı geliştirilmiştir. Bu çalışmada yapılan analizler, sentetik veya simüle edilmiş hasta verileri kullanılarak geliştirilen yapay zekâ modellerinin, hastaya özgü zaman serisi tahminlerinde etkili sonuçlar üretebildiğini göstermektedir (Köse et al., 2025).

Modelleme çalışmalarında, sentetik hasta verilerine dayalı tedavi simülasyonlarının, hastalardaki biyolojik parametrelerin ne durumda olduğunu ve tedavi yanıtlarını yeniden üretebildiğini göstermektedir. Hematolojik parametrelerin zamansal değişimini modellenmesi, araştırmacıların farklı kemoterapi stratejilerinin zaman içinde hasta sonuçlarını nasıl etkilediğini analiz edebilmektedir (Le vd., 2019). Simülasyon tabanlı yaklaşımlar, araştırmacıların gerçek hastalarda test edilmesi zor ve güvenli olmayan tedavi senaryolarını değerlendirmelerine olanak tanıdığı için tıbbi araştırmalar için önemli bir avantaj sağlamaktadır (Krittawong C., vd., 2027). Farklı ilaç dozaj planlamaları, tedavi programları veya çevresel koşullar, klinik uygulamaya geçmeden önce sanal hasta grupları kullanılarak test edilmektedir.

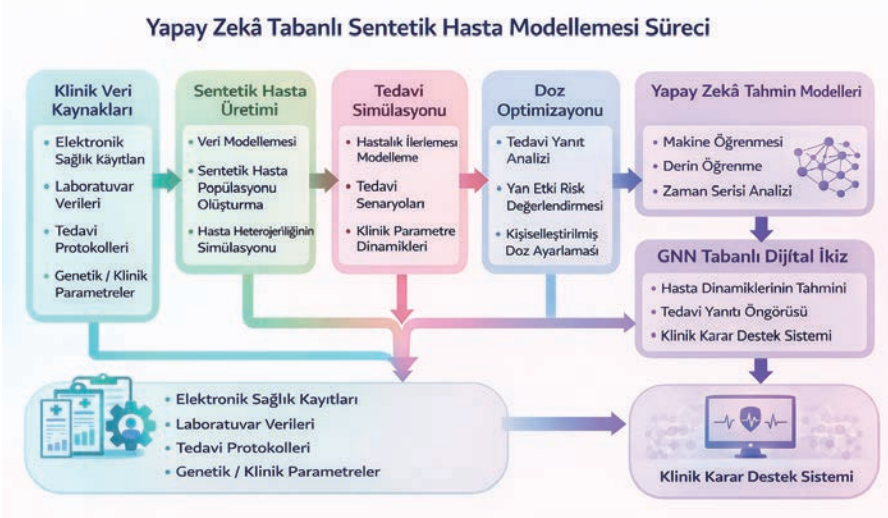
### 4.3 Doz Optimizasyon Stratejileri

Sentetik hasta modellemesinin en önemli uygulamalarından biri doz optimizasyonudur. Uygun ilaç dozunun belirlenmesi, kemoterapi tedavisinin en karmaşık durumlarında biri olmaktadır. Bu durumun sebebi etkin bir tedavi ile toksik yan etkiler arasındaki etkili ve güvenli doz aralığı öok dar kalmaktadır. Mevcuttaki doz stratejileri basit kurallar çerçevesinde kalmaktadır. Bununla birlikte, bu yaklaşımlar ilacın metabolizmasında genetik farklılıklar ve tedaviye verilen bireysel yönde yanıtlar arasındaki farklılıkları tam olarak açıklayamamaktadır. Sonuç olarak, kişiselleştirilmiş doz optimizasyonu, tedavi sürecinde önemli bir konu haline gelmektedir. Sentetik hasta simülasyonları, araştırmacıların farklı dozlama stratejilerinin büyük sanal hasta gruplarında tedavi sonuçlarını nasıl etkilediğini analiz etmelerini sağlamaktadır. Çok sayıda simüle edilmiş hastada tedavi yanıtlarını değerlendirerek, hesaplamalı modeller, toksisite risklerini en aza indirirken tedavi yanıtının etkisini koruyan dozlama planlamaları belirlemektedir. Optimizasyon algoritmaları, en uygun tedavi sonuçlarını üreten doz stratejilerini otomatik olarak belirlemektedir (Miotto, R., 2016). Algoritmalar ve pekiştirmeli öğrenme yaklaşımları da dahil olmak üzere çeşitli hesaplamalı optimizasyon yöntemleri, kişiselleştirilmiş tedavi planlaması için araştırılmaktadır (Tan ve ark., 2022).

### 4.4 Yapay Zekâ Tabanlı Tahmin Modelleri

Yapay zekâ, sentetik hasta veri kümelerinin analizinde ve klinik karar destek sistemleri için tahmin modelleri geliştirmede çok önemli bir rol oynamaktadır. Makine öğrenme algoritmaları, hasta özellikleri ile tedavi sonuçları arasındaki ilişkileri belirleyerek, hastalık ilerlemesini veya tedavi yanıtını tahmin edebilen tahmin modellerinin geliştirilmesini sağlamaktadır. Bununla birlikte makine öğrenmesi teknikleri, sağlık verilerinin karmaşıklığı, parametre değişkenleri arasındaki doğrusal olmayan etkenler gibi durumları eğiterek daha gelişmiş modelleme yaklaşımlarını sunmaktadır. Bu modeller, büyük ölçekli veri kümelerini işleyebilmekte ve mevcuttaki istatistiksel modeller tarafından kolayca işlenemeyen hasta verilerindeki karmaşık kalıpları işlemekte ve sonuçlar çıkarabilmektedir. Sentetik hasta veri kümeleri üzerinde eğitildiğinde, derin öğrenme modelleri tedavi değişkenlerinin sonuçları nasıl etkilediğini öğrenebilmektedir (Obermeyer, Z., 2016). Güncel çalışmalarda sentetik sağlık hizmetlerindeki veri kümeleri üzerinde eğitilen yapay zekâ modelleri, gerçek klinik veriler üzerinde eğitilen modellere kıyasla benzer tahmin performansı elde edebildiğini göstermektedir. Bu bulgu, hasta gizliliğini korurken yapay zekâ tabanlı sağlık araştırmalarını ilerletmek için sentetik hasta modellemesinin potansiyelini vurgulamaktadır (Tucker vd., 2020).

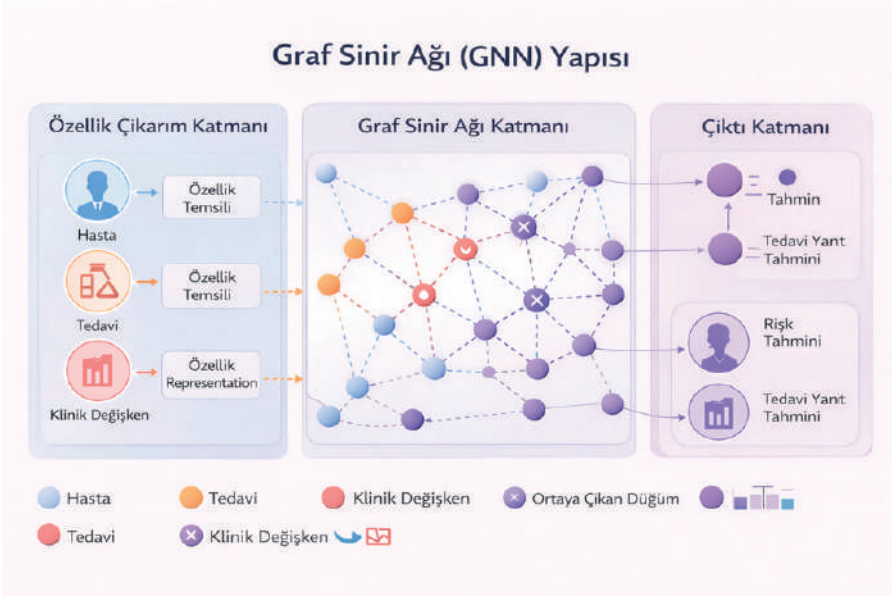
Yapılan çalışmalardaki derin öğrenme yaklaşımları, tahmine dayalı modelleme yeteneklerini geliştirmektedir. Derin sinir ağları, tıbbi görüntüler, karmaşık hasta kayıtları gibi sağlık verilerini otomatik olarak öğrenebilmektedir. Evrimsel sinir ağları (CNN) ve tekrarlayan sinir ağları (RNN) gibi mimariler, hastalık teşhisi ve hasta takibi de dahil olmak üzere çeşitli tıbbi tahmin görevlerinde başarıyla uygulanmaktadır (LeCun, Bengio ve Hinton, 2015). Grafik tabanlı öğrenme yöntemleri biyomedikal veri analizinde giderek daha fazla ilgi görmektedir. GNN mimarisi hastalar, tedaviler ve klinik değişkenler arasındaki karmaşık ilişkilerin grafik yapılar içinde temsil edilmesine olanak tanımaktadır. Bu yaklaşım, özellikle birbirine bağlı sağlık verilerini modellemek ve hastaların zaman içindeki gidişatını analiz etmek için kullanılmaktadır (Hamilton, Ying ve Leskovec, 2017). İlişkisel veri yapılarını tahmine dayalı modelleme algoritmalarıyla entegre ederek, grafik tabanlı yöntemler, gelişmiş klinik karar destek sistemleri geliştirmek için yeni fırsatlar sunmaktadır. Genel olarak, yapay zekâ tabanlı tahmine dayalı modelleme, sağlık sistemlerini veri odaklı ortamlara dönüştürmede kritik bir rol oynamaktadır. Bu modeller, erken hastalık tespiti, kişiselleştirilmiş tedavi planlaması ve risk tahmini sağlayarak doktorlar desteklemektedir. Hesaplama gücü ve veri erişilebilirliği artmaya devam ettikçe, tahmine dayalı modelleme tekniklerinin klinik karar verme süreçlerini iyileştirmede ve hassas tıp alanında ilerleme sağlamada daha da etkili hale gelmesi beklenmektedir (Rieke, N.,2020). Hasta modelleme ili ilgili süreç Şekil 1'de görsel olarak verilmektedir.



*Şekil 1. Yapay Zekâ Tabanlı Sentetik Hasta Modellemesi Süreci. (Şekil, bölümde sunulan kavramsal çerçeveyi göstermek amacıyla yazarlar tarafından yapay zekâ destekli bir görselleştirme aracı kullanılarak oluşturulmuştur).*

#### 4.5 Grafik Sinir Ağları ve Dijital İkiz Modellemesi

Grafik Sinir Ağları (GNN), ilişkisel ve yapılandırılmış verileri analiz etmek için tasarlanmış nispeten yeni bir derin öğrenme yöntemleri sınıfını temsil etmektedir. GNN modelleri veri kümelerini düğümler ve kenarlardan oluşan grafikler olarak temsil etmektedir (Rajkomar, A., vd., 2019). Sağlık uygulamalarında, grafik yapıları hastalar, klinik ölçümler veya zamansal gözlemler arasındaki ilişkileri temsil etmektedir. Tedavi sırasında ölçülen parametreler, bir grafik yapısı içindeki birbirine bağlı düğümler olarak temsil edilmektedir. Araştırmalar grafik tabanlı derin öğrenme modellerinin özellikle büyük ölçekli sağlık veri kümelerini analiz etmek ve hastalık ilerleme modellerini tahmin etmek için etkili olduğunu göstermektedir (Ahmedt-Aristizabal vd., 2022). Bu modeller, hastalık tahmini, tedavi sonucu tahmini ve hastalar arasındaki benzerliklerin analizi de dahil olmak üzere çeşitli biyomedikal uygulamalarda kullanılmaktadır. Lösemi araştırmalarında, grafik tabanlı modeller, WBC ve ANC gibi hematolojik parametreleri zamansal değişiklikleri analiz etmek için kullanılmıştır. Ardışık klinik gözlemlerdeki grafik düğümler modellenerek, GNN modelleri tedavi parametrelerinin zaman içinde tedavi parametrelerini nasıl etkilediğini öğrenebilmektedir. Bu gelişme, hastaya özgü tedavi yanıtlarını simüle eden dijital ikiz tabanlı klinik karar destek sistemlerinin geliştirilmesini sağlamaktadır. Dijital ikiz sistemleri, gerçek veya sentetik veri girdilerini kullanarak hastalık ilerlemesini ve tedavi yanıtını sürekli olarak simüle eden bireysel hastaların hesaplamalı modellerini temsil etmektedir. Bu sistemler, doktorların potansiyel tedavi planlarını klinik uygulamada uygulamadan önce sanal bir ortamda değerlendirmelerine olanak tanımaktadır. Sonuç olarak, GNN tabanlı dijital ikiz modelleri, kişiselleştirilmiş tıpta yeni nesil klinik karar destek sistemleri için umut vadeden bir yönü temsil etmektedir. Aşağıdaki Şekil 2 görselinde GNN yapısının tedavi sürecindeki yapısı gösterilmektedir.



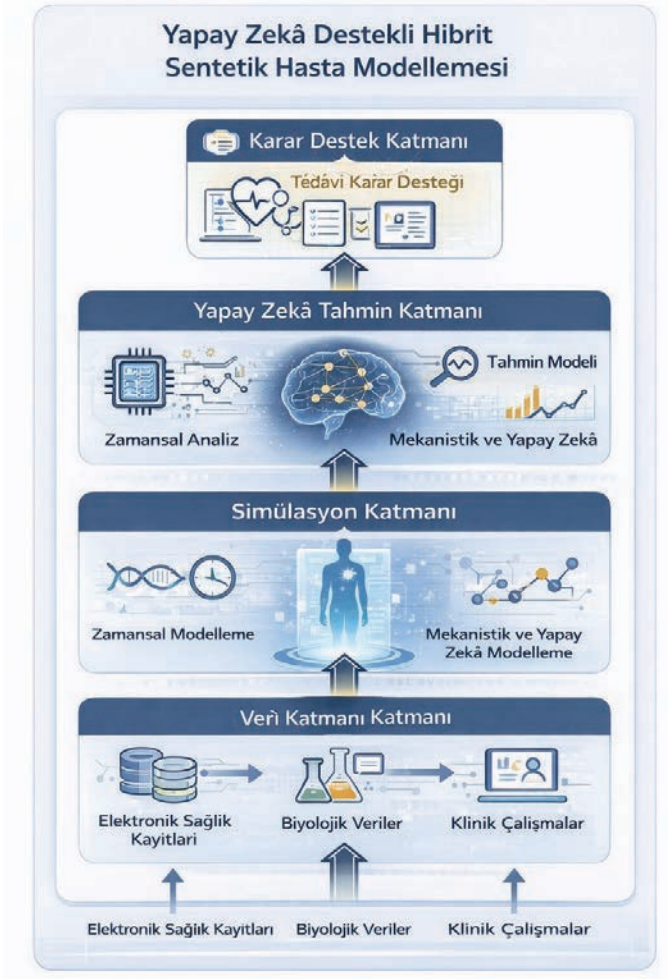
*Şekil 2. Grafik Sinir Ağı (GNN) mimarisi ve klinik veri tabanlı tahmin süreci. (Görsel yazarlar tarafından yapay zekâ destekli bir görselleştirme aracı kullanılarak hazırlanmıştır).*

#### 4.6 Hesaplamalı Hasta Modellerinin Yapay Zekâ ile Entegrasyonu

Hesaplamalı hasta modellerinin yapay zekâ ile entegrasyonu, modern klinik karar destek sistemlerinde önemli bir gelişmeyi temsil etmektedir. Hesaplamalı hasta modelleri, araştırmacıların biyolojik özelliklerinin matematiksel olarak tanımlayarak fizyolojik süreçleri ve tedavi yanıtlarını simüle etmelerini sağlamaktadır. Hasta modelleri yapay zekâ algoritmalarıyla birleştirildiğinde, bu modeller kişiselleştirilmiş tedavi planlamasını destekleyebilen tahmin sistemlerinin geliştirilmesini mümkün kılmaktadır. Lösemi tedavisi araştırmalarında, hesaplamalı hasta modelleri kemoterapi ilaçlarının farmakokinetik ve farmakodinamik davranışlarını simüle etmek için yaygın olarak kullanılmaktadır. Bu modeller tipik olarak ilaç emilimi, metabolizma ve biyolojik yanıt süreçlerini temsil ederken, vücut yüzey alanı, genetik farklılıklar ve fizyolojik değişkenlik gibi hastaya özgü özellikleri de içermektedir. Bu süreçleri simüle ederek, hesaplamalı modeller tedavi stratejilerinin zaman içinde hematolojik yanıtları nasıl etkilediğini anlamak için bir çerçeve sağlamaktadır.

Çalışmalar, kişiselleştirilmiş farmakokinetik ve farmakodinamik modellerin, kemoterapi ilaçlarının WBC ve ANC gibi hematolojik parametreler üzerindeki

etkilerini başarıyla simüle edebildiğini göstermektedir. Özellikle, çocukluk çağı akut lenfoblastik lösemisinin tedavisi için geliştirilen hesaplamalı modeller, genetik özellikler, beslenme durumu, D vitamini seviyeleri ve fiziksel aktivite gibi hastaya özgü değişkenlerin dahil edilmesinin, tedavi yanıtı simülasyonlarının doğruluğunu önemli ölçüde artırabileceğini göstermiştir (Köse, vd., 2025). Bu hesaplamalı hasta modelleri, sentetik hasta veri kümeleri oluşturmak için önemli bir temel sağlamaktadır. Farklı hasta koşulları altında ilaç metabolizması ve hematolojik özelliklerin simülasyonu ile çok sayıda sanal hasta profili üretilebilmektedir. Aşağıdaki şekilde önerilen mimari, klinik karar destek sistemlerinde sentetik hasta modellemesi için yapay zekâ odaklı hibrit bir çerçeveyi göstermektedir. Süreç, elektronik sağlık kayıtları, biyolojik veri kümeleri ve klinik çalışma verileri gibi birden fazla sağlık verisi kaynağını entegre eden veri katmanıyla başlamaktadır (Shortliffe vd., 2018). Bu heterojen veri kaynakları daha sonra simülasyon katmanında işlenmektedir. Modelleme ve zamansal analiz teknikleri kullanılarak hasta dinamikleri simüle edilmekte ve sanal hasta temsilleri oluşturulmaktadır. Elde edilen simüle edilmiş hasta yörüngeleri daha sonra yapay zekâ tahmin katmanında analiz edilmektedir. Yapay zekâ modelleri, tahmine dayalı yorumlar üretmek için oluşturulan veri kümelerindeki zamansal kalıpları ve ilişkileri öğrenmektedir (Yu vd., 2018). Sistem, model tahminlerini yorumlanabilir klinik içgörülere dönüştüren ve sağlık uzmanlarının tedavi stratejilerini ve hasta yönetimi kararlarını değerlendirmede yardımcı olabilecek karar destek katmanı aracılığıyla klinik olarak ilgili öneriler sunmaktadır.



*Şekil 3. Yapay zekâ destekli hibrit sentetik hasta modelleme mimarisi (Görsel, bölümde sunulan kavramsal çerçeveyi göstermek amacıyla yazarlar tarafından yapay zekâ destekli bir görselleştirme aracı kullanılarak hazırlanmıştır)*

### 5. Yapay Zekaya Dayalı Sentetik Hasta Modellemesindeki Zorluklar ve Gelecek Çalışmalar

Yapay zekaya dayalı sentetik hasta modellemesi, klinik karar destek sistemleri için önemli avantajlar sunmasının yanında birçok önemli zorluk devam etmektedir. En kritik konulardan biri, sentetik hasta verilerinin gerçekliği ve güvenilirliğidir. Yapay zeka yöntemleri, gerçek klinik verileri taklit eden büyük veri kümeleri üretebilse de, bu veri kümelerinin biyolojik süreçleri ve klinik değişkenliği doğru bir şekilde temsil etmesini sağlamak

büyük bir zorluk olmaya devam etmektedir. Sentetik hasta verileri gerçek fizyolojik özellikleri yansıtmadığında, bu veri kümeleri üzerinde eğitilmiş tahmin modelleri güvenilir olmayan klinik tahminler üretebilmektedir. Bir diğer önemli zorluk ise model doğrulamasıyla ilgili olmaktadır. Sentetik hasta simülasyonlarının güvenilirliğini ve klinik uygulanabilirliğini sağlamak için gerçek klinik gözlemlere karşı doğrulanması gerekmektedir. Bu süreç, yüksek kaliteli klinik veri kümelerine erişimi ve sağlık uzmanları ve veri bilimcileri arasında yakın işbirliğini gerektirmektedir. Titiz bir doğrulama olmadan, hesaplamalı tahminler klinik güvenilirlikten yoksun olabilmekte ve gerçek dünya tıbbi uygulamaları için uygun olmayabilmektedir.

Veri karmaşıklığı, yapay zekâ destekli klinik karar destek sistemlerinin geliştirilmesinde de önemli bir engel teşkil etmektedir. Modern sağlık verileri, elektronik sağlık kayıtları, laboratuvar ölçümleri ve yaşam tarzıyla ilgili değişkenler de dahil olmak üzere birçok kaynaktan gelmektedir. Bu çeşitli veri kaynaklarını birleşik bir modelleme çerçevesine entegre etmek teknik olarak zorlu olmaya devam etmekte ve gelişmiş veri entegrasyon stratejileri gerektirmektedir. Ayrıca yorumlanabilirlik sağlık alanındaki yapay zekâ modelleri için önemli bir endişe kaynağı olmaya devam etmektedir. Birçok derin öğrenme algoritması kara kutu sistemleri olarak işlev görmektedir. Bu durum doktorların tahminlerin nasıl oluşturulduğunu anlamasını zorlaştırmaktadır. Klinik karar destek sistemlerinin yaygın olarak benimsenmesi için, gelecekteki araştırmalar, sentetik hasta modellemesinde kullanılan yapay zekâ modellerinin şeffaflığını ve yorumlanabilirliğini iyileştirmeye odaklanmalıdır. Yapay zekâ tabanlı sentetik hasta modellemesi alanındaki gelecekteki araştırmaların hem sanal hasta gruplarında gerçekçiliğini hem de hesaplamalı modellerin tahmin yeteneklerini geliştirmeye odaklanması beklenmektedir. Gelecek araştırmalarda sağlık verilerindeki karmaşık ilişkileri modelleyebilen daha gelişmiş yapay zekâ mimarilerinin geliştirilmesi beklenmektedir. Özellikle, zamansal ve ilişkisel veri yapıları için tasarlanmış derin öğrenme yöntemleri, hastalık ilerlemesinin ve tedavi sonuçlarının tahminini iyileştirebilmektedir. Bu modeller, klinik parametrelerin daha doğru tahmin edilmesini sağlayabilmekte ve klinik karar destek sistemlerinin performansını arttırabilmektedir.

## 6. Sonuç

Yapay zekâ, gelişmiş veri analizi, tahmine dayalı modelleme ve karar destek mekanizmalarını mümkün kılarak modern sağlık hizmetlerinde giderek daha etkili bir araç haline gelmiştir. Kanser tedavisi gibi karmaşık tıbbi alanlarda, klinik karar verme genellikle hasta özellikleri, tedavi stratejileri, biyolojik yanıtlar ve hastalık ilerleme modelleri de dahil olmak üzere birbirine bağlı birçok faktörün değerlendirilmesini içerir. Bu nedenle, hesaplama yöntemlerinin

sağlık arařtırmalarına giderek daha fazla entegre edilmesi, klinik karar verme süreçlerinin doğruluğunu ve verimliliğini artırmak için yeni fırsatlar yaratmıştır. Bu bağlamda, sentetik hasta modellemesi, arařtırmacıların gerçekçi hasta popülasyonlarını simüle etmelerine ve kontrollü bir hesaplama ortamında tedavi dinamiklerini analiz etmelerine olanak tanıyan umut vadeden bir yaklaşım olarak ortaya çıkmıştır. Sentetik hasta modellemesinin en önemli avantajlarından biri, klinik veri kullanılabilirliğiyle ilgili sınırlamaları ele alabilmesidir. Tıbbi veri kümeleri, gizlilik düzenlemeleri, etik hususlar ve kurumsal veri paylaşım politikaları nedeniyle sıklıkla kısıtlanmaktadır. Bu sınırlamalar, eğitim için büyük ve çeşitli veri kümeleri gerektiren güvenilir yapay zekâ modellerinin geliştirilmesini engelleyebilir. Sentetik hasta üretimi, gerçek hasta popülasyonlarının istatistiksel özelliklerini ve klinik özelliklerini taklit eden yapay veri kümeleri üreterek bu zorluğa etkili bir çözüm sunmaktadır. Sonuç olarak, arařtırmacılar hassas hasta verilerine doğrudan güvenmeden farklı tedavi senaryolarını inceleyebilir, potansiyel klinik sonuçları değerlendirebilir ve tahmin modelleri geliştirebilirler. Sentetik hasta modellemesinin bir diğer önemli katkısı ise hasta heterojenliğini yakalayabilme yeteneğidir. Lösemi gibi hastalıklarda, genetik farklılıklar, fizyolojik koşullar, çevresel faktörler ve tedavi toleransı düzeyleri nedeniyle bireyler arasında tedavi yanıtları önemli ölçüde değişebilir. Geleneksel istatistiksel yaklaşımlar bu karmaşıklığı tam olarak yakalayamayabilir. Buna karşılık, sentetik hasta simülasyonları, arařtırmacıların hastaya özgü değişkenliği hesaplama modellerine dahil etmelerine ve bu farklılıkların tedavi sonuçlarını nasıl etkilediğini analiz etmelerine olanak tanır. Ayrıca, sentetik hasta modellemesinin yapay zeka teknikleriyle entegrasyonu, gelişmiş klinik karar destek sistemlerinin geliştirilmesi için yeni olanaklar sunmaktadır. Sentetik veri kümeleri üzerinde eğitilmiş makine öğrenimi modelleri, hasta özellikleri ve tedavi sonuçları arasındaki karmaşık ilişkileri belirleyerek, klinisyenlerin tedavi stratejilerini değerlendirmelerine ve hasta yanıtlarını tahmin etmelerine yardımcı olabilir. Genel olarak, yapay zeka tabanlı sentetik hasta modellemesi, veri odaklı klinik karar destek sistemlerini geliştirmek ve daha kişiselleştirilmiş sağlık çözümlerinin geliştirilmesini desteklemek için güçlü bir çerçeve oluşturmaktadır.

### **Teşekkür**

Bu çalışma, Türkiye Bilimsel ve Teknolojik Arařtırma Kurumu (TÜBİTAK) tarafından desteklenen “STING: Dijital İkiz Yönelimli Derin Öğrenme ile Çocukluk Çağı Akut Lösemisi İçin İlaç Yeniden Konumlandırma Karar Destek Sistemi Geliştirilmesi” başlıklı proje ile ilişkilidir (Proje No: 123E383).

Ayrıca bu çalışma, Öğr. Gör. Emine Betül Sürücü'nün Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı'nda

(Prof. Dr. Utku Köse danışmanlığında) yürütmekte olduğu “Akciğer Kanseri Tedavisinde Derin Öğrenme Tabanlı İlaç Yeniden Konumlandırma ve Sentetik Hasta Modelleme Destekli Değerlendirme Sistemi” başlıklı doktora tezi ile ilişkilidir.

## Kaynaklar

- Ahmedt-Aristizabal, D., Armin, M. A., Denman, S., Fookes, C., & Petersson, L. (2022). A survey on graph-based deep learning for computational histopathology. *Computerized Medical Imaging and Graphics*, 95, 102027.
- Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., & Escobar, G. (2014). Big data in health care: Using analytics to identify and manage high-risk and high-cost patients. *Health Affairs*, 33(7), 1123–1131.
- Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. *JAMA*, 319(13), 1317–1318.
- Chen, R. J., et al. (2021). Synthetic data in machine learning for medicine and healthcare. *Nature Biomedical Engineering*, 5, 493–497.
- Choi, E., Schuetz, A., Stewart, W. F., & Sun, J. (2017). Using recurrent neural network models for early detection of heart failure onset. *Journal of the American Medical Informatics Association*, 24(2), 361–370.
- Dash, S., Shakyawar, S. K., Sharma, M., & Kaushik, S. (2019). Big data in healthcare: Management, analysis and future prospects. *Journal of Big Data*, 6(1), 54.
- Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future Healthcare Journal*, 6(2), 94–98.
- Esteva, A., et al. (2019). A guide to deep learning in healthcare. *Nature Medicine*, 25, 24–29.
- Evans, W. E., et al. (2018). Precision medicine in pediatric acute lymphoblastic leukemia. *Journal of Clinical Oncology*, 36(21), 2125–2133.
- Goodfellow, I., et al. (2014). Generative adversarial networks. *Advances in Neural Information Processing Systems*.
- Hamilton, W. L., Ying, Z., & Leskovec, J. (2017). Inductive representation learning on large graphs. *NeurIPS*.
- Johnson, A. E. W., et al. (2016). MIMIC-III, a freely accessible critical care database. *Scientific Data*, 3.
- Kingma, D. P., & Welling, M. (2014). Auto-encoding variational Bayes. *International Conference on Learning Representations*.
- Kose, U., Tükel, G. O., Surucu, E. B., Karatas, M. E., & Uysal, I. (2025, October). Graph Neural Networks-Based Digital Twin Modeling of WBC and ANC Dynamics for Personalized Time-Series Prediction. In *2025 Medical Technologies Congress (TIPTEKNO)* (pp. 1–4). IEEE.
- Krittanawong, C., Johnson, K. W., Rosenson, R. S., Wang, Z., Aydar, M., & Narayan, S. M. (2017). Deep learning for cardiovascular medicine: A practical primer. *European Heart Journal*, 40(25), 2058–2073.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521, 436–444.

- Miotto, R., Li, L., Kidd, B. A., & Dudley, J. T. (2016). Deep patient: An unsupervised representation to predict the future of patients from electronic health records. *Scientific Reports*, 6, 26094.
- Mould, D. R., & Upton, R. N. (2012). Basic concepts in population modeling. *CPT: Pharmacometrics & Systems Pharmacology*.
- Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future—big data, machine learning, and clinical medicine. *New England Journal of Medicine*, 375, 1216–1219.
- Rajkumar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine*, 380, 1347–1358.
- Rieke, N., et al. (2020). The future of digital health with federated learning. *NPJ Digital Medicine*, 3, 119.
- Shickel, B., Tighe, P., Bihorac, A., & Rashidi, P. (2018). Deep EHR: A survey of recent advances in deep learning techniques for electronic health record analysis. *IEEE Journal of Biomedical and Health Informatics*, 22(5), 1589–1604.
- Shortliffe, E. H., & Sepúlveda, M. J. (2018). Clinical decision support in the era of artificial intelligence. *JAMA*, 320(21), 2199–2200.
- Sutton, R. T., et al. (2020). An overview of clinical decision support systems. *NPJ Digital Medicine*.
- Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 25, 44–56.
- Tucker, A., Wang, Z., Rotalinti, Y., & Myles, P. (2020). Generating high-fidelity synthetic patient data for assessing machine learning healthcare software. *NPJ Digital Medicine*, 3(1), 147.
- Viceconti, M., et al. (2021). Digital twins in medicine. *Nature Reviews Biomedical Engineering*.
- Yu, K. H., Beam, A. L., & Kohane, I. S. (2018). Artificial intelligence in healthcare. *Nature Biomedical Engineering*, 2, 719–731.



## Bilgisayar Bilimleri ve Mühendisliğinde Dönüşüm: Üretken Yapay Zekâ, Güvenli Sistemler ve Ölçeklenebilir Hesaplama

Furkan Atlan<sup>1</sup>

### Özet

Bu çalışma, 2010-2026 döneminde bilgisayar bilimleri ve mühendisliğinde yaşanan dönüşümü; üretken yapay zekâ, güvenli yapay zekâ sistemleri ve ölçeklenebilir hesaplama altyapıları ekseninde bütüncül biçimde incelemektedir. İlk olarak, veri yoğun dünyada artan hesaplama gereksinimleri, donanım sınırları ve yazılım merkezli optimizasyon yaklaşımları ele alınmakta; sağlık, enerji, finans ve savunma gibi alanlarda ortaya çıkan disiplinler arası etkiler tartışılmaktadır. Ardından üretken yapay zekâ ve temel modeller kapsamında büyük dil modelleri, çok modlu sistemler, Edge AI, hafif model mimarileri ve açıklanabilir yapay zekâ yaklaşımları değerlendirilmekte; model güvenilirliği ve halüsinasyon problemi, çağdaş yapay zekâ sistemlerinin temel sınırlılıkları arasında konumlandırılmaktadır. Bölümün güvenlik odaklı kısmında adversarial saldırılar, veri/model zehirlenme, yapay zekâ ile siber güvenlik kesişimi, düzenleyici çerçeveler ve algoritmik adalet konuları ele alınmaktadır. Son olarak GPU/TPU hızlandırılmalı sistemler, bulut ve hibrit mimariler, HPC kümeleri, Slurm tabanlı kaynak yönetimi, edge-cloud iş bölümü ve enerji verimli hesaplama başlıkları üzerinden modern hesaplama altyapılarının teknik ve stratejik boyutları tartışılmaktadır. Genel olarak bölüm, yapay zekâ çağında güvenilir, açıklanabilir, sürdürülebilir ve ölçeklenebilir bilişim sistemlerinin geliştirilmesinin bilgisayar bilimlerinin temel yönelimlerinden biri hâline geldiğini ortaya koymaktadır.

1 Dr. Öğr. Üyesi, Burdur Mehmet Akif Ersoy Üniversitesi Yönetim Bilişim Sistemleri Bölümü, fatlan@mehmetakif.edu.tr , <https://orcid.org/0000-0003-1602-1941>

## 1. Bilgisayar Bilimlerinde Paradigma Değişimi

### 1.1. 2010-2026 Yılları Arasındaki Teknolojik Kırılmalar

2010-2026 yılları arası dönem, bilgisayar bilimlerinde hem kuramsal yaklaşımların hem de uygulama alanlarının önemli ölçüde yeniden şekillendiği bir zaman dilimi olarak değerlendirilmektedir. Bu süreçte veri üretim hızının artması, hesaplama altyapılarının gelişmesi ve yapay zekâ (AI) algoritmalarındaki ilerlemeler, bilgisayar bilimlerini yalnızca teknik bir disiplin olmaktan çıkararak çok sayıda alanla kesişen disiplinler arası bir araştırma alanına dönüştürmüştür. Özellikle büyük veri (Big Data) ekosisteminin ortaya çıkışı, bulut bilişim altyapılarının yaygınlaşması ve dağıtık sistem mimarilerinin olgunlaşması, yüksek hacimli verilerin işlenmesini mümkün kılmış ve veri odaklı karar verme süreçlerinin bilimsel ve endüstriyel alanlarda yaygınlaşmasına katkı sağlamıştır (Bibri, 2019).

Bu dönemde dikkat çeken en önemli teknolojik kırılmalardan biri AI ve derin öğrenme alanında yaşanmıştır. Özellikle 2012 yılında derin sinir ağlarının görüntü tanıma problemlerinde gösterdiği performans artışı, AI araştırmalarında yeni bir paradigma oluşturmuştur. Takip eden yıllarda konvolüsyonel sinir ağları (CNN), tekrarlayan sinir ağları (RNN) ve transformer tabanlı mimariler gibi modellerin geliştirilmesi; doğal dil işleme, bilgisayarla görme ve konuşma tanıma gibi alanlarda önemli ilerlemelerin gerçekleşmesini sağlamıştır (Chinnaiyan vd., 2025). 2020'li yıllarla birlikte ise büyük ölçekli dil modelleri (Large Language Models-LLM) ve üretken AI (Generative AI) sistemleri, bilgi üretimi, içerik oluşturma ve insan-makine etkileşimi gibi alanlarda yeni araştırma ve uygulama fırsatları ortaya çıkarmıştır (Zhou vd., 2024).

Donanım tarafında yaşanan gelişmeler de bu dönüşümü hızlandıran önemli faktörlerden biri olmuştur. Grafik işlem birimleri (GPU), tensör işlem birimleri (TPU) ve özel amaçlı AI hızlandırıcılarının geliştirilmesi, büyük ölçekli AI modellerinin eğitilmesini mümkün kılmıştır. Bunun yanında yüksek performanslı hesaplama (High Performance Computing - HPC) altyapılarının ve paralel işlem mimarilerinin gelişmesi, bilimsel hesaplamalarda ve veri yoğun uygulamalarda önemli performans kazanımları sağlamıştır (Deelman vd., 2025). Ayrıca uç bilişim (Edge Computing) ve nesnelerin interneti (IoT) teknolojileri sayesinde hesaplama süreçleri merkezi sistemlerden dağıtık ve akıllı cihazlara doğru kaymaya başlamıştır.

2010-2026 döneminde bilgisayar bilimlerindeki dönüşüm yalnızca teknik gelişmelerle sınırlı kalmamış, aynı zamanda güvenlik, etik ve düzenleyici çerçeveler açısından da yeni tartışmaları beraberinde getirmiştir. AI sistemlerinin

karar verme süreçlerindeki rolünün artması; açıklanabilir AI (Explainable AI), algoritmik adalet ve veri güvenliği gibi konuların önemini artırmıştır (Altukhi ve Pradhan, 2025). Bunun yanı sıra siber güvenlik tehditlerinin karmaşıklaşması, kuantum hesaplama araştırmalarının hız kazanması ve sürdürülebilir bilişim yaklaşımlarının gündeme gelmesi, bilgisayar bilimlerinin gelecekteki araştırma yönelimlerini belirleyen temel unsurlar arasında yer almaktadır. Bu bağlamda 2010-2026 yılları arası dönem, bilgisayar bilimlerinde paradigma değişimlerinin yaşandığı ve dijital dönüşümün küresel ölçekte hız kazandığı kritik bir kırılma noktası olarak değerlendirilmektedir.

## 1.2. Donanım Sınırları ve Yazılım Merkezli Optimizasyon

Bilgisayar bilimlerinde son on beş yılda yaşanan en önemli dönüşümlerden biri, donanım ölçeklenmesinin fiziksel sınırlarına yaklaşılmasıyla birlikte performans artışının giderek yazılım temelli optimizasyonlara dayanır hâle gelmesidir. Moore Yasası'nın yavaşlaması, transistör boyutlarının küçülmesiyle ortaya çıkan enerji tüketimi, ısınma ve güvenilirlik sorunları, işlemci mimarilerinin yalnızca donanım geliştirmeleriyle performans kazanmasını zorlaştırmıştır. Özellikle modern çok çekirdekli sistemlerde artan entegrasyon yoğunluğu, işlemcileri geçici ve kalıcı hatalara karşı daha hassas hâle getirmiştir. Bu durum, sistem doğruluğunu korumak amacıyla mimari düzeyde hata toleransı ve yazılım destekli güvenilirlik mekanizmalarının geliştirilmesini gerekli kılmıştır (Keller, 2025).

Donanım sınırlamalarının belirginleşmesiyle birlikte araştırmacılar, performans ve güvenilirlik hedeflerine ulaşmak için donanım-yazılım ortak tasarımına dayalı optimizasyon stratejilerine yönelmiştir. Çok çekirdekli işlemcilerde paralel yürütme yetenekleri önemli performans avantajları sunsa da bu mimariler, yazılım tarafından etkin şekilde yönetilmediğinde beklenen verimliliği sağlayamamaktadır. Bu nedenle görev zamanlama, hata tespiti, redundant çoklu iş parçacığı (redundant multithreading) ve geri alma (rollback) mekanizmaları gibi yazılım tabanlı teknikler, sistem güvenilirliğini artırmak için yaygın olarak kullanılmaktadır. Yazılım seviyesinde gerçekleştirilen bu tür optimizasyonlar, donanım maliyetini artırmadan hata toleransı sağlayabilmeleri nedeniyle özellikle modern işlemci mimarilerinde önemli bir araştırma alanı hâline gelmiştir (Keller, 2025; De Souza, 2025).

Benzer şekilde, günümüz dijital ürün ve sistem mimarilerinde donanım ile yazılım bileşenlerinin ayrıştırılması (hardware-software decoupling) önemli bir tasarım yaklaşımı olarak ortaya çıkmıştır. Yazılım tanımlı ürünler (Software-Defined Products) yaklaşımında sistem fonksiyonlarının büyük bölümü yazılım katmanında tanımlanmakta ve donanım yalnızca temel altyapıyı

sağlamaktadır (Barwasser vd., 2024). Bu yaklaşım sayesinde ürün özellikleri, donanım değişikliğine ihtiyaç duyulmadan yazılım güncellemeleri aracılığıyla geliştirilebilmekte ve sistemler daha esnek hâle getirilebilmektedir. Özellikle otomotiv, tüketici elektroniği ve tıbbi cihazlar gibi alanlarda bu yaklaşımın yaygınlaşması, performans iyileştirmelerinin giderek yazılım merkezli optimizasyonlara dayandığını göstermektedir (Lee vd., 2025).

Sonuç olarak, modern bilgisayar sistemlerinin gelişimi yalnızca donanım performansının artırılmasıyla açıklanamayacak kadar karmaşık bir hâl almıştır. Enerji verimliliği, güvenilirlik, ölçeklenebilirlik ve maliyet gibi faktörler, sistem tasarımında yazılımın rolünü giderek daha kritik hâle getirmiştir. Bu nedenle günümüz bilgisayar mimarilerinde donanım altyapısı ile yazılım optimizasyonlarının birlikte ele alındığı bütünlük tasarım yaklaşımları ön plana çıkmaktadır (Lee vd., 2025). Gelecekte AI destekli sistem tasarımı, donanım soyutlama katmanları ve modüler yazılım mimarileri gibi yaklaşımların, donanım sınırlamalarını dengeleyen temel optimizasyon araçları olarak daha da önem kazanması beklenmektedir.

### **1.3. Veri Yoğun Dünyada Hesaplama Gereksinimleri**

Günümüzde üretilen veri miktarı tarihsel olarak görülmemiş bir hızla artmaktadır. İnternet, mobil cihazlar, sensör ağları ve IoT sistemleri tarafından üretilen veri akışı; metin, görüntü, video ve sensör verileri gibi farklı türlerde büyük veri ekosistemlerini ortaya çıkarmıştır. Bu gelişmeler, geleneksel bilgi işlem mimarilerinin kapasitesini zorlamakta ve veri yoğun (data-intensive) hesaplama paradigmasının önemini artırmaktadır. Veri yoğun hesaplama; büyük veri kümelerinin toplanması, depolanması, işlenmesi ve analiz edilmesini kapsayan çok aşamalı bir süreç olup özellikle makine öğrenmesi, grafik analizi ve veri madenciliği gibi yöntemler bu süreçte temel rol oynamaktadır. Bu nedenle modern bilgi işlem sistemlerinin yalnızca hesaplama gücü değil, aynı zamanda veri erişimi ve veri işleme kapasitesi açısından da ölçeklenebilir olması gerekmektedir (Akarvardar ve Wong, 2023).

Büyük veri çağında hesaplama gereksinimlerinin artmasının en önemli nedenlerinden biri, veri hacminin (volume), veri üretim hızının (velocity) ve veri çeşitliliğinin (variety) sürekli büyümesidir. Örneğin küresel veri evreninin 2020-2025 yılları arasında yaklaşık 3,5 kat artacağı öngörülmektedir (Evans vd., 2024). Bu hızlı büyüme, veri analizi süreçlerinin yalnızca depolama ve erişim açısından değil, aynı zamanda yüksek bant genişliği ve paralel işlem kapasitesi açısından da yeni altyapılar gerektirmesine yol açmaktadır. Veri yoğun uygulamalar genellikle milyonlarca basit işlemin eş zamanlı olarak yürütülmesini gerektirdiğinden GPU, FPGA ve özel amaçlı hızlandırıcılar gibi paralel mimariye dayalı donanımlar bu alanda önemli bir rol oynamaktadır.

Geleneksel süper bilgisayar sistemleri uzun süre yalnızca hesaplama yoğun (compute-intensive) bilimsel problemlerin çözümüne odaklanmış olsa da günümüzde veri yoğun uygulamalar HPC altyapılarının temel kullanım alanlarından biri hâline gelmiştir. Özellikle iklim modelleme, genom analizi, mühendislik tasarımı ve AI uygulamaları gibi alanlarda büyük veri analitiği ile HPC sistemlerinin entegrasyonu kritik bir ihtiyaç hâline gelmiştir. Bu entegrasyon sayesinde çok büyük veri kümeleri paralel işlem teknikleri kullanılarak kısa sürede analiz edilebilmekte ve karmaşık problemlere yönelik yeni çözümler geliştirilebilmektedir (Pyzer-Knapp vd., 2022).

Son yıllarda veri yoğun uygulamaların yaygınlaşmasıyla birlikte süper bilgisayar mimarilerinin tasarım hedefleri de değişmeye başlamıştır. Güncel HPC sistemleri yalnızca işlem gücünü artırmayı değil, aynı zamanda veri işleme kapasitesini, depolama altyapısını ve veri aktarım hızını da optimize etmeyi amaçlamaktadır (Navaux vd., 2023). Büyük veri analitiği ve AI uygulamalarının HPC ortamlarında yaygınlaşması, veri işleme süreçlerinin sistem performansındaki en kritik darboğazlardan biri hâline gelmesine neden olmuştur. Bu nedenle modern süper bilgisayar mimarileri; yüksek bant genişliğine sahip depolama sistemleri, paralel veri işleme mekanizmaları ve veri merkezleri ile entegre çalışan veri-yoğun altyapılar geliştirmeye yönelmektedir.

#### **1.4. Disiplinler Arası Dönüşüm (Sağlık, Enerji, Finans, Savunma)**

2010 sonrasında bilgisayar bilimlerinde yaşanan teknolojik gelişmeler yalnızca bilişim alanını değil, aynı zamanda farklı sektörleri de derinden etkileyen disiplinler arası bir dönüşümü beraberinde getirmiştir. Büyük veri analitiği, AI, HPC ve bulut bilişim gibi teknolojiler; sağlık, enerji, finans ve savunma gibi kritik sektörlerde karar verme süreçlerini yeniden şekillendirmiştir. Bu dönüşüm, veri odaklı yaklaşımların farklı disiplinlerle entegrasyonunu mümkün kılarak daha karmaşık problemlerin çözülebilmesine olanak sağlamaktadır (Egger ve Yu, 2022). Böylece bilgisayar bilimleri, yalnızca teknik bir alan olmaktan çıkarak çok disiplinli araştırmaların merkezinde yer alan stratejik bir bilim dalı hâline gelmiştir.

Sağlık alanında bu dönüşüm özellikle AI destekli tanı sistemleri, tıbbi görüntü analizi ve kişiselleştirilmiş tıp uygulamaları ile belirginleşmiştir. Derin öğrenme tabanlı algoritmalar, radyoloji görüntülerinde tümör tespiti, patoloji görüntülerinde hücresel yapıların analizi ve genom verilerinin yorumlanması gibi alanlarda yüksek doğruluk oranları sağlayarak klinik karar destek sistemlerinin gelişimine katkı sunmaktadır (Thirunavukarasu ve Kotej, 2024). Bunun yanında büyük veri analitiği sayesinde hastane bilgi sistemlerinden, biyomedikal sensörlerden ve genomik çalışmalardan elde edilen veriler entegre

edilerek hastalıkların erken teşhisi ve tedavi süreçlerinin optimize edilmesi mümkün hâle gelmiştir.

Enerji sektöründe ise bilgisayar bilimlerinin sunduğu veri analitiği ve AI teknikleri, enerji üretim ve dağıtım süreçlerinin daha verimli hâle getirilmesinde önemli rol oynamaktadır. Akıllı şebekeler (smart grids), sensör ağları ve IoT tabanlı enerji izleme sistemleri sayesinde enerji tüketimi gerçek zamanlı olarak analiz edilebilmekte ve üretim planlaması daha etkin biçimde yapılabilmektedir. Özellikle yenilenebilir enerji kaynaklarının entegrasyonu, enerji talep tahmini ve şebeke optimizasyonu gibi konularda makine öğrenmesi algoritmaları önemli avantajlar sağlamaktadır (Maheshwari vd., 2022). Bu sayede enerji sistemleri daha sürdürülebilir, güvenilir ve verimli bir yapıya kavuşmaktadır.

Finans ve savunma sektörleri de bilgisayar bilimlerindeki ilerlemelerden önemli ölçüde etkilenmiştir. Finans alanında algoritmik ticaret sistemleri, dolandırıcılık tespit mekanizmaları ve risk analizi modelleri büyük veri ve makine öğrenmesi teknikleri ile geliştirilmektedir (Kumar vd., 2025). Bu sistemler milyonlarca finansal işlemi gerçek zamanlı olarak analiz ederek piyasa davranışlarını modelleyebilmekte ve yatırım kararlarını destekleyebilmektedir. Savunma alanında ise büyük veri analitiği, siber güvenlik, otonom sistemler ve karar destek platformları modern askeri operasyonların önemli bileşenleri hâline gelmiştir. Özellikle AI destekli gözetleme sistemleri, tehdit analizi ve stratejik planlama araçları savunma teknolojilerinde yeni bir paradigma oluşturmuştur (Weber vd., 2024). Bu gelişmeler, bilgisayar bilimlerinin farklı disiplinlerle bütünleşerek küresel ölçekte stratejik bir dönüşüm yarattığını göstermektedir.

Savunma sanayi alanında bilgisayar bilimleri ve özellikle AI teknolojileri, modern askeri stratejilerin önemli bileşenlerinden biri hâline gelmiştir. AI ve makine öğrenmesi tabanlı sistemler; istihbarat toplama, gözetleme, keşif (ISR), siber güvenlik ve hedef tespiti gibi kritik askeri faaliyetlerde kullanılmaktadır. Bu teknolojiler sayesinde büyük veri kümeleri hızlı biçimde analiz edilerek savaş alanındaki durumsal farkındalık artırılmakta ve komuta kontrol süreçlerinde daha hızlı ve doğru kararlar alınabilmektedir. Özellikle otonom sistemler, insansız hava araçları ve akıllı savunma platformları gibi teknolojiler askeri operasyonların etkinliğini artıran önemli araçlar olarak öne çıkmaktadır (Khan vd., 2021).

Günümüzde ABD, Çin ve Rusya gibi teknolojik açıdan gelişmiş ülkeler savunma sistemlerine AI tabanlı yetenekler entegre etmeye yönelik yoğun yatırımlar yapmaktadır. Bu çerçevede geliştirilen otonom silah sistemleri, insan-makine iş birliği ile çalışan karar destek mekanizmaları ve ağ tabanlı savunma platformları, modern savaş doktrinlerinde giderek daha fazla yer almaktadır. Ayrıca AI destekli sistemler, yüksek hızda veri işleyebilme kabiliyeti sayesinde

savaş alanında insanın reaksiyon süresinin ötesinde karar alma süreçlerine katkı sağlayabilmektedir (Khan vd., 2021; Weng, 2024). Bu durum, gelecekte askeri güç dengesinin yalnızca geleneksel silah sistemleriyle değil, aynı zamanda gelişmiş yazılım ve algoritmik yeteneklerle de belirleneceğini göstermektedir.

## 2. Üretken AI ve Temel Modeller

Üretken AI ve temel modeller (Foundation Models), son yıllarda AI araştırmalarında ortaya çıkan en önemli paradigmalardan birini temsil etmektedir. Üretken AI; metin, görüntü, ses ve video gibi yeni içerikler üretebilen ve büyük veri kümeleri üzerinde eğitilmiş derin öğrenme modellerine dayanan sistemleri ifade eder. Bu sistemler, özellikle transformer tabanlı mimariler sayesinde dil üretimi, görüntü sentezi, kod üretimi ve çok modlu içerik oluşturma gibi görevlerde insan benzeri çıktılar üretebilmektedir. Temel modeller ise çok büyük ölçekli veri setleri üzerinde önceden eğitilmiş ve farklı görevler için yeniden uyarlanabilen genel amaçlı AI modelleridir. Bu modeller, transfer öğrenme ve ince ayar (fine-tuning) yöntemleri sayesinde doğal dil işleme, bilgisayarlı görü ve konuşma işleme gibi birçok farklı uygulama alanında kullanılabilir (Fui-Hoon Nah vd., 2023; Waqas vd., 2023). Günümüzde üretken AI sistemlerinin önemli bir bölümü bu temel modeller üzerine inşa edilmekte olup, bu yaklaşım AI geliştirme süreçlerini hızlandırmakta ve farklı disiplinlerdeki uygulamaların daha ölçeklenebilir ve erişilebilir hâle gelmesini sağlamaktadır.

### 2.1. Büyük Dil Modelleri (LLM) ve Çok Modlu Sistemler

Büyük Dil Modelleri (LLM'ler), doğal dili yalnızca istatistiksel örüntüler düzeyinde değil, bağlam, anlam ilişkileri ve görev niyeti düzeyinde de işleyebilen, çoğunlukla Transformer temelli ve milyarlarca parametreyle eğitilmiş temel modellerdir (Chen vd., 2024). Bu modeller; çok büyük metin koleksiyonları üzerinde ön eğitimden geçirilerek metin üretimi, çeviri, özetleme ve soru-cevaplama gibi görevlerde yüksek başarı göstermekte, ayrıca ölçek büyüdükçe bağlam içi öğrenme, yönerge izleme ve çok adımlı akıl yürütme gibi daha önce küçük modellerde belirgin olmayan "ortaya çıkan yetenekler" sergileyebilmektedir. Bu nedenle LLM'ler, yalnızca klasik doğal dil işleme araçları olarak değil; bilgiye dayalı karar destek, etkileşimli yardımcı sistemler ve genel amaçlı AI ajanlarının bilişsel çekirdeği olarak da değerlendirilmektedir. Nitekim son dönem çalışmalar, GPT, LLaMA ve PaLM gibi ailelerin LLM ekosistemini şekillendirdiğini; bu modellerin geniş ölçekli ön eğitim, ince ayar ve hizalama süreçleri sayesinde dil anlama ve üretme kapasitesini belirgin biçimde ileri taşıdığını göstermektedir (Chen vd., 2024; Minaee vd., 2024).

Çok modlu sistemler ise LLM'lerin bu dilsel yeteneklerini metin dışındaki veri türleriyle birleştirerek görüntü, ses, video ve sensör çıktıkları gibi heterojen girdileri ortak bir anlamsal uzayda işleyebilen daha kapsamlı yapılar ortaya koymaktadır. Bu çerçevede büyük çok modlu modeller (LMM/MLM), yalnızca görsel betimleme ya da görsel soru-cevap gibi görevleri yerine getirmekle kalmayıp, farklı modaliteler arasındaki ilişkileri öğrenerek çapraz-modal akıl yürütme, çok modlu diyalog ve bağlama duyarlı karar verme yetenekleri geliştirmektedir. Wang ve arkadaşlarına (2024) göre, KOSMOS-1, Gemini, BLIP-2, Flamingo, MiniGPT-4 ve LLaVA gibi örnekler; metin ile görüntü arasındaki hizalamanın, dondurulmuş görsel kodlayıcılar ile LLM'lerin ara katmanlar üzerinden bütünleştirilmesinin ve çok modlu ön eğitim stratejilerinin bu dönüşümde merkezi rol oynadığını göstermektedir. Dahası, otonom sürüşte LiDAR, kamera ve dil girdilerinin birlikte yorumlanması ya da insan-robot etkileşiminde konuşma, bakış, duruş ve nesne konumlarının doğal dil üzerinden işlenmesi, çok modlu sistemlerin fiziksel dünyaya ilişkin daha zengin ve uygulanabilir bir AI anlayışı sunduğunu ortaya koymaktadır (Wang vd., 2024; Li vd., 2025). Bu nedenle çok modlu sistemler, LLM'lerin metin merkezli sınırlarını aşarak üretken AI'yı daha kapsayıcı, etkileşimli ve gerçek dünya ile uyumlu bir yöne taşımaktadır.

## 2.2. Edge AI ve Hafif Model Mimarileri

Edge AI, AI modellerinin veriyi uzak bulut merkezlerine taşımadan doğrudan cihaz üzerinde ya da ağın kenarındaki düğümlerde çalıştırılmasını ifade eder. Bu yaklaşımın temel gerekçesi; gecikmeyi azaltmak, gerçek zamanlı karar vermeyi mümkün kılmak, bant genişliği kullanımını düşürmek ve özellikle sağlık, endüstriyel IoT, akıllı şehirler, otonom sistemler ve siber güvenlik gibi alanlarda yerel veri işleme avantajı sağlamaktır. Ancak kenar cihazlar; işlem gücü, bellek, enerji tüketimi ve ağ erişimi bakımından bulut altyapılarına göre belirgin sınırlılıklara sahiptir. Bu nedenle Edge AI ekosisteminde başarı, yalnızca güçlü modeller geliştirmekle değil; sınırlı kaynaklarda kabul edilebilir doğruluk, düşük gecikme ve enerji verimliliği arasında dengeli bir mimari kurabilmekle ilişkilidir (Singh ve Gill, 2023). Edge ortamları; heterojen donanımlar, dağıtık yapı, sınırlı CPU/bellek kapasitesi ve zaman zaman kesintili bağlantılar nedeniyle klasik büyük AI modelleri için uygun değildir; bu yüzden model sıkıştırma, yerel çıkarım ve edge-to-cloud hibrit mimariler temel tasarım ilkeleri hâline gelmiştir.

Bu bağlamda hafif model mimarileri, Edge AI'nın uygulanabilirliğini sağlayan temel yapı taşlarıdır. MobileNet ailesi, derinlik ayrılabilir evrişimler sayesinde işlem yükünü ve parametre sayısını azaltırken; SqueezeNet, “fire module” tasarımıyla çok daha küçük model boyutlarında rekabetçi performans

sunmaktadır. TinyML yaklaşımı ise mikrodenetleyici sınıfı cihazlarda dahi çıkarımı mümkün kılarak ultra düşük güç tüketimli uygulamaların önünü açmaktadır. Bunun yanında budama (pruning), nicemleme (quantization) ve bilgi damıtma (knowledge distillation) gibi sıkıştırma teknikleri, hafif modellerin edge cihazlara dağıtılabirliğini daha da güçlendirmektedir (Babalola vd., 2024). Güncel çalışmalarda yalnızca klasik hafif CNN'ler değil; EfficientNetV2 türevleri, MobileViTv2, EdgeViTs ve EdgeNeXt gibi yeni nesil mimariler de öne çıkmakta; hatta örnek bir çalışmada geliştirilen “Linge” modeli yalnızca 7.63 MB parametre boyutuyla edge sunucular üzerinde çalıştırılabilirken yüksek doğruluk ve AUC değerleri elde etmiştir. Bu durum, hafif mimarilerin artık yalnızca “küçük model” değil; bellek erişim maliyetini azaltan, dikkat mekanizmalarını seçici biçimde kullanan ve gerçek dünya edge senaryolarına göre yeniden tasarlanan özel mimari çözümler olarak ele alınması gerektiğini göstermektedir (Zhou vd., 2024).

### 2.3. Açıklanabilir AI (Explainable AI-XAI)

XAI, özellikle derin öğrenme ve diğer karmaşık makine öğrenmesi modellerinin “kara kutu” niteliğini azaltmayı; modelin hangi veri, örüntü, özellik ya da mantıksal ilişki üzerinden sonuca ulaştığını insanlar için anlaşılır hâle getirmeyi amaçlayan yöntemler bütünüdür (Bilal vd., 2025). Bu yönüyle XAI, yalnızca teknik bir şeffaflık aracı değil; aynı zamanda güven, hesap verebilirlik ve düzenleyici uyum açısından kritik bir çerçevedir. Angelov ve arkadaşlarının (2021) analitik incelemesinde vurgulandığı üzere, yorumlanabilirlik tek başına kara kutu modellerin doğurduğu tüm sorunları karşılamaya yetmez; kullanıcıların güvenini kazanmak ve kararların nedenlerini anlamlandırmak için açıklanabilirlik gereklidir. Aynı çalışmada NIST’in dört temel XAI ilkesi de öne çıkarılmaktadır: sistemin kararına dair bir gerekçe sunması, bu gerekçenin kullanıcı için anlamlı olması, açıklamanın sistemin gerçek işleyişini doğru yansıtması ve modelin bilgi sınırlarını tanıyabilmesi. Bu çerçeve, XAI’nın yalnızca model içi tekniklerin toplamı olmadığını; yerel ve küresel açıklamalar, post-hoc yaklaşımlar ve açıklanabilir-tasarım ilkeleri arasında kurulan bütüncül bir güven mimarisi olduğunu göstermektedir.

Güncel literatür ise XAI’nın etkili olabilmesi için açıklamaların yalnızca teknik olarak üretilmesinin yeterli olmadığını, insan-merkezli ve bağlama duyarlı olması gerektiğini göstermektedir (Budhkar vd., 2025; Nandan vd., 2025; Nikiforidis vd., 2025). Sağlık alanına odaklanan çalışmada, XAI’nın veri, muhakeme ve karar süreçlerine ilişkin içgörü sağlayarak insan anlayışını ve güveni artırdığı; hata ve önyargıların görünür kılınmasına katkı sunduğu belirtilmektedir. Bununla birlikte mevcut yaklaşımların çoğu hâlen algoritma-merkezlidir; oysa özellikle sağlık gibi yüksek riskli alanlarda açıklamanın hangi paydaş için,

hangi karar anında ve hangi etik gereksinimlerle üretildiği belirleyicidir. Bu nedenle insan-merkezli açıklanabilir AI yaklaşımı, teknik açıklamayı sosyo-teknik boyut, insan değerleri ve kullanıcı ihtiyaçlarıyla birleştirmektedir. LLM tabanlı güncel XAI yazını da bu yönelimi desteklemekte; post-hoc açıklamalar, içkin yorumlanabilirlik ve insan-merkezli anlatısal açıklamalar arasında bir ayırım yaparak, iyi bir açıklamanın yalnızca “doğru” değil, aynı zamanda anlaşılır, bağlama uygun ve kullanıcı tarafından denetlenebilir olması gerektiğini vurgulamaktadır (van Leersum vd., 2025). Dolayısıyla XAI, çağdaş AI sistemlerinde performans ile şeffaflık arasındaki gerilimi yönetmeye çalışan; güvenilir, adil ve kullanıcıyla iş birliği kurabilen sistemlerin inşasında temel rol oynayan bir yaklaşım olarak değerlendirilmelidir.

#### **2.4. Model Güvenilirliği ve Halüsinasyon Problemi**

Model güvenilirliği, AI sistemlerinin farklı bağlamlarda tutarlı, doğrulanabilir ve görevin gerektirdiği doğruluk düzeyinde çıktılar üretebilme kapasitesiyle ilişkilidir. Büyük dil modelleri bağlamında bu güvenilirlik sorunu, doğrudan model mimarisinin sınırlarıyla bağlantılıdır; çünkü LLM’ler gerçeği denetleyen bir mekanizma üzerinden değil, önceki belirteçlere dayanarak sonraki belirteci olasılıksal biçimde üreten otoregresif yapılar olarak çalışır (Ahn, 2025). Bu nedenle özellikle uzmanlık gerektiren alanlarda uzun menzilli tutarlılık, teknik doğruluk ve bağlama sadakat her zaman garanti edilememektedir. Halüsinasyon, yalnızca geçici bir mühendislik kusuru değil; kimi çalışmalara göre LLM mimarisinin yapısal bir sonucu olarak ortaya çıkmakta ve olgusal yanlışlık, yanlış yorumlama, kısmi yanlışlık ve uydurma gibi farklı biçimlerde görülebilmektedir. Ayrıca güncel değerlendirmeler, soru-cevap görevlerinde modellerin hâlen anlamlı düzeyde halüsinasyon üretme eğiliminde olduğunu göstermekte; örneğin bazı benchmark’larda bu oranların yaklaşık %17 ile %55 arasında değişebildiği raporlanmaktadır (Ahn, 2025; Yang vd., 2025). Literatürde halüsinasyon; girdiyle çelişen, bağlamla çelişen ve yerleşik dünya bilgisiyle çelişen çıktılar olarak sınıflandırılmakta, ayrıca veri toplama, eğitim ve çıkarım aşamalarındaki kalite sorunları, önyargılar, uzun bağlam işleme sınırlılıkları ve örnekleme rastlantısallığı da bu problemin başlıca nedenleri arasında gösterilmektedir (Yang vd., 2025b).

Halüsinasyon probleminin güvenilirlik üzerindeki etkisi, özellikle tıp, hukuk, bilimsel araştırma ve savunma gibi yüksek riskli alanlarda daha belirgin hâle gelmektedir; çünkü akıcı ve ikna edici görünen ama yanlış olan içerikler, uzman olmayan kullanıcıları kolayca yanıltabilmektedir. Bu nedenle güncel çalışmalar, model güvenilirliğini artırmak için çok katmanlı doğrulama ve denetim stratejileri önermektedir. İleri düzey istem mühendisliği teknikleri arasında Chain-of-Thought, Self-Consistency, Decomposition ve Chain-of-

Verification gibi yaklaşımlar, yanıtların daha sistematik üretilmesine katkı sağlarken; Retrieval-Augmented Generation (RAG) sistemleri modeli güvenilir veri tabanlarıyla destekleyerek çıktılarını olgusal zemine oturtmayı amaçlamaktadır (Ahn, 2025; Yang vd., 2025). Bununla birlikte Yang ve arkadaşlarına (2025) göre, RAG'ın da hataları tamamen ortadan kaldırmadığı açıkça görülmektedir. Daha yeni yaklaşımlar arasında ise çok-etmenli tartışma ve oylama mekanizmaları, tekrar sorgulama, hata günlüğü ve öz-yansıma süreçleri ile metamorfik ilişkiler üzerinden çalışan sıfır-kaynaklı halüsinasyon tespit yöntemleri öne çıkmaktadır. Bu çalışmaların ortak sonucu, model güvenilirliğinin tek başına model ölçeğiyle çözülemeyeceğini; güvenilir kullanım için görevle uyumlu araç seçimi, sistematik doğrulama, çapraz denetim ve özellikle insan gözetiminin vazgeçilmez olduğunu göstermesidir (Yang vd., 2025). Dolayısıyla halüsinasyon problemi, yalnızca teknik bir üretim hatası değil; güven, hesap verebilirlik ve gerçek dünya uygulamalarında AI sistemlerinin sınırlarını belirleyen temel bir güvenilirlik meselesidir.

## 2.5. Akademik ve endüstriyel kullanım alanları

Akademik kullanım alanlarında XAI, özellikle kararın yalnızca doğru olmasının değil, neden doğru olduğunun da gösterilmesi gereken yüksek etkili araştırma alanlarında öne çıkmaktadır. Kalasampath ve arkadaşlarına (2025) göre, son dönem XAI uygulamalarının en yoğun biçimde sağlık alanında toplandığını; kanser tanısı, COVID-19 yönetimi, tıbbi görüntüleme, nörobilim ve genel klinik karar destek süreçlerinde açıklanabilirliğin güven, hesap verebilirlik ve uzman doğrulaması açısından kritik görüldüğünü göstermektedir. Bunun yanında çevre ve tarım, finans, hukuk, eğitim, sosyal bakım, ulaşım ve siber güvenlik gibi alanlarda da akademik çalışmalar hızla artmaktadır. Eğitimde öğrenci performansının öngörülmesi ve öğrenme analitikleri, sosyal medyada yanlış bilgi ve nefret söylemi analizi, hukukta karar destek, siber güvenlikte saldırı ve anomali tespiti gibi örnekler, XAI'nın yalnızca teknik bir yorumlama aracı değil; alan uzmanlarının model davranışını değerlendirmesine imkân veren disiplinler arası bir araştırma altyapısı hâline geldiğini ortaya koymaktadır. Ayrıca bu literatürde SHAP ve LIME gibi yerel açıklama yöntemlerinin baskın olduğu, görselleştirme ve özellik önem derecelendirmesinin en yaygın açıklama biçimleri arasında yer aldığı görülmektedir (Kalasampath vd., 2025; Saarela vd., 2024).

### 3. Güvenli ve Etik AI Sistemleri

#### 3.1. Düşmanca (Adversarial) Saldırıları

Adversarial saldırılar, makine öğrenmesi ve özellikle derin öğrenme modellerinin giriş verisine insan tarafından çoğu zaman fark edilemeyecek kadar küçük fakat stratejik bozulmalar eklenerek hatalı karar vermeye zorlanmasıdır. Qiu ve arkadaşlarına (2019) göre, bu saldırıların eğitim aşamasında veri zehirlenme, etiket manipülasyonu ve özellik bozma biçiminde; test aşamasında ise beyaz kutu ve siyah kutu saldırıları olarak ortaya çıktığı belirtilmektedir. Ayrıca bu saldırıların yalnızca görüntü sınıflandırma ile sınırlı kalmadığı; doğal dil işleme, siber güvenlik, bulut servisleri ve fiziksel dünya uygulamalarında da ciddi tehdit oluşturduğu vurgulanmaktadır. Bu durum, AI sistemlerinde yüksek doğruluk oranlarının tek başına yeterli olmadığını; model dayanıklılığı, savunma mekanizmaları ve güvenli dağıtım stratejilerinin de en az performans kadar önemli olduğunu göstermektedir (Qiu vd., 2019).

#### 3.2. Model zehirlenme (Data/Model Poisoning)

Data/model poisoning, AI sistemlerinin eğitim verisini, doğrulama kümelerini ya da doğrudan modelin mantığını kasıtlı biçimde bozarak performansı düşürmeyi veya belirli yanlış kararlar üretmesini sağlamayı amaçlayan saldırıları ifade eder. Kaynağa göre veri zehirlenme; eğitim ya da doğrulama verisine yeni örnekler ekleme, mevcut içerik, etiket veya özellikleri değiştirme ya da veriyi silme yoluyla gerçekleştirilebilir ve hedefli ya da hedefsiz biçimde uygulanabilir. Model poisoning ya da mantık bozma saldırıları ise modelin algoritmasını, kodunu, gradyanlarını, kurallarını veya karar prosedürlerini değiştirerek doğruluğu azaltabilir ya da sistemi kötü niyetli çıktılar üretmeye yöneltebilir. Bu nedenle poisoning saldırıları, yalnızca veri kalitesine yönelik bir tehdit değil; veri işleme, model geliştirme ve dağıtım aşamalarının tamamını etkileyebilen, AI tabanlı yazılımların güvenilirliğini ve bütünlüğünü doğrudan zedeleyen kritik bir güvenlik sorunudur (Kumar vd., 2024).

#### 3.3. AI Güvenliği ve Siber Güvenlik Kesişimi

AI güvenliği ile siber güvenliğin kesişimi, AI'nin yalnızca korunması gereken bir hedef değil, aynı zamanda siber savunmayı dönüştüren etkin bir araç hâline gelmesiyle ortaya çıkmaktadır. Sontan ve Samuel'e (2024) göre AI; tehdit tespiti, zafiyet analizi ve olay müdahalesi gibi alanlarda büyük veri hacimlerini yüksek hızda işleyerek anomali ve saldırı örüntülerini geleneksel yöntemlere göre daha hızlı saptayabilmekte, böylece kurumların savunma kapasitesini artırmaktadır. Bununla birlikte, bu bütünleşmenin yeni riskler de ürettiği

vurgulanmaktadır: AI sistemlerinin şeffaflık eksikliği, önyargı, veri gizliliği, etik sorunlar ve saldırganlar tarafından kötüye kullanılabilme ihtimali, AI destekli siber güvenlik çözümlerinin dikkatli biçimde tasarlanmasını zorunlu kılmaktadır. Bu nedenle AI güvenliği ile siber güvenliğin kesişimi, bir yandan akıllı otomasyon ve öngörüsül savunma fırsatları sunarken, diğler yandan güvenilirlik, açıklanabilirlik, mahremiyet ve hesap verebilirlik ilkelerini merkeze alan çok katmanlı bir güvenlik anlayışını gerekli kılan stratejik bir alan olarak değlerlendirilmektedir (Sontan ve Samuel, 2024).

### 3.4. Regölasyonlar (AI Act, Veri Koruma Politikaları)

Regölasyonlar bağlamında AI ile veri koruma politikalarının kesişimi, teknolojik yenilik ile bireysel hakların korunması arasında denge kurma çabasına dayanmaktadır. Yanamala ve Suryadevara'ya (2023) göre, özellikle Genel Veri Koruma Yönetmeliği'nin (General Data Protection Regulation - GDPR), kişisel verilerin işlenmesi, şeffaflık, veri minimizasyonu ve hesap verebilirlik ilkeleri üzerinden küresel veri yönetişimini derinden etkilediği; buna benzer düzenlemelerin de farklı ülkelerde yaygınlaştığı vurgulanmaktadır. Aynı metin, Avrupa Birliği'nin önerdiği AI Act yaklaşımının ise özellikle yüksek riskli AI uygulamalarına yönelik yükümlülükler getirerek temel haklar, güvenlik ve etik uygunluk ekseninde yeni bir düzenleyici çerçeve oluşturduğunu belirtmektedir. Bu çerçevede regölasyonlar, yalnızca veri ihlallerini önlemeyi değil; aynı zamanda AI sistemlerinin şeffaf, adil, izlenebilir ve insan haklarıyla uyumlu biçimde geliştirilmesini sağlamayı amaçlamaktadır. Dolayısıyla AI Act ve veri koruma politikaları birlikte değlerlendirildiğinde, güncel AI yönetişiminin temelini oluşturan iki tamamlayıcı sütun olarak öne çıkmaktadır (Yanamala ve Suryadevara, 2023).

### 3.5. Etik Çerçeveler ve Algoritmik Adalet

Etik çerçeveler, AI sistemlerinin hangi ilkelere göre geliştirilmesi ve değlerlendirilmesi gerektiğini tanımlayan normatif bir yapı sunar. Prem'e göre (2023), güncel AI etik çerçevelerinin çoğunda insan özerkliği ve gözetimi, teknik sağlamlık ve güvenlik, mahremiyet ve veri yönetişimi, şeffaflık, çeşitlilik-ayrımcılık yapmama-adalet, toplumsal fayda ve hesap verebilirlik gibi ortak ilkelerin öne çıktığını göstermektedir. Ancak aynı çalışma, bu çerçevelerin çoğunlukla ilke düzeyinde kaldığını; yani "etik AI" hedeflerini tanımlasa da bu hedeflerin somut sistem tasarımına ve uygulanabilir teknik gereksinimlere nasıl dönüştürüleceğinin çoğu zaman belirsiz olduğunu vurgulamaktadır. Bu nedenle etik çerçeveler, AI yönetişimi için gerekli bir başlangıç zemini sağlasa da gerçek etki yaratabilmeleri için ilke listelerinin pratik tasarım kararları,

denetim araçları ve bağlama duyarlı uygulama mekanizmalarıyla desteklenmesi gerekmektedir.

Algoritmik adalet ise meseleyi yalnızca teknik doğruluk ya da önyargı azaltma problemi olarak değil, algoritmik kararların toplumda kimleri nasıl etkilediğini sorgulayan daha geniş bir adalet perspektifiyle ele alır. “Theorising Algorithmic Justice” çalışmasına göre (Marjanovic vd., 2022) algoritmik adalet; “adaletin konusu nedir, kim adaletin öznesidir, adaletsizlikler nasıl üretilir ve bu ihtilaflar nasıl giderilir?” soruları etrafında düşünülmelidir. Marjanovic ve arkadaşları (2025), algoritmik adaletsizliklerin yalnızca yanlış sınıflandırmadan ibaret olmadığını; maldistribution, misrecognition ve misrepresentation gibi ekonomik, sosyo-kültürel ve siyasal boyutlar taşıdığını, ayrıca datafication, black-boxing ve algoritmik inscrutability gibi süreçlerle görünmez biçimde üretilebildiğini göstermektedir. Bu bakımdan algoritmik adalet, etik ilkelerin toplumsal sonuçlarını görünür kılar; özellikle kırılgan gruplar üzerindeki etkileri, yanlış pozitiflerin yarattığı zararları ve insan gözetimi olmadan işleyen karar sistemlerinin doğurabileceği yapısal eşitsizlikleri tartışmaya açar.

## 4. Yüksek Performanslı ve Dağıtık Hesaplama Sistemleri

### 4.1. GPU/TPU Hızlandırılmalı Sistemler

GPU/TPU hızlandırılmalı sistemler, derin öğrenme ve büyük ölçekli veri işleme süreçlerinde ortaya çıkan yüksek paralel hesaplama gereksinimini karşılamak üzere geliştirilen temel hızlandırma altyapıları arasında yer almaktadır. Geleneksel CPU mimarileri daha genel amaçlı ve sıralı işlem odaklı çalışırken, GPU’lar binlerce çekirdekle eşzamanlı işlem yürütebilme yetenekleri sayesinde özellikle matris çarpımları, tensör işlemleri ve derin sinir ağı eğitimi gibi yoğun hesaplama gerektiren görevlerde belirgin performans üstünlüğü sağlamaktadır. TPU’lar ise tensör tabanlı işlemler için özel olarak tasarlanmış uygulamaya özgü tümleşik devreler olarak, özellikle büyük ölçekli yapay sinir ağı eğitim ve çıkarım süreçlerinde süreyi azaltma, ölçeklenebilirliği artırma ve işlem verimliliğini yükseltme açısından öne çıkmaktadır (Emmanuel vd., 2025). CPU, GPU ve TPU platformlarını karşılaştıran çalışmalar, özellikle dağıtık derin öğrenme senaryolarında GPU ve TPU’nun klasik işlemcilerle göre çok daha kısa eğitim süreleri sunduğunu; bazı iş yüklerinde TPU’nun daha dengeli ve yüksek verimli bir performans sergileyebildiğini göstermektedir. Bu nedenle GPU/TPU tabanlı hızlandırma sistemleri, yalnızca AI uygulamalarının performansını artıran teknik bileşenler değil; aynı zamanda büyük modellerin eğitilebilirliğini, deneysel çalışmaların uygulanabilirliğini ve modern yüksek performanslı hesaplama altyapılarının sürdürülebilirliğini belirleyen stratejik unsurlar olarak değerlendirilmektedir (Kimm vd., 2021).

## 4.2. Bulut Bilişim ve Hibrit Mimariler

Bulut bilişim, işlemci, depolama ve ağ kaynaklarının internet üzerinden isteğe bağlı ve ölçeklenebilir biçimde sunulmasını mümkün kılan bir hizmet modeli olarak, özellikle esnek kaynak kullanımı, hızlı ölçeklenme ve kullanım kadar ödeme yaklaşımı sayesinde modern bilgi işlem altyapılarının temelini oluşturmuştur (Aktas, 2018). Ancak tekil bulut yapıları zamanla sağlayıcı bağımlılığı, entegrasyon güçlüğü, güvenlik kaygıları ve farklı iş yüklerinin aynı ortamda verimli biçimde yönetilememesi gibi sınırlılıkları görünür kılmıştır. Bu nedenle hibrit mimariler, özel ve genel bulut kaynaklarını bir araya getirerek hem hassas verilerin daha kontrollü ortamlarda tutulmasına hem de değişken iş yüklerinin dış kaynaklarla elastik biçimde desteklenmesine olanak tanıyan daha dengeli bir çözüm olarak öne çıkmıştır. Bu yaklaşım, özellikle ölçeklenebilirlik, maliyet optimizasyonu, iş sürekliliği ve mevzuata uyum gereksinimlerinin birlikte ele alınması gereken kurumsal yapılarda önem kazanmaktadır (Cherukuri, 2019).

Hibrit mimarilerin asıl gücü, yalnızca farklı altyapıları birleştirmesinde değil, bu bileşenleri yönetilebilir ve birlikte çalışabilir bir yapıya dönüştürmesinde yatmaktadır. Literatürde bu amaçla önerilen mimariler; gerçek zamanlı izleme, olay akışı işleme, otomatik ölçekleme, önleyici bakım ve iş yükü yönlendirme gibi işlevleri ortak bir katmanda toplayarak bulut platformundan bağımsız yönetim anlayışını desteklemektedir (Mateescu vd., 2011). HPC ile bulutun kesiştiği yapılarda ise “elastic cluster” benzeri modeller, yerel yüksek performanslı kaynakları gerektiğinde bulut kaynaklarıyla genişleterek öngörülebilir yürütme, kapasite artışı ve cloud bursting olanağı sağlamaktadır. Böylece hibrit mimariler, bir yandan kurumsal BT’de esneklik ve güvenliği birlikte sunarken, diğer yandan bilimsel hesaplama ve büyük ölçekli iş yüklerinde performans ile kaynak verimliliği arasında işlevsel bir köprü kurmaktadır (Mateescu vd., 2011; Aktas, 2018).

## 4.3. HPC Kümeleri ve Slurm Tabanlı Kaynak Yönetimi

HPC kümeleri, çok sayıda işlem düğümünün yüksek hızlı ağlar üzerinden bir araya getirilerek büyük ölçekli ve yoğun hesaplama gerektiren iş yüklerinin paralel biçimde yürütülmesini sağlayan altyapılardır. Bu yapılar; CPU, GPU, bellek ve depolama gibi heterojen kaynakların birlikte kullanılmasına imkân vererek genomik analiz, görüntü işleme, bilimsel simülasyon ve büyük model çıkarımı gibi alanlarda hesaplama süresini anlamlı biçimde azaltmaktadır. Bu tür ortamlarda temel sorun, mevcut kaynakların çok kullanıcı ve çok işli senaryolarda verimli, adil ve ölçeklenebilir biçimde tahsis edilmesidir. Slurm tabanlı kaynak yönetimi bu noktada, düğüm, CPU, GPU, RAM ve iş kuyruğu

yönetimini merkezi bir planlama mantığıyla düzenleyerek HPC altyapısının işletim omurgasını oluşturmaktadır. Slurm'un Linux tabanlı kurulum, kimlik doğrulama için Munge bağımlılığı, slurm.conf üzerinden denetleyici düğüm ve adres tanımları, servislerin yeniden başlatılması ve sbatch ile iş gönderimi gibi bileşenleri, dağıtık hesaplama kaynaklarının ortak bir hizmet yapısı altında yönetilmesini mümkün kılmaktadır (Curia-Alcantara vd., 2024).

Güncel çalışmalarda (Doosthosseini vd., 2024; Decker vd., 2025) Slurm yalnızca klasik toplu iş planlayıcısı olarak değil, aynı zamanda büyük dil modelleri gibi modern AI iş yüklerini çok düğümlü HPC ortamlarında ölçeklenebilir biçimde yürütmeyi sağlayan bir orkestrasyon katmanı olarak da ele alınmaktadır. Özellikle heterojen kümelerde CPU, GPU ve bellek tahsisinin dinamik olarak yapılması, konteyner tabanlı mikro servislerin entegrasyonu, REST tabanlı çıkarım uç noktalarının oluşturulması ve yük dengeleme mekanizmalarının devreye alınması, Slurm'un çağdaş HPC mimarilerindeki rolünü genişletmiştir. Nitekim ölçeklenebilir LLM çıkarım mimarilerini inceleyen çalışma, küçük modellerin daha yüksek eşzamanlı istekleri düşük gecikmeyle karşılayabildiğini, büyük modellerde ise doygunluk noktasına çok daha erken ulaşıldığını göstermekte; bu durum da kaynak planlama, kuyruk yönetimi, yeniden sıraya alma ve hata toleransı gibi Slurm işlevlerinin performans kadar kritik olduğunu ortaya koymaktadır (Luiz vd., 2025). Böylece Slurm tabanlı kaynak yönetimi, HPC kümelerinde yalnızca işlem dağıtımını yapan bir yazılım bileşeni değil; performans, ölçeklenebilirlik, güvenilirlik ve hizmet sürekliliğini birlikte yöneten stratejik bir yönetim katmanı hâline gelmektedir.

#### 4.4. Edge-Cloud İş Bölümü

Edge-Cloud iş bölümü, AI ve veri yoğun uygulamalarda görevlerin gecikme, bant genişliği, enerji tüketimi ve hesaplama gereksinimlerine göre kenar ve bulut katmanları arasında paylaştırılmasını ifade eder. Bu yapıda bulut tarafı, büyük veri birikimi, yüksek işlem gücü ve model eğitimi gibi hesaplama yoğun süreçler için uygun bir ortam sunarken; edge katmanı, düşük gecikme gerektiren gerçek zamanlı çıkarım, yerel karar verme ve bağlama duyarlı hizmetler açısından kritik rol oynamaktadır (Yao vd., 2022). Literatür, tek başına bulutun büyük veri aktarımı ve gecikme sorunları nedeniyle, tek başına edge'nin ise sınırlı işlem gücü, bellek ve enerji kaynakları nedeniyle tüm gereksinimleri karşılamakta yetersiz kaldığını göstermektedir (Ma vd., 2025). Bu nedenle edge-cloud iş bölümü, özellikle gecikmeye duyarlı görevlerin edge'de, yüksek hesaplama maliyetli eğitim, koordinasyon ve küresel optimizasyon işlemlerinin ise bulutta yürütülmesine dayanan tamamlayıcı bir mimari anlayış olarak öne çıkmaktadır. Ayrıca çok atlamalı görev aktarımı, heterojen kaynak yapısı, ağ topolojisi

ve görev türlerindeki çeşitlilik, bu iş bölümünün sabit değil; dinamik görev zamanlama, kaynak ticareti, hesaplama offloading'i ve dijital ikiz destekli izleme mekanizmalarıyla sürekli optimize edilen bir süreç olduğunu göstermektedir. Böylece edge-cloud iş bölümü, yalnızca teknik bir dağıtım yaklaşımı değil; endüstriyel IoT, akıllı hizmetler ve modern AI sistemlerinde performans, ölçeklenebilirlik ve kaynak verimliliği arasında denge kuran temel bir mimari ilke hâline gelmektedir (Li vd., 2024).

#### 4.5. Enerji Verimli Hesaplama (Green Computing)

Green Computing, yüksek performanslı bilgi işlem ve AI ekosistemlerinde artan işlem gücü ihtiyacının çevresel etkilerini azaltmayı amaçlayan bir yaklaşım olarak öne çıkmaktadır. Bu çerçevede Green AI, yalnızca model doğruluğunu değil, eğitim ve çıkarım süreçlerinin enerji tüketimi, karbon ayak izi ve kaynak verimliliği boyutlarını da değerlendirmeyi gerekli kılar. Sistemik incelemeler, özellikle 2020 sonrasında bu alandaki çalışmaların hızla arttığını; enerji tüketiminin izlenmesi, hiper parametre ayarlaması, model kıyaslama, dağıtım stratejileri ve doğruluk-enerji dengesi gibi başlıkların öne çıktığını göstermektedir (Verdecchia vd., 2023). Bulut bilişim tarafında ise veri merkezlerinin sürekli çalışan sunucular, soğutma sistemleri ve depolama altyapıları nedeniyle ciddi elektrik tüketimi ve karbon salımı ürettiği; bu nedenle akıllı iş yükü planlaması, dinamik kaynak tahsisi, otomatik ölçekleme, veri sıkıştırma, aynı verinin tekrar eden kopyalarını tespit edip tekilleştirme işlemi (deduplikasyon) ve AI destekli soğutma mekanizmalarının enerji verimliliği açısından kritik olduğu anlaşılmaktadır. Dolayısıyla enerji verimli hesaplama, yalnızca donanım tasarrufu sağlayan teknik bir optimizasyon alanı değil; algoritma tasarımı, yazılım mimarisi, veri merkezi yönetimi ve sürdürülebilir bulut altyapılarının birlikte ele alındığı bütüncül bir dönüşüm alanı olarak değerlendirilmelidir (Oloruntoba vd., 2024).

## Kaynaklar

- Bibri, S. E. (2019). On the sustainability of smart and smarter cities in the era of big data: an interdisciplinary and transdisciplinary literature review. *Journal of Big Data*, 6(1), 25.
- Chinnaiyan, B., Balasubaramanian, S., Jeyabalu, M., & Warriar, G. S. (2025). AI Applications—Computer Vision and Natural Language Processing. *Model optimization methods for efficient and edge AI: Federated learning architectures, frameworks and applications*, 25-41.
- Zhou, P., Wang, L., Liu, Z., Hao, Y., Hui, P., Tarkoma, S., & Kangasharju, J. (2024). A survey on generative ai and llm for video generation, understanding, and streaming. *arXiv preprint arXiv:2404.16038*.
- Deelman, E., Dongarra, J., Hendrickson, B., Randles, A., Reed, D., Seidel, E., & Yelick, K. (2025). High-performance computing at a crossroads. *Science*, 387(6736), 829-831.
- Altukhi, Z. M., & Pradhan, S. (2025). Systematic literature review: Explainable AI definitions and challenges in education. *arXiv preprint arXiv:2504.02910*.
- Keller, J. M. (2025). Architectural and Software-Based Fault Tolerance in Multicore and Lockstep Processing Systems: A Comprehensive Reliability-Centric Analysis. *Academic Research Library for International Journal of Computer Science & Information System*, 10(11), 103-108.
- De Souza, A. D. C., & de Freitas, H. C. (2025). A Performance Analysis of System Rollback Techniques. In *Simpósio em Sistemas Computacionais de Alto Desempenho (SSCAD)* (pp. 9-16). SBC.
- Barwasser, A., Lentjes, J., Riedel, O., Zimmermann, N., Dangelmaier, M., & Zhang, J. (2023). Method for the development of Software-Defined Manufacturing equipment. *international journal of production research*, 61(19), 6467-6484.
- Lee, J., Kim, J., Park, S. J., Song, B., & Moon, S. K. (2025). Software-Defined Product Architecture: Status, Challenges, and Future Perspectives. In *2025 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)* (pp. 1088-1093). IEEE.
- Akarvardar, K., & Wong, H. S. P. (2023). Technology prospects for data-intensive computing. *Proceedings of the IEEE*, 111(1), 92-112.
- Evans, M. R., Oliver, D., Zhou, X., & Shekhar, S. (2024). Spatial big data: Case studies on volume, velocity, and variety. In *Big Data* (pp. 115-138). CRC Press.
- Pyzer-Knapp, E. O., Pitera, J. W., Staar, P. W., Takeda, S., Laino, T., Sanders, D. P., ... & Curioni, A. (2022). Accelerating materials discovery using artificial intelligence, high performance computing and robotics. *npj Computational Materials*, 8(1), 84.

- Navaux, P. O. A., Lorenzon, A. F., & da Silva Serpa, M. (2023). Challenges in high-performance computing. *Journal of the Brazilian Computer Society*, 29(1), 51-62.
- Egger, R., & Yu, J. (2022). Data science and interdisciplinarity. *Applied Data Science in Tourism: Interdisciplinary Approaches, Methodologies, and Applications*, 35-49.
- Thirunavukarasu, R., & Kotei, E. (2024). A comprehensive review on transformer network for natural and medical image analysis. *Computer Science Review*, 53, 100648.
- Maheshwari, S., Shetty, S., Ratnakar, R., & Sanyal, S. (2022). Role of computational science in materials and systems design for sustainable energy applications: An industry perspective. *Journal of the Indian Institute of Science*, 102(1), 11-37.
- Kumar, S., Sharma, D., Rao, S., Lim, W. M., & Mangla, S. K. (2025). Past, present, and future of sustainable finance: insights from big data analytics through machine learning of scholarly research. *Annals of Operations Research*, 345(2), 1061-1104.
- Weber, P., Carl, K. V., & Hinz, O. (2024). Applications of Explainable Artificial Intelligence in Finance-a systematic review of Finance, Information Systems, and Computer Science literature: P. Weber et al. *Management Review Quarterly*, 74(2), 867-907.
- Khan, A., Imam, I., & Azam, A. (2021). Role of Artificial Intelligence in Defence Strategy. *Strategic Studies*, 41(1), 19-40.
- Weng, Y. (2024). Big data and machine learning in defence. *International Journal of Computer Science and Information Technology*, 16(2), 25-35.
- Fui-Hoon Nah, F., Zheng, R., Cai, J., Siau, K., & Chen, L. (2023). Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. *Journal of information technology case and application research*, 25(3), 277-304.
- Waqas, A., Bui, M. M., Glassy, E. F., El Naqa, I., Borkowski, P., Borkowski, A. A., & Rasool, G. (2023). Revolutionizing digital pathology with the power of generative artificial intelligence and foundation models. *Laboratory investigation*, 103(11), 100255.
- Chen, Z., Xu, L., Zheng, H., Chen, L., Tolba, A., Zhao, L., ... & Feng, H. (2024). Evolution and Prospects of Foundation Models: From Large Language Models to Large Multimodal Models. *Computers, Materials & Continua*, 80(2).
- Minaee, S., Mikolov, T., Nikzad, N., Chenaghlu, M., Socher, R., Amatriain, X., & Gao, J. (2024). Large language models: A survey. *arXiv preprint arXiv:2402.06196*.
- Wang, C., Hasler, S., Tanneberg, D., Ocker, F., Joubin, F., Ceravola, A., ... & Gienger, M. (2024, May). Lami: Large language models for multi-modal

- human-robot interaction. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (pp. 1-10).
- Li, J., Li, J., Yang, G., Yang, L., Chi, H., & Yang, L. (2025). Applications of large language models and multimodal large models in autonomous driving: A comprehensive review. *Drones*, 9(4), 238.
- Singh, R., & Gill, S. S. (2023). Edge AI: a survey. *Internet of Things and Cyber-Physical Systems*, 3, 71-92.
- Babalola, O., Raji, O. M. O., Akande, J. O., Abdulkareem, A. O., Anyah, V., Samson, A., & Folorunso, S. (2024). AI-powered cybersecurity in edge computing: Lightweight neural models for anomaly detection. *International Journal of Multidisciplinary Research and Growth Evaluation*, 5(2), 1130-1138.
- Zhou, F., Hu, S., Du, X., Wan, X., & Wu, J. (2024). A lightweight neural network model for disease risk prediction in edge intelligent computing architecture. *Future Internet*, 16(3), 75.
- Bilal, A., Ebert, D., & Lin, B. (2025). Llms for explainable ai: A comprehensive survey. *arXiv preprint arXiv:2504.00125*.
- Angelov, P. P., Soares, E. A., Jiang, R., Arnold, N. I., & Atkinson, P. M. (2021). Explainable artificial intelligence: an analytical review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 11(5), e1424.
- Budhkar, A., Song, Q., Su, J., & Zhang, X. (2025). Demystifying the black box: A survey on explainable artificial intelligence (XAI) in bioinformatics. *Computational and Structural Biotechnology Journal*, 27, 346-359.
- Nandan, M., Mitra, S., & De, D. (2025). GraphXAI: a survey of graph neural networks (GNNs) for explainable AI (XAI). *Neural Computing and Applications*, 37(17), 10949-11000.
- Nikiforidis, K., Kyrtoglou, A., Vafeiadis, T., Kotsiopoulos, T., Nizamis, A., Ioannidis, D., ... & Sarigiannidis, P. (2025). Enhancing transparency and trust in AI-powered manufacturing: A survey of explainable AI (XAI) applications in smart manufacturing in the era of industry 4.0/5.0. *ICT Express*, 11(1), 135-148.
- van Leersum, C. M., & Maathuis, C. (2025). Human centred explainable AI decision-making in healthcare. *Journal of Responsible Technology*, 21, 100108.
- Ahn, S. (2025). A guide to evade hallucinations and maintain reliability when using large language models for medical research: a narrative review. *Annals of Pediatric Endocrinology & Metabolism*, 30(3), 115-118.
- Yang, B., Mamun, M. A. A., Zhang, J. M., & Uddin, G. (2025). Hallucination detection in large language models with metamorphic relations. *URL <https://arxiv.org/abs/2502.15844>*.

- Yang, Y., Ma, Y., Feng, H., Cheng, Y., & Han, Z. (2025). Minimizing hallucinations and communication costs: Adversarial debate and voting mechanisms in llm-based multi-agents. *Applied Sciences*, 15(7), 3676.
- Kalasampath, K., Spoorthi, K. N., Sajeev, S., Kuppa, S. S., Ajay, K., & Maruthamuthu, A. (2025). A literature review on applications of explainable artificial intelligence (XAI). *IEEE access*, 13, 41111-41140.
- Saarela, M., & Podgorelec, V. (2024). Recent applications of explainable AI (XAI): A systematic literature review. *Applied Sciences*, 14(19), 8884.
- Qiu, S., Liu, Q., Zhou, S., & Wu, C. (2019). Review of artificial intelligence adversarial attack and defense technologies. *Applied Sciences*, 9(5), 909.
- Kumar, V., Mayo, J., & Bahiss, K. (2024). Admin: Attacks on dataset, model and input. a threat model for ai based software. *arXiv preprint arXiv:2401.07960*.
- Sontan, A. D., & Samuel, S. V. (2024). The intersection of Artificial Intelligence and cybersecurity: Challenges and opportunities. *World Journal of Advanced Research and Reviews*, 21(2), 1720-1736.
- Yanamala, A. K. Y., & Suryadevara, S. (2023). Advances in data protection and artificial intelligence: Trends and challenges. *International Journal of Advanced Engineering Technologies and Innovations*, 1(01), 294-319.
- Prem, E. (2023). From ethical AI frameworks to tools: a review of approaches. *AI and Ethics*, 3(3), 699-716.
- Marjanovic, O., Cecez-Kecmanovic, D., & Vidgen, R. (2022). Theorising algorithmic justice. *European Journal of Information Systems*, 31(3), 269-287.
- Emmanuel, F. C., Henry, O. N., & Chibuzo, O. B. (2025). A survey comparing specialized hardware and evolution in cpu, gpu and tpu for neural network. *IRE Journals*, 8.
- Kimm, H., Paik, I., & Kimm, H. (2021). Performance comparison of tpu, gpu, cpu on google colab over distributed deep learning. In *2021 IEEE 14th International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoc)* (pp. 312-319). IEEE.
- Aktas, M. S. (2018). Hybrid cloud computing monitoring software architecture. *Concurrency and Computation: Practice and Experience*, 30(21), e4694.
- Cherukuri, B. R. (2019). Future of cloud computing: Innovations in multi-cloud and hybrid architectures. *World J. Adv. Res. Rev*, 1(1), 068-081.
- Mateescu, G., Gentsch, W., & Ribbens, C. J. (2011). Hybrid computing-where HPC meets grid and cloud computing. *Future Generation Computer Systems*, 27(5), 440-453.
- Curia-Alcantara, N. E., Inga-Coveñas, C., & Aucchuasi, W. (2024). Methodology for the Implementation of Slurm-based HPC Services. In *2024 3rd International Conference on Automation, Computing and Renewable Systems (ICACRS)* (pp. 771-775). IEEE.

- Doosthosseini, A., Decker, J., Nolte, H., & Kunkel, J. M. (2024). Chat ai: A seamless slurm-native solution for hpc-based services. *arXiv preprint arXiv:2407.00110*.
- Decker, J., Metje, S., & Kunkel, J. (2025). Running Kubernetes workloads on rootless HPC systems using Slurm. *CLOUD COMPUTING*, *111*, 2025.
- Luiz, A. D. L., Kurlekar, S. V., & Georges, M. (2025). Scalable Engine and the Performance of Different LLM Models in a SLURM based HPC architecture. *arXiv preprint arXiv:2508.17814*.
- Yao, J., Zhang, S., Yao, Y., Wang, F., Ma, J., Zhang, J., ... & Yang, H. (2022). Edge-cloud polarization and collaboration: A comprehensive survey for AI. *IEEE Transactions on Knowledge and Data Engineering*, *35*(7), 6866-6886.
- Ma, Q., Qin, Y., Zhu, C., Gao, L., & Chen, X. (2025). Joint resource trading and task scheduling in edge-cloud computing networks. *IEEE Transactions on Networking*, *33*(3), 994-1008.
- Li, X., Chen, B., Fan, J., Kang, J., Ye, J., Wang, X., & Niyato, D. (2024). Cloud-edge-end collaborative intelligent service computation offloading: A digital twin driven edge coalition approach for industrial IoT. *IEEE Transactions on Network and Service Management*, *21*(6), 6318-6330.
- Verdecchia, R., Sallou, J., & Cruz, L. (2023). A systematic review of Green AI. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *13*(4), e1507.
- Oloruntoba, O. (2024). Green cloud computing: AI for sustainable database management. *World Journal of Advanced Research and Reviews*, *23*(03), 3242-3257.

## Histogram Tabanlı Çok Seviyeli Eşikleme Tekniklerinin Renk Nicemleme Üzerindeki Etkileri

Mahmut Kılıçaslan<sup>1</sup>

### Özet

Bu bölümde RGB renk uzayında renk indirgeme problemi, çok seviyeli histogram tabanlı eşikleme yöntemleri kullanılarak ele alınmıştır. Otsu, Kapur entropi ve histogramın ağırlık merkezi tabanlı eşikleme teknikleri ile her bir renk kanalı için optimum eşik değerleri belirlenmiş ve bu eşik değerleri kullanılarak renk uzayı alt sınıflara ayrılmıştır. Elde edilen her alt sınıfta yer alan piksel değerleri, ilgili sınıfın ortalama yoğunluk değeri ile temsil edilmiştir. Bu yaklaşım sayesinde görüntünün temel karakteristik özellikleri korunurken veri boyutunda önemli ölçüde azalma sağlanmıştır. Çalışmada farklı eşikleme yöntemlerinin renk dağılımı, sınıf içi homojenlik ve görsel kalite üzerindeki etkileri karşılaştırmalı olarak incelenmiştir. Elde edilen bulgular, çok seviyeli eşikleme tabanlı renk indirgeme yöntemlerinin görüntü işleme uygulamalarında etkili bir ön işleme adımı olarak kullanılabileceğini göstermektedir.

### 1. Giriş

Dijital görüntü işleme alanında renk indirgeme, görüntülerin daha düşük veri boyutu ile temsil edilmesini sağlarken görsel bilginin mümkün olduğunca korunmasını amaçlayan önemli bir ön işleme adımıdır [1]. Renk indirgeme yaklaşımları bölütleme ve içerik tabanlı görüntü erişimi gibi farklı alanlarda da kullanılmaktadır [2-4]. Özellikle depolama, iletim ve hesaplama maliyetlerinin azaltılması gereken uygulamalarda etkili renk nicemleme yöntemlerine ihtiyaç duyulmaktadır. Histogram tabanlı eşikleme teknikleri, gri seviye görüntülerde bölütleme amacıyla yaygın olarak kullanılmasına rağmen, bu yöntemlerin renk uzayına uyarlanarak renk indirgeme problemi için kullanılması literatürde sınırlı

1 Doç. Dr., Ankara Üniversitesi Bilişim Teknolojileri Meslek Yüksekokulu, İstatistik Bölümü, Büyük Veri Analistliği Programı, m.kilicaslan@ankara.edu.tr, ORCID ID: 0000-0003-1117-7736

sayıda çalışmada ele alınmıştır. Çok seviyeli eşikleme yöntemleri, görüntü histogramını istatistiksel ve bilgi kuramsal ölçütlere göre analiz ederek optimum eşik değerlerini belirlemeye olanak tanımaktadır. Otsu yöntemi sınıflar arası varyansı maksimize etmeye dayanırken, Kapur yöntemi maksimum entropi ilkesini esas almakta, histogramın ağırlık merkezi yaklaşımı ise dağılımın moment özelliklerinden yararlanmaktadır. Histogramın ağırlık merkezine dayalı indirgeme, özellikle kenar görüntülerinin eşiklenmesinde oldukça başarılı sonuçlar vermiştir [5-7]. Bu farklı optimizasyon kriterleri, elde edilen alt sınıfların yapısını ve dolayısıyla renk indirgeme performansını doğrudan etkilemektedir. Bu bölümde, söz konusu eşikleme teknikleri RGB renk kanallarına ayrı ayrı uygulanarak renk uzayı çok seviyeli olarak bölütlenmekte ve her bir alt sınıf sınıf ortalaması ile temsil edilmektedir. Böylece hem renk nicemleme gerçekleştirilmekte hem de farklı eşikleme yöntemlerinin renk indirgeme performansı üzerindeki etkileri karşılaştırmalı olarak analiz edilmektedir.

Renk indirgeme temel olarak 2 aşamadan oluşmaktadır. Bu adımlar palet tasarımı ve piksel eşleme (pixel mapping) olarak adlandırılır [8]. Palet tasarımı hesaplama bakımından zor ve kritik bir adımdır. Bu sebeple birçok makale palet tasarımına odaklanılmışlardır. Palet tasarımı için uygulanan birçok makale bulunmaktadır [9-11]. Median-Cut algoritması popüler renk indirgeme yaklaşımlarından biridir. Median cut renk indirgeme algoritması, sayısal görüntülerdeki yüksek renk çeşitliliğini sınırlı sayıda temsilci renge dönüştürmek amacıyla kullanılan sezgisel ve yinelemeli bir renk kuantalama yöntemidir. Algoritma, görüntüde bulunan tüm pikselleri RGB renk uzayında tek bir küme olarak ele almakta ve her yinelemede en geniş renk dağılımına sahip bileşeni (R, G veya B) belirleyerek ilgili eksen boyunca ortanca değere göre kümeyi iki alt kümeye ayırmaktadır. Bu bölme işlemi, hedeflenen palet boyutuna ulaşıncaya kadar sürdürülmekte ve elde edilen her alt küme, içerdiği renklerin ortalaması alınarak tek bir temsilci renk ile ifade edilmektedir. Böylece renk uzayı, her biri yaklaşık eşit sayıda piksel içeren kutulara ayrılmakta ve hem hesaplama maliyeti düşük hem de doğal görüntüler için görsel olarak tatmin edici bir palet oluşturulmaktadır. Median cut yöntemi; uygulama kolaylığı, hızlı çalışması ve ek parametre gerektirmemesi nedeniyle özellikle palet tabanlı görüntü gösterimi, görüntü sıkıştırma ve düşük bit düzeyli görüntü işleme uygulamalarında yaygın olarak tercih edilmekte, ancak insan görme sisteminin algısal özelliklerini doğrudan dikkate almaması nedeniyle bazı durumlarda optimal renk dağılımını sağlayamamaktadır [12]. Octree renk indirgeme algoritması ise bir diğer renk indirgeme yöntemidir. İlgili algoritma ağaç tabanlı bir renk kuantalama yöntemidir. Bu yöntemde RGB renk uzayı, her seviye bir renk bileşenine ait bit bilgisini temsil eden sekiz dallı bir ağaç yapısı

(octree) biçiminde tasarlar ve görüntüdeki her piksel, renk bileşenlerinin en anlamlı bittene en az anlamlı bite doğru ilerlenmesiyle ağaca yerleştirilmektedir. Her düğüm, kendisine karşılık gelen renk kümesine ait piksel sayısını ve toplam renk değerlerini tutarak istatistiksel bilgi biriktirmekte, hedeflenen palet boyutuna ulaşmak için ise ağaç alt seviyelerden başlayarak kontrollü biçimde budanmaktadır. Octree algoritması, özellikle tek geçişte çalışabilmesi, dinamik olarak palet oluşturabilmesi ve gerçek zamanlı görüntü işleme uygulamalarına uygunluğu nedeniyle gömülü sistemler, grafik donanımları ve palet tabanlı görüntü sıkıştırma süreçlerinde yaygın olarak kullanılmakta; buna karşın budama stratejisine bağlı olarak bazı durumlarda algısal açıdan optimal olmayan renk dağılımları üretebilmektedir [13]. Popüler yaklaşımlardan bir diğeri de k-means yaklaşımı ile gerçekleştirilen renk indirgeme yaklaşımlarıdır [14-17]. K-means tabanlı renk indirgeme yaklaşımı, benzer renkleri aynı küme altında toplayarak sınırlı sayıda temsilci renge dönüştüren, bir yöntemidir. Bu yöntemde görüntüyü oluşturan pikseller RGB renk uzayında birer veri noktası olarak ele alınmakta ve önceden belirlenen K adet küme merkezi ile yinelemeli bir süreç yürütülmektedir. İlk aşamada rastgele veya belirli bir sezgisel yaklaşımla seçilen küme merkezleri, her pikselin Öklid uzaklığına göre en yakın merkeze atanması ve ardından her küme için yeni merkezin o kümeyle ait renklerin ortalaması alınarak güncellenmesi adımlarıyla sürekli olarak iyileştirilmektedir. Bu süreç, küme merkezlerinde anlamlı bir değişim kalmayınca ya da belirlenen yineleme sayısına ulaşıncaya kadar devam etmektedir. Sonuç olarak her küme merkezi paleti temsil eden bir renk olarak kullanılmaktadır. Söz konusu yöntem, renk dağılımını veri temelli olarak modellemesi ve algısal açıdan daha dengeli sonuçlar üretebilmesi nedeniyle yüksek görsel kalite gerektiren renk indirgeme uygulamalarında etkili bir çözüm sunmaktadır. Diğer taraftan başlangıç küme merkezlerine duyarlılığı, yinelemeli yapısı nedeniyle yüksek hesaplama maliyeti ve büyük görüntülerde işlem süresinin artması gibi sınırlılıklara da sahip bulunmaktadır. Fuzzy C-Means (FCM) tabanlı renk indirgeme yöntemi kümeleme tabanlı olarak sınıflandırılan yaklaşımlardandır [18, 19]. Görüntüde ki renklerin keskin sınırlarla değil, bulanık üyelik dereceleri ile kümelere atanmasını esas almaktadır. Bu yöntemde her piksel RGB renk uzayında bir veri noktası olarak ele alınmakta ve önceden belirlenen küme sayısı kadar renk kümesi oluşturulmaktadır. K-means algoritmasından farklı olarak her piksel yalnızca tek bir kümeyle atanmak yerine, tüm kümelere belirli üyelik katsayıları ile bağlıdır ve bu katsayılar piksel ile küme merkezi arasındaki uzaklığa bağlı olarak hesaplanmaktadır. Elde edilen küme merkezleri paleti temsil eden renkleri oluştururken, pikseller genellikle en yüksek üyelik derecesine sahip oldukları kümeyle atanarak yeni görüntü elde edilmektedir. 2023 yılında Kılıçaslan ve İncetaş tarafından renk

indirgeme problemine yönelik olarak histogram tabanlı çok seviyeli eşikleme yaklaşımını kullanan ve temsilci renk sayısını otomatik olarak belirleyebilen yeni bir renk kuantalama yöntemi önerilmiştir[20]. Geliştirilen model üç temel aşamadan oluşmaktadır. İlk aşamada RGB renk uzayının her bir bileşenine ait histogramlar ağırlık merkezi (Center of Gravity – CoG) tabanlı çok seviyeli eşikleme ile bölütlenmiştir. Sonraki aşamada bu alt bölgelerin merkez noktaları kullanılarak küme merkezleri ve renk paleti oluşturulmuş, son aşamada ise her piksel Öklid uzaklığına göre en yakın küme merkezine yeniden atanarak nihai indirgenmiş görüntü elde edilmektedir. Önerilen yöntem, klasik bölme tabanlı tekniklerin geri dönüşsüz bölme problemi ile kümeleme tabanlı yöntemlerin başlangıç değerlerine duyarlılık ve kullanıcıya bağlı küme sayısı belirleme gereksinimini ortadan kaldırmayı hedeflemiştir.

Bu çalışmada, Otsu, Kapur ve histogramın ağırlık merkezi tabanlı çok seviyeli eşikleme yöntemleri kullanılarak RGB renk uzayına ait histogramlar üzerinden eşik değerleri belirlenmiş ve bu eşiklere bağlı olarak renk uzayı alt bölgelere ayrılmıştır. Gerçekleştirilen bölütleme işleminin ardından, her bir alt kümeyle düşen piksellerin renk değerlerinin ortalaması alınarak temsilci renkler elde edilmiş ve böylece renk indirgeme işlemi gerçekleştirilmiştir. Literatürde benzer biçimde eşikleme temelli bölütleme yaklaşımları farklı görüntü işleme problemleri için kullanılmıştır. Ancak bu çalışmada, Otsu, Kapur ve ağırlık merkezi tabanlı eşikleme tekniklerinin renk kuantalama başarımına etkileri; elde edilen görüntülerin görsel kalitesi ve nicel performans ölçütleri üzerinden analiz edilerek yöntemlerin güçlü ve zayıf yönleri tartışılmıştır. Böylece eşikleme tabanlı renk indirgeme yaklaşımlarının hem kuramsal hem de uygulamaya dönük karşılaştırmalı bir değerlendirmesi gerçekleştirilmiştir.

## 2. Görüntü Eşikleme Teknikleri

Eşikleme, bir görüntüyü anlamlı ve analiz edilebilir alt bölgelere ayırmak amacıyla piksel yoğunluk dağılımını temel alan en temel ve etkili bölütleme yaklaşımlarından biridir. Bu yöntemlerde temel amaç, görüntü histogramından yararlanarak sınıflar arasındaki ayrımı en iyi temsil eden eşik değerlerini otomatik olarak belirlemek ve böylece farklı özelliklere sahip piksel gruplarını birbirinden ayırmaktır. Tek seviyeli eşikleme yöntemleri görüntüyü iki sınıfa ayırırken, çok seviyeli eşikleme teknikleri daha karmaşık sahnelerin birden fazla homojen bölgeye ayrılmasına olanak tanımaktadır. Otsu, Kapur ve histogramın ağırlık merkezi gibi yöntemler, eşik değerlerini belirlerken sırasıyla sınıflar arası varyansın eniyilenmesi, entropinin maksimize edilmesi ve histogram dağılımının merkezinin dikkate alınması gibi farklı optimizasyon ölçütlerini kullanmaktadır. Bu bölümde ele alınan eşikleme teknikleri, yalnızca görüntü

bölütleme amacıyla değil, aynı zamanda renk uzayının alt bölgelere ayrılması yoluyla renk indirgeme sürecine temel oluşturacaktır.

Kapur [21] (entropi) tarafından önerilen amaç fonksiyonu;

$$J(t) = H_x + H_y \quad (1)$$

şeklinde ifade edilmektedir. Burada  $H_x$  ve  $H_y$  histogramın entropilerini temsil etmekte olup,

$$H_x = -\sum_{i=0}^{t-1} \frac{p_i}{\omega_0} \ln \frac{p_i}{\omega_0}, \omega_0 = \sum_{i=0}^{t-1} p_i \quad (2)$$

$$H_y = -\sum_{i=t}^{255} \frac{p_i}{\omega_1} \ln \frac{p_i}{\omega_1}, \omega_1 = \sum_{i=t}^{255} p_i$$

$$p_i = h_i / (M \times N)$$

olarak hesaplanmaktadır. Burada  $h_i$  i. seviyede piksel sayısını temsil ederken,  $p_{i\text{ ilgili}}$  pikselin bulunma olasılığıdır. Otsu [22] yaklaşımının amaç fonksiyonu ise;

$$j(t) = \sigma_x + \sigma_y \quad (3)$$

şeklinde ifade edilmektedir. Söz konusu eşitlikte  $\sigma$  değerleri sınıflar arası varyansı temsil etmektedir.

$$\sigma_x = \omega_0(\mu_0 - \mu_t) \text{ ve } \sigma_y = \omega_1(\mu_1 - \mu_t) \quad (4)$$

şeklinde formüle edilmiştir. Histogramın ağırlık merkezini [23] dikkate alan eşikleme yaklaşımı da başarılı metotlardan biridir. Kısmi olasılık dağılımlarının toplamı ve kısmi ortalamalar Eşitlik 5'teki gibi hesaplanmaktadır.

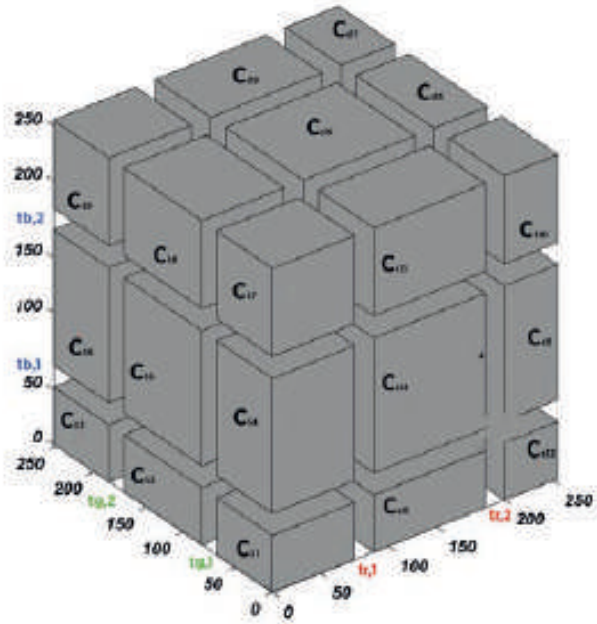
$$\omega_0 = \sum_{i=0}^{\mu_t-1} h_i, \mu_0 = \sum_{i=0}^{\mu_t-1} \frac{ih_i}{\omega_0} \quad (5)$$

$$\omega_1 = \sum_{i=\mu_t}^{255} h_i, \mu_1 = \sum_{i=\mu_t}^{255} \frac{ih_i}{\omega_1}$$

### 3. Eşikleme Teknikleri ile Renk Uzayı Bölütleme ve Renk İndirgeme

Bu çalışmada renk indirgeme prosedürü, RGB renk uzayının her bir kanalına ayrı ayrı uygulanan çok seviyeli histogram tabanlı eşikleme yöntemleri ile gerçekleştirilmiştir. Bu kapsamda görüntüye ait kırmızı (R), yeşil (G) ve mavi

(B) kanallarının her biri bağımsız bir gri seviye görüntü olarak ele alınmış ve her kanal için histogram dağılımı elde edilmiştir. Elde edilen histogramlar üzerinden Otsu, Kapur ve histogramın ağırlık merkezi tabanlı eşikleme yöntemleri kullanılarak optimum eşik değerleri belirlenmiştir. Çok seviyeli eşikleme sürecinde, her bir renk kanalına ait yoğunluk değerleri  $L=256$  gri seviye olacak şekilde aralığında tanımlanmıştır. Seçilen eşikleme yöntemine bağlı olarak  $k$  adet eşik değeri  $T = \{t_1, t_2, \dots, t_k\}$  şeklinde elde edilmiştir. Bu eşik değerler ile renk kanalı,  $c_0 = [0, t_1 - 1], c_1 = [t_1, t_2 - 1], \dots, c_k = [t_k, L - 1]$  olmak üzere  $k+1$  adet sınıfa ayrılmıştır. Tüm renk kanalları kombinasyonu ile renk uzayı  $C_{altisinf} = (k+1)^3$  kadar alt kümeye ayrılmıştır. Şekil 1 bölütlenmiş renk uzayını temsil etmektedir. Bahsi geçen bu işlemler referans [21], [24] ve [25]'de de yapılmıştır. Ancak ilgili çalışmalarda yapılan renk indirgeme yöntemlerinin detaylı bir karşılaştırılması bulunmamaktadır.






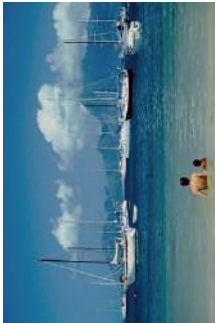





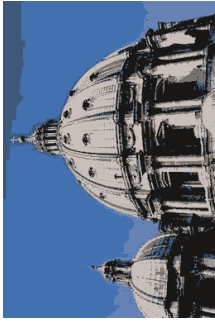

Şekil 1. Bölütlenmiş renk uzayı [20]

#### 4. Deneysel Bulgular ve Yorumlar

Önerilen çok seviyeli eşikleme tabanlı renk indirgeme yaklaşımının performansı, yaygın olarak kullanılan Corel-1000 veri tabanından rastgele seçilen 10 farklı doğal görüntü üzerinde deneysel olarak değerlendirilmiştir. Çalışmada, RGB renk uzayını oluşturan her bir kanal bağımsız olarak ele alınmış ve Otsu, Kapur entropi ve histogramın ağırlık merkezi tabanlı eşikleme

yöntemleri kullanılarak 2 eşik seviyesinde bölütleme işlemi gerçekleştirilmiştir. İki eşik değeri kullanıldığında her bir renk kanalı üç alt sınıfa ayrılmış ve kanalların birleşimi ile toplamda maximum 27 renk seviyeli bir temsil elde edilmiştir. Benzer şekilde 3 eşik için gerçekleştirilen sonuçlarda ise renk uzayı 64 alt sınıfa ayrılmıştır. Gerçekleştirilen deneysel çalışmalar kapsamında, farklı eşikleme yöntemlerinin renk dağılımı, sınıf içi homojenlik, görsel kalite ve detay koruma başarımı üzerindeki etkileri karşılaştırmalı olarak incelenmiştir. Ayrıca eşik sayısının artırılmasının renk temsil gücü ve veri indirgeme oranı üzerindeki rolü analiz edilmiştir. Bu değerlendirmeler sayesinde, varyans tabanlı, entropi tabanlı ve ortalama tabanlı eşikleme yaklaşımlarının RGB renk uzayını bölütleme karakteristikleri ortaya konularak hangi yöntemin hangi koşullar altında daha başarılı sonuçlar ürettiği tartışılmıştır. Yöntemlerin performanslarının nesnel olarak karşılaştırılabilmesi amacıyla tepe sinyal-gürültü oranı (PSNR) ve yapısal benzerlik indeksi (SSIM) metrikleri kullanılmıştır. Tablo 1’de Corel 1K verisetinden seçilen rastgele orijinal görüntüler ve yöntemlere ait iki eşik ile olan indirgeme sonuçları görsel olarak verilmiştir. Tablo 2 ve Tablo 3’te ise sırayla iki ve üç eşiklere ait performanslar verilmiştir.

Tablo 1. Görüntülerin farklı eşikleme teknikleri ile indirgenmiş görsel çıktıları

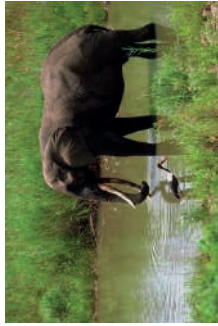
Görüntü No	Orjinal	Otsu	Kapur	Ağırlık Merkezi
1				
2				
3				



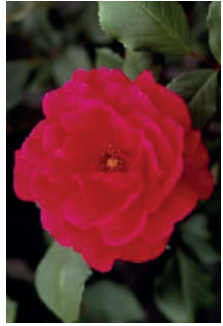
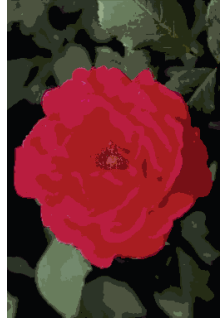
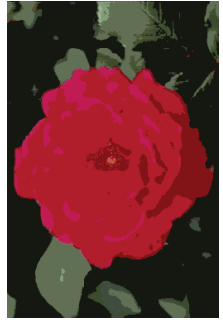
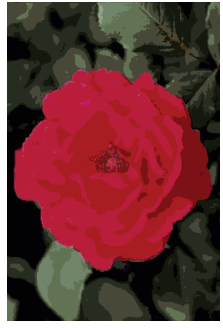
4



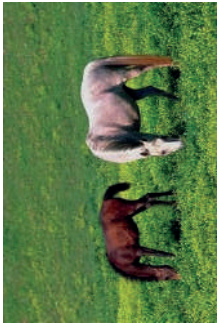
5



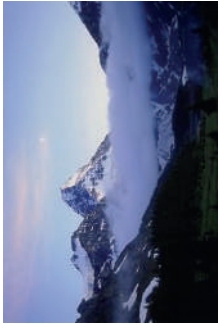
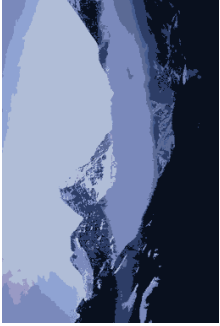
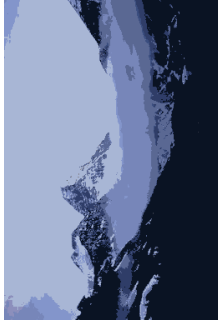
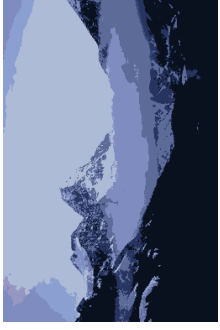
6



7



8



9



10

**Tablo 2. Renk indirgeme algoritmalarının karşılaştırmalı performansları (2 eşik)**

Görüntü No	Otsu			Kapur			Ağırlık Merkezi		
	SSIM	PSNR	Renk Sayısı	SSIM	PSNR	Renk Sayısı	SSIM	PSNR	Renk Sayısı
1	0,9482	23,0852	20	0,9476	23,028	20	0,9438	22,7413	20
2	0,924	25,1837	16	0,9195	24,6445	16	0,9135	24,5761	17
3	0,9645	24,5579	14	0,9602	24,0351	12	0,9652	24,6017	16
4	0,9695	23,188	20	0,9745	24,1589	18	0,9707	23,421	20
5	0,9902	28,6592	10	0,98	25,682	10	0,9846	26,6115	15
6	0,9482	24,6864	16	0,9362	23,8153	16	0,9458	24,4828	16
7	0,9621	28,4616	19	0,909	24,716	18	0,9575	27,9412	19
8	0,9323	25,0614	20	0,913	24,0673	15	0,9305	24,96	19
9	0,9839	25,6371	9	0,9791	24,5375	14	0,9823	25,2206	13
10	0,9535	22,6336	20	0,9481	22,1679	19	0,9512	22,4056	20

**Tablo 3. Renk indirgeme algoritmalarının karşılaştırmalı performansları (3 eşik)**

Görüntü No	Otsu			Kapur			Ağırlık Merkezi		
	SSIM	PSNR	Renk Sayısı	SSIM	PSNR	Renk Sayısı	SSIM	PSNR	Renk Sayısı
1	0,9715	25,6008	39	0,9679	25,0934	36	0,9697	35,3595	42
2	0,963	27,9132	31	0,9442	26,4053	29	0,9615	27,6127	31
3	0,971124	28,2895	22	0,9792	26,7975	22	0,9851	28,2365	22
4	0,9846	26,1936	36	0,9809	25,082	38	0,9838	36,011	36
5	0,9948	31,5546	16	0,9889	28,5976	19	0,9899	28,4671	23
6	0,9689	26,8454	28	0,9612	25,8823	23	0,9683	26,7557	26
7	0,9791	30,7847	36	0,966	28,3792	31	0,9992	27,8748	42
8	0,9626	27,5591	34	0,9413	25,7927	25	0,9822	29,5229	45
9	0,989	27,315	17	0,9905	27,9105	22	0,9905	27,9946	20
10	0,9741	25,0706	38	0,9718	24,7133	34	0,9753	25,2704	41

Tablo 2 ve Tablo 3 birlikte değerlendirildiğinde, eşik sayısındaki artışın tüm yöntemler için performansı belirgin biçimde iyileştirdiği görülmektedir. İki eşik kullanılan durumda elde edilen SSIM ve PSNR değerlerinin, üç eşik kullanılan duruma kıyasla daha düşük kaldığı; buna karşılık üç eşik ile hem yapısal benzerliğin hem de tepe sinyal-gürültü oranının yükseldiği tespit edilmiştir. Bu artış, renk uzayının daha fazla alt kümeye ayrılması sayesinde görüntünün özgün yapısının daha başarılı temsil edilmesinden kaynaklanmaktadır. İki eşik durumunda yöntemler karşılaştırıldığında, SSIM ölçütüne göre Otsu

yönteminin çoğu görüntüde en yüksek ya da en yüksek değere çok yakın sonuçlar ürettiği görülmektedir. Kapur yöntemi bazı görüntülerde Otsu yöntemini geride bıraksa da genel eğilim Otsu'nun daha kararlı ve yüksek yapısal benzerlik sağladığını göstermektedir. Ağırlık merkezi yöntemi ise SSIM açısından genellikle diğer iki yöntemin bir miktar gerisinde kalmıştır. PSNR değerleri açısından da benzer bir durum söz konusu olup, iki eşik seviyesinde Otsu yöntemi çoğu görüntüde en yüksek veya en yüksek değere çok yakın sonuçlar üretmiştir. Üç eşik kullanılan durumda ise tüm yöntemlerin performansının yükseldiği, ancak yöntemler arasındaki farkların kısmen azaldığı gözlenmiştir. SSIM değerleri incelendiğinde, bazı görüntülerde Kapur ve ağırlık merkezi yöntemlerinin Otsu yöntemini geçtiği ve özellikle ağırlık merkezi yönteminin belirli görüntülerde oldukça yüksek yapısal benzerlik değerlerine ulaştığı görülmektedir. Bununla birlikte Otsu yöntemi genel ortalama dikkate alındığında yine yüksek ve dengeli sonuçlar üretmeye devam etmiştir. PSNR bakımından değerlendirildiğinde üç eşik seviyesinde en dikkat çekici artışın ağırlık merkezi yönteminde gerçekleştiği ve bazı görüntülerde diğer yöntemlere kıyasla daha yüksek yeniden yapılandırma kalitesi sağladığı belirlenmiştir. Renk sayıları açısından değerlendirme yapıldığında, iki eşik durumunda yöntemlerin benzer sayıda renk ile temsil sağladığı, üç eşik durumunda ise özellikle ağırlık merkezi yönteminin daha fazla renk içeren bir temsil oluşturduğu görülmektedir. Bu durum, söz konusu yöntemin görüntü detaylarını koruma eğiliminin daha yüksek olduğunu, ancak renk indirgeme oranı açısından diğer yöntemlere göre daha düşük bir sıkıştırma sağladığını göstermektedir.

Genel olarak değerlendirildiğinde, düşük eşik sayılarında Otsu yönteminin hem SSIM hem de PSNR açısından daha kararlı ve başarılı sonuçlar ürettiği, eşik sayısı arttıkça Kapur ve ağırlık merkezi yöntemlerinin rekabetçi hale geldiği ve özellikle ağırlık merkezi yönteminin yüksek görüntü kalitesi sağladığı sonucuna ulaşılmıştır. Bu bulgular, eşik sayısı ile görüntü kalitesi arasında güçlü bir ilişki bulunduğunu ve uygun eşik sayısı ile yöntem seçiminin uygulamanın hedeflerine (yüksek kalite ya da yüksek indirgeme oranı) bağlı olarak yapılması gerektiğini göstermektedir.

Tablolardaki değerlerin genel ortalamaları dikkate alındığında, eşik sayısındaki artışın tüm yöntemler için performans kazanımı sağladığını göstermektedir. İki eşik durumunda SSIM değerleri yöntemlere göre yaklaşık olarak 0.94–0.96 iken, üç eşik seviyesinde bu değerler 0.96–0.99 seviyesine yükseldiği görülmektedir. Benzer şekilde PSNR değerleri iki eşikte 23–26 dB aralığında kalırken, üç eşikte 25–31 dB seviyelerine ulaşarak önemli bir artış göstermiştir. Bu artış, yeniden oluşturulan görüntü kalitesinde gözle görülür bir iyileşmeye karşılık gelmektedir. Yöntem bazında ortalama değerlere bakıldığında iki eşik seviyesinde Otsu'nun hem SSIM hem PSNR açısından en yüksek ortalamaya

sahip olduğu, Kapur yönteminin buna çok yakın ancak biraz daha düşük kaldığı, ağırlık merkezi yönteminin ise özellikle SSIM bakımından sınırlı bir farkla diğerlerinin gerisinde yer aldığı belirlenmiştir. Üç eşik seviyesinde ise yöntemler arasındaki farkların azaldığı, hatta bazı görüntülerde Kapur ve ağırlık merkezi yöntemlerinin Otsu'yu geçtiği, ancak genel ortalama dikkate alındığında Otsu'nun dengeli ve yüksek performansını koruduğu görülmektedir. Renk sayılarının ortalaması incelendiğinde iki eşikten üç eşik seviyesine geçişte yaklaşık iki katına varan bir artış olduğu ve özellikle ağırlık merkezi yönteminin daha fazla renk ile temsil sağlayarak en yüksek PSNR değerlerine ulaştığı dikkat çekmektedir. Bu bulgular birlikte değerlendirildiğinde, düşük eşik sayılarında Otsu'nun kalite–indirgeme dengesi açısından daha avantajlı olduğu, eşik sayısı arttıkça ise yöntemler arasındaki performans farkının azaldığı ve yüksek görüntü kalitesi hedeflenen uygulamalarda ağırlık merkezi ve Kapur yöntemlerinin de güçlü alternatifler haline geldiği sayısal olarak ortaya konmuştur.

## 5. Sonuç

Bu çalışmada, Otsu, Kapur ve histogramın ağırlık merkezi yöntemleri kullanılarak gerçekleştirilen çok seviyeli eşikleme tabanlı renk indirgeme işlemlerinin performansı iki ve üç eşik durumu için karşılaştırmalı olarak incelenmiştir. Elde edilen bulgular, eşik sayısının artırılmasının tüm yöntemler için hem SSIM hem de PSNR değerlerinde belirgin iyileşme sağladığını ve buna bağlı olarak yeniden oluşturulan görüntü kalitesini artırdığını göstermiştir. İki eşik seviyesinde Otsu yöntemi kalite–renk indirgeme dengesi açısından daha kararlı ve yüksek ortalama performans sunarken, üç eşik seviyesinde yöntemler arasındaki farkın azaldığı ve özellikle ağırlık merkezi yönteminin daha yüksek PSNR değerleri ile dikkat çektiği belirlenmiştir. Bununla birlikte artan eşik sayısının renk sayısını yükselterek indirgeme oranını azalttığı görülmüştür. Sonuç olarak, düşük renk sayısı ile kabul edilebilir görüntü kalitesinin hedeflendiği uygulamalarda Otsu yöntemi öne çıkarken, daha yüksek görsel kalite gerektiren durumlarda eşik sayısının artırılmasıyla birlikte Kapur ve ağırlık merkezi yöntemlerinin de etkili alternatifler sunduğu ortaya konmuştur. Bu bulgular, renk indirgeme problemlerinde yöntem ve eşik sayısı seçiminin uygulama gereksinimlerine bağlı olarak yapılması gerektiğini göstermektedir.

## Kaynakça

- [1] M. E. Celebi, "Forty years of color quantization: a modern, algorithmic survey," *Artificial Intelligence Review*, vol. 56, no. 12, pp. 13953–14034, 2023.
- [2] M.O. İncetaş, and M Meriçelli, "Renk İndirgeme ve Kenar Belirleme Yardımıyla SRG Tabanlı Görüntü Bölütleme", *Mühendislik Uygulamalarında Yenilikçi ve Multidisipliner Çalışmalar: Özgür Yayınları*, 2024, pp. 75-84.
- [3] M.O. İncetaş, "Content-Based Image Retrieval Based on Block Truncation Code using Fuzzy-C Means Quantization and Edge Detection", *Current Research in Engineering*, Gece Kitaplığı, 2023, pp. 289-303.
- [4] M.O. İncetaş, M. Kiliçaslan, and T.R. Farshi, "Image retrieval with SNN-based multi-level thresholding", *Gümüşhane Üniversitesi Fen Bilimleri Dergisi*, pp. 98–108, 2022.
- [5] M.O. İncetaş, R. Demirci, and H.G. Yavuzcan, "Automatic segmentation of color images with transitive closure". *AEU-international Journal of Electronics and Communications*, 68(3), pp. 260–269, 2014.
- [6] M.O. İncetaş, E. Veske, N. Emre, and R. Demirci, "Automatic cells counting in Natt-Herrick stained fish blood", *Aquaculture Studies*, 17(3), pp. 283-294, 2017.
- [7] M.O. İncetaş, R. Demirci, and H.G. Yavuzcan, "Automatic color edge detection with similarity transformation", *Gazi University Journal of Science*, 32(2), pp. 458–469, 2019.
- [8] A. Abernathy and M. E. Celebi, "The incremental online k-means clustering algorithm and its application to color quantization," *Expert Systems with Applications*, vol. 207, p. 117927, 2022.
- [9] L. Brun and A. Trémeau, "Color quantization," in *Digital color imaging handbook*: CRC press, 2017, pp. 589–637.
- [10] Z. Xiang, "Color quantization," in *Handbook of approximation algorithms and metaheuristics*: Chapman and Hall/CRC, 2018, pp. 691–709.
- [11] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," *ACM computing surveys (CSUR)*, vol. 31, no. 3, pp. 264–323, 1999.
- [12] P. Heckbert, "Color image quantization for frame buffer display," *ACM Siggraph Computer Graphics*, vol. 16, no. 3, pp. 297–307, 1982.
- [13] M. Gervautz and W. Purgathofer, "A simple method for color quantization: Octree quantization," in *New Trends in Computer Graphics: Proceedings of CG International'88*, 1988: Springer, pp. 219–231.
- [14] M. E. Celebi, "Fast color quantization using weighted sort-means clustering," *Journal of the Optical Society of America A*, vol. 26, no. 11, pp. 2434–2443, 2009.

- [15] M. E. Celebi, "Improving the performance of k-means for color quantization," *Image and Vision Computing*, vol. 29, no. 4, pp. 260–271, 2011.
- [16] S.-C. Huang, "An efficient palette generation method for color image quantization," *Applied Sciences*, vol. 11, no. 3, p. 1043, 2021.
- [17] S. Thompson, M. E. Celebi, and K. H. Buck, "Fast color quantization using MacQueen's k-means algorithm," *Journal of Real-Time Image Processing*, vol. 17, no. 5, pp. 1609–1624, 2020.
- [18] L. Szilágyi, G. Dénesi, and C. Enăchescu, "Fast color quantization via fuzzy clustering," in *International Conference on Neural Information Processing*, 2016: Springer, pp. 95–103.
- [19] Q. Wen and M. E. Celebi, "Hard versus fuzzy c-means clustering for color quantization," *EURASIP Journal on Advances in Signal Processing*, vol. 2011, no. 1, p. 118, 2011.
- [20] M. Kılıçaslan and M. O. İncetaş, "Adaptive color quantization method with multi-level thresholding," *International Journal of Computational Intelligence Systems*, vol. 16, no. 1, p. 7, 2023.
- [21] J. N. Kapur, P. K. Sahoo, and A. K. Wong, "A new method for gray-level picture thresholding using the entropy of the histogram," *Computer vision, graphics, and image processing*, vol. 29, no. 3, pp. 273–285, 1985.
- [22] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, no. 285-296, 1979.
- [23] R. Demirci and Ü. Okur, "Renkli Görüntülerin Ortalama Tabanlı Çok Seviyeli Eşiklenmesi," *Duzce University Journal of Science and Technology*, vol. 7, no. 1, pp. 664–676, 2019.
- [24] M. Kılıçaslan, U. Tanyeri, and R. Demirci, "Renkli görüntüler için tek boyutlu histogram," *Duzce University Journal of Science and Technology*, vol. 6, no. 4, pp. 1094–1107, 2018.
- [25] M. Kılıçaslan, U. Tanyeri, and R. Demirci, "Tekrarlı Ortalama Yardımıyla Renk İndirgeme ve Görüntü Erişimi," *Duzce University Journal of Science and Technology*, vol. 8, no. 1, pp. 1042–1057, 2020.



# **Bilgisayar Bilimleri ve Mühendisliđi Alanında Güncel Arařtırmalar**

**Editör:**

**Doç. Dr. Mürsel Ozan İNCETAŐ**