YAPAY ZEKÂ VE DİJİTAL TEKNOLOJİLERDE GÜNCEL ARAŞTIRMALAR:

ALGORİTMALAR, GÖRÜNTÜ İŞLEME VE AKILLI UYGULAMALAR

EDİTOR
PROF. DR. EYYÜP GÜLBANDILAR





BİDGE Yayınları

BİDGE Yayınları

Yapay Zekâ ve Dijital Teknolojilerde Güncel Araştırmalar: Algoritmalar, Görüntü İşleme Ve Akıllı Uygulamalar

Editör: PROF. DR. EYYÜP GÜLBANDILAR

ISBN: 978-625-5057-86-0

1. Baskı
Sayfa Düzeni: Gözde YÜCEL
Yayınlama Tarihi: 2026-10-25
BİDGE Yayınları

Bu eserin bütün hakları saklıdır. Kaynak gösterilerek tanıtım için yapılacak kısa alıntılar dışında yayıncının ve editörün yazılı izni olmaksızın hiçbir yolla çoğaltılamaz.

Sertifika No: 71374
Yayın hakları © BİDGE Yayınları
www.bidgeyayinlari.com.tr - bidgeyayinlari@gmail.com
Krc Bilişim Ticaret ve Organizasyon Ltd. Şti.
Güzeltepe Mahallesi Abidin Daver Sokak Sefer Apartmanı No: 7/9 Çankaya / Ankara

İÇİNDEKİLER

DERİN ÖĞRENME TABANLI YOLO11 MİMARİSİ İLE KUMAŞ HATA TESPİTİ 1
İLKER YILDIZ, FİLİZ YILDIZ

ZOR TRAKEAL ENTÜBASYONUN TAHMİNİNDE YAPAY ZEKÂ YAKLAŞIMLARI: KLİNİK VERİLERDEN GÖRÜNTÜ TABANLI DERİN ÖĞRENMEYE 47
NACİYE NUR ARSLAN

UYDU GÖRÜNTÜSÜ VERİ SETLERİNDE BENZERLİK TESPİTİ İÇİN HİBRİT SİSTEM 72
MUSTAFA KURÇ, AYHAN İSTANBULLU

ÜRETKEN YAPAY ZEKÂ ÇAĞINDA KATALOGLAMA VE SINIFLANDIRMA: OLANAKLAR, SINIRLILIKLAR VE İNSAN DENETİMİ 91
HÜLYA AKYÜZ

YOLO MODELLERİ İLE FİŞ ANALİZİNDE SAHTECİLİK TESPİTİ 180
TOLGA TUNÇEL, EMRE KILINÇ

ÜRETKEN GÖRSEL MODELLERİN KİŞİSELLEŞTİRİLMESİNDE YENİDEN ÜRETİLEBİLİRLİK VE DEĞERLENDİRME 213
FURKAN GÖZÜKARA

ÇOK KRİTERLİ KARAR VERME YÖNTEMLERİNDEN TOPSIS, GRA, VIKOR, WASPAS VE SAW İLE PANDEMİ DÖNEMİ BİST İNŞAAT FİRMALARININ PERFORMANS DEĞERLENDİRMELERİ 247
GÜRKAN KAPLAN

GÖMÜLÜ SİSTEMLERDE DERİN ÖĞRENME: DONANIM PLATFORMLARI, MODEL UYARLAMA VE

İÇİNDEKİLER

PERFORMANS DEĞERLENDİRMESİ 273
VOLKAN YAMAÇLI, SEMİH KAHVECİ, RAMAZAN AKKURT

BÖLÜM 1

DERİN ÖĞRENME TABANLI YOLO11 MİMARİSİ İLE KUMAŞ HATA TESPİTİ

İlker YILDIZ[1]
Filiz YILDIZ[2]

Giriş

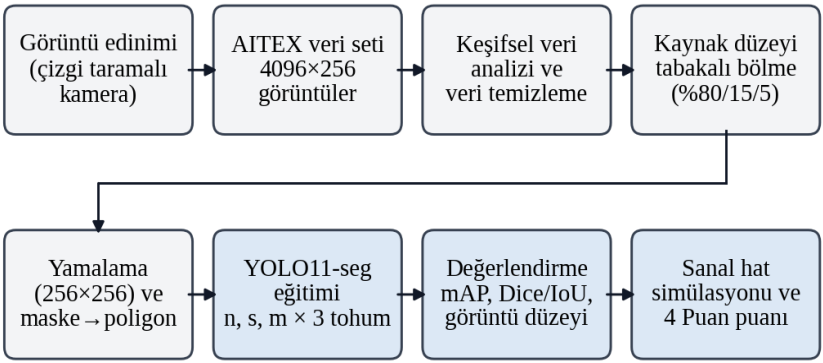
Tekstil üretiminde kalite kontrol, ürünün pazar değerini doğrudan belirleyen kritik bir aşamadır. Kumaş yüzeyindeki hatalar görünümü bozmanın yanı sıra kesim planlamasını güçleştirir, konfeksiyon firesini artırır ve kumaşın ikinci kalite olarak sınıflandırılmasına neden olur; ikinci kalite kumaşın satış fiyatının %45–65 oranında düşebildiği Srinivasan et al. (1992) tarafından bildirilmiş ve sonraki çalışmalarda aktarılmıştır (Habib et al., 2014). Endüstride muayene günümüzde de büyük ölçüde ışıklı masalarda çalışan operatörlerce yapılmaktadır. Yorgunluk ve kişiye bağlı karar ölçütleri nedeniyle insan muayenesinin başarısı %60–70 düzeyinde kalırken otomatik görme sistemlerinin %90'ın üzerinde

[1] Dr. Öğr. Üyesi, Bolu Abant İzzet Baysal Üniversitesi, Bolu Teknik Bilimler Meslek Yüksekokulu, Bilgisayar Teknolojileri Bölümü, Orcid: 0000-0002-1575-2673
[2] Dr. Öğr. Üyesi, Bolu Abant İzzet Baysal Üniversitesi, Bolu Teknik Bilimler Meslek Yüksekokulu, Tekstil, Giyim, Ayakkabı ve Deri Bölümü, Orcid: 0000-0001-7757-8888

başarıya ulaşabildiği bildirilmektedir (Mao & Hong, 2025; Ngan et al., 2011). Hata türlerinin çokluğu ve aynı türdeki hataların biçimsel çeşitliliği kural tabanlı muayeneyi de zorlaştırdığından (Kumar, 2008; Ngan et al., 2011), derin öğrenme tabanlı bilgisayarla görü yöntemleri tekstil kalite kontrolünün başlıca araştırma alanlarından biri haline gelmiştir (Hanbay et al., 2016; Kahraman & Durmuşoğlu, 2023).

Bu bölümde kumaş hata tespiti, kuramsal temellerinden uygulamaya uzanan bütüncül bir çerçevede ele alınmaktadır. Uygulamada, dokuma kumaş görüntülerinden oluşan AITEX veri seti (Silvestre-Blanes et al., 2019) üzerinde, Ultralytics tarafından geliştirilen ve literatürde YOLOv11 olarak da anılan YOLO11 (Jocher & Qiu, 2024; Khanam & Hussain, 2024) örnek bölütleme (instance segmentation) modelleri eğitilmiştir. AITEX üzerinde YOLO ailesiyle yapılan çalışmalar hataları sınırlayıcı kutularla (bounding box) tespit etmeye odaklanırken (Seidel et al., 2022; Waiandt & Seibel Júnior, 2024), bu çalışmada hatalar piksel düzeyinde bölütlenmekte ve maskelerden elde edilen boyutlar 4 Puan Sistemi ile ilişkilendirilmektedir. Çalışmanın katkıları; AITEX veri setinin kaynak görüntü düzeyinde bölünmüş, yeniden üretilebilir bir örnek bölütleme sürümünün hazırlanması, üç model ölçeğinin (n, s, m) çok tohumlu olarak doğruluk ve hız yönünden karşılaştırılması ve tam kumaş şeridi üzerinde maske tabanlı otomatik puanlamanın gösterilmesidir. Çalışmanın uçtan uca iş akışı Şekil 1’de verilmiştir.

Şekil 1. Çalışmanın uçtan uca iş akışı



Kumaş Hataları ve Kalite Değerlendirme Standartları

Dokuma, Örme ve Terbiye Hataları

Kumaş hataları, oluştukları üretim aşamasına göre genel olarak üç grupta incelenir. **Dokuma hataları**, çözgü (warp) ve atkı (weft, pick) ipliklerinin kopması, gevşemesi ya da yanlış bağlantı yapmasından kaynaklanır; kopuk çözgü (broken end), kopuk atkı (broken pick) ve atkı kıvrılması (weft curling) bu grubun tipik örnekleridir. **Örme hataları** ilmek oluşumundaki düzensizliklerden doğar; delik bu grubun tipik bir örneğidir (Das Gupta et al., 2024). **Terbiye hataları** ise boyama ve apre gibi terbiye işlemleri sırasında oluşur. AITEX veri setini hazırlayan çalışmada endüstride izlenen hata türleri kodlarıyla birlikte listelenmiştir; bunların 12'si veri setinde örneklenmiştir (Silvestre-Blanes et al., 2019). Bu 12 hata türü ve bu çalışmada ait oldukları morfolojik grup Tablo 1'de verilmiştir.

Tablo 1. AITEX veri setindeki hata türleri ve bu çalışmada kullanılan morfolojik gruplar

Kod	Hata (İngilizce)	Morfolojik grup
002	Broken end	Çizgisel
006	Broken yarn	Çizgisel
010	Broken pick	Çizgisel
016	Weft curling	Çizgisel
023	Crease	Çizgisel
036	Weft crack	Çizgisel
019	Fuzzy ball	Noktasal
025	Warp ball	Noktasal
027	Knots	Noktasal
029	Contamination	Noktasal
030	Nep	Noktasal
022	Cut selvage	Kenar

Kaynak: Hata kodları ve adları Silvestre-Blanes et al. (2019) çalışmasından alınmış, morfolojik gruplar yazarlar tarafından belirlenmiştir.

Görsel Tespitteki Zorluklar

Kumaş hatalarının görüntüden tespiti, genel nesne tespitinden birkaç yönden ayrılır. İlk olarak hatalar, periyodik ve yoğun bir doku üzerinde yer alır. Doku örüntüsü bazen hatanın kendisinden daha belirgindir, bu da hatanın arka plandan ayrılmasını zorlaştırır (Ngan et al., 2011). İkinci olarak hataların boyut ve biçim yelpazesi çok geniştir: hav topu gibi noktasal hatalar yalnızca birkaç piksel kaplarken, kopuk atkı gibi hatalar kumaşı boydan boya kesebilir. Nitekim bu çalışmadaki veri analizinde maske bileşenlerinin ortanca boyutu 7×8 piksel bulunmuş, en büyük bileşenin genişliği ise 3607 piksele ulaşmıştır. Üçüncü olarak aynı sınıfa ait hatalar farklı kumaşlarda farklı görünür; sınıflar arası benzerlik yüksek, sınıf içi benzerlik düşüktür (Kumar, 2008). Son olarak hatalı örnek sayısı hatasız örneklere göre azdır ve hata türleri arasında belirgin bir dengesizlik bulunur; bu durum özellikle gözetimli öğrenme yöntemleri için ciddi bir güçlük oluşturur (Kahraman & Durmuşoğlu, 2023).

4 Puan Sistemi

Kumaş muayenesinin sonuçlarını sayısal bir kaliteye dönüştürmek için ticarette puanlama sistemleri kullanılır. ASTM D5430 standardı, görsel muayene sonucunda kumaşa sayısal bir derece atanmasına yönelik yöntemleri tanımlamaktadır (ASTM International, 2017). En yaygın kullanılan yöntem olan **4 Puan Sistemi**'nde her hataya, uzunluğuna göre 1 ile 4 arasında ceza puanı verilir; delik ve açıklıklar boyutlarına göre ayrıca puanlanır (Tablo 2). Bir doğrusal yardaya (yd) toplamda en fazla 4 puan verilir.

Tablo 2. 4 Puan Sisteminde hata boyutuna göre ceza puanları

Hata boyutu	**Ceza puanı**
3 inç (≈76 mm) ve altı	1
3 inçten büyük, 6 inç (≈152 mm) ve altı	2
6 inçten büyük, 9 inç (≈229 mm) ve altı	3
9 inçten büyük	4
Delik ve açıklıklar: 1 inç (≈25 mm) ve altı	2
Delik ve açıklıklar: 1 inçten büyük	4

Kaynak: ASTM International (2017) esas alınarak hazırlanmıştır.

Farklı genişlikteki topların karşılaştırılabilmesi için toplam ceza puanı 100 yd² alana normalize edilir:

$$P_{100} = \frac{3600 \times \sum_{i=1}^{n} p_i}{L \times W} \quad (1)$$

Burada P_{100} 100 yd² başına düşen ceza puanını, p_i i'nci hataya verilen puanı, n muayene edilen toplam hata sayısını, L muayene edilen kumaş uzunluğunu (yd), W kesilebilir kumaş genişliğini (inç) göstermektedir. 3600 katsayısı, 100 yd² alanın yarda × inç biriminde karşılığıdır (100 × 36). Bu çalışmada, hata uzunluğuna dayalı olduğu ve doğrudan maske ölçümüyle ilişkilendirilebildiği için 4 Puan Sistemi kullanılmıştır.

Görüntü Edinimi, Veri Seti ve Deneysel Protokol

Kamera ve Aydınlatma

Hareket halindeki kumaşın muayenesinde genellikle çizgi taramalı (line-scan) kameralar tercih edilir. Alan taramalı (area-scan) kameralar bir çerçevenin tamamını aynı anda yakalar. Çizgi taramalı kamera ise tek satırlık bir algılayıcıyla kumaş genişliği boyunca bir satır görüntü alır; kumaş ilerledikçe ardışık satırlar birleşerek kesintisiz bir görüntü oluşturur. Bu yapı, geniş ve sürekli hareket eden yüzeylerde çerçeveler arası örtüşme ve dikiş sorunlarını ortadan kaldırır. AITEX veri setini oluşturan sistemde 4096 piksellik bir çizgi taramalı kamera, doğrusal LED aydınlatma ve kumaş hareketini izleyen bir kodlayıcı (encoder) kullanılmıştır. Kameranın çözünürlüğü milimetre başına dört piksel olarak bildirilmiştir (Silvestre-Blanes et al., 2019).

Kumaş genişliği doğrultusundaki uzamsal çözünürlük, kameranın görüş alanı ve algılayıcıdaki piksel sayısıyla belirlenir:

$$r_x = \frac{W_{GA}}{N_x} \tag{2}$$

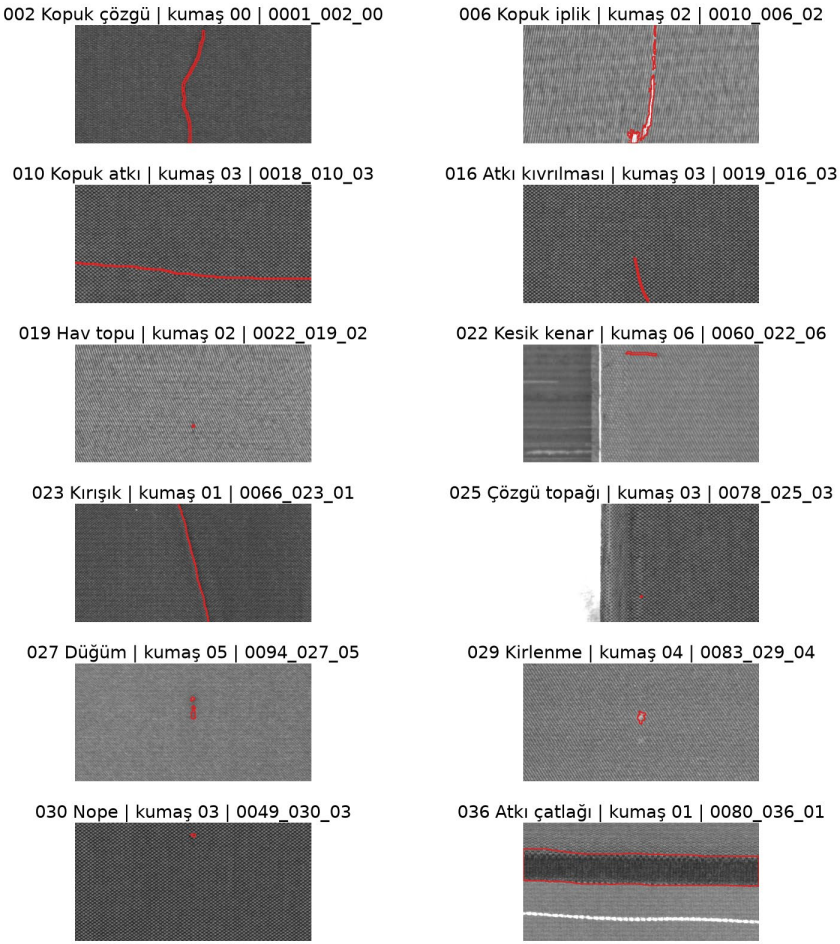
Burada r_x genişlik doğrultusundaki piksel boyutunu (mm/piksel), W_{GA} kameranın görüş alanının genişliğini (mm), N_x algılayıcıdaki piksel sayısını göstermektedir. AITEX sistemi için $N_x = 4096$ ve $r_x = 0{,}25$ mm/piksel (4 piksel/mm) alındığında görüş alanı yaklaşık 1024 mm olarak bulunur.

AITEX Veri Seti

AITEX Fabric Image Database, yedi farklı dokuma kumaş yapısından alınmış 4096×256 piksel boyutunda gri tonlu görüntülerden oluşur. Makalede 140 hatasız ve 105 hatalı olmak üzere 245 görüntü, hatalı görüntüler için elle hazırlanmış bölütleme maskeleri ve 12 hata türü tanımlanmıştır. Hatasız görüntüler her kumaş türünden 20’şer adettir (Silvestre-Blanes et al., 2019).

Veri setinin resmi sürümü indirilerek yapılan keşifsel analizde makaledeki sayılardan bazı sapmalar ve tutarsızlıklar saptanmıştır. Hatalı görüntü klasöründe 106, hatasız görüntü klasöründe 141 dosya ve 107 maske dosyası bulunmuştur. Yedi kumaş türü dışında kalan ve maskesi olmayan bir görüntü (0100_025_08), maskesi tamamen boş bir görüntü (0106_010_03) ve ad biçimi hatalı, fazladan bir hatasız görüntü (0018_00_01) dışlanmış; aynı görüntüye ait iki ayrı maskesi bulunan iki görüntüde maskeler birleştirilmiştir. Ayrıca 92 maskenin kenarlarında ara gri tonlar, bazı dosyalarda saydamlık kanalı ve bir görüntüde farklı genişlik (3796 piksel) bulunmuştur. 147 görüntüde ise kumaş dışı beyaz arka plan yer almaktadır; bu sorunların nasıl ele alındığı Uygulama I'de açıklanmıştır. Hatalı görüntülerin kumaş türlerine dağılımı da dengesizdir: kumaş 02 ve 03 hatalı görüntülerin yaklaşık üçte ikisini oluştururken, kumaş 05'te yalnızca bir hatalı görüntü vardır. Her hata türünden birer örnek Şekil 2'de gösterilmiştir.

Şekil 2. AITEX veri setindeki 12 hata türünden birer örnek (maske sınırları kırmızı ile gösterilmiştir)



Kaynak: Silvestre-Blanes et al. (2019) tarafından yayımlanan veri setinden yazarlar tarafından üretilmiştir.

Deneysel Protokol

Değerlendirmenin güvenilirliği için bölme işlemi yamalamadan **önce** ve **kaynak görüntü düzeyinde** yapılmıştır. Böylece aynı kumaş şeridinden çıkan yamaların farklı kümelere

düşmesi ve modelin test sırasında eğitimde gördüğü dokuyla değerlendirilmesi, yani veri sızıntısı önlenmiştir. Dışlamalardan sonra kalan 244 görüntü (104 hatalı, 140 hatasız), hata grubuna göre tabakalı olarak %80 eğitim, %15 doğrulama ve %5 test kümelerine ayrılmıştır. Tabakalar çizgisel, noktasal ve kenar hata grupları ile hatasız görüntülerden oluşur. Her tabakanın içinde görüntüler kumaş koduna göre sıralanmış ve seçim, rastgele başlangıçlı sistematik örneklemeyle (tohum 42) yapılmıştır. Böylece kumaş türleri de kümelere dengeli dağılmıştır. Çok küçük olan kenar grubu için doğrulama ve test kümelerinde en az birer görüntü bulunması güvence altına alınmıştır. Bölme listeleri bir kez oluşturulmuş ve SHA-256 özet değeriyle kilitlenmiştir; sonraki tüm deneyler aynı listeyi kullanır. Model ve hiperparametre seçimleri yalnızca doğrulama kümesine göre yapılmış, test kümesi yalnızca nihai raporlama için kullanılmıştır. Elde edilen dağılım Tablo 3'te verilmiştir.

Tablo 3. Kaynak görüntülerin kümelere dağılımı

Grup	**Eğitim**	**Doğrulama**	**Test**	**Toplam**
Çizgisel	28	5	2	35
Noktasal	48	9	3	60
Kenar	7	1	1	9
Hatasız	112	21	7	140
Toplam	**195 (%79,9)**	**36 (%14,8)**	**13 (%5,3)**	**244**

Kaynak: Yazarlar tarafından hazırlanmıştır.

Piksel Ölçeği ve Varsayımlar

Maskeden ölçülen hata boyutlarının 4 Puan Sistemi ile puanlanabilmesi için piksellerin fiziksel uzunluğa dönüştürülmesi gerekir. AITEX için bildirilen milimetre başına dört piksel değeri (Silvestre-Blanes et al., 2019) esas alınarak dönüşüm şöyle yapılmıştır:

$$\ell_{mm} = \ell_{px} \cdot r$$
$$\ell_{in} = \frac{\ell_{mm}}{25{,}4} \qquad (3)$$

Burada ℓ_{px} hatanın piksel cinsinden uzunluğunu, r piksel boyutunu (0,25 mm/piksel), ℓ_{mm} ve ℓ_{in} ise hatanın sırasıyla milimetre ve inç cinsinden uzunluğunu göstermektedir.

Bu dönüşüm, bildirilen çözünürlüğün tüm görüntüler için geçerli olduğu ve piksellerin kare olduğu (hareket doğrultusunda da 0,25 mm/piksel) varsayımlarına dayanır; bulgular bu varsayımlar çerçevesinde yorumlanmalıdır.

Bu varsayımlar altında 4096×256 piksellik bir görüntü, kumaş üzerinde yaklaşık 1024 mm × 64 mm'lik bir alana karşılık gelir. Maske bileşenlerinin ortanca boyutu olan 7×8 piksel ise yaklaşık 1,8 mm × 2,0 mm'dir.

Kuramsal Arka Plan

Kumaş hatalarının otomatik olarak tespit edilmesi üzerine yapılan çalışmalar, kullanılan temsil biçimine göre iki ana döneme ayrılabilir. İlk dönemde görüntüden elle tasarlanmış doku öznitelikleri çıkarılmış ve bu öznitelikler istatistiksel ya da makine öğrenmesi yöntemleriyle sınıflandırılmıştır. İkinci dönemde ise öznitelikler, derin sinir ağları tarafından doğrudan veriden öğrenilmeye başlanmıştır (Hanbay et al., 2016).

Klasik Yöntemler

Dokuma kumaşlar, atkı ve çözgü ipliklerinin düzenli tekrarından oluşan periyodik bir dokuya sahiptir. Klasik yöntemler, hatanın bu düzeni yerel olarak bozduğu varsayımıyla dokuyu sayısal özniteliklerle tanımlar ve bu özniteliklerin normalden sapmasını arar (Hanbay et al., 2016; Kumar, 2008; Ngan et al., 2011). İstatistiksel yaklaşımların en yaygın örneği, belirli yön ve uzaklıktaki piksel çiftlerinin gri düzey birlikteliklerinden doku öznitelikleri çıkaran gri düzey eş oluşum matrisidir (GLCM). Spektral yaklaşımlarda Gabor filtreleri, belirli bir yön ve frekansa duyarlı oldukları için atkı ve çözgü doğrultusundaki kopukluklara

güçlü tepki verir. Bu doku seçiciliğinden güncel modellerde de yararlanılmaktadır; örneğin Xu et al. (2023) Swin-Unet tabanlı bir bölütleme modelinin ilk katmanına Gabor filtresi eklemiştir. Dalgacık dönüşümü görüntüyü farklı ölçek ve yönlerde alt bantlara ayırarak konum bilgisini koruyan bir çözümleme sunar. Yerel ikili örüntüler (LBP) ise her pikselin komşularını merkez piksele göre ikili olarak kodlayan, aydınlatma değişimlerine görece dayanıklı bir tanımlayıcıdır; LBP histogramları genellikle destek vektör makineleri (SVM) gibi sınıflandırıcılarla kullanılır.

Klasik yöntemler az veriyle çalışabilir ve kararları yorumlanabilir. Buna karşılık öznitelikleri ve eşik değerleri çoğunlukla belirli bir kumaş türüne göre ayarlanır, doku ya da desen değiştiğinde yeniden ayarlanmaları gerekir ve çoğunlukla hatanın piksel düzeyindeki kesin sınırlarını vermezler (Hanbay et al., 2016; Kumar, 2008). Derin öğrenmeye geçişin temel gerekçesi bu genelleme sorunudur.

Evrişimli Sinir Ağlarının Temelleri

Evrişimli sinir ağları (convolutional neural networks, CNN), öznitelikleri elle tasarlamak yerine eğitim verisinden öğrenir (LeCun et al., 1998). Temel yapı taşı olan evrişim katmanında öğrenilebilir küçük çekirdekler görüntü üzerinde kaydırılır ve her konumda ağırlıklı bir toplam hesaplanır. Aynı çekirdeğin her konumda kullanılması (ağırlık paylaşımı) parametre sayısını azaltır ve bir hatanın görüntünün neresinde olursa olsun aynı biçimde tanınmasını sağlar. Katmanlar derinleştikçe ilk katmanlar kenar ve doku gibi yerel örüntüleri, derin katmanlar ise daha geniş bağlamı temsil eder. Adımlı evrişim ve havuzlama bu kapsama alanını genişletirken çözünürlüğü düşürür; bu da AITEX'teki gibi birkaç piksel genişliğindeki hataların derin katmanlarda kaybolmasına yol açabilir. Bu nedenle modern tespit modelleri, farklı derinlikteki

öznitelik haritalarını birleştiren öznitelik piramitleri kullanır (Lin et al., 2017).

Sınıflandırma, Tespit ve Örnek Bölütleme

Derin öğrenme tabanlı kumaş muayenesi, çıktının ayrıntı düzeyine göre üç görev türünde ele alınabilir (Kahraman & Durmuşoğlu, 2023). **Görüntü sınıflandırma** (image classification) bir görüntünün ya da yamanın hatalı olup olmadığını belirler, ancak hatanın yerini vermez; Kahraman ve Durmuşoğlu (2022) kapsül ağlarıyla, Geze ve Akbaş (2024) ise önceden eğitilmiş CNN'lerle bu görevi ele almıştır. **Nesne tespiti** (object detection) her hatayı bir sınırlayıcı kutu ve sınıf etiketiyle konumlandırır; YOLO gibi tek aşamalı yöntemler konum ve sınıfı tek bir ileri geçişte tahmin ettiklerinden iki aşamalı yöntemlere göre genellikle daha hızlıdır (Redmon et al., 2016). **Bölütleme** (segmentation) ise her pikselin hataya ait olup olmadığını belirler: anlamsal bölütleme (semantic segmentation) pikselleri yalnızca sınıflara ayırır ve U-Net bu alanın temel mimarisidir (Ronneberger et al., 2015); örnek bölütleme (instance segmentation) ise aynı sınıftaki ayrı hataları da birbirinden ayırarak her birine ayrı bir maske verir (Bolya et al., 2019).

4 Puan Sistemi her hatanın ayrı ayrı sayılmasını ve uzunluğunun ölçülmesini gerektirir. Sınıflandırma bu bilgiyi vermez; kutu tabanlı tespit, eğik veya kıvrımlı çizgisel hatalarda gerçek uzunluğu yansıtmaz; anlamsal bölütleme ise birbirine değen hataları ayırt etmez. Bu çalışmada örnek bölütlemenin seçilmesinin nedeni, her hatanın hem ayrı bir nesne olarak sayılmasına hem de maskesinden fiziksel boyutunun ölçülmesine olanak tanımasıdır.

Dikkat Mekanizmaları ve Transformer Modelleri

Evrişim katmanları yerel bir pencereye baktığından, görüntünün birbirinden uzak bölgeleri arasındaki ilişkileri ancak çok sayıda katmandan sonra modelleyebilir. Transformer

mimarisindeki öz-dikkat (self-attention) mekanizması ise her konumun diğer tüm konumlarla ilişkisini doğrudan hesaplar (Vaswani et al., 2017). Bu modeller geniş bağlamı iyi modeller, ancak genellikle çok veri ve yüksek hesaplama gücü gerektirir; kumaş hatası literatüründe bu nedenle karma yaklaşımlar öne çıkmaktadır. Mao ve Hong (2026), YOLOv11'in omurga ve boyun kısmına çok başlı öz-dikkat modülleri ekleyerek DETR tabanlı modellerle karşılaştırmıştır. Bu çalışmada kullanılan YOLO11 da evrişim tabanlı yapısının içine hafif bir dikkat bloğu (C2PSA) yerleştirir (bkz. YOLO11-Seg Mimarisi).

Yalnızca hatasız örneklerle eğitilip bu örneklerin dağılımından sapan bölgeleri hata olarak işaretleyen anomali tespiti yöntemleri de kumaş alanında kullanılmaktadır (Thomine & Snoussi, 2024; Zhou et al., 2021); ancak bu yöntemler hata türünü vermediğinden bu çalışmada uygulanmamıştır.

Literatürdeki Çalışmaların Karşılaştırılması

AITEX veri setini kullanan seçilmiş çalışmalar Tablo 4'te özetlenmiştir. Tablodaki sonuçlar yazarların kendi raporlarından aktarılmıştır. Çalışmalar farklı veri alt kümeleri (örneğin yalnızca hatalı görüntüler), farklı yamalama ve veri artırma işlemleri, farklı bölme oranları ve farklı ölçütler kullandığından sayılar birbirleriyle ve bu çalışmanın sonuçlarıyla doğrudan karşılaştırılamaz.

Tablo 4. AITEX veri setini kullanan seçilmiş çalışmalar

Çalışma	Yöntem	Görev	Veri ve protokol	Bildirilen sonuç
Zhou et al. (2021)	VAE + Gauss karışım modeli	Anomali tespiti (yarı gözetimli)	AITEX ve DAGM 2007	Karşılaştırılan yöntemlerden genel olarak daha iyi başarım
Seidel et al. (2022)	YOLOv5	Kutu tespiti	AITEX, yalnızca hatalı görüntüler; tek örnekli sınıflar çıkarılmış; veri artırma; %65/20/15	mAP50-95: %56,06; kesinlik: %89,66; duyarlılık: %79,66
Cheng et al. (2023)	SCUNet (U-Net türevi)	Anlamsal bölütleme	AITEX'ten 106 görüntü, kırpma ve veri artırma	Doğruluk: %98,01; mIoU: %34,32
Xu et al. (2023)	Swin-Unet + Gabor filtresi	Anlamsal bölütleme	Kırpılmış AITEX (yalnızca hatalı görüntüler, 256×256)	IoU: 0,549 (U-Net: 0,522)
Waiandt ve Seibel Júnior (2024)	YOLOv8s-P2	Kutu tespiti	Artırılmış AITEX (802 hatalı örnek), %65/20/15	mAP50: %85,11
Thomine ve Snoussi (2024)	Ters damıtma	Anomali tespiti (gözetimsiz)	MVTec AD, AITEX, TILDA ve üretim verisi	AUROC ile raporlanmıştır
Bu çalışma	YOLO11-seg (n, s, m)	Örnek bölütleme	AITEX, hatalı ve hatasız görüntüler; kaynak düzeyinde tabakalı %80/15/5; 3 tohum	Bkz. Uygulama IV (Tablo 9–12)

Tablo 4, AITEX üzerinde yapılan YOLO çalışmalarının kutu tespitine odaklandığını ve genellikle yalnızca hatalı görüntüleri kullandığını göstermektedir. Hatasız görüntülerin değerlendirmeye katılmaması yanlış alarm oranının ölçülmesini engeller. Bu çalışma iki açıdan bu çalışmalardan ayrılmaktadır. Birincisi, maske düzeyinde örnek bölütleme yapılmakta ve maskeler 4 Puan Sistemi ile puanlamada kullanılmaktadır. İkincisi, hatasız görüntüler de değerlendirmeye dahil edilmekte ve bölme, veri sızıntısını önlemek için kaynak görüntü düzeyinde yapılmaktadır.

YOLO Ailesi ve YOLO11-Seg Mimarisi

YOLOv1'den YOLO11'e Gelişim

YOLO (You Only Look Once), nesne tespitini görüntü üzerinde tek bir ileri geçişte çözülen bir bağlanım (regression) problemi olarak ele alır (Redmon et al., 2016). İlk sürümde görüntü bir ızgaraya bölünmüş ve her hücre, merkezi kendi içine düşen nesnelerin kutularını ve sınıf olasılıklarını doğrudan tahmin etmiştir. Bu yaklaşım iki aşamalı yöntemlere göre çok daha hızlıdır, ancak ilk sürümlerde küçük ve birbirine yakın nesnelerde başarım düşük kalmıştır. Sonraki sürümlerde bu sorun adım adım ele alınmıştır (Mao & Hong, 2025). YOLOv2 ve YOLOv3 önceden tanımlı çapa kutuları (anchor boxes) ile çok ölçekli çıkışlar, YOLOv4 ve YOLOv5 ise CSP (Cross Stage Partial) tabanlı omurgalar ve PANet tipi boyun yapıları getirmiştir; YOLOv5 ile Ultralytics'in PyTorch tabanlı eğitim altyapısı yaygınlaşmıştır. YOLOv8 çapasız (anchor-free) ve ayrık (decoupled) bir başlık getirmiş ve örnek bölütlemeyi bu yapıya taşımış, YOLO11 ise bu yapıyı koruyarak C3k2 ve C2PSA bloklarıyla öznitelik çıkarımını güncellemiştir (Jocher & Qiu, 2024; Khanam & Hussain, 2024).

Kumaş hatası literatüründe de YOLO sürümleri yaygın biçimde kullanılmakta ve küçük hatalar için çoğunlukla özel modüllerle uyarlanmaktadır (Mao & Hong, 2025). Bu çalışmada YOLO11'in tercih edilmesinin nedeni, örnek bölütlemeyi hızlı bir tek aşamalı yapıda sunması ve n, s, m gibi farklı ölçeklerde aynı mimarinin karşılaştırılmasına olanak tanımasıdır.

Omurga, Boyun ve Temel Bloklar: C3k2, SPPF ve C2PSA

YOLO11-seg üç bölümden oluşur (Şekil 3). **Omurga (backbone)** giriş görüntüsünden giderek azalan çözünürlükte öznitelik haritaları çıkarır. **Boyun (neck)** bu haritaları yukarıdan aşağıya ve aşağıdan yukarıya yollarla birleştirerek her ölçekte hem ayrıntı hem bağlam bilgisi taşıyan öznitelikler üretir (Lin et al.,

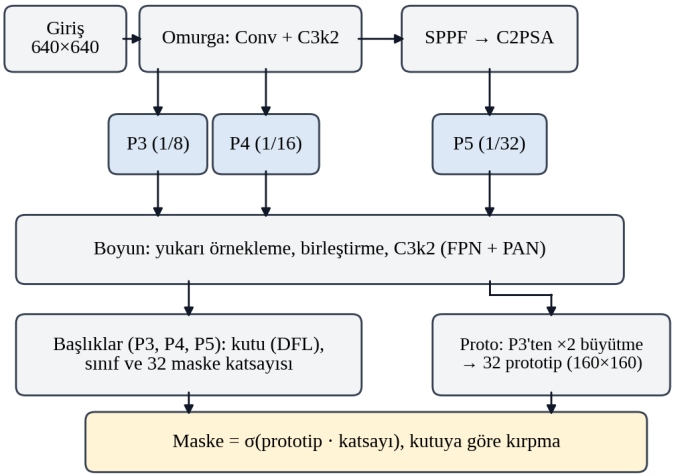
2017). **Başlık (head)** ise giriş boyutunun 1/8, 1/16 ve 1/32'si çözünürlüğündeki üç öznitelik haritası (P3, P4, P5) üzerinde tahmin yapar. Mimaride öne çıkan üç blok şunlardır (Jocher & Qiu, 2024; Khanam & Hussain, 2024):

- **C3k2:** YOLOv8'deki C2f bloğunun yerini alan, CSP tabanlı bir öznitelik çıkarma bloğudur. Giriş kanalları iki kola ayrılır; kollardan biri art arda darboğaz (bottleneck) alt bloklarından geçerken diğeri doğrudan aktarılır ve sonunda kollar birleştirilir. Model ölçeğine bağlı olarak iç bloklar standart darboğaz ya da daha derin C3k alt blokları olarak seçilir. Bu yapı, gradyan akışını korurken parametre sayısını düşük tutar.
- **SPPF (Spatial Pyramid Pooling – Fast):** Omurganın sonunda yer alır. Aynı öznitelik haritasına art arda üç kez 5×5 en büyük havuzlama (max pooling) uygulanır ve ara çıktılar birleştirilir. Böylece farklı büyüklükteki alıcı alanlardan gelen bağlam bilgisi düşük maliyetle bir araya getirilir.
- **C2PSA:** SPPF'nin ardından eklenen ve konuma duyarlı dikkat (position-sensitive attention) bloklarını CSP yapısı içinde kullanan bir modüldür. Çok başlı dikkat ve ileri beslemeli katmanlar içerir. Bu blok, modelin en düşük çözünürlüklü öznitelik haritasında görüntünün önemli bölgelerine odaklanmasını sağlar. Khanam ve Hussain (2024) bu dikkat mekanizmasının küçük ve kısmen örtülü nesnelerin tespitine katkı sağladığını belirtmektedir.

Model ölçekleri (n, s, m), aynı mimarinin derinlik ve genişlik çarpanlarıyla elde edilir. Ultralytics yapılandırma dosyasına göre derinlik çarpanı üç ölçekte de 0,50, genişlik çarpanı ise sırasıyla 0,25, 0,50 ve 1,00'dir (Jocher & Qiu, 2024). Ölçeklerin

parametre sayıları Uygulama IV'te, varyant karşılaştırmasıyla birlikte verilmiştir.

Şekil 3. YOLO11-seg mimarisinin genel yapısı



Kaynak: Jocher ve Qiu (2024) ile Khanam ve Hussain (2024) esas alınarak yazarlar tarafından hazırlanmıştır.

Segmentasyon Başlığı ve Kayıp Fonksiyonu

YOLO11-seg, örnek maskelerini YOLACT'ın önerdiği prototip yaklaşımıyla üretir (Bolya et al., 2019). Bu yaklaşımda her örnek için ayrı bir maske ağı çalıştırılmaz. Bunun yerine görüntünün tamamı için ortak bir **prototip maske** kümesi üretilir ve her örneğin maskesi bu prototiplerin ağırlıklı birleşimi olarak elde edilir. Ultralytics uygulamasında Proto modülü, P3 öznitelik haritasını iki kat büyüterek giriş çözünürlüğünün 1/4'ünde (640 piksellik girişte 160×160) $k = 32$ prototip üretir. Tespit başlığı ise her aday kutu için sınıf ve konum tahminine ek olarak 32 boyutlu bir katsayı vektörü tahmin eder (Jocher & Qiu, 2024). Bir örneğin maskesi şöyle hesaplanır:

$$M = \sigma(\mathbf{P}\,\mathbf{c}) \tag{4}$$

Burada $\mathbf{P} \in \mathbb{R}^{(H_p W_p) \times k}$ satırları piksellere, sütunları prototiplere karşılık gelen prototip matrisini; $\mathbf{c} \in \mathbb{R}^{k}$ ilgili örneğin maske katsayılarını; σ sigmoid fonksiyonunu; M ise $H_p \times W_p$ boyutuna yeniden düzenlenen ve değerleri 0 ile 1 arasında olan örnek maskesini göstermektedir. Elde edilen maske, tahmin edilen kutunun dışında kalan bölgeleri sıfırlanarak kırpılır ve giriş boyutuna büyütülür. Prototipler bütün örnekler için bir kez hesaplandığından maske üretiminin ek maliyeti düşüktür.

Model, dört bileşenden oluşan bir kayıp fonksiyonuyla eğitilir:

$$\mathcal{L} = \lambda_{box}\,\mathcal{L}_{CIoU} + \lambda_{seg}\,\mathcal{L}_{mask} + \lambda_{cls}\,\mathcal{L}_{BCE} + \lambda_{dfl}\,\mathcal{L}_{DFL} \qquad (5)$$

Burada $\mathcal{L}_{CIoU}$ kutu konum kaybını (Zheng et al., 2020), $\mathcal{L}_{mask}$ maske kaybını, $\mathcal{L}_{BCE}$ sınıflandırma kaybını, $\mathcal{L}_{DFL}$ dağılım odak kaybını (Li et al., 2020), λ katsayıları ise bu kayıpların ağırlıklarını göstermektedir. CIoU ve DFL kayıpları ile kayıp ağırlıkları Ultralytics varsayılanlarıyla kullanılmıştır (Jocher & Qiu, 2024): $\lambda_{box} = 7{,}5$, $\lambda_{cls} = 0{,}5$ ve $\lambda_{dfl} = 1{,}5$. Maske kaybı, uygulamada kutu kaybıyla aynı katsayıyla ölçeklenir ($\lambda_{seg} = \lambda_{box}$).

Sınıflandırma ve maske kayıpları. Sınıflandırma kaybı, her sınıf için ayrı ayrı hesaplanan ikili çapraz entropidir (binary cross-entropy, BCE). Maske kaybı da piksel bazında ikili çapraz entropiyle hesaplanır. Ultralytics uygulamasında bu kayıp yalnızca gerçek kutunun içinde kalan pikseller üzerinden ortalanır ve kutu alanına göre normalize edilir. Böylece küçük ve büyük hatalar kayba benzer ölçüde katkı yapar. Gerçek maskeler eğitimde `mask_ratio` parametresiyle belirlenen oranda küçültülmüş olarak kullanılır. Bu oran prototip çözünürlüğünden farklı olduğunda prototipler kayıp hesabı için gerçek maskenin boyutuna büyütülür (Jocher & Qiu, 2024). Bu ayrıntı, ince çizgisel hataların maske

etiketlerinin korunması açısından önemlidir ve Uygulama III'te ele alınmıştır.

Uygulama I: Veri Setinin Hazırlanması

Ham AITEX görüntüleri, YOLO11-seg modelinin doğrudan kullanabileceği biçimde değildir: görüntüler çok geniştir (4096×256), maskeler piksel tabanlıdır ve bazı görüntülerde birden fazla maske bulunmaktadır. Hazırlama adımları aşağıda açıklanmıştır.

Maskelerin ikili hale getirilmesi. Maskelerin çoğunun kenarlarında ara gri tonlar bulunduğundan maskeler 127 eşik değeriyle ikili hale getirilmiştir. Bu eşik, sıfırdan büyük tüm pikselleri hata saymaya göre toplam hata alanını yalnızca %0,9 azaltmaktadır. İki maskesi olan görüntülerde maskeler birleştirilmiştir.

Yamalama. 4096×256'lık şeritler 256×256 piksellik kare yamalara bölünmüştür. Eğitim kümesinde yamalar %50 örtüşmeli (128 piksel adım) çıkarılmıştır. Böylece yama sınırına denk gelen hatalar en az bir yamada bütün olarak görülür ve eğitim örneği sayısı artar. Doğrulama ve test kümelerinde örtüşmesiz (256 piksel adım) yamalar kullanılmıştır; bu sayede her piksel değerlendirmede yalnızca bir kez sayılır. Son yama her zaman görüntünün sağ kenarına hizalanmış, böylece farklı genişlikteki (3796 piksel) görüntü de tamamen kapsanmıştır.

Kumaş dışı alanların ayıklanması. Bazı görüntülerin kenarlarında kumaşın bittiği beyaz arka plan bulunmaktadır. Sütun ortalaması 230'un üzerinde olan sütunlar kumaş dışı kabul edilmiş, yarısından fazlası kumaş dışı olan ve hata içermeyen 663 yama veri setinden çıkarılmıştır.

Sınıf yapısı. 12 hata türünden üçünde yalnızca birer görüntü bulunduğundan (Tablo 1), her türü ayrı bir sınıf olarak modellemek

güvenilir bir değerlendirmeye izin vermez. Hata türleri bu nedenle görünüşlerine göre üç morfolojik grupta birleştirilmiştir: ipliğin ya da kumaşın bir doğrultu boyunca bozulduğu **çizgisel** hatalar, sınırlı bir bölgede topaklanan **noktasal** hatalar ve kumaş kenarında oluşan **kenar** hataları. Kırışık (crease) hatası, kumaşta bir doğrultu boyunca uzanan iz bıraktığı için çizgisel gruba alınmıştır. Bu karar, kenar sınıfının çok az örnekle temsil edilmesi pahasına, grupların görsel olarak tutarlı olmasını gözetmektedir.

Maskeden poligona dönüştürme. YOLO11-seg etiket biçimi, her nesne için sınıf numarasını ve yama boyutuna göre normalize edilmiş bir poligonu içerir. Her yamadaki maske bağlantılı bileşenlerine ayrılmış ve her bileşenin dış konturu bir poligon olarak kaydedilmiştir. 4 pikselden küçük bileşenler kenar yumuşatma artığı kabul edilerek etiketlenmemiştir. Tek piksel kalınlığındaki çizgiler geçerli bir poligon oluşturamadığı için bu bileşenlerde sınırlayıcı dikdörtgen kullanılmıştır. Dönüşümün doğruluğu, poligonlar yeniden maskeye çevrilip özgün maskeyle karşılaştırılarak ölçülmüştür. Kesişim/birleşim (IoU) ortalaması tüm hatalı yamalarda 0,997 çıkmıştır.

Hatasız yamaların dengelenmesi. Hatasız yamalar hatalı yamalardan çok daha fazladır; eğitim kümesinde 5189 hatasız aday yama bulunmaktadır. Modelin arka plana aşırı uyum sağlamaması için eğitim kümesine, hatalı yama sayısı kadar (1:1) hatasız yama seçilmiştir. Seçim, kumaş türlerinin oranı korunarak ve sabit tohumla yapılmıştır. Doğrulama ve test kümelerinde ise kumaş dışı yamalar dışındaki tüm hatasız yamalar tutulmuştur. Böylece bu kümelerdeki hatasız yama oranı gerçek üretim koşullarına yakın kalır ve yanlış alarm oranı gerçekçi biçimde ölçülür. Hazırlanan veri setinin son hali Tablo 5'te verilmiştir.

Tablo 5. Hazırlanan YOLO11-seg veri setinde yama ve örnek sayıları

Küme	Kaynak görüntü	Yama	Hatalı yama	Hatasız yama	Çizgisel örnek	Noktasal örnek	Kenar örnek
Eğitim	195	562	281	281	229	109	17
Doğrulama	36	514	33	481	24	12	1
Test	13	181	8	173	7	4	1
Toplam	**244**	**1257**	**322**	**935**	**260**	**125**	**19**

Kenar grubunun doğrulama ve test kümelerinde birer örnekle temsil edilmesi ve test kümesinde yalnızca 8 hatalı yama bulunması, sonuçların çok tohumlu ve şerit düzeyinde değerlendirilmesini gerektirmiştir.

Uygulama II: Ön İşleme ve Veri Artırma

Tüm adımlar Ultralytics eğitim altyapısı içinde, her eğitim adımında rastgele olarak uygulanmıştır. Diske artırılmış kopyalar yazılmamıştır; bu nedenle doğrulama ve test kümeleri özgün halleriyle kalmıştır.

Ön işleme. AITEX görüntüleri tek kanallı gri tonlu görüntülerdir. Renkli COCO veri setinde önceden eğitilmiş ağırlıklarla uyumlu olması için gri kanal üç kanala kopyalanmıştır. 256×256 piksellik yamalar modele girmeden önce 640×640 piksele büyütülmüş ve piksel değerleri [0, 1] aralığına ölçeklenmiştir. Yamalar kare olduğundan büyütme sırasında kenar dolgusu gerekmemiştir. Büyütmenin amacı, ortanca boyutu yaklaşık 7×8 piksel olan hataların modelin en ince tahmin ölçeğinde (1/8) birden fazla hücreye karşılık gelmesini sağlamaktır.

Veri artırma. Uygulanan dönüşümler ve gerekçeleri Tablo 6'da özetlenmiştir. Dönüşümler, kumaşın hatta karşılaşabileceği değişimleri taklit edecek ve hatanın anlamını (örneğin atkı ya da çözgü doğrultusunu) değiştirmeyecek biçimde seçilmiştir.

Tablo 6. Eğitimde uygulanan veri artırma dönüşümleri

Dönüşüm	Değer	Gerekçe
Parlaklık (HSV-V) değişimi	±%30	Aydınlatma yoğunluğundaki ve kumaş yansıtıcılığındaki farklar
Renk tonu ve doygunluk (HSV-H, HSV-S)	0 (kapalı)	Görüntüler gri tonlu
Yatay ve düşey çevirme	Her biri 0,5 olasılıkla	Dokunun atkı/çözgü doğrultuları korunur; hata yönü anlamını yitirmez
Öteleme	±%10	Kumaşın hat üzerindeki konum kaymaları
Ölçekleme	±%30	Kamera uzaklığı ve iplik sıklığındaki küçük farklar
Döndürme, eğme, perspektif	0 (kapalı)	Atkı ve çözgü doğrultularının anlamı korunur
Mozaik (4 yamanın birleştirilmesi)	1,0; son 10 epokta kapalı	Bir eğitim görüntüsünde daha fazla hata ve bağlam; son epoklarda gerçek dağılıma uyum
MixUp, Copy-Paste	0 (kapalı)	Yarı saydam veya yapay yerleştirilmiş hatalar gerçekçi değildir

Uygulama III: Model Eğitimi

Eğitim Ortamı ve Tekrarlanabilirlik

Eğitimler Kaggle platformunda tek bir NVIDIA Tesla T4 GPU (14,6 GB) üzerinde; Python 3.12, PyTorch 2.10 (CUDA 12.8) ve Ultralytics 8.4.164 ile yapılmıştır. GPU çıkarım süreleri de aynı GPU'da ölçülmüş, şerit düzeyindeki değerlendirme ve CPU hız ölçümleri ise Intel Core i5-1035G1 (4 çekirdek, 8 iş parçacığı, 1,0 GHz temel frekans) işlemcili bir dizüstü bilgisayarda yürütülmüştür. Üç model ölçeği (n, s, m) COCO veri setinde önceden eğitilmiş Ultralytics ağırlıklarından başlatılmış (aktarımlı öğrenme, transfer learning) ve üç farklı tohumla (42, 43, 44) eğitilmiştir. Ultralytics'in deterministik çalışma kipi açık tutulmuş ve eğitimler arasında yalnızca model ölçeği ile tohum değiştirilmiştir. Deterministik kip aynı sonuçları yalnızca aynı donanım ve yazılım ortamında güvence altına alır; farklı ortamlarda (örneğin Colab ve Kaggle) sonuçlar farklılaşabilir.

Hiperparametreler

Kullanılan hiperparametreler Tablo 7'de verilmiştir. Her eğitimde, Ultralytics'in doğrulama uygunluk ölçütünü (kutu ve maske mAP50 ile mAP50-95 değerlerinin ağırlıklı toplamı) en yüksek yapan epoğun ağırlıkları nihai model olarak saklanmıştır.

Tablo 7. Tüm eğitimlerde kullanılan hiperparametreler

Hiperparametre	Değer	Hiperparametre	Değer
Giriş boyutu	640×640	Optimizasyon algoritması	SGD (momentum 0,937)
Yığın boyutu (batch)	16	Başlangıç öğrenme oranı	0,01
En fazla epok	150	Son öğrenme oranı çarpanı	0,01 (doğrusal azalma)
Erken durdurma sabrı	50 epok	Ağırlık bozunumu	0,0005
Isınma	3 epok	Karma duyarlıklı eğitim (AMP)	Açık
Maske küçültme oranı (mask_ratio)	2	Kayıp ağırlıkları (kutu / sınıf / DFL)	7,5 / 0,5 / 1,5
Sınıf sayısı	3 (çizgisel, noktasal, kenar)	Tohumlar	42, 43, 44

Ön Deneyler: Maske Çözünürlüğü ve Sınıf Yapısı

Tablo 7'deki değerlerin büyük bölümü Ultralytics varsayılanlarıdır. İki tasarım kararı ise ön deneylerle verilmiştir: maske küçültme oranı ve sınıf yapısı. Ön deneylerde YOLO11n-seg, Google Colab ortamında tek bir tohumla (42) ve en fazla 150 epokla eğitilmiş ve doğrulama kümesinde karşılaştırılmıştır. Sonuçlar Tablo 8'de verilmiştir.

Tablo 8. Ön deney sonuçları (YOLO11n-seg, tohum 42.

Deney	Maske küçültme oranı	Sınıf yapısı	Kutu mAP50	Maske mAP50	Maske mAP50-95
A	4 (varsayılan)	3 grup	0,220	0,066	0,020
B	2	3 grup	**0,573**	**0,394**	**0,087**
C	2	Tek sınıf	0,481	0,110	0,024

Varsayılan ayarla (A) model hataların yerini kısmen bulabilmiş, ancak maskeleri neredeyse hiç üretememiştir (maske mAP50 = 0,066). Bunun nedeni, varsayılan oranda gerçek

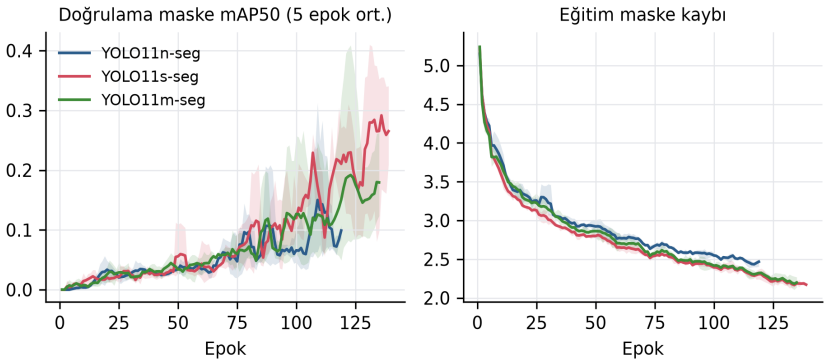
maskelerin eğitimde dört kat küçültülmesidir. Bu küçültmede 640 piksellik girişte 160×160'a inen etiketlerde, genişliği bir iki piksel olan çizgisel hatalar büyük ölçüde kaybolmaktadır. Oranın 2'ye indirilmesiyle (B) etiketler 320×320 çözünürlükte tutulmuş, prototipler de kayıp hesabında bu boyuta büyütülmüştür (bkz. Segmentasyon Başlığı ve Kayıp Fonksiyonu). Maske mAP50 değeri 0,066'dan 0,394'e yükselmiştir. Oranın 1'e indirilmesi ise T4 GPU'nun belleğini doğrulama aşamasında aştığından uygulanamamıştır. Aynı maske oranıyla, tüm hata türlerinin tek bir "hata" sınıfında birleştirildiği deney (C), üç morfolojik grubun kullanıldığı deneyden (B) belirgin biçimde düşük sonuç vermiştir (maske mAP50: 0,110'a karşı 0,394). Farklı görünümdeki hataların tek sınıfta toplanması, modelin her gruba özgü görünümü öğrenmesini zorlaştırmış görünmektedir. Bu nedenle ana eğitimlerde maske küçültme oranı 2 ve üç sınıflı yapı kullanılmıştır. Ön deney B ile aynı ayarlara sahip olan ana eğitim (YOLO11n-seg, tohum 42) Kaggle ortamında 0,157 maske mAP50 değerine ulaşmıştır. Donanım ve yazılım ortamındaki farkların yol açmış olabileceği bu fark, Uygulama IV'te görülen yüksek tohum duyarlılığıyla tutarlıdır. Ön deneyler bu nedenle yalnızca ayarların birbirine göre karşılaştırılması için kullanılmış, raporlanan sonuçlar ana eğitimlerden alınmıştır.

Eğitim Süreci ve Epok Sayısının Yeterliliği

Dokuz eğitimin (3 ölçek × 3 tohum) doğrulama maske mAP50 eğrileri ve eğitim maske kayıpları Şekil 4'te gösterilmiştir. Maske kaybı üç ölçekte de düzenli olarak azalmaktadır. Doğrulama maske mAP50 değeri ise ilk 50 epokta çok düşük kalmakta, yaklaşık 75. epoktan sonra belirgin biçimde yükselmektedir. Bu gecikme, modelin önce hataların yerini (kutuları) öğrendiğini, ince maske ayrıntılarını ise eğitimin ilerleyen aşamalarında öğrendiğini düşündürmektedir. Eğrilerdeki geniş bantlar tohumlar arasındaki

değişkenliğin yüksek olduğunu göstermekte ve çok tohumlu değerlendirmenin gerekliliğini desteklemektedir.

Şekil 4. Doğrulama maske mAP50 ve eğitim maske kaybı



Eğitimlerin bir kısmı 119–139. epoklar arasında erken durmuş, diğerleri 150 epoğu tamamlamıştır; bir eğitim T4 GPU'da ortalama 50 (n), 67 (s) ve 97 dakika (m) sürmüştür. 150 epokluk üst sınırın yeterliliği, YOLO11s-seg'in tohum 42 ile 300 epoka kadar eğitilmesiyle ayrıca sınanmıştır. Daha uzun eğitim (en iyi epok 196) doğrulama maske mAP50 değerini artırmamış (0,391'e karşı 0,381), eğitim süresini ise 64 dakikadan 104 dakikaya çıkarmıştır; bu nedenle 150 epokluk üst sınır korunmuştur.

Uygulama IV: Performans Değerlendirmesi

Modeller iki düzeyde değerlendirilmiştir. İlk düzeyde, nesne tespiti ve bölütleme literatüründe yaygın olan ortalama kesinlik ölçütleri, eğitimde kullanılan 256×256 piksellik yamalar üzerinde hesaplanmıştır. İkinci düzeyde, yamalar üzerindeki tahminler özgün 4096×256 piksellik şeritlerde yeniden birleştirilmiş ve piksel ile görüntü düzeyinde ölçütler hesaplanmıştır. İkinci düzey, bir muayene sisteminin sahada karşılaşacağı durumu daha iyi yansıtır: kamera kumaşı yama yama değil, şerit şerit görür ve kararın bütün şerit için verilmesi gerekir. Doğrulama kümesinde 36 şerit (15 hatalı, 21 hatasız), test

kümesinde 13 şerit (6 hatalı, 7 hatasız) bulunmaktadır. Aksi belirtilmedikçe sonuçlar üç tohumun (42, 43, 44) ortalaması ± standart sapması olarak verilmiştir.

Kutu ve Maske Ortalama Kesinliği (mAP)

Bir tahminin doğru sayılması için tahmin ile gerçek hata arasındaki kesişim/birleşim oranının (intersection over union, IoU) belirli bir eşiği aşması gerekir. Güven eşiği değiştirildikçe elde edilen kesinlik (precision)–duyarlılık (recall) eğrisinin altında kalan alan ortalama kesinliği (average precision, AP) verir. mAP50, IoU eşiği 0,50 iken sınıflar üzerinden ortalaması alınan AP'yi; mAP50-95 ise 0,50 ile 0,95 arasındaki on IoU eşiği üzerinden ayrıca ortalanan değeri ifade eder (Lin et al., 2014). IoU'nun sınırlayıcı kutular üzerinden hesaplanmasıyla **kutu mAP**, piksel maskeleri üzerinden hesaplanmasıyla **maske mAP** elde edilir. Hesaplamalar Ultralytics kütüphanesinin doğrulama modülüyle yapılmıştır (Jocher & Qiu, 2024).

Üç varyantın yama düzeyindeki sonuçları Tablo 9'da verilmiştir.

Tablo 9. Yama düzeyinde kutu ve maske mAP değerleri (üç tohum, ortalama ± standart sapma)

Ölçüt	YOLO11n-seg	YOLO11s-seg	YOLO11m-seg
Doğrulama kutu mAP50	0,549 ± 0,015	**0,593 ± 0,024**	0,482 ± 0,133
Doğrulama kutu mAP50-95	0,242 ± 0,082	**0,329 ± 0,015**	0,219 ± 0,109
Doğrulama maske mAP50	0,203 ± 0,179	**0,387 ± 0,022**	0,305 ± 0,148
Doğrulama maske mAP50-95	0,065 ± 0,076	**0,219 ± 0,053**	0,103 ± 0,075
Parametre sayısı (milyon)	2,84	10,08	22,36
GPU çıkarım süresi (ms/yama, T4)	5,4	14,6	37,6

Not: GPU süresi yama başına yalnızca model çıkarımını kapsar.

Doğrulama kümesinde YOLO11s-seg dört ölçütün dördünde de en yüksek değere ulaşmıştır. Daha önemlisi, maske ölçütlerinde tohumlar arasındaki sapması en düşük olan varyanttır. YOLO11n-seg'in maske mAP50 değeri tohuma bağlı olarak 0,052

ile 0,401 arasında değişmiştir. YOLO11m-seg'de ise tohum 42 ile yapılan eğitim diğer iki tohumun belirgin biçimde gerisinde kalmıştır (doğrulama kutu mAP50: 0,331; diğer tohumlarda 0,529 ve 0,585). Bu bulgu, küçük veri setlerinde tek tohumla yapılan karşılaştırmaların yanıltıcı olabileceğini göstermektedir.

Tabloda dikkat çeken ikinci nokta, kutu ve maske değerleri arasındaki farktır. YOLO11s-seg doğrulamada hataların yerini kutu düzeyinde 0,593 mAP50 ile bulurken, maske düzeyinde bu değer 0,387'ye düşmektedir. Sınıf bazındaki sonuçlar (Tablo 10) bu farkın büyük ölçüde çizgisel hatalardan kaynaklandığını göstermektedir: çizgisel sınıfta kutu mAP50 0,456 iken maske mAP50 yalnızca 0,012'dir.

Tablo 10. YOLO11s-seg için sınıf bazında doğrulama sonuçları (üç tohum ortalaması)

Sınıf	**Doğrulama örnek sayısı**	**Kutu mAP50**	**Maske mAP50**	**Maske mAP50-95**
Çizgisel	24	0,456 ± 0,059	0,012 ± 0,007	0,003 ± 0,002
Noktasal	12	0,326 ± 0,034	0,152 ± 0,061	0,042 ± 0,018
Kenar	1	0,995 ± 0,000	0,995 ± 0,000	0,611 ± 0,148

Not: Kenar sınıfı doğrulama kümesinde tek örnekle temsil edildiğinden bu sınıfın değerleri yorumlanmamalıdır.

Çizgisel hatalardaki bu düşüş, maske IoU'sunun ince yapılar karşısındaki hassasiyetiyle açıklanabilir. Genişliği w piksel olan bir çizgi, tahminde çizgiye dik doğrultuda d piksel kaydırılırsa iki maskenin IoU değeri yaklaşık olarak şöyle olur:

$$\mathrm{IoU}(d) \approx \frac{w-d}{w+d} \qquad 0 \le d < w \tag{6}$$

Burada w çizginin piksel cinsinden genişliğini, d ise tahmin edilen maskenin gerçek maskeye göre kaymasını göstermektedir. AITEX'teki çizgisel hataların bir kısmı yalnızca birkaç piksel genişliğindedir. Örneğin $w = 2$ piksellik bir çizgide yalnızca 1 piksellik kayma IoU'yu 1/3'e düşürür ve tahmin, hata doğru yerde

bulunmuş olsa bile 0,50 eşiğinde yanlış sayılır. Kutu IoU'su çizginin uzunluğu boyunca hesaplandığı için bu kaymadan çok daha az etkilenir. Kenar sınıfının yüksek değerleri ise doğrulama kümesinde tek bir örneğe dayandığı için yorumlanmamalıdır.

Test kümesindeki mAP değerleri (0,018–0,120) doğrulama değerlerinin çok altındadır. Bunun iki nedeni vardır. Birincisi, test kümesinde yalnızca 8 hatalı yama ve 12 hata örneği bulunmaktadır (Tablo 5); tek bir örneğin kaçırılması ölçütü önemli ölçüde değiştirir. İkincisi, değerlendirmede örtüşmesiz yamalar kullanıldığından uzun hatalar yama sınırlarında parçalanmakta ve her parça ayrı bir örnek olarak sayılmaktadır. Bu nedenlerle yama düzeyindeki test mAP değerleri bu çalışmada yalnızca metinde ve bilgi amaçlı verilmiş, modellerin asıl başarımı aşağıdaki şerit düzeyindeki ölçütlerle değerlendirilmiştir.

Piksel Düzeyinde Dice ve IoU

Şerit düzeyindeki değerlendirmede her şerit, eğitimdeki gibi 256×256 piksellik örtüşmesiz yamalara bölünmüş (son yama sağ kenara hizalanmıştır) ve yamalar modele toplu olarak verilmiştir. Güven puanı t eşiğini aşan tüm örnek maskeleri şerit boyutundaki boş bir maskeye birleşim (mantıksal VEYA) işlemiyle yerleştirilmiştir. Kumaş dışı beyaz sütunlar (bkz. Uygulama I) hem gerçek hem de tahmin maskesinden çıkarılmıştır. Elde edilen şerit maskeleri gerçek maskelerle piksel piksel karşılaştırılmış ve Dice katsayısı ile IoU hesaplanmıştır:

$$\begin{aligned} \text{Dice}_{\mu} &= \frac{2\sum_k DP_k}{2\sum_k DP_k + \sum_k YP_k + \sum_k YN_k} \\ \text{IoU}_{\mu} &= \frac{\sum_k DP_k}{\sum_k DP_k + \sum_k YP_k + \sum_k YN_k} \end{aligned} \tag{7}$$

Burada k kümedeki şeritleri; DP_k, YP_k ve YN_k ise k. şeritteki doğru pozitif, yanlış pozitif ve yanlış negatif piksel sayılarını göstermektedir. Toplamlar hatasız şeritleri de kapsar; bu

nedenle hatasız şeritlerdeki yanlış alarmlar mikro (μ) ölçütleri doğrudan düşürür. Ayrıca yalnızca hatalı şeritler üzerinden hesaplanan şerit başına Dice değerlerinin ortalaması (makro Dice) da raporlanmıştır.

Sonuçlar Tablo 11'de verilmiştir. Doğrulama kümesinde en yüksek mikro Dice değerine YOLO11s-seg ulaşmıştır (0,383 ± 0,048). Test kümesinde ise YOLO11m-seg (0,521 ± 0,064) diğer iki varyantın önündedir; n ve s varyantlarının test değerleri birbirine çok yakındır (0,425 ve 0,428). Yama düzeyindeki test mAP değerlerinin aksine şerit düzeyindeki Dice değerleri doğrulama ile aynı büyüklüktedir. Bu durum, şerit düzeyindeki değerlendirmenin küçük test kümesinde daha kararlı bir ölçüm sağladığını düşündürmektedir.

Tablo 11. Şerit düzeyinde piksel ve görüntü ölçütleri

Küme	Ölçüt	YOLO11n-seg	YOLO11s-seg	YOLO11m-seg
	Çalışma eşiği t^*	0,053 ± 0,045	0,060 ± 0,036	0,107 ± 0,124
Doğrulama	Mikro Dice	0,325 ± 0,049	**0,383 ± 0,048**	0,340 ± 0,060
	Mikro IoU	0,195 ± 0,035	**0,238 ± 0,036**	0,206 ± 0,043
	Görüntü doğruluğu	0,759 ± 0,197	0,824 ± 0,058	**0,833 ± 0,048**
	Görüntü duyarlılığı	**0,978 ± 0,038**	0,911 ± 0,038	0,911 ± 0,102
	Yanlış alarm oranı	0,397 ± 0,364	0,238 ± 0,126	**0,222 ± 0,120**
Test	Mikro Dice	0,425 ± 0,077	0,428 ± 0,057	**0,521 ± 0,064**
	Mikro IoU	0,272 ± 0,063	0,274 ± 0,047	**0,354 ± 0,059**
	Makro Dice	**0,496 ± 0,030**	0,489 ± 0,049	0,484 ± 0,086
	Görüntü doğruluğu	0,692 ± 0,077	0,821 ± 0,118	**0,846 ± 0,077**
	Görüntü duyarlılığı	0,944 ± 0,096	0,944 ± 0,096	0,944 ± 0,096
	Yanlış alarm oranı	0,524 ± 0,218	0,286 ± 0,247	**0,238 ± 0,218**

Görüntü Düzeyinde Doğruluk ve Yanlış Alarm Oranı

Üretim hattında operatör için en önemli bilgi, bir kumaş bölümünün hatalı olup olmadığıdır. Bu nedenle her şerit, içinde güven puanı t eşiğini aşan en az bir tahmin bulunduğunda "hatalı", aksi durumda "hatasız" olarak sınıflandırılmıştır. Görüntü doğruluğu doğru sınıflandırılan şeritlerin oranı, görüntü duyarlılığı hatalı şeritlerden yakalananların oranı, yanlış alarm oranı (YAO) ise hatasız şeritlerden hatalı olarak işaretlenenlerin oranıdır. Bu

tanım tahminin gerçek hatayla aynı yerde olup olmadığına bakmaz. Hatalı bir şeritte hatadan uzak bir noktada verilen yanlış alarm da o şeridi "yakalanmış" sayar. Hatanın gerçekten bulunup bulunmadığı, Uygulama V'te hata düzeyinde eşleştirme ile ayrıca incelenmiştir.

Tablo 11'e göre üç varyant da testteki hatalı şeritlerin %94,4'ünü yakalamıştır. Varyantları ayıran ölçüt yanlış alarm oranıdır. YOLO11n-seg testteki hatasız şeritlerin yarısından fazlasını (%52,4) hatalı olarak işaretlerken bu oran YOLO11s-seg'de %28,6'ya, YOLO11m-seg'de %23,8'e düşmektedir. Test kümesinde yalnızca 7 hatasız şerit bulunduğundan tek bir şeridin yanlış işaretlenmesi yanlış alarm oranını 0,143 değiştirmektedir. Bu nedenle s ve m arasındaki fark istatistiksel olarak anlamlı kabul edilmemelidir.

YOLO11s-seg'in görüntü düzeyindeki kararları, üç tohumun toplamı olarak Tablo 12'deki karmaşıklık matrisinde (confusion matrix) verilmiştir. Test kümesinde 18 hatalı şerit kararının 17'si doğrudur; 21 hatasız şerit kararının 6'sında yanlış alarm oluşmuştur.

Tablo 12. YOLO11s-seg için görüntü düzeyinde karmaşıklık matrisi

Küme	**Gerçek durum**	**Tahmin: hatalı**	**Tahmin: hatasız**	**Toplam**
Doğrulama	Hatalı	41	4	45
	Hatasız	15	48	63
Test	Hatalı	17	1	18
	Hatasız	6	15	21

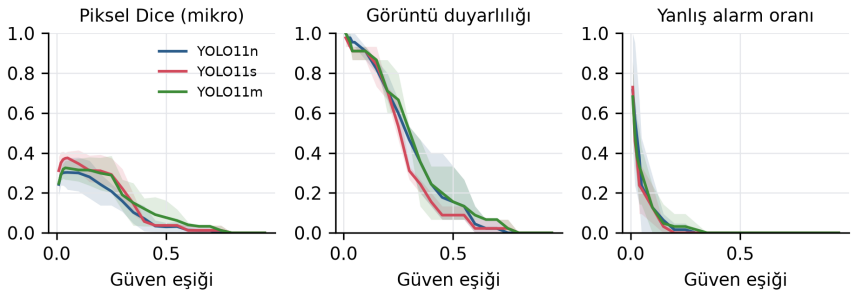
Güven Eşiği Analizi

Modelin her tahmini 0 ile 1 arasında bir güven puanı taşır ve hangi tahminlerin kabul edileceği bir eşikle belirlenir. Ultralytics'in varsayılan eşiği (0,25) bu çalışma için sabit alınmamış, eşik her eğitim için doğrulama kümesinde 0,01–0,05 aralığında 0,01, 0,05–0,95 aralığında 0,05 adımla taranmış ve mikro Dice değerini en yüksek yapan değer çalışma eşiği (t^*)

olarak seçilmiştir. Seçilen eşik test kümesinde değiştirilmeden kullanılmıştır.

Taramanın sonuçları Şekil 5'te verilmiştir. Mikro Dice değeri üç varyantta da 0,04–0,05 civarında en yüksek düzeye ulaşmakta, 0,30'un üzerinde hızla düşmektedir. YOLO11s-seg için ortalama t^* değeri 0,060 ± 0,036'dır. Bu düşük değer, modellerin doğru tahminlere bile görece düşük güven puanı atadığını göstermektedir.

Şekil 5. Güven eşiğine göre doğrulama kümesinde mikro Dice, görüntü duyarlılığı ve yanlış alarm oranı (çizgiler üç tohumun ortalaması, gölgeli alanlar en küçük ve en büyük değerler arasıdır)

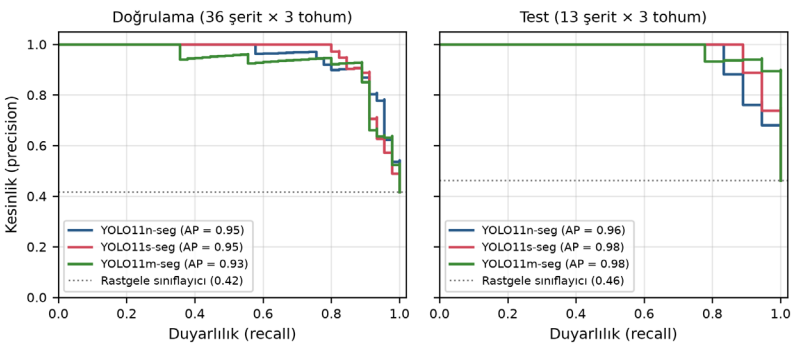


Şekil 5, eşik seçiminin bir tercih meselesi olduğunu da göstermektedir. Piksel düzeyinde en iyi bölütlemeyi veren eşik, yanlış alarm açısından en iyi eşik değildir. YOLO11s-seg için doğrulama kümesinde eşik 0,05'ten 0,15'e çıkarıldığında mikro Dice 0,376'dan 0,313'e düşmektedir. Buna karşılık yanlış alarm oranı 0,222'den 0,032'ye inmekte ve görüntü doğruluğu 0,833'ten 0,917'ye yükselmektedir; görüntü duyarlılığı ise 0,911'den 0,844'e gerilemektedir. Dolayısıyla hata alanının ve uzunluğunun doğru ölçülmesinin önemli olduğu puanlama uygulamalarında düşük eşik, yalnızca hatalı bölümün işaretlenmesinin yeterli olduğu ve operatörün yanlış alarmlarla meşgul edilmemesinin istendiği uygulamalarda ise daha yüksek bir eşik tercih edilebilir. Bu

çalışmada 4 Puan Sistemi puanlaması hata uzunluğuna dayandığından doğrulamada seçilen eşik kullanılmıştır.

Eşik seçiminden bağımsız bir değerlendirme için her şeride, kumaş bölgesindeki tahminlerinin en yüksek güven puanı atanmış ve görüntü düzeyinde kesinlik–duyarlılık (precision–recall, PR) eğrileri çizilmiştir (Şekil 6). Doğrulama kümesinde YOLO11s-seg için ortalama kesinlik (AP) 0,957 ± 0,006, ROC eğrisi altındaki alan (AUROC) 0,950 ± 0,008 bulunmuştur; n ve m varyantlarının değerleri benzer, ancak tohumlar arası sapmaları daha yüksektir (m için AP 0,930 ± 0,089). Bu sonuçlar modellerin hatalı ve hatasız şeritleri güçlü biçimde ayırabildiğini göstermektedir. Tablo 11 ve 12'deki yanlış alarmlar bu nedenle ayırma gücünden çok, piksel düzeyinde bölütleme için seçilen düşük çalışma eşiğinden kaynaklanmaktadır; görüntü düzeyindeki karar için ayrı ve daha yüksek bir eşik seçilmesi yanlış alarm oranını azaltabilir. Test kümesindeki değerler (YOLO11s-seg için AP 0,983 ± 0,030) 13 şeride dayandığından ihtiyatla yorumlanmalıdır.

Şekil 6. Görüntü düzeyinde kesinlik–duyarlılık eğrileri



Varyant Karşılaştırması: Doğruluk ve Hız

Hız ölçümleri Tablo 9 ve Tablo 13'te verilmiştir.

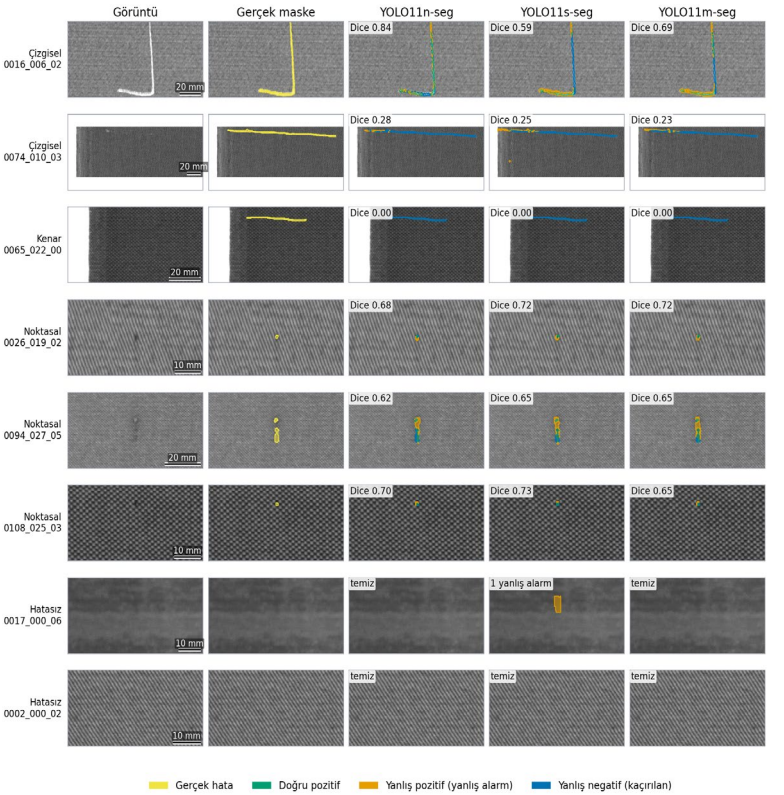
Ana model seçiminde doğrulama kümesi esas alınmıştır. YOLO11s-seg doğrulama kümesinde hem en yüksek mikro Dice ve

maske mAP değerlerine hem de tohumlar arasında en düşük maske mAP sapmasına sahiptir. Bu nedenle çalışmanın ana modeli olarak seçilmiştir. YOLO11m-seg'in test kümesindeki üstünlüğü (mikro Dice farkı 0,093) dikkat çekicidir. Ancak bu fark iki nedenle bir seçim ölçütü olarak kullanılmamıştır. Birincisi, test kümesine göre model seçmek, test kümesinin bağımsız bir ölçüm olma niteliğini ortadan kaldırır. İkincisi, fark tohumlar arasındaki standart sapmanın yaklaşık 1,5 katıdır ve 13 şeritlik bir test kümesinde, üç tohumla güvenilir biçimde ayırt edilemez. Bununla birlikte bu bulgu, daha büyük modellerin daha geniş bir veri setiyle yeniden değerlendirilmesi gerektiğine işaret etmektedir. YOLO11s-seg, CPU'da YOLO11m-seg'den yaklaşık 2,8 kat daha hızlıdır. YOLO11n-seg'e göre ise yaklaşık 2,3 kat daha yavaştır ama çok daha kararlı sonuç vermektedir.

Hata Analizi

Sayısal ölçütlerin arkasındaki davranışı görmek için her varyantın doğrulamada en iyi sonucu veren tohumu seçilmiş ve bu modellerin test şeritlerindeki tahminleri, doğru pozitif, yanlış pozitif ve yanlış negatif pikseller ayrı renklerle gösterilerek incelenmiştir (Şekil 7). Kullanılan eğitimler ve eşikler şunlardır: YOLO11n-seg tohum 42 ($t^* = 0{,}10$), YOLO11s-seg tohum 42 ($t^* = 0{,}10$), YOLO11m-seg tohum 44 ($t^* = 0{,}25$). Şekilde her hatanın çevresinden alınan bir kesit gösterilmektedir; küçük hatalar büyütülmüş ve her satıra 0,25 mm/piksel varsayımıyla ölçek çubuğu eklenmiştir. Panellerdeki Dice değerleri kesitin değil, şeridin tamamının değeridir.

Şekil 7. Test şeritlerinde üç varyantın tahminleri (sarı: gerçek hata; yeşil: doğru pozitif; turuncu: yanlış pozitif; mavi: yanlış negatif)



Kaynak: Silvestre-Blanes et al. (2019) tarafından yayımlanan veri seti kullanılarak yazarlar tarafından üretilmiştir.

Şekil 7'den dört gözlem çıkmaktadır:

- **Noktasal hatalar** (0026, 0094 ve 0108 numaralı şeritler) üç varyant tarafından da doğru yerde ve benzer biçimde bölütlenmiştir (Dice 0,62–0,73). Hatalar 1–2 mm boyutunda olmasına karşın model konumu doğru bulmakta, sınırları ise bir iki piksel geniş ya da dar çizmektedir.
- **Uzun çizgisel hatada** (0074, yaklaşık 155 mm) modeller hatanın yalnızca sol ucunu bulmakta, geri kalanını

kaçırmaktadır (Dice 0,23–0,28). Hata üç yamaya yayılmakta ve her yamada ayrı ayrı, düşük güvenle tahmin edilmektedir. Bu örnek, yama tabanlı çıkarımın uzun ve düşük kontrastlı hatalarda bir sınırlılığıdır.

- **Kenar hatası** (0065) bu üç modelin hiçbiri tarafından bulunamamıştır (Dice 0). Kenar grubunun eğitim kümesinde yalnızca 7 kaynak görüntü ve 17 örnekle temsil edilmesi (Tablo 5) bu sonucun olası temel nedenidir.
- **Yanlış alarm örneğinde** (0017, hatasız) YOLO11s-seg, belirgin yatay yoğunluk değişimleri içeren bir dokuda küçük bir bölgeyi hata olarak işaretlemiştir. Aynı şeridi n ve m varyantları temiz bulmuştur. Bu da yanlış alarmların büyük ölçüde modele ve tohuma özgü olduğuna işaret etmektedir.

Bu analiz, modelin neye bakarak karar verdiğini değil, verdiği kararın nerede doğru ve nerede yanlış olduğunu göstermektedir. Grad-CAM gibi gradyan tabanlı açıklanabilirlik yöntemleri bu çalışmada uygulanmamıştır.

Uygulama V: Hat Simülasyonu ve Otomatik Puanlama

Ana model (YOLO11s-seg) bir muayene hattında çalışıyormuş gibi kullanılmış ve tahmin ettiği maskelerden kumaşın 4 Puan Sistemi'ne göre otomatik olarak puanlanması denenmiştir. Amaç, model başarımını yalnızca bilgisayarlı görü ölçütleriyle değil, tekstil kalite kontrolünde kullanılan bir karar ölçütüyle de değerlendirmektir.

Tam Şerit Üzerinde Sanal Hat Simülasyonu

AITEX görüntüleri çizgi tarama kamerasıyla elde edilmiş 4096×256 piksellik şeritlerdir. Bu çalışmada test şeritleri, hat üzerinde ardışık olarak gelen kareler gibi işlenmiştir. Şeritler,

piksel düzeyindeki değerlendirmede açıklanan yamalı çıkarım adımlarıyla işlenmiştir.

Sistemin bir üretim hattına yetişip yetişemeyeceği, bir şeridin işlenme süresinin kumaşın o şerit boyunca ilerlemesi için geçen süreden kısa olmasına bağlıdır. Buna göre, işleme süresinin karşılayabileceği en yüksek hat hızı şöyle hesaplanmıştır:

$$v_{\max} = \frac{H \cdot r}{t_{\text{şerit}}} \tag{8}$$

Burada $v_{\max}$ eşdeğer hat hızını, H şeridin hareket doğrultusundaki piksel sayısını (256), r piksel boyutunu (0,25 mm/piksel; bkz. Denklem 2), $t_{\text{şerit}}$ ise bir şeridin işlenme süresini göstermektedir. Bir şerit hareket doğrultusunda yaklaşık 64 mm kumaşa karşılık gelir. GPU süresi Ultralytics'in bildirdiği ön işleme, çıkarım ve son işleme sürelerinin toplamının 16 yama ile çarpılmasıyla bulunmuştur. Bu değer görüntünün diskten okunması ve maskelerin birleştirilmesi gibi adımları içermediğinden bir üst sınır olarak yorumlanmalıdır. Sonuçlar Tablo 13'te verilmiştir.

Tablo 13. Şerit başına işleme süresi ve eşdeğer hat hızı

Varyant	CPU süresi (s/şerit)	CPU eşdeğer hız (m/dk)	GPU süresi (ms/şerit, T4)	GPU eşdeğer hız (m/dk)
YOLO11n-seg	2,60 ± 0,53	1,5	167 ± 4	23,0
YOLO11s-seg	5,93 ± 0,67	0,6	295 ± 2	13,0
YOLO11m-seg	16,66 ± 0,33	0,2	646 ± 2	5,9

Not: Süreler şerit (16 yama) başınadır; GPU süresi ön işleme, çıkarım ve son işleme sürelerinin toplamıdır.

CPU'da elde edilen hızlar (0,2–1,5 m/dk) bir muayene hattı için yetersizdir; bu donanım yalnızca geliştirme ve çevrimdışı analiz için uygundur. T4 GPU'da YOLO11s-seg yaklaşık 13 m/dk'lık bir hıza karşılık gelmektedir. Hız; daha güçlü bir GPU, daha düşük giriş çözünürlüğü (bu çalışmada 256 piksellik yamalar

640 piksele büyütülmektedir) veya yarım duyarlıklı hesaplama ile artırılabilir. Ancak bu seçeneklerin doğruluk üzerindeki etkisi bu çalışmada test edilmemiştir.

Maske Tabanlı Ölçüm ve 4 Puan Sistemi Puanlaması

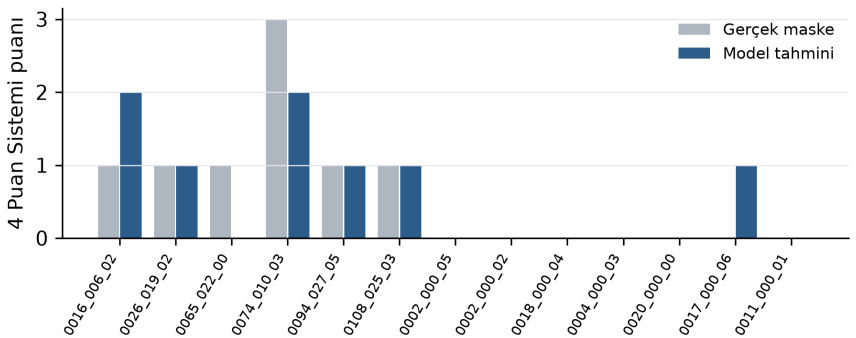
Tahmin maskesindeki her bağlantılı bileşen bir hata adayı olarak ele alınmıştır. Aynı hatanın birbirine yakın parçalara bölünmüş olabileceği göz önüne alınarak maske önce 5 mm (20 piksel) genişletilmiş, ardından bağlantılı bileşenlere ayrılmıştır. Her hatanın uzunluğu, bileşeni çevreleyen en küçük alanlı döndürülmüş dikdörtgenin uzun kenarı olarak ölçülmüş ve Denklem 3 ile inç cinsine çevrilmiştir. Hataya verilecek puan, ASTM D5430 standardındaki uzunluk sınıflarına (Tablo 2) göre belirlenmiştir (ASTM International, 2017). Aynı işlem gerçek maskelere de uygulanmış ve modelin verdiği puanlar, gerçek maskelerden hesaplanan puanlarla karşılaştırılmıştır. Bir tahmin, bir gerçek hatayla örtüşüyorsa o hatayı bulmuş sayılmıştır. Şeritlerin toplam puanı Denklem 1 ile 100 yd² başına puana (P_{100}) çevrilmiştir. Bu nedenle uygulanan yöntem, standardın uzunluğa dayalı kısmının bir uyarlamasıdır ve iki basitleştirme içerir. Birincisi, deliklere uzunluktan bağımsız verilen puanlar uygulanmamıştır, çünkü veri setinde delik türünü ayırt eden bir sınıf bulunmamaktadır. İkincisi, "doğrusal yarda başına en fazla 4 puan" kuralı uygulanmamıştır, çünkü test şeritleri aynı kumaş topunun ardışık bölümleri değildir. Sonuçlar Tablo 14'te, şerit bazındaki puanlar ise Şekil 8'de verilmiştir.

Tablo 14. YOLO11s-seg ile 4 Puan Sistemi puanlama sonuçları

Ölçüt	**Ortalama ± SS**
Çalışma eşiği t^*	0,060 ± 0,036
Bulunan gerçek hata (toplam 6)	5,67 ± 0,58
Hata düzeyinde duyarlılık	0,944 ± 0,096
Yanlış hata sayısı	4,0 ± 2,6
— bunların hatasız şeritteki kısmı	2,0 ± 1,7
Eşleşen hatalarda uzunluk hatası (MAE, mm)	35,6 ± 4,4
Puanı tam tutan şerit oranı	0,766 ± 0,061
Şerit başına puan hatası (MAE)	0,33 ± 0,12
Toplam puan (gerçek = 8)	11,0 ± 3,0
P_{100} (gerçek = 900,2)	1237,8 ± 337,6

Not: SS, standart sapma; MAE, ortalama mutlak hata.

Şekil 8. Test şeritlerinde gerçek maskeden ve YOLO11s-seg tahmininden (tohum 42) hesaplanan 4 Puan Sistemi puanları



Model, altı gerçek hatanın ortalama 5,67'sini bulmuş ve şeritlerin ortalama %76,6'sında gerçek puanla aynı puanı vermiştir. Buna karşılık toplam puan gerçek değerin üzerindedir (11,0'a karşı 8). Bunun iki nedeni vardır. Birincisi, özellikle tohum 44'te hatasız şeritlerde verilen yanlış alarmlar puanlamaya ek hata olarak girmektedir. İkincisi, çizgisel hatalar birden fazla parçaya bölünerek ayrı hatalar gibi sayılabilmektedir. Uzun çizgisel hatanın (0074, gerçek uzunluk ≈155 mm, 3 puan) yalnızca bir bölümünün bulunması ise puanı düşürmektedir. Bu iki yöndeki hatalar toplamda birbirini dengeleyebilir. Nitekim tohum 42'de toplam

puan gerçek değere tam olarak eşit çıkmıştır (8), ancak Şekil 8 bu eşitliğin iki şeritteki fazla ve iki şeritteki eksik puanın birbirini götürmesinden kaynaklandığını göstermektedir. Bu nedenle otomatik puanlamanın başarısı yalnızca toplam puanla değil, şerit ve hata düzeyindeki uyumla birlikte değerlendirilmelidir.

P_{100} değerleri (gerçek: 900,2) gerçek bir kumaş topunun kalitesini temsil etmez. AITEX, hatalı görüntülerin oranı yüksek tutularak hazırlanmış bir veri setidir ve test şeritleri toplam yalnızca 0,74 m² kumaşa karşılık gelir. Bu değerler yalnızca modelin ve gerçek maskenin aynı ölçütle karşılaştırılmasını sağlamak için verilmiştir. Ayrıca bütün uzunluk ve puan hesapları 0,25 mm/piksel ölçek varsayımına dayanmaktadır. Gerçek bir hatta bu değerin kamera ve optik düzeneğe göre ölçülmesi gerekir.

Tartışma ve Sınırlılıklar

Bu çalışmanın en belirgin bulgusu, aynı mimarinin farklı ölçeklerinin küçük bir veri setinde çok farklı kararlılık gösterebilmesidir (bkz. Uygulama IV). Tablo 4'teki çalışmalarda tohumlar arası değişkenlik raporlanmamaktadır. Bu bulgu, tek tohumla verilen bir değerin bir modelin başarımını olduğundan iyi ya da kötü gösterebileceğine işaret etmektedir.

İkinci önemli bulgu, değerlendirme düzeyinin sonuçları güçlü biçimde etkilemesidir. Ayrıca hatasız görüntülerin değerlendirmeye katılması, yalnızca hatalı görüntülerle yapılan değerlendirmelerde görülemeyen bir sorunu ortaya çıkarmıştır: yanlış alarmlar. Ana model, testteki yedi hatasız şeridin ortalama %28,6'sında yanlış alarm vermiştir. Bir muayene sisteminin sahada kabul görmesi için bu oranın düşürülmesi gerekir. Güven eşiği analizi (Şekil 5), eşiğin yükseltilmesiyle yanlış alarmların büyük ölçüde azaltılabileceğini, ancak bunun hata yakalama oranını ve bölütleme kalitesini düşüreceğini göstermektedir. Görüntü düzeyindeki yüksek AUROC değerleri (Şekil 6) ise yanlış

alarmların büyük ölçüde modelin ayırma gücünden değil, düşük çalışma eşiğinden kaynaklandığını göstermektedir.

Hata türleri açısından sonuçlar belirgin bir ayrım göstermektedir. Noktasal hatalar, boyutları 1–2 mm düzeyinde olmasına karşın doğru yerde ve tutarlı biçimde bölütlenmiştir. Uzun çizgisel hatalar yalnızca kısmen bulunmuştur. Kenar hatası, ana modelin doğrulamada en iyi sonucu veren tohumu (42) tarafından hiç bulunamamış, diğer iki tohumda ise yalnızca kısmen bulunmuştur. Bunun başlıca nedenleri, kenar grubunun eğitimde çok az örnekle temsil edilmesi ve yama tabanlı çıkarımın uzun hataları parçalamasıdır. Bu gözlemler, maske IoU'sunun ince yapılara olan hassasiyetiyle (bkz. Uygulama IV) birlikte düşünüldüğünde, çizgisel hatalar için yalnızca IoU tabanlı ölçütlerin yetersiz kaldığını ve uzunluk gibi uygulamaya yönelik ölçütlerin de raporlanması gerektiğini göstermektedir.

4 Puan Sistemi ile yapılan otomatik puanlama, bölütleme maskelerinin doğrudan bir kalite kararına dönüştürülebileceğini göstermiştir. Das Gupta et al. (2024) tarafından örme kumaşlar için önerilen yaklaşımda olduğu gibi, puanlamanın gerçek bir kalite kontrol sürecine taşınması, kumaş topunun bütünü üzerinde ve fiziksel olarak kalibre edilmiş bir kamera düzeneğiyle yapılmalıdır.

Sınırlılıklar

Bu çalışmanın bulguları bazı sınırlılıklar çerçevesinde değerlendirilmelidir. İlk olarak, %5'lik test kümesi yalnızca 13 kaynak görüntüden (6 hatalı, 7 hatasız) oluştuğundan test sonuçlarının belirsizliği yüksektir; çok tohumlu değerlendirme ve doğrulama kümesi bu etkiyi azaltmakla birlikte bağımsız ve daha büyük bir test kümesinin yerini tutmaz. İkinci olarak, deneyler yalnızca AITEX veri setiyle ve yalnızca YOLO11-seg ölçekleri arasında yapılmış; farklı kumaş türlerine, görüntüleme koşullarına ve mimarilere genellenebilirlik sınanmamıştır. Üçüncü olarak,

uzunluk ve puan hesapları görüntüler üzerinde bağımsız olarak doğrulanamayan 0,25 mm/piksel ve kare piksel varsayımına dayanmakta; puanlamada delik puanları ile yarda başına en fazla 4 puan kuralı uygulanmamaktadır. AITEX'in hatalı görüntü oranı yüksek olduğundan 100 yd^2 başına puanlar gerçek bir kumaş topunu temsil etmez. Son olarak, çalışma eşiği küçük bir doğrulama kümesinde seçilmiş, varyant farkları için istatistiksel anlamlılık testi uygulanmamış, hiperparametreler sistematik olarak aranmamış ve hız ölçümleri görüntü okuma ile maske birleştirme adımlarını kapsamamaktadır.

Sonuç

Bu bölümde, tekstil kalite kontrolünde yapay zeka destekli kumaş hata tespiti kuramsal temellerinden başlanarak uçtan uca bir uygulamayla ele alınmıştır. Kuramsal kısımda kumaş hataları, 4 Puan Sistemi, klasik doku analizi yöntemleri ve derin öğrenme tabanlı sınıflandırma, tespit ve bölütleme yaklaşımları özetlenmiştir. Uygulama kısmında ise AITEX veri seti, veri sızıntısını önleyen kaynak düzeyinde tabakalı bir protokolle hazırlanmış, YOLO11-seg'in üç ölçeği çok tohumlu olarak eğitilmiş ve modeller hem yama hem de tam şerit düzeyinde değerlendirilmiştir.

Elde edilen sonuçlar şöyle özetlenebilir. YOLO11s-seg doğrulamada en yüksek ve en kararlı başarımı göstermiş ve ana model olarak seçilmiştir. Ana model test kümesindeki hatalı şeritlerin %94,4'ünü yakalamış ve 0,428 mikro Dice değerine ulaşmıştır. Maske tabanlı 4 Puan Sistemi puanlaması gerçek hataların %94,4'ünü bulmuş ve şeritlerin %76,6'sında doğru puanı vermiştir. YOLO11m-seg test kümesinde daha yüksek Dice değerine (0,521) ulaşmış, ancak doğrulamada bu üstünlüğü göstermemiş ve CPU'da yaklaşık 2,8 kat daha yavaş çalışmıştır. Görüntü düzeyindeki PR eğrileri (doğrulamada AUROC $\approx$ 0,95)

modellerin hatalı ve hatasız şeritleri güçlü biçimde ayırdığını, yanlış alarmların ise büyük ölçüde düşük çalışma eşiğinden kaynaklandığını göstermiştir. Noktasal hatalar başarıyla bölütlenmiş; uzun çizgisel hatalar, kenar hataları ve hatasız kumaştaki yanlış alarmlar ise geliştirilmesi gereken alanlar olarak belirlenmiştir.

Çalışma, bölütleme tabanlı modellerin yalnızca hatanın yerini değil, fiziksel boyutunu da ölçerek endüstride kullanılan kalite standartlarına doğrudan bağlanabileceğini göstermektedir. Bu yaklaşımın üretim hattında güvenle kullanılabilmesi için daha büyük ve çeşitli veri setleri, kalibre edilmiş görüntüleme düzenekleri ve yanlış alarm oranının düşürülmesi gerekmektedir.

Kaynakça

ASTM International. (2017). *Standard test methods for visually inspecting and grading fabrics* (ASTM D5430-13(2017)).

Bolya, D., Zhou, C., Xiao, F., & Lee, Y. J. (2019). YOLACT: Real-time instance segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 9156–9165). IEEE. https://doi.org/10.1109/ICCV.2019.00925

Cheng, L., Yi, J., Chen, A., & Zhang, Y. (2023). Fabric defect detection based on separate convolutional UNet. *Multimedia Tools and Applications, 82*(2), 3101–3122. https://doi.org/10.1007/s11042-022-13568-7

Das Gupta, A., Sadek, Z., Hossain, M. S., Toha, T. R., Mondol, A., Habiba, S. U., & Alam, S. M. M. (2024). An approach to automatic fault detection in four-point system for knitted fabric with our benchmark dataset Isl-Knit. *Heliyon, 10*(17),

Article e35931. https://doi.org/10.1016/j.heliyon.2024.e35931

Geze, R. A., & Akbaş, A. (2024). Detection and classification of fabric defects using deep learning algorithms. *Politeknik Dergisi, 27*(1), 371–378. https://doi.org/10.2339/politeknik.1386458

Habib, M. T., Faisal, R. H., Rokonuzzaman, M., & Ahmed, F. (2014). Automated fabric defect inspection: A survey of classifiers. *International Journal in Foundations of Computer Science & Technology, 4*(1), 17–25. https://doi.org/10.5121/ijfcst.2014.4102

Hanbay, K., Talu, M. F., & Özgüven, Ö. F. (2016). Fabric defect detection systems and methods—A systematic literature review. *Optik, 127*(24), 11960–11973. https://doi.org/10.1016/j.ijleo.2016.09.110

Jocher, G., & Qiu, J. (2024). *Ultralytics YOLO11* (Version 11.0.0) [Computer software]. https://github.com/ultralytics/ultralytics

Kahraman, Y., & Durmuşoğlu, A. (2022). Classification of defective fabrics using capsule networks. *Applied Sciences, 12*(10), Article 5285. https://doi.org/10.3390/app12105285

Kahraman, Y., & Durmuşoğlu, A. (2023). Deep learning-based fabric defect detection: A review. *Textile Research Journal, 93*(5–6), 1485–1503. https://doi.org/10.1177/00405175221130773

Khanam, R., & Hussain, M. (2024). *YOLOv11: An overview of the key architectural enhancements* (arXiv:2410.17725). arXiv. https://doi.org/10.48550/arXiv.2410.17725

Kumar, A. (2008). Computer-vision-based fabric defect detection: A survey. *IEEE Transactions on Industrial Electronics, 55*(1), 348–363. https://doi.org/10.1109/TIE.1930.896476

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE, 86*(11), 2278–2324. https://doi.org/10.1109/5.726791

Li, X., Wang, W., Wu, L., Chen, S., Hu, X., Li, J., Tang, J., & Yang, J. (2020). Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection. In *Advances in Neural Information Processing Systems 33* (pp. 21002–21012). Curran Associates.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2117–2125). IEEE. https://doi.org/10.1109/CVPR.2017.106

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.), *Computer Vision – ECCV 2014* (pp. 740–755). Springer. https://doi.org/10.1007/978-3-319-10602-1_48

Mao, M., & Hong, M. (2025). YOLO object detection for real-time fabric defect inspection in the textile industry: A review of YOLOv1 to YOLOv11. *Sensors, 25*(7), Article 2270. https://doi.org/10.3390/s25072270

Mao, M., & Hong, M. (2026). Lightweight transformer-enhanced YOLOv11 for real-time fabric defect detection: A systematic comparison with DETR-based architectures.

Electronics, 15(15), Article 3244. https://doi.org/10.3390/electronics15153244

Ngan, H. Y. T., Pang, G. K. H., & Yung, N. H. C. (2011). Automated fabric defect detection—A review. *Image and Vision Computing, 29*(7), 442–458. https://doi.org/10.1016/j.imavis.2011.02.002

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). IEEE. https://doi.org/10.1109/CVPR.2016.91

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (pp. 234–241). Springer. https://doi.org/10.1007/978-3-319-24574-4_28

Seidel, R., Seibel Júnior, H., & Komati, K. S. (2022). Textile defect detection using YOLOv5 on AITEX dataset. In *Anais do XIX Encontro Nacional de Inteligência Artificial e Computacional (ENIAC 2022)* (pp. 763–774). SBC. https://doi.org/10.5753/eniac.2022.227396

Silvestre-Blanes, J., Albero-Albero, T., Miralles, I., Pérez-Llorens, R., & Moreno, J. (2019). A public fabric database for defect detection methods and results. *Autex Research Journal, 19*(4), 363–374. https://doi.org/10.2478/aut-2019-0035

Thomine, S., & Snoussi, H. (2024). Distillation-based fabric anomaly detection. *Textile Research Journal, 94*(5–6), 552–565. https://doi.org/10.1177/00405175231206820

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017).

Attention is all you need. In *Advances in Neural Information Processing Systems 30* (pp. 5998–6008). Curran Associates.

Waiandt, C. A., & Seibel Júnior, H. (2024). Defect detection in the textile industry using the YOLOv8 model. In *Anais do XV Workshop de Sistemas de Informação (WSIS 2024)* (pp. 74–79). SBC. https://doi.org/10.5753/wsis.2024.33676

Xu, H., Liu, C., Duan, S., Ren, L., Cheng, G., & Hao, B. (2023). A fabric defect segmentation model based on improved Swin-Unet with Gabor filter. *Applied Sciences, 13*(20), Article 11386. https://doi.org/10.3390/app132011386

Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., & Ren, D. (2020). Distance-IoU loss: Faster and better learning for bounding box regression. *Proceedings of the AAAI Conference on Artificial Intelligence, 34*(7), 12993–13000. https://doi.org/10.1609/aaai.v34i07.6999

Zhou, Q., Mei, J., Zhang, Q., Wang, S., & Chen, G. (2021). Semi-supervised fabric defect detection based on image reconstruction and density estimation. *Textile Research Journal, 91*(9–10), 962–972. https://doi.org/10.1177/0040517520966733

BÖLÜM 2

Zor Trakeal Entübasyonun Tahmininde Yapay Zekâ Yaklaşımları: Klinik Verilerden Görüntü Tabanlı Derin Öğrenmeye

Naciye Nur Arslan[1]

Giriş

Havayolu yönetimi, genel anestezi uygulamalarının ve kritik hasta bakımının temel bileşenlerinden biridir (Apfelbaum vd., 2022). Trakeal entübasyon, havayolu açıklığının korunması ve yeterli ventilasyonun sağlanması amacıyla yaygın olarak kullanılan bir yöntemdir. Bununla birlikte bazı hastalarda anatomik, fizyolojik veya uygulamaya ilişkin faktörler nedeniyle entübasyon işlemi güçleşebilmekte ve ciddi komplikasyonlara yol açabilmektedir. Zor trakeal entübasyon, tekrarlayan entübasyon girişimleri, hipoksemi, havayolu travması ve diğer istenmeyen klinik sonuçlarla ilişkili olabilmektedir (Frerk vd., 2015). Bu nedenle zor entübasyon riskinin ameliyat öncesinde belirlenmesi, uygun havayolu yönetim stratejisinin planlanması, gerekli ekipmanın hazırlanması ve hasta güvenliğinin artırılması açısından büyük önem taşımaktadır.

[1]Arş. Gör. Dr., Kütahya Dumlupınar Üniversitesi, Mühendislik Fakültesi, Yazılım Mühendisliği Bölümü. ORCID: 0000-0002-3208-7986

Zor havayolunun ameliyat öncesi değerlendirilmesinde geleneksel olarak Mallampati sınıflaması, üst dudak ısırma testi, tiromental mesafe, sternomental mesafe, ağız açıklığı ve boyun hareketliliği gibi klinik ve antropometrik parametrelerden yararlanılmaktadır (De Rosa vd., 2025). Bu yöntemler, uygulama kolaylıkları ve düşük maliyetleri nedeniyle klinik pratikte yaygın biçimde kullanılmaktadır. Ancak tek bir değerlendirme yönteminin zor entübasyonu yüksek duyarlılık ve özgüllükle tahmin etmesi her zaman mümkün olmamaktadır (J. H. Kim vd., 2022). Ayrıca ölçümlerin bir bölümünün uygulayıcı deneyiminden etkilenmesi, hastalar arasındaki anatomik farklılıklar ve zor entübasyonun çok faktörlü yapısı, geleneksel değerlendirme yaklaşımlarının öngörü gücünü sınırlandırmaktadır. Bu durum, birden fazla klinik değişkenin birlikte değerlendirildiği daha kapsamlı tahmin yöntemlerine duyulan ihtiyacı artırmıştır.

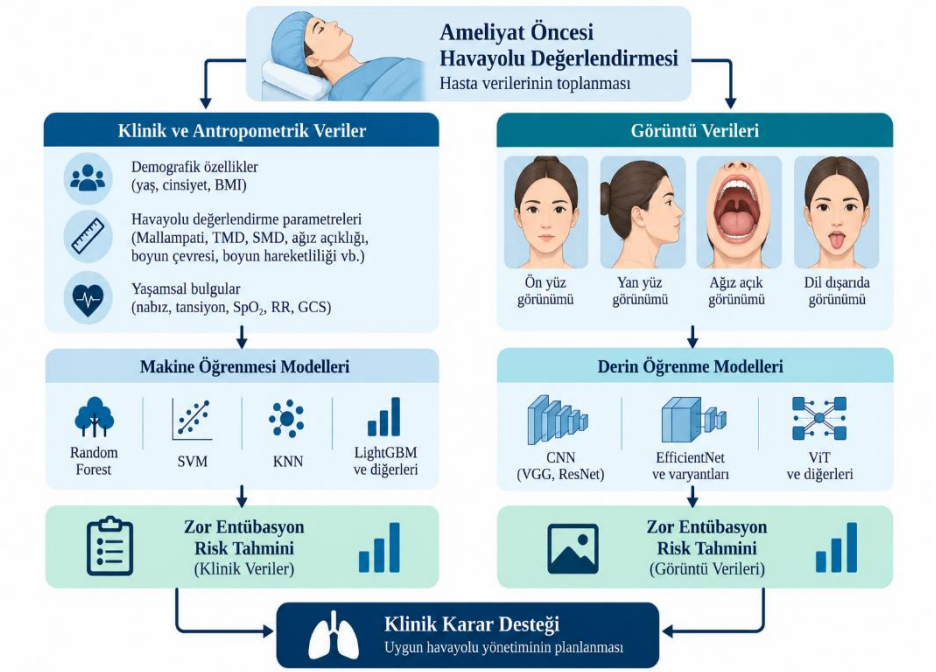
Son yıllarda yapay zekâ ve makine öğrenmesi alanındaki gelişmeler, tıbbi verilerden anlamlı örüntülerin çıkarılmasına ve klinik karar süreçlerinin desteklenmesine yönelik yeni olanaklar sunmuştur (Shen vd., 2026). Geleneksel istatistiksel yaklaşımlardan farklı olarak makine öğrenmesi algoritmaları, çok sayıda değişken arasındaki karmaşık ve doğrusal olmayan ilişkileri modelleyebilmekte ve bu ilişkilerden yararlanarak bireysel risk tahminleri oluşturabilmektedir. Zor entübasyon tahmininde gerçekleştirilen ilk yapay zekâ çalışmalarında ağırlıklı olarak yaş, cinsiyet, vücut kitle indeksi, Mallampati sınıflaması, tiromental mesafe, boyun çevresi ve ağız açıklığı gibi klinik ve antropometrik değişkenlerden oluşan yapılandırılmış veri kümeleri kullanılmıştır. Bu kapsamda Random Forest, Support Vector Machine, K-Nearest Neighbors, LightGBM ve çeşitli topluluk öğrenmesi algoritmaları değerlendirilmiştir. Klinik değişkenlerin birlikte analiz edilmesi, geleneksel değerlendirme yöntemleriyle belirlenmesi güç olabilen

ilişkilerin öğrenilmesine ve zor havayolu riskinin daha kapsamlı biçimde incelenmesine olanak sağlamıştır.

Derin öğrenme yöntemlerinin gelişmesiyle birlikte zor entübasyon tahmininde kullanılan veri türleri de çeşitlenmiştir. Özellikle evrişimli sinir ağlarının görüntülerden otomatik özellik çıkarabilme yeteneği, yüz ve baş-boyun anatomisinin doğrudan analiz edilmesine dayalı yaklaşımların geliştirilmesini sağlamıştır. Görüntü tabanlı modeller, geleneksel klinik ölçümlerde kullanılan sınırlı sayıdaki anatomik parametrenin ötesine geçerek mandibular yapı, yüz konturu, ağız çevresi ve boyun anatomisine ilişkin karmaşık görsel örüntüleri öğrenmeyi amaçlamaktadır. Bu doğrultuda başlangıçta tek bir yüz görüntüsüne dayalı sınıflandırma modelleri geliştirilirken, sonraki araştırmalarda farklı anatomik pozisyonlardan elde edilen görüntülerin birlikte değerlendirildiği çoklu görünüm temelli derin öğrenme yaklaşımları ön plana çıkmıştır (Xia vd., 2024).

Görüntü tabanlı zor entübasyon tahmininde VGG, ResNet ve EfficientNet gibi evrişimli sinir ağı mimarilerinin yanı sıra dikkat mekanizmaları, kontrastif öğrenme ve özellik birleştirme stratejilerinden yararlanılmıştır. Farklı anatomik pozisyonlardan elde edilen görüntülerin birlikte değerlendirilmesi, tek bir görüntüde yeterince temsil edilemeyen anatomik özelliklerin öğrenilmesine katkı sağlayabilecek bir yaklaşım olarak ele alınmıştır. Ayrıca akıllı telefonlarla elde edilen standart yüz görüntülerinin kullanılması, görüntü tabanlı karar destek sistemlerinin daha erişilebilir biçimde geliştirilmesine yönelik araştırmaları teşvik etmiştir. Bununla birlikte görüntüleme koşullarının standartlaştırılması, veri kümesi büyüklüğü ve farklı hasta gruplarında model performansının korunması gibi sorunlar önemini sürdürmektedir.

Şekil 1 Zor entübasyon tahmininde klinik ve görüntü tabanlı yapay zekâ yaklaşımları



Bu kitap bölümünün amacı, zor trakeal entübasyonun ameliyat öncesi tahmininde kullanılan yapay zekâ yaklaşımlarını veri türleri ve modelleme stratejileri açısından bütüncül biçimde incelemektir. Bu kapsamda öncelikle klinik ve antropometrik verilere dayalı geleneksel makine öğrenmesi yöntemleri ele alınmakta, ardından yüz ve baş-boyun görüntülerini kullanan derin öğrenme modelleri değerlendirilmektedir (Şekil 1). Son olarak mevcut araştırmaların metodolojik ve teknik sınırlılıkları değerlendirilerek gelecekte geliştirilecek güvenilir, açıklanabilir ve klinik uygulamalara uyarlanabilir yapay zekâ tabanlı karar destek sistemlerine yönelik araştırma yönelimleri ortaya konulmaktadır.

Klinik Verilere Dayalı Yapay Zekâ Yaklaşımları

Görüntü tabanlı derin öğrenme yaklaşımlarının yaygınlaşmasından önce, zor havayolu tahminine yönelik yapay zekâ araştırmaları ağırlıklı olarak klinik, demografik ve antropometrik değişkenlerden oluşan yapılandırılmış (tabular) veri kümeleri üzerinde gerçekleştirilmiştir. Bu yaklaşımların temel amacı, geleneksel yatak başı havayolu değerlendirme yöntemlerinin sınırlı öngörü gücünü geliştirmek ve farklı klinik değişkenler arasındaki karmaşık ilişkileri ortaya çıkarmaktır. Geleneksel değerlendirme yöntemleri çoğunlukla belirli anatomik özellikleri ayrı ayrı ele alırken, makine öğrenmesi algoritmaları çok sayıda değişkeni eş zamanlı değerlendirerek doğrusal olmayan ilişkileri ve etkileşimleri modelleyebilmektedir. Bu kapsamda Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Light Gradient Boosting Machine (LightGBM) ve çeşitli topluluk öğrenmesi yöntemleri zor havayolu tahmininde kullanılmıştır. Klinik verilere dayalı makine öğrenmesi çalışmalarının temel özellikleri Tablo 1'de sunulmuştur.

Tablo 1 Klinik verilere dayalı yapay zekâ çalışmalarının karşılaştırılması

Çalışma	Örneklem	En iyi model	Performans
Yamanaka vd. (2022)	10.741 (543 zor vaka)	Ridge + RF Ensemble	AUC: 0,74; Duy.: 0,67;
Kim vd. (2022)	616 (64 zor vaka)	LightGBM	AUC: 0,71; Duy.: 0,85
Sezari vd. (2024)	622 (91 zor vaka)	KNN	Duy.: 0,82; AUC: 0,87;
García-García vd. (2024)	623 (53 zor vaka)	XGBoost	AUC: 0,746-0,766
Demir Şenoğlu vd. (2025)	329 (62 zor vaka)	SVM	AUC: 0,85; Duy.: 0,67

Yamanaka ve arkadaşları tarafından gerçekleştirilen çok merkezli çalışmada, 10.741 hastaya ait klinik veriler kullanılarak yedi farklı makine öğrenmesi algoritmasının tahmin performansı

değerlendirilmiştir. Veri kümesinde 543 zor havayolu olgusu bulunmakta olup demografik özelliklerin yanı sıra modifiye LEMON kriterleri, boyun hareketliliği ve yaşamsal bulgular gibi çok sayıda klinik parametre kullanılmıştır. Çalışmada Ridge Regression ve Random Forest algoritmalarının birleştirildiği topluluk öğrenmesi yaklaşımı öne çıkmış; 0,74 AUC, 0,67 duyarlılık ve 0,70 özgüllük değerleri bildirilmiştir. Ayrıca değişkenlerin tahmin performansına katkısını değerlendirmek amacıyla delta C-istatistiğine dayalı önem analizi gerçekleştirilmiştir. Geniş örneklem büyüklüğü ve çok merkezli tasarım, çalışmanın farklı klinik özelliklere sahip hasta gruplarını kapsaması açısından önemli bir avantaj sağlamaktadır (Yamanaka vd., 2022).

Kim ve arkadaşlarının çalışmasında, 616 hastadan elde edilen klinik ve antropometrik veriler kullanılarak LightGBM tabanlı bir zor entübasyon tahmin modeli geliştirilmiştir. Çalışmada yaş, vücut kitle indeksi, Mallampati sınıflaması, sternomental mesafe ve boyun çevresi gibi geleneksel havayolu değerlendirme parametrelerinden yararlanılmıştır. Toplam 64 zor entübasyon olgusunun bulunduğu veri kümesinde modelin AUC değeri 0,71, duyarlılığı 0,85 ve kesinliği (precision) 0,18 olarak bildirilmiştir. Yüksek duyarlılığa karşın düşük kesinlik değeri, pozitif olarak sınıflandırılan hastaların önemli bir bölümünün gerçekte zor entübasyon yaşamamış olabileceğine işaret etmektedir. Çalışmada ayrıca ağaç tabanlı bölünme sayılarına dayalı özellik önem analizi gerçekleştirilerek klinik değişkenlerin model kararlarına katkısı incelenmiştir (J. H. Kim vd., 2022).

Sezari ve arkadaşları tarafından gerçekleştirilen araştırmada, 622 hastaya ait 24 farklı klinik değişken kullanılarak yedi makine öğrenmesi algoritması karşılaştırılmıştır. Değerlendirilen parametreler arasında yaş, vücut kitle indeksi, ağız açıklığı, tiromental ve sternomental mesafeler, boyun fleksiyon ve ekstansiyon açıları, hyomental mesafe, Mallampati sınıflaması ve

diş yapısına ilişkin özellikler yer almaktadır. Toplam 91 zor entübasyon olgusunun bulunduğu çalışmada KNN algoritması öne çıkmış; 0,87 doğruluk, 0,82 duyarlılık, 0,87 AUC ve 0,85 F1-skoru bildirilmiştir. Ayrıca özelliklerin tahmin üzerindeki etkilerini değerlendirmek amacıyla ağaç tabanlı özellik önem sıralaması kullanılmıştır. Bu bulgular, uygun klinik değişkenlerin seçilmesi ve birlikte değerlendirilmesi durumunda geleneksel makine öğrenmesi algoritmalarının zor havayolu tahmininde etkili sonuçlar üretebildiğini göstermektedir (Sezari vd., 2024).

García-García ve arkadaşlarının çalışması, klinik ve morfolojik özelliklerin birlikte değerlendirilmesinin yanı sıra model kararlarının yorumlanabilirliğine odaklanması bakımından dikkat çekmektedir. Araştırmada yüz morfolojisini temsil eden anatomik ölçümler ile demografik değişkenler kullanılarak XGBoost başta olmak üzere farklı makine öğrenmesi algoritmaları değerlendirilmiştir. En başarılı model için 0,766 AUC değeri bildirilmiş ve özellik önem analizleri aracılığıyla tahmin sürecine katkı sağlayan değişkenler incelenmiştir. Bu yaklaşım, yapay zekâ modellerinin yalnızca risk sınıflandırması yapmakla sınırlı kalmayıp klinik açıdan anlamlı anatomik özelliklerin belirlenmesine de katkı sağlayabileceğini göstermektedir (García-García vd., 2024).

Demir Şenoğlu ve arkadaşları ise geleneksel klinik değerlendirme parametrelerini ultrasonografik ölçümlerle birleştirerek zor entübasyon tahminini incelemiştir. Çalışmada 329 hastaya ait demografik özellikler, Mallampati sınıflaması, üst dudak ısırma testi, ağız açıklığı, boyun çevresi, tiromental ve sternomental mesafe ölçümleri kullanılmıştır. Bunlara ek olarak cilt-hyoid, cilt-epiglot ve cilt-trakea mesafeleri ile dil kalınlığı gibi ultrasonografik parametreler değerlendirilmiştir. Toplam 62 zor entübasyon olgusunun bulunduğu araştırmada SVM algoritmasıyla %89,39 doğruluk, 0,85 AUC, 0,69 F1-skoru ve 0,67 duyarlılık bildirilmiştir. Klinik ve ultrasonografik ölçümlerin birlikte kullanılması, klasik

fizik muayene bulgularının daha ayrıntılı anatomik bilgilerle desteklenebileceğini göstermektedir. Bununla birlikte ultrasonografik ölçümlerin modele sağladığı ek katkının belirlenebilmesi için yalnızca klinik değişkenleri kullanan modellerle karşılaştırmalı değerlendirme yapılması önem taşımaktadır (Demir Senoglu vd., 2025).

Klinik verilere dayalı zor entübasyon tahmin çalışmaları birlikte değerlendirildiğinde, yaş, cinsiyet, vücut kitle indeksi, Mallampati sınıflaması, tiromental mesafe, sternomental mesafe, boyun çevresi ve ağız açıklığı gibi geleneksel değerlendirme parametrelerinin sıklıkla kullanıldığı görülmektedir. Bununla birlikte çalışmalar arasında örneklem büyüklüğü, zor entübasyonun tanımlanmasında kullanılan kriterler, değişken sayısı ve model doğrulama yöntemleri açısından önemli farklılıklar bulunmaktadır. Bu nedenle farklı araştırmalarda bildirilen performans değerlerinin doğrudan karşılaştırılması, belirli bir algoritmanın diğerlerinden kesin olarak üstün olduğu sonucuna ulaşmak için yeterli değildir.

Klinik veri tabanlı makine öğrenmesi modellerinin önemli avantajları arasında düşük maliyet, rutin ameliyat öncesi değerlendirme süreçlerine kolay entegrasyon ve özellik önem analizleri aracılığıyla belirli düzeyde yorumlanabilirlik sağlanabilmesi yer almaktadır. Bununla birlikte bu modellerin önemli bir bölümü gözlemci bağımlı fizik muayene ölçümlerine dayanmaktadır. Ayrıca klinik değişkenler, üst havayolu anatomisinin karmaşık yapısını ve anatomik özellikler arasındaki uzamsal ilişkileri doğrudan temsil edememektedir. Bu sınırlılıklar, araştırmacıları yüz, baş ve boyun anatomisinin görüntüler üzerinden otomatik olarak analiz edildiği derin öğrenme yaklaşımlarına yöneltmiştir.

Görüntü Tabanlı Derin Öğrenme Yaklaşımları

Klinik ve antropometrik değişkenlere dayalı makine öğrenmesi yöntemlerinin ardından, zor trakeal entübasyonun tahmininde görüntü tabanlı derin öğrenme yaklaşımları giderek önem kazanmıştır. Geleneksel havayolu değerlendirme yöntemleri belirli anatomik mesafelerin ölçülmesine ve fizik muayene bulgularının yorumlanmasına dayanırken, görüntü tabanlı modeller yüz, baş ve boyun anatomisine ilişkin özelliklerin doğrudan görüntüler üzerinden öğrenilmesine olanak sağlamaktadır. Özellikle evrişimli sinir ağları (Convolutional Neural Networks, CNN), görüntülerdeki kenar, doku, şekil ve uzamsal örüntüleri hiyerarşik biçimde öğrenebilme yetenekleri sayesinde bu alanda yaygın olarak kullanılmaya başlanmıştır. Bu yaklaşımlarda tek bir yüz görüntüsünden tahmin gerçekleştiren modellerin yanı sıra farklı anatomik pozisyonlardan elde edilen görüntüleri birlikte değerlendiren çoklu görünüm temelli mimariler de geliştirilmiştir. Görüntü tabanlı çalışmaların temel özellikleri Tablo 2'de sunulmuştur.

Hayasaka ve arkadaşları (2021), standartlaştırılmış yüz görüntülerinden zor entübasyon tahmini gerçekleştirmek amacıyla CNN tabanlı bir model geliştirmiştir. Çalışmada 202 hastadan elde edilen 16 farklı anatomik pozisyona ait yüz görüntüleri değerlendirilmiş ve VGG-16 tabanlı transfer öğrenme yaklaşımından yararlanılmıştır. Araştırmacılar, farklı yüz pozisyonlarının tahmin performansı üzerindeki etkilerini karşılaştırarak belirli anatomik görünümlerin zor entübasyonun öngörülmesinde daha fazla bilgi sağlayabileceğini göstermiştir. En başarılı görüntü konfigürasyonu için 0,864 AUC, %80,5 doğruluk, %81,8 duyarlılık ve %83,3 özgüllük bildirilmiştir. Ayrıca sınıf aktivasyon haritaları kullanılarak modelin tahmin sırasında dikkate aldığı anatomik bölgeler görselleştirilmiştir. Çalışma, klinik ölçümlerin doğrudan modele girilmesine gerek kalmadan yüz

görüntülerinden öğrenilen özelliklerin zor entübasyon tahmininde kullanılabileceğini göstermesi açısından önemli bir erken dönem araştırmasıdır. Bununla birlikte farklı pozisyonların değerlendirilmesi, tüm pozisyonların eş zamanlı olarak ortak bir model içerisinde birleştirildiği anlamına gelmemektedir (Hayasaka vd., 2021).

Tablo 2. Görüntü tabanlı zor entübasyon tahmin çalışmalarının karşılaştırılması

Çalışma	Örneklem	Poz	Model	Performans
Hayasaka vd. (2021)	202	16	VGG-16 tabanlı CNN	AUC: 0,864; duy.: %81,8
Tavolara vd. (2021)	152	11	CNN + Attention-MIL	AUC: 0,7105
Wang vd. (2023)	1.000	5	ResNet-18	AUC: 0,9435; duy:%89,58
Li vd. (2024)	188	5	MCE-Net	AUC: 0,767; duy.: %77,50
Wu vd. (2024)	225	5	DMF-Net	Acc: %77,92; duy.: %82,50
Xia vd. (2024)	5.849	7	ResNet-18 + LightGBM	AUC: 0,779; duy.: 0,757
Kim vd. (2024)	3.053	4	EfficientNet-B5	AUC: 0,81–0,88
García-García vd. (2024)	623	2	DCNN	AUC: 0,59
Koca vd. (2026)	109	16	CNN / EfficientNet	Acc: %88,64; F1: %87,28

Not: AUC, ROC eğrisi altında kalan alanı; duy., duyarlılığı ifade etmektedir.

Tavolara ve arkadaşları (2021), yalnızca frontal yüz görüntülerinden zor entübasyon riskini belirlemek amacıyla çok aşamalı bir topluluk öğrenmesi mimarisi geliştirmiştir. Araştırmada öncelikle CASIA-WebFace veri kümesinden yararlanılarak farklı yüz bölgelerine yönelik CNN tabanlı özellik çıkarıcılar oluşturulmuştur. Ardından yüzün 11 anatomik bölgesinden elde edilen özellikler, dikkat mekanizmasına dayalı çoklu örnek öğrenmesi (Attention-based Multiple Instance Learning)

yaklaşımıyla değerlendirilmiştir. Son aşamada farklı modellerin tahminleri çoğunluk oylaması yöntemiyle birleştirilerek hasta düzeyinde sınıflandırma gerçekleştirilmiştir. Toplam 152 hastanın değerlendirildiği çalışmada 0,7105 AUC değeri bildirilmiştir. Modelin farklı karar eşiklerinde yüksek duyarlılık ile yüksek özgüllük arasında belirgin bir değişim göstermesi, klinik kullanımda karar eşiği seçiminin önemini ortaya koymaktadır. Bu çalışma, yüzün farklı anatomik bölgelerinden öğrenilen özelliklerin birleştirilmesine dayanan yaklaşımıyla tek görüntü kullanan derin öğrenme modellerinin geliştirilmesine katkı sağlamıştır (Tavolara vd., 2021).

Erken dönem görüntü tabanlı araştırmalar, yüz anatomisinin zor entübasyon tahmininde kullanılabilecek önemli bilgiler içerdiğini göstermiştir. Bununla birlikte frontal görüntüler veya sınırlı sayıdaki anatomik pozisyonlar, mandibular yapı, submandibular bölge, boyun konturu ve servikal hareket açıklığı gibi farklı anatomik özelliklerin tamamını temsil etmekte yetersiz kalabilmektedir. Bu nedenle sonraki araştırmalarda birden fazla görüntünün birlikte değerlendirilmesine ve farklı açılardan elde edilen anatomik bilgilerin bütünleştirilmesine yönelik yöntemler geliştirilmiştir.

Wang ve arkadaşları (2023), farklı açılardan elde edilen yüz görüntülerini kullanarak yarı denetimli öğrenme temelli bir zor entübasyon tahmin yaklaşımı önermiştir. Çalışmada 1.000 hastaya ait beş farklı anatomik pozisyondan elde edilen görüntüler değerlendirilmiş ve ResNet-18 mimarisinden yararlanılmıştır. Yarı denetimli öğrenme yaklaşımı sayesinde etiketlenmemiş görüntülerden de yararlanılması amaçlanmıştır. Araştırmada 0,9435 AUC, %90 doğruluk, %89,58 duyarlılık ve %90,13 özgüllük bildirilmiştir. Bu sonuçlar, çoklu görünüm bilgilerinin ve yarı denetimli öğrenme stratejilerinin zor entübasyon tahmininde kullanılabileceğini göstermektedir. Bununla birlikte yüksek

performans değerlerinin farklı hasta popülasyonlarında doğrulanması, modelin genellenebilirliğinin belirlenmesi açısından önem taşımaktadır (Wang vd., 2023).

Li ve arkadaşları (2024), çoklu yüz görüntülerinden elde edilen özelliklerin daha etkili öğrenilmesini sağlamak amacıyla Multi-view Contrastive Learning yaklaşımına dayanan MCE-Net mimarisini geliştirmiştir. Çalışmada 188 hastaya ait beş farklı anatomik pozisyondan elde edilen görüntüler kullanılmıştır. Kontrastif öğrenme, aynı hastaya veya benzer anatomik özelliklere ait görüntü temsillerinin birbirine yaklaştırılması ve farklı özelliklere sahip temsillerin ayrıştırılması yoluyla daha ayırt edici özelliklerin öğrenilmesini amaçlamaktadır. Geliştirilen model için 0,767 AUC, %77,50 duyarlılık ve %75,92 özgüllük bildirilmiştir. Bu yaklaşım, çoklu görüntüler arasındaki ilişkilerin yalnızca özelliklerin birleştirilmesiyle değil, temsil öğrenme aşamasında da değerlendirilebileceğini göstermektedir (Li vd., 2024).

Wu ve arkadaşları (2024), farklı anatomik görünümler arasındaki tamamlayıcı bilgileri öğrenmek amacıyla Dual-Path Multi-View Fusion Network (DMF-Net) mimarisini önermiştir. Araştırmada 225 hastaya ait beş farklı pozisyondan elde edilen yüz görüntüleri kullanılmıştır. Modelde farklı görüntülerden elde edilen özelliklerin bütünleştirilmesi için çift yollu öğrenme ve dikkat mekanizmalarından yararlanılmıştır. Geliştirilen yaklaşım %82,50 duyarlılık, %75,62 özgüllük, %77,92 doğruluk ve %71,35 F1-skoru elde etmiştir. Bu çalışma, çoklu görünüm bilgilerinin dikkat mekanizmaları aracılığıyla bütünleştirilmesinin zor entübasyon tahminindeki potansiyelini ortaya koymaktadır (Wu vd., 2024).

Xia ve arkadaşları (2024), görüntü tabanlı özelliklerle klinik değişkenleri birleştiren hibrit bir tahmin sistemi geliştirmiştir. Çalışmada 5.849 hastaya ait yedi farklı anatomik pozisyondan elde edilen görüntüler değerlendirilmiş, ResNet-18 mimarisiyle çıkarılan görsel özellikler LightGBM algoritması kullanılarak klinik bilgilerle

birlikte analiz edilmiştir. Model için 0,779 AUC, 0,757 duyarlılık ve 0,721 özgüllük bildirilmiştir. Ayrıca Grad-CAM++ yöntemiyle modelin görüntülerde hangi anatomik bölgelere odaklandığı incelenmiştir. Bu araştırma, görüntülerden öğrenilen yüksek boyutlu özelliklerin yapılandırılmış klinik verilerle birleştirilmesi bakımından önem taşımaktadır. Bununla birlikte klinik ve görüntü verilerinin birlikte kullanılması nedeniyle bu yaklaşımın yalnızca görüntü tabanlı bir model olarak değil, çok modlu öğrenme kapsamında da değerlendirilmesi gerekmektedir (Xia vd., 2024).

Kim ve arkadaşları (2024), akıllı telefonlarla elde edilen yüz ve boyun görüntülerinden zor havayolu tahmini gerçekleştirmek amacıyla EfficientNet-B5 mimarisini kullanmıştır. Çalışmada 3.053 hastaya ait dört farklı anatomik pozisyon değerlendirilmiş ve 0,81–0,88 aralığında AUC değerleri bildirilmiştir. Ayrıca Grad-CAM yöntemi kullanılarak modelin tahmin sürecinde dikkate aldığı görüntü bölgeleri görselleştirilmiştir. Akıllı telefonlarla elde edilen görüntülerin kullanılması, derin öğrenme tabanlı havayolu değerlendirme sistemlerinin özel görüntüleme ekipmanlarına bağımlılığını azaltabilecek bir yaklaşım sunmaktadır. Bununla birlikte görüntü kalitesi, çekim mesafesi, aydınlatma koşulları ve hasta pozisyonunun standartlaştırılması, bu sistemlerin klinik ortamlarda güvenilir biçimde uygulanabilmesi açısından önemlidir (J.-H. Kim vd., 2024).

García-García ve arkadaşlarının (2024) araştırmasında, doğrudan yüz görüntülerinden tahmin gerçekleştiren derin evrişimli sinir ağı yaklaşımı değerlendirilmiştir. Toplam 623 hastanın yer aldığı çalışmada iki farklı görüntü konfigürasyonu kullanılmış ve görüntü tabanlı model için 0,59 AUC değeri bildirilmiştir. Bu sonuç, görüntü tabanlı derin öğrenme yaklaşımlarının her veri kümesinde yüksek ayırt edici performans göstermediğini ortaya koymaktadır. Model başarısının görüntüleme protokolü, örneklem büyüklüğü,

hedef değişken tanımı ve eğitim stratejisi gibi faktörlerden etkilenebileceği dikkate alınmalıdır (García-García vd., 2024).

Koca ve arkadaşları (2026), farklı anatomik pozisyonlardan elde edilen yüz görüntülerini CNN ve EfficientNet tabanlı mimarilerle değerlendirerek zor entübasyon tahminine yönelik bir yaklaşım sunmuştur. Çalışmada 109 hastaya ait 16 farklı pozisyon değerlendirilmiş ve en başarılı model için %88,64 doğruluk ile %87,28 F1-skoru bildirilmiştir. Araştırma, çoklu anatomik görünümlerin değerlendirilmesine yönelik güncel çalışmalar arasında yer almakla birlikte sınırlı örneklem büyüklüğü, sonuçların daha geniş ve bağımsız hasta gruplarında doğrulanmasını gerekli kılmaktadır (Koca vd., 2026).

Görüntü tabanlı zor entübasyon tahmin araştırmaları birlikte değerlendirildiğinde, yöntemlerin tek görüntüye dayalı CNN modellerinden çoklu görünüm, kontrastif öğrenme ve dikkat mekanizmaları kullanan mimarilere doğru çeşitlendiği görülmektedir. Bu gelişim, farklı anatomik pozisyonlardan elde edilen bilgilerin birlikte değerlendirilmesine olanak sağlamıştır. Bununla birlikte incelenen çalışmaların örneklem büyüklükleri, hedef değişkenleri ve doğrulama yöntemleri farklı olduğundan, daha fazla görüntü kullanılmasının veya daha karmaşık mimarilerin tek başına üstün tahmin performansı sağladığı sonucuna ulaşılamamaktadır.

Görsel açıklanabilirlik yöntemleri, modellerin tahmin sırasında hangi anatomik bölgelere duyarlı olduğunu incelemeye yardımcı olmaktadır. Ancak bu görselleştirmeler anatomik özelliklerle tahminler arasında nedensel ilişki bulunduğunu kanıtlamamaktadır. Görüntü tabanlı modellerin klinik kullanım açısından değerlendirilmesinde, açıklanabilirlik bulgularının yanı sıra farklı hasta gruplarında doğrulama sonuçlarının da dikkate alınması gerekmektedir.

Mevcut Yaklaşımların Sınırlılıkları ve Gelecek Araştırma Yönelimleri

Zor trakeal entübasyonun tahmininde kullanılan yapay zekâ yaklaşımları, son yıllarda makine öğrenmesi ve derin öğrenme alanındaki gelişmelerle birlikte önemli ilerlemeler kaydetmiştir. Klinik ve antropometrik değişkenlere dayalı geleneksel makine öğrenmesi modellerinden, yüz ve boyun görüntülerini analiz eden derin öğrenme mimarilerine ve farklı veri türlerini birleştiren çok modlu sistemlere kadar çeşitli yöntemler geliştirilmiştir. Bununla birlikte literatürde bildirilen umut verici performans sonuçlarına rağmen, bu modellerin rutin klinik uygulamalarda güvenilir biçimde kullanılmasını sınırlandıran önemli metodolojik ve teknik sorunlar bulunmaktadır. Özellikle veri kümelerinin büyüklüğü ve çeşitliliği, hedef değişkenlerin tanımlanmasındaki farklılıklar, sınıf dengesizliği, doğrulama stratejileri, açıklanabilirlik ve klinik genellenebilirlik konuları dikkat çekmektedir.

Veri Kümesi Büyüklüğü ve Çeşitliliği

Mevcut araştırmaların önemli sınırlılıklarından biri, modellerin çoğunlukla tek merkezden elde edilen ve sınırlı sayıda zor entübasyon olgusu içeren veri kümeleri üzerinde geliştirilmesidir. Zor entübasyonun görece düşük görülme sıklığı nedeniyle, toplam örneklem büyüklüğü yeterli görünse bile pozitif sınıfa ait hasta sayısı sınırlı kalabilmektedir. Bu durum özellikle çok sayıda öğrenilebilir parametre içeren derin öğrenme modellerinde aşırı öğrenme riskini artırmaktadır. Modelin eğitim verilerindeki örüntüleri başarılı biçimde öğrenmesi, farklı hasta popülasyonlarında aynı performansı göstereceği anlamına gelmemektedir.

Görüntü tabanlı araştırmalarda bu sorun daha belirgin hâle gelebilmektedir. Yüz ve boyun anatomisinin yaş, cinsiyet, vücut yapısı ve bireysel anatomik farklılıklardan etkilenmesi, eğitim veri

kümesinin çeşitliliğini önemli kılmaktadır. Ayrıca görüntülerin farklı cihazlarla, aydınlatma koşullarında ve çekim mesafelerinde elde edilmesi, modellerin öğrenmiş olduğu özelliklerin değişmesine neden olabilmektedir. Bu nedenle gelecekte gerçekleştirilecek araştırmalarda farklı merkezlerden, farklı hasta gruplarından ve standartlaştırılmış görüntüleme protokollerinden elde edilen daha geniş veri kümelerinin oluşturulması önem taşımaktadır.

Hedef Değişken Tanımlarındaki Farklılıklar

Zor entübasyon tahminine yönelik yapay zekâ çalışmalarında karşılaşılan temel metodolojik sorunlardan biri, zor havayolunun farklı klinik kriterler kullanılarak tanımlanmasıdır. Literatürde Cormack–Lehane sınıflaması, entübasyon girişim sayısı, alternatif havayolu cihazı kullanımı ve entübasyonun gerçekleştirilmesi için gereken ek müdahaleler gibi farklı ölçütlerden yararlanılmaktadır. Bununla birlikte bu kriterler her zaman aynı klinik durumu temsil etmemektedir. Örneğin Cormack–Lehane III–IV derecesi zor laringoskopiyi gösterirken, birden fazla entübasyon girişimi gerçekleştirilmesi teknik güçlüklerin yanı sıra uygulayıcı deneyimi ve kullanılan ekipmanlardan da etkilenebilmektedir.

Hedef değişkenin tanımlanma biçimi, yalnızca pozitif ve negatif sınıflardaki hasta sayısını değil, modelin öğrenmeye çalıştığı klinik örüntüleri de doğrudan etkilemektedir. Aynı hasta popülasyonu ve aynı algoritma kullanılsa bile farklı hedef tanımları altında geliştirilen modellerin duyarlılık, özgüllük, ROC-AUC ve F1-skoru değerleri değişebilmektedir. Dolayısıyla farklı çalışmalarda bildirilen performans farklılıklarının bir bölümü kullanılan algoritmalardan ziyade hedef değişkenlerin heterojenliğinden kaynaklanabilir.

Gelecekteki araştırmalarda zor laringoskopi, zor trakeal entübasyon ve başarısız entübasyon gibi klinik sonuçların açık biçimde ayrıştırılması, kullanılan hedef tanımlarının ayrıntılı olarak

raporlanması ve mümkün olduğunda birden fazla klinik sonucun ayrı ayrı değerlendirilmesi önerilmektedir. Böylece geliştirilen modellerin hangi klinik problemi tahmin ettiği daha açık biçimde ortaya konulabilecek ve araştırmalar arasındaki karşılaştırılabilirlik artırılabilecektir.

Sınıf Dengesizliği ve Model Performansının Değerlendirilmesi

Zor entübasyon olgularının kolay entübasyon olgularına kıyasla daha az sayıda bulunması, veri kümelerinde sınıf dengesizliğine yol açmaktadır. Bu durum, makine öğrenmesi modellerinin çoğunluk sınıfına yönelmesine ve zor entübasyon olgularını belirlemede yetersiz kalmasına neden olabilmektedir. Özellikle doğruluk değerinin tek başına değerlendirilmesi, dengesiz veri kümelerinde yanıltıcı sonuçlar oluşturabilir. Zor entübasyon olgularının az olduğu bir veri kümesinde, hastaların büyük bölümünü kolay entübasyon olarak sınıflandıran bir model yüksek doğruluk elde etmesine rağmen klinik açıdan yetersiz duyarlılık gösterebilir.

Bu sorunun azaltılması amacıyla rastgele aşırı örnekleme, rastgele eksik örnekleme, Synthetic Minority Over-sampling Technique (SMOTE), Adaptive Synthetic Sampling (ADASYN) ve sınıf ağırlıklandırma gibi yöntemlerden yararlanılabilmektedir. Bununla birlikte yeniden örnekleme işlemlerinin eğitim verileri üzerinde uygulanması ve doğrulama ya da test verilerinin özgün dağılımının korunması gerekmektedir. Aksi durumda sentetik veya tekrarlanan örneklerin değerlendirme verilerine taşınması, veri sızıntısına ve performansın olduğundan yüksek tahmin edilmesine neden olabilir.

Model performansının değerlendirilmesinde ROC-AUC değerinin yanı sıra duyarlılık, özgüllük, kesinlik, F1-skoru, dengeli doğruluk ve precision–recall eğrisi altında kalan alan gibi ölçütlerin birlikte raporlanması önemlidir. Ayrıca modellerin ürettiği risk

olasılıklarının gerçek olay sıklıklarıyla uyumunu değerlendirmek amacıyla kalibrasyon analizlerinin gerçekleştirilmesi gerekmektedir. Klinik uygulamalarda zor entübasyon olgularının gözden kaçırılmasının sonuçları dikkate alındığında, karar eşiklerinin yalnızca istatistiksel performansa göre değil, yanlış negatif ve yanlış pozitif tahminlerin klinik etkileri göz önünde bulundurularak belirlenmesi uygun olacaktır.

Veri Sızıntısı ve Doğrulama Stratejileri

Yapay zekâ tabanlı zor entübasyon tahmin modellerinin güvenilirliğini etkileyen bir diğer önemli konu, eğitim ve değerlendirme süreçlerinde veri sızıntısının önlenmesidir. Özellikle aynı hastaya ait birden fazla görüntünün kullanıldığı araştırmalarda, görüntülerin rastgele olarak eğitim ve test kümelerine dağıtılması aynı hastanın farklı görüntülerinin her iki kümede de bulunmasına yol açabilir. Bu durumda model, daha önce karşılaşmadığı hastalara genellenebilir anatomik özellikler öğrenmek yerine hastaya özgü görsel örüntülerden yararlanabilir. Sonuç olarak bildirilen performans değerleri gerçek klinik genellenebilirliği yansıtmayabilir.

Bu sorunun önlenmesi amacıyla veri ayrımının görüntü düzeyinde değil, hasta düzeyinde gerçekleştirilmesi gerekmektedir. Benzer şekilde özellik seçimi, normalizasyon, eksik veri tamamlama ve yeniden örnekleme gibi işlemler de yalnızca ilgili eğitim kümesi üzerinde öğrenilmelidir. Hiperparametre optimizasyonunun değerlendirme verilerinden bağımsız yürütülmesi ve mümkün olduğunda iç içe çapraz doğrulama veya bağımsız test kümelerinin kullanılması, performans tahminlerinin güvenilirliğini artıracaktır.

Bununla birlikte iç doğrulama yöntemleri modelin aynı veri kaynağındaki performansını değerlendirmek için yararlı olsa da farklı klinik merkezlerdeki başarısını tek başına garanti etmemektedir. Bu nedenle gelecekte geliştirilecek modellerin

bağımsız hastanelerden elde edilen veriler üzerinde dış doğrulamaya tabi tutulması ve mümkün olduğunda prospektif çalışmalarla değerlendirilmesi önem taşımaktadır.

Açıklanabilirlik ve Klinik Güvenilirlik

Derin öğrenme modellerinin yüksek boyutlu verilerden karmaşık ilişkileri öğrenebilmesi önemli bir avantaj sağlamakla birlikte, model kararlarının hangi özelliklere dayandığının anlaşılması güç olabilmektedir. Bu durum, özellikle hasta güvenliğinin ön planda olduğu anesteziyoloji uygulamalarında önemli bir sınırlılık oluşturmaktadır. Klinik karar destek sistemlerinin yalnızca risk tahmini üretmesi değil, bu tahminlerin hangi klinik veya anatomik özelliklerle ilişkili olduğuna dair anlaşılabilir bilgiler sunması da önemlidir.

Tabular veri kullanan modellerde SHapley Additive exPlanations (SHAP), permütasyon temelli özellik önemi ve benzeri yöntemler aracılığıyla değişkenlerin tahminlere katkısı incelenebilmektedir. Görüntü tabanlı modellerde ise Class Activation Mapping (CAM), Gradient-weighted Class Activation Mapping (Grad-CAM) ve Grad-CAM++ gibi yöntemlerle modelin tahmin sırasında hangi görüntü bölgelerine duyarlı olduğu görselleştirilebilmektedir. Bununla birlikte bu açıklamaların model davranışını yaklaşık olarak temsil ettiği ve tek başına nedensel ilişki ortaya koymadığı dikkate alınmalıdır.

Gelecekteki çalışmalarda açıklanabilirlik yöntemlerinin yalnızca görsel örneklerle sunulması yerine, farklı hasta gruplarında tutarlılıklarının değerlendirilmesi ve klinik uzman görüşleriyle karşılaştırılması yararlı olacaktır. Ayrıca modelin tahmin belirsizliğini gösterebilmesi, güvenilir olmayan durumlarda kullanıcıyı uyarabilmesi ve karar destek çıktılarının klinisyen tarafından yorumlanabilir biçimde sunulması, güvenli klinik uygulamalar açısından önem taşımaktadır.

Çok Modlu Öğrenme ve Gelişmiş Model Mimarileri

Gelecekteki araştırmalar açısından öne çıkan yönelimlerden biri, farklı veri kaynaklarının aynı tahmin sistemi içerisinde bütünleştirilmesidir. Klinik ve antropometrik değişkenler, hastanın genel özellikleri ile belirli anatomik ölçümleri temsil ederken; yüz ve boyun görüntüleri, anatomik yapıların uzamsal ilişkilerine yönelik tamamlayıcı bilgiler sağlayabilmektedir. Ultrasonografik ölçümler ve diğer görüntüleme yöntemleri de doğrudan fizik muayene ile değerlendirilemeyen anatomik özelliklerin modele dahil edilmesine olanak tanıyabilir.

Bu doğrultuda geliştirilecek çok modlu modellerde klinik değişkenlerin tabular öğrenme algoritmalarıyla, görüntülerin ise CNN veya Vision Transformer tabanlı mimarilerle işlenmesi ve elde edilen özelliklerin ortak bir tahmin katmanında birleştirilmesi mümkündür. Erken birleştirme, özellik düzeyinde birleştirme ve karar düzeyinde birleştirme gibi farklı stratejilerin karşılaştırılması, veri kaynaklarının birbirini ne ölçüde tamamladığının belirlenmesine katkı sağlayabilir.

Özellikle farklı anatomik pozisyonlardan elde edilen görüntülerin birlikte değerlendirilmesi, tek bir görüntüyle temsil edilemeyen yapısal özelliklerin öğrenilmesi açısından önemli bir araştırma alanıdır. Bununla birlikte çoklu görüntü kullanımının hesaplama maliyetini, veri toplama süresini ve model karmaşıklığını artırabileceği dikkate alınmalıdır. Bu nedenle parametre paylaşımı, dikkat mekanizmaları ve verimli özellik birleştirme stratejileri gibi yaklaşımlar, gelecekte geliştirilecek sistemlerde araştırılabilecek yöntemler arasında yer almaktadır.

Klinik Uygulamaya Entegrasyon ve Gelecek Perspektifi

Yapay zekâ tabanlı zor entübasyon tahmin sistemlerinin klinik uygulamalarda kullanılabilmesi için yüksek ayırt edici performans tek başına yeterli değildir. Modellerin farklı hasta

gruplarında güvenilir sonuçlar üretmesi, veri toplama süreçlerinin standartlaştırılması, tahminlerin klinisyenler tarafından yorumlanabilmesi ve mevcut ameliyat öncesi değerlendirme süreçlerine uyum sağlaması gerekmektedir. Ayrıca kullanılan verilerin güvenliği, hasta mahremiyeti ve sistemlerin klinik kullanım sırasında izlenmesi de dikkate alınmalıdır.

Akıllı telefonlarla elde edilen standart yüz görüntülerini klinik bilgilerle birleştiren sistemler, düşük maliyetli ve erişilebilir karar destek uygulamalarının geliştirilmesi açısından potansiyel taşımaktadır. Bununla birlikte bu tür sistemlerin klinik yararının belirlenebilmesi için yalnızca tahmin performansının değil, havayolu yönetim planlaması, hazırlık süreçleri ve hasta sonuçları üzerindeki etkilerinin de değerlendirilmesi gerekmektedir. Prospektif ve çok merkezli araştırmalar, yapay zekâ modellerinin gerçek klinik koşullardaki etkinliğinin belirlenmesinde önemli rol oynayacaktır.

Sonuç

Bu kitap bölümünde incelenen araştırmalar, zor trakeal entübasyonun ameliyat öncesi tahmininde klinik verilere dayalı makine öğrenmesi ve görüntü tabanlı derin öğrenme yöntemlerinin kullanılabileceğini ortaya koymaktadır. Klinik modeller, rutin havayolu değerlendirmesinde elde edilen yapılandırılmış değişkenleri birlikte analiz edebilme avantajına sahipken, görüntü tabanlı modeller yüz ve baş-boyun anatomisine ilişkin özelliklerin doğrudan öğrenilmesine olanak sağlamaktadır. Bununla birlikte mevcut bulgular, bu iki yaklaşımın birbirine karşı kesin bir üstünlüğünü göstermemektedir.

İncelenen çalışmaların en önemli metodolojik farklılıklarından biri, tahmin edilen klinik sonuçların aynı biçimde tanımlanmamasıdır. Zor laringoskopi, zor trakeal entübasyon, zor videolaringoskopi ve ilk girişimde başarısızlık gibi sonuçlar

birbiriyle ilişkili olmakla birlikte farklı klinik durumları temsil etmektedir. Bu nedenle araştırmalarda bildirilen performans değerleri, kullanılan veri türü veya algoritmanın üstünlüğünü belirlemek amacıyla doğrudan karşılaştırılmamalıdır.

Geliştirilen modellerin klinik uygulanabilirliği, ayırt edici performanslarının yanı sıra farklı hasta popülasyonlarında güvenilir sonuçlar üretmelerine ve tahminlerinin klinik açıdan yorumlanabilmesine bağlıdır. Bu doğrultuda hedef değişkenlerin açık biçimde tanımlandığı, hasta düzeyinde uygun veri ayrımının gerçekleştirildiği ve bağımsız veri kümeleriyle doğrulanan araştırmaların önemi devam etmektedir. Klinik ve görüntü tabanlı bilgilerin birlikte değerlendirilmesi de gelecekte araştırılabilecek bir yaklaşım olmakla birlikte, sağladığı ek katkının doğrudan karşılaştırmalarla gösterilmesi gerekmektedir.

Sonuç olarak yapay zekâ, zor havayolu değerlendirmesinde geleneksel yöntemleri tamamlayabilecek bir karar destek aracı olarak değerlendirilebilir. Bu alandaki ilerlemenin temel ölçütü yalnızca daha yüksek tahmin performansına ulaşılması değil, geliştirilen modellerin klinik açıdan anlamlı sonuçları güvenilir biçimde öngörebilmesi ve gerçek havayolu yönetimi süreçlerine ölçülebilir katkı sağlaması olacaktır.

Kaynakça

Apfelbaum, J. L., Hagberg, C. A., Connis, R. T., Abdelmalak, B. B., Agarkar, M., Dutton, R. P., Fiadjoe, J. E., Greif, R., Klock, P. A., Mercier, D., Myatra, S. N., O'Sullivan, E. P., Rosenblatt, W. H., Sorbello, M., & Tung, A. (2022). 2022 American Society of Anesthesiologists Practice Guidelines for Management of the Difficult Airway *. *Anesthesiology*, *136*(1), 31-81. https://doi.org/10.1097/ALN.0000000000004002

De Rosa, S., Bignami, E., Bellini, V., & Battaglini, D. (2025). The Future of Artificial Intelligence Using Images and Clinical Assessment for Difficult Airway Management. *Anesthesia & Analgesia*, *140*(2), 317-325. https://doi.org/10.1213/ANE.0000000000006969

Demir Senoglu, G., Kara Timucin, E., & Kırısoglu, S. (2025). Evaluation of anthropometric and ultrasonographic measurements with different machine learning methods in predicting difficult intubation: A prospective observational study. *BMC Anesthesiology*, *25*(1), 496. https://doi.org/10.1186/s12871-025-03370-x

Frerk, C., Mitchell, V. S., McNarry, A. F., Mendonca, C., Bhagrath, R., Patel, A., O'Sullivan, E. P., Woodall, N. M., & Ahmad, I. (2015). Difficult Airway Society 2015 guidelines for management of unanticipated difficult intubation in adults. *BJA: British Journal of Anaesthesia*, *115*(6), 827-848. https://doi.org/10.1093/bja/aev371

García-García, F., Lee, D.-J., Mendoza-Garcés, F. J., & García-Gutiérrez, S. (2024). Reliable prediction of difficult airway for tracheal intubation from patient preoperative photographs by machine learning methods. *Computer Methods and Programs in Biomedicine*, *248*, 108118. https://doi.org/10.1016/j.cmpb.2024.108118

Hayasaka, T., Kawano, K., Kurihara, K., Suzuki, H., Nakane, M., & Kawamae, K. (2021). Creation of an artificial intelligence model for intubation difficulty classification by deep learning (convolutional neural network) using face images: An observational study. *Journal of Intensive Care*, *9*(1), 38. https://doi.org/10.1186/s40560-021-00551-x

Kim, J. H., Choi, J. W., Kwon, Y. S., & Kang, S. S. (2022). Predictive model for difficult laryngoscopy using machine learning: Retrospective cohort study. *Brazilian Journal of*

Anesthesiology (English Edition), *72*(5), 622-628. https://doi.org/10.1016/j.bjane.2021.06.016

Kim, J.-H., Jung, H.-S., Lee, S.-E., Hou, J.-U., & Kwon, Y.-S. (2024). Improving difficult direct laryngoscopy prediction using deep learning and minimal image analysis: A single-center prospective study. *Scientific Reports*, *14*(1), 14209. https://doi.org/10.1038/s41598-024-65060-x

Koca, E., Kutlusoy, S., Er, M. B., & Koca, T. (2026). Predicting Difficult Tracheal Intubation Using Multi-Angle Photographic Analysis with Convolutional Neural Networks and EfficientNet. *Diagnostics*, *16*(7), 1042. https://doi.org/10.3390/diagnostics16071042

Li, X., Peng, B., Yao, Y., Zhang, G., & Xie, Z. (2024). Difficult airway assessment with Multi-View Contrastive Representation Prior and Ensemble Classification. *Biomedical Signal Processing and Control*, *98*, 106738. https://doi.org/10.1016/j.bspc.2024.106738

Sezari, P., Kohzadi, Z., Dabbagh, A., Jafari, A., Khoshtinatan, S., Mottaghi, K., Kohzadi, Z., & Rahmatizadeh, S. (2024). Unravelling intubation challenges: A machine learning approach incorporating multiple predictive parameters. *BMC Anesthesiology*, *24*(1), 453. https://doi.org/10.1186/s12871-024-02842-w

Shen, Y., Wu, Y., Qiu, Y., & Wu, J. (2026). AI-based decision models for difficult airway assessment: From research innovation to clinical implementation—a narrative review. *Frontiers in Medicine*, *13*, 1818061. https://doi.org/10.3389/fmed.2026.1818061

Tavolara, T. E., Gurcan, M. N., Segal, S., & Niazi, M. K. K. (2021). Identification of difficult to intubate patients from frontal face images using an ensemble of deep learning models. *Computers in Biology and Medicine*, *136*, 104737. https://doi.org/10.1016/j.compbiomed.2021.104737

Wang, G., Li, C., Tang, F., Wang, Y., Wu, S., Zhi, H., Zhang, F., Wang, M., & Zhang, J. (2023). A fully-automatic semi-supervised deep learning model for difficult airway assessment. *Heliyon*, *9*(5), e15629. https://doi.org/10.1016/j.heliyon.2023.e15629

Wu, J., Yao, Y., Zhang, G., Li, X., & Peng, B. (2024). Difficult Airway Assessment Based on Multi-View Metric Learning. *Bioengineering*, *11*(7), 703. https://doi.org/10.3390/bioengineering11070703

Xia, M., Jin, C., Zheng, Y., Wang, J., Zhao, M., Cao, S., Xu, T., Pei, B., Irwin, M. G., Lin, Z., & Jiang, H. (2024). Deep learning-based facial analysis for predicting difficult videolaryngoscopy: A feasibility study. *Anaesthesia*, *79*(4), 399-409. https://doi.org/10.1111/anae.16194

Yamanaka, S., Goto, T., Morikawa, K., Watase, H., Okamoto, H., Hagiwara, Y., & Hasegawa, K. (2022). Machine Learning Approaches for Predicting Difficult Airway and First-Pass Success in the Emergency Department: Multicenter Prospective Observational Study. *Interactive Journal of Medical Research*, *11*(1), e28366. https://doi.org/10.2196/28366

BÖLÜM 3

UYDU GÖRÜNTÜSÜ VERİ SETLERİNDE BENZERLİK TESPİTİ İÇİN HİBRİT SİSTEM

MUSTAFA KURÇ[1]
AYHAN İSTANBULLU[2]

Giriş

Gelişen uzaktan algılama teknolojileriyle birlikte, farklı kaynaklardan elde edilen uydu görüntülerinin hacmi ve çeşitliliği muazzam bir artış göstermiştir. Bu durum, görüntü veri setleri içinde yer alan benzer kısımların, ortak alanların ve örtüşen bölgelerin geleneksel yöntemlerle tespit edilmesini zaman ve maliyet açısından oldukça zorlaştırmaktadır. Klasik bilgisayarlı görü yaklaşımları, büyük arama uzaylarında yüksek hesaplama maliyeti yaratırken; tek başına derin öğrenme yaklaşımları ise pikseller arası hassas geometrik doğrulamada yetersiz kalabilmektedir.

Bu çalışmada; uydu görüntülerinden oluşan veri setindeki görüntü ikilileri arasında otonom benzerlik ve ortak alan tespiti gerçekleştiren, derin öğrenme mimarileri ile klasik bilgisayarlı görü

[1] Yüksek Lisans Öğrencisi, Balıkesir Üniversitesi,Bilgisayar ve Bilişim Mühendisliği Anabilim Dalı, Orcid: 0009-0002-3319-1184
[2] Prof. Dr., Balıkesir Üniversitesi,Bilgisayar ve Bilişim Mühendisliği Anabilim Dalı, Orcid: 0000-0002-7066-4238

algoritmalarının hibrit entegrasyonuna dayalı bir sistem geliştirilmiştir. Çalışmanın temel amacı, arama uzayını derin öğrenmeyle optimize edip, nihai eşleşmeyi geometrik algoritmalarla doğrulayarak ölçekten bağımsız bir görüntü benzerlik sistemi sunmaktır.

Uydu görüntülerinin elde edilmesinde Google Earth Engine (Gorelick et al., 2017) bulut tabanlı coğrafi analiz platformu kullanılmıştır. Görüntüler, Copernicus Sentinel-2 Surface Reflectance Harmonized (S2_SR_HARMONIZED) veri kümesinden (European Space Agency, 2025) Google Earth Engine API aracılığıyla temin edilmiştir. Veri kümesi çalışma bölgesi, tarih aralığı (01.06.2025–01.09.2025) ve bulutluluk oranı (%5'in altında) kriterlerine göre filtrelenmiş, ardından medyan bileşik görüntü oluşturulmuştur. Son olarak kırmızı (B4), yeşil (B3) ve mavi (B2) bantları kullanılarak RGB görüntüler oluşturulmuştur. İstanbul/Marmara Bölgesi referans alınarak (başlangıç: 41.0082° K, 29.0100° D) coğrafi olarak $0.054°$ enlem, $0.072°$ boylam adımıyla 40×50 boyutlarında düzenli bir ızgara (grid) ağı oluşturulmuştur. Bu ızgara üzerinden belirlenen 2.000 farklı koordinat noktasının her birinden iki farklı ölçekte (yakın ve uzak görünüm) olmak üzere toplam 4.000 adet görüntü üretilmiş ve veri kümesine kaydedilmiştir.

Bu doğrultuda, ImageNet (Deng et al., 2009) ile önceden eğitilmiş ResNet50 derin öğrenme modeli (He et al., 2016) bu veri seti üzerinden Google Colab (Google, 2026a) ortamında, transfer öğrenme (transfer learning) ve ince ayar (fine-tuning) süreçlerine tabi tutulmuştur. Modelin eğitiminde sınıflandırma gücünden ziyade, modelin veri tabanındaki görüntülere ait ölçekten bağımsız öznitelik vektörlerini öğrenmesi amaçlanmıştır. Modelin vektör çıkarıcı özelliği, klasik bilgisayarlı görü algoritmalarıyla hibrit kullanılarak web tabanlı bir benzerlik bulma uygulaması geliştirilmiştir.

Literatür Özeti ve Temel Kavramlar

Görüntü işleme literatüründe benzerlik tespiti için klasik RGB farkı tespiti ve SSIM gibi algoritmaların kullanıldığı çalışmalar bulunmaktadır (Kurç et al., 2026); Bu yöntemler, basit bilgisayarlı görü görevlerinde hesaplama yükü açısından avantajlı olsalar da görüntüler arasında ortak alanların ve benzerliklerin bulunması çalışmalarında atmosferik etkiler, sensör gürültüleri ve zamansal aydınlatma değişimleri nedeniyle görüntülerin doğrudan kıyaslanmasında yetersiz kalmaktadır. Bu yaklaşımların temel kısıtı; görüntüler arasında çözünürlük (ölçek), rotasyon (açı) veya belirgin aydınlatma farkları meydana geldiğinde, benzerlik skorlarının doğruluğunun önemli ölçüde azalmasıdır.

Klasik SSIM (Yapısal Benzerlik İndeksi) (Wang et al., 2004), piksel değerlerinin mutlak farklarına odaklanmak yerine; görüntünün parlaklık, kontrast ve yapısal bileşenlerini ayrıştırarak inceler. Bu sebeple SSIM, aydınlatma ve parlaklık değişimlerine karşı geleneksel piksel tabanlı yöntemlere göre daha yüksek doğruluk gösterir. Ancak SSIM, hala piksellerin uzamsal dizilimine ve global hizalamaya bağımlı (intensity-based) bir metriktir. Bu durum, görüntüler arasında ölçekleme (scale), rotasyon (rotation) veya bakış açısı (affine) gibi değişiklikler meydana geldiğinde SSIM'in benzerlik skorlarında düşüşlere yol açmaktadır. Bu sebeple SSIM; SIFT gibi görüntünün geometrisinden bağımsız, yerel ve invariant (değişmez) öznitelik noktalarını (keypoints) temel alan özellik tabanlı algoritmaların gerisinde kalmaktadır.

Bu zafiyetler, modern sistemlerde piksel kaymalarından ve ölçek değişimlerinden bağışık özellik tabanlı bilgisayarlı görü algoritmalarının ve derin öznitelik uzaylarının (deep feature spaces) kullanılması gerekliliğini ortaya koymuştur. SIFT (Scale-Invariant Feature Transform) algoritması (Lowe, 2004), görüntüdeki belirgin köşeleri ve yapısal kilit noktaları (keypoints) tespit edip, ölçek ve açı farklılıklarını başarılı bir şekilde tolere edebilmektedir. Ancak

görüntüler eşleştirilirken, SIFT algoritması doku farklılıklarını yanlış yorumlayarak hatalı eşleşmeler üretebilmektedir. Bu sorunu aşmak ve sadece doğru geometrik bağları korumak için RANSAC (Random Sample Consensus) algoritması (Fischler & Bolles, 1981) kullanılmaktadır. RANSAC, SIFT'in bulduğu eşleşme noktaları arasından rastgele alt kümeler seçerek iki görüntü arasında homografi matrisi kurmaya çalışır. Bu geometrik modele uymayan yalancı eşleşmeleri filtreleyerek, görüntüler arasındaki doğru eşleşmeleri korur.

Önceden eğitilmiş ResNet50 (Residual Networks) gibi modeller, pikselleri veya lokal köşeleri tek tek incelemek yerine görüntünün genel semantik yapısını anlayarak onu çok boyutlu tek bir özellik vektörüne (embedding) sıkıştırabilmektedir.

Literatürdeki bu iki farklı yöntemin (klasik bilgisayarlı görü ve derin öğrenme) birleştirilmesine dayanan hibrit sistemler bulunmaktadır (Hamzehei et al., 2023; Zhao et al., 2021). Bu çalışmada söz konusu yaklaşıma benzer olarak, uydu görüntülerinde benzerlik ve ortak alan tespiti problemlerine, derin öğrenme ve geleneksel bilgisayarlı görü yöntemlerini birleştiren hibrit bir çözüm getirilmektedir.

Metodoloji ve Sistem Mimarisi

Geliştirilen sistemin çalışma mekanizması üç temel aşamadan oluşmaktadır:

1. **Arama Uzayının Daraltılması (Kaba Filtreleme):** İnce ayar sürecine tabi tutulan derin öğrenme modelinin öznitelik çıkarıcı kabiliyetinden yararlanılarak, veri tabanındaki uydu görüntüleri için 2048 boyutlu öznitelik vektörleri üretilmektedir. Çalışma esnasında veri tabanından bir görüntü seçildiğinde bu seçilen görüntünün öznitelik vektörleri ve diğer görüntülerin vektörleri arasında kosinüs benzerlikleri hesaplanarak bu

görüntüye en benzer olmaya aday/en fazla ortak alan barındıran diğer 5 görüntü filtrelenir. Bu sayede, aynı kaynaktan gelen büyük verinin doğrusal ($1 \rightarrow N$) arama karmaşıklığı ortadan kaldırılır.

2. **Öznitelik Noktalarının Tespiti ve Geometrik Doğrulama (Hassas Arama):** Seçilen aday görüntüler üzerinde SIFT (Scale-Invariant Feature Transform) (Lowe, 2004) algoritması çalıştırılarak ölçek ve piksel değişimlerinden bağımsız anahtar noktalar saptanmaktadır. Ardından, RANSAC (Random Sample Consensus) (Fischler & Bolles, 1981) algoritması ile bu noktalar arasında bir Homografi Matrisi kurulmakta, konumsal ve geometrik tutarlılık doğrulanarak aykırı değerler (outliers) elenmektedir.
3. **Görselleştirme ve Arayüz Entegrasyonu:** Geometrik doğrulamadan başarıyla geçen ve kaynak görüntüyle en büyük ortak alanı barındıran optimum görüntü tespit edilerek, iki görüntü arasındaki ortak alanın sınırları hesaplanmakta; ortak bölgeler bir sınırlayıcı kutu (bounding box) ile işaretlenerek web arayüzü üzerinden kullanıcıya sunulmaktadır.

Geliştirilen web tabanlı uygulama kullanıcıya üç farklı opsiyon sunmaktadır:

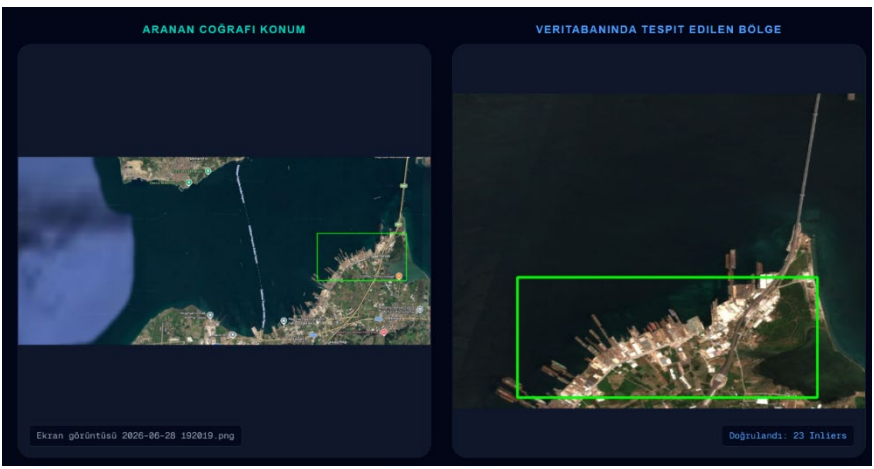
a. **Mod 1 (Hibrit Arama Modu):** Veri tabanından seçilen bir görüntünün, derin öğrenme ve SIFT&RANSAC filtrelerinden geçirilerek, bu görüntü ile en çok ortak alan barındıran başka bir görüntünün otomatik bulunması.

b. **Mod 2 (Dış Kaynaktan Gelen Harita Görüntüsünün Veri Tabanındaki Karşılığının Bulunması):** Google Haritalar (Google, 2026b) gibi harici platformlardan ekran görüntüsü (screenshot) olarak alınan (veri

tabanındaki uydu görüntülerinin kapsadığı sınırlar dahilinde olmak kaydıyla) bir girdinin sisteme dinamik olarak yüklenmesiyle, anlık öznitelik çıkarımı yapılarak veri tabanındaki en iyi eşleşen karşılığı arasında ortak alanlarının tespitinin yapılması ve ortak sınırların dinamik olarak işaretlenmesi.

c. **Mod 3 (Klasik SIFT&RANSAC Modu):** Kullanıcının veri tabanından manuel olarak seçeceği herhangi iki görüntü arasında derin öğrenme modeli ile kaba filtreleme aşaması devre dışı bırakılarak doğrudan SIFT&RANSAC algoritmaları ile ortak alan belirlemesi yapılması.

Resim 1: Google Haritalar'dan Yüklenen Bir Ekran Görüntüsünün Veri Tabanındaki Karşılığının Bulunduğu Bir Örnek



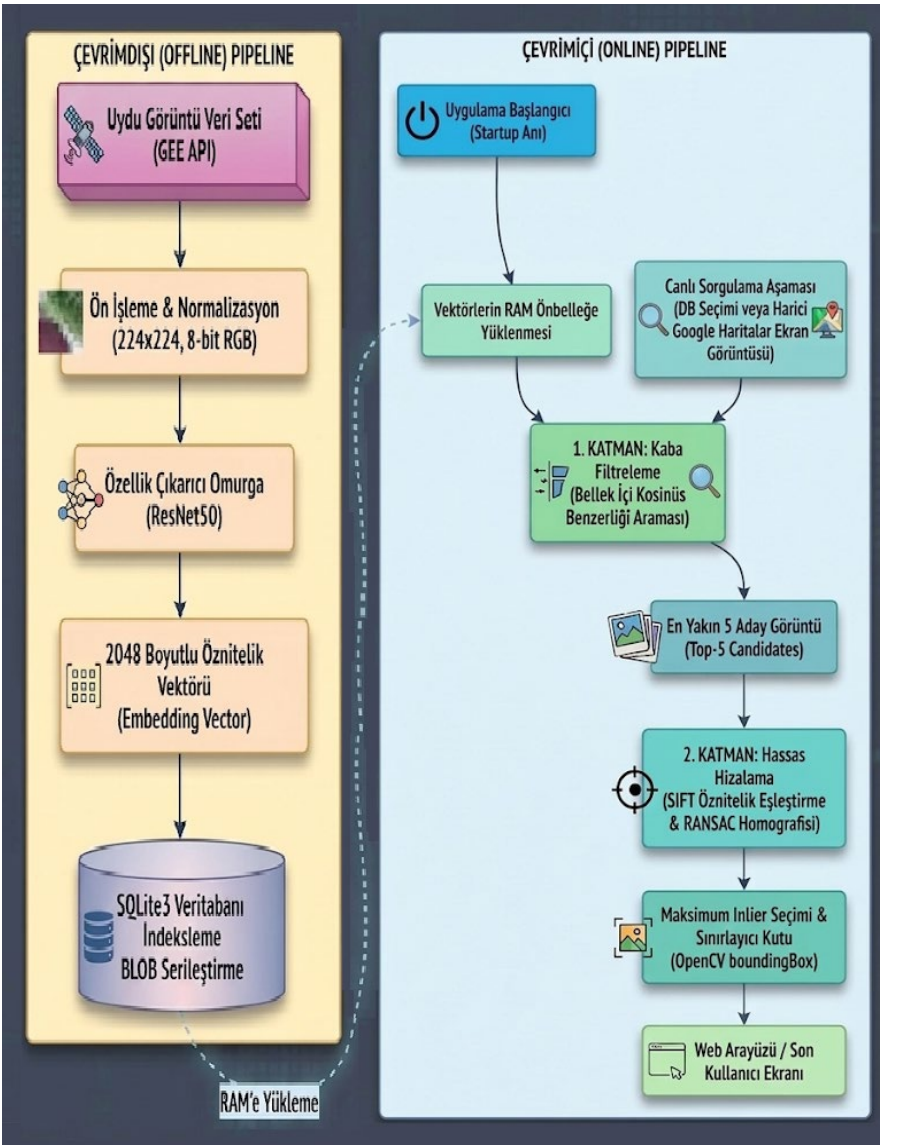
Kaynak: Geliştirilen uygulamadan alınan ekran görüntüsüdür.

Sistem Mimarisi ve Yöntem

4000 uydu görüntüsü üzerinde yerel öznitelik tabanlı veya derin öğrenme modelleriyle gerçek zamanlı çıkarım (real-time inference) yapmak yüksek hesaplama maliyeti doğurmaktadır. Bu kısıtı aşmak amacıyla; çevrimdışı öznitelik indeksleme ve çevrimiçi hiyerarşik

sorgulama olmak üzere iki aşamalı hibrit bir sistem mimarisi kurgulanmıştır.

Şekil 1: Sistem Akış Diyagramı



Kaynak: Google Gemini kullanılarak oluşturulmuştur.

Çevrimdışı Veri Hazırlığı ve İndeksleme Hattı

Google Earth Engine (GEE) üzerinden elde edilen 600×600 çözünürlüğündeki 16-bit GeoTIFF uydu görüntüleri, Min-Max normalizasyonu ile 8-bit biçimine dönüştürülmüş ve ilk üç bandı alınarak RGB renk uzayına aktarılmıştır. 224×224 boyutuna ölçeklenen görüntüler, ImageNet (Deng et al., 2009) dağılımına göre normalize edilerek derin ağ girdisine hazır hale getirilmiştir.

Ön işlemeden geçen görüntüler, ölçekten bağımsız öznitelik çıkarımı için ince ayar (fine-tuning) yapılmış ResNet50 mimarisine beslenmiştir. Ağın küresel ortalama havuzlama katmanından elde edilen 2048 boyutlu anlamsal öznitelik vektörleri (embeddings) çıkarılmıştır. Elde edilen vektörler ile sıkıştırılmış ham görüntüler, ikili veri (BLOB) formatında SQLite3 (Gaffney et al., 2022) veritabanına indekslenerek vektör tabanlı bir arama havuzu oluşturulmuştur.

Hibrit Benzerlik Sistemi

Çalışma zamanındaki arama performansı, kaba filtreleme ve hassas geometrik doğrulama adımlarından oluşan iki katmanlı bir mekanizmayla sağlanmaktadır. Ortak alan tespiti yapılan bölgeler OpenCV (Bradski, 2000) tabanlı sınırlayıcı kutu (bounding box) koordinatlarıyla çerçevelenir ve görsel sonuçlar kullanıcıya aktarılır.

Web Arayüzü Sistemi

- Arka Yüz (Backend): Algoritmik iş yükü, bellek içi vektör önbellekleme ve model çıkarımları Python (Python Software Foundation, 2026) tabanlı FastAPI (Lathkar, 2023) çatısı altında servis edilmiştir.
- Ön Yüz (Frontend): Kullanıcı deneyimi, analitik metriklerin takibi ve eşleşen alanların anlık görselleştirilmesi Next.js (Vercel, 2026), React (Meta Open Source, 2026), TypeScript (Microsoft, 2026) ve

TailwindCSS (Tailwind Labs, 2026) kullanılarak geliştirilen bir gösterge paneli üzerinden sağlanmaktadır.

Bulgular

4000 imajlık veri havuzunun çevrimdışı indekslenmesi (prepare_db.py), FastAPI backend servisinin ayağa kaldırılması (main.py) ve performans testlerinin gerçekleştirildiği yerel donanım altyapısı; Windows 11 Home Single Language 64-bit işletim sistemine, AMD Ryzen 7 5800H (~3.2GHz, 16 CPUs) işlemciye, 16 GB RAM (16384 MB) bellek bileşenine ve işlemciye entegre AMD Radeon(TM) Graphics grafik işlem birimine sahiptir.

Sistem performansını ölçmek adına, her üç operasyonel mod 10'ar örnekle test edilmiş ve her birinde derin öğrenme ve SIFT&RANSAC doğrulamasının çalışma zamanları hesaplanarak ortalamaları alınmış, bu ortalama süreler Tablo 1'de sunulmuştur.

Tablo 1: Operasyonel Uygulama Modlarının Ortalama Zaman Maliyeti

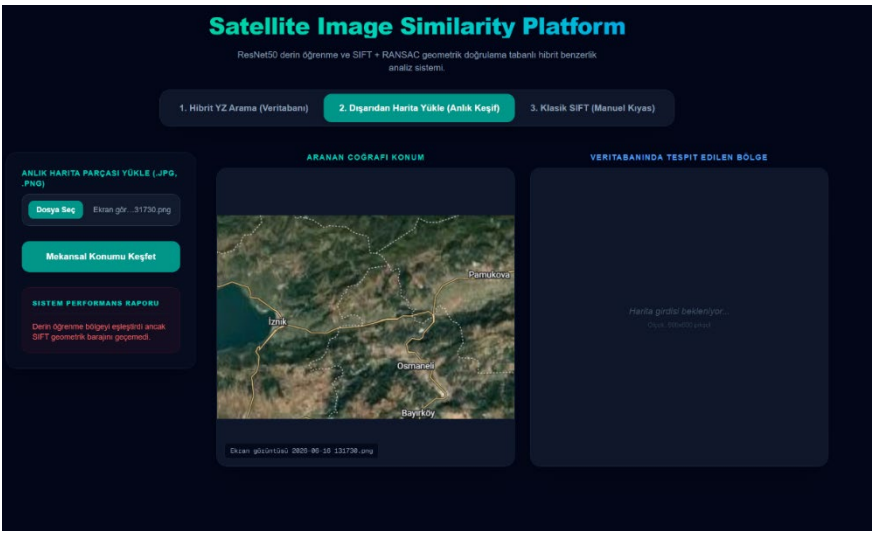
Operasyonel Mod Tanımı	**Çağrılan Backend Endpoint Yapısı**	**Kaba Filtreleme/Model İnferans Süresi**	**Hassas Arama (SIFT+RANSAC) Süresi**	**Toplam Yanıt Süresi**
Mod 1: Hibrit Yapay Zeka Modu	/hybrid_search	100.88 ms	1952.82 ms	2053.7 ms
Mod 2: Harici Harita Entegrasyonu	/search_by_uploaded_image	603.72 ms	2483.00 ms	3086.72 ms
Mod 3: Klasik SIFT Modu	/analyze	-	415.48 ms	415.48 ms

Kaynak: Yazarlar tarafından oluşturulmuştur

Tablo 1'de gösterilen bulgular, derin öğrenme tabanlı kaba filtreleme katmanının (Mod 1 ve Mod 2), sistemi veri havuzunun geneli ölçeğinde ne kadar büyük bir hesaplama yükünden ve zamansal darboğazdan kurtardığının en somut mühendislik ispatıdır. Mod 3 senaryosunda iki imaj arasındaki tekil kıyaslama süresi (415.48 ms) düşük görünse de, sistemin 4000 imajlık veri havuzunun tamamında derin öğrenme katmanı olmaksızın klasik SIFT algoritmasıyla doğrusal arama ($O(N)$ karmaşıklığında) yapmaya zorlanması durumunda toplam işlem süresi teorik olarak 4000×415.48 ms ≈ 1662 saniye (yaklaşık 27.7 dakika) seviyesine ulaşacaktır.

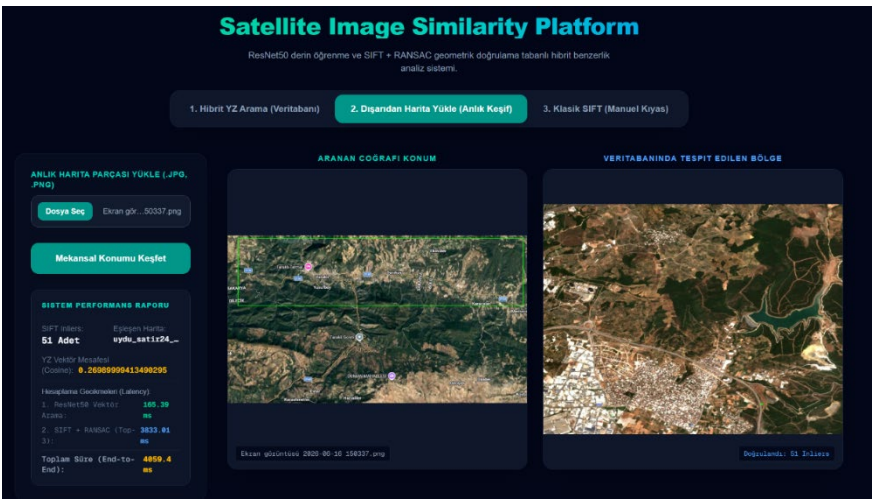
Mod 2'de yapılan 10 denemenin 1 adedinde sistem hassas hizalama katmanında çıktı üretememiş (Resim 2); 1 adedinde ise SIFT anahtar noktalarının uzamsal dağılım manipülasyonundan dolayı yanlış bölgeyi sınırlayıcı kutu içine alarak hatalı pozitif (false positive) sonuç üretmiştir (Resim 3). Bu durumların temel mühendislik gerekçesi; harici platform ekran görüntülerinin veritabanındaki uydu verilerine kıyasla farklı bir kırpma oranına, atmosferik renk skalasına ve çözünürlüğe sahip olmasıdır.

Resim 2: Mod 2'de Eşleşme Bulunamayan Başarısız Deneme Örneği



Kaynak: Geliştirilen uygulamadan alınan ekran görüntüsüdür.

Resim 3: Mod 2'de Hatalı Eşleşme Bulunan Bir Örnek



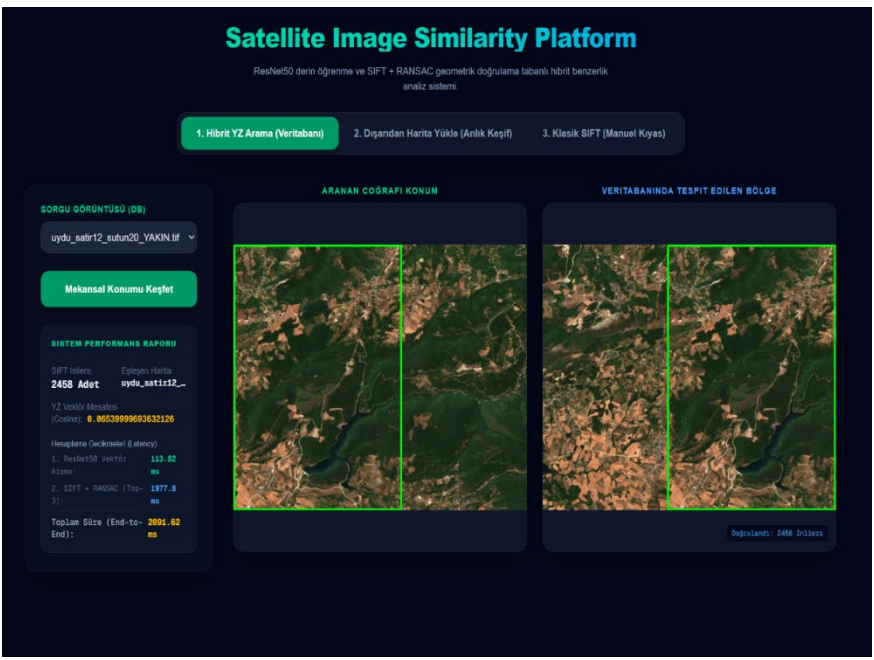
Kaynak: Geliştirilen uygulamadan alınan ekran görüntüsüdür.

Yapay zeka destekli ön filtreleme katmanı sayesinde, tüm arama uzayı Kosinüs benzerliğiyle milisaniyeler içinde elenerek hassas hizalama yükü yalnızca en olası 5 adaya indirgenmiş; böylece

kütüphane ölçeğinde zamansal maliyet bazında ~%99.8 oranında bir hesaplama tasarrufu elde edilmiştir. Bu mimari, geleneksel bilgisayarlı görü algoritmalarının büyük veri setlerinde doğrusal arama yaparken kilitlenmesi (computational bottleneck) problemini büyük oranda çözmektedir.

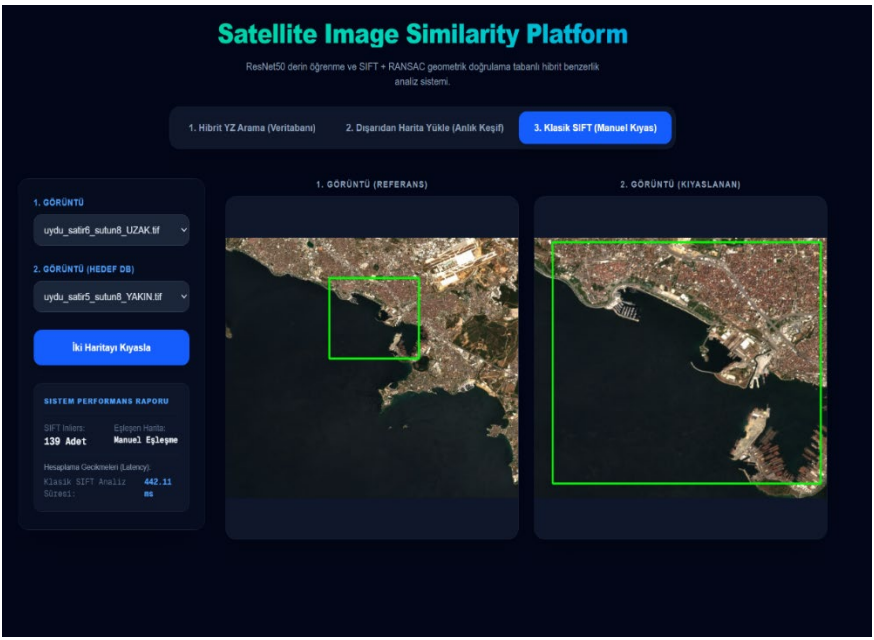
Geliştirilen web tabanlı uygulamanın canlı test senaryolarında ürettiği operasyonel başarım çıktıları ve sınırlayıcı kutu (bounding box) doğrulama örnekleri Resim 4, Resim 5 ve Resim 6'da sunulmuştur.

Resim 4: Hibrit Arama Moduna ait Başarılı Eşleşme Örneği



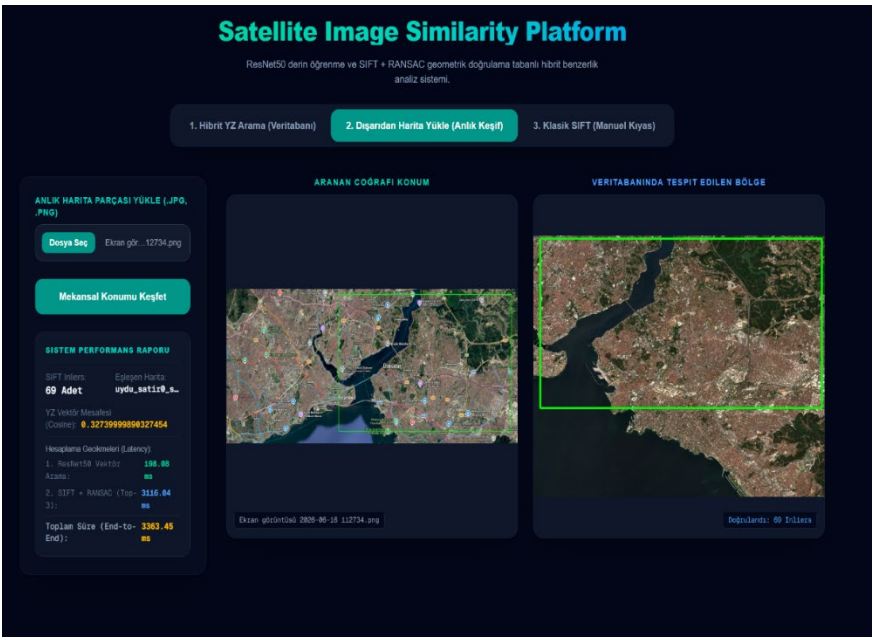
Kaynak: Geliştirilen uygulamadan alınan ekran görüntüsüdür.

Resim 5: Klasik SIFT&RANSAC Manuel Eşleştirme Modunda Elde Edilen Ölçekten Bağımsız Başarılı Eşleşme Örneği



Kaynak: Geliştirilen uygulamadan alınan ekran görüntüsüdür.

Resim 6: Dışarıdan Harita Yükleme (Anlık Keşif) Modunda Elde Edilen Başarılı Eşleşme Örneği



Kaynak: Geliştirilen uygulamadan alınan ekran görüntüsüdür.

Resim 4'te görüldüğü üzere; sistem, veri tabanından seçilen görüntünün en çok ortak alan barındıran benzerini 2458 adet geometrik doğrulama noktasıyla saptamış ve yeşil sınırlayıcı kutu içine almıştır. Resim 5'te sunulan SIFT&RANSAC Modu senaryosunda ise; iki farklı harita arasındaki ölçek (scale) farklılıklarına rağmen, algoritmanın özellik tabanlı ayırt edicilik kabiliyeti sayesinde ölçekten bağımsız geometrik benzerlikler 139 inlier nokta ile başarıyla yakalanmış ve doğru ortak bölge sınırlandırılmıştır. Resim 6'da ise Google Haritalar'dan anlık yüklenen ekran görüntüsünün, uydu görüntüsü veri tabanındaki benzer karşılığı 69 inlier kilit noktasıyla doğrulanarak ortak alan tespiti başarıyla tamamlanmıştır.

Sonuç

Bu çalışmada, uzaktan algılama görüntü arşivlerinde ortak alanların ve örtüşen uydu görüntülerinin hızlı ve yüksek doğrulukla

tespit edilmesini sağlayan, derin öğrenme ile klasik bilgisayarlı görü algoritmalarını birleştiren iki katmanlı bir benzerlik tespit sistemi geliştirilmiştir.

Google Earth Engine üzerinden temin edilen 4000 görüntülük uydu veri seti üzerinde yürütülen deneysel çalışmalarda, doğrusal arama yükü ortadan kaldırılmış ve hibrit benzerlik sorgulamaları yaklaşık 2053.7 milisaniye gibi makul bir sürede tamamlanmıştır.

Elde edilen bulgular, önerilen hibrit sistemin görüntü verileri üzerinde ölçek, çözünürlük ve arama karmaşıklığı problemlerine hızlı ve ölçeklenebilir bir mühendislik çözümü sunduğunu kanıtlamaktadır.

Dipnot: Çalışmada kullanılan uydu verileri, Google Earth Engine (Gorelick, et al. 2017) platformu aracılığıyla işlenmiş değiştirilmiş "Copernicus Sentinel [2025]" verilerini (European Space Agency, 2025) içermektedir.

Kaynakça

Bradski, G. (2000). The OpenCV library. *Dr. Dobb's Journal: Software Tools for the Professional Programmer*, *25*(11), 120–123.

Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition* (ss. 248–255). IEEE. https://doi.org/10.1109/CVPR.2009.5206848

European Space Agency. (2025). *Copernicus Sentinel-2 Surface Reflectance Harmonized (S2_SR_HARMONIZED)* [Data set]. Google Earth Engine. https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR_HARMONIZED

Fischler, M., & Bolles, R. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, *24*(6), 381–395. https://doi.org/10.1145/358669.358692

Gaffney, K., Prammer, M., Brasfield, L., Hipp, D., Kennedy, D., & Patel, J. (2022). SQLite: Past, present, and future. *Proceedings of the VLDB Endowment*, *15*(12), 3535–3547. https://doi.org/10.14778/3554821.3554842

Google. (2026a). *Google Colaboratory* [Bilgisayar yazılımı]. https://colab.research.google.com

Google. (2026b). *Google Maps imagery of Marmara region* [Map]. https://www.google.com/maps

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Hamzehei, S., Bai, J., Raimondi, G., Tripp, R., Ostroff, L., & Nabavi, S. (2023). 3D Biological/biomedical image registration with enhanced feature extraction and outlier detection. *Proceedings of the 14th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics* (ss. 1–10). ACM. https://dl.acm.org/doi/abs/10.1145/3584371.3612965

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (ss. 770–778). IEEE. https://doi.org/10.1109/CVPR.2016.90

Kurç, M., İstanbullu, A., & Haney, N. (2026, 5 Şubat). *Büyük ölçekli görüntü ve video veri kümelerinde çoklu geometrik benzerlik tespiti için algoritmik yaklaşım* [Bildiri sunumu]. Akademik Bilişim 2026, İstanbul, Türkiye, 99-101 https://inetd.org.tr/yayin/2026-ab24-bildiri-kitabi.pdf

Lathkar, M. (2023). *High-Performance Web Apps with FastAPI*. Apress. https://doi.org/10.1007/978-1-4842-9178-8

Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, *60*(2), 91–110. https://doi.org/10.1023/B:VISI.0000029664.99615.94

Meta Open Source. (2026). *React Documentation*. https://react.dev

Microsoft. (2026). *TypeScript Documentation*. https://www.typescriptlang.org/docs/

Python Software Foundation. (2026). *Python Language Reference*. https://www.python.org

Tailwind Labs. (2026). *Tailwind CSS Documentation*. https://tailwindcss.com/docs

Vercel. (2026). *Next.js Documentation*. https://nextjs.org/docs

Wang, Z., Bovik, A., Sheikh, H., & Simoncelli, E. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, *13*(4), 600–612. https://doi.org/10.1109/TIP.2003.819861

Zhao, X., Li, H., Wang, P., & Jing, L. (2021). An image registration method using deep residual network features for multisource high-resolution remote sensing images. *Remote Sensing*, *13*(17), Makale 3425. https://doi.org/10.3390/rs13173425

BÖLÜM 4

ÜRETKEN YAPAY ZEKÂ ÇAĞINDA KATALOGLAMA VE SINIFLANDIRMA: OLANAKLAR, SINIRLILIKLAR VE İNSAN DENETİMİ

HÜLYA AKYÜZ[1]

Giriş

Kütüphanelerin temel işlevlerinden biri, bilgi kaynaklarını yalnızca depolamak değil, kullanıcıların bu kaynakları tanımlayabilmesini, ayırt edebilmesini, bulabilmesini ve aralarındaki ilişkileri keşfedebilmesini sağlayacak biçimde düzenlemektir. Kataloglama, metadata (üstveri) oluşturma, konu analizi, indeksleme ve sınıflandırma bu düzenleme faaliyetinin birbirini tamamlayan bileşenleridir. Söz konusu süreçler, kaynakların bibliyografik özelliklerinin tanımlanmasını, içeriklerinin temsil edilmesini, benzer konuların bir araya getirilmesini ve kaynaklar arasında tutarlı erişim noktalarının oluşturulmasını sağlamaktadır. Bu nedenle kataloglama ve sınıflandırma, yalnızca kütüphanelerin teknik işleyişini destekleyen

[1] Dr. Öğretim Görevlisi (Tokat Gaziosmanpaşa Üniversitesi, Kütüphane ve Dokümantasyon Daire Başkanlığı, Tokat/Türkiye, Orcid: 0000-0001-5869-0815

işlemler değil, bilgiye erişimin niteliğini doğrudan etkileyen mesleki faaliyetlerdir.

Kataloglama ve sınıflandırma süreçleri, bilgi kaynaklarına ilişkin verilerin belirli alanlara aktarılmasının ötesinde mesleki bilgi ve bağlamsal değerlendirme gerektirmektedir. Bibliyografik kayıtların oluşturulmasında kaynakların doğru tanımlanması, kişi ve kurum adlarının tutarlı biçimde kullanılması, eserler arasındaki ilişkilerin belirlenmesi ve kataloglama standartlarına uygunluğun sağlanması önem taşımaktadır. Konu analizi ve sınıflandırmada ise bir eserin temel konusunun ikincil konularından ayrılması, kapsamının belirlenmesi, uygun kontrollü konu terimlerinin seçilmesi ve sınıflandırma sistemindeki hiyerarşik yerinin saptanması gerekmektedir. Dolayısıyla bu işlemlerin niteliği, kullanılan teknolojik araçların yanı sıra bilgi profesyonellerinin mesleki uzmanlığına ve değerlendirme süreçlerine de bağlıdır.

Geleneksel olarak kataloglama ve sınıflandırma süreçleri, kataloglama kuralları, sınıflandırma sistemleri, konu başlığı listeleri ve yetke dosyaları gibi standartlaştırılmış araçlar aracılığıyla yürütülmüştür. Bilgisayar teknolojilerinin kütüphanelerde kullanılmaya başlanmasıyla bibliyografik kayıtlar elektronik ortama aktarılmış; çevrim içi kamuya açık kataloglar (Online Public Access Catalog – OPAC), paylaşımcı kataloglama uygulamaları ve toplu katalog sistemleri bibliyografik verilere erişimi ve bu verilerin yeniden kullanılmasını kolaylaştırmıştır (Reitz, 2004, aktaran International Federation of Library Associations and Institutions [IFLA], t.y.; IFLA, 2026). Türkiye'de bu gelişmelerin önemli örneklerinden biri, TÜBİTAK Ulusal Akademik Ağ ve Bilgi Merkezi (ULAKBİM) tarafından sunulan Türkiye Ulusal Toplu Kataloğu'dur (TO-KAT). TO-KAT, farklı kütüphane ve kurumların koleksiyonlarında bulunan kitap, dergi ve diğer bilgi kaynaklarına ait bibliyografik kayıtların ortak bir arayüz üzerinden taranmasını ve sorgulanmasını sağlayan, bibliyografik veri paylaşımını destekleyen

ulusal bir toplu katalog sistemidir (TÜBİTAK Ulusal Akademik Ağ ve Bilgi Merkezi [ULAKBİM], 2006).

Bibliyografik verilerin elektronik ortamda oluşturulması, düzenlenmesi ve paylaşılmasına yönelik bu gelişmelerin ardından, üretken yapay zekâ ve büyük dil modelleri kataloglama ve sınıflandırma süreçlerinde yeni kullanım olanaklarını gündeme getirmiştir. Üretken yapay zekâyı önceki otomasyon uygulamalarından ayıran temel özelliklerden biri, doğal dilde sunulan girdilerden yeni metinler, özetler, açıklamalar ve öneriler oluşturabilmesidir. Büyük dil modelleri, geniş veri kümelerinden öğrendikleri dilsel ve anlamsal örüntüler doğrultusunda, verilen bağlama uygun olasılıksal çıktılar üretmektedir. Bu özellikler, bibliyografik verilerin oluşturulması ve düzenlenmesinde çeşitli görevlerin desteklenmesine olanak tanımaktadır. Bir eserin kapak, eser adı sayfası, içindekiler veya özet bilgilerinden bibliyografik öğelerin çıkarılması, eksik metadata alanlarına ilişkin önerilerin oluşturulması, konu terimlerinin belirlenmesi ve sınıflandırma numaralarının önerilmesi bu kapsamda değerlendirilebilir. Ex Libris tarafından Alma kataloglama ortamı için geliştirilen yapay zekâ destekli metadata yardımcısı, söz konusu teknolojilerin mevcut kütüphane iş akışlarına dâhil edilmesine yönelik uygulama örneklerinden biridir (Ex Libris, 2025). Bununla birlikte yapay zekâ tarafından üretilen çıktıların dilsel açıdan tutarlı görünmesi, bibliyografik ve anlamsal açıdan doğru oldukları anlamına gelmemektedir. Bu durum, kataloglama ve sınıflandırma süreçlerinde insan denetiminin önemini artırmaktadır.

Kütüphane ve bilgi bilimi literatüründe üretken yapay zekâya yönelik ilgi, büyük dil modellerinin yaygınlaşmasıyla birlikte artmıştır. Üretken yapay zekâ ile kataloglama ve metadata uygulamaları arasındaki ilişkiyi ele alan çalışmalar ile kurumsal arşivlerdeki doğuştan dijital koleksiyonların metadata yönetiminde bu teknolojilerin kullanımını inceleyen araştırmalar, üretken yapay

zekânın bibliyografik tanımlama ve metadata yönetimindeki kullanımına ilişkin farklı yaklaşımlar ortaya koymaktadır (Roncaglia, 2026; Tian, 2026). Ayrıca metadata çıkarımı, betimleyici kataloglama, konu analizi ve kalite kontrol işlevlerini bir araya getiren büyük dil modeli tabanlı kataloglama yaklaşımları geliştirilmiştir (Tan et al., 2026). Sınıflandırma alanında ise ChatGPT-4o, DeepSeek ve Gemini 2.0 modellerinin Dewey Onlu Sınıflama numarası atama performansları (Reshi et al., 2026) ile büyük dil modellerinin insanlar tarafından gerçekleştirilen konu indekslemesini değerlendirme kapasitesi araştırılmıştır (Włodarczyk & Zdunek, 2026). Bu çalışmalar, üretken yapay zekânın teknik hizmetlerdeki kullanımının yalnızca kuramsal düzeyde tartışılmadığını, somut uygulamalar ve deneysel değerlendirmeler üzerinden de incelendiğini göstermektedir.

Bununla birlikte üretken yapay zekânın kataloglama ve sınıflandırma süreçlerinde kullanımını inceleyen araştırmalar, bu teknolojilerin sunduğu olanakların yanı sıra çeşitli sınırlılıkları da gündeme getirmektedir. Bilgi profesyonellerinin deneyimlerine dayanan çalışmalarda, bu araçların iş süreçlerine sağlayabileceği katkıların yanı sıra kullanım sırasında karşılaşılan güçlükler de ele alınmıştır (Oladokun et al., 2025). Kontrollü konu sözlükleriyle gerçekleştirilen indeksleme araştırmalarında ise konu terimlerinin doğruluğu, bağlamsal yorumlama ve kontrollü sözlüklerle uyum açısından sorunlar belirlenmiştir (Golub et al., 2026). Bu bulgular, üretken yapay zekânın kataloglama ve sınıflandırma süreçlerindeki potansiyelinin yalnızca işlem gerçekleştirebilme kapasitesiyle değil, üretilen çıktıların niteliği ve güvenilirliğiyle birlikte değerlendirilmesi gerektiğine işaret etmektedir.

Türkiye'de yapay zekânın kütüphane hizmetlerindeki kullanımına ilişkin literatürde, bu teknolojilerin potansiyel uygulama alanlarını ve kütüphane hizmetlerine etkilerini ele alan derleme ve bütünleştirici incelemeler, üretken yapay zekânın bilgi

üretimi ve yönetimi üzerindeki etkilerini tartışan kavramsal çalışmalar ve üniversite kütüphanelerinde etik, sorumlu ve kurumsal politikalarla uyumlu kullanımı vurgulayan mesleki rehberler bulunmaktadır (Sivri, 2023; Demir, 2025; ANKOS Yapay Zekâ Araştırma Grubu, 2025; Kayalı, 2026). Söz konusu çalışmalar ve rehberler, Türkiye'deki bilgi ve belge yönetimi alanında yapay zekâ kullanımına ilişkin tartışmalar için bir çerçeve sunmakla birlikte, kataloglama ve sınıflandırma uygulamalarının yerel koşullar çerçevesinde değerlendirilmesi gerekliliğini de gündeme getirmektedir. Bu çalışmalardan önce, yapay zekânın kütüphanelerdeki potansiyel kullanım alanlarını inceleyen derleme niteliğindeki çalışmalar da yayımlanmıştır (Sivri, 2023).

Üretken yapay zekânın kataloglama ve sınıflandırma süreçlerinde kullanımı, yalnızca teknik bir otomasyon meselesi değildir. Bibliyografik verilerin doğruluğu, kataloglama standartlarına uygunluğu, konu temsilinin güvenilirliği, algoritmik önyargı, veri gizliliği ve mesleki sorumluluk gibi unsurlar, bu teknolojilerin kütüphane uygulamalarına nasıl dâhil edileceği konusunda belirleyici olmaktadır. Yapay zekâ sistemleri bibliyografik veri çıkarma, konu terimi önerme ve sınıflandırma numarası üretme gibi işlemlerde bilgi profesyonellerine destek sağlayabilse de bu çıktıların doğrulanması, standartlara uygunluğunun değerlendirilmesi ve gerekli düzeltmelerin yapılması insan denetimini gerektirmektedir. Bu çerçevede çalışmanın amacı, üretken yapay zekânın kataloglama ve sınıflandırma süreçlerindeki kullanım alanlarını, sunduğu olanakları ve karşılaşılan sınırlılıkları ilgili literatür ve uygulama örnekleri üzerinden incelemek; insan denetiminin gerekliliğini ve bilgi profesyonellerinin değişen rolünü değerlendirmektir.

Kataloglama ve Sınıflandırmada Üretken Yapay Zekâ

Üretken yapay zekâ, kataloglama ve sınıflandırma süreçlerinde özellikle yapılandırılmamış kaynaklardan veri çıkarma, bibliyografik kayıt taslakları oluşturma, konu terimleri önerme ve sınıflandırma kararlarını destekleme amacıyla kullanılmaktadır. Bu bölümde söz konusu kullanım alanları, teknik hizmetlerdeki dönüşümün devamı olarak ele alınmakta; bibliyografik tanımlama ile konu analizi ve sınıflandırmanın birbirinden farklı gereksinimleri üzerinde durulmaktadır.

Teknik Hizmetlerin Dönüşen Yapısı

Kataloglama ve sınıflandırmanın teknolojik dönüşümü üretken yapay zekâ ile başlamamıştır. MARC tabanlı makinece okunabilir kataloglama, çevrim içi ortak kataloglar, kopya kataloglama, yetke denetimi, otomatik veri doğrulama, bağlantılı veri ve varlık temelli bibliyografik modeller teknik hizmetlerin iş yapma biçimini uzun süredir değiştirmektedir. MARC tabanlı geleneksel bibliyografik kayıt yapılarından bağlantılı veri yaklaşımlarına uzanan dönüşüm, FRBR gibi kavramsal modellerin, RDA gibi kataloglama standartlarının ve BIBFRAME gibi bibliyografik veri modellerinin geliştirilmesiyle yeni bir boyut kazanmıştır (Roncaglia, 2026). Bu açıdan üretken yapay zekâ, kataloglama geleneğinden kopuş değil; otomasyon ve semantik zenginleştirme arayışının yeni bir aracıdır. Ancak bu dönüşüm yalnızca teknik bir mesele değildir; MARC'dan BIBFRAME'e geçiş süreci, organizasyonel kültür ve insan faktörlerini de içeren karmaşık bir süreçtir (Flinchbaugh et al., 2019). Değişime karşı direnç, öğrenme isteksizliği ve başarısızlık korkusu gibi psikososyal etmenler teknik geçiş süreçlerini zorlaştırabilmektedir.

Üretken yapay zekânın farklılığı, önceden açık biçimde programlanmamış görevlerde doğal dil üzerinden esnek çıktılar üretebilmesidir. Geleneksel kural tabanlı sistemler belirli alanların doğrulanması, veri biçimlerinin dönüştürülmesi veya tanımlı

sözlüklerle eşleştirme gibi görevlerde yüksek öngörülebilirlik sunarken, büyük dil modelleri belirsiz ve yapılandırılmamış içerik üzerinde çalışabilir. Büyük dil modellerinin ayırt edici yönü, doğal dilde verilen istemlerle eser adı sayfası, özet, içindekiler ve tam metin gibi farklı kaynaklardan alan önerileri ve açıklamalar oluşturabilmeleridir. Bu farklılık, kataloglamada hangi işlemlerin otomatikleştirileceği sorusunu görev bazında ele almayı gerektirir (Roncaglia, 2026; Tian, 2026).

Kullanım biçimleri, kaynaktan doğrudan okunabilen bilgilerin çıkarılmasından konu yorumuna dayanan önerilere kadar değişmektedir. Örneğin, eser adı ve yayın bilgisi için hazırlanan bir taslağın kontrolü, bir kaynağa uygun konu başlığının veya sınıflandırma numarasının belirlenmesine kıyasla daha düşük düzeyde uzman değerlendirmesi gerektirir. CNMARC bağlamında geliştirilen çok ajanlı kataloglama yaklaşımı, veri çıkarımı, betimleyici kataloglama, konu analizi ve kalite kontrolün ayrı görevler olarak düzenlenebileceğini göstermektedir (Tan et al., 2026). Bu nedenle üretken yapay zekânın katkısı, tüm kataloglama sürecinin tek adımda devredilmesinden çok belirli aşamaların desteklenmesinde aranmalıdır.

Bu değişim bilgi profesyonellerinin yetkinliklerini de etkilemektedir. Kataloglama ve metadata alanındaki güncel mesleki yetkinlikler, geleneksel standart bilgisinin yanında teknolojik araçları değerlendirme ve veri kalitesini gözetme becerilerini öne çıkarmaktadır (Amerikan Kütüphane Derneği Core [ALA Core], 2023). İnsan denetiminin kurumsal düzeyde nasıl düzenleneceği ise dördüncü bölümde ayrıca ele alınacaktır.

Bibliyografik Tanımlama ve Metadata Üretimi

Bibliyografik tanımlamada üretken yapay zekânın başlıca kullanım alanlarından biri, kaynağa ilişkin bilgileri belirleyerek bibliyografik kayıt oluşturma sürecine ilk taslağı sağlamasıdır. Eser

adı, sorumluluk bildirimi, yayıncı, yayın tarihi ve tanımlayıcılar gibi öğeler eser adı sayfası, kapak, dosya içeriği veya mevcut kayıtlardan çıkarılabilir. Bunun yanında kısa kayıtların zenginleştirilmesi, açıklama ve özet taslaklarının hazırlanması ya da eksik alanların belirlenmesi gibi işlemler de desteklenebilir (Ex Libris, 2025; Tian, 2026). Ancak kaynakta bulunan bilginin çıkarılması ile kaynağın içeriği hakkında yeni bir açıklama üretilmesi birbirinden farklı görevlerdir.

Kurumsal arşivlerde metadata yönetimini inceleyen bir çalışma, üretken yapay zekâ araçlarının başlık, özet, anahtar kelime ve tam metin bağlantısı gibi öğelerde farklı başarı düzeyleri gösterebildiğine dikkat çekmektedir. Sonuçlar, üretken yapay zekânın etkinliğinin görev türüne göre değiştiğini de ortaya koymaktadır. Modeller, sözdizimsel ve yapısal olarak tanımlanmış görevlerde iyi performans gösterirken, anlamsal olarak karmaşık bağlamlarda model performansı düşmektedir. Bağlamsal doğruluk ve kalite güvencesi için insan doğrulaması gerekli ve yapay zekânın profesyonel konu uzmanlığına yardımcı olabileceği ancak onun yerini alamayacağı vurgulanmaktadır (Tian, 2026).

Mevcut kütüphane sistemlerine entegre edilen uygulamalar bu işlevlerin somut örneklerini sunmaktadır. Alma Metadata Editor için geliştirilen AI Metadata Assistant, bibliyografik metadata önerilerini bilgi profesyonelinin incelemesine sunmakta; önerilerin kabul edilmesine, değiştirilmesine veya reddedilmesine olanak tanımaktadır (Ex Libris, 2025). Benzer biçimde Belçika Kraliyet Kütüphanesi bağlamındaki retrospektif kataloglama yaklaşımı, yapay zekânın birikmiş kataloglama işlerinin işlenmesi ve mevcut kayıtların geliştirilmesi amacıyla iş akışına eklenebileceğini tartışmaktadır (Lowagie, 2025). Bu örneklerde teknoloji, kayıtların niteliğine ilişkin mesleki kararı bütünüyle üstlenen bir sistemden çok kataloglama faaliyetini destekleyen bir yardımcı olarak konumlanmaktadır.

Üretken yapay zekâ destekli kayıt oluşturma, farklı diller ve kaynak türleri açısından da araştırılmaktadır. ChatGPT-4 ile gerçekleştirilen bibliyografik tanımlama çalışmasında standartlara uyum ve hata düzeltme olanakları incelenirken Sinhala ve Tamil dillerindeki materyallerde güçlükler bildirilmiştir (Gamage & Wanigasooriya, 2024). Arapça materyaller için MARC 21 kaydı oluşturulmasına odaklanan çalışma da dil ve yazı sistemi özelliklerinin kataloglama süreçlerinde ayrıca değerlendirilmesi gerektiğini göstermektedir (Alwreikat et al., 2026). Bu bulgular, çok dilli ortamlarda kayıt üretiminin yalnızca metin tanımaya indirgenemeyeceğine işaret etmektedir.

Bibliyografik tanımlamanın bir diğer boyutu, oluşturulan verilerin mevcut kataloglarla ve yetke dosyalarıyla uyumudur. Üretken yapay zekâ farklı ad biçimlerini veya alanlar arasındaki olası ilişkileri önerebilse de önerilen kişi, kurum ve eser erişim noktalarının geçerli yetke kayıtlarıyla karşılaştırılması gerekir. İnsan–büyük dil modeli iş birliğine dayalı metadata modelleme yaklaşımları, veri yapılarının kavramsal olarak düzenlenmesine yönelik yeni olanaklar sunmaktadır (Bagchi, 2025). Bununla birlikte bu tür önerilerin uygulanabilirliği, kullanılan kataloglama standardı ve kurumsal veri altyapısıyla birlikte değerlendirilmelidir.

Üretken yapay zekânın bibliyografik tanımlamadaki sınırlılıklarını aşmak için kullanılan yaklaşımlardan biri Retrieval-Augmented Generation (RAG) yöntemidir. RAG, büyük dil modellerinin güvenilir bilgi tabanlarıyla desteklenmesini sağlayarak halüsinasyon riskini azaltmakta ve üretilen çıktıların kaynaklara dayandırılmasını mümkün kılmaktadır. Bibliyografik tanımlamada RAG, yetke dosyaları, kontrollü sözlükler ve kurumsal kataloglar gibi güvenilir kaynaklarla entegre edilerek kullanılabilir. Bu yaklaşım, özellikle konu terimi önerisi ve sınıflandırma numarası atama gibi anlamsal yorum gerektiren görevlerde üretken yapay zekânın güvenilirliğini artırmaktadır (Lewis et al., 2020).

Sonuç olarak üretken yapay zekâ, bibliyografik tanımlamada ilk kayıt taslağını hazırlama ve mevcut metadatayı zenginleştirme gibi görevlerde kullanılabilir. Bu olanakların kalıcı bir iyileşmeye dönüşmesi, üretilen alanların kaynağa dayalı olarak doğrulanmasına ve kayıtların kullanılan standartlarla uyumlu olmasına bağlıdır. Bibliyografik doğruluk ve tutarlılık sorunları üçüncü bölümde ayrıntılı biçimde ele alınacaktır.

Konu Analizi, İndeksleme ve Sınıflandırma

Konu analizi, bilgi kaynağının temel konusunu belirlemeyi ve bu içeriği uygun terimlerle temsil etmeyi amaçlar. İndeksleme, belirlenen kavramların erişim terimlerine dönüştürülmesini; sınıflandırma ise kaynağın bir sınıflandırma sistemindeki kavramsal ve hiyerarşik konumunun belirlenmesini içerir. Bu süreçler, bibliyografik tanımlamadaki veri çıkarımından farklı olarak disipliner bağlamın ve konunun kapsamının yorumlanmasını gerektirir (Lancaster, 2003; Hjørland, 2018).

Almanya'daki Leibniz Ekonomi Bilgi Merkezi'nde (ZBW) yürütülen çalışmalar sonucunda geliştirilen AutoSE hizmeti, makine öğrenmesine dayalı otomatik konu indeksleme sistemlerinin kütüphanelerin günlük iş akışlarına entegre edilmesine örnek oluşturmaktadır. ZBW bünyesinde yürütülen otomatik konu indeksleme çalışmaları, bu tür sistemlerin kurumsal iş akışlarına dâhil edilmesinde model performansının yanı sıra metadata kalitesi ve uygulama süreçlerinin de önem taşıdığını göstermektedir (Kasprzik, 2024). Üretken yapay zekâya dayalı büyük dil modelleri ise bilgi kaynaklarının içeriklerinden hareketle aday konu terimleri oluşturmak amacıyla kullanılabilmektedir. Bununla birlikte üretilen terimlerin anlamsal açıdan uygun görünmesi, kontrollü sözlüklerin tanımlarına ve indeksleme politikalarına uygun oldukları anlamına gelmemektedir. Nitekim İsveç'teki tarihsel LGBTQ+ edebiyat eserleri üzerinde gerçekleştirilen bir araştırmada, büyük dil

modellerinin daha özgül kavramlar yerine genel kavramları tercih edebildiği, kontrollü sözlük tanımlarını ve indeksleme politikalarını her zaman izleyemediği ve tarihsel bağlamı yorumlamada güçlük yaşayabildiği belirlenmiştir (Golub et al., 2026).

Bu çalışmalar, otomatik konu indeksleme sistemlerinin değerlendirilmesinde yalnızca teknik performansın değil, üretilen konu terimlerinin bibliyografik standartlara ve kurumsal indeksleme politikalarına uygunluğunun da dikkate alınması gerektiğini ortaya koymaktadır. Büyük dil modelleri tarafından önerilen konu terimlerinin kontrollü sözlüklerdeki karşılıklarının, kavramsal özgüllük düzeylerinin ve kaynakların konusal kapsamını temsil etme yeterliliklerinin uzmanlar tarafından değerlendirilmesi önem taşımaktadır. Dolayısıyla üretken yapay zekâ, konu indeksleme sürecinde bağımsız bir karar mekanizmasından ziyade, bilgi profesyonelinin değerlendirmesine sunulan aday terimlerin oluşturulmasını destekleyen bir araç olarak ele alınmalıdır.

Büyük dil modelleri yalnızca konu terimi üretmek için değil, mevcut indeksleme kararlarını değerlendirmek için de incelenmektedir. Polonya'da kütüphane ve bilgi bilimi alanında yayımlanan 40 makaleye atanmış konu tanımlayıcıları üzerinden yürütülen araştırmada, modellerin farklı değerlendirme ölçütlerinde orta ile yüksek düzey arasında değişen sonuçlar verdiği; buna karşılık yanlış atanmış tanımlayıcıları belirleme konusunda sınırlılıkların sürdüğü bildirilmiştir (Włodarczyk & Zdunek, 2026). Bu çalışma, otomatik değerlendirme ile uzman değerlendirmesinin aynı şey olmadığını göstermesi bakımından önemlidir.

Sınıflandırma numarası önerme görevi, konu analizine ek olarak seçilen sistemin notasyon yapısını ve uygulama talimatlarını dikkate almayı gerektirir. Dewey Onlu Sınıflama (DOS) gibi hiyerarşik sistemlerde geniş konu alanını belirlemek ile ayrıntılı alt sınıfı doğru seçmek farklı zorluk düzeylerindedir. ChatGPT-4o, DeepSeek ve Gemini 2.0 modellerinin Dewey numarası atama

performansını inceleyen çalışma, genel amaçlı büyük dil modellerinin bu alandaki olanaklarını ve sınırlılıklarını deneysel olarak değerlendirmiştir (Reshi et al., 2026).

Üretken yapay zekânın konu analizi ve sınıflandırmadaki katkısı, aday konu terimleri ve sınıflandırma numaraları önererek bilgi profesyonelinin değerlendireceği seçenekleri görünür kılmasında ortaya çıkabilir. Özellikle kontrollü sözlüklerle eşleştirme, notasyon kurallarına uygunluk ve yerel terminoloji gibi konularda önerilerin ayrıca sınanması gerekir. Bu sorunların güvenilirlik, önyargı ve insan denetimi boyutları izleyen bölümlerde ele alınacaktır.

Üretken Yapay Zekâ Çıktılarının Güvenilirliği ve Sınırlılıkları

Üretken yapay zekânın kataloglama ve sınıflandırma süreçlerinde kullanımı, üretilen çıktıların doğruluğu, güvenilirliği ve ilgili standartlarla uyumu açısından çeşitli sorunları gündeme getirmektedir. Bibliyografik bilgilerin hatalı veya gerçeğe aykırı biçimde üretilmesi, kontrollü sözlüklerle uyumsuz konu terimlerinin önerilmesi, kaynak içeriğinin yanlış yorumlanması ve algoritmik önyargılar bu teknolojilerin kullanımında dikkate alınması gereken başlıca sınırlılıklar arasındadır. Bu bölümde söz konusu sorunlar bibliyografik doğruluk, standartlara uyum, semantik yorum, önyargı, temsil ve açıklanabilirlik boyutlarıyla ilgili literatür çerçevesinde ele alınmaktadır.

Bibliyografik doğruluk ve uydurulmuş bilgi

Üretken yapay zekânın bibliyografik kayıt oluşturma süreçlerinde kullanılması, üretilen bilgilerin güvenilirliği ve doğrulanabilirliği açısından çeşitli sorunları gündeme getirmektedir. Büyük dil modelleri, gerçekte bulunmayan kaynaklara ilişkin bibliyografik bilgiler üretebildikleri gibi mevcut yayınların yazar, eser adı, yayın yılı, yayıncı ve tanımlayıcı bilgilerini de hatalı sunabilmektedir (Walters & Wilder, 2023). Literatürde halüsinasyon

olarak nitelendirilen bu durum, üretilen bilgilerin dilsel ve biçimsel açıdan tutarlı görünmesine rağmen gerçek kaynaklarla doğrulanamaması nedeniyle önem taşımaktadır. Bibliyografik doğruluk açısından temel sorun, oluşturulan bilgilerin belirli bir kayıt biçimine uygun görünmesinden çok, ilgili kaynaklarla karşılaştırıldığında gerçeği ne ölçüde yansıttığıdır.

Büyük dil modellerinin bibliyografik referans üretimindeki doğruluğu çeşitli deneysel araştırmalarda incelenmiştir. ChatGPT-3.5 ve GPT-4 tarafından 42 farklı konuda oluşturulan toplam 636 bibliyografik referansın değerlendirildiği bir araştırmada, GPT-3.5 tarafından üretilen referansların %55'inin, GPT-4 tarafından üretilenlerin ise %18'inin gerçekte bulunmayan kaynaklara ait olduğu belirlenmiştir. Ayrıca gerçek kaynaklara ilişkin referanslarda da bibliyografik hatalar tespit edilmiştir (Walters & Wilder, 2023). Bu bulgular, daha gelişmiş modellerin kullanılmasının uydurma referans sorununu tamamen ortadan kaldırmadığını göstermektedir.

Benzer şekilde, ChatGPT tarafından doğa bilimleri ve beşerî bilimler alanlarında oluşturulan referansların karşılaştırıldığı bir araştırmada, kaynakların varlığı ile bibliyografik bilgilerinin doğruluğunun birbirinden ayrı değerlendirilmesi gerektiği ortaya konmuştur. Araştırmada, gerçekte mevcut yayınlara ilişkin referanslarda dahi dijital nesne tanımlayıcılarının (DOI) hatalı üretilebildiği belirlenmiştir (Mugaanyi et al., 2024). Anatomi alanında üç büyük dil modelinin ürettiği 1.800 referansın incelendiği başka bir araştırmada ise modeller arasında halüsinasyon oranlarının farklılaştığı ve bibliyografik bileşenlerin doğruluğunun değişkenlik gösterdiği bildirilmiştir (Ülkir & Paslı, 2026). Bu çalışmalar, bibliyografik doğruluğun yalnızca bir yayının varlığının kontrol edilmesiyle sınırlı olmadığını; yazar, eser adı, yayın yılı ve tanımlayıcı gibi bilgilerin de ayrı ayrı doğrulanmasını gerektirdiğini göstermektedir.

Kütüphanecilik literatürüne odaklanan bir araştırmada, ChatGPT-5 tarafından oluşturulan 200 bibliyografik girdi incelenmiştir. Araştırmada girdilerin %74'ü tamamen doğru bulunurken %20,5'inde DOI hataları ve %4'ünde birden fazla metadata hatası belirlenmiştir (Oladokun et al., 2026). Bu bulgular, bibliyografik bilgilerin önemli bir bölümünün doğru üretilebilmesine karşın özellikle tanımlayıcılar ve ayrıntılı yayın bilgilerinde doğrulama gereksiniminin devam ettiğini göstermektedir. Dolayısıyla bir bibliyografik girdinin biçimsel olarak eksiksiz görünmesi, içerdiği bilgilerin güvenilir olduğunu tek başına ortaya koymamaktadır.

Üretken yapay zekânın bibliyografik doğruluğuna ilişkin sorunlar, yalnızca akademik referans üretimiyle sınırlı değildir. Kütüphane kataloglama uygulamalarını ele alan araştırmalarda, büyük dil modellerinin bibliyografik kayıt oluşturma, mevcut kayıtları düzenleme ve kataloglama süreçlerini destekleme potansiyeli de incelenmektedir. ChatGPT-4'ün MARC 21, AACR2 ve RDA standartları çerçevesinde değerlendirildiği on kullanım senaryosunda, mevcut bibliyografik kayıtlardaki hataların düzeltilmesi ve kataloglama eğitimine destek sağlanması bakımından olumlu sonuçlar elde edilmiştir. Ayrıca CATMELK adlı özel GPT uygulamasıyla kitapların çeşitli sayfalarından alınan görseller kullanılarak RDA uyumlu MARC 21 kayıtları oluşturulabilmiş; ancak Singhalaca ve Tamilce bibliyografik bilgilerin işlenmesinde sınırlılıklar gözlenmiştir (Gamage & Wanigasooriya, 2024). Arapça materyallere yönelik bir başka araştırmada ise 13 kitap için ChatGPT tarafından oluşturulan MARC 21 kayıtları, profesyonel bilgi profesyonellerinin hazırladığı kayıtlarla karşılaştırılmıştır. Yalnızca eser adı verildiğinde yeterli bibliyografik kayıt oluşturulamamış; yazar, yayıncı, yayın yeri ve yayın tarihi gibi bilgiler sağlandığında temel MARC 21 alanlarını içeren kayıtlar üretilebilmiştir. Bununla birlikte, bazı alanlarda yer

tutucu ifadelerin kullanıldığı ve özellikle kodlanmış alanlarda eksikliklerin bulunduğu belirlenmiştir (Alwreikat et al., 2026). Her iki araştırmanın bulguları, üretken yapay zekânın bibliyografik kayıt oluşturma ve mevcut kayıtları düzenleme süreçlerinde destekleyici bir araç olarak kullanılabileceğini; ancak çıktıların doğruluğunun sağlanan bibliyografik bilgilerin yeterliliğine, materyallerin diline ve kataloglama standartlarına uygunluğuna bağlı olarak değişebildiğini ortaya koymaktadır.

Uydurulmuş bibliyografik bilgilerin dikkat çekici özelliklerinden biri, gerçek yayınlara ait unsurların bir araya getirilmesiyle inandırıcı ancak doğrulanamayan referansların oluşturulabilmesidir. Üretken yapay zekâ tarafından oluşturulan uydurma akademik referansların özellikleri ve benzer referansların yeniden üretilebilmesi sorunu literatürde ele alınmıştır (Spennemann, 2026). Bu tür referanslar, yazar, eser adı ve yayın bilgileri bakımından gerçek kaynaklara benzeyebildikleri için doğrulama yapılmadığında fark edilmeleri güçleşebilmektedir. Bununla birlikte, akademik referans üretiminde gözlenen bu sorunların doğrudan kütüphane kataloglama sistemlerinde aynı sıklıkta ortaya çıktığı varsayılamaz.

Bibliyografik doğrulukla ilgili sorunlar, uydurulmuş referansların oluşturulmasıyla sınırlı olmayıp bibliyografik verilerin standartlara uygun biçimde üretilmesi, dönüştürülmesi ve yeniden kullanılması süreçlerini de kapsamaktadır. Üretken yapay zekânın olasılıksal çalışma yapısı ile bibliyografik verilerin standartlaştırılmış ve doğrulanabilir olma gerekliliği arasındaki farklılık, bu teknolojilerin kütüphane uygulamalarına uyarlanmasında dikkate alınması gereken bir sorun olarak değerlendirilmektedir (Roncaglia, 2026). RDA öğelerinin Dublin Core terimleri ve MARC 21 alanlarıyla eşleştirilmesini inceleyen bir araştırmada, veri standartları arasındaki yapısal farklılıklardan kaynaklanan tutarsızlıklar belirlenmiştir (Taniguchi, 2025). Üretken

yapay zekâ modellerinin bibliyografik metadata oluşturma performansını karşılaştıran bir başka araştırmada ise modellerin geçersiz konu başlıkları ve URI'ler üretebildiği, kataloglama kurallarını tutarlı biçimde uygulayamadığı görülmüştür (Heng et al., 2026). Bu çalışmalar birlikte değerlendirildiğinde, bibliyografik doğruluğun korunmasında yapay zekâ tarafından üretilen bilgilerin doğrulanmasının yanı sıra veri standartlarının yapısal özelliklerinin ve farklı sistemler arasındaki uyumun da dikkate alınması gerektiği anlaşılmaktadır.

Kütüphane kataloglarında bibliyografik kayıtların paylaşılması ve yeniden kullanılması, kataloglama süreçlerinde verimlilik sağlamakla birlikte kayıtların doğruluğunun korunmasını da önemli hâle getirmektedir. WorldCat kapsamında bibliyografik verilerin yerel kütüphane kataloglarına aktarılabilmesi ve farklı kurumlar tarafından yeniden kullanılabilmesi, paylaşımcı kataloglamanın kurumsal düzeydeki işleyişini ortaya koymaktadır (OCLC, 2010). Öte yandan, üretken yapay zekâ tarafından oluşturulan bibliyografik referanslarda uydurulmuş veya hatalı bilgilerin bulunabildiği deneysel araştırmalarla gösterilmiştir (Walters & Wilder, 2023). Bu iki durum birlikte değerlendirildiğinde, üretken yapay zekâ tarafından oluşturulan hatalı bilgilerin doğrulanmadan katalog kayıtlarına aktarılması hâlinde, söz konusu yanlışlıkların kayıt paylaşımı ve yeniden kullanım süreçlerinde başka sistemlere taşınması olası bir risk olarak değerlendirilebilir. Ancak bibliyografik referans üretiminde belirlenen hatalar, bu hataların kütüphane katalogları arasında fiilen yayıldığını doğrudan kanıtlamamaktadır. Bu nedenle söz konusu risk, gerçekleştiği deneysel olarak gösterilmiş bir sonuçtan ziyade, bibliyografik veri paylaşımının işleyişi ve üretken yapay zekânın bilinen doğruluk sorunları temelinde yapılan bir değerlendirme olarak ele alınmalıdır.

Sonuç olarak üretken yapay zekâ destekli bibliyografik tanımlamada doğruluk, yalnızca oluşturulan kaydın biçimsel bütünlüğüyle değil, içerdiği bilgilerin güvenilir kaynaklarla doğrulanabilirliğiyle birlikte değerlendirilmelidir. Akademik referans üretimi ve kataloglama uygulamalarına ilişkin araştırmalar, büyük dil modellerinin bibliyografik bilgi üretiminde yararlı olabildiğini, ancak yanlış veya uydurulmuş bilgilerin ortaya çıkma olasılığının devam ettiğini göstermektedir. Bu nedenle bibliyografik doğruluğun korunması, kayıt oluşturma aşamasından başlayarak bilgilerin kontrol edilmesi, aktarılması ve yeniden kullanılması süreçlerine kadar uzanan bir değerlendirmeyi gerektirmektedir.

Standartlar, kontrollü sözlükler ve tutarlılık

Kataloglama standartları, bibliyografik verilerin tutarlı biçimde tanımlanması, düzenlenmesi ve farklı bilgi sistemleri arasında kullanılabilmesi açısından temel bir işlev üstlenmektedir. MARC 21, bibliyografik verilerin makine tarafından işlenebilir bir yapıda kodlanmasını sağlarken RDA, kaynakların tanımlanması ve aralarındaki ilişkilerin belirlenmesine yönelik kurallar sunmaktadır. Üretken yapay zekâ modellerinin bibliyografik kayıt oluşturabilmesi, bu standartların bütün kurallarını ve kurumlara özgü uygulama kararlarını doğru biçimde yerine getirebildikleri anlamına gelmemektedir. Büyük dil modellerinin MARC 21, AACR2 ve RDA çerçevesinde bibliyografik kayıt oluşturma, metadata çıkarma ve mevcut kayıtlardaki hataları düzeltme kapasiteleri uygulamalı araştırmalarda değerlendirilmiştir (Gamage & Wanigasooriya, 2024). Arapça materyaller için ChatGPT aracılığıyla oluşturulan MARC 21 kayıtlarının insan bilgi profesyonelleri tarafından hazırlanan kayıtlarla karşılaştırılması da standartlara uygunluğun ayrıca incelenmesi gereken bir değerlendirme ölçütü olduğunu ortaya koymaktadır (Alwreikat et al., 2026). Bu çerçevede alan ve alt alan seçimi, göstergelerin kullanımı, kodlanmış bilgiler ve ilişki belirteçleri gibi yapısal ayrıntıların doğruluğu, bibliyografik kaydın

yalnızca biçimsel olarak eksiksiz görünmesinden bağımsız olarak değerlendirilmelidir.

Standartlara uygunluk sorunu, bibliyografik verilerin farklı metadata yapıları arasında dönüştürülmesinde de önem kazanmaktadır. RDA öğelerinin Dublin Core terimleri ve MARC 21 alanlarıyla eşleştirilmesini inceleyen bir araştırmada, eşleştirmelerdeki karmaşıklığın azaltılması amacıyla öğeler arasındaki hiyerarşik ve kapsama ilişkilerinden yararlanılması değerlendirilmiştir (Taniguchi, 2025). Bu çalışma doğrudan üretken yapay zekâ kullanımını incelememekle birlikte, bibliyografik metadata dönüşümünün yalnızca alan adlarının karşılıklı olarak eşleştirilmesinden ibaret olmadığını göstermektedir. Dolayısıyla üretken yapay zekâ destekli metadata oluşturma ve dönüştürme süreçlerinde, kaynak ve hedef standartların yapısal özelliklerinin dikkate alınması ve oluşturulan eşleştirmelerin doğrulanması gerekmektedir.

Üretken yapay zekânın kataloglama standartlarıyla birlikte kullanılmasına yönelik uygulamalardan biri, Alma kütüphane yönetim sistemi içerisinde geliştirilen AI Metadata Assistant'tır. Bu uygulama, bibliyografik kayıtların oluşturulması veya mevcut kayıtların zenginleştirilmesi amacıyla metadata önerileri sunmakta; bilgi profesyonelinin önerileri incelemesine, kabul etmesine, düzeltmesine veya reddetmesine olanak tanımaktadır. Ayrıca yöneticiler tarafından belirli metadata taleplerinin sınırlandırılması ve bilgi profesyonelinin konu başlıkları, özet veya içindekiler bilgisi gibi belirli metadata türlerine yönelik öneri istemesi desteklenmektedir (Ex Libris, 2025). Sistemin MARC 21 kayıtları ve Library of Congress konu başlıklarıyla çalışması, üretken yapay zekânın kataloglama süreçlerinde tamamen bağımsız bir karar mekanizması olarak değil, mevcut standartlar ve bilgi profesyoneli denetimiyle birlikte kullanılmasına yönelik bir uygulama örneği sunmaktadır. Bununla birlikte, uygulamanın belirli standartları ve

kontrollü sözlükleri desteklemesi, oluşturulan bütün önerilerin doğruluğunu kendiliğinden güvence altına almamaktadır.

Kontrollü sözlükler ve yetke denetimi, üretken yapay zekâ destekli konu analizinde ayrıca değerlendirilmesi gereken alanlardır. Serbest metin üretimine dayanan bir model, kaynağın içeriğiyle anlamsal bakımdan ilişkili bir konu terimi önerebilse de bu terimin kullanılan kontrollü sözlükte yetkili biçimiyle yer alması, doğru konu başlığına karşılık gelmesi ve gerekli alt bölümlerle uygun biçimde birleştirilmesi ayrı bir doğrulama gerektirmektedir. Örneğin Library of Congress konu başlıklarında yetkili terimlerin seçimi ile coğrafi, dönemsel ve biçimsel alt bölümlerin kullanımı, yalnızca anlamsal benzerlik üzerinden değerlendirilemez. Üretken yapay zekâ destekli metadata uygulamalarında konu sözlüklerinin kapsamının kontrollü biçimde genişletilmesinin planlanması da bu alanın ayrı bir değerlendirme ve kalite kontrolü gerektirdiğine işaret etmektedir (Ex Libris, 2025). Bu nedenle konu başlığı önerilerinin doğruluğu, yalnızca kaynağın konusunu yansıtma düzeyiyle değil, ilgili kontrollü sözlüğün yetkili terimleri ve kullanım kurallarıyla uyumu bakımından da incelenmelidir.

Sınıflandırma numarasının kapsamı, notasyon derinliği, tablo kullanımı ve sentez kuralları, önerilen numaranın doğruluğunu etkileyen unsurlardır. Bu nedenle üretken yapay zekâ tarafından önerilen sınıflandırma numaralarının yalnızca konu bakımından uygunluğu değil, kullanılan sınıflandırma sisteminin yapısal kurallarıyla uyumu da değerlendirilmelidir. Büyük dil modellerinin Dewey sınıflandırma numarası atama performansının araştırmalara konu olması, bu önerilerin doğruluğunun ayrıca değerlendirilmesinin önemini göstermektedir (Reshi et al., 2026). Burada belirtilen olası hata türleri, incelenen araştırmaların tamamında ayrı ayrı deneysel olarak gösterilmiş bulgular şeklinde değil, sınıflandırma sistemlerinin kural temelli yapısından hareketle belirlenen denetim gereksinimleri olarak ele alınmalıdır.

Tutarlılık, bibliyografik kayıtların niteliği açısından standartlara uygunluk kadar önemli bir başka boyuttur. Aynı tür kaynakların benzer kataloglama kararlarıyla temsil edilmesi, katalog içerisinde düzenli ve öngörülebilir bir bibliyografik yapı oluşturulmasına katkı sağlamaktadır. Kütüphaneler, yerel kataloglama politikaları, yetke dosyaları ve ortak uygulama kararları aracılığıyla bilgi profesyonelinin arasında ortaya çıkabilecek yorum farklılıklarını azaltmaya çalışmaktadır. Bununla birlikte, büyük dil modellerinin çıktıları aynı girdiler karşısında dahi farklılık gösterebilmektedir. Nitekim AI Metadata Assistant'ın resmî kullanım belgelerinde, aynı kaynak için yinelenen işlemlerde farklı metadata önerileri alınabileceği ve model çıktılarının zaman içerisinde değişebileceği belirtilmektedir (Ex Libris, t.y.). Bu özellik, üretken yapay zekâ destekli kataloglama süreçlerinde çıktıların yalnızca tek tek doğrulanmasını değil, koleksiyon genelinde tutarlı biçimde kullanılmasının da değerlendirilmesini gerektirmektedir.

Aynı kaynağa farklı işlemlerde birbirinden farklı konu başlıkları veya sınıflandırma numaraları önerilmesi, bu önerilerin denetlenmeden kullanılması durumunda katalog içi tutarlılığı olumsuz etkileyebilecek olası bir risk oluşturmaktadır (Ex Libris, t.y.). Bu nedenle yapay zekâ destekli kataloglama iş akışlarında kullanılan modelin sürümünün, istemlerin, başvurulan veri kaynaklarının ve bilgi profesyoneli tarafından yapılan düzeltmelerin mümkün olduğunca izlenebilir olması önem taşımaktadır. Üretken yapay zekâdan elde edilen önerilerin kurumsal kataloglama politikaları, kontrollü sözlükler ve sınıflandırma kurallarıyla karşılaştırılması, bibliyografik verilerin doğruluğunun ve tutarlılığının korunmasına yönelik temel bir denetim yaklaşımı olarak değerlendirilebilir.

Semantik yorumun sınırları

Üretken yapay zekânın bibliyografik metadata oluşturma süreçlerindeki başarısı, gerçekleştirilen görevin niteliğine göre farklılık gösterebilmektedir. Dijital koleksiyonların metadata yönetimini inceleyen bir araştırmada, yapılandırılmış metadata oluşturma görevleri ile anlamsal anahtar kelime belirleme görevlerinin farklı performans düzeyleri sergilediği bildirilmiştir (Tian, 2026). Bu ayrım, kaynaktan doğrudan elde edilebilen bibliyografik bilgiler ile kaynağın içeriğinin yorumlanmasını gerektiren konu verilerinin aynı otomasyon yaklaşımıyla değerlendirilmemesi gerektiğine işaret etmektedir. Yazar, eser adı veya yayın tarihi gibi açıkça tanımlanabilen bilgiler uygun kaynaklar üzerinden doğrulanabilirken, konu analizi eserin amacı, kapsamı, disiplini ve ele alınış biçiminin değerlendirilmesini gerektirmektedir. Dolayısıyla üretken yapay zekânın bibliyografik bilgileri doğru biçimde çıkarabilmesi, eserin konusunu aynı doğruluk düzeyinde temsil edebileceği anlamına gelmemektedir.

Semantik (anlambilim), sözcük dizilimlerinin cümleye kattığı anlam, cümle yapılarının anlaşılması ve cümle anlamının ortaya çıkarılmasıdır (Adalı, 2020, s. 273). Kütüphanelerde konu analizi açısından semantik çözümleme, bir eserde kullanılan sözcüklerin yalnızca sözlük anlamlarıyla değil, bağlam içinde taşıdıkları anlamlarla birlikte değerlendirilmesini gerektirmektedir. Bu süreçte karşılaşılan temel güçlüklerden biri, metinde kullanılan kavramlarla eserin asıl konusu arasındaki ilişkinin her zaman doğrudan olmamasıdır. Bir kavramın eser adında bulunması, o kavramın eserin temel inceleme konusu olduğunu tek başına göstermeyebilir. Benzer şekilde, bir eserin konusu eser adında açıkça belirtilmemiş ya da farklı kavramlarla ifade edilmiş olabilir. Bu nedenle konu analizi, sözcüklerin belirlenmesinin ötesinde kavramların bağlam içerisinde yorumlanmasını, eserin hangi disiplin çerçevesinde ele alındığının belirlenmesini ve konusal kapsamının değerlendirilmesini gerektirmektedir. Büyük dil modellerinin

Dewey Onlu Sınıflama Sistemi kapsamında sınıflandırma numarası atama performansının karşılaştırmalı olarak araştırılması, üretken yapay zekâ ile gerçekleştirilen sınıflandırmanın ayrıca değerlendirilmesi gereken bir uygulama alanı olduğunu göstermektedir (Reshi et al., 2026).

Örneğin "çocuklarda kalp hastalıkları" biçimindeki bir eser adı, ilk değerlendirmede pediatri ve kardiyoloji alanlarıyla ilişkilendirilebilir. Ancak eserin doğumsal kalp hastalıklarına, cerrahi tedavi yöntemlerine, yoğun bakım süreçlerine veya belirli bir hastalık grubuna odaklanması, konu analizi ve sınıflandırma kararlarını etkileyebilir. Bu durumda yalnızca eser adından hareketle yapılan bir değerlendirme, kaynağın ayrıntılı konusal kapsamını belirlemek için yeterli olmayabilir. İçindekiler, özet, önsöz ve gerektiğinde eserin ilgili bölümlerinin incelenmesi, konunun daha doğru belirlenmesine katkı sağlayabilir. Benzer şekilde, Dewey sınıflandırma numarası yalnızca eser adındaki bir kavramın karşılığı olarak değerlendirilmemelidir. Kaynağın konusal odağı, ele alınış biçimi ve sınıflandırma sistemine ilişkin talimatlar birlikte dikkate alınmalıdır. Dolayısıyla üretken yapay zekâ tarafından önerilen bir numaranın anlamsal bakımdan uygun görünmesi, bu numaranın kaynak incelemesi ve sınıflandırma kurallarıyla doğrulanması gereğini ortadan kaldırmamaktadır.

Semantik yorumun bir diğer boyutu, aynı kaynağın farklı kavramsal çerçeveler içerisinde temsil edilebilmesidir. Bir eser birden fazla disiplinle ilişkili olabileceği gibi, içerdiği kavramlar arasında temel konu ile ikincil konuların ayrılması da gerekebilir. Üretken yapay zekâ tarafından önerilen konu terimlerinin anlam bakımından ilgili olması, bunların kaynağın temel konusunu yeterli düzeyde temsil ettiğini tek başına göstermemektedir. Nitekim büyük dil modellerinin insan tarafından gerçekleştirilen konu indeksleme işlemlerini değerlendirmek amacıyla kullanılması da araştırmalara konu olmuştur (Włodarczyk & Zdunek, 2026). Bununla birlikte, bu

tür çalışmaların varlığı, modellerin her koşulda bilgi profesyonelleri ile eşdeğer konu analizi gerçekleştirebildiği şeklinde yorumlanmamalıdır. Konu temsilinin niteliği, önerilen terimlerin anlamsal uygunluğunun yanı sıra kaynağın kapsamını, temel konusunu ve disiplin bağlamını ne ölçüde yansıttığı üzerinden değerlendirilmelidir.

Bu sınırlılıklar, üretken yapay zekânın konu analizi ve sınıflandırma süreçlerinde kullanılmasının yararsız olduğu anlamına gelmemektedir. Büyük dil modelleri, olası konu terimlerinin belirlenmesi, uzun içindekiler listelerinin özetlenmesi, kavramlar arasındaki ilişkilerin karşılaştırılması ve sınıflandırma seçeneklerinin bilgi profesyoneline sunulması gibi görevlerde destekleyici araçlar olarak değerlendirilebilir. Ancak bu kullanım alanları, önerilerin doğruluğunun ve uygulamadaki başarısının ayrıca sınanmasını gerektirmektedir. Özellikle kaynağın konusal kapsamının belirsiz olduğu, birden fazla disiplinle ilişkili bulunduğu veya ayrıntılı sınıflandırma kararlarının verilmesini gerektirdiği durumlarda insan değerlendirmesinin önemi artmaktadır. Bu nedenle üretken yapay zekâ destekli konu analizi ve sınıflandırma süreçlerinde otomasyon düzeyinin, görevin doğrulanabilirliği, konusal belirsizlik ve olası hata sonuçları dikkate alınarak belirlenmesi uygun bir yaklaşım olarak değerlendirilebilir.

Önyargı, temsil ve açıklanabilirlik

Kataloglama ve sınıflandırma, yalnızca bibliyografik bilgilerin teknik olarak düzenlenmesinden ibaret değildir. Bilgi kaynaklarının hangi kavramlarla temsil edildiği, hangi konu başlıkları altında bir araya getirildiği ve hangi sınıflandırma konumlarına yerleştirildiği, kullanıcıların bu kaynaklara erişimini ve kaynaklar arasındaki ilişkileri keşfetmesini etkileyebilmektedir. Bu nedenle üretken yapay zekânın kataloglama süreçlerinde kullanılması, teknik verimlilik ve doğruluğun yanı sıra mesleki etik

ilkeler açısından da değerlendirilmelidir. Kataloglama Etiği Kuralları çerçevesinde yapay zekâ kullanımını inceleyen bir araştırmada, kataloglama alanındaki uygulama çalışmalarının veri doğruluğu, gizlilik ve önyargı gibi etik sorunları her zaman sistematik biçimde ele almadığı belirtilmekte; sorumlu yapay zekâ kullanımına yönelik mesleki yaklaşımların geliştirilmesi de tartışılmaktadır (Dover & Grzegorski, 2025). Benzer şekilde kataloglama ve metadata profesyonellerinin yapay zekâ uygulamalarına ilişkin etik çerçeveleri kurumsal iş akışlarına uyarlaması ve doğru, kültürel açıdan duyarlı metadata üretme sorumluluğunu sürdürmesi gerektiği vurgulanmaktadır (Sussmeier & Henry, 2025).

Üretken yapay zekâ modellerinin eğitim verilerindeki tarihsel, kültürel ve dilsel örüntüleri yeniden üretme potansiyeli, bu sistemlerin bibliyografik konu temsilinde oluşturabileceği risklerin başında gelmektedir. Yapay zekâ risklerini ele alan NIST (National Institute of Standards and Technology) çerçevesinde önyargı; sistematik, hesaplamalı/istatistiksel ve insan-bilişsel olmak üzere üç kategoride incelenmektedir. Sistematik önyargılar kurumsal uygulamalar ve toplumsal yapılardan, hesaplamalı ve istatistiksel önyargılar veri dağılımlarından ve algoritmik süreçlerden, insan-bilişsel önyargılar ise bireylerin yapay zekâ sistemlerini algılama ve kullanma biçimlerinden kaynaklanabilmektedir. Ayrıca yapay zekâ sistemlerinin mevcut önyargıları daha geniş ölçekte yeniden üretme ve olumsuz etkilerini artırma riski bulunduğu belirtilmektedir (National Institute of Standards and Technology [NIST], 2023). Kataloglama bağlamında bu riskler, baskın dil ve kültürlere ait terimlerin öncelenmesi, yerel kavramların yeterince temsil edilmemesi veya belirli topluluklara ilişkin kaynakların genelleyici konu başlıklarıyla tanımlanması şeklinde ortaya çıkabilir. Amerikan Kütüphane Derneği (ALA) tarafından yayımlanan rehberde de yapay zekâ destekli keşif ve teknik hizmet uygulamalarının belirli

konuların, yazarların, görüşlerin, dillerin ve toplulukların görünürlüğünü azaltabilecek örüntüler bakımından düzenli olarak incelenmesi önerilmektedir (ALA, 2026).

Önyargının belirlenebilmesi için tekil hata örneklerinin ötesine geçen sistematik değerlendirme yöntemlerine ihtiyaç vardır. Bu kapsamda farklı dillerde, disiplinlerde ve kültürel bağlamlarda hazırlanmış kaynaklardan oluşan bir değerlendirme kümesi oluşturulabilir. Üretken yapay zekâ tarafından önerilen konu başlıkları ve sınıflandırma numaraları, uzman bilgi profesyonelinin doğruladığı kayıtlar ve kullanılan kontrollü sözlüklerle karşılaştırılabilir. Değerlendirme sırasında doğru konu terimi önerme oranı, yetkisiz terim kullanımı, konusal özgüllük düzeyi ve sınıflandırma hataları gibi göstergeler incelenebilir. Bu göstergelerin kaynakların dili, konusu ve ilgili kültürel bağlamlara göre ayrı ayrı değerlendirilmesi, belirli gruplarda yoğunlaşan hata örüntülerinin belirlenmesine yardımcı olabilir. NIST çerçevesinin MEASURE 2.11 maddesinde de yapay zekâ sistemlerindeki adalet ve önyargı risklerinin değerlendirilmesi ve sonuçların belgelenmesi öngörülmektedir (NIST, 2023).

Bununla birlikte, her hatalı konu başlığı veya sınıflandırma numarası doğrudan önyargı göstergesi olarak değerlendirilmemelidir. Önyargı değerlendirmesinde belirli kaynak gruplarının diğerlerine kıyasla sistematik biçimde daha düşük doğrulukla temsil edilip edilmediği ve bu farklılığın hangi koşullarda ortaya çıktığı incelenmelidir. Örneğin, konu kapsamı bakımından Türkçe ve İngilizce kaynaklara yönelik konu başlığı önerileri, doğruluk ve ayrıntı düzeyi açısından karşılaştırılabilir. Benzer şekilde, yerel kültüre ilişkin kavramların daha genel terimlerle değiştirilip değiştirilmediği veya belirli konuların sistematik biçimde daha geniş sınıflara yönlendirilip yönlendirilmediği değerlendirilebilir. Ancak bu karşılaştırmalarda kaynakların konu güçlüğü, kullanılan metadatanın kapsamı ve değerlendirme

kümesinin dağılımı dikkate alınmalıdır. Ayrıca gruplar arasında benzer doğruluk değerlerinin elde edilmesi, bütün kullanıcılar açısından adil ve erişilebilir bir sistem oluşturulduğunu tek başına göstermemektedir. NIST çerçevesinde de adaletin yalnızca istatistiksel dengeden ibaret olmadığı; erişilebilirlik ve mevcut toplumsal eşitsizliklerin etkilerinin de dikkate alınması gerektiği belirtilmektedir (NIST, 2023).

Önyargıların değerlendirilmesinde yalnızca teknik ölçütlere dayanılması yeterli olmayabilir. Kataloglama kararlarının kaynaklarda temsil edilen topluluklar üzerindeki etkilerinin anlaşılabilmesi için bilgi profesyonelinin, alan uzmanlarının ve gerektiğinde ilgili kullanıcı topluluklarının görüşlerinden yararlanılması önem taşımaktadır. NIST çerçevesinin GOVERN 5.1 maddesi, yapay zekâ sistemlerinin geliştirilmesi veya uygulanmasından sorumlu ekiplerin dışındaki paydaşlardan olası bireysel ve toplumsal etkiler hakkında geri bildirim toplanmasına, bu geri bildirimlerin değerlendirilmesine ve uygun görülenlerin süreçlere dâhil edilmesine yönelik kurumsal uygulamaların geliştirilmesini öngörmektedir (NIST, 2023). Bu yaklaşım kataloglama süreçlerine uyarlandığında, kültürel açıdan duyarlı metadata oluşturulması ve sorunlu konu temsillerinin belirlenmesi bakımından katılımcı bir değerlendirme mekanizması kurulmasına katkı sağlayabilir. Kataloglama ve metadata profesyonellerinin etik yapay zekâ uygulamalarında sahip oldukları uzmanlığın kurumsal politikalara yansıtılması da bu açıdan önem taşımaktadır (Sussmeier & Henry, 2025).

Açıklanabilirlik, özellikle konu analizi ve sınıflandırma kararlarının denetlenmesi açısından önem taşımaktadır. NIST çerçevesinde şeffaflık, açıklanabilirlik ve yorumlanabilirlik birbirini destekleyen ancak farklı işlevleri bulunan özellikler olarak ele alınmaktadır. Şeffaflık, sistemin işleyişine ve kullanımına ilişkin bilgilerin erişilebilirliğiyle; açıklanabilirlik, sistemin nasıl

çalıştığının anlaşılmasıyla; yorumlanabilirlik ise üretilen çıktının kullanım bağlamındaki anlamının değerlendirilmesiyle ilişkilidir (NIST, 2023). Kataloglama uygulamalarında bu ayrım, yalnızca yapay zekâ tarafından hangi konu başlığının veya sınıflandırma numarasının önerildiğinin bilinmesiyle yetinilmemesi gerektiğini göstermektedir. Önerinin hangi bibliyografik verilere dayandığı, ilgili kontrollü sözlük veya sınıflandırma kurallarıyla nasıl ilişkilendirildiği ve bilgi profesyoneli tarafından hangi gerekçeyle kabul edildiği ya da değiştirildiği de izlenebilir olmalıdır.

Bir bilgi profesyoneli belirli bir konu başlığını veya sınıflandırma numarasını seçme gerekçesini eserin temel konusu, kullanılan kontrollü sözlüğün kuralları veya sınıflandırma şemasındaki kapsam notları ve talimatlar üzerinden açıklayabilir. Büyük dil modelleri, uygun görünen bir öneri üretse bile, bu önerinin hangi doğrulanabilir kaynaklara ve kurallara dayandığını her zaman açıkça gösteremeyebilir. Modelden karara ilişkin açıklama istenmesi, bilgi profesyonelinin öneriyi değerlendirmesine yardımcı olabilir; ancak üretilen açıklamanın da model çıktısı olduğu ve gerçek karar sürecini doğru biçimde yansıttığının varsayılamayacağı dikkate alınmalıdır. Bu nedenle açıklanabilirlik, yalnızca modelin gerekçe üretmesiyle sınırlandırılmamalı; sunulan gerekçelerin yetke kayıtları, kontrollü sözlükler ve sınıflandırma talimatları üzerinden doğrulanabilmesiyle birlikte değerlendirilmelidir. Yapay zekâ sistemlerinin etik açıdan değerlendirilmesinde mesleki standartların ve kataloglama uzmanlığının dikkate alınması, bu denetim yaklaşımıyla uyumludur (Dover & Grzegorski, 2025).

Türkçe ve diğer İngilizce dışı diller açısından önyargı ve temsil sorunlarının ayrıca değerlendirilmesi gerekmektedir. Büyük dil modellerinin çok dilli metinleri işleyebilmesi, farklı dillerde aynı düzeyde bibliyografik konu analizi gerçekleştirebildikleri anlamına gelmemektedir. Türkiye'deki kütüphaneler açısından yalnızca Türkçe eser adlarının anlaşılması değil, Türkçe konu

terminolojisinin, yerel yetke uygulamalarının ve kullanılan sınıflandırma kurallarının doğru biçimde uygulanması da önem taşımaktadır. Bu nedenle Türkçe kaynaklar üzerinde yapılacak uygulamalı değerlendirmelerde konu terimlerinin doğruluğu, terminolojik uygunluk, konusal özgüllük, yerel kavramların korunması ve sınıflandırma tutarlılığı gibi ölçütlerin birlikte incelenmesi yararlı olacaktır. Bu tür çalışmalar, üretken yapay zekânın Türkçe bibliyografik veriler üzerindeki kullanım olanaklarının ve sınırlılıklarının ortaya konulmasına katkı sağlayabilir.

İnsan Denetimi ve Bilgi Profesyonelinin Rolü

Üretken yapay zekânın kataloglama ve sınıflandırma süreçlerinde sunduğu olanaklar, bibliyografik doğruluk, standartlara uygunluk, semantik yorum ve etik sorumluluk açısından insan denetiminin önemini ortadan kaldırmamaktadır. Önceki bölümde ele alınan sınırlılıklar, yapay zekâ destekli uygulamalarda yalnızca teknik performansın değil, üretilen çıktıların mesleki ilkelere ve kurumsal kataloglama politikalarına uygunluğunun da değerlendirilmesi gerektiğini göstermektedir. Bu bağlamda bilgi profesyonelinin rolü, bibliyografik kayıtları doğrudan oluşturan kişi olmanın ötesine geçerek yapay zekâ tarafından sunulan önerileri değerlendiren, doğrulayan ve gerektiğinde düzelten bir uzmanlık alanını da kapsamaktadır. Bu bölümde insan-döngüde yaklaşımı, mesleki sorumluluk ve yapay zekâ destekli kataloglama süreçlerinde bilgi profesyonellerinin üstlenebileceği görevler ele alınmaktadır.

İnsan-döngüde yaklaşımı

Yapay zekâ sistemlerinde kararların tamamen otomatik olarak alınması, insan uzmana bırakılması veya yapay zekâ önerilerinin insan kararlarını desteklemek amacıyla kullanılması gibi farklı insan–yapay zekâ etkileşim biçimleri bulunmaktadır. Üretken yapay zekânın teknik hizmetlerde sorumlu kullanımı

açısından öne çıkan yaklaşımlardan biri, human-in-the-loop (HITL) olarak adlandırılan insan-döngüde yaklaşımıdır. Bu yaklaşımda yapay zekâ sistemi belirli işlemleri gerçekleştirirken veya öneriler üretirken insan uzman iş akışının belirlenen aşamalarında çıktıları değerlendirme ve kararları denetleme görevini üstlenmektedir (NIST, 2023). Kataloglama ve sınıflandırma süreçlerinde insan-döngüde yaklaşımının temel amacı, insan müdahalesini otomasyonun önünde bir engel olarak görmek değil, yapay zekânın bilgi işleme ve içerik üretme kapasitesi ile bilgi profesyonelinin bağlamsal değerlendirme yetkinliğini, kataloglama standartlarına hâkimiyetini ve mesleki sorumluluğunu bir araya getirmektir. Böylece insan denetimi, yalnızca hataların giderilmesine yönelik sonradan gerçekleştirilen bir işlem değil, yapay zekâ destekli iş akışının tasarımına dâhil edilen bir unsur olarak değerlendirilmektedir.

ALA'nın yapay zekâ rehberinde, kataloglama, sınıflandırma ve sağlama süreçlerinde yapay zekâ kullanımının insan gözetimi ve denetimi içerecek biçimde yapılandırılması; çalışanların sistem tarafından oluşturulan önerileri kabul etme, düzenleme veya reddetme yetkisini koruması gerektiği belirtilmektedir. Rehberde yapay zekânın taslak bibliyografik betimleme, aday konu terimleri oluşturma, metadata zenginleştirme ve başlangıç düzeyinde sınıflandırma önerileri sunma gibi destekleyici görevlerde kullanılabileceği ifade edilmektedir. Özellikle konu başlıkları, sınıflandırma, yetke çalışmaları, İngilizce dışındaki materyaller ve hassas konularla ilgili işlemlerde uzman değerlendirmesinin sürdürülmesi gerektiği vurgulanmaktadır (ALA, 2026).

ALA (2026) rehberinde vurgulanan insan gözetimi ve denetimi ilkeleri doğrultusunda, yapay zekâ destekli kataloglama süreçlerinde insan katılımı farklı biçimlerde yapılandırılabilir. İnsan onayına dayalı uygulamalarda sistem tarafından oluşturulan bibliyografik kayıt veya metadata önerisi, bilgi profesyonelinin

kabul ya da ret kararına sunulmaktadır. İnsan değerlendirmesine dayalı uygulamalarda ise öneriler yalnızca onaylanmakla sınırlı kalmamakta; doğruluk, standartlara uygunluk ve konusal temsil gibi ölçütler üzerinden incelenerek gerektiğinde düzeltilmektedir. İnsan–yapay zekâ iş birliğine dayalı uygulamalarda bilgi profesyoneli, işlemin farklı aşamalarında modele yönlendirme yapmakta, ek bibliyografik bilgiler sağlamakta, alternatif öneriler istemekte ve nihai kararı vermektedir. Burada açıklanan uygulama biçimleri, birbirini dışlayan kesin kategoriler olarak değil, kataloglama iş akışlarında insan katılımının farklı işlevlerini açıklamak amacıyla kullanılan bir değerlendirme çerçevesi olarak ele alınmalıdır. Örneğin bir bilgi profesyoneli, aynı kayıt üzerinde yapay zekâyla etkileşimli biçimde çalışabilir, önerilen konu başlıklarını değerlendirebilir ve son aşamada kaydı onaylayabilir.

İnsan-döngüde yaklaşımı, üretken yapay zekânın teknik hizmetleri tamamen bağımsız biçimde yürütebileceği varsayımı ile hata olasılığı nedeniyle bu teknolojinin hiçbir yarar sağlamayacağı düşüncesi arasında, görevin niteliğine dayanan bir uygulama olanağı sunmaktadır. Bu kapsamda otomasyon ve insan denetiminin düzeyi, işlemin doğrulanabilirliği, karmaşıklığı ve olası hata sonuçlarına göre farklılaştırılabilir. Kaynaktan açıkça elde edilebilen bibliyografik bilgilerin çıkarılması daha yüksek otomasyon düzeyiyle yürütülebilirken konu analizi ve sınıflandırma önerilerinin daha yoğun uzman incelemesi gerektirmesi mümkündür. Yetke denetimi, kültürel açıdan hassas konu temsili ve karmaşık sınıflandırma kararlarında ise insanın yönlendirici rolünün korunması önem taşımaktadır. Bu görev ayrımı, yapay zekâ sistemlerinin kullanım bağlamına göre değerlendirilmesi ve risklerin yönetilmesi yaklaşımıyla uyumludur (NIST, 2023).

İnsan denetiminin uygulamadaki örneklerinden biri, Alma kütüphane yönetim sistemi içerisinde kullanılan AI Metadata Assistant'tır. Bu uygulama, bilgi profesyoneline bibliyografik kayıt

oluşturma veya mevcut kayıtları zenginleştirme amacıyla metadata önerileri sunmaktadır. Bilgi profesyoneli, önerileri inceleyebilmekte, kabul edebilmekte, düzenleyebilmekte veya reddedebilmektedir. Ayrıca belirli metadata türleri için öneri talep edilmesi ve yöneticiler tarafından kullanılabilecek özelliklerin sınırlandırılması desteklenmektedir (Ex Libris, 2025; Ex Libris, t.y.). Böylece üretken yapay zekâ, bilgi profesyonelinin kullandığı metadata düzenleme ortamına yardımcı bir araç olarak dâhil edilmekte ve önerilerin değerlendirilmesi mevcut kataloglama iş akışı içinde gerçekleştirilebilmektedir. Bununla birlikte, uygulamanın insan onayına olanak tanıması, her önerinin yeterli düzeyde incelendiğini tek başına göstermemektedir.

İnsan denetiminin yalnızca işlemin sonunda gerçekleştirilen kısa bir kontrol olarak anlaşılması yeterli değildir. Etkili bir insan-döngüde uygulamasında uzmanların hangi görevlerin otomatikleştirileceğine, hangi kaynak verilerin kullanılacağına, hangi metadata alanları için öneri oluşturulacağına ve hangi durumlarda zorunlu inceleme yapılacağına ilişkin kararlara katılması önemlidir. Bu açıdan insan müdahalesi işlem öncesinde sistemin kullanım koşullarının belirlenmesini, işlem sırasında önerilerin değerlendirilmesini ve işlem sonrasında kayıtların doğrulanmasını kapsayabilir. Yapay zekâ sistemlerine ilişkin sorumlulukların tanımlanması, risklerin değerlendirilmesi ve süreçlerin belgelenmesi güvenilir yapay zekâ yönetişiminin temel bileşenleri arasında yer almaktadır (NIST, 2023).

Bununla birlikte, iş akışında bir insanın bulunması, denetimin her durumda etkili biçimde gerçekleşeceğini garanti etmemektedir. Otomatik sistemlerin önerilerine aşırı güvenilmesi, kullanıcıların bu önerileri yeterince sorgulamadan kabul etmesine veya sistemden uyarı gelmediğinde gerekli kontrolleri yapmamasına neden olabilmektedir. Otomasyon yanlılığı olarak adlandırılan bu durum, ihmal ve icra hataları üzerinden açıklanmaktadır. İhmal

hatası, otomatik sistemin bir sorunu bildirmemesi nedeniyle kullanıcının gerekli müdahaleyi gerçekleştirmemesi; icra hatası ise sistemin yanlış önerisinin, diğer güvenilir bilgilerle çelişmesine rağmen izlenmesidir (Skitka et al., 1999). Otomatik izleme yardımcısının kullanıldığı deneysel bir araştırmada, otomasyon desteği alan katılımcıların destek almayan katılımcılara kıyasla daha fazla ihmal hatası yaptığı belirlenmiştir. Otomasyon kullanılan koşulda ilgili olayların doğru tespit edilme oranı %59 iken otomasyon kullanılmayan koşulda bu oran %97 olarak bulunmuştur. Ayrıca katılımcıların, otomatik sistemin hatalı yönlendirmelerini diğer göstergeler aksini ortaya koymasına rağmen izleyebildiği görülmüştür (Skitka et al., 1999). Bu bulgular, insanın karar sürecine dâhil edilmesinin otomatik sistemlere aşırı güvenme riskini tek başına ortadan kaldırmadığını göstermektedir. Kataloglama süreçlerinde de benzer bir riskin ortaya çıkması mümkündür. Örneğin yapay zekâ tarafından önerilen bir sınıflandırma numarasının, eserin temel konusu ve ilgili sınıflandırma kurallarıyla karşılaştırılmadan kabul edilmesi icra hatasına; sistemin herhangi bir sorun bildirmemesi nedeniyle eksik veya hatalı bibliyografik alanların kontrol edilmemesi ise ihmal hatasına benzer bir duruma yol açabilir. Bu nedenle insan onayının gerçek bir kalite kontrol işlevi görebilmesi için bilgi profesyonellerinin önerileri değerlendirebilecek bilgiye, zamana ve gerekli doğrulama kaynaklarına erişiminin sağlanması önemlidir. İnsan denetiminin niteliği, yalnızca son kararın bir insan tarafından verilmesiyle değil, bu kararın yeterli incelemeye dayanmasıyla ilişkilidir.

İnsan-döngüde yaklaşımının etkililiği, uygulama sonuçlarının düzenli olarak değerlendirilmesiyle de desteklenebilir. Bu kapsamda yapay zekâ tarafından önerilen kayıtların kabul, düzeltme ve ret oranları; en sık karşılaşılan hata türleri; bilgi profesyonellerinin müdahale ettiği metadata alanları ve son kayıtların kalite kontrol sonuçları incelenebilir. Ancak yüksek kabul

oranı tek başına başarılı bir insan–yapay zekâ iş birliğinin göstergesi olarak değerlendirilmemelidir. Kabul edilen önerilerin doğruluğu ve kataloglama standartlarına uygunluğu da ayrıca ölçülmelidir. Böylece insan denetimi yalnızca kurumsal bir ilke olarak kalmayıp sonuçları izlenebilen ve gerektiğinde geliştirilebilen bir kalite güvence mekanizmasına dönüştürülebilir.

Mesleki sorumluluk ve etik

Üretken yapay zekânın kataloglama ve sınıflandırma süreçlerine dâhil edilmesi, bilgi profesyonellerinin mesleki sorumluluklarını ortadan kaldırmamakta; bu sorumlulukların teknoloji destekli çalışma ortamlarında yeniden değerlendirilmesini gerekli kılmaktadır. Kataloglama ve sınıflandırma kararları, bilgi kaynaklarının bibliyografik olarak tanımlanmasının yanı sıra erişilebilirliğini, bulunabilirliğini ve kullanıcılar tarafından anlamlandırılmasını da etkilemektedir. IFLA tarafından yayımlanan Kütüphaneciler ve Diğer Bilgi Çalışanları için Etik Kurallar, bilgiye erişim, bireylere ve topluma karşı sorumluluk, gizlilik, şeffaflık, tarafsızlık ve mesleki yeterlilik gibi ilkeleri kütüphanecilik mesleğinin etik temelleri arasında ele almaktadır (IFLA, 2012). Kataloglama Etiği Kuralları çerçevesinde yapay zekâ kullanımını değerlendiren bir çalışmada, yapay zekâ teknolojilerinin etik kataloglama ilkelerini hem destekleyebileceği hem de bu ilkelerle çelişebileceği belirtilmektedir. Metadata üretiminde hız ve kapasitenin artırılması, veri örüntülerinin ve tutarsızlıkların belirlenmesi gibi olanakların bilgi kaynaklarına erişimi destekleyebileceği; buna karşılık önyargı, veri doğruluğu, gizlilik, şeffaflık ve kurumsal kaynaklara erişimde eşitsizlik gibi sorunların etik riskler oluşturabileceği tartışılmaktadır. Çalışmada ayrıca yapay zekâ uygulamalarının planlama, uygulama ve bakım aşamalarında etik ilkelerin değerlendirilmesine yönelik bir kontrol listesinin geliştirilmesi önerilmektedir (Dover & Grzegorski, 2025). Bu yaklaşım, üretken yapay zekâ destekli kataloglama uygulamalarında

etik sorumluluğun yalnızca ortaya çıkan hataların düzeltilmesiyle sınırlı olmadığını, teknolojinin seçimi ve kullanımı boyunca gözetilmesi gereken bir yükümlülük olduğunu göstermektedir.

Mesleki sorumluluğun temel boyutlarından biri, bibliyografik kayıtların doğruluğunun ve güvenilirliğinin korunmasıdır. Üretken yapay zekâ tarafından önerilen eser adı, sorumluluk bildirimi, yayın bilgisi, konu başlığı veya sınıflandırma numarasının dilsel açıdan tutarlı görünmesi, söz konusu bilginin bibliyografik açıdan doğru olduğunu garanti etmemektedir. Bu nedenle bilgi profesyonelinin görevi, önerileri yalnızca biçimsel olarak incelemek değil, gerektiğinde bilgi kaynağının kendisi, güvenilir bibliyografik kayıtlar, yetke dosyaları ve ilgili kataloglama standartları üzerinden doğrulamaktır. Özellikle paylaşımcı kataloglama uygulamalarında bir kaydın farklı kurumlar tarafından yeniden kullanılabilmesi, doğrulama ve kalite kontrol sorumluluğunu daha önemli hâle getirmektedir. Kütüphanelerde yapay zekâ kullanımına yönelik mesleki rehberde de kataloglama ve sınıflandırma gibi teknik hizmetlerde uzmanların önerileri değerlendirme, değiştirme veya reddetme yetkisini koruması gerektiği vurgulanmaktadır (ALA, 2026). Bu bağlamda yapay zekâ tarafından üretilen bibliyografik bilgilerin kullanılması, söz konusu bilgilerin doğruluğuna ilişkin mesleki sorumluluğu ortadan kaldırmamaktadır.

Mesleki etik sorumluluğun bir diğer boyutu, bilgi kaynaklarının adil ve kültürel açıdan duyarlı biçimde temsil edilmesidir. Konu başlıklarının ve sınıflandırma numaralarının belirlenmesinde kullanılan kavramlar, farklı toplulukların, dillerin ve kültürel bağlamların katalog içinde nasıl görünür olacağını etkileyebilmektedir. Mesleki etik sorumluluğun bir diğer boyutu, bilgi kaynaklarının adil ve kültürel açıdan duyarlı biçimde temsil edilmesidir. IFLA'nın meslek etiği ilkelerinde, bilgiye erişim hakkının ayrımcılık yapılmaksızın korunması, dilsel azınlıkların

kendi dillerinde bilgiye erişim haklarına saygı gösterilmesi, içeriğin kullanıcıların ihtiyaç duydukları bilgilere bağımsız biçimde ulaşmalarını sağlayacak şekilde düzenlenmesi, koleksiyon ve hizmet süreçlerinde tarafsızlığın gözetilmesi ve kişisel inançlarla mesleki görevlerin birbirinden ayrılması gerektiği vurgulanmaktadır (IFLA, 2012). Bu ilkeler kataloglama ve sınıflandırma süreçleri açısından değerlendirildiğinde, konu başlıklarının ve sınıflandırma numaralarının belirlenmesinde yalnızca teknik standartlara uygunluğun değil, kullanılan kavramların farklı toplulukları ve kültürel bağlamları nasıl temsil ettiğinin de dikkate alınması gerektiği söylenebilir. Kataloglama ve metadata profesyonellerinin yapay zekâya ilişkin etik yaklaşımları kendi iş akışlarına uyarlaması ve doğru, kültürel açıdan duyarlı metadata üretme konusundaki uzmanlıklarını koruması gerektiği de ilgili literatürde tartışılmaktadır (Sussmeier & Henry, 2025). Bu çerçevede üretken yapay zekâ tarafından önerilen konu başlıkları ve sınıflandırma numaraları, yalnızca biçimsel ve teknik uygunlukları açısından değil, bilgi kaynaklarının konusal ve kültürel özelliklerini doğru yansıtıp yansıtmadıkları bakımından da değerlendirilmelidir.

Gizlilik ve veri güvenliği, üretken yapay zekâ destekli kataloglama uygulamalarında ayrıca ele alınması gereken etik sorumluluklardır. Kataloglama işlemlerinde kullanılan tüm bilgiler kamuya açık olmayabilir. Yayımlanmamış materyallere ilişkin kayıtlar, bağışçı ve hak sahibi bilgileri, kurumsal belgeler veya özel koleksiyonlara ait erişim kısıtlamaları, verilerin yapay zekâ sistemlerine aktarılmasından önce değerlendirilmesini gerektirebilir. IFLA'nın meslek etiği kuralları, kullanıcı gizliliğine ve kişisel verilerin korunmasına saygı gösterilmesini temel mesleki ilkeler arasında değerlendirmektedir (IFLA, 2012). Yapay zekâ teknolojilerinin kütüphanelerde kullanımına ilişkin IFLA bildirisinde de mahremiyetin ve temel hakların korunması, teknolojinin sorumlu biçimde benimsenmesinin koşulları arasında

ele alınmaktadır (IFLA, 2020). Bu nedenle bilgi profesyonellerinin etik sorumluluğu, yalnızca yapay zekâ tarafından oluşturulan metadatanın içeriğini değil, bu metadatanın hazırlanması sırasında kullanılan verilerin hangi sistemlere, hangi amaçlarla ve hangi koşullarda aktarıldığını da kapsamaktadır.

Bireysel mesleki sorumlulukların etkili biçimde yerine getirilebilmesi, kurumsal düzeyde hesap verebilirlik mekanizmalarının oluşturulmasını da gerektirmektedir. Bir bibliyografik kaydın hazırlanmasında yapay zekâdan yararlanılması, kaydın doğruluğuna ve kurumsal kataloglama politikalarına uygunluğuna ilişkin sorumluluğun bütünüyle teknoloji sağlayıcısına devredilebileceği anlamına gelmemektedir. Bu nedenle kütüphanelerde yapay zekâ kullanımına ilişkin görev ve yetkilerin tanımlanması, hangi işlemlerin uzman onayı gerektirdiğinin belirlenmesi, hataların bildirilmesine yönelik süreçlerin oluşturulması ve gerektiğinde kayıtların düzeltilmesine olanak sağlayan uygulamaların geliştirilmesi önem taşımaktadır. NIST'in yapay zekâ risk yönetimi çerçevesinde, sorumlulukların belirlenmesi, risk yönetiminin kurumsal süreçlere entegre edilmesi ve değerlendirme sonuçlarının belgelenmesi güvenilir yapay zekâ yönetişiminin unsurları arasında ele alınmaktadır (NIST, 2023). Kataloglama bağlamında bu yaklaşım, bireysel mesleki kararların kurumsal politika ve kalite güvence mekanizmalarıyla desteklenmesi gerektiğine işaret etmektedir.

Kurumsal sorumluluğun uygulamaya aktarılmasında, uluslararası etik ilkelerin kütüphanelerin kendi hizmet koşullarına uygun değerlendirme ölçütlerine dönüştürülmesi önemlidir. IFLA tarafından yayımlanan Entry Point to Libraries and AI belgesi, kütüphanelerin yapay zekâ teknolojilerini benimserken olası yararları ve riskleri tartışabilmeleri için değerlendirme soruları ve senaryolar sunmaktadır. Belge, tek başına kesin kararlar veren bir araç olarak değil, meslektaşlar ve kurumlar arasında eleştirel

değerlendirmeyi ve ortak karar almayı destekleyen bir kaynak olarak tasarlanmıştır (IFLA, 2025). Bu yaklaşım kataloglama süreçlerine uyarlandığında, bir yapay zekâ aracının kullanılmasına karar verilmeden önce veri güvenliği, kataloglama standartlarıyla uyumluluk, insan müdahalesinin kapsamı, teknoloji sağlayıcısına bağımlılık ve hata düzeltme olanakları gibi ölçütlerin değerlendirilmesi önerilebilir. Böylece etik ilkelerin yalnızca genel birer ilke olarak benimsenmesi yerine, teknoloji seçimi ve kullanımına ilişkin kurumsal kararlarda dikkate alınması mümkün olabilir.

Türkiye'deki üniversite kütüphaneleri açısından bu değerlendirmeleri yerel uygulamalarla ilişkilendiren kaynaklardan biri, ANKOS Yapay Zekâ Araştırma Grubu tarafından hazırlanan Üniversite Kütüphaneleri İçin Üretken Yapay Zekâ Kullanım Rehberi ve Politikası başlıklı çalışmadır. Rehber, üretken yapay zekânın kütüphanelerde etik ve etkin kullanımına yönelik açıklamalar sunmakta; kataloglama, araştırma desteği, veri yönetimi ve mesleki yeterliliklerin geliştirilmesi gibi konuları ele almaktadır. Belgedeki politika yaklaşımı ise üretken yapay zekâ uygulamalarının sorumluluk, hesap verebilirlik, şeffaflık ve kullanıcı odaklılık ilkeleri doğrultusunda kurumsal süreçlere dâhil edilmesini amaçlamaktadır (ANKOS Yapay Zekâ Araştırma Grubu, 2025). Bu çerçeve, Türkiye'deki üniversite kütüphanelerinin üretken yapay zekâya ilişkin kurumsal politikalar geliştirirken uluslararası etik ilkeleri yerel hizmet gereksinimleriyle birlikte değerlendirmelerine katkı sağlayabilecek bir başvuru kaynağıdır.

Türkiye'deki kütüphane uygulamalarında bibliyografik veri kalitesi ve kataloglama tutarlılığına ilişkin sorunlar da mesleki sorumluluk açısından önem taşımaktadır. Kataloglama etiğini inceleyen bir araştırmada, Ulusal Toplu Katalog (TO-KAT) üzerinden değerlendirilen aynı esere ait kayıtlarda farklı ve tutarsız konu başlıklarının kullanılması, farklı sınıflandırma numaralarının

verilmesi ve aynı kütüphane tarafından aynı esere ilişkin farklı kayıtların oluşturulması gibi sorunlar belirlenmiştir (Beygirci, 2022). Bu bulgular, bibliyografik kayıtların doğruluğu ve tutarlılığına ilişkin sorunların üretken yapay zekâ uygulamalarından önce de mevcut olduğunu göstermektedir. Dolayısıyla üretken yapay zekâ destekli kataloglama ve sınıflandırma sistemlerinin kullanımında, yalnızca teknolojik araçların yeterliliğinin değil, mevcut bibliyografik verilerin niteliğinin ve bu veriler üzerindeki mesleki denetimin de değerlendirilmesi önem taşımaktadır.

Mesleki etik açıdan dikkat edilmesi gereken bir başka konu, yapay zekâ kullanımının bilgi profesyonellerinin uzmanlığı ve çalışma koşulları üzerindeki etkisidir. Otomatik metadata üretiminin yaygınlaşması, bilgi profesyonellerinin görevlerinin kayıt oluşturmaktan yapay zekâ önerilerini değerlendirmeye, kalite kontrolü gerçekleştirmeye ve sistemlerin kullanım koşullarını belirlemeye doğru genişlemesine yol açabilir. Bununla birlikte, bu dönüşümün çalışanları yalnızca otomatik önerileri onaylayan kişilere dönüştüren veya mesleki uzmanlığı ikincil konuma yerleştiren bir iş akışı biçiminde yapılandırılmaması önemlidir. Kataloglama ve metadata profesyonellerinin yapay zekâ uygulamaları karşısındaki mesleki konumlarını ele alan bir çalışmada, etik çerçevelerin iş süreçlerine uyarlanması ve mesleki uzmanlığın korunmasıyla ilgili sorunlar tartışılmaktadır (Sussmeier & Henry, 2025). IFLA'nın yapay zekâ bildirisinde de otomasyonun kütüphane çalışanlarının becerileri ve mesleki rolleri üzerindeki etkilerinin dikkate alınması gerektiği belirtilmektedir (IFLA, 2020). Bu nedenle yapay zekâ destekli kataloglama süreçlerinde verimlilik beklentileri ile mesleki özerklik, uzmanlık ve çalışma koşullarının birlikte değerlendirilmesi gerekmektedir.

Mesleki sorumluluğun sürdürülebilmesi, bilgi profesyonellerinin yapay zekâ teknolojilerini eleştirel biçimde değerlendirebilecek yeterliliklere sahip olmasını da

gerektirmektedir. Yapay zekâ destekli kataloglama uygulamalarında yalnızca kataloglama standartlarının bilinmesi yeterli olmayabilir; kullanılan sistemlerin işleyişi, sınırlılıkları ve veri işleme koşulları hakkında bilgi sahibi olunması da önem taşımaktadır. Yapay zekâ okuryazarlığı, yapay zekâ teknolojilerinin nasıl çalıştığını, sınırlılıklarını ve toplumsal etkilerini anlama becerisini kapsamaktadır (Long & Magerko, 2020). Kataloglama bağlamında bu yeterlilik, bir sistemin hangi görevlerde kullanılmasının uygun olduğunu değerlendirebilme, önerilen bibliyografik bilgileri eleştirel biçimde inceleyebilme ve gerektiğinde sistemin kullanımına ilişkin mesleki gerekçelerle itiraz edebilme yetkinlikleriyle ilişkilendirilebilir. IFLA'nın yapay zekâya ilişkin yaklaşımında da kütüphanelerin teknolojiyi sorumlu biçimde kullanmanın yanı sıra kullanıcıların algoritmik sistemleri anlamalarına ve eleştirel değerlendirmelerine destek olabilecekleri belirtilmektedir (IFLA, 2020). Bu çerçevede yapay zekâ okuryazarlığı, yalnızca teknik bir beceri değil, mesleki sorumluluğun bilinçli biçimde yerine getirilmesini destekleyen bir yeterlilik olarak değerlendirilebilir.

Sonuç olarak üretken yapay zekânın kataloglama ve sınıflandırma süreçlerinde etik kullanımı, yalnızca teknolojik araçların doğru seçilmesine veya insan onayının iş akışına eklenmesine bağlı değildir. Bibliyografik doğruluğun korunması, bilgi kaynaklarının adil ve kültürel açıdan duyarlı biçimde temsil edilmesi, verilerin gizliliğinin gözetilmesi, kurumsal sorumlulukların tanımlanması ve mesleki uzmanlığın sürdürülmesi birlikte değerlendirilmelidir. Bu çerçevede bilgi profesyoneli, yapay zekâ tarafından oluşturulan önerileri kullanan bir uygulayıcı olmanın ötesinde, teknolojinin hangi koşullarda kullanılacağını değerlendiren, etik ilkelerin uygulanmasını gözeten ve bibliyografik kayıtların güvenilirliğini koruyan bir karar verici konumundadır.

Bilgi profesyonelinin değişen yetkinlikleri

Yapay zekâ destekli kataloglama ve sınıflandırma uygulamaları, bilgi profesyonelinin uzmanlığını ortadan kaldırmaktan çok bu uzmanlığın kapsamını ve gerektirdiği yetkinlikleri değiştirebilir. Geleneksel kataloglama ve sınıflandırma bilgisi önemini korumaktadır; çünkü yapay zekâ tarafından üretilen bir bibliyografik kaydın veya önerilen sınıflandırma numarasının doğruluğunu değerlendirebilmek, ilgili standartlar ve bilgi düzenleme ilkeleri hakkında uzmanlık gerektirmektedir. Bununla birlikte, veri okuryazarlığı, yapay zekâ okuryazarlığı, istem tasarımı, çıktı değerlendirme ve hata analizi gibi yetkinlikler de mesleki uygulamalarda önem kazanmaktadır. Veri okuryazarlığı, verilerin toplanması, analiz edilmesi, yorumlanması ve karar verme süreçlerinde kullanılmasına ilişkin bilgi ve becerileri kapsamaktadır (Mandinach & Gummer, 2013). Yapay zekâ okuryazarlığı ise bireylerin yapay zekâ teknolojilerini eleştirel biçimde değerlendirmelerini, bu teknolojilerle etkili iletişim ve iş birliği kurmalarını ve yapay zekâyı farklı bağlamlarda bir araç olarak kullanmalarını sağlayan yetkinlikler bütünü olarak tanımlanmaktadır (Long & Magerko, 2020). İstem tasarımı, büyük dil modellerinden beklenen çıktıyı elde etmek amacıyla girdilerin yapılandırılmasıyla ilişkili bir beceri olup bilgi okuryazarlığı eğitimi bağlamında da ele alınmaktadır (Lo, 2023).

Bu yetkinliklerin kataloglama ve sınıflandırma süreçlerine yansıması, yapay zekâ araçlarını kullanabilmenin ötesinde, bu araçların ürettiği bibliyografik bilgileri eleştirel biçimde değerlendirebilmeyi de gerektirmektedir. Akademik kütüphane çalışanlarına yönelik hazırlanan yapay zekâ yetkinlikleri çerçevesinde etik değerlendirme, bilgi ve anlama, analiz ve değerlendirme ile kullanım ve uygulama olmak üzere dört temel boyut ele alınmaktadır (Association of College and Research Libraries [ACRL], 2025). Bu çerçeve kataloglama ve sınıflandırma süreçlerine uyarlandığında veri okuryazarlığı, kullanılan

bibliyografik kayıtların kaynağını, niteliğini, eksikliklerini ve tutarsızlıklarını belirleme açısından; yapay zekâ okuryazarlığı ise sistemlerin işleyişini, olanaklarını ve sınırlılıklarını anlama açısından önem taşımaktadır. İstem tasarımı, bibliyografik tanımlama, konu analizi veya sınıflandırma gibi görevlerde modelden beklenen çıktının kapsamını, biçimini ve kullanılacak ölçütleri açıkça belirtmeye yardımcı olabilir. Bununla birlikte uygun bir istem hazırlanması çıktının doğruluğunu garanti etmediğinden, bilgi profesyonelinin yapay zekâ önerilerini yetke dosyaları, sınıflandırma şemaları ve kataloglama standartlarıyla karşılaştırabilmesi önemini korumaktadır. Dolayısıyla değişen mesleki yetkinlikler, geleneksel kataloglama bilgisinin yerini alan bağımsız beceriler olarak değil, bu bilgiyi yapay zekâ destekli iş akışlarında kullanmayı sağlayan tamamlayıcı yeterlilikler olarak değerlendirilmelidir.

Kütüphanecilerin üretken yapay zekâ araçlarını mesleki faaliyetlerinde kullanma deneyimlerini inceleyen bir araştırmada, 272 katılımcının %66,5'inin bu araçlardan yararlandığı belirlenmiştir. Araştırmada katılımcıların üretken yapay zekâ kullanımında karşılaştıkları başlıca güçlükler arasında doğruluk ve güvenilirlik (%41), etik, gizlilik ve güvenlik (%25,7) ile istem oluşturma ve iletişim sorunları (%21) yer almaktadır. Ayrıca katılımcıların %55,9'u üretken yapay zekâya uyum sağlamak için eğitim ve öğretimi önemli bir gereksinim olarak belirtmiştir (Luo, 2025). Bu bulgular, üretken yapay zekânın kütüphane hizmetlerinde benimsenmesinin yalnızca teknolojik araçların erişilebilirliğine bağlı olmadığını; bilgi profesyonellerinin mesleki yetkinliklerinin geliştirilmesini, güvenilirlik sorunlarının değerlendirilmesini ve kurumsal destek mekanizmalarının oluşturulmasını da gerektirdiğini göstermektedir.

ALA Core tarafından güncellenen kataloglama ve metadata profesyonellerinin yetkinliklerine yapay zekâ ve büyük dil

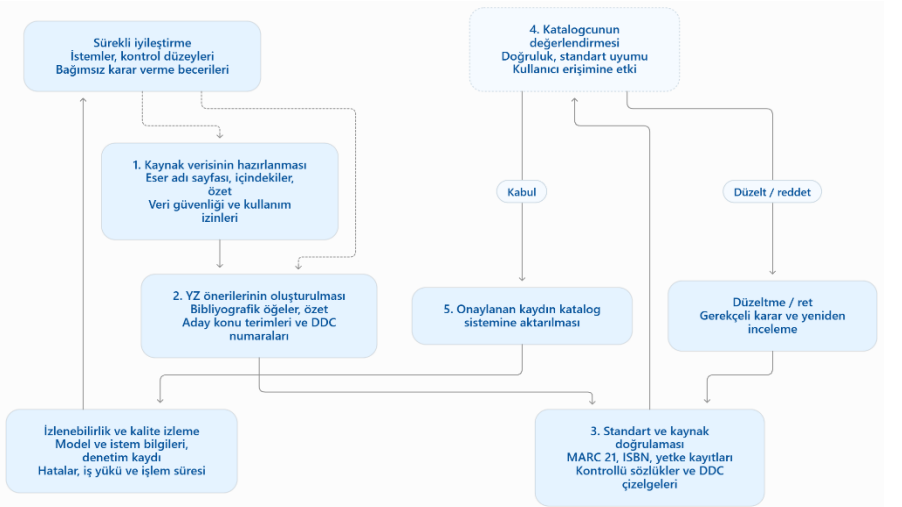
modellerine ilişkin unsurların eklenmesi, bu dönüşümün mesleki yeterlilikler düzeyinde de ele alındığını göstermektedir (ALA Core, 2023). Bu gelişmeler doğrultusunda bilgi profesyonelinin gelecekteki rolü, yalnızca bibliyografik kayıt üreten bir uzman olarak değil; metadata kalitesini yöneten, otomasyon süreçlerini denetleyen ve bilgi düzenleme sistemlerinde insan kullanıcılar ile algoritmik araçlar arasında aracılık eden bir uzman olarak da değerlendirilebilir. Bu dönüşümün mesleki eğitime yansıtılması için kataloglama öğretiminde yalnızca kuralların aktarılmasıyla yetinilmemesi, otomatik sistemlerin önerilerinin hangi gerekçelerle kabul veya reddedileceğinin de ele alınması önerilebilir. Örneğin öğrencilerin ve çalışanların aynı bilgi kaynağına farklı yapay zekâ modelleri tarafından önerilen metadata ve sınıflandırma numaralarını karşılaştırmaları, hata türlerini belirlemeleri ve önerileri yetke dosyaları ile kataloglama standartları üzerinden doğrulamaları sağlanabilir. Böylece yapay zekâ, mesleki uzmanlığın yerine geçen bir araç olarak değil, uzmanlığın eleştirel biçimde kullanılmasını destekleyen bir eğitim bileşeni olarak değerlendirilebilir.

İnsan–yapay zekâ iş birliği iş akışı

Kataloglama ve sınıflandırma süreçlerinde üretken yapay zekânın kullanımı, bibliyografik kayıt oluşturma işleminin bütünüyle otomatikleştirilmesinden ziyade, insan uzmanlığının farklı aşamalarda korunduğu kontrollü bir iş akışı çerçevesinde ele alınabilir. ALA'nın yapay zekâ kullanım rehberinde, teknik hizmetlerde yapay zekâ destekli uygulamaların insan denetimini içermesi ve bilgi profesyonelinin otomatik önerileri kabul etme, düzenleme veya reddetme yetkisini koruması gerektiği belirtilmektedir (ALA, 2026). Benzer biçimde Ex Libris tarafından geliştirilen Alma AI Metadata Assistant uygulamasında, yapay zekâ tarafından oluşturulan metadata önerileri bilgi profesyonelinin değerlendirmesine sunulmakta ve bilgi profesyoneline bu önerileri

kabul etme, düzeltme veya reddetme olanağı tanınmaktadır (Ex Libris, t.y.). Bu yaklaşımlardan hareketle kataloglama ve sınıflandırma süreçleri için kaynak verisinin hazırlanması, yapay zekâ önerilerinin oluşturulması, standart ve kaynak doğrulaması, uzman değerlendirmesi ve kalite izlemeden oluşan beş aşamalı bir iş akışı önerilmektedir. Söz konusu iş akışı, mevcut bir uygulamanın doğrudan aktarımı olmayıp ilgili rehberler ve uygulama örneklerinden yararlanılarak bu çalışma kapsamında geliştirilmiş kavramsal bir modeldir. Önerilen iş akışının aşamaları Şekil 1'de gösterilmektedir.

Şekil 1. Kataloglama ve sınıflandırmada önerilen insan–yapay zekâ iş birliği iş akışı



Kaynak: ALA (2026), Ex Libris (t.y.), NIST (2023) ve ANKOS Yapay Zekâ Araştırma Grubu (2025) temel alınarak yazar tarafından oluşturulmuştur.

İlk aşamada, bibliyografik kaydın oluşturulmasında kullanılacak kaynak verileri belirlenir ve bu veriler hazırlanır. Eser adı sayfası, kapak, içindekiler, özet ve erişim izni bulunan dijital tam metin gibi kaynaklardan yararlanılabilir. Bu aşamada verilerin okunabilirliği, eksiksizliği, ilgili esere ait olup olmadığı ve kullanım izinleri değerlendirilmelidir. Özellikle lisanslı içeriklerin, kişisel

verilerin ve kamuya açık olmayan belgelerin üçüncü taraf yapay zekâ sistemlerine aktarılmasında kurumsal veri koruma politikaları ve ilgili hukuki yükümlülükler dikkate alınmalıdır. Üniversite kütüphanelerinde üretken yapay zekânın kullanımına ilişkin ANKOS rehberinde de etik kullanım, veri güvenliği ve kurumsal sorumlulukların belirlenmesi üzerinde durulmaktadır (ANKOS Yapay Zekâ Araştırma Grubu, 2025).

İkinci aşamada, üretken yapay zekâ sistemine açıkça tanımlanmış ve sınırlandırılmış görevler verilir. Bu görevler kaynak üzerinde bulunan bibliyografik öğelerin çıkarılması, kısa bir özet taslağı hazırlanması, aday konu terimlerinin belirlenmesi veya aday sınıflandırma numaralarının önerilmesi biçiminde düzenlenebilir. Modelden hangi verileri kullanmasının beklendiği, çıktıyı hangi biçimde sunacağı ve kaynakta bulunmayan bilgileri tahmin etmek yerine belirsiz olarak belirtmesi gerektiği istem (prompt) içerisinde açıklanmalıdır. Ayrıca doğrudan kaynaktan çıkarılan bilgiler ile modelin yorumlayarak oluşturduğu öneriler birbirinden ayrılmalıdır. Böylece sonraki aşamalarda hangi alanların kaynak üzerinden, hangilerinin konu analizi ve sınıflandırma araçlarıyla doğrulanacağı belirlenebilir. Akademik kütüphane çalışanlarının yapay zekâ yetkinliklerine ilişkin çerçevede de görevlerin niteliğine uygun, açık ve etkili istemlerin oluşturulması ile istemlerin çıktı kalitesini geliştirmek amacıyla yeniden düzenlenmesi mesleki yeterlilikler arasında ele alınmaktadır (ACRL, 2025).

Üçüncü aşamada, yapay zekâ çıktıları bibliyografik standartlar ve güvenilir başvuru araçları kullanılarak doğrulanır. ISBN'nin biçimsel uygunluğu kontrol edilebilir; ancak biçimsel olarak geçerli bir ISBN'nin ilgili esere ait olduğu ayrıca doğrulanmalıdır. Eser adı, sorumluluk bildirimi, yayın bilgileri ve diğer betimleyici alanlar kaynak üzerindeki bilgilerle karşılaştırılır. MARC 21 alanlarının yapısal uygunluğu, kişi ve kurum adlarının yetke kayıtlarıyla eşleşmesi, önerilen konu terimlerinin kontrollü

sözlüklerdeki karşılığı ve Dewey Onlu Sınıflama numaralarının ilgili çizelgelerle uyumu incelenir. Bu aşamada kural tabanlı kontrollerden ve güvenilir dış başvuru kaynaklarından yararlanılabilir. Bununla birlikte yapısal açıdan geçerli bir bibliyografik kaydın içerik ve konu temsili bakımından da doğru olduğu varsayılmamalıdır. Bu aşamada biçimsel ve yapısal uygunluk kontrolleri, uygun doğrulama araçlarıyla otomatik olarak gerçekleştirilebilirken, kaynakların konusal kapsamına ve bibliyografik bilgilerin doğruluğuna ilişkin değerlendirmelerde bilgi profesyonelinin incelemesi gerekebilir. Otomatik kontrollerin sonuçları ve belirlenen belirsizlikler, dördüncü aşamada gerçekleştirilecek uzman değerlendirmesine temel oluşturur. Yapay zekâ araçlarının belirli kütüphane görevlerine uygunluğunun, doğruluk ve güvenilirlik gibi ölçütler üzerinden değerlendirilmesi, akademik kütüphane çalışanlarının mesleki yetkinlikleri arasında da yer almaktadır (ACRL, 2025).

Dördüncü aşamada bilgi profesyoneli, doğrulama sonuçlarını ve yapay zekâ önerilerini birlikte değerlendirerek nihai mesleki kararını verir. Bilgi profesyonelinin değerlendirmesi; önerinin bibliyografik doğruluğu, kataloglama standartlarına uygunluğu, yetke kayıtları ve kontrollü sözlüklerle uyumu, sınıflandırma numarasının konusal doğruluğu ve kullanıcıların bilgi kaynaklarına erişimi üzerindeki olası etkileri gibi ölçütlere dayanmalıdır. Bu değerlendirme sonucunda öneriler kabul edilebilir, düzeltilebilir veya reddedilebilir; belirsiz durumlarda ek kaynak incelemesi yapılabilir. Özellikle konu analizi, ayrıntılı sınıflandırma, yetke denetimi ve kültürel açıdan duyarlı kavramların temsili gibi işlemlerde uzman değerlendirmesi belirleyici olmalıdır. Kabul edilen, düzeltilen veya reddedilen önerilere ilişkin kararlar, kurumun olanakları doğrultusunda kayıt altına alınmalı; özellikle düzeltme ve ret kararları ile yüksek riskli işlemlerde karar gerekçeleri belgelenmelidir. Böylece mesleki kararların izlenebilirliği desteklenebilir ve yapay zekâ sisteminin hangi kataloglama

görevlerinde başarılı veya yetersiz kaldığı sistematik biçimde değerlendirilebilir. Bu yaklaşım, bilgi profesyonelinin otomatik öneriler üzerindeki karar yetkisini koruyan uygulamalarla ve yapay zekâ sistemlerinde hesap verebilirlik ile izlenebilirliği öne çıkaran risk yönetimi ilkeleriyle uyumludur (ALA, 2026; Ex Libris, t.y.; NIST, 2023).

Beşinci aşamada bilgi profesyoneli tarafından onaylanan bibliyografik kayıt katalog sistemine aktarılır ve kalite izleme süreci yürütülür. Bu süreçte yapay zekâ destekli işlemlere ilişkin asgari kayıt bilgilerinin belirlenmesi önemlidir. Kullanılan modelin adı ve mümkünse sürümü, işlemin tarihi, uygulanan istemin veya istem şablonunun sürümü, yapay zekâ tarafından önerilen metadata alanları, gerçekleştirilen düzeltmeler ve denetimi yapan yetkili personelin kurumsal kaydı gibi bilgiler izlenebilirlik amacıyla tutulabilir. Özellikle yüksek riskli veya sonradan uyuşmazlık doğurabilecek kararların gerekçeleriyle birlikte belgelenmesi, hesap verebilirliği ve kalite geliştirme çalışmalarını destekleyebilir. Bununla birlikte kayıtların kapsamı, erişim yetkileri ve saklama süreleri kurumsal gereksinimler ve veri güvenliği ilkeleri doğrultusunda belirlenmelidir. NIST tarafından geliştirilen Yapay Zekâ Risk Yönetimi Çerçevesi, yapay zekâ sistemlerinin belgelenmesini, performanslarının izlenmesini ve risklerin yaşam döngüsü boyunca değerlendirilmesini öngörmektedir (NIST, 2023). Alma AI Metadata Assistant uygulamasında da yapay zekâ destekli oluşturulan veya zenginleştirilen kayıtların belirlenebilmesi ve üretilen metadatanın kalitesine ilişkin geri bildirim sağlanması için işlevler bulunmaktadır (Ex Libris, t.y.).

Önerilen iş akışında bütün görevlerin aynı risk düzeyinde değerlendirilmesi uygun değildir. Kaynak üzerinde açıkça bulunan bibliyografik bilgilerin çıkarılması, çoğu durumda doğrudan kaynakla karşılaştırılarak doğrulanabilirken; özet oluşturma, konu analizi ve ayrıntılı sınıflandırma önerileri daha kapsamlı anlamsal

değerlendirme gerektirebilir. Bununla birlikte risk düzeyi yalnızca işlemin türüne göre değil, hatanın olası etkisine, tespit edilebilirliğine ve düzeltilme güçlüğüne göre belirlenmelidir. Yapay zekâ sistemlerine ilişkin risklerin kullanım bağlamı içerisinde tanımlanması, ölçülmesi ve yönetilmesi gerektiği NIST çerçevesinde de vurgulanmaktadır (NIST, 2023). Bu yaklaşımdan hareketle kurumlar, kataloglama görevlerine özgü kontrol düzeyleri belirleyebilir ve özellikle yanlış konu temsili veya hatalı sınıflandırma gibi kullanıcıların bilgi kaynaklarına erişimini etkileyebilecek kararlar için ayrıntılı uzman incelemesi öngörebilir.

İş akışının tasarımında otomasyon yanlılığı ve beceri erozyonu riskleri de dikkate alınmalıdır. Otomatik önerilere aşırı güvenilmesi, bilgi profesyonelinin hatalı önerileri yeterince sorgulamadan kabul etmesine; konu analizi ve sınıflandırma kararlarının sürekli otomatik sistemlere bırakılması ise bağımsız mesleki değerlendirme becerilerinin yeterince kullanılmamasına yol açabilir. Otomasyon yanlılığına ilişkin araştırmalar, karar destek sistemlerinden gelen önerilere aşırı güvenmenin insan kararlarını olumsuz etkileyebileceğini göstermektedir (Skitka et al., 1999). Bu risklerin kataloglama alanında aynı biçimde ve aynı düzeyde ortaya çıktığı varsayılmamakla birlikte, iş akışının bağımsız değerlendirme becerilerini koruyacak biçimde tasarlanması önemlidir. Bu amaçla bilgi profesyonelinin belirli görevlerde önce bağımsız konu analizi veya sınıflandırma kararı vermesi, ardından kendi değerlendirmesini yapay zekâ önerisiyle karşılaştırması ve gerektiğinde öneriyi gerekçeli biçimde reddedebilmesi desteklenebilir. Böylece insan denetimi, yalnızca otomatik önerilerin sonradan onaylanmasıyla sınırlı kalmayıp mesleki muhakemenin sürdürülmesine de katkı sağlayabilir.

İş akışının etkililiğinin değerlendirilebilmesi için kalite kontrolün ölçülebilir göstergelerle desteklenmesi önemlidir. Yapay zekâ önerilerinin kabul, düzeltme ve reddedilme oranları; alan

bazında hata türleri, kayıt başına toplam işlem süresi; uzman doğrulamasına ayrılan zaman ve sonradan tespit edilen hatalar izlenebilir. Ancak kabul oranının yüksek olması tek başına bibliyografik doğruluğun yüksek olduğu anlamına gelmemektedir. Benzer biçimde, yalnızca otomatik öneri üretiminin süresini ölçmek, kataloglama sürecinin bütünü açısından zaman tasarrufu sağlandığını göstermeyebilir. Bu nedenle mümkün olduğunda yapay zekâ destekli iş akışı ile mevcut kataloglama süreci, benzer kaynak türleri ve kalite ölçütleri kullanılarak karşılaştırılmalıdır. İnsan denetiminin gerektirdiği düzeltme emeği de toplam iş yüküne dâhil edilmelidir. Yapay zekâ sistemlerinin performansının ölçülmesi, kullanım sürecinde izlenmesi ve değerlendirme sonuçlarının risk yönetiminde kullanılması, NIST çerçevesinin temel bileşenleri arasında yer almaktadır (NIST, 2023).

Kalite izleme sonuçları, iş akışının sabit bir yapı olarak değil, düzenli biçimde gözden geçirilen bir süreç olarak değerlendirilmesine olanak sağlayabilir. Belirli bir metadata alanında sürekli hata üretilmesi durumunda istemin yeniden düzenlenmesi, kullanılan başvuru kaynaklarının gözden geçirilmesi veya ilgili görevin otomatik öneri kapsamından çıkarılması düşünülebilir. Ayrıca model sürümündeki değişiklikler, kullanılan kataloglama standartlarının güncellenmesi ve yeni materyal türlerinin sisteme dâhil edilmesi, mevcut kontrollerin yeniden değerlendirilmesini gerektirebilir. Bu nedenle kalite izlemenin yalnızca bireysel bilgi profesyoneli performansını ölçmeye yönelik bir araç olarak değil, kullanılan sistemin ve iş akışının geliştirilmesine hizmet eden kurumsal bir mekanizma olarak tasarlanması önemlidir. Bu yaklaşım, yapay zekâ sistemlerinin performansının ve ortaya çıkan risklerin kullanım süreci boyunca izlenmesi ve elde edilen değerlendirmelerin risk yönetimine yansıtılması ilkeleriyle uyumludur (NIST, 2023).

Sonuç olarak önerilen insan–yapay zekâ iş birliği iş akışının temel amacı, kataloglama ve sınıflandırma süreçlerinden insan uzmanlığını çıkarmak değil, bu uzmanlığın gerekli olduğu karar noktalarını belirginleştirmektir. Tekrarlayan veri çıkarımı ve biçimlendirme işlemleri otomasyonla desteklenebilirken, belirsiz konu analizi, karmaşık sınıflandırma, yetke denetimi ve kültürel temsil gibi işlemler daha kapsamlı mesleki değerlendirme gerektirebilir. Bununla birlikte hangi görevlerin otomasyonla destekleneceği, kurumun kaynakları, personel yetkinlikleri ve kabul edilebilir risk düzeyi dikkate alınarak belirlenmelidir. Böylece önerilen model, verimlilik ile bibliyografik kaliteyi birlikte gözeten, insan karar yetkisini koruyan ve sürekli iyileştirmeye açık bir uygulama çerçevesi sunabilir.

Kurumsal Politika ve Kalite Güvencesi

Üretken yapay zekâ araçlarının kataloglama ve sınıflandırma süreçlerinde bireysel çalışan tercihlerinden kurumsal iş akışlarına dönüşmesi için açık ve uygulanabilir politikaların oluşturulması gerekmektedir. Hangi araçların kullanılabileceği, hangi veri türlerinin bu araçlara aktarılabileceği, yapay zekâ tarafından hangi metadata alanlarının üretilebileceği, zorunlu doğrulama adımları, sorumluluk dağılımı ve hata düzeltme mekanizmaları kurumsal düzeyde tanımlanmalıdır. Kütüphanelerde yapay zekâ kullanımına ilişkin rehberlerde de mesleki değerlerin korunması, veri güvenliği, insan denetimi ve kurumsal sorumlulukların belirlenmesi vurgulanmaktadır (ALA, 2026; ANKOS Yapay Zekâ Araştırma Grubu, 2025).

Kurumsal politika aynı zamanda şeffaflık ve izlenebilirliği desteklemelidir. Yapay zekâ aracılığıyla oluşturulan veya zenginleştirilen kayıtların kurum içi sistemlerde uygun biçimde işaretlenmesi, kullanılan model veya hizmetin belirlenebilmesi ve belirli aralıklarla örnekleme yoluyla kalite denetimi yapılması

düşünülebilir. Model güncellemeleri ve kullanım koşullarındaki değişiklikler performansı etkileyebileceğinden, bir kez gerçekleştirilen pilot uygulamanın sonuçları kalıcı kabul edilmemelidir. Yeni model sürümleri, farklı materyal türleri ve değişen kataloglama standartları doğrultusunda yeniden değerlendirme yapılmalıdır. Bu yaklaşım, yapay zekâ sistemlerinin yaşam döngüsü boyunca izlenmesini ve risklerinin düzenli olarak değerlendirilmesini öngören ilkelerle uyumludur (NIST, 2023).

Tedarikçi tarafından geliştirilen yapay zekâ sistemlerinde kurumun denetim kapasitesi ayrıca önem taşımaktadır. Modelin bütün teknik ayrıntılarına erişilemese bile hangi verilerin dış sisteme gönderildiği, verilerin saklanma koşulları, model eğitimi amacıyla kullanılıp kullanılmadığı, önerilerin hangi kaynaklarla doğrulanabildiği, hataların nasıl raporlandığı ve yapay zekâ özelliğinin gerektiğinde devre dışı bırakılıp bırakılamadığı değerlendirilmelidir. Yapay zekâ destekli hizmetlerin tedarik ve değerlendirme süreçlerinde gizlilik, doğruluk, erişim ve kullanıcı hakları açısından incelenmesi gerektiği de vurgulanmaktadır (ALA, 2026).

Bu ilkeler doğrultusunda üniversite kütüphanelerinin teknik hizmetler birimlerinde uygulanabilecek örnek politika hükümleri aşağıda sunulmaktadır. Bu hükümler, ilgili rehberlerden doğrudan aktarılmış bağlayıcı kurallar değil; kataloglama ve sınıflandırma süreçlerinin gereksinimleri doğrultusunda bu çalışma kapsamında geliştirilmiş önerilerdir. Önerilen görev dağılımı, kurumların idari yapılanmaları ve yetki düzenlemeleri doğrultusunda uyarlanabilir.

Kapsam ve Sorumluluklar

1.Bu politika, kütüphanenin teknik hizmetler biriminde üretken yapay zekâ araçlarının kataloglama, sınıflandırma, metadata üretimi ve konu analizi süreçlerinde kullanımını kapsar.

2.Kurumsal verilerin işlenmesinde yalnızca yetkili birimler tarafından kullanım amacı, güvenlik koşulları ve veri işleme özellikleri değerlendirilerek onaylanmış yapay zekâ araçları kullanılmalıdır.

3.Yapay zekâ tarafından oluşturulan veya zenginleştirilen bibliyografik kayıtların mesleki doğrulamasından yetkili kataloglama personeli sorumlu olmalıdır. Teknik hizmetler birim sorumlusu, iş akışının uygulanmasını, kalite kontrol faaliyetlerini ve hata düzeltme süreçlerini koordine etmelidir. Kütüphane yönetimi, yapay zekâ kullanım politikasının onaylanmasından, görev ve yetkilerin belirlenmesinden ve gerekli durumlarda uygulamanın sınırlandırılması veya durdurulmasına ilişkin kararların alınmasından sorumlu olmalıdır. Bilgi işlem ve ilgili veri güvenliği birimleri ise sistemin teknik güvenliğini, erişim yetkilerini ve veri işleme koşullarını kendi görev alanları çerçevesinde denetlemelidir.

4.Teknik hizmetler personeline yapay zekâ okuryazarlığı, istem tasarımı, çıktı doğrulama, hata analizi ve veri güvenliği konularında görevlerine uygun eğitim verilmelidir.

Veri Yönetimi ve Gizlilik

5.Yapay zekâ araçlarına aktarılabilecek veri türleri önceden tanımlanmalıdır. Kişisel veriler, lisanslı içerikler ve kamuya açık olmayan belgeler, ilgili mevzuat, lisans koşulları ve kurumsal veri koruma politikaları doğrultusunda işlenmelidir.

6.Harici yapay zekâ hizmetlerine aktarılan verilerin saklanma süreleri, kullanım amaçları, üçüncü taraflarla paylaşım koşulları ve model eğitimi için kullanılıp kullanılmadığı tedarikçi sözleşmeleri ve hizmet koşulları üzerinden değerlendirilmelidir.

7.Hassas verilerin işlenmesinde kurum içi sunucularda çalışan modeller dâhil olmak üzere alternatif teknolojik çözümler;

veri güvenliği, teknik yeterlilik, bakım gereksinimleri ve maliyet açısından karşılaştırılarak değerlendirilmelidir.

Doğrulama ve Kalite Kontrol

8.Yapay zekâ tarafından oluşturulan betimleyici bibliyografik alanlar, ilgili bilgi kaynağı veya güvenilir bibliyografik başvuru kaynaklarıyla karşılaştırılarak doğrulanmalıdır. Doğrulanmamış yapay zekâ önerileri nihai bibliyografik kayıt olarak kullanılmamalıdır. Bu doğrulama işlemlerinin tamamlanmasından yetkili kataloglama personeli sorumlu olmalıdır.

9.Yapay zekâ tarafından önerilen yetke biçimleri, kontrollü konu başlıkları ve sınıflandırma numaraları, ilgili yetke dosyaları, kontrollü sözlükler ve sınıflandırma çizelgeleri kullanılarak uzman kataloglama personeli tarafından değerlendirilmelidir.

10.Kataloglama görevleri, olası hata maliyeti, hatanın tespit edilebilirliği ve kullanıcıların bilgi kaynaklarına erişimi üzerindeki etkileri dikkate alınarak risk düzeylerine ayrılmalıdır. Risk düzeylerinin belirlenmesi ve uygun kontrol yöntemlerinin uygulanması, teknik hizmetler birim sorumlusunun koordinasyonunda yürütülmeli; yüksek riskli işlemler için ayrıntılı uzman incelemesi öngörülmelidir.

İzlenebilirlik ve Hesap Verebilirlik

11.Yapay zekâ destekli işlemlerde kullanılan model veya hizmetin adı, mümkün olduğunda sürümü, işlem tarihi, istem şablonu ve denetim sonucuna ilişkin bilgiler, kurumun teknik olanakları ve veri koruma politikaları doğrultusunda kayıt altına alınmalıdır. Özellikle önemli düzeltme ve ret kararlarının gerekçeleri belgelenmelidir. Bu kayıtların oluşturulması ve korunmasına ilişkin sorumluluklar kurumsal düzeyde belirlenmelidir.

12.Kurumsal politika, teknik hizmetler birim sorumlusunun koordinasyonunda ve ilgili kurumsal birimlerin katılımıyla yılda en

az bir kez gözden geçirilmeli; değerlendirme sonuçları kütüphane yönetimine sunulmalıdır. Model sürümlerinde, materyal türlerinde, kullanılan standartlarda veya veri güvenliği koşullarında önemli değişiklikler meydana geldiğinde ayrıca değerlendirme yapılmalıdır.

13.Yapay zekâ destekli bibliyografik kayıtlarda tespit edilen hataların personel ve kullanıcılar tarafından bildirilebilmesi, incelenmesi ve düzeltilmesi için uygun bildirim kanalları, sorumlu personel ve işlem basamakları belirlenmelidir. Bibliyografik hataların incelenmesi ve düzeltilmesi teknik hizmetler biriminin sorumluluğunda yürütülmeli; sistem güvenliğini veya veri korumayı ilgilendiren bildirimler ilgili kurumsal birimlere yönlendirilmelidir.

14.Sistematik bibliyografik hatalar, ciddi veri güvenliği sorunları veya kabul edilemez kalite kayıpları tespit edildiğinde ilgili yapay zekâ uygulamasının kullanımı geçici olarak sınırlandırılabilmeli veya durdurulabilmelidir. Bu karar, teknik hizmetler birim sorumlusunun değerlendirmesi ve gerektiğinde bilgi işlem ile veri güvenliği birimlerinin görüşleri doğrultusunda, kurumun yetki düzenlemelerine uygun olarak kütüphane yönetimi tarafından alınmalıdır.

Risk Temelli Kalite Güvencesi

Kurumsal politika hükümlerinin uygulanmasında, üretken yapay zekâ destekli kataloglama ve sınıflandırma işlemlerinin risk düzeylerinin dikkate alınması önem taşımaktadır. Kataloglama ve sınıflandırma süreçlerinde üretken yapay zekânın farklı amaçlarla kullanılması, işlemlerin aynı düzeyde insan denetimi gerektirmediğini göstermektedir. Kaynak üzerinden doğrudan doğrulanabilen betimleyici bibliyografik alanlarla konu analizi, yetke denetimi ve sınıflandırma gibi yorumlama gerektiren işlemler arasında risk ve doğrulama gereksinimleri bakımından farklılıklar bulunmaktadır. Bununla birlikte risk düzeyi yalnızca işlemin türüne değil, kullanılan verilerin niteliğine, olası hatanın etkisine ve

kurumun denetim kapasitesine göre de değişebilmektedir. Yapay zekâ risklerinin kullanım bağlamı içerisinde değerlendirilmesi ve uygun kontrol mekanizmalarının geliştirilmesi, risk yönetimi yaklaşımının temel ilkeleri arasındadır (NIST, 2023).

Bu doğrultuda yüksek riskli işlemlerde ayrıntılı uzman doğrulaması, daha düşük riskli işlemlerde ise uygun kontrol ve örnekleme yöntemleri uygulanabilir. Bununla birlikte örnekleme yoluyla yapılan kalite denetimi, nihai bibliyografik kayıtların oluşturulmasından önce gerçekleştirilmesi gereken zorunlu doğrulamanın yerine geçmemelidir. Teknik hizmetler birim sorumlusunun koordinasyonunda yürütülecek kalite güvencesi çalışmalarında bibliyografik doğruluk, konu erişiminin uygunluğu, düzeltme ve ret oranları ile işlem süreleri izlenebilir. Kataloglama personeli tarafından belirlenen hata türleri ve düzeltme gereksinimleri değerlendirilerek sistematik sorunlar tespit edilebilir. Ciddi veri güvenliği sorunlarında ilgili teknik birimlerin değerlendirmesi alınmalı; sistematik bibliyografik hatalar veya kabul edilemez kalite kayıpları karşısında uygulamanın sınırlandırılması ya da geçici olarak durdurulması kütüphane yönetiminin kararına sunulmalıdır. Bu yaklaşım, risklerin sürekli izlenmesi ve değişen koşullara göre yönetilmesi ilkeleriyle uyumludur (NIST, 2023).

Böylece kurumsal politika, yalnızca yapay zekâ kullanımına ilişkin kuralları belirleyen bir belge olmaktan çıkarak bibliyografik kalitenin, mesleki sorumluluğun ve kullanıcıların bilgiye erişiminin korunmasını destekleyen sürekli bir denetim mekanizmasına dönüşebilir.

Teknik Hizmetlerde Uygulama Senaryoları

Üretken yapay zekânın kataloglama ve sınıflandırma süreçlerindeki kullanımını yalnızca kuramsal olanaklar üzerinden değerlendirmek, teknolojinin mesleki iş akışlarına nasıl dâhil

edilebileceğini açıklamak açısından yeterli olmayabilir. Bu nedenle üretken yapay zekânın teknik hizmetlerdeki olası kullanım biçimleri, farklı kataloglama ve metadata yönetimi senaryoları üzerinden ele alınabilir. Aşağıda sunulan altı senaryo, ilgili literatür ve mevcut teknolojik uygulamalardan hareketle oluşturulmuş örneklerdir. Her senaryoda yapay zekânın üstlenebileceği görevler, olası yararlar, temel riskler ve gerekli insan denetimi birlikte değerlendirilmektedir. İlk senaryoda ayrıca önceki bölümde önerilen beş aşamalı insan–yapay zekâ iş birliği iş akışının kataloglama sürecinde nasıl uygulanabileceğini göstermek amacıyla temsili bir örneğe yer verilmiştir.

Birinci senaryo: Yeni gelen basılı kitabın kataloglanması

Bu senaryoda bilgi profesyoneli, eserin eser adı, sayfası, kapak ve içindekiler gibi bibliyografik bilgi içeren bölümlerini dijital görüntüler aracılığıyla sisteme sağlayabilir. Üretken yapay zekâ, bu görüntülerden eser adı, sorumluluk bildirimi, yayın bilgileri ve ISBN gibi unsurları çıkarmak; içindekiler bilgisini düzenlemek, özet taslağı hazırlamak ve aday konu terimleri önermek amacıyla kullanılabilir. Bibliyografik veriler ve kaynak görüntülerinden yararlanarak yeni kayıt oluşturma veya mevcut kayıtları zenginleştirme önerileri sunan yapay zekâ destekli metadata araçları, bu kullanım biçimine örnek oluşturmaktadır. Bu araçlarda önerilen metadataların bilgi profesyonelinin incelemesine sunulması, insan denetiminin iş akışındaki yerini göstermektedir (Ex Libris, t.y.).

Temsili kataloglama örneği: Yeni gelen ve çocuklarda kalp hastalıklarının tanı ve tedavisine odaklanan bir kitabın kataloglanması düşünüldüğünde, ilk aşamada bilgi profesyoneli eserin eser adı sayfasını, yayın bilgilerini ve içindekiler bölümünü inceleyerek yapay zekâ sistemine aktarılabilecek kaynak verilerini belirler. İkinci aşamada üretken yapay zekâdan bu veriler üzerinden betimleyici bibliyografik alanları çıkarması, kısa bir özet taslağı hazırlaması, aday konu terimleri oluşturması ve olası Dewey Onlu

Sınıflama (DOS) numaralarını önermesi istenir. Model, eserin kapsamına ilişkin yorumuna bağlı olarak pediatri veya kardiyolojiyle ilişkili farklı sınıflandırma adayları sunabilir. Üçüncü aşamada eserden çıkarılan betimleyici bilgiler kaynak görüntüleriyle karşılaştırılır; ISBN, yayın bilgileri ve diğer alanların doğruluğu kontrol edilir. Önerilen konu terimleri ve sınıflandırma adayları ise ilgili kontrollü sözlükler ve DOS çizelgeleri üzerinden incelenir. Dördüncü aşamada bilgi profesyoneli, eserin baskın konusunu içindekiler ve gerektiğinde tam metin üzerinden değerlendirerek önerilerin uygunluğuna karar verir. Kaynağın konusal kapsamını yeterince temsil etmeyen önerileri reddedebilir, hatalı betimleyici alanları düzeltebilir veya uygun önerileri kabul edebilir. Beşinci aşamada bilgi profesyoneli tarafından onaylanan bibliyografik kayıt katalog sistemine aktarılır; gerçekleştirilen düzeltmeler ve değerlendirme sonuçları, kurumsal politika doğrultusunda kalite izleme sürecine dâhil edilir. Bu örnek, belirli bir yapay zekâ sistemiyle gerçekleştirilmiş deneysel bir uygulamayı değil, önerilen iş akışının olası işleyişini göstermektedir.

Risk değerlendirmesi: Temel riskler, kaynakta bulunan bilgilerin yanlış çıkarılması, eksik bırakılması veya kaynakta bulunmayan bilgilerin model tarafından üretilmesidir. Eser adı ve sorumluluk bildirimi gibi doğrudan görüntü üzerinden doğrulanabilen alanlar görece düşük risk taşırken yayın bilgileri, seri ilişkileri ve çoklu sorumluluk bildirimleri daha karmaşık değerlendirme gerektirebilir. Konu ve sınıflandırma önerileri ise yorumlayıcı işlemler olduğundan daha yüksek risk taşımaktadır. Bu nedenle betimleyici alanlar kaynakla karşılaştırılmalı, anlamsal öneriler ilgili başvuru kaynaklarıyla doğrulanmalı ve nihai kayıt bilgi profesyonelinin onayından sonra sisteme aktarılmalıdır. Önerilen genel risk düzeyi: Orta; konu analizi ve sınıflandırma işlemlerinde yüksek.

İkinci senaryo: Kısa veya eksik bibliyografik kayıtların zenginleştirilmesi

Kütüphane kataloglarında geçmiş dönemlerde oluşturulmuş bazı kayıtlar yalnızca temel bibliyografik bilgileri içerebilir. Bu kayıtlarda özet, içindekiler notu, ayrıntılı konu erişimi veya ilişkili eser bilgileri bulunmayabilir. Üretken yapay zekâ, eserin kendisinden veya güvenilir bilgi kaynaklarından sağlanan veriler doğrultusunda eksik alanlar için öneriler oluşturabilir. Mevcut kısa bibliyografik kayıtların yapay zekâ destekli zenginleştirilmesi, Alma AI Metadata Assistant'ın belgelenmiş işlevleri arasındadır. Sistem, önerilen alanların bilgi profesyoneli tarafından incelenmesine ve uygun görülen değişikliklerin kayda aktarılmasına olanak tanımaktadır (Ex Libris, t.y.). Bu süreçte bilgi profesyoneli, mevcut kayıt ile önerilen alanları karşılaştırarak hangi bilgilerin korunacağına, değiştirileceğine veya ekleneceğine karar vermelidir.

Risk değerlendirmesi: Temel risk, mevcut doğru bibliyografik bilgilerin hatalı önerilerle değiştirilmesi veya yeni eklenen bilgilerin farklı bir eser ya da baskıya ait olmasıdır. Üretilen özet ve konu terimlerinde içerik sapmaları da görülebilir. Bu nedenle özgün kayıt korunmalı, önerilen alanlar mevcut bilgilerden ayırt edilmeli ve değişiklikler doğrulanmadan nihai kayda aktarılmamalıdır. Özellikle toplu zenginleştirme işlemlerinde hataların çok sayıda kayda yayılma olasılığı dikkate alınmalıdır. Önerilen genel risk düzeyi: Orta; toplu ve denetimsiz işlemlerde yüksek.

Üçüncü senaryo: Retrospektif kataloglama

Kart kataloglarda veya eski otomasyon sistemlerinde bulunan bibliyografik kayıtların güncel kataloglama ortamlarına aktarılması, özellikle geniş koleksiyonlarda yoğun insan emeği gerektirebilmektedir. Yapay zekâ destekli kataloglama uygulamalarına ilişkin bir çalışma, Belçika Kraliyet Kütüphanesi

bağlamından hareketle kataloglama iş akışlarında otomasyonun kullanılabileceği alanları ele almaktadır (Lowagie, 2025). Bu tür uygulamalarda optik karakter tanıma (OCR) ile elde edilen metinlerin düzenlenmesi, bibliyografik alan adaylarının belirlenmesi ve olası kayıt eşleşmelerinin önerilmesi amacıyla yapay zekâdan yararlanılabilir. Bununla birlikte OCR, kayıt eşleştirme ve üretken yapay zekâ işlemlerinin farklı teknolojik işlevler olduğu ve her aşamanın ayrı doğrulama gerektirdiği göz önünde bulundurulmalıdır. Bilgi profesyoneli, özellikle belirsiz kayıt eşleşmelerini ve farklı basımlara ilişkin bilgileri özgün kaynaklarla karşılaştırarak değerlendirmelidir.

Risk değerlendirmesi: Eski katalog kartlarının fiziksel durumu, kullanılan yazı biçimleri, eksik bilgiler ve tarihsel kataloglama uygulamalarındaki farklılıklar hatalı veri aktarımına neden olabilir. Özellikle kişi adlarının yanlış eşleştirilmesi, farklı basımların birleştirilmesi ve tarihsel verilerin güncel kurallara uyarlanırken özgün anlamının kaybolması önemli risklerdir. Model, OCR çıktılarındaki düzensizlikleri yorumlamaya yardımcı olsa da tarihsel bağlamı her zaman doğru değerlendiremeyebilir. Bu nedenle özgün kayıt görüntüleri korunmalı, belirsiz eşleştirmeler uzman incelemesine alınmalı ve toplu değişiklikler öncesinde örneklem denetimi yapılmalıdır. Önerilen genel risk düzeyi: Yüksek.

Dördüncü senaryo: Dijital koleksiyonlar ve kurumsal arşivlerde metadata oluşturulması

Doğuştan dijital materyallerde geleneksel basılı kaynaklarda bulunan eser adı sayfası gibi sabit bilgi kaynakları her zaman mevcut olmayabilir. Bu nedenle dosya adı, dosya içeriği, gömülü metadataler, web sayfası ve kurumsal bağlamın birlikte değerlendirilmesi gerekebilir. Üretken yapay zekâ araçlarının doğuştan dijital koleksiyonlarda eser adı, özet, anahtar kelime ve tam metin URL'si gibi alanların çıkarılması veya oluşturulmasında kullanılabileceği araştırılmıştır. Bu tür uygulamalarda

yapılandırılmış veri çıkarımı ile anlamsal yorumlama gerektiren görevlerin farklı doğrulama ihtiyaçları bulunduğu görülmektedir (Tian, 2026). Bu süreçte bilgi profesyoneli veya ilgili metadata uzmanı, önerilen alanları kaynak dosya ve kurumsal kayıtlarla karşılaştırmalı; erişim bağlantılarını ve kullanım koşullarını kontrol etmelidir.

Risk değerlendirmesi: Temel riskler, materyalin bağlamsal bilgilerinin eksik yorumlanması, yanlış dosya eşleştirmesi, hatalı bağlantı üretilmesi ve içerikle uyumsuz anahtar kelimelerin önerilmesidir. Dosya adı veya web sayfası gibi kaynaklar tek başına yeterli olmayabilir. Ayrıca erişim kısıtlaması bulunan belgelerin harici yapay zekâ hizmetlerine aktarılması gizlilik ve lisans sorunları doğurabilir. Bu nedenle üretilen alanlar kaynak dosyayla karşılaştırılmalı, bağlantılar kontrol edilmeli ve veri aktarım izinleri değerlendirilmelidir. Önerilen genel risk düzeyi: Orta; hassas içerik ve yetkisiz veri aktarımı söz konusu olduğunda yüksek.

Beşinci senaryo: Konu analizi ve sınıflandırma desteği

Üretken yapay zekâ, bilgi profesyoneline doğrudan tek bir Dewey Onlu Sınıflama (DOS) numarası sunan sistem olarak değil, konu analizini destekleyen bir ön değerlendirme aracı olarak da kullanılabilir. Modelden kaynağın baskın ve ikincil konularını belirlemesi, önerilerini kaynakta yer alan bilgilerle ilişkilendirmesi, birden fazla aday sınıf sunması ve bilgi profesyonelini ilgili sınıflandırma çizelgelerine yönlendirmesi istenebilir. Bununla birlikte modelin sunduğu açıklamalar, kararın hangi iç işlemler sonucunda üretildiğinin kesin kanıtı olarak kabul edilmemelidir. Büyük dil modellerinin Dewey Onlu Sınıflama numarası atama performansını inceleyen bir araştırma, bu araçların sınıflandırma desteğinde kullanılma potansiyelini ve özellikle ayrıntılı sınıflandırma kararlarında karşılaşılabilecek sınırlılıkları ortaya koymaktadır (Reshi et al., 2026). Bilgi profesyoneli, aday numaraları

güncel çizelgeler ve uygulama notlarıyla karşılaştırarak eserin baskın konusuna uygun nihai sınıflandırma kararını vermelidir.

Risk değerlendirmesi: Başlıca riskler, eserin baskın konusunun yanlış belirlenmesi, aşırı genel veya uygun olmayan sınıflandırma numaralarının önerilmesi, sınıflandırma hiyerarşisinin hatalı yorumlanması ve konu terimlerinin kontrollü sözlüklerle uyumsuz olmasıdır. Modelin doğrulanmamış fakat ikna edici açıklamalar üretmesi de bilgi profesyonelinin değerlendirmesini etkileyebilir. Bu nedenle önerilen numaralar güncel sınıflandırma çizelgeleri ve uygulama notları üzerinden kontrol edilmeli, nihai konu ve sınıflandırma kararı bilgi profesyoneli tarafından verilmelidir. Önerilen genel risk düzeyi: Yüksek.

Altıncı senaryo: Bibliyografik kayıtların kalite kontrolü

Üretken yapay zekâ yalnızca yeni metadata oluşturmak için değil, mevcut kayıtların tutarlılığını değerlendirmek amacıyla da kullanılabilir. Örneğin sistem, bibliyografik kayıt ile özet arasındaki olası konu uyumsuzluklarını işaretleyebilir, benzer eserlerin farklı konu başlıklarıyla temsil edildiği durumları belirleyebilir veya eksik erişim noktalarını bilgi profesyonelinin dikkatine sunabilir. Büyük dil modellerinin insan tarafından gerçekleştirilen konu indekslemesini değerlendirme kapasitesini inceleyen bir araştırma, bu araçların belirli değerlendirme görevlerinde yararlı olabildiğini, ancak yanlış atanmış konu tanımlayıcılarını belirlemede sınırlılıklar taşıdığını göstermektedir (Włodarczyk & Zdunek, 2026). Bu senaryoda yapay zekâ tarafından işaretlenen olası hatalar doğrudan düzeltilmemeli; bilgi profesyoneli tarafından özgün kaynak ve ilgili standartlar üzerinden incelenmelidir.

Risk değerlendirmesi: Temel riskler, doğru kayıtların hatalı olarak işaretlenmesi (yanlış pozitif) ve mevcut hataların tespit edilememesidir (yanlış negatif). Yanlış pozitif sonuçlar gereksiz düzeltmelere ve iş yüküne neden olabilirken yanlış negatif sonuçlar

katalogdaki hataların devam etmesine yol açabilir. Otomatik değerlendirmelerin sorgulanmadan kabul edilmesi, doğru bibliyografik kayıtların bozulması riskini de artırabilir. Bu nedenle yapay zekânın işaretlediği kayıtlar kesin hata olarak değil, uzman incelemesine alınacak kayıtlar olarak değerlendirilmelidir. Önerilen genel risk düzeyi: Orta; otomatik toplu düzeltmelerde yüksek.

Bu senaryolar birlikte değerlendirildiğinde üretken yapay zekânın kataloglama süreçlerindeki olası katkısının, bibliyografik kayıt oluşturma işlemini bütünüyle üstlenmesinden çok, farklı alt görevleri desteklemesinde ortaya çıkabileceği görülmektedir. Veri çıkarımı, taslak oluşturma, aday üretme, karşılaştırma, kayıt eşleştirme ve olası hata belirleme gibi işlevler mesleki iş akışlarını destekleyebilir. Buna karşılık yetke kararları, karmaşık konu analizi, sınıflandırma talimatlarının yorumlanması ve yerel kataloglama politikalarına bağlı işlemler daha yoğun insan uzmanlığı gerektirmektedir.

Burada sunulan risk düzeyleri, belirli bir kütüphane sisteminde ölçülmüş hata oranlarına veya deneysel risk puanlarına dayanmamaktadır. Değerlendirmeler; olası hata türleri, hataların kullanıcı erişimine etkileri ve gerekli doğrulama işlemleri dikkate alınarak bu çalışma kapsamında oluşturulmuştur. Dolayısıyla gerçek uygulamalarda risk düzeylerinin kurumsal veri yapısı, kullanılan teknoloji, işlem hacmi ve mevcut kontrol mekanizmaları doğrultusunda yeniden belirlenmesi gerekmektedir.

Yapay Zekâ Destekli Kataloglamanın Değerlendirilmesi

Bir yapay zekâ aracının kataloglama ve sınıflandırma süreçlerinde kullanılabilir olup olmadığı yalnızca genel doğruluk yüzdesiyle belirlenemez. Bibliyografik kayıtlar, farklı özelliklere sahip alanlardan ve mesleki karar süreçlerinden oluştuğu için değerlendirme çok boyutlu gerçekleştirilmelidir. Üretken yapay zekâ ile metadata oluşturma ve Dewey Onlu Sınıflama (DOS) numarası

atama üzerine gerçekleştirilen araştırmalar, performansın görevin niteliğine ve değerlendirme düzeyine göre değişebildiğini göstermektedir (Tian, 2026; Reshi et al., 2026). Bu doğrultuda bibliyografik doğruluk, hata türleri, verimlilik, tutarlılık, açıklanabilirlik, farklı koleksiyon gruplarındaki performans ve mesleki kullanılabilirlik birlikte ele alınmalıdır. Aşağıda sunulan yedi değerlendirme boyutu ve ölçüm göstergeleri, ilgili araştırmalar ile yapay zekâ risk yönetimi ilkelerinden yararlanılarak bu çalışma kapsamında geliştirilmiş önerilerdir.

Birinci boyut: Bibliyografik doğruluk ve anlamsal uygunluk

Bibliyografik kayıtların değerlendirilmesi alan bazında yapılmalıdır. Eser adı, yayıncı, yayın tarihi ve ISBN gibi betimleyici alanlarda kaynakla tam eşleşme veya alan doğruluğu ölçülebilir. Bununla birlikte noktalama, yazım farklılıkları ve kataloglama kurallarının gerektirdiği biçimsel düzenlemeler dikkate alınmalıdır. Özet ve konu terimleri gibi anlamsal alanlarda içeriğin doğru temsil edilmesi, kontrollü sözlüklerle uyum ve konu kapsamının yeterliliği uzman değerlendirmesini gerektirir. Sınıflandırmada ise yalnızca ana sınıfın doğru belirlenmesi değil, notasyonun gerekli ayrıntı düzeyine ulaşması ve sınıflandırma çizelgelerindeki kurallara uygunluğu da incelenmelidir. Metadata oluşturma ve DOS numarası atama üzerine yapılan çalışmalar, yapılandırılmış veri çıkarımı ile anlamsal yorumlama gerektiren görevlerin farklı performans özellikleri taşıdığını göstermektedir (Tian, 2026; Reshi et al., 2026).

Önerilen ölçüm göstergeleri: Alan bazında doğruluk oranı, tam eşleşme oranı, uzman tarafından uygun bulunan konu terimi oranı, kontrollü sözlük uyum oranı, ana sınıf doğruluğu, ayrıntılı sınıflandırma doğruluğu ve geçerli notasyon oranı.

İkinci boyut: Hata türleri ve halüsinasyon

Bibliyografik bir alanın boş bırakılması ile yanlış bilgiyle doldurulması aynı sonuçları doğurmayabilir. Eksik alanlar bilgi profesyoneli tarafından fark edilerek tamamlanabilirken gerçekte bulunmayan bir yayıncı, yanlış ISBN veya kaynak içeriğiyle uyumsuz bir konu başlığı doğruymuş gibi kabul edilebilir. Bu nedenle üretken yapay zekâ değerlendirmesinde yalnızca doğruluk oranı değil, eksik veri, yanlış veri, uydurma bilgi ve hatalı önerilerin güven verici biçimde sunulması da dikkate alınmalıdır. Büyük dil modellerinin ürettiği bibliyografik atıfların incelendiği bir araştırmada, var olmayan kaynakların üretilebildiği ve gerçek kaynaklara ait bibliyografik bilgilerde de önemli hatalar görülebildiği ortaya konulmuştur (Walters & Wilder, 2023).

Önerilen ölçüm göstergeleri: Eksik alan oranı, yanlış alan oranı, kaynakta bulunmayan bilgi üretme oranı, yanlış pozitif oranı, yanlış negatif oranı ve uzman incelemesinde tespit edilen kritik hata sayısı. Güven puanı sunan sistemlerde yüksek güvenle verilen hatalı önerilerin oranı ayrıca değerlendirilebilir.

Üçüncü boyut: İşlem verimliliği ve iş yükü

Yapay zekâ destekli bir sistemin bibliyografik kayıt taslağını kısa sürede oluşturması, toplam kataloglama süresinin mutlaka azaldığı anlamına gelmez. Bilgi profesyonelinin önerileri doğrulamak, hataları düzeltmek ve eksik alanları tamamlamak için harcadığı süre toplam kazanımı azaltabilir. Ayrıca gereğinden fazla aday konu başlığı veya sınıflandırma numarası sunulması, karar verme sürecini uzatabilir. Doğuştan dijital koleksiyonlarda metadata oluşturulmasına yönelik bir araştırma, üretken yapay zekânın bazı yapılandırılmış görevlerde işlem verimliliğini artırabildiğini, ancak bu yararın görevin niteliğine bağlı olduğunu göstermektedir (Tian, 2026). Bu nedenle verimlilik değerlendirmesinde yalnızca modelin çıktı üretme süresi değil, insan incelemesi dâhil bütün işlem süresi dikkate alınmalıdır. Karşılaştırmanın benzer materyal türleri ve benzer güçlük düzeyindeki kayıtlar üzerinde yapılması önemlidir.

Önerilen ölçüm göstergeleri: Kayıt başına ortalama toplam işlem süresi, yapay zekâ çıktı üretim süresi, insan doğrulama ve düzeltme süresi, birim zamanda tamamlanan kayıt sayısı, kayıt başına düzeltme sayısı ve geleneksel iş akışına göre net zaman tasarrufu oranı.

Dördüncü boyut: Tutarlılık ve sürdürülebilir performans

Aynı tür bilgi kaynaklarına benzer kataloglama kararlarının verilmesi, katalog bütünlüğünün ve öngörülebilir bilgi erişiminin korunması açısından önemlidir. Üretken yapay zekâ sistemleri aynı girdiye farklı zamanlarda veya farklı model sürümlerinde değişen yanıtlar verebilir. Bu durum, genel doğruluk düzeyi kabul edilebilir olsa bile bibliyografik kayıtlar arasında tutarsızlıklara neden olabilir. Yapay zekâ sistemlerinin kullanım süresince izlenmesi, performanslarının değerlendirilmesi ve değişen risk koşullarına göre kontrol mekanizmalarının güncellenmesi, yapay zekâ risk yönetimi yaklaşımının temel unsurları arasındadır (NIST, 2023). Bu doğrultuda kurumlar, önceden belirlenmiş bir test kümesini farklı zamanlarda ve model güncellemelerinden sonra yeniden kullanarak performans değişimlerini izleyebilir. Özellikle konu başlığı, yetke biçimi ve sınıflandırma alanlarında önceki sürüme göre ortaya çıkan kalite kayıpları ayrıca incelenmelidir.

Önerilen ölçüm göstergeleri: Aynı girdiye verilen yanıtların tutarlılık oranı, benzer eserlerde konu başlığı uyum oranı, model sürümleri arasındaki alan bazında doğruluk farkı, güncelleme sonrasında ortaya çıkan yeni hata sayısı ve tekrarlanan testlerde performans değişimi.

Beşinci boyut: Açıklanabilirlik ve denetlenebilirlik

Bilgi profesyoneli, yapay zekâ tarafından önerilen bibliyografik bilgilerin hangi kaynak verilerle ilişkilendirildiğini inceleyebilmelidir. Özellikle konu başlığı ve sınıflandırma önerilerinde ilgili metin bölümlerinin, yetke kayıtlarının ve

sınıflandırma çizelgesi bilgilerinin gösterilmesi uzman doğrulamasını kolaylaştırabilir. Bununla birlikte modelin ürettiği doğal dil açıklamaları, kararın gerçekten hangi iç işlemler sonucunda oluşturulduğunu kanıtlamaz. Yapay zekâ sistemlerinin şeffaflık, açıklanabilirlik ve hesap verebilirlik açısından değerlendirilmesi, güvenilir yapay zekâ yaklaşımının önemli unsurları arasında yer almaktadır. Kütüphanelerde yapay zekâ kullanımına yönelik mesleki rehberler de insan denetiminin ve kurumsal sorumluluğun korunmasını vurgulamaktadır (NIST, 2023; ALA, 2026). Bu nedenle açıklanabilirlik değerlendirmesinde yalnızca gerekçenin anlaşılır olması değil, önerilen bilginin doğrulanabilir kaynaklarla ilişkilendirilebilmesi ve yapılan işlemlerin izlenebilmesi önem taşımaktadır.

Önerilen ölçüm göstergeleri: Doğrulanabilir kaynak dayanağı bulunan önerilerin oranı, kaynakla eşleştirilebilen alan oranı, işlem kaydı tutulabilen öneri oranı, uzman tarafından yeterli bulunan gerekçe oranı ve kaynağı doğrulanamadığı için reddedilen öneri sayısı.

Altıncı boyut: Farklı dil ve materyal türlerinde performans eşitliği

Bir yapay zekâ sistemi İngilizce monografilerde başarılı sonuçlar üretirken Türkçe eserlerde, Osmanlı Türkçesi materyallerde, çocuk kitaplarında, görsel-işitsel kaynaklarda veya yerel koleksiyonlarda daha düşük performans gösterebilir. Genel ortalama değerler bu farklılıkları gizleyebileceğinden pilot uygulamalar, kurum koleksiyonunun gerçek çeşitliliğini temsil eden örneklerle yürütülmelidir. Yapay zekâ sistemlerinin farklı kullanım bağlamlarında değerlendirilmesi ve olası eşitsizliklerin dikkate alınması, güvenilir yapay zekâ ilkeleriyle ilişkilidir. Kütüphanelerde yapay zekâ kullanımında mesleki değerlerin ve kullanıcıların eşit erişim haklarının korunması da önem taşımaktadır (NIST, 2023;

ALA, 2026). Bu nedenle dil, materyal türü ve koleksiyon özellikleri bakımından ayrı performans sonuçları raporlanmalıdır.

Önerilen ölçüm göstergeleri: Dil bazında doğruluk oranı, materyal türüne göre alan doğruluğu, koleksiyon grupları arasındaki performans farkı, her gruptaki kritik hata oranı ve değerlendirme örnekleminin koleksiyon dağılımını temsil etme düzeyi.

Yedinci boyut: İnsan kabulü ve mesleki kullanılabilirlik

Yapay zekâ destekli kataloglama sistemlerinin başarısı yalnızca teknik performansla sınırlı değildir. Bilgi profesyonellerinin önerileri ne ölçüde yararlı bulduğu, hangi alanlarda düzeltme yaptığı, hangi önerileri reddettiği ve sistemin iş akışını nasıl etkilediği de değerlendirilmelidir. Kullanıcı arayüzünün anlaşılabilirliği, önerilerin kolay karşılaştırılabilmesi ve hatalı bilgilerin hızlı biçimde düzeltilebilmesi, mesleki kullanılabilirliği etkileyen unsurlardır. Kütüphanelerde yapay zekâ uygulamalarının insan merkezli yaklaşımla değerlendirilmesi ve çalışanların mesleki uzmanlığının korunması gerektiği vurgulanmaktadır (ALA, 2026). Bununla birlikte yüksek kabul oranı tek başına bibliyografik doğruluğun göstergesi değildir. Bilgi profesyonellerinin yanlış önerileri sorgulamadan kabul etmesi de mümkün olduğundan kabul ve ret kararları bağımsız kalite değerlendirmeleriyle birlikte incelenmelidir.

Önerilen ölçüm göstergeleri: Öneri kabul oranı, öneri düzeltme oranı, öneri ret oranı, ret gerekçelerinin dağılımı, bilgi profesyoneli memnuniyet düzeyi, algılanan iş yükü ve kullanım sırasında karşılaşılan sorunların sıklığı.

Yukarıda tanımlanan yedi değerlendirme boyutunun uygulamaya aktarılabilmesi, göstergelerin ortak bir ölçüm süreci içerisinde ele alınmasını gerektirmektedir. Bu kapsamda izlenebilecek örnek değerlendirme süreci aşağıda açıklanmaktadır.

Ölçüm ve değerlendirme süreci: Önerilen göstergelerin uygulamada kullanılabilmesi için öncelikle farklı konu alanlarını, dilleri ve materyal türlerini temsil eden bir pilot değerlendirme kümesi oluşturulmalıdır. Bu kümedeki bibliyografik kayıtlar, ilgili bilgi kaynakları ve kataloglama standartları temel alınarak uzman personel tarafından doğrulanmalı ve karşılaştırma için referans veri olarak kullanılmalıdır. Yapay zekâ tarafından üretilen veya zenginleştirilen kayıtlar bu referanslarla karşılaştırılarak alan bazında doğruluk, eksik ve yanlış bilgi oranları, konu terimi uygunluğu ve sınıflandırma başarısı hesaplanabilir. Metadata oluşturma ve sınıflandırma görevlerinin farklı değerlendirme gereksinimleri taşıması, bu alanların ayrı göstergelerle incelenmesini gerekli kılmaktadır (Tian, 2026; Reshi et al., 2026). Verimlilik değerlendirmesinde ise benzer özelliklere sahip kayıtlar üzerinde geleneksel kataloglama ve yapay zekâ destekli kataloglama süreçlerinin toplam işlem süreleri, doğrulama gereksinimleri ve düzeltme sayıları karşılaştırılmalıdır. Güvenilirliğin değerlendirilmesi amacıyla aynı test kümesi farklı zamanlarda ve model güncellemelerinden sonra yeniden kullanılmalı; kritik hata türleri ve performans değişimleri izlenmelidir. Bu yaklaşım, yapay zekâ sistemlerinin yaşam döngüsü boyunca değerlendirilmesi ve risklerinin izlenmesine ilişkin ilkelerle uyumludur (NIST, 2023). Kurumlar, pilot uygulama öncesinde işlem türlerinin risk düzeylerine göre kabul edilebilir hata oranlarını, asgari kalite koşullarını ve beklenen verimlilik kazanımlarını belirlemelidir. Böylece değerlendirme sonuçları yalnızca sayısal performans göstergeleri olarak kalmayıp sistemin kullanılmasına, sınırlandırılmasına veya yeniden düzenlenmesine ilişkin kurumsal kararlara dayanak oluşturabilir.

Bu yedi boyutun birlikte değerlendirilmesi, yapay zekâ destekli kataloglama sistemlerinin yalnızca teknik doğruluk açısından değil, bibliyografik kalite, mesleki iş yükü, kurumsal

sürdürülebilirlik ve kullanıcıların bilgiye erişimi bakımından da incelenmesini sağlayabilir. Bununla birlikte önerilen göstergelerin tamamının her kurumda aynı kapsamda kullanılması zorunlu değildir. Kütüphaneler, koleksiyonlarının özellikleri, personel kapasitesi ve yapay zekâdan yararlanılan işlemlerin risk düzeylerine göre öncelikli göstergeleri belirleyebilir.

Bu değerlendirme çerçevesi, kütüphanelerin "Yapay zekâ kullanıyor muyuz?" sorusundan "Hangi görevde, hangi kalite düzeyinde ve hangi denetim mekanizmasıyla kullanıyoruz?" sorusuna geçişini desteklemektedir. Böylece üretken yapay zekânın kataloglama ve sınıflandırma süreçlerine entegrasyonu, bibliyografik kaliteyi, mesleki sorumluluğu ve kullanıcıların bilgiye erişimini birlikte gözeten sistematik bir değerlendirme yaklaşımıyla ele alınabilir.

Türkiye'de Üretken Yapay Zekâ Destekli Kataloglama ve Sınıflandırmada Araştırma Öncelikleri

Türkiye'deki bibliyografik veri altyapısı ve kataloglama uygulamalarına ilişkin araştırmalar, üretken yapay zekânın kataloglama ve sınıflandırma süreçlerine uyarlanabileceği alanların belirlenmesi açısından önemli bir başlangıç noktası sunmaktadır. Üretken yapay zekâ ve büyük dil modellerinin bibliyografik veri üretimi, konu analizi ve sınıflandırma kararlarının desteklenmesinde kullanılabilmesi, mevcut kataloglama uygulamalarının, bibliyografik veri kalitesinin ve mesleki gereksinimlerin birlikte değerlendirilmesini gerekli kılmaktadır. Bu nedenle Türkiye'de gelecekte yürütülebilecek araştırmaların belirlenmesinden önce, ilgili alanlarda gerçekleştirilmiş çalışmaların ortaya koyduğu bulguların ele alınması önem taşımaktadır.

Türkiye'de üniversite kütüphanelerinin kataloglama uygulamalarına ilişkin araştırmalar, kurumlar arasında iş süreçleri ve bibliyografik kayıtların oluşturulma biçimleri bakımından

farklılıklar bulunduğunu göstermektedir. Bazı üniversite kütüphanelerinde diğer kütüphanelerin kataloglarından bibliyografik kayıt aktarımına başvurulurken özgün kataloglama uygulamaları da kullanılabilmektedir (Çakmak, 2019). Kataloglama standartlarının benimsenmesine ilişkin kurumsal deneyimler ise ortak kuralların uygulanmasının teknik altyapının yanı sıra personelin bilgi birikimi ve kurumsal uygulamalarla da ilişkili olduğunu düşündürmektedir (Çakmak, 2018). Bu bulgular, Türkiye'deki kataloglama süreçlerinin tek biçimli bir yapıya sahip olmadığını ve yeni teknolojilerin uygulanmasında kurumsal farklılıkların dikkate alınması gerektiğini göstermektedir.

Türkiye'de kataloglama eğitimi ve araştırmaların mesleki uygulamalara yansıması da incelenen konular arasındadır. Bilgi ve Belge Yönetimi bölümlerindeki kataloglama eğitimine ilişkin bir çalışmada, eğitim olanakları ve program içeriklerinin değerlendirilmesi gerektiğine işaret edilmektedir (Bayter, 2023). Kütüphanecilik ve bilgibilim alanındaki araştırmaların uygulamaya yansımasını ele alan bir başka çalışmada ise araştırmacılar ile uygulayıcılar arasındaki iş birliğinin güçlendirilmesi gerektiği vurgulanmaktadır (Zencir & Öztürk, 2022). Bu çalışmalar, kataloglama süreçlerinin teknolojik dönüşümünde yalnızca kullanılan araçların değil, mesleki eğitim ve araştırma–uygulama ilişkisinin de önem taşıdığını düşündürmektedir.

Türkiye'de otomatik konu sınıflandırmasına ilişkin araştırmalar da gelecekteki çalışmalar için bir temel oluşturmaktadır. Türkçe eser adlarından hareketle Dewey Onlu Sınıflama (DOS) tabanlı konu sınıflandırmasının makine öğrenmesi yöntemleriyle gerçekleştirilebilirliğini inceleyen bir doktora çalışmasında, eser adlarının konu sınıflandırmasında kullanılabildiği; ancak çok sayıda sınıfın bulunması, veri dengesizliği ve alt sınıflardaki örnek yetersizliğinin model performansını sınırladığı ortaya konmuştur (Akyüz, 2026). Bu çalışma üretken yapay zekâya değil, makine

öğrenmesiyle otomatik sınıflandırmaya odaklanmaktadır. Bununla birlikte ortaya koyduğu veri ve sınıflandırma sorunları, büyük dil modelleriyle yürütülecek karşılaştırmalı araştırmalar açısından değerlendirilebilir.

Türkiye'deki kataloglama uygulamaları, mesleki eğitim ve otomatik sınıflandırma üzerine gerçekleştirilen bu çalışmalar, üretken yapay zekâ destekli teknik hizmet araştırmalarının hangi sorunlardan hareketle tasarlanabileceğini göstermektedir. Ancak mevcut kataloglama süreçlerinin incelenmesi veya geleneksel makine öğrenmesiyle sınıflandırma yapılması, üretken yapay zekâ modellerinin bibliyografik kayıt oluşturma ve sınıflandırma performansının doğrudan değerlendirildiği anlamına gelmemektedir. Bu nedenle gelecekteki araştırmalarda, mevcut çalışmalardan elde edilen bilgi birikiminin üretken yapay zekâya özgü deneysel tasarımlarla geliştirilmesi önem taşımaktadır. Aşağıda Türkiye bağlamında yürütülebilecek araştırmalar için öncelikli alanlar ve bu alanlarda kullanılabilecek değerlendirme yaklaşımları ele alınmaktadır.

Türkçe bibliyografik verilerden metadata oluşturulması, öncelikli araştırma alanlarından biri olarak değerlendirilebilir. Bu kapsamda tıp, hukuk, sosyal bilimler ve edebiyat gibi farklı disiplinlerden, çeşitli yayınevlerine ait Türkçe kitaplar seçilebilir. Örneğin bir ön araştırmada 50–100 kitaplık bir örneklem oluşturularak farklı yayın yılları, baskı özellikleri ve tipografik düzenlerin temsil edilmesi sağlanabilir. Örneklem büyüklüğü, araştırmanın kapsamına ve değerlendirme amacına göre belirlenmelidir. Seçilen kitapların eser adı sayfası, verso ve kapak görüntüleri aynı koşullarda farklı üretken yapay zekâ modellerine sunularak bibliyografik bilgileri çıkarma becerileri karşılaştırılabilir. Bilgi profesyonelleri tarafından doğrulanmış kayıtlar referans alınarak MARC 21 alanlarının doğruluk oranları hesaplanabilir. Eser adı, sorumluluk bildirimi, yayıncı, yayın tarihi ve ISBN gibi alanlar

ayrı ayrı değerlendirilirken kaynakta bulunmayan bilgi üretimi, eksik alan oranı ve bilgi profesyonelinin düzeltme süresi de ölçülebilir. Böylece yalnızca otomatik veri üretiminin başarısı değil, insan denetimi sonrasındaki toplam işlem verimliliği de değerlendirilebilir.

Konu analizi ve kontrollü terim önerisi, araştırılması gereken ikinci alanı oluşturmaktadır. Türkçe eser özetleri veya içindekiler bilgileri kullanılarak modellerin önerdiği konu terimleri, bilgi profesyonellerinin seçtiği kontrollü terimlerle karşılaştırılabilir. Farklı konu alanlarından seçilen eserler üzerinde uzman değerlendirmesi yapılarak terimlerin anlamsal uygunluğu, yetke kayıtlarıyla eşleşmesi ve konu kapsamını temsil etme düzeyi incelenebilir. Ayrıca uzmanlar arasındaki değerlendirme farklılıklarının belirlenmesi, karşılaştırmada kullanılan referans verilerin güvenilirliğini değerlendirmeye katkı sağlayabilir.

Türkçe bibliyografik verilerin işlenmesinde dilin yapısal özellikleri ayrıca dikkate alınmalıdır. Türkçenin eklemeli yapısı, aynı kavramın farklı çekim ve türetim biçimleriyle ifade edilmesine neden olabilir. Örneğin “sınıflandırma”, “sınıflandırılması” ve “sınıflandırmalar” biçimleri arasında anlamsal ilişki bulunmasına rağmen sözcük düzeyinde farklılıklar vardır. Bunun yanında eş anlamlı terimler, alanlara özgü kısaltmalar, birleşik ifadeler ve tarihsel yazım farklılıkları konu analizini güçleştirebilir. Bu nedenle araştırmalarda modellerin yalnızca genel doğruluk düzeyleri değil, farklı dilsel özellikler içeren eser adları ve konu ifadeleri üzerindeki performansları da incelenmelidir. Yapay zekâ sistemlerinde dilsel ve kültürel çeşitliliğin gözetilmesi, uluslararası etik ilkeler açısından da önem taşımaktadır (UNESCO, 2021).

DOS numarası tahmini, üretken yapay zekâ bağlamında geliştirilmesi gereken bir diğer araştırma alanıdır. Büyük dil modellerinin DOS numarası atama performansını ele alan araştırmalar, sınıflandırma doğruluğunun ayrıntılı notasyon

düzeyinde ayrıca değerlendirilmesinin önemine işaret etmektedir (Reshi et al., 2026). Türkiye'de gerçekleştirilecek çalışmalarda aynı test kümesi üzerinde geleneksel makine öğrenmesi yöntemleri ile büyük dil modelleri karşılaştırılabilir. Yalnızca eser adının kullanıldığı koşul ile özet ve içindekiler gibi ek bilgilerin sağlandığı koşul ayrı ayrı test edilerek ana sınıf doğruluğu, ayrıntılı notasyon doğruluğu ve geçersiz numara üretme oranı ölçülebilir. Böylece farklı yöntemlerin hangi koşullarda daha başarılı olduğu ve ek bibliyografik bilgilerin sınıflandırma performansını nasıl etkilediği değerlendirilebilir.

İnsan–yapay zekâ iş birliğinin kalite ve verimlilik üzerindeki etkisi, kontrollü pilot çalışmalarla araştırılabilir. Benzer güçlük düzeyindeki bibliyografik kayıtlar, yalnızca bilgi profesyoneli, yalnızca yapay zekâ ve yapay zekâ destekli bilgi profesyoneli koşullarında işlenerek karşılaştırılabilir. İşlem süresi, alan doğruluğu, düzeltme sayısı, sınıflandırma tutarlılığı ve bilgi profesyonelinin algıladığı iş yükü birlikte ölçülebilir. Katılımcıların deneyim düzeyi ve materyal türü gibi değişkenlerin dikkate alınması, sonuçların daha sağlıklı yorumlanmasına yardımcı olabilir. Yapay zekâ sistemlerinde insan denetiminin ve kullanım bağlamına uygun risk değerlendirmesinin önemi, risk yönetimine ilişkin çalışmalarda da ele alınmaktadır (Autio et al., 2024).

Mesleki yetkinlikler ve kurumsal hazırlık düzeyi, gelecekteki araştırmaların bir başka boyutunu oluşturmaktadır. Üretken yapay zekâ sistemlerinin sunduğu bibliyografik önerilerin doğrulanabilmesi, kataloglama standartları, yetke denetimi ve sınıflandırma ilkelerine ilişkin mesleki bilgi gerektirmektedir. Bu nedenle üniversite kütüphanelerinde çalışan bilgi profesyonellerinin yapay zekâ araçlarını kullanma deneyimleri, kataloglama bilgi düzeyleri, güven algıları ve eğitim gereksinimleri anketler, görüşmeler ve uygulamalı değerlendirmelerle araştırılabilir. Kurumların veri güvenliği, onaylı araç kullanımı, sorumluluk

dağılımı ve kalite kontrol uygulamaları ise belge analiziyle karşılaştırılabilir. Elde edilecek bulgular, hizmet içi eğitim programlarının geliştirilmesinde ve kurumsal yapay zekâ kullanım politikalarının oluşturulmasında değerlendirilebilir. Üniversite kütüphanelerinde üretken yapay zekâ kullanımına yönelik etik ve kurumsal politika gereksinimleri, Türkiye'deki mesleki rehberlerde de ele alınmaktadır (ANKOS Yapay Zekâ Araştırma Grubu, 2025).

Kullanıcı perspektifi, teknik hizmet araştırmalarının sonuçlarını değerlendirmede ayrıca dikkate alınmalıdır. Bibliyografik kayıtların doğruluğu ve konu erişim noktalarının uygunluğu, kullanıcıların aradıkları kaynaklara ulaşmasını etkileyebilir. Bu nedenle yapay zekâ ile zenginleştirilmiş kayıtların kullanıldığı katalog arayüzleri ile mevcut kayıtların kullanıldığı arayüzler, aynı bilgi arama görevleri üzerinden karşılaştırılabilir. Kullanıcıların ilgili kaynağa ulaşma başarısı, arama süresi, sonuçların konu bakımından uygunluğu ve memnuniyet düzeyi ölçülebilir. Farklı kullanıcı gruplarının deneyimlerinin incelenmesi, yapay zekâ uygulamalarının yalnızca kataloglama sürecine değil, bilgiye erişim kalitesine etkisini de ortaya koyabilir. Kütüphanelerde yapay zekâ kullanımına yönelik uluslararası mesleki yaklaşımlar, erişim eşitliği ve kullanıcı gereksinimlerinin dikkate alınmasını öne çıkarmaktadır (IFLA, 2025).

Türkiye'de ortak kataloglar ve paylaşımlı bibliyografik veri ortamları, yapay zekâ destekli kalite kontrol uygulamaları açısından önemli araştırma olanakları sunmaktadır. Büyük bibliyografik veri kümelerinde yinelenen kayıtların, eksik erişim noktalarının, farklı yetke biçimlerinin ve sınıflandırma tutarsızlıklarının belirlenmesi amacıyla yapay zekâ destekli tarama yöntemleri geliştirilebilir. Bu yöntemlerin başarısı, uzmanlar tarafından önceden belirlenmiş hatalı ve doğru kayıt örnekleri üzerinde test edilerek değerlendirilebilir. Yanlış uyarı oranı ve gerçekten sorunlu kayıtları belirleme başarısı birlikte incelenmelidir. Modelin kayıtları doğrudan değiştirmesi

yerine uzman incelemesi gerektiren kayıtları önceliklendirmesi, daha kontrollü bir uygulama yaklaşımı sağlayabilir.

Maliyet ve ölçeklenebilirlik, araştırmaların uygulamaya aktarılmasında göz önünde bulundurulması gereken diğer unsurlardır. Bir yapay zekâ sisteminin ekonomik açıdan uygunluğu yalnızca abonelik veya kullanım ücretleriyle belirlenemez. Teknik altyapı, sistem entegrasyonu, personel eğitimi, insan doğrulaması, bakım ve olası hata düzeltme maliyetleri de toplam maliyete dâhil edilmelidir. Pilot uygulamalarda farklı büyüklükteki kayıt kümeleri kullanılarak kayıt başına toplam işlem maliyeti, insan emeği ve işlem kapasitesindeki değişim karşılaştırılabilir. Ayrıca veri miktarı arttığında doğruluk ve işlem süresinin nasıl değiştiği incelenmelidir. Yapay zekâ destekli metadata iş akışlarında kaynak kullanımı, sürdürülebilirlik ve insan denetiminin birlikte değerlendirilmesi gerektiği de vurgulanmaktadır (Program for Cooperative Cataloging, 2025).

Türkiye'de kataloglama uygulamaları, mesleki eğitim ve otomatik sınıflandırma üzerine gerçekleştirilen araştırmalar, üretken yapay zekâ destekli kataloglama ve sınıflandırma çalışmalarının geliştirilebileceği alanların belirlenmesi açısından önemli bir temel sunmaktadır. Bu doğrultuda Türkçe bibliyografik verilerden metadata oluşturulması, kontrollü konu erişimi, sınıflandırma performansı, insan–yapay zekâ iş birliği, mesleki yetkinlikler ve kurumsal hazırlık düzeyi gelecekteki araştırmalar için öncelikli konular olarak değerlendirilebilir. Bununla birlikte Türkiye'de üretken yapay zekâ destekli kataloglamaya ilişkin çalışmaların bulunmadığı yönünde kesin bir değerlendirme yapılabilmesi, kapsamlı ve güncel bir literatür taramasını gerektirmektedir.

Sonuç olarak Türkiye'de üretken yapay zekâ destekli kataloglama ve sınıflandırma araştırmalarının yalnızca uluslararası modellerin Türkçe veriler üzerinde denenmesiyle sınırlı kalmaması önemlidir. Türkçenin dilsel özellikleri, yerel konu terminolojisi,

kataloglama standartları, kullanıcı gereksinimleri, mesleki iş akışları ve kurumsal maliyetler araştırma tasarımlarına dâhil edilmelidir. Akademisyenler ile kataloglama uzmanlarının birlikte yürüteceği, gerçek bibliyografik veriler ve açık değerlendirme ölçütleri kullanan çalışmalar hem Türkiye'deki uygulamaların geliştirilmesine hem de üretken yapay zekâ literatürüne özgün katkılar sunabilir.

Sonuç ve Değerlendirme

Üretken yapay zekâ, kütüphanelerde kataloglama ve sınıflandırma süreçlerini dönüştürme potansiyeline sahip olmakla birlikte, bu dönüşüm yalnızca insan emeğinin otomasyonla değiştirilmesi üzerinden değerlendirilmemelidir. Bu bölümde incelenen literatür ve uygulama örnekleri, büyük dil modellerinin bibliyografik veri çıkarma, metadata oluşturma ve zenginleştirme, özet ve anahtar kelime üretme, konu terimi önerme, indeksleme kararlarını değerlendirme ve sınıflandırma numarası tahmini gibi farklı teknik hizmet görevlerinde kullanılmaya başlandığını göstermektedir (Roncaglia, 2026; Tian, 2026; Reshi et al., 2026; Włodarczyk & Zdunek, 2026). Alma AI Metadata Assistant gibi mevcut kütüphane sistemlerine entegre edilen uygulamalar da üretken yapay zekânın yalnızca deneysel araştırmalarla sınırlı kalmadığını, kataloglama iş akışlarına yardımcı bir araç olarak dâhil edilmeye başlandığını ortaya koymaktadır (Ex Libris, t.y.). Bununla birlikte bu gelişmeler, üretken yapay zekânın kataloglama ve sınıflandırma süreçlerinin tamamını bağımsız ve güvenilir biçimde yürütebildiği anlamına gelmemektedir.

Üretken yapay zekânın teknik hizmetlerdeki kullanım olanakları, gerçekleştirilen görevin niteliğine göre farklılaşmaktadır. Kaynak üzerinde açıkça bulunan bibliyografik bilgilerin çıkarılması ile eserin temel konusunun belirlenmesi, kontrollü konu terimlerinin seçilmesi veya ayrıntılı sınıflandırma numarasının oluşturulması aynı düzeyde anlamsal değerlendirme ve uzmanlık

gerektirmemektedir. Yapılandırılmış metadata görevleri ile anlamsal yorumlama gerektiren görevlerde farklı performans düzeylerinin gözlenmesi, otomasyonun bütün kataloglama işlemlerine aynı biçimde uygulanamayacağını göstermektedir (Tian, 2026). Benzer şekilde kontrollü konu indeksleme ve Dewey Onlu Sınıflama üzerine gerçekleştirilen araştırmalar, önerilen terimlerin ve sınıflandırma numaralarının doğruluğunun ilgili standartlar ve konu bağlamı içerisinde ayrıca değerlendirilmesi gerektiğine işaret etmektedir (Golub et al., 2026; Reshi et al., 2026). Bu nedenle üretken yapay zekânın kullanım düzeyi, işlemin doğrulanabilirliği, konusal karmaşıklığı ve olası hataların sonuçları dikkate alınarak belirlenmelidir.

Üretken yapay zekânın başlıca sınırlılıklarından biri, doğru görünmek ile doğru olmak arasındaki ayrımdır. Dilsel ve biçimsel açıdan tutarlı görünen bir bibliyografik bilginin doğruluğu, yalnızca sunuluş biçiminden hareketle kabul edilemez. Büyük dil modellerinin gerçekte bulunmayan bibliyografik referanslar üretebildiği deneysel araştırmalarda gösterilmiştir (Walters & Wilder, 2023). Bununla birlikte akademik referans üretiminde belirlenen bu hataların kütüphane kataloglama sistemlerinde aynı sıklıkta ortaya çıktığı varsayılamaz. Kataloglama açısından temel gereklilik, kaynakta açıkça bulunan bilgilerin çıkarılması ile modelin eksik bilgileri üretmesi veya yorumlamasının birbirinden ayrılması ve önerilen bibliyografik alanların güvenilir kaynaklarla doğrulanmasıdır. Konu analizi ve sınıflandırmada ise önerilerin anlamsal açıdan uygun görünmesi yeterli değildir; kontrollü sözlüklerin yetkili terimleri, sınıflandırma sistemlerinin hiyerarşik yapısı ve uygulama talimatlarıyla uyumun da sağlanması gerekmektedir. Bu bakımdan bibliyografik doğruluk, standartlara uygunluk ve katalog içi tutarlılık, üretken yapay zekâ kullanımında birbirini tamamlayan kalite ölçütleridir.

Üretken yapay zekâ destekli kataloglama uygulamalarının değerlendirilmesi, teknik doğruluk ve işlem verimliliğiyle sınırlı tutulmamalıdır. Bibliyografik kayıtların ve konu erişim noktalarının oluşturulmasında ortaya çıkabilecek önyargılar, farklı dil ve kültürlerin temsili, veri gizliliği, şeffaflık ve kurumsal hesap verebilirlik de dikkate alınmalıdır. Kataloglama etiği üzerine gerçekleştirilen çalışmalar, yapay zekâ teknolojilerinin metadata üretiminde çeşitli olanaklar sunarken doğruluk, önyargı, gizlilik ve şeffaflık bakımından yeni değerlendirme gereksinimleri oluşturduğunu ortaya koymaktadır (Dover & Grzegorski, 2025; Sussmeier & Henry, 2025). Yapay zekâ sistemlerinin kullanım bağlamına özgü risklerinin belirlenmesi, değerlendirilmesi ve izlenmesi ise güvenilir yapay zekâ yönetişiminin temel unsurları arasında yer almaktadır (NIST, 2023). Bu ilkeler kataloglama ve sınıflandırma süreçlerine uyarlandığında, bir sistemin yalnızca doğru kayıt üretip üretmediğinin değil, bilgi kaynaklarını adil biçimde temsil edip etmediğinin, hangi verileri işlediğinin ve ortaya çıkan hatalardan kimin sorumlu olduğunun da değerlendirilmesi gerekmektedir.

Bu bölümde benimsenen temel yaklaşım, üretken yapay zekânın kataloglama ve sınıflandırmada bağımsız bir karar verici olarak değil, bilgi profesyonelinin çalışmalarını destekleyen ve uzman denetimi altında kullanılan bir karar destek aracı olarak konumlandırılmasıdır. Bu doğrultuda önerilen beş aşamalı insan–yapay zekâ iş birliği iş akışı; kaynak verisinin hazırlanması, yapay zekâ önerilerinin oluşturulması, standart ve kaynak doğrulaması, uzman değerlendirmesi ve kalite izleme süreçlerini bir araya getirmektedir. İş akışının temel amacı, insan denetimini yalnızca son aşamada gerçekleştirilen biçimsel bir onay işlemi olmaktan çıkararak kataloglama sürecinin farklı aşamalarına yerleştirmektir. Bununla birlikte iş akışında insanın bulunması, denetimin her durumda etkili biçimde gerçekleşeceğini garanti etmemektedir.

Otomatik sistemlerin önerilerine aşırı güvenilmesinin karar hatalarına yol açabildiği, üretken yapay zekâ ve kataloglama bağlamı dışında gerçekleştirilen araştırmalarda gösterilmiştir (Skitka et al., 1999). Bu nedenle bilgi profesyonelinin önerileri değerlendirebilecek mesleki bilgiye, yeterli zamana ve güvenilir doğrulama kaynaklarına sahip olması, insan denetiminin niteliği açısından önem taşımaktadır.

Bu süreçte yüksek kaliteli bibliyografik kayıtlar, yetke dosyaları, kontrollü sözlükler ve sınıflandırma şemaları önemini kaybetmek yerine yapay zekâ çıktılarının doğrulanması ve değerlendirilmesinde yeni bir işlev kazanabilir. Bu kaynaklar, üretken yapay zekâ tarafından önerilen bilgilerin doğruluğunu ve standartlara uygunluğunu incelemek için güvenilir başvuru noktaları oluşturmaktadır. Dolayısıyla yapay zekâ destekli kataloglama sistemlerinin geliştirilmesi, mevcut bilgi düzenleme altyapısının korunması ve iyileştirilmesi gereksinimini ortadan kaldırmamaktadır. Bununla birlikte söz konusu kaynakların da güncel, tutarlı ve güvenilir olması gerekmektedir. Hatalı veya eksik bibliyografik verilerin yapay zekâ destekli süreçlerde kullanılması, mevcut sorunların yeniden üretilmesine veya daha geniş kayıt kümelerine yayılmasına neden olabilir. Bu nedenle üretken yapay zekâ uygulamalarının başarısı yalnızca kullanılan modelin özelliklerine değil, yararlanılan bibliyografik veri altyapısının niteliğine de bağlıdır.

İnsan denetiminin sürdürülebilirliği, bireysel uzmanlığın yanı sıra kurumsal politika ve kalite güvence mekanizmalarının oluşturulmasını da gerektirmektedir. Bu bölümde geliştirilen örnek teknik hizmetler politikası; görev ve sorumlulukların belirlenmesi, veri güvenliği, bibliyografik doğrulama, izlenebilirlik, hata bildirimi ve risk temelli denetim gibi konuları birlikte ele almaktadır. Ayrıca yeni gelen kitapların kataloglanması, eksik kayıtların zenginleştirilmesi, retrospektif kataloglama, dijital koleksiyonlarda

metadata oluşturulması, konu analizi ve sınıflandırma desteği ile bibliyografik kayıtların kalite kontrolü için sunulan uygulama senaryoları, insan denetiminin farklı görevlerde nasıl yapılandırılabileceğini somutlaştırmaktadır. Bu politika hükümleri ve senaryolar, bütün kütüphanelerde aynı biçimde uygulanması gereken kesin kurallar veya deneysel olarak doğrulanmış modeller değil; literatür ve mevcut uygulamalardan hareketle bu çalışma kapsamında geliştirilmiş önerilerdir. Kurumların koleksiyon özellikleri, teknik altyapıları, personel kapasiteleri ve kabul edilebilir risk düzeyleri doğrultusunda uyarlanmaları gerekmektedir.

Yapay zekâ destekli kataloglama sistemlerinin başarısının değerlendirilebilmesi için bibliyografik kalite ile iş süreçlerine ilişkin sonuçların birlikte ölçülmesi önem taşımaktadır. Bu bölümde önerilen yedi boyutlu değerlendirme çerçevesi, sistem performansının yalnızca genel doğruluk veya işlem hızı üzerinden değerlendirilmesinin yeterli olmayacağı düşüncesine dayanmaktadır. Bibliyografik alanların doğruluğu, standartlara uygunluk, konu ve sınıflandırma kararlarının niteliği, çıktıların tutarlılığı, verimlilik, farklı kaynak gruplarındaki performans ve mesleki kullanılabilirlik gibi ölçütlerin birlikte incelenmesi, yapay zekâ destekli iş akışlarının güçlü ve zayıf yönlerinin daha kapsamlı biçimde belirlenmesine katkı sağlayabilir. Özellikle önerilerin kabul edilme oranı veya kayıt oluşturma süresindeki azalma, nihai bibliyografik doğruluğun tek başına göstergesi olarak değerlendirilmemelidir. Doğrulama ve düzeltme için harcanan zaman, kritik hata türleri ve kullanıcıların bilgiye erişimine olası etkiler de değerlendirmeye dâhil edilmelidir. Böylece sistemlerin kullanımına ilişkin kararlar yalnızca teknolojik olanaklara değil, ölçülebilir kalite göstergelerine ve kurumsal gereksinimlere dayandırılabilir.

Üretken yapay zekânın teknik hizmetlerde kullanımı, bilgi profesyonellerinin mesleki yetkinliklerinin yeniden

değerlendirilmesini de gerektirmektedir. Kataloglama standartları, yetke denetimi, konu analizi ve sınıflandırma bilgisi önemini korurken veri okuryazarlığı, yapay zekâ okuryazarlığı, istem tasarımı, çıktı değerlendirme ve hata analizi gibi yeterlilikler bu uzmanlığı tamamlamaktadır. Akademik kütüphane çalışanlarına yönelik yapay zekâ yetkinlikleri çerçevesinde etik değerlendirme, bilgi ve anlama, analiz ve değerlendirme ile kullanım ve uygulama boyutlarının ele alınması, teknolojinin mesleki kullanımında eleştirel değerlendirme becerilerinin önemini göstermektedir (ACRL, 2025). Bu bağlamda bilgi profesyonelinin rolü, yalnızca bibliyografik kayıt oluşturan bir uygulayıcı olmaktan çıkarak otomatik önerileri değerlendiren, metadata kalitesini yöneten, kurumsal standartlara uygunluğu denetleyen ve teknolojinin bilgiye erişim üzerindeki etkilerini sorgulayan bir uzmanlık alanına doğru genişleyebilir. Bu dönüşüm, kataloglama ve sınıflandırma bilgisini daha az değil, farklı biçimde gerekli kılmaktadır. Bir modelin hatasını belirleyebilmek için doğru kataloglama uygulamasının ne olduğunu bilmek; otomasyon süreçlerini yapılandırabilmek için hangi işlemlerin rutin ve doğrudan doğrulanabilir, hangilerinin bağlamsal değerlendirme ve uzman muhakemesi gerektirdiğini ayırt edebilmek önem taşımaktadır. Bu nedenle mesleki uzmanlık, yalnızca bibliyografik kayıtların oluşturulmasında değil, yapay zekâ destekli iş akışlarının tasarlanması, denetlenmesi ve değerlendirilmesinde de belirleyici bir rol üstlenmektedir.

Türkiye açısından üretken yapay zekâ destekli kataloglama ve sınıflandırma, yerel bibliyografik veri yapıları, Türkçenin dilsel özellikleri ve kurumsal uygulamalar doğrultusunda geliştirilmesi gereken bir araştırma alanı sunmaktadır. Türkiye'deki kataloglama uygulamaları, mesleki eğitim ve otomatik konu sınıflandırması üzerine gerçekleştirilen çalışmalar, gelecekteki araştırmalar için bir başlangıç noktası oluşturmaktadır (Çakmak, 2019; Bayter, 2023; Akyüz, 2026). Bununla birlikte geleneksel makine öğrenmesine

dayalı sınıflandırma araştırmaları ile üretken yapay zekâ uygulamalarının birbirinden ayrılması gerekmektedir. Gelecekte Türkçe bibliyografik verilerden metadata oluşturulması, kontrollü konu terimi önerileri, ayrıntılı Dewey numarası tahmini, insan–yapay zekâ iş birliğinin kalite ve verimlilik üzerindeki etkileri, kurumsal hazırlık düzeyi ve kullanıcıların bilgiye erişim deneyimleri karşılaştırmalı çalışmalarla incelenebilir. Bu araştırmalarda yalnızca model performansına değil, kullanılan verilerin niteliğine, mesleki doğrulama gereksinimlerine, maliyetlere ve sonuçların kütüphane iş akışlarına aktarılabilirliğine de odaklanılması önem taşımaktadır. Türkiye'de üretken yapay zekâ destekli kataloglama araştırmalarının mevcut kapsamına ilişkin kesin değerlendirmeler ise sistematik ve güncel literatür incelemelerine dayandırılmalıdır.

Sonuç olarak üretken yapay zekâ, kataloglama ve sınıflandırma faaliyetlerini ortadan kaldıran bir teknoloji olarak değil, bu faaliyetlerde kullanılan araçları, iş akışlarını ve mesleki sorumlulukların uygulanma biçimini dönüştürebilecek bir gelişme olarak değerlendirilmelidir. Bu dönüşümün güvenilir ve sürdürülebilir olması; bibliyografik standartların korunmasına, kontrollü bilgi düzenleme araçlarının kullanılmasına, kurumsal risk yönetimine, ölçülebilir kalite değerlendirmelerine ve etkili insan denetimine bağlıdır. Üretken yapay zekânın hız ve içerik üretme kapasitesi, bilgi profesyonelinin bağlamsal değerlendirme yetkinliği ve mesleki sorumluluğuyla birlikte ele alındığında teknik hizmetlerin geliştirilmesine katkı sağlayabilir. Bu çerçevede kataloglama ve sınıflandırmanın geleceği, insan ile yapay zekâ arasında bir yer değiştirme ilişkisinden çok, görevlerin niteliğine ve risk düzeyine göre düzenlenen, uzman denetimine dayalı bir iş birliği modeli üzerinden şekillenebilir. Bu modelde yapay zekâ önerir ve destekler; bilgi profesyoneli doğrular, bağlama yerleştirir ve nihai sorumluluğu üstlenir.

Kaynakça/References

Adalı, E. (2020). *Türkçe Doğal Dil İşleme.* Akçağ.

Akyüz, H. (2026). *Makine öğrenme ile kütüphanelerde konu sınıflandırma çalışması (Tez No: 1025463) [Doktora tezi]. Ankara Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.* YÖK Ulusal Tez Merkezi.

Alwreikat, A., Musleh, W., Momani, H., Alshibly, H., & Al-Shboul, M. (2026). Transforming Arabic Bibliographic Records: Implementing ChatGPT in MARC 21 Record Creation. *Cataloging & Classification Quarterly, 64*(1–2), 28–49. https://doi.org/10.1080/01639374.2026.2619172

American Library Association. (2026). *Guidance on the use of artificial intelligence in libraries*. https://www.ala.org/tools/standards-and-guidelines/guidance-use-artificial-intelligence-libraries

American Library Association Core. (2023). *Revised competencies for cataloging & metadata professional librarians*. https://alair.ala.org/server/api/core/bitstreams/4a980621-0f7b-4f8a-bfed-d0c8775866fb/content

ANKOS Yapay Zekâ Araştırma Grubu. (2025). *Üniversite kütüphaneleri için üretken yapay zekâ kullanım rehberi ve politikası.* https://doi.org/10.5281/zenodo.15104933

Association of College and Research Libraries. (2025). *AI competencies for academic library workers*. https://www.ala.org/sites/default/files/2025-10/acrl_ai_competencies.pdf

Autio, C., Schwartz, R., Dunietz, J., Jain, S., Stanley, M., Tabassi, E., Hall, P., & Roberts, K. (2024). *Artificial intelligence risk management framework: Generative artificial intelligence*

profile (NIST AI 600-1). https://doi.org/10.6028/NIST.AI.600-1

Bagchi, M. (2025). Toward Generative AI–Driven Metadata Modeling: A Human–Large Language Model Collaborative Approach. *Library Trends, 73*, 297–322.

Bayter, M. (2023). Türkiye'de Kataloglama Eğitimi. *Külliye, 4*(2), 59–80. https://doi.org/10.48139/aybukulliye.1267015

Beygirci, B. (2022). Bilginin Düzenlenmesinde Kataloglama Etiği: İlkeler ve Sorunlar. *Bilgi Yönetimi, 5*(2), 322-339. https://doi.org/10.33721/by.1160843

Çakmak, T. (2018). "RDA Transition in Turkey: Perspectives and Experiences of the Libraries". *JLIS, 9*(1). https://doi.org/10.4403/jlis.it-12438

Çakmak, T. (2019). Cataloguing practices at university libraries: Analysis of current conditions and practices in Turkey. *The Electronic Library, 37*(1), 155-172. https://doi.org/10.1108/EL-06-2018-0122

Demir, G. (2025). Artificial Intelligence (AI) and Robotic Technologies Implementation in Libraries: An Integrative Literature Review. *Bilgi Yönetimi, 8*(1), 77-98. https://doi.org/10.33721/by.1645554

Dover, A., & Grzegorski, J. (2025). Artificial Intelligence Through the Lens of the Cataloguing Code of Ethics. *Cataloging & Classification Quarterly, 63*(6–7), 600–620. https://doi.org/10.1080/01639374.2025.2544137

Ex Libris. (2025). *Alma 2025 release notes: May 2025 release*. https://knowledge.ag-software.clarivate.com/alma/en-us/Release_Notes/2025/May_2025.html

Ex Libris. (tarih yok). *The AI Metadata Assistant in the Metadata Editor*. https://knowledge.ag-software.clarivate.com/alma/en-us/Product_Documentation/Metadata_Management/The_AI_Metadata_Assistant_in_the_Metadata_Editor.html

Flinchbaugh, M., Thomas, C., Tench, R., Sipe, V., Moskal, R., Aldana, L., & Owusu, E. (2019). *Transforming Acquisitions and Collection Services.* https://doi.org/10.2307/j.ctv2x00vjh

Gamage, R., & Wanigasooriya, P. (2024). Using Generative AI for Bibliographic Description: A Study with ChatGPT 4. *Journal of the University Librarians Association of Sri Lanka*.

Golub, K., Falk, O., & Humlesjö, S. (2026). Beyond the gold standard: evaluating AI for subject indexing of Swedish LGBTQ + fiction. *Journal of Documentation, 82*(7), 526–543. https://doi.org/10.1108/JD-06-2026-0363

Heng, G., Lampron, P., & Han, M. (2026). Generative AI Meets Cataloging Practice: Findings from a Comparative Pilot Study. *Information Technology and Libraries, 45*(2). https://doi.org/10.5860/ital.v45i2.17499

Hjørland, B. (2018). Indexing: Concepts and theory. *Knowledge Organization, 45*(7), 609–639.

International Federation of Library Associations and Institutions. (2012). *IFLA code of ethics for librarians and other information workers*. https://www.ifla.org/g/faife/professional-codes-of-ethics-for-librarians/

International Federation of Library Associations and Institutions. (2020). *IFLA statement on libraries and artificial*

intelligence. https://repository.ifla.org/handle/20.500.14598/1646

International Federation of Library Associations and Institutions. (2025). *IFLA entry point to libraries and AI*. https://www.ifla.org/news/release-of-ifla-entry-point-to-libraries-and-ai-document/

International Federation of Library Associations and Institutions. (2026). *Libraries – Cataloguers*. https://www.ifla.org/g/bibliography/libraries-cataloguers/

International Federation of Library Associations and Institutions. (tarih yok). *Glossary of terms and abbreviations and useful links (O–R)*. https://www.ifla.org/g/bibliography/glossary-of-terms-and-abbreviations-and-useful-links-o-r/

Kasprzik, A. (2024). The automation of subject indexing at ZBW and the role of metadata in times of large language models. *Procedia Computer Science, 249*, 160-166. https://doi.org/10.1016/j.procs.2024.11.059

Kayalı, N. (2026). Üretken Yapay Zekâ Çağında Bilginin Dönüşümü: Bilgi Yönetimi Perspektifinden Kavramsal Bir Tartışma. *Bilgi Yönetimi, 9*(1), 1-18. https://doi.org/10.33721/by.1914407

Lancaster, F. (2003). *Indexing and abstracting in theory and practice.* Facet Publishing.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. https://doi.org/10.48550/arXiv.2005.11401

Lo, L. (2023). The CLEAR path: A framework for enhancing information literacy through prompt engineering. *The Journal of Academic Librarianship, 49*. https://doi.org/10.1016/j.acalib.2023.102720

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (s. 1–16). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376727

Lowagie, H. (2025). AI-powered cataloguing: A practical guide to building a cataloguing application with Power Apps. *Canadian Journal of Academic Librarianship, 12*, 1-3. https://www.researchgate.net/publication/410714412_AI-Powered_Cataloguing_A_Practical_Guide_to_Building_a_Cataloguing_Application_with_Power_Apps_by_Hannes_Lowagie

Luo, L. (2025). Use of Generative AI in Aiding Daily Professional Tasks: A Survey of Librarians' Experiences. *Library Trends, 73*(3), 381–398. https://dx.doi.org/10.1353/lib.2025.a961200

Mandinach, E., & Gummer, E. (2013). A systemic view of implementing data literacy in educator preparation. *Educational Researcher, 42*(1), 30–37. https://doi.org/10.3102/0013189X12459803

Mugaanyi, J., Cai, L., Cheng, S., Lu, C., & Huang, J. (2024). Evaluation of large language model performance and reliability for citations and references in scholarly writing: Cross-disciplinary study. *Journal of Medical Internet Research, 26*. https://doi.org/10.2196/52935

National Institute of Standards and Technology. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST AI 100-1)*. https://doi.org/10.6028/NIST.AI.100-1

OCLC. (2010). *WorldCat rights and responsibilities for the OCLC cooperative*. https://www.oclc.org/en/worldcat/cooperative-quality/policy.html

Oladokun, B., Irivike, F., Enakrire, R., & Ajani, Y. (2025). Use and application of ChatGPT-4o by cataloguers and bibliographers from Botswana, Nigeria, and South Africa. *Journal of Library Metadata, 25*(4), 277–289. https://doi.org/10.1080/19386389.2025.2546702

Oladokun, B., Ogunjimi, B., Olatunbosun, I., Adefila, E., Abdul, A., Ebhonu, S., Omoniyi, Y., & Enakrire, R. (2026). Assessing metadata quality: Analysis of bibliographic entries in librarianship literature generated by ChatGPT-5. *Journal of Library Metadata, 26*(1), 51–83. https://doi.org/10.1080/19386389.2025.2598503

Program for Cooperative Cataloging. (2025). *PCC task group on AI and machine learning in cataloging and metadata: Final report*. https://www.loc.gov/aba/pcc/taskgroup/AI-and-Machine-Learning-TG-final-report.pdf

Reshi, A., Lone, Y., Ayub, T., & Gul, S. (2026). Applicability of large language models (LLMs) to library classification: Analysis of ChatGPT-4o, DeepSeek and Gemini 2.0 for assigning Dewey Decimal Classification (DDC) numbers. *Journal of Librarianship and Information Science*. https://doi.org/10.1177/09610006261442170

Roncaglia, G. (2026). Cataloguing, metadata, and generative AI: Early experiences and future perspectives. *JLIS.it, 17*(1), 106–127. https://doi.org/10.36253/jlis.it-693

Sivri, E. (2023). Kütüphanelerde yapay zekâ'nın geleceği: Farklı alanlardaki potansiyel uygulamalar ve yeni kullanım alanları oluşturma. *Kütüphane Arşiv ve Müze Araştırmaları Dergisi, 4*(2), 175–184. https://doi.org/10.59116/lamre.1299783

Skitka, L., Mosier, K., & Burdick, M. (1999). Does automation bias decision-making? *International Journal of Human-Computer Studies, 51*(5), 991–1006. https://doi.org/10.1006/ijhc.1999.0252

Spennemann, D. H. R. (2026). How unique are hallucinated citations offered by generative artificial intelligence models? Publications, 14(3), Article 38. https://doi.org/10.3390/publications14030038

Sussmeier, S., & Henry, J. (2025). Mind the gap!: How do we ethically bridge the divide between the cataloging/metadata community and the world of AI? *Journal of Library Metadata, 25*(4), 241–259. https://doi.org/10.1080/19386389.2025.2525720

Tan, Z., Yang, X., Chen, Q., & Chen, T. (2026). Research on large language model-based bibliographic cataloging agent in the CNMARC context. *Publications, 14*(1), Article 19. https://doi.org/10.3390/publications14010019

Taniguchi, S. (2025). Addressing and reducing complexity in metadata mappings: RDA to DC and MARC 21. *Cataloging & Classification Quarterly, 63*(4), 240–266. https://doi.org/10.1080/01639374.2025.2511804

Tian, T. (2026). Generative AI for born-digital collections: A case study of metadata management in an institutional repository. *Cataloging & Classification Quarterly, 64*(3/4), 123–147. https://doi.org/10.1080/01639374.2026.2661928

TÜBİTAK Ulusal Akademik Ağ ve Bilgi Merkezi. (2006). *TO-KAT: Ulusal Toplu Katalog*. https://veri.ulakbim.gov.tr/hizmet/14

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. https://www.unesco.org/en/legal-affairs/recommendation-ethics-artificial-intelligence

Ülkir, M., & Paslı, B. (2026). Reference hallucination, citation reliability, and readability of large language models in anatomy-related question answering. Clinical Anatomy. https://doi.org/10.1002/ca.70187

Walters, W., & Wilder, E. (2023). Fabrication and errors in the bibliographic citations generated by ChatGPT. *Scientific Reports, 13*. https://doi.org/10.1038/s41598-023-41032-5

Włodarczyk, B., & Zdunek, M. (2026). Large language models as judges for human subject indexing. *Cataloging & Classification Quarterly, 64*(5/6), 251–280. https://doi.org/10.1080/01639374.2026.2714375

Zencir, M., & Öztürk, T. (2022). Türkiye'de kütüphanecilik ve bilgibilim alanında araştırma-uygulama boşluğu: Makaleler kapsamında yazarlık ve iş birliğinin bir analizi. *Türk Kütüphaneciliği, 36*(4), 451–471. https://doi.org/10.24146/tk.1144164

BÖLÜM 5

YOLO MODELLERİ İLE FİŞ ANALİZİNDE SAHTECİLİK TESPİTİ

1. Tolga TUNÇEL[1]
2. Emre KILINÇ[2]

Giriş

Fiş sahteciliği, işletmeleri ve tüketicileri eşit koşullarda etkileyerek küresel ölçekte ekonomik olarak büyük bir endişe haline gelmiştir. Bu tür durumlarla karşılaşmak, hem ekonomik hem de itibar kaybına yol açmaktadır. Fiş sahteciliğinin tespitinde kullanılan geleneksel yöntemler, belirli bir dereceye kadar etkili olsa da genellikle zaman alıcı veya hata yapmaya açıktır. Bu bağlamda, fiş sahtekarlığının tespiti için etkili, verimli, yenilikçi ve otomatik bir sisteme ihtiyaç duyulmaktadır.

Bilgisayarlı görü ve derin öğrenme tabanlı algoritmaların oldukça gelişmesi, desen tanıma, sahtecilik ve anormallik gibi konuların tespitinde oldukça iyi sonuçlar vererek bu konulara yaklaşımı değiştirmiştir. Gerçek zamanlı nesne tespiti konusunda iyi

[1] Öğr. Gör., Ankara Üniversitesi Bilişim Teknolojileri Meslek Yüksekokulu, Büyük Veri Analistliği, Orcid: 0000-0002-7427-2071
[2] Dr. Öğr. Üyesi, Çukurova Üniversitesi, Yapay Zekâ ve Makine Öğrenmesi Bölümü, Orcid: 0000-0002-5250-9322

sonuçlar veren YOLO (You Only Look Once) algoritma ailesi de bu algoritmalar arasında sayılabilmektedir. YOLO ailesi, nesne tespiti konusunda önemli iyileştirmeler sunarak, gerçek zamanlı ve yüksek doğrulukta fiş sahteciliği tespiti yapabilme kapasitesine sahiptir. Bu çalışma, YOLO ailesinde seçilmiş YOLO algoritmaları ile fiş sahteciliğini tespit etme potansiyelini incelemeyi amaçlamaktadır.

Araştırma, ilk olarak, fiş sahteciliğini tespit edebilen mevcut yöntemlere dair bir literatür taraması ile başlar. Daha sonra, Roboflow platformunda yayımlanan Receipt Fraud veri seti üzerinden (projectORCA, 2024), çalışma kapsamında seçilmiş YOLO ailelerine ait algoritmaların eğitimi ve test edilmesi süreci detaylandırılır. Çalışmanın sonunda, seçilen algoritmaların tespit sonuçlarının performansı değerlendirilerek, fiş sahteciliğine yönelik öneriler sunulmaktadır.

Bu girişim, fiş sahteciliği tespitinin YOLO ailesi gibi gelişmiş yapay zekâ teknolojilerinin potansiyelini ortaya koymaktadır. Çalışmanın literatüre katacağı en büyük faydalardan birisi de nesne tespitindeki teorik gelişmeler ile sahtecilik önlemeyi birleştirerek aradaki boşluğu kapatmasıdır. Bu çalışma, aynı zamanda, yapay zekâ tabanlı güvenlik sistemlerinin geliştirilmesi ve uygulanması alanında gelecekte yapılacak araştırmalara temel oluşturmayı hedeflemektedir.

Literatür Taraması

Görüntü işleme tabanlı sahtecilik tespitinin önemli uygulama alanlarından biri gıda ürünleridir. Bu alanda görüntülerin dokusal ve spektral özelliklerinden yararlanan farklı yaklaşımlar geliştirilmiştir. Bu doğrultuda Jahanbakhshi vd. (2021), zencefil tozunda yapılan sahteciliği tespit etmek amacıyla görüntü işleme yöntemleri ve derin öğrenme tabanlı CNN modelleri kullanmışlardır. Önerdikleri sistemin zencefil tozunun yüzey dokusunu analiz ederek sahtecilik belirtilerini yüksek bir doğruluk ile otomatik bir şekilde yapılması

hedeflenmişlerdir. Önerdikleri modelin etkinliğini kanıtlamak için saf zencefil tozu ve saf nohut tozlarından %10, %20, %30, %40, %50 karışımlar yapılarak toplamda 7 farklı sınıfta veri seti oluşturmuşlardır. Oluşturdukları veri setini kuvvetlendirmek için döndürme, yansıtma ve renk manipülasyonu gibi veri artırma teknikleri kullanarak veri setini eğitime hazır etmişlerdir. Eğitim için oluşturulan modelin 8 konvolüsyon katmanı, 4 batch normalizasyon katmanı, 1 havuzlama katmanı ve 3 tam bağlantılı katmanı olduğu belirtilmiştir. CNN kapsamında ise gated pooling fonksiyonu kullanılarak, max pooling ve average pooling yöntemleri birleştirilmiş ve bununla birlikte performans artışı hedeflemişlerdir. Görüntüler 30x30, 40x40 ve 50x50 olarak parçalanarak aynı görüntülere fakat farklı piksel boyutlarına sahip 3 veri seti elde edilmiş ve yöntemin etkinliğini değerlendirmek için bu veri kümeleri üzerinden değerlendirme yapılmıştır. Özellikle 50x50 piksel görüntüsünden oluşan veriseti ile yapılan eğitim sonucunda %99,70 gibi harika bir sonuç elde edildiği, gated pooling öncesi önerilen yöntemin KNN'in arkasında kalırken gated pooling ile KNN'in önüne geçtiği de vurgulanmıştır. Önerilen sistemin zencefil tozunda hile tespitinde etkili bir yöntem olduğu savunulmaktadır.

Standart görüntülerden çıkarılan özelliklerin değerlendirilmesi, gıda sahteciliğinde geleneksel makine öğrenmesi yaklaşımlarının kullanımına da olanak sağlamaktadır. Bu doğrultuda Nargesi ve Kheiralipour (2024), kara ve kırmızı biberlerde deniz köpüğü karıştırılarak yapılan sahtekarlığın tespit edilmesi için görüntü işleme ve makine öğrenimi yöntemleri kullanılmış ve performans değerlendirilmesi yapılmıştır. Çalışmada yerel pazardan temin edilen kara biber ve kırmızı biberler toz haline getirilmiş ve deniz köpüğünü sahtecilik maddesi olarak kullanılmıştır. Deniz köpüğü karışımlara %0, %5, %15, %30 ve %50 oranlarında eklenmiş ve örnekler 25°C'de ışık almayan bir ortamda muhafaza edilmiştir. Veri setindeki görüntüler, bir aydınlatma kutusu içerisinde

15 cm mesafeden Xiaomi Note 10 Pro Max kullanarak toplamda 180 görüntü toplanmıştır. Kara biber ve kırmızı biber için ortalama, medyan, standart sapma, enerji, entropi, kontrast gibi özellik çıkarımları yapılmıştır. Toplamda 266 özellik belirlenmiş fakat 266 özellik içerisinden en etkili 17 özellik seçilmiştir. Sahtekarlık tespiti için 17 giriş, 13 gizli ve 5 çıkış nöronuna sahip bir ANN modeli ve SVM yöntemleri kullanılmış ve karşılaştırmalar yapılmıştır. Yapılan karşılaştırmalarda ANN için kara biberde doğruluk %98,22, kırmızı biberde ise %99,11 , SVM'de ise Siyah biberde doğruluk %97,77, kırmızı biberde %99,11 elde edilmiştir. Kara biberde yapılan sahtecilikte ANN daha iyi bir sonuç verirken, kırmızı biberde her iki yöntemde aynı sonucu verdiği görülmüştür. Biberlerdeki sahtecilik tespitinde görüntü işleme ve makine öğrenimi yöntemlerinin etkinliği savunulmaktadır.

Benzer biçimde, standart görüntüleri temel alan ancak farklı derin öğrenme mimarilerinin karşılaştırıldığı çalışmalar da bulunmaktadır. Besharati vd. (2024), İran siyah çayında yapılan sahteciliğin tespit edilmesi için derin öğrenme tabanlı bir yaklaşım önermişlerdir. Çalışma kapsamında oluşturulan veri setinde, Rasht şehrindeki bir fabrikadan temin edilmiş 2 kg Momtaz türü İran çayı kullanılmıştır. %100 gerçek çay, %85 gerçek çay + %15 atık, %55 gerçek çay + %45 atık, %100 atık olacak şekilde 4 sınıf belirlenmiş ve Canon EOS 4000D kamerayla laboratuvar ışıklandırması altında, sabit bir mesafeden görüntüler toplanmıştır. Toplanan veri setini genişletmek için 2 yöntem kullanılmıştır. İlk olarak, görüntüler küçük parçalara (440x440 piksel) boyutlarına bölünmüş ve bu bölüm patched olarak isimlendirilmiştir. Bir diğeri ise unpatched olarak isimlendirilen, tam görüntüler küçültülerek (660x440 piksel) boyutuna indirgenerek veri setine entegre edilmiştir. Sahtecilik tespitinde, ağ tasarımı ve parametre optimizasyonu ile hem yüksek doğruluk hem de düşük bellek kullanımı sağlayan, RegNet800MF, düşük bellek gereksinimi ve yüksek işlem hızı ile öne çıkan

MobileNet V3 sinir ağı modeli, daha hızlı eğitim ve daha yüksek doğruluk için derinlik, genişlik ve çözünürlüğü dengeli bir şekilde ölçekleyen EfficientNet V2, düşük hesaplama maliyeti ve kanal karıştırma mekanizması ile verimliliği artıran ShuffleNet V2 ve görsellerdeki yerel ve küresel ilişkileri etkili bir şekilde yakalamak için kayan pencere mekanizması kullanan Swin Transformer V2 olacak şekilde 5 farklı derin öğrenme modeli kullanılmış ve karşılaştırmalar yapılmıştır. Patched'da %95 ile en iyi sonucu MobileNet V3 verirken, unpatched'da %90.63 ile EfficinetNetV2S en iyi sonucu verdiği görülmüştür. Özellikle MobileNet V3'ün hem patched hem unpatchedda kısa sürelerde çalışarak çok iyi sonuçlar elde ettiği vurgulanmıştır. Derin öğrenme ve görüntü işleme yöntemlerinin sahte çayı tespit etmek için güvenilir bir çözüm sunduğunu savunulmuştur.

Gıda sahteciliğinin tespitinde standart görüntülere dayalı yöntemlerin yanında, spektral verilerin kullanıldığı yaklaşımlar da geliştirilmiştir. Bu kapsamda Zhang vd. (2024), hiperspektral görüntüleme ve tek sınıf öğrenme yöntemi kullanarak mısır tohumlarında yapılan sahteciliği ayırt etmeye odaklanmışlardır. Çalışmada tespit için 20 farklı çeşit mısır tohumundan 500'er adet tohum alınarak toplamda 10000 adet tohum elde edilmiş ve 382.2-1026.7 nm spektral aralığında çalışan bir hiperspektral görüntüleme cihazından analiz yapılmıştır. Elde edilen görüntüler siyah beyaz düzeltme işlemine tabi tutulmuştur. Gri tonlama görüntüleri 729.1 nm'de maskeleme için kullanılarak tohumlar ve arkaplanlar arasındaki kontrast artırılmıştır. Daha sonra ikili maskelemeyle belirlenen bir eşik değeri ile arkaplan bilgisi çıkarılarak yalnızca tohumların ayrıştırılması sağlanmıştır. Her bir tohumun ayrı ayrı analiz edilmesi için belirlenmiş bölgesi yani ROI tanımlanmış ve her bir tohumun spektrum değeri hesaplanmıştır. Çalışma kapsamında tanıtılan tek sınıf öğrenme modelinin, spektral ve mekânsal bilgi çıkarım modüllerini, bant dikkati modülünü ve hiperküre öğrenme

bileşenlerini birleştirdiği belirtilmiştir. Bant dikkat modülünün, spektral özellik çıkarım modülünden önce geldiği ve gereksiz bantlardan kaynaklanan gürültüyü azaltmak için kullanılmıştır. Tohumların kimyasal yapısını temsil eden spektral özellikleri çıkarmak için spektral özellik çıkarım modülünün kullanılmış ve mekânsal özellik çıkarım modülüyle tohumların fiziksel özelliklerini tanımlamak için mekânsal özellikleri elde edildiği belirtilmiştir. Gerçek tohumları içeren minimum hacimli bir hiperküre oluşturduğunu bununda test sırasında, hiperkürenin dışına düşen tohumların sahte olarak sınıflandırıldığı belirtilmiştir. Önerilen yöntemin performansını değerlendirmek için OCSVM, Isolation Forest, CAE, Deep SVDD, OCCNN gibi farklı yöntemler ile karşılaştırmalar yapılmış ve önerilen yöntemin, gerçek çeşitlerin %93,70'ini doğru bir şekilde tanımlarken, sahte çeşitlerin %94,28'ini doğru bir şekilde sınıflandırarak diğer yöntemlerin önüne geçtiği gözlemlenmiştir. Önerilen yöntemin tarım ve gıda sektörlerindeki sahtecilikle mücadele için kullanılabileceği savunulmaktadır. Hiperspektral görüntüleme ile sahtecilik tespitine yönelik bir diğer çalışmada Cheng vd. (2024), Manuka balında yapılan sahteciliği hiperspektral görüntüleme ve derin öğrenme temelli GANomaly modeli ile tespitine odaklanılmıştır. Çalışmada tespit için 5 farklı markadan 18 UMF dereceli saf Manuka balı örneğiyle birlikte şurup ile karıştırılmış ve uzun süre saklanmış bal örnekleri de test amacıyla kullanılmıştır. Bal örneklerinin 20–25 °C sıcaklıkta 6 yıl boyunca saklandığı ve bu süre zarfında kristalizasyon süreci geçirdiği belirtilmiştir. Her bir bal örneği, 399.40–1063.79 nm dalga boyu aralığında vis-NIR SOC710-VP hiperspektral görüntüleme cihazı ile taranmış ve toplanan hiperspektral görüntüler, beyaz referans kalibrasyonu yapılarak analiz edilmiş ayrıca her görüntüden, örneklenmiş 100 piksel değerinin ortalaması alınarak spektral profiller elde edilmiştir. Hiperspektral görüntüleme cihazı ile taranan veriler GANomaly modeline aktarılarak saf bal örneklerinin özelliklerini öğrenilmiş ve bu özellikler dışında kalan

örnekler sahte olarak tanımlanmıştır. Modelin etkinliğini değerlendirmek için geleneksel DD-SIMCA ve OC-PLS tek sınıf sınıflandırıcıları ile karşılaştırma yapılmıştır. GANomaly modelinin, test setindeki saf ve sahte bal örneklerini %100 gibi harika bir doğruluk oranı ile ayırt edebildiği ve diğer geleneksel yöntemlerin önünde olduğu görülmüştür. Yeni Zelanda Manuka balı sahteciliğini tespit etmek için önerilen yöntemin sahtecilik tespitinde muhteşem sonuçlar aldığı savunulmaktadır. Benzer şekilde hiperspektral görüntülemeden yararlanan Wu vd. (2023), kinoa ununda katkı maddesi olup olmadığını belirlemek için taşınabilir bir hiperspektral görüntüleme sistemini kullanarak küresel kalibrasyon modeli oluşturmayı amaçlamışlardır. Çalışma kapsamında deneyler için kullanılacak veri setinde, kinoa, buğday, pirinç, soya ve mısır tohumları, her bir tohum türü için kurutulmuş, öğütülmüş ve 250 mikron boyutunda bir elek kullanılarak aynı parçacık boyutuna getirilmiştir. Kinoa ununa farklı oranlarda (%0-98) katkı maddesi eklenerek toplamda 150 numune hazırlanmış ve 114 numune kalibrasyon için, kalan 36 numune ise doğrulama için kullanılmıştır. Anlamlı verileri elde etmek için 400–1000 nm'de çalışan taşınabilir bir hiperspektral görüntüleme sistemi kullanılmıştır. Her taramada 512x512 piksel çözünürlüğe ve 204 spektral bantlı bir 3D hiper-küp oluşturulmuştur. Görüntülemeden önce sistem, beyaz ve siyah referanslarla kalibre edilmiş ve her numunenin spektral profili, bölgesel maskeleme yapılarak arka plandan ayrılmış ve yalnızca numune alanı üzerinden hesaplama yapılarak veriler elde edilmiştir. Spektral bant sayısını azaltmak ve veri işlemeyi hızlandırmak için, BOSS yöntemi ile 13 önemli dalga boyu seçilmiştir. Bu seçilen dalga boyları, kinoa unundaki katkı maddelerini belirlemede kilit rol oynadığı ve gelecekte taşınabilir birçok bantlı görüntüleme sistemi tasarımında kullanılabileceği vurgulanmıştır. Önerdikleri bu sistemin kinoa unundaki sahteciliğin tespitinde kullanılabileceği savunulmuştur.

Bu hiperspektral yaklaşımları tamamlayan bir diğer çalışma, badem türlerinin ayrıştırılması üzerine gerçekleştirilmiştir. Bu doğrultuda Torres-Rodríguez vd. (2022), tatlı badem partilerinde yapılan sahteciliği hiperspektral görüntüleme ile önlemeyi ele almışlardır. Çalışmada kullanılan örnek grubunun, 97 tatlı badem ve 61 acı badem örneği kullanılarak %5, %10, %15 ve %20 oranında karıştırılarak 500'er gramlık 84 parti oluşturulduğu bilgisi verilmiştir. Analiz öncesi partilerin ısısı 20 derecede sabitlenmiştir. Görüntü alınacak hiperspektral görüntüleme cihazının 946.6–1648 nm dalga boyunda çalışan bir cihaz olduğu bilgisi verilerek görüntülerin yansıma modunda bir CCD kamera ve konveyör bant sistemi kullanılarak toplanmıştır. Elde edilen yansıma değerleri, siyah ve beyaz referanslarla normalize edilmiştir. Acı ve tatlı bademlerin ayrımı için PLS-DA algoritması kullanılmış ve tatlı ve acı bademleri %99'a %100 gibi harika skorlarla ayırt ettiği gözlemlenmiştir. Eğitim setinde 117 de 116 örnek doğru sınıflandırılırken, doğrulama setindeki tüm örnekler doğru analiz edilmiştir. Sonuç olarak, HSI teknolojisi, tatlı badem partilerindeki acı bademlerin tespiti için uygulanabilir bir yöntem olduğu savunulmuştur.

Gıda alanındaki çalışmaların dışında, görüntü işleme yöntemleri sınav sahteciliğinin tespitinde de kullanılmıştır. Görüntü işleme yöntemlerinin sahtecilik tespitinde kullanımına yönelik çalışmalardan birinde Okorie vd. (2023), sınav sahtekarlıklarını tespit etmek için optik sensör ve görüntü işleme sistemine dayalı bir dolandırıcılık tespit sistemi geliştirmişlerdir. Geliştirdikleri sistemi 3 temel bileşen olarak tanımlamışlardır. İlk bileşen olarak tanımladıkları görüntü segmentasyonunda, optik sensörlerden gelen görüntüler ile sınav salonundaki kişilerin tespiti yapılarak, görüntülerin belirli bölümleri analiz edilerek kimlik sahtekarlığının belirlenmesine odaklanılmıştır. Bir diğer bileşen olarak morfolojik operasyonlar ile görüntülerdeki gürültüleri azaltmak ve daha net

görüntüler elde etmek için genişletme gibi morfolojik işlemler uygulanmışlardır. Bu işlemle birlikte kağıt üzerindeki istenmeyen detaylar kaldırılarak sahte kimlikler daha kolay bir şekilde tespit edilmesini hedeflemişlerdir. Son olarak da sınavda bulunan optik sensörler ile verileri alınan kişilerin görüntüsü ile kayıtlı veriler veri tabanında karşılaştırma yapılmaktadır. Sistemin etkinliğini değerlendirmek için hem geliştirilen yöntemle hem de geleneksel yöntemle testler yapılmış, geliştirilen yöntemle 1 saniyede 130 sınav sahtekarlığı tespit edilirken geleneksel yöntemle yalnızca 82 sınav sahtekarlığı tespit edilmiştir. Böylece önerilen sistemin sınav sahtekarlıklarında daha güvenilir bir yöntem olduğu savunulmuştur.

Optik sensör temelli sınav güvenliği çalışmalarından sonra, şüpheli davranışları doğrudan nesne tespitiyle izleyen YOLO tabanlı yöntemler de ele alınmıştır. Daha önce Okorie vd. (2023) tarafından ele alınan optik sensör tabanlı yaklaşımdan farklı olarak Padhiyar vd. (2023), sınav salonlarındaki şüpheli öğrenci davranışlarını izlemek için YOLO kullanmışlardır. Çalışmada kullanılacak veri setinde CIFAR-100, SCUT-HEAD gibi veri setlerinden ve kampüs video kayıtlarından yararlanılmıştır. Çalışmada önerilen sistemin, kamera aracılığıyla, öğrencilerin yüz, üst vücut ve baş hareketlerini algıladığı, daha sonra her 5-10 karede bir öğrencinin pozisyonu hesaplandığı ve euclidean mesafesi ile yer değişimleri analiz edildiği son olarak da mesafe, belirlenen eşik değerleri aştığında alarm üreten bir sistem olduğu açıklanmıştır. Çalışmada kullanılan YOLO'nun etkinliğini kanıtlamak için Cascade ile karşılaştırma yapılmıştır. FPS'den biraz ödün verilse de Cascade'e göre %27 gibi daha iyi bir sonuç elde edilmiştir. YOLO'nun sınav salonlarında yapılan sahtekarlık tespitinde iyi sonuçlar verdiği savunulmuştur. YOLO algoritmasının güvenlik uygulamalarındaki kullanımına yönelik bir diğer çalışmada Katragada vd. (2023), banka ATM'lerinin güvenliği için YOLOv5 kullanılarak, ATM kameralarından elde edilen canlı görüntüler üzerinde düşük maliyetli ve etkili bir güvenlik sistemi

geliştirmişlerdir. Çalışma kapsamında deneyler için kullanılacak veri setinin, card trapper, card scammer, hammer gibi tehdit unsurları ve cüzdan telefon gibi normal unsurlardan oluşan 11.000 adet görüntüden olduğu bilgisi verilmiştir. Normal unsurların veri setine dahil edilmesinin ana sebebinin yanlış pozitifleri düşürmek olduğu vurgulanmıştır. Çalışmada kullanılan sistem mimarisinin, bir kullanıcı arayüzü için web sunucusu, bir de modeli çalıştırma için Amazon EC2 sunucusu içeren multi bir sunucu olduğu bilgisi verilmiştir. Veritabanının, şüpheli aktiviteleri kaydettiği ve kullanıcının ekranına yansıttığı vurgulanmıştır. Modelin etkinliği test etmek için temel model ve multimodel üzerinden karşılaştırmalar yapılmıştır. Çalışma kapsamında geliştirilen model Özellikle card trapper ve face mask kategorilerinde %99 gibi harika sonuçlar verdiği gözlemlenmiştir. Çalışma kapsamında önerilen yöntemin, ATM'lerin güvenliğinde, düşük maliyetli ve gerçek zamanlı çalışması nedeniyle kullanılabileceği savunulmaktadır. Sınav güvenliğine yönelik YOLO tabanlı çalışmalarda ise fiziksel sınav salonlarının yanı sıra çevrimiçi sınav ortamları da ele alınmıştır. Bu kapsamda Singh vd. (2024), online sınavlarda yapılabilecek sahtekarlığın önüne geçmek için YOLO modeli ile geliştirilen yeni bir çevrimiçi sınav gözetim sisteminin tanıtmışlardır. Geliştirilen sınav gözetim modelini 4 aşamada tanımlamışlardır. İlk olarak, sınav sırasında kamera ve mikrofon gibi cihazlardan veri toplanarak, bu veriler yüz tanıma için, yüz tanıma algoritmaları (LBPH) ve gerçek kişi olup olmadığını belirleyen spoofing tespiti, ağız hareketleri izleme, ses analizi ve YOLO ile mobil cihaz algılama ve kişi sayımı gibi çeşitli bilgisayarla görme teknikleri kullanılarak işlenmiştir. Elde edilen sonuçlara göre "Sınavı Sonlandır" veya "Devam Et" kararı verilerek, şüpheli durumlar tespit edildiğinde sınav otomatik olarak sonlandırılan bir sistem olduğu vurgulanmıştır. Yöntemin etkinliğini değerlendirmek için, çeşitli senaryolar ve 70 video görüntüsü üzerinde testler yapılmış ve sistem, özellikle, birden fazla kişinin algılanması ve

mobil cihaz tespitinde çok başarılı sonuçlar elde etmiştir. Ancak özellikle yüz spoofing'de yanlış pozitif sonuçlar görülmüştür. Örneğin, bir fotoğrafın gerçek bir kişi olarak algılanması gibi hatalar yapılmıştır. Sonuç olarak önerilen sistemin çevrimiçi sınavlarda akademik dürüstlüğü sağlamada etkili bir araç olduğu belirtilmiştir.

Sınav ve ATM gibi fiziksel güvenlik ortamlarının yanı sıra sahtecilik tespiti dijital veriler üzerinde de uygulanmaktadır. Bazı çalışmalarda görüntü işleme ve derin öğrenme yöntemlerinden yararlanmak amacıyla farklı veri türleri görüntü biçimine dönüştürülmektedir. Bu kapsamda Zhou vd. (2023), Ethereum akıllı sözleşmelerinde yapılan hileli durumların tespiti için ECA-EfficientNet modeli ve veri artırımı yöntemi geliştirmişlerdir. Çalışma kapsamında kullanılan EfficientNet'in hileli akıllı sözleşmelerin tespitini yaparken, ilk olarak akıllı sözleşmeden gelen bytecode'daki her baytın bir RGB değeri ile eşleştirilerek 3 kanallı görüntülere dönüştürdüğü vurgulanmış. Daha sonra RGB kanalları yer değiştirilerek veri seti zenginleştirildiği, devamında görsellerden özellik çıkarımları yapıldığı ve Softmax fonksiyonu ile sözleşmeler beş sınıfa (Ponzi, honeypot, kimlik avı, kumar, normal) ayrıldığı vurgulanmıştır. Çalışmada kullanılan veri seti, ponzi, honeypot, phishing, kumar ve normal sözleşmelerin Ethereum'dan ve literatürden toplanarak, çok sınıflı sınıflandırma için kullanılan fraud veriseti ve sadece ponzi sözleşmelerini içeren ponzi veriseti oluşturulmuştur. Yöntemin etkinliğini değerlendirmek için fraud veriseti üzerinden literatürdeki diğer yöntemler ile karşılaştırma yapılmış ve en iyi sonucu verdiği görülmüştür. Aynı zamanda önerilen yöntemi ponzi veriseti üzerinde hem literatürdeki yöntemler hem de ponzi veriseti kullanıldığı farklı çalışmalar ile karşılaştırılmış ve önerilen yöntemin ponzi veriseti için en iyi yöntem olduğu görülmüştür. Önerilen yöntemin Ethereum'daki hileli akıllı sözleşmeleri etkili bir şekilde tespit ettiği savunulmuştur.

Dijital verilerden sahtecilik tespiti yapılmasının yanında, yapay zekâ ile sahte görüntülerin oluşturulması da ayrı bir araştırma sorunudur. Bu çerçevede Gu vd. (2022), bilimsel yayınlarda yapay zekâ destekli görsel sahtekarlık ele alınmıştır. Çalışmada, styleGAN modeli ile gerçeğe çok yakın sahte görüntüler oluşturulabildiği, sinGAN modeli ile tek bir görüntü üzerinden eğitim yapılarak birçok yeni görüntü oluşturabileceği ve görüntülerin manipüle edilerek sahte görüntülerin oluşturulabildiği vurgulanmıştır. Yapay zekâ destekli görsel sahtekarlığın tespitinin hem insan hem de otomatik sistemler için zor olduğu belirtilmiştir. Özellikle insanların, doğal görüntülerin gerçekliğini daha iyi değerlendirebilirken, bilimsel görüntülerde yanılabilmesinin çok olası olduğu vurgulanmıştır. Yapay zekâ destekli görsel sahtekarlığın düşük maliyet ve kolay erişimi olduğuna değinilmiş ve gelişmiş yapay zekâ modellerinin açık kaynak olarak paylaşılmasının, kötüye kullanım riskini artırdığı belirtilmiştir. Literatürdeki görüntü işleme araçlarının, yapay zekâ tarafından üretilen görüntüleri güvenilir bir şekilde tespit edememesinden yakınılmıştır. Fakat bu yapılan sahteciliğe karşı bazı çözüm önerileri sunulmuştur. İlk olarak araştırmacılardan, çalışmada kullandıkları görüntülerin ham verilerini sağlamaları istenebileceği belirtilmiştir. Devamında, özellikle bilimsel görüntüler için özel tespit araçlarının geliştirilmesine öncelik verilebileceği söylenmiştir. Son olarak yapay zekâ modelleri yayınlanırken, potansiyel kötüye kullanım etkileri değerlendirilmesinin zaruri olduğu vurgulanmıştır. Sonuç olarak yapay zekâ destekli sahtecilik teknolojilerinin bilimsel yayınlar için ciddi bir tehdit oluşturduğu vurgulanmıştır.

Görsel sahteciliğin belgelere yönelik önemli bir boyutu da kimlik kartlarının taklit edilmesi veya değiştirilmesidir. Bu alanda tespit algoritmalarının geliştirilmesinin yanı sıra eğitim verilerinin oluşturulmasına yönelik yöntemler de araştırılmıştır. Bu doğrultuda Benalcazar vd. (2023), sahte kimlik kartlarını önlemek için sentetik kimlik kartı görüntüleri üretilmiş ve sahtecilik tespitinde kullanılan

derin sinir ağlarını eğitmek için yeni bir veri artırma yaklaşımı önerilmiştir. Sahte kimlik tespitini iyileştirmek için Şili'ye ait kimlik kartlarını temel alarak sentetik görüntü üretimine odaklanmıştır. Şili kimlik kartları, Bona Fide (orijinal), Composite (dijital veya fiziksel olarak değiştirilmiş), Print (baskı) ve Screen (ekran görüntüsü) olmak üzere dört sınıfa ayrılmış ve bu kimlik kartları, veri setinin oluşturulması ve sahtecilik tespiti algoritmalarının eğitilmesinde temel veri kaynağı olarak kullanılmıştır. Sentetik kimlik kartı görüntülerini üretmek için şablon tabanlı görüntüleme, doku transferi ve GAN modelleri olarak 3 farklı yöntem geliştirilmiştir. İlk olarak şablon tabanlı kimlik kartı üretiminde, yüksek kaliteli orijinal kimlik kartlarından şablonlar oluşturularak üzerine rastgele yüz, imza ve bilgiler yerleştirilmiştir. Kullanılan yüzlerin FERET veri kümesinden, imzaların ise Kaggle imza veri kümesinden alındığı belirtilmiştir. Doku transferinde ise 50 farklı renk tonunun, yazıcı ve ekran kaynaklı gürültülerin izolasyonu için kullanıldığı belirtilmiş ve renk paletinin her bir rengi, QR kodu ile işaretlenerek izole edilmiş dokular oluşturulmuştur. Dokular, kimlik kartının boyutuna göre kırpılmış ve birleştirmeler yapılmıştır. Bu şekilde toplam 10.000 doku izolasyonu gerçekleştirilmiştir. Son olarak GAN modellerinden, StyleGAN2 ve CycleGAN kullanılarak hem sahte hem orijinal kimlik kartı görüntüleri üretilmiş ve bu görüntüler ile MobileNetV2 tabanlı sahtecilik tespit ağları, kompozit (bütünleşik) tespit ve kaynak tespiti için ayrı ayrı eğitilmiştir. Yapılan eğitimden sonra sonuçlar ele alındığında, kompozit tespit sonuçlarına göre, sentetik görüntülerin, yalnızca gerçek verilerle eğitim alan modellere kıyasla %4 daha yüksek hata oranı verdiği görülmüştür. Kaynak tespitine göre, yazıcı kaynaklı sahtecilik tespitinde CycleGAN yöntemi, yalnızca %1 performans kaybıyla gerçek verilerle benzer sonuçlar üretmiştir. Son olarak fid analizine göre, CycleGAN ve doku transferi yöntemiyle üretilen görüntülerin, orijinal görüntülere en çok benzediği (FID $\approx$ 20) saptanmıştır. Sentetik kimlik kartı üretiminin, geniş veri kümeleri oluşturmada maliyet ve gizlilik

kaygılarını azaltan bir alternatif olarak öne çıktığı vurgulanmıştır. Özellikle yazıcı ve ekran kaynaklı saldırılarda, CycleGAN gibi yöntemlerin etkili olduğu kanıtlanmıştır. Ancak, kompozit sahtecilik tespitinde sentetik verilerin hala gerçek verilere kıyasla dezavantajlı olduğu belirtilmiştir.

Bu çalışmalar, görüntü işleme ve derin öğrenme yöntemlerinin farklı sahtecilik türlerinde kullanılabildiğini göstermektedir. Ancak incelenen uygulamaların veri yapıları ve tespit hedefleri farklılık göstermektedir. Bu çalışma ise fiş görüntülerindeki "damage" ve "fraud" etiketlerinin YOLO modelleriyle tespit edilmesine odaklanmaktadır.

Materyal ve Yöntem

Bu çalışmanın amacı, ticari faaliyetleri tehdit eden sahte fişleri tespit etmek için YOLOv5, YOLOv8, YOLOv9 ve YOLOv10 modellerinin performanslarını karşılaştırmalı olarak değerlendirmektir. Araştırmanın temelini, Roboflow platformunda yayımlanan Receipt Fraud veri seti üzerinde eğitilen YOLOv5, YOLOv8, YOLOv9 ve YOLOv10 modellerinin performanslarının karşılaştırılması oluşturmaktadır. Bu bölümde veri setinin özellikleri, karşılaştırılan YOLO modellerinin eğitim süreçleri, temel mimari özellikleri ve performans değerlendirme ölçütleri açıklanmaktadır.

Veri Seti

Bu çalışmanın temelini oluşturan veri seti, kullanılabilir gerçek zamanlı fiş görüntülerini içerir. Veri seti, Roboflow platformundan, günlük hayattan toplanan 527 fiş görüntüsünü ve fiş üzerinden etiketlenmiş 'damage' ve 'fraud' etiketlerini içerir (projectORCA, 2024). Veri setini geliştirmek için döndürme, gri ölçekleme ve doygunluk gibi veri artırma teknikleri kullanılarak veri seti 1265 görüntüye genişletilmiştir. Veri seti öncelikle eğitim, doğrulama ve test kümelerine ayrılmıştır. Ardından yalnızca eğitim

kümesindeki görüntülere veri artırma teknikleri uygulanmıştır. Böylece aynı görüntünün artırılmış sürümlerinin farklı veri kümelerinde yer alması önlenerek veri sızıntısı riski azaltılmıştır. Veri artırma örnekleri Şekil 1'de gösterilmektedir.

Şekil 1 Veri artırma örnekleri a) Ham veri b) Veri artırma teknikleri ile oluşturulan veri



Bu teknikler, modellerin genelleme yeteneğini önemli ölçüde artırmaktadır. Bu yöntemler, aşırı öğrenmeyi önlemek ve çeşitli çevresel koşullara dayanıklı olmasını sağlamak için kullanılmıştır. Ayrıca, veri çeşitliliğini artırarak modelin geniş bir veri kümesi ile eğitilmesine olanak sağlayarak daha güçlü ve genelleştirilebilir sonuçlar elde edilmesini hedeflemiştir. Veri seti, değişken ışık koşullarında, çeşitli açılardan ve çeşitli senaryolarda çekilmiştir. Bu sayede modelin gerçek dünya koşullarına uyum sağlanması amaçlanmıştır. Toparlanan görüntüler 640 x 640 piksele indirgenmiştir.

Veri artırma sonrasında veri kümesi, 1.107 eğitim, 105 doğrulama ve 53 test görüntüsü olmak üzere toplam 1.265 görüntüden oluşmuştur. Veri kümelerinin dağılımı Tablo 1'de sunulmuştur. Modelin genelleme yeteneğini test etmek için ayrılan

test seti, model parametrelerini doğrulamak için kullanılacak doğrulama seti ile dengeli bir şekilde dağıtılmıştır. Bu yapılandırma, modelin çeşitli koşullara ve çevresel değişikliklere karşı dayanıklılığını değerlendirmek için çok önemlidir.

Tablo 1 Veri kümesinin eğitim, doğrulama ve test kümelerine dağılımı

Toplam Veri	Eğitim Seti	Doğrulama Seti	Test Seti
1265	1107	105	53

Model Eğitimi

Model eğitimi, Intel i5-12400F işlemci, 16 GB RAM ve NVIDIA GeForce RTX 3070 Ti ekran kartına sahip bir bilgisayarda gerçekleştirilmiştir. Derin öğrenme algoritmalarının gerektirdiği hesaplama yükünü aşmak için bu güçlü donanım seçilmiştir. Eğitim GPU üzerinden gerçekleştirilerek, işlem sürelerini önemli ölçüde kısaltması amaçlanmıştır.

YOLO modellerinin eğitiminde Adam optimizasyon algoritması kullanılmıştır. Eğitim sürecinde epoch sayısı 100, başlangıç öğrenme oranı 0,001, momentum parametresi 0,937 ve parti boyutu 8 olarak belirlenmiştir. Bu hiperparametreler Tablo 2'de gösterilmektedir. Çalışmada, sahtecilik tespiti yapabilmek için “damage” ve “fraud” etiketleri roboflow platformu kullanılarak gerçekleştirilmiştir. Eğitim sürecinde karşılaşılan zorluklar arasında eğitimin, elektrik kesintileri nedeniyle durması başı çekmektedir. Bu tür durumlar, model eğitiminin süresini uzatmış ve süreci yavaşlatmıştır.

Tablo 2 Eğitim hiperparametreleri

Hiperparametreler	Parametre Ayarları
Epoch	100
Başlangıç Öğrenme Oranı	1×10^{-3}
Momentum	0.937
Parti Boyutu	8
Optimizasyon Algoritması	Adam

YOLO

YOLO algoritma ailesi, gerçek zamanlı nesne tespitinde, hızı ve doğruluk oranlarıyla diğer nesne tespit algoritmalarına göre öne çıkmaktadır. Nesneleri tanımlamak için kayan pencereler veya bölge önerileri kullanan geleneksel yöntemlerin aksine YOLO, birden fazla sınırlayıcı ve bunlara karşılık gelen sınıf olasılıklarını tahmin etmek için tek bir evrişimli sinir ağı (CNN) kullanmaktadır. Tek atış olarak isimlendirilen bu yöntem, YOLO'nun tüm görüntüleri aynı anda işlemesini sağlayarak, fiş sahteciliğinin tespiti gibi yüksek hızlı uygulamalar için mükemmel bir araç olarak öne çıkmaktadır. Tablo 3'te genel YOLO algoritmasının sözde kodu verilmiştir.

YOLO'nun temel yaklaşımında, görüntüler üzerinde oluşturulan özellik haritaları aracılığıyla nesnelerin konumları ve sınıfları tahmin edilmektedir. Nesne konumlarının belirlenmesi ve tahminlerin işlenmesi, kullanılan YOLO sürümünün mimarisine göre farklılık gösterebilmektedir.

YOLO ailesinin sunduğu yeniliklerden bir tanesi de, birleşik bir mimariye sahip olması ve nesne tanımayı bir regresyon olarak ele almasıdır. YOLO, sınıflandırma, yerelleştirme ve güven tahminlerini birleştiren tek bir kayıp fonksiyonunu optimize ederek hem doğruluğu hem de hesaplama verimliliğini dengelemektedir. YOLO ailesinde nesne konumlarının belirlenmesinde, kullanılan model sürümüne bağlı olarak bağlantı kutusu tabanlı (anchor-based) ve bağlantı kutusundan bağımsız (anchor-free) yaklaşımlar

kullanılmaktadır. Bu sayede, metinsel ifadeler, sayısal tutarsızlık ve hasar gibi fişlerde karşılaşılabilecek küçük ve karmaşık ayrıntıları tespit etmede yardımcı olmaktadır.

YOLO ailesi temel olarak güçlü bir yapıda olsa da her yeni modeliyle güncellemiştir. Çalışma kapsamında kullanılan YOLOv5, gelişmiş bağlantı kutusu seçimi, daha iyi veri artırma teknikleri ve çapraz aşamalı kısmi ağlar (CSPNet) aracılığıyla güncellenerek daha gelişmiş özellik çıkarımı yapmayı hedeflemiştir. YOLOv8 ve YOLOv9'da, gelişmiş derin sinir ağları, optimize edilmiş aktivasyon fonksiyonları ve görüntünün kritik alanlarına daha iyi odaklanmak için dikkat mekanizmaları eklenerek daha da güçlü bir algoritma nesli olarak çıkış sağlamıştır. YOLOv10'da ise artırılmış ölçeklenebilirlikle eski nesillerin yetenekleri daha da fazla geliştirilmiştir. YOLO mimarisinin esnek yapısı bu güncellemelere olanak sağlamaktadır. Bu nedenle çok farklı alanlarda aktif olarak gerçek zamanlı nesne tespiti için kullanılabilmektedir. Chen vd. (2023) sınıf içi öğrenci durumlarını tespit etmek için, Shukhratov vd. (2024) plastik atıklarının tespitinde, Hou vd. (2024) iletim hatlarının uç noktası hasar tespitinde, Sun vd. (2024) tütün bitkileri zararlılarının tespitinde, Li vd. (2023) makine kullanan işçilerin eldiven giyme durumunun tespitinde, Niu vd. (2024) mısır tohum kalitesinin tespitinde YOLO ailesi kullanılmıştır.

Bu model, farklı koşullarda yüksek adaptasyon yeteneği, güçlü ve esnek yapısı sayesinde, tarımdan sahtecilik tespitine kadar farklı alanlarda kullanılmış ve güçlü sonuçlar vermiştir. Modelin genel başarımı, sahtecilik faaliyetlerin tespitinde verimliliği artırabilir ve büyük ölçüde maliyet tasarrufu sağlayabilir.

Tablo 3 YOLO tabanlı nesne tespitinin genel işlem adımları

Algoritma 1: YOLO

$\boldsymbol{T} \leftarrow YOLO\ modeli$ (önceden eğitilmiş $model$)
$\boldsymbol{input_image} \leftarrow Giriş$ görüntüsü
$\boldsymbol{grid} \leftarrow$ Görüntü $S\ x\ S$ ızgarasına böl
$\boldsymbol{E = 0} \leftarrow Başlangıç\ durumu$ ($hiçbir\ nesne\ tespit\ edilmedi$)

while $(E > mintreshhold)\ and\ (overlappingbox)$ ***do***
  $candidate_boxes \leftarrow T(input_image, grid)$
  $candidate_classes \leftarrow tahmin\ edilen\ sınıflar$
  $candidate_confidence \leftarrow her\ kutunun\ güven\ skoru$

  $\Delta E \leftarrow candidate_confidence - E$
  if Accept $(\Delta E, T)$ **then**
    $E \leftarrow candidate_confidence$
    $final_detections \leftarrow Güncellenmiş\ kutular\ ve\ sınıflar$
  end
  $grid \leftarrow$ Çözünürlük $güncelle$
end
$final_detections \leftarrow Non - Maximum\ Suppression(final_detection)$
return final_detections

Performans Değerlendirme

Modelin çalışmasını değerlendirmek için bir dizi istatistiksel metrik kullanılmıştır. Bu ölçümler, modelin fiş sahteciliğinin doğru ve tutarlı bir şekilde belirleme kapasitesini artırmak için çok önemlidir. Modelin, eğitim ve doğrulama aşamalarında elde edilen sonuçlar, performans değerlendirme kriteleri üzerinden yapılmıştır. Çalışmada kullanılan performans metrikleri arasında precision (kesinlik), recall (duyarlılık), f1-skoru, mAP (mean Average Precision) ve confusion matrix (karışıklık matrisi) yer almaktadır.

Kesinlik

Kesinlik, modelin pozitif olarak tahmin ettiği durumların gerçekte pozitif olanların oranını hesaplar. Yüksek bir kesinlik değeri, modelin yanlış pozitif sayısının düşük olduğunu gösterir, bu da modelin, olmayan bir şeyi varmış gibi görünmesini gösterir. Eşitlik 1'de kesinlik hesabı gösterilmiştir.

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

TP, tahmin edilen değerin doğru olduğunu gösterir, ancak FP, yanlış pozitifi gösterir. FP, yanlış pozitifi temsil eder. Yanlış pozitif ise yanlış tahmin değeridir. Kesinlik metriği, modelin ne kadar "hassas" olduğunu gösterir. Özellikle, fiş sahteciliğinin tespiti gibi kritik uygulamalarda yüksek kesinlik, önemsiz hatalardan kaçınmak için önemlidir.

Duyarlılık

Duyarlılık, gerçekte pozitif olan örneklerin ne kadarının model tarafından doğru olarak pozitif olarak işaretlendiğini hesaplar. Yüksek bir duyarlılık değeri, modelin gerçek pozitifleri kaçırmadığını, yani çok az sayıda yanlış negatif hata yaptığını göstermektedir. Eşitlik 2'de duyarlılık hesabı gösterilmiştir.

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

2'de TP, doğru pozitifi belirtirken FN, yanlış negatifi gösterir.

F1-Skoru

F1-Skoru, kesinlik ve duyarlılığın harmonik ortalamasıdır ve iki ölçeği dengeli bir şekilde birleştirir. Bu metrik, özellikle veri seti dengesiz olduğunda yani bir sınıfın diğerlerinden çok daha az ya da çok daha fazla örneği olduğunda modelin genel performansını değerlendirmek için kullanılır. F1-Skoru hesabı Eşitlik 3'te gösterilmiştir.

$$F_1 = 2\ x\ \frac{Precision\ x\ Recall}{Precision + Recall} \quad (3)$$

F1-Skoru, her iki metriği eşit ağırlıkta hesaba katarak, modelin genel doğruluk ve güvenilirlik dengesini sunar.

mAP

Ortalama Kesinlik Ortalaması (mean Average Precision, mAP), nesne tespiti performansının değerlendirilmesinde kullanılan bir ölçüttür. Her sınıf için kesinlik–duyarlılık eğrisi üzerinden hesaplanan AP (ortalama kesinlik) değerlerinin ortalaması alınarak elde edilmektedir. Bu çalışmada, kesişimin birleşime oranı (Intersection over Union, IoU) eşik değeri 0,5 olarak belirlenen mAP@0.5 ölçütü kullanılmıştır. Eşitlik 4'te AP hesabı gösterilmiştir.

$$AP = \sum_n (R_n - R_{n-1})\ x\ P_n \quad (4)$$

Burada P_n, n'inci noktanın hassasiyetini, R_n, n'inci noktanın geri çağrımı ve R_{n-1} bir önceki noktanın geri çağırımını temsil eder.

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \quad (5)$$

4'te N sınıf sayısını gösterirken AP_i, i'nci sınıfın AP değerini temsil eder. Bu nedenle mAP, tüm sınıflar için AP değerinin genel değerini alır. Özellikle nesne tespiti ve segmentasyon görevlerinde yaygın olarak kullanılan önemli bir metriktir. Modelin tüm sınıflar arasında tutarlı olup olmadığını belirlemek için kullanılır.

Karışıklık Matrisi

Karışıklık matrisi, bir sınıflandırma modelinin başarısını, çeşitli sınıflar arasındaki doğru ve yanlış tahminleri gösterir. Bu matris, modelin her sınıf için ne kadar iyi çalıştığını görsel bir şekilde sunmaktadır. Özellikle birden fazla sınıf olduğu durumlarda, modelin her bir sınıfı nasıl tanımladığını ve sık yapılan hataları anlamak için çok değerlidir. Bu matris, kesinlik, duyarlılık ve F1-skoru gibi diğer performans metriklerinin hesaplanmasına temel sağlamaktadır.

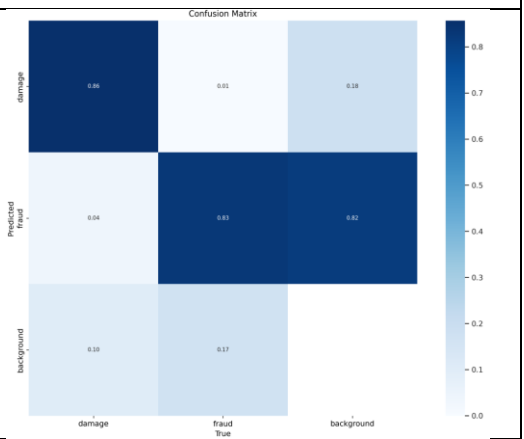
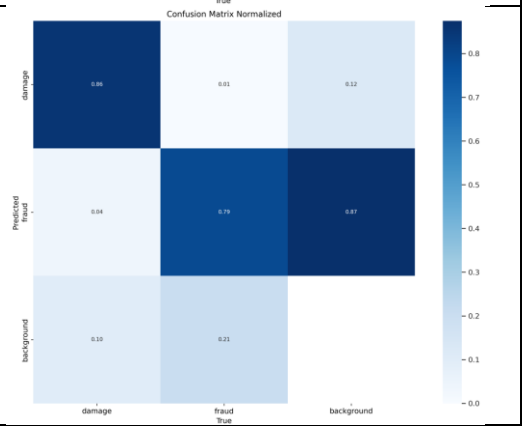
Performans değerlendirme süreci, modelin saha uygulamalarına ne kadar uygun olduğunu ve sahtecilik tespiti için gerçek potansiyelini belirlemek için temel bir adımdır. Bu süreç,

modelin geliştirilmesine yardımcı olmak için gerekli bilgileri sağlar ve pratik uygulamalarda modelin nasıl optimize edilebileceğini göstermektedir.

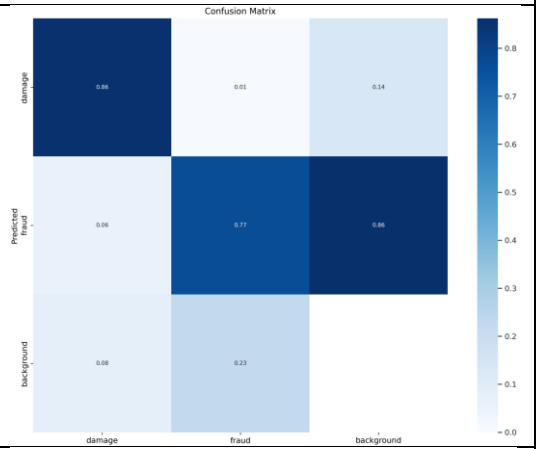
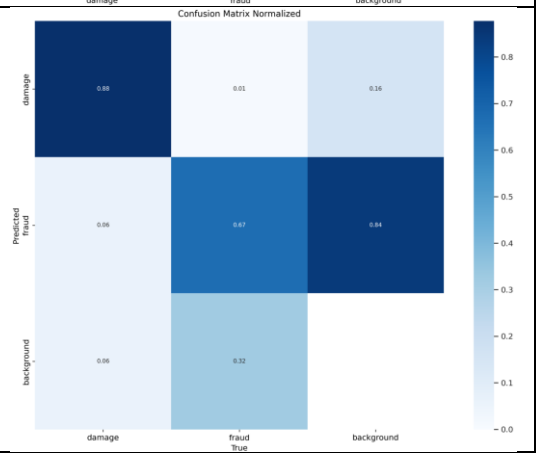
Deneysel Sonuçlar

Elde edilen sonuçlar, modellerin, fiş sahteciliğini tespit etme kapasitesini geniş bir yelpazede sergilemektedir. Çalışma kapsamında incelenen YOLO modelleri genel olarak, fiş sahteciliğinin tespitinde yüksek bir doğruluk oranı göstermiştir.

Tablo 4 YOLO modellerine ait karışıklık matrisleri

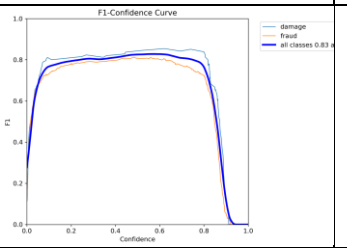
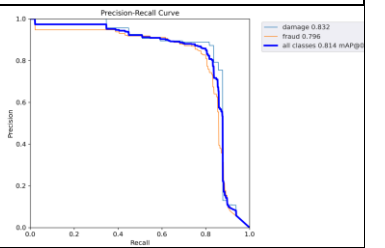
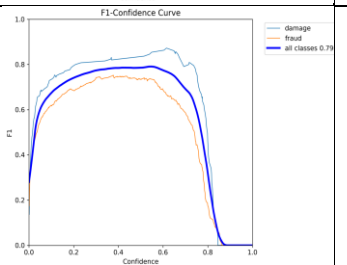
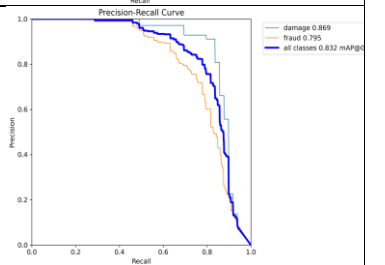
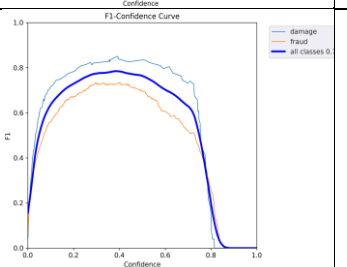
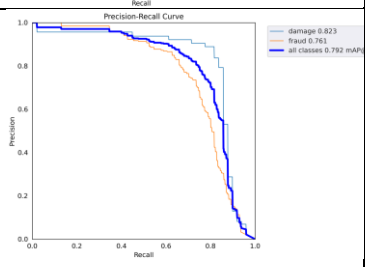
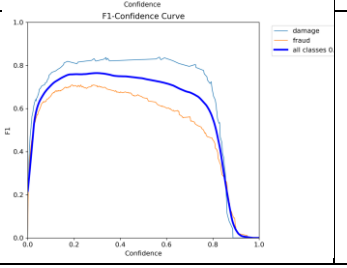
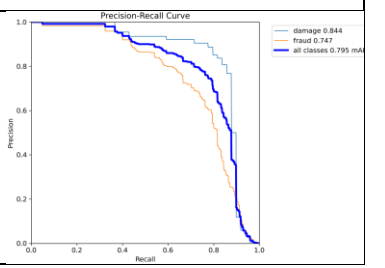
Model	Karışıklık Matrisi
YOLOv5	
YOLOv8	

Tablo 4 (devamı)

Model	Karışıklık Matrisi
YOLOv9	
YOLOv10	

Tablo 4'te YOLOv5, YOLOv8, YOLOv9 ve YOLOv10 modellerine ait normalize edilmiş karışıklık matrisleri sunulmuştur. Matrisler incelendiğinde, modellerin damage ve fraud etiketli bölgeleri farklı oranlarda tespit ettiği görülmektedir. Özellikle fraud sınıfında, arka planla ilişkili eşleşme hatalarının bulunduğu gözlemlenmiştir. Genel performans karşılaştırması ise Tablo 8'de sunulan mAP@0.5 değerleri üzerinden gerçekleştirilmiştir.

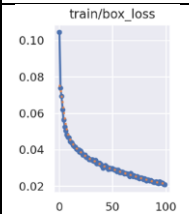
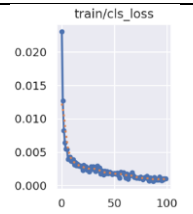
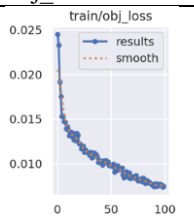
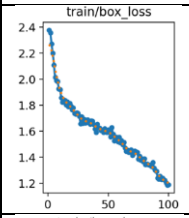
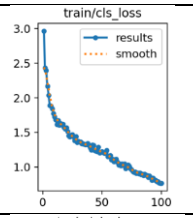
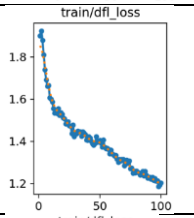
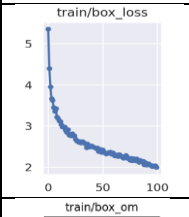
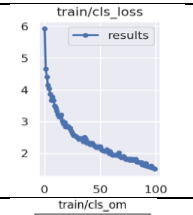
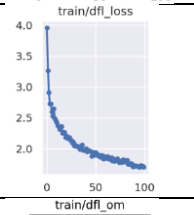
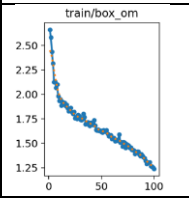
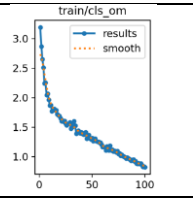
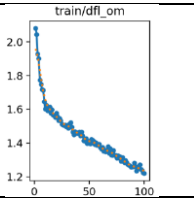
Tablo 5 Modellerin eğitim sonrası F1-Curve ve PR-Curve grafikleri

Model	F1-Curve	PR-Curve
YOLOv5		
YOLOv8		
YOLOv9		
YOLOv10		

Tablo 5'te modellerin eğitim sonrası F1 ve PR curve'leri gösterilmiştir. Genel performans ele alındığında en iyi sonucu YOLOv8 ve YOLOv5 verdiği görülmüştür. YOLOv5 "fraud" sınıfında küçük bir üstünlük sağlasa da "damage" sınıfında YOLOv8 daha üstündür. YOLOv10 ve YOLOv9 "fraud" sınıfında düşük

değerler göstermesiyle dikkat çekmiştir. Damage sınıfı için en yüksek AP@0.5 değeri %86,9 ile YOLOv8 modelinde elde edilirken, YOLOv10 modeli %84,4 değerine ulaşmıştır. Genel denge ve tutarlılık açısından en iyi sonuçları YOLOv8 ve YOLOv5 verdiği görülmüştür.

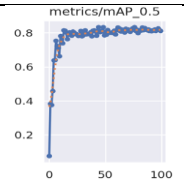
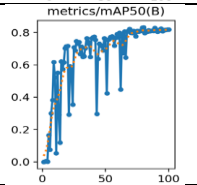
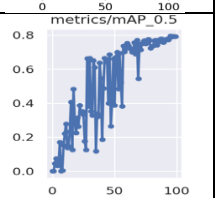
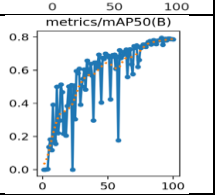
Tablo 6 Modellerin eğitim sonrası loss grafikleri

Model	Box_loss	Cls_loss	Dfl_loss / obj_loss
YOLOv5			
YOLOv8			
YOLOv9			
YOLOv10			

Eğitim sürecindeki kayıplar, modelin öğrenme sürecinin sağlıklı ilerlediğini göstermektedir. Tablo 6'da modelin öğrenme sürecindeki eğitimin kayıpları gösterilmiştir. Eğitim kayıpları incelendiğinde YOLOv5 modeli için hızlı ve düzenli bir azalma gözlemlenmiştir. YOLOv8 modelinde box ve cls loss'ta istikrarlı bir

düşüş görülürken, dfl_loss'ta ufak dalgalanmalar görülmüştür. YOLOv9, diğer modellere göre daha yüksek başlangıç kayıplarına sahip olsa da, eğitim ilerledikçe box, cls ve dfl_loss'ta düzenli düşüş yaşandığı gözlemlenmiştir. YOLOv10'da ise benzer şekilde loss grafikleri düzenli olarak azaldığı görülmüştür. Eğitim kayıplarındaki genel azalma, modellerin eğitim sürecinde optimizasyonun ilerlediğini göstermektedir.

Tablo 7 Modellerin eğitim sonrası mAP50 grafikleri

Model	mAP50
YOLOv5	
YOLOv8	
YOLOv9	
YOLOv10	

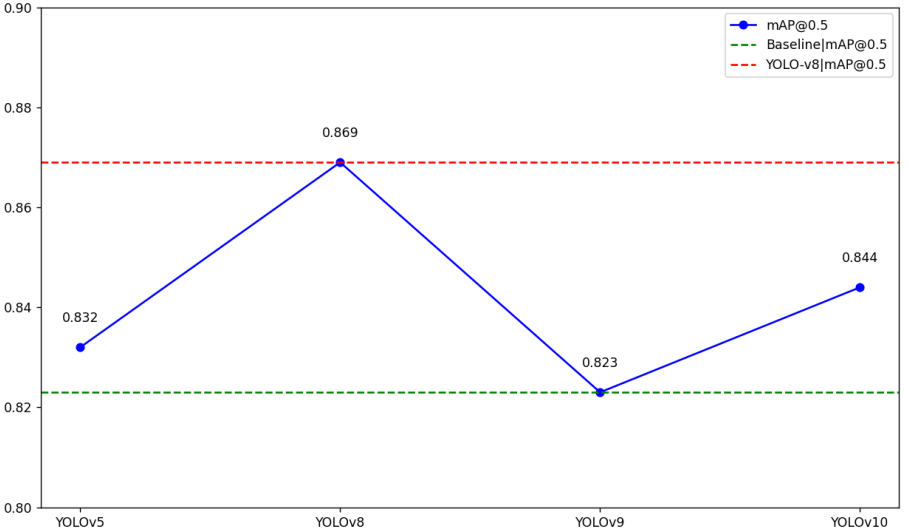
Modelin mAP50 değeri, ortalama doğruluğu gösterir ve modelin genel etkinliğini ölçmede kritik bir göstergedir. Tablo 7'de bu çalışma kapsamında modellerin eğitim süresince mAP50 grafiği verilmiştir. Eğitim sonunda YOLOv5 ile yapılan eğitimde %81.4,

YOLOv8'in %83.2, YOLOv9'un %79.2 ve YOLOv10'un %79.5 mAP50 skoru ile tamamladığı gözlemlenmiştir. Bu sonuçlar modellerin, azımsanamayacak doğruluk ve tutarlılıkla nesneleri tespit edebildiğini göstermektedir.

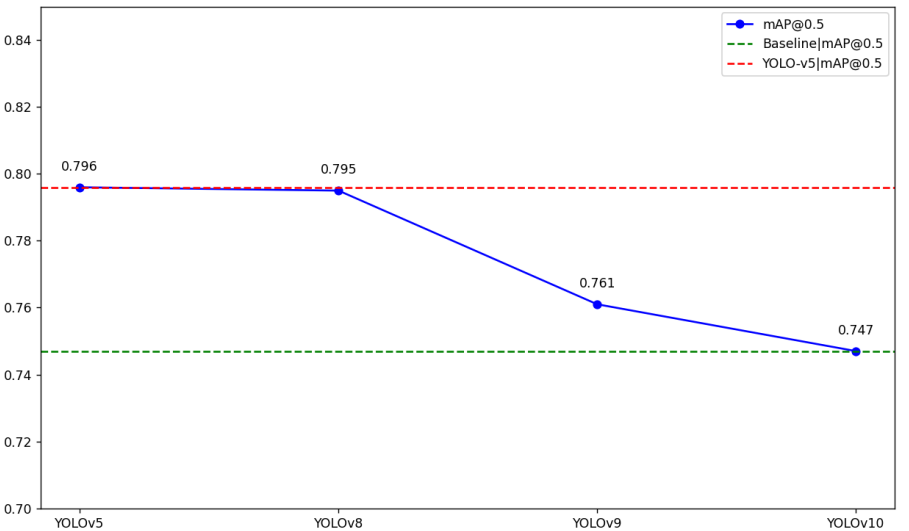
Bu sonuçlar, çalışma kapsamında kullanılan YOLO modellerinin fiş sahteciliği tespitinde yüksek performans gösterdiğini ve gerçek dünya senaryolarında güvenilir sonuçlar elde edebileceğini ortaya koymaktadır. Ancak, yanlış pozitif oranının azaltılması ve daha da geniş veri setleri üzerinde test yapılması modelin performansını daha da artırabilir.

Çalışma kapsamında YOLO modellerinin etkinliğini ölçmek için farklı YOLO aynı veri seti üzerinden eğitim yapılmış ve "damage","fraud" ve "genel" mAP değerleri elde edilmiştir. "Damage" için elde edilen sonuçlar Şekil 2'de, "Fraud" için elde edilen sonuçlar Şekil 3'te, modellerin genel performansları ise Şekil 4'te ve tüm durumlar Tablo 8'de gösterilmiştir. "Damage" sınıfı için elde edilen sonuçlar incelendiğinde %86.9 ile YOLOv8 en iyi sonucu verirken %84.4 ile YOLOv10 onu takip etmiştir. Bu sınıf için %82.3 ile YOLOv9 en az başarı gösteren sınıf olmuştur. "Fraud" sınıfı için ise %79.6 ile YOLOv5 en iyi sonucu verirken %79.5 ile YOLOv8 modeli hemen arkasından gelmiştir. %74.7 ile YOLOv10 bu modeller arasında en kötü sonucu vermiştir. Modellerin genel performanslarına bakıldığında %83.2 ile YOLOv8 diğer modellerin önünde olduğu görülürken, %81.4 ile YOLOv5 en iyi ikinci sonucu veren model olmuştur. Genel sonuçlarda %79.2 performans veren YOLOv9 çalışma kapsamında incelenen YOLO modelleri arasında en kötü sonucu vererek diğer YOLO modellerinin gerisinde kalmıştır.

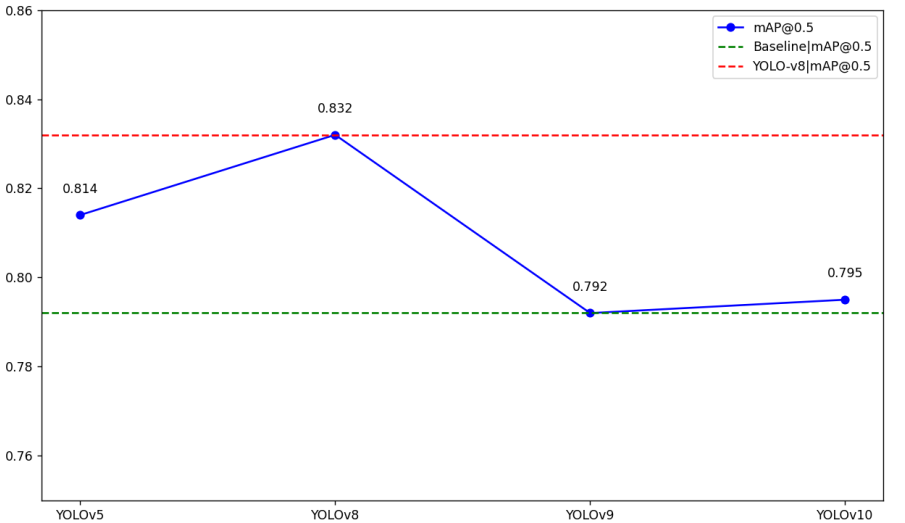
Şekil 2 Damage sonuçlarının karşılaştırılması



Şekil 3 Fraud sonuçlarının karşılaştırılması



Şekil 4 Genel sonuçların karşılaştırılması



Tablo 8 Model sonuçlarının karşılaştırılması

Tespit Türü	Model	mAP@0.5
Damage	YOLOv5	0.832
	YOLOv8	**0.869**
	YOLOv9	0.823
	YOLOv10	0.844
Fraud	YOLOv5	**0.796**
	YOLOv8	0.795
	YOLOv9	0.761
	YOLOv10	0.747
Genel	YOLOv5	0.814
	YOLOv8	**0.832**
	YOLOv9	0.792
	YOLOv10	0.795

Sonuçlar

Bu çalışmada, fiş sahteciliğinin tespiti için yüksek tespit yeteneğine sahip YOLO modelleri kullanılmıştır. Modeller, Roboflow platformunda yayımlanan Receipt Fraud veri seti temel alınarak hazırlanmış görüntüler üzerinde eğitilmiş ve elde edilen

sonuçlar farklı performans ölçütleriyle karşılaştırılmıştır. Modeller, YOLOv5 ile yapılan eğitimde %81.4, YOLOv8'in %83.2, YOLOv9'un %79.2 ve YOLOv10'un %79.5 mAP skorları elde ederek tatmin edici sonuçlar almıştır. Bu sonuçlar, fiş sahteciliği tespiti için yüksek bir başarı elde ettiğini ortaya koymuştur. Genel çerçevede YOLOv8 ve YOLOv5 ile yapılan eğitimlerin YOLOv9 ve YOLOv10'dan daha etkili olduğu görülmüştür. YOLOv9 ve YOLOv10'un YOLOv8 ve YOLOv5'in gerisinde kalmasının birkaç olası nedeni olabilir. YOLOv9 ve YOLOv10 modellerinin daha düşük performans göstermesi, veri kümesinin özellikleri, model yapılandırmaları ve eğitim koşullarıyla ilişkili olabilir. Ancak bu etkenlerin performans üzerindeki etkisini belirlemek için farklı veri kümeleri ve hiperparametre yapılandırmalarıyla ek deneyler gerçekleştirilmesi gerekmektedir. Bu çalışmanın başlıca sınırlılıkları, değerlendirmelerin tek bir veri kümesi üzerinde gerçekleştirilmesi ve özellikle bağımsız test kümesindeki görüntü sayısının sınırlı olmasıdır. Gelecek çalışmalarda farklı veri kümeleri üzerinde karşılaştırmalar yapılarak modellerin genelleme yeteneklerinin daha kapsamlı incelenmesi hedeflenmektedir. Sonuç olarak, bu çalışmada, farklı YOLO modellerinin gerçek hayat senaryolarına çözüm sunduğu gözlemlenmiş ve bu alana katkı sağlamıştır.

Kaynakça

Benalcazar, D., Tapia, J. E., Gonzalez, S., & Busch, C. (2023). Synthetic ID card image generation for improving presentation attack detection. *IEEE Transactions on Information Forensics and Security, 18*, 1814–1830. https://doi.org/10.1109/TIFS.2023.3255585

Besharati, M. S., Pourdarbani, R., Sabzi, S., Sotoudeh, D., Ahmaditeshnizi, M., & García-Mateos, G. (2024). Counterfeit detection of Iranian black tea using image processing and deep learning based on patched and unpatched images. *Horticulturae, 10(7)*, 665. https://doi.org/10.3390/horticulturae10070665

Chen, H., Zhou, G., & Jiang, H. (2023). Student behavior detection in the classroom based on improved YOLOv8. *Sensors, 23(20),* 8385. https://doi.org/10.3390/s23208385

Cheng, J., Zhang, G., Abdulla, W., & Sun, J. (2024). Advancing fraud detection in New Zealand Mānuka honey: Integrating hyperspectral imaging and GANomaly-based one-class classification. *Food Bioscience, 60*, 104428. https://doi.org/10.1016/j.fbio.2024.104428

Gu, J., Wang, X., Li, C., Zhao, J., Fu, W., Liang, G., & Qiu, J. (2022). AI-enabled image fraud in scientific publications. *Patterns, 3(7),* 100511. https://doi.org/10.1016/j.patter.2022.100511

Hou, C., Li, Z., Shen, X., & Li, G. (2024). Real-time defect detection method based on YOLO-GSS at the edge end of a transmission line. *IET Image Processing*, 1–13. https://doi.org/10.1049/ipr2.13028

Jahanbakhshi, A., Abbaspour-Gilandeh, Y., Heidarbeigi, K., & Momeny, M. (2021). Detection of fraud in ginger powder using an automatic sorting system based on image processing technique and deep learning. *Computers in Biology and Medicine, 136*, 104764. https://doi.org/10.1016/j.compbiomed.2021.104764

Katragada, A. S., Vuribindi, K., & Erickson, V. L. (2023). Real-time suspicious activity detection on ATMs using multimodel YOLO object detection. *In 2023 International Conference on Computational Science*

and Computational Intelligence (CSCI). https://doi.org/10.1109/CSCI62032.2023.00204

Li, S., Huang, H., Meng, X., Wang, M., Li, Y., & Xie, L. (2023). A glove-wearing detection algorithm based on improved YOLOv8. *Sensors, 23(24),* 9906. https://doi.org/10.3390/s23249906

Nargesi, M. H., & Kheiralipour, K. (2024). Visible feature engineering to detect fraud in black and red peppers. *Scientific Reports, 14(1),* 25417. https://doi.org/10.1038/s41598-024-76617-1

Niu, S., Xu, X., Liang, A., Yun, Y., Li, L., Hao, F., Bai, J., & Ma, D. (2024). Research on a lightweight method for maize seed quality detection based on improved YOLOv8. *IEEE Access, 12,* 32927–32937. https://doi.org/10.1109/ACCESS.2024.3365559

Okorie, E., Nwaukwa, C. J., Dom-anyanwu, C. P., & Ezika, N. D. (2023). An image processing fraud detecting system using optical sensor. *International Journal of Research, 10(08),* 54–57. https://doi.org/10.5281/zenodo.8223443

Padhiyar, P., Parmar, K., Parmar, N., & Degadwala, S. (2023). Visual distance fraudulent detection in exam hall using YOLO detector. *In International Conference on Inventive Computation Technologies (ICICT).* https://doi.org/10.1109/ICICT57646.2023.10134271

projectORCA. (2024). *Receipt Fraud Dataset* (Version 2) [Data set]. Roboflow. https://universe.roboflow.com/projectorca/receipt-fraud/dataset/2

Shukhratov, I., Pimenov, A., Stepanov, A., Mikhailova, N., Baldycheva, A., & Somov, A. (2024). Optical detection of plastic waste through computer vision. *Intelligent Systems with Applications, 22,* 200341. https://doi.org/10.1016/j.iswa.2024.200341

Singh, T., Nair, R. R., Babu, T., & Duraisamy, P. (2024). Enhancing academic integrity in online assessments: Introducing an effective online exam proctoring model using YOLO. *Procedia Computer Science, 235,* 1399–1408. https://doi.org/10.1016/j.procs.2024.04.131

Sun, D., Zhang, K., Zhong, H., Xie, J., Xue, X., Yan, M., Wu, W., & Li, J. (2024). Efficient tobacco pest detection in complex environments using an enhanced YOLOv8 model. *Agriculture, 14(3),* 353. https://doi.org/10.3390/agriculture14030353

Torres-Rodríguez, I., Sánchez, M.-T., Entrenas, J.-A., Vega-Castellote, M., Garrido-Varo, A., & Pérez-Marín, D. (2022). Hyperspectral imaging for the detection of bitter almonds in sweet almond batches. *Applied Sciences, 12(10),* 4842. https://doi.org/10.3390/app12104842

Wu, Q., Mousa, M. A. A., Al-Qurashi, A. D., Ibrahim, O. H. M., Abo-Elyousr, K. A. M., Rausch, K., Abdel Aal, A. M. K., & Kamruzzaman, M. (2023). Global calibration for non-targeted fraud detection in quinoa flour using portable hyperspectral imaging and chemometrics. *Current Research in Food Science, 6,* 100483. https://doi.org/10.1016/j.crfs.2023.100483

Zhang, L., Wei, Y., Liu, J., An, D., & Wu, J. (2024). Maize seed fraud detection based on hyperspectral imaging and one-class learning. *Engineering Applications of Artificial Intelligence, 133,* 108130. https://doi.org/10.1016/j.engappai.2024.108130

Zhou, X., Yang, W., Wang, L., Wei, F., HaiLaTi, K., & Liao, Y. (2023). The detection of fraudulent smart contracts based on ECA-EfficientNet and data enhancement. *Computers, Materials & Continua, 77(3),* 4074–4092. https://doi.org/10.32604/cmc.2023.040253

BÖLÜM 6

ÜRETKEN GÖRSEL MODELLERİN KİŞİSELLEŞTİRİLMESİNDE YENİDEN ÜRETİLEBİLİRLİK VE DEĞERLENDİRME

FURKAN GÖZÜKARA[1]

Giriş

Metinden görüntü üreten modeller, belirli bir kişiyi, nesneyi veya tasarımı farklı sahnelerde göstermeyi mümkün kılar. Aynı ürünün değişen kullanım koşullarında sunulması ve bir karakterin farklı kompozisyonlarda tanınabilir kalması, kişiselleştirmenin başlıca kullanım alanları arasındadır. Bu uygulamalarda öznenin ayırt edici özellikleri korunurken duruşunun, aydınlatmasının ve çevresinin değişebilmesi beklenir. Eğitim fotoğrafındaki arka planı veya pozu sürekli yineleyen bir model, yeni sahnelere aktarım bakımından sınırlı kalır. Kişiselleştirme başarısını değerlendirirken özneye sadakat ile eğitim görüntüsüne bağlılığı ayırt etmek gerekir.

DreamBooth ve Textual Inversion, az sayıdaki referans görüntüsünü önceden eğitilmiş bir metinden görüntüye üretim sistemine aktarmanın farklı yollarını sunar (Gal et al., 2023; Ruiz et al., 2023). Düşük ranklı uyarlama, öğrenilen parametre sayısını ve

[1] Dr. Öğr. Üyesi, Mersin Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, ORCID: 0000-0001-9379-2163

saklanan güncellemenin boyutunu azaltır. Görsel istem adaptörleri ise önceden öğrenilmiş bir bağlantı üzerinden referans görüntüsünü üretime katar; böylece her yeni özne için optimizasyon gerekmeyebilir (Hu et al., 2022; Ye et al., 2023). Bu yöntemler aynı üretim iş akışında bir arada kullanılabilse de öğrenme süreçleri ve maliyetleri farklıdır.

Kişiselleştirme yöntemlerini karşılaştırırken sonuçların hangi koşullarda yeniden elde edilebildiği de araştırılır. Yeniden üretilebilirlik, açıklanan veri ve hesaplama düzeniyle çalışmanın temel çıktılarının elde edilebilmesini; sağlamlık ise koşullar makul ölçüde değiştiğinde ana çıkarımın korunmasını ifade eder. Aynı görüntüyü bit düzeyinde üretmek, belirli bir hesaplama düzeninin denetlenmesini sağlar. Uygulama açısından ise yöntemler arasındaki farkın başka tohum değerlerinde (random seed), özne türlerinde ve istemlerde sürüp sürmediği de önem taşır. Bu iki soruyu birlikte ele almak, yeniden üretilebilirlik literatüründe veri, kod ve deney koşullarına verilen önemle uyumludur (Pineau et al., 2021).

Bu bölümde kişiselleştirme mekanizmaları, değerlendirme çalışmaları ve belgeleme ilkeleri literatür üzerinden eleştirel olarak incelenmektedir. Yöntemler öğrendikleri bileşen ve kullandıkları veri bakımından, ölçütler ise yakaladıkları özellikler ve sınırlılıkları bakımından karşılaştırılır. Farklı protokollerle elde edilmiş puanlar ortak bir başarı sıralamasına elvermediği için tartışmanın odağında karşılaştırma koşulları yer alır. Son bölümlerde bu tartışmadan yola çıkılarak karşılaştırmalı bir deneyin veri bölmesi, hesaplama bütçesi ve ölçüm düzeni bakımından nasıl kurulabileceği gösterilmektedir.

Difüzyon Tabanlı Üretimin Teknik Temelleri

Difüzyon modellerinde ileri süreç, veri örneğine adım adım gürültü ekler. Üretim modeli ise bu bozulmayı tersine çevirmeyi öğrenir. Denoising Diffusion Probabilistic Models çalışması, gürültü giderme hedefleriyle yüksek nitelikli görüntüler üretilebildiğini

göstermiştir (Ho et al., 2020). Üretim sırasında örnek, öğrenilmiş dağılım ve koşullandırma bilgisi doğrultusunda ardışık dönüşümlerden geçer. Bu süreç yeni görüntüler oluşturabildiği gibi, bazı koşullarda eğitim örneklerine çok yakın sonuçlar da verebilir.

Gizil difüzyon modellerinde (latent diffusion models) gürültü giderme, piksel uzayı yerine bir kodlayıcının oluşturduğu daha düşük boyutlu gösterimde yürütülür. Kod çözücü, son gizil örneği görüntüye dönüştürür. Bu tasarımla hesaplama maliyeti azaltılırken görsel ayrıntının korunması amaçlanır (Rombach et al., 2022). Üretim sistemi; gürültü tahmin ağı, görüntü kodlayıcısı ve kod çözücüsü, metin kodlayıcısı, sözcükleştirici (tokenizer) ve koşullandırma bağlantılarından oluşur. Bu bileşenlerden biri değişirse aynı uyarlama ağırlıkları farklı sonuçlar verebilir.

Basitleştirilmiş bir gürültü tahmini hedefi şöyle yazılabilir:

$$L = \mathrm{E}[\|\varepsilon - \varepsilon_\theta(z_t, t, c)\|_2^2]$$

Burada z_t, t zaman adımında gürültülenmiş gizil gösterimi; ε eklenen gürültüyü; c koşullandırma bilgisini; ε_θ ise θ parametreli tahmin ağını belirtir. Beklenti, eğitim örnekleri, zaman adımları ve gürültü çekilişleri üzerindedir. Farklı modellerde tahmin hedefi, zamanın parametreleştirilmesi ve eğitim ağırlıklandırması değişebilir. Bu seçimler öğrenme sürecini etkilediğinden, modeller arasındaki farkı yorumlamak için kullanılan düzenin bilinmesi gerekir.

Çıkarımda kullanılan örnekleyici, adım sayısı ve koşullandırma kuvveti sonucu doğrudan etkiler. Sınıflandırıcısız yönlendirme (classifier-free guidance), koşullu ve koşulsuz tahminleri birleştirerek üretimin koşula bağlılığını değiştirir (Ho & Salimans, 2022). Yönlendirme değerinin kimlik sadakati ve görüntü doğallığı üzerindeki etkisi, kullanılan model ve isteme göre değişebilir. Bu ayarlar değiştiğinde aynı kontrol noktası (checkpoint), yani aynı model ağırlıkları farklı görüntüler üretebilir;

bu yüzden uyarlamalar karşılaştırılırken temel model, örnekleyici ve yönlendirme ayarları da kaydedilir.

Kişiselleştirme Yöntemleri ve Karşılaştırma Birimi

Textual Inversion, yeni kavrama ayrılan yer tutucu simgenin (placeholder token) gömme vektörünü öğrenir. Gömme, metin kodlayıcısının girişinde kullanılır ve referans görüntülerindeki difüzyon kaybını azaltacak biçimde optimize edilir; metin kodlayıcısının geri kalanı ile gürültü giderme ağı sabittir (Gal et al., 2023). Özgün yöntemin tek simge gömmesi ile birden çok vektör kullanan uygulamaları, öğrenilen temsilin kapsamı bakımından ayrılır. Kavramın ne ölçüde korunacağı, bu kapasiteyle birlikte veri seçimine ve optimizasyona bağlıdır. Yöntemin yeniden uygulanabilmesi için simge-gömme eşleştirmesi, vektör sayısı, başlatma sözcüğü, istem şablonları ve metin kodlayıcısı bilinmelidir.

DreamBooth, az sayıdaki özne görüntüsünü ayırt edici bir tanımlayıcı ve sınıf adı içeren istemlerle eşleştirerek modeli özneye uyarlar. Özgün yöntemde sınıfa özgü önsel koruma (class-specific prior preservation) için, uyarlama öncesindeki dondurulmuş modelden sınıf adıyla görüntüler üretilir. Bu görüntüler üzerinden hesaplanan yardımcı kayıp, sınıf adının yeni özneye daralmasını ve çeşitlilik kaybını azaltmayı amaçlar; özne kaybına göre ağırlığı ayrı bir katsayıyla belirlenir (Ruiz et al., 2023). Bu nedenle DreamBooth uygulamalarını karşılaştırmak için güncellenen modüller kadar tanımlayıcı-sınıf eşleştirmesi ve önsel koruma düzeni de bilinmelidir. Sınıf görüntülerinin sayısı, üretildikleri kontrol noktası, üretim ayarları ve yardımcı kaybın ağırlığı, aynı yöntem adı altında değişebilen tercihlerdir. Önsel korumanın kaldırılması da özgün yöntemi değiştiren bir uygulama tercihidir.

LoRA, ağırlık güncellemesini düşük ranklı matrislerle temsil eden bir uyarlama tekniğidir. İlk olarak büyük dil modelleri için geliştirilmiş, daha sonra görsel kişiselleştirmede de kullanılmıştır

(Hu et al., 2022). DreamBooth özneye ilişkin eğitim düzenini, LoRA ise güncellemenin nasıl parametreleştirileceğini tanımlar. Bu nedenle DreamBooth benzeri bir eğitim hem tam ağırlık güncellemesiyle hem de LoRA ile yürütülebilir. İkisinin etkisini ayırmak için eğitim hedefi ve veri düzeni sabit tutulup güncellenen parametre uzayı değiştirilebilir. Birden fazla bileşen değişiyorsa sonuç, bu değişikliklerin birlikte etkisini yansıtır.

IP-Adapter, referans görüntüsünü bir görsel kodlayıcıyla temsil eder; bu temsili metin koşulundan ayrılmış çapraz dikkat dallarıyla (decoupled cross-attention) gürültü giderme ağına aktarır. Özgün eğitimde temel difüzyon modeli ve görsel kodlayıcı dondurulurken görüntü projeksiyonu ile ek görsel dikkat parametreleri öğrenilir. Eğitilmiş adaptör, yeni bir özne için yeniden optimizasyon yapılmadan kullanılabilir (Ye et al., 2023). Buradaki eğitim gerektirmeyen aşama, özne başına kullanımdır. Referanstan taşınan içerik ve üslubun özne kimliğini ne ölçüde koruduğu ayrıca değerlendirilir. ControlNet ise kenar, derinlik veya poz gibi uzamsal koşulları üretime bağlar (L. Zhang et al., 2023). Kimlik uyarlamasıyla birlikte kullanıldığında sahnenin yapısını denetlemeyi sağlar. Bu yaklaşımları karşılaştırırken özne başına eğitim, değişen parametreler ve koşullandırma türü ayrı ayrı ele alınır.

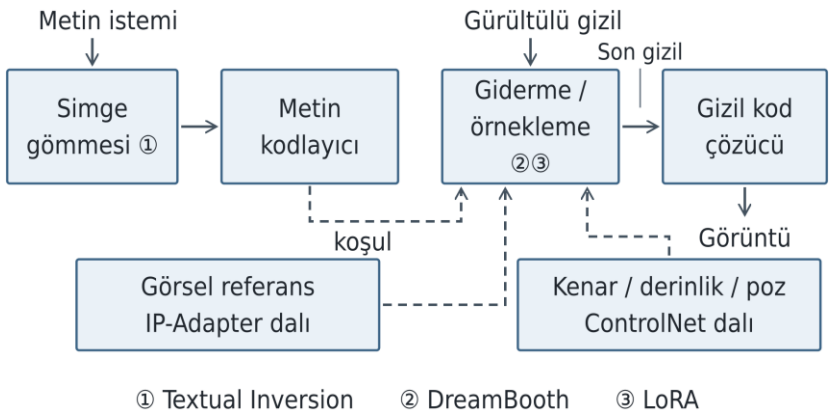
Tablo 1, yöntemleri yeni özne için gereken işlem ve kaydedilecek bilgiler bakımından karşılaştırır. Şekil 1’de bu yöntemlerin üretim hattına nasıl bağlandığı görülmektedir; düz çizgiler ana akışı, kesikli çizgiler koşullandırma bağlantılarını gösterir. Textual Inversion simge gömmesine, DreamBooth ve LoRA ise gösterilen örnekte gürültü giderme ağına uygulanır. IP-Adapter ve ControlNet üretime ek koşullar sağlar. Kullanılan eğitim düzenine bağlı olarak DreamBooth ve LoRA uygulamalarında metin kodlayıcısı da uyarlanabilir.

Tablo 1. Özne uyarlaması ve koşullandırıcı kullanımının ayrımı

Yaklaşım	**Yeni özne için işlem**	**Kritik kayıt**
Textual Inversion	Kavrama ait metin gömmesi optimize edilir	Simge, gömme boyutu, metin kodlayıcı
DreamBooth tabanlı tam uyarlama	Seçilen temel model ağırlıkları optimize edilir	Güncellenen modüller, önsel koruma
LoRA ile özne uyarlama	Düşük ranklı güncellemeler optimize edilir	Hedef modüller, rank, ölçek
Önceden eğitilmiş IP-Adapter	Özneye özgü optimizasyon olmadan görüntüyle koşullandırma	Adaptör sürümü, referanslar, koşul ağırlığı
Önceden eğitilmiş ControlNet	Kenar, derinlik veya pozla uzamsal denetim	Kontrol modeli, ön işleyici, koşul ağırlığı

Kaynak: Gal et al. (2023), Hu et al. (2022), Ruiz et al. (2023), Ye et al. (2023) ve L. Zhang et al. (2023) temel alınarak yazar tarafından hazırlanmıştır.

Şekil 1. Gizil üretim yolunda müdahale noktaları



Kaynak: Gal et al. (2023), Hu et al. (2022), Rombach et al. (2022), Ruiz et al. (2023), Ye et al. (2023) ve L. Zhang et al. (2023) temel alınarak yazar tarafından hazırlanmıştır.

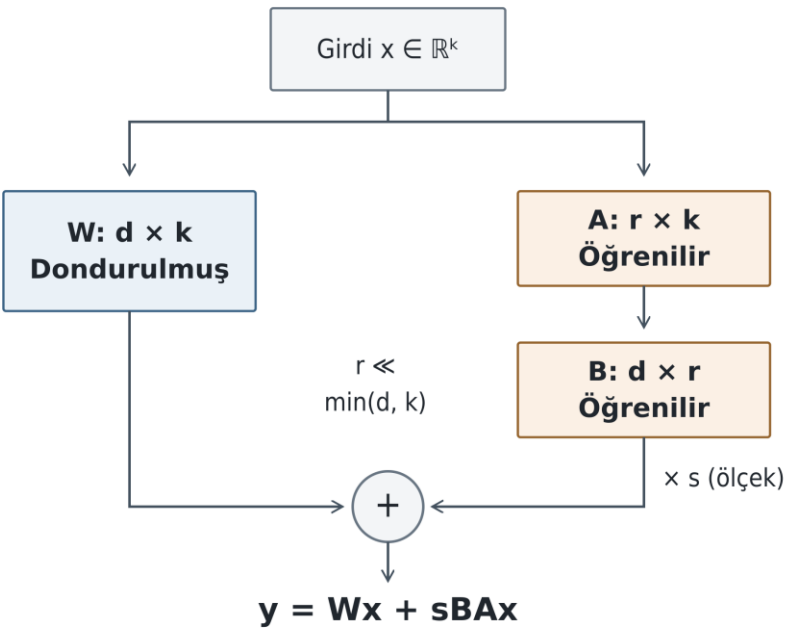
Düşük Ranklı Uyarlamanın Kapasitesi ve Sınırları

LoRA'da bir ağırlık matrisinin güncellemesi, iki küçük matrisin çarpımıyla temsil edilir. Şekil 2'deki paralel hesaplama yollarına karşılık gelen birleşik ağırlık matrisi şöyle yazılır:

$$W' = W + sBA$$

Bu ifadede W, $d \times k$ boyutlu dondurulmuş temel ağırlığı; B, $d \times r$ boyutlu matrisi; A, $r \times k$ boyutlu matrisi ve s ölçek katsayısını temsil eder. r, uyarlamanın rank sınırını belirler. Bu katmanda tam güncelleme $d \times k$ parametre gerektirirken düşük ranklı gösterim $r(d + k)$ öğrenilebilir parametre kullanır (Hu et al., 2022). Buradaki parametre sayısı, tek bir matrisin güncellemesini tanımlar. Toplam bellek gereksinimi ayrıca etkinleştirmelere (activations), temel ağırlıklara ve eğitim uygulamasına bağlıdır.

Şekil 2. LoRA katmanındaki iki paralel hesaplama yolu



Kaynak: Hu et al. (2022) temel alınarak yazar tarafından hazırlanmıştır.

Şekil 2, aynı girdinin hem dondurulmuş temel yolda hem de öğrenilen A ve B matrislerinden geçen ek yolda işlendiğini gösterir; s ile ölçeklenen BA çarpımı, $d \times k$ boyutlu güncellemeyi ($\Delta W = sBA$) oluşturur.

Rank, uyarlamanın temsil kapasitesini belirleyen etmenlerden biridir. Sonuç üzerindeki etkisi, hedef katmanlar, öğrenme oranı, eğitim süresi ve veri çeşitliliğiyle birlikte ortaya çıkar. Yüksek rank daha geniş bir güncelleme uzayı sunar; bu kapasitenin kimlik sadakatine katkısı ise deneysel olarak sınanır. Düşük rankta da aşırı öğrenme görülebilir. Aynı rankla farklı katmanları uyarlayan iki deneyin parametre sayıları ve müdahale alanları değişeceğinden, rank değerine hedef modül listesi ve toplam eğitilebilir parametre sayısı eşlik etmelidir.

Özgün LoRA düzeninde A matrisi rastgele Gauss değerleriyle, B matrisi ise sıfırla başlatılır. Böylece başlangıçtaki BA güncellemesi sıfır olur ve eğitim sırasında α/r ile ölçeklenir (Hu et al., 2022). Yukarıdaki eşitlikte s ile gösterilen eğitim ölçeği ve matrislerin başlatılma biçimi uygulamalar arasında değişebilir. Bu tercihler bilinmeden aynı rank değerine sahip iki uyarlamanın eğitim düzeni eşdeğer kabul edilemez. Çıkarımda adaptör etkisini ayarlamak için kullanılan ek çarpan, eğitim ölçeğinden ayrı bir parametredir. Bu çarpanın azaltılması, çıktıyı temel modelin davranışına yaklaştırabilir; kimlik ve istem uyumu üzerindeki etkisi kullanılan veriye bağlıdır. Birden fazla adaptörün birleştirilmesi veya güncellemenin temel ağırlıklara kaynaştırılması durumunda uygulanan dönüşüm ve etkin toplam ölçek de önem kazanır.

Düşük ranklı güncellemelerin küçük dosyalarda saklanması, adaptörlerin paylaşılmasını kolaylaştırır. Çalıştırılabilmeleri için uygun temel ağırlıklar, mimari, metin kodlayıcısı ve hedef katman eşleştirmeleri gerekir. Adaptör başka bir kontrol noktasına yüklenebilse de temel ağırlıkların değişmesi çıktıları etkileyebilir.

Bu yüzden aktarımın başarısı, yeni modeldeki kimlik sadakati ve istem uyumu ölçümleriyle değerlendirilir.

Veri Kapsamı, Özne Tanımı ve Açıklama Kalitesi

Az görüntüyle yapılan kişiselleştirmede her örnek, öğrenilen temsil üzerinde daha fazla ağırlık taşıyabilir. Veri seçimi bu yüzden öznenin nasıl tanımlandığıyla başlar. Bir üründe model numarasını ayırt eden biçimsel ayrıntılar öne çıkarken, insan yüzünde biyometrik özellikler ve görüntüleme koşulları önem kazanır. Hayvanlar ve çizilmiş karakterler de kendilerine özgü ayırt edici özellikler taşır. Stil öğrenimi ise farklı içeriklerde ortak bir görsel anlatımın korunmasını hedefler. Bu amaçlar ayrı değerlendirme soruları gerektirir; aynı protokolde birleştirildiklerinde puanın neyi ölçtüğü belirsizleşebilir.

Veri çeşitliliği, görüntülerin hangi koşullarda çekildiğiyle yakından ilişkilidir. Aynı çekim dizisindeki çok sayıda kare, farklı açı ve ortamları içeren daha küçük bir koleksiyondan daha az bağımsız bilgi taşıyabilir. Çekim oturumu, arka plan, poz, aydınlatma, ölçek ve örtülme bilgileri bu farkı açıklamaya yardımcı olur. Özne sürekli aynı odada görünüyorsa model, özneyle ortamı birlikte öğrenebilir. Yeni bir arka plan istendiğinde ortaya çıkan başarısızlığı değerlendirirken bu veri bağımlılığını da hesaba katmak gerekir.

Görüntü açıklamaları, özneye ve sahneye ait özelliklerin öğrenilmesini etkiler. Her görüntüye aynı kısa açıklamanın verilmesiyle görünür bağlamın tek tek anlatılması, modele farklı sinyaller sunar. Otomatik açıklama araçları veri hazırlamayı hızlandırabilir; buna karşılık yanlış nesne adları, görünmeyen özellikler veya tekrar eden kalıplar üretebilir. Kullanılan araç ve sürüm ile insan düzeltmesinin kapsamı bilindiğinde, bu hataların sonuçlara olası katkısı incelenebilir.

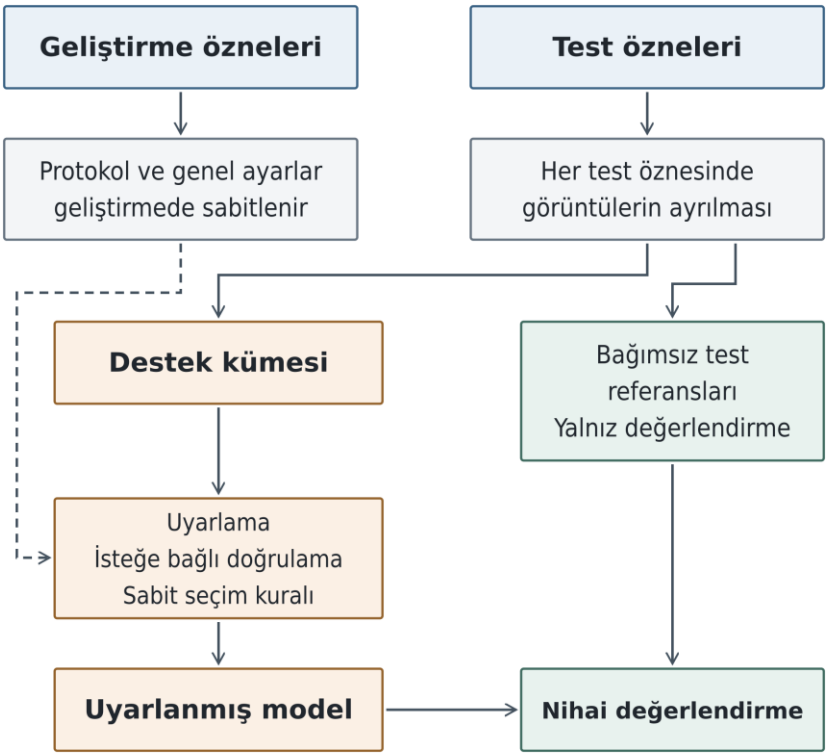
Veri kümesi belgeleri, örneklerin toplanma amacını, kullanım sınırlarını ve olası önyargıları açıklar (Gebru et al., 2021). Kişiselleştirmede bu bilgi; hak sahipliğini, izinleri, kopyaların temizlenmesini, kırpma kurallarını ve dışlanan görüntüleri de kapsar. Mahremiyet nedeniyle görüntüler paylaşılamasa bile seçim ve işleme süreci belgelenebilir. Ancak bağımsız araştırmacının aynı veriye erişememesi, çalışmayı yeniden yürütme olanağını sınırlar.

Veri Sızıntısı, Ezberleme ve Kimliğin Korunması

Kişiselleştirme çalışmalarında veri sızıntısı en az üç düzeyde ortaya çıkabilir. Birincisi, eğitim ve değerlendirme kümeleri arasında aynı görüntünün ya da küçük dönüşümlerle değiştirilmiş kopyalarının bulunmasıdır. İkincisi, ayrı dosyalar olsalar bile aynı çekim dizisinin çok benzer karelerinin iki kümeye dağılmasıdır. Üçüncüsü, temel modelin ön eğitiminde test öznesinin veya ilgili görsellerin yer almış olabilmesidir. İlk iki risk yerel veri düzenlemesiyle azaltılabilirken, üçüncüsü özellikle ön eğitim verisi açıklanmayan modellerde tamamen dışlanamayabilir.

Veri bölmesinin birimi, araştırılan genelleme türüne göre seçilir. Aynı öznenin yeni bağlamlara aktarımında, eğitimde görülmemiş çekim koşullarından referanslar gerekir. Başka öznelere genelleme araştırılıyorsa hiperparametre seçiminde kullanılan özneler son testten ayrılır. Şekil 3, özne düzeyindeki bu ayrımı her test öznesi içindeki görüntü bölmesiyle birlikte gösterir. Çekim oturumu veya kaynak koleksiyonu üzerinden bölme, birbirine çok yakın karelerin eğitim ve teste dağılmasını azaltır. Yalnızca dosyaları rastgele ayırmak, bu bağımlılıkları koruyabilir ve bilinen öznelere göre ayarlanmış bir işlem hattının başarısını olduğundan geniş yorumlamaya yol açabilir.

Şekil 3. Özne ve görüntü düzeyinde veri ayrımı



Kaynak: Pineau et al. (2021) ve Somepalli et al. (2023) temel alınarak yazar tarafından hazırlanmıştır.

Şekil 3'teki şemada geliştirme özneleriyle belirlenen protokol, test aşamasındaki uyarlama ve seçim işlemlerinde değiştirilmeden kullanılır. Her test öznesinin görüntüleri destek kümesine ve bağımsız değerlendirme referanslarına ayrılır. Erken durdurma veya kontrol noktası seçimi gerekiyorsa doğrulama için destek kümesinin bir bölümü kullanılır; bu görüntüler gradyan güncellemelerine katılmaz. Seçim kuralı test başlamadan sabitlendiği için bütün test öznelerine aynı biçimde uygulanır.

Özneye sadakat ile eğitim görüntüsünün tekrarı, farklı değerlendirmeler gerektirir. Bir çıktı, eğitim fotoğrafındaki yüzü çok

iyi korurken aynı duruş ve kompozisyona bağlı kalabilir. Difüzyon modellerinden eğitim örneklerinin çıkarılabildiği de gösterilmiştir (Carlini et al., 2023). Bu çalışmanın Stable Diffusion deneyi, eğitim verisinde en çok yinelenen görüntülere öncelik veren hedefli bir aramaya dayanır. Bildirilen başarı oranı bu arama koşulları için geçerlidir; rastgele bir istemin sızıntı olasılığına veya küçük bir kişiselleştirme deneyine doğrudan genellenemez. Kişiselleştirilmiş modelin eğitim örneklerini ne ölçüde tekrarladığı, kendi verisi ve kullanım koşulları içinde araştırılmalıdır.

Üretilen görüntülerin eğitim kümesindeki yakın komşularını incelemek, bu araştırma için pratik bir başlangıçtır. Görsel gömme benzerliği, algısal uzaklık ve insan incelemesi birlikte kullanılabilir. Ancak bir özne türünde anlamlı olan yakınlık eşiği, başka bir türde aynı yorumu desteklemeyebilir; eşik seçiminin gerekçesi bu yüzden önemlidir. Taramanın bulguları kullanılan temsil, arama koleksiyonu ve yöntemle sınırlıdır. Eşleşme bulunmaması yalnızca bu kapsamda şüpheli yakınlık saptanmadığını gösterir. Temsilin ayırt edemediği ayrıntılar veya farklı çıkarım saldırıları için aynı sonuca varılamaz.

Hesaplama Bütçesi ve Bellek Ödünleşimleri

Kişiselleştirmenin bellek maliyeti, eğitim boyunca bellekte tutulan bileşenlerin toplam kullanımına bağlıdır. Temel ağırlıklar, eğitilebilir ağırlıklar, gradyanlar, iyileştirici (optimizer) durumları, etkinleştirmeler ve geçici çalışma alanları bu kullanıma katkıda bulunur. Dondurulan ağırlıklar için gradyan ve iyileştirici durumu tutulmasına gerek kalmaz; buna rağmen öğrenilebilir bileşenlere ulaşan hesaplama yollarında ara değerler gerekebilir. LoRA'nın sağladığı parametre tasarrufunun toplam belleğe ve süreye ne ölçüde yansıdığı, model ve eğitim düzeni üzerinde ölçülür.

Şekil 4, eğitim ve standart çıkarım sırasında bellekte bulunan bileşenleri ayırır. Eğitimde gradyanlar ve iyileştirici durumları da tutulurken standart çıkarım bunlara ihtiyaç duymaz. Her bileşenin

kapladığı alan, kullanılan model ve uygulamaya göre değişir. Tepe bellek, ölçüm aralığında toplam kullanımın ulaştığı en yüksek değerdir. Bileşenlerin en yüksek kullanımları farklı anlara denk gelebileceği için bu değerleri ayrı ayrı toplamak tepe belleği vermez.

Şekil 4. Eğitim ve çıkarım belleğinin kavramsal bileşenleri



Kaynak: Chen et al. (2016) ve Hu et al. (2022) temel alınarak yazar tarafından hazırlanmıştır.

Karma duyarlık (mixed precision), etkinleştirmelerin yeniden hesaplanması (activation checkpointing), verimli dikkat işlemleri ve CPU belleğine aktarma (CPU offloading) farklı bellek kalemlerini etkiler. Örneğin etkinleştirmeleri yeniden hesaplamak belleği azaltırken işlem süresini uzatabilir. Gradyan biriktirme, etkin yığın boyutunu (effective batch size) mikro-yığın boyutundan ayırır; toplam işlem yükü ise işlenen mikro-yığınlara bağlıdır. Mikro-yığın boyutu, biriktirme adımı sayısı ve aygıt sayısı bilinmeden etkin yığın boyutu yorumlanamaz. Aynı eğitim düzeni farklı donanımlarda

farklı süre ve bellek sonuçları verebildiğinden, bir aygıtta ölçülen maliyet diğerine doğrudan taşınamaz.

Maliyet hesabının kapsamı, karşılaştırmanın amacına göre belirlenir. Model indirme, veri ön işleme, sınıf görüntüsü üretimi, eğitim, doğrulama ve adaptör kaydetme süreleri ayrı kalemler olarak verilebilir ya da tanımlı bir toplamda birleştirilebilir. Önceden eğitilmiş görsel adaptörlerde özne başına işlem kısa sürebilir; adaptörün ön eğitimi ise ayrı bir maliyettir. Kullanıcı açısından yeni bir özne eklemenin maliyeti, yöntem geliştirme açısından toplam kaynak tüketimi öne çıkabilir. Bu ayrım, farklı kapsamlarla hesaplanmış sürelerin aynı tabloda karıştırılmasını önler.

Süre ve bellek sonuçlarının yorumlanabilmesi için donanım modeli, bellek kapasitesi, sayısal duyarlık, yazılım sürümleri ve ölçüm yöntemi bilinmelidir. Tepe kullanım kısa süreli bellek gereksinimini, ortalama kullanım ise zaman içindeki genel düzeyi anlatır. Enerji tüketimi, ölçüme katılan donanım bileşenlerine ve ölçüm süresine; parasal maliyet ise bunların yanında fiyatlandırmaya bağlıdır. Üreticinin belirttiği güç değeri, deney sırasında tüketilen enerjiyi tek başına göstermez. Bu nedenle enerji karşılaştırması, kapsamı tanımlanmış tüketim ölçümlerine dayanmalıdır.

Çıkarım Yapılandırmasının Sonuçlara Etkisi

Eğitim sonrasında seçilen örnekleyici, gürültü çizelgesi, üretim adımı, yönlendirme katsayısı ve görüntü boyutu aynı adaptörün çıktısını değiştirebilir. DDIM, eğitim hedefiyle ilişkili farklı bir üretim süreci kurarak örnekleme verimliliğinin değişebildiğini göstermiştir (Song et al., 2021). Kişiselleştirme karşılaştırmasında çıkarım ayarlarının sabit tutulması, uyarlama farkının yorumlanmasını kolaylaştırır. Çıkarım yönteminin etkisi de araştırılıyorsa bu ayarlar ayrı deney etmenleri olarak ele alınır.

Aynı tohum değeri, farklı işlem hatlarında farklı rastgele sayı akışlarıyla sonuçlanabilir. Rastgele sayı üretecinin çalıştığı aygıt,

çağrı sırası ve araya giren işlemler bu akışı etkiler. Ortak başlangıç gürültüsüyle karşılaştırma yapılacaksa tohum listesine ek olarak gizillerin oluşturulma yolu da eşleştirilir ve denetlenir. Başlangıç tensörlerinin saklanması bu denetimi kolaylaştırabilir. Mimariler veya gizil boyutlar farklı olduğunda ise bire bir gürültü eşlemesinin uygulanabilirliği ayrıca değerlendirilir.

İstem seçimi, yöntem karşılaştırmasının sonucunu etkileyebilir. Yönteme özel uzun açıklamalar, negatif istemler veya elle seçilmiş anahtar sözcükler kullanıldığında, uyarlama yöntemiyle istem mühendisliğinin katkıları iç içe geçer. Ortak istemlerle yapılan deney kontrollü bir karşılaştırma sunarken, her yönteme göre ayarlanmış istemler uygulamada erişilebilen sonucu gösterir. Bu iki düzen ayrı raporlanabilir. Her iki durumda da yalnızca en iyi tohumu veya başarılı örnekleri seçmek, üretimin genel davranışını olduğundan iyi gösterebilir.

Temel mimari değiştiğinde uygun çıkarım düzeni de değişebilir. SDXL, metin kodlayıcıları, koşullandırması ve yüksek çözünürlüklü üretimi bakımından önceki gizil difüzyon uygulamalarından ayrılır (Podell et al., 2024). Çalışmada, temel modelin gizil çıktısını işleyen isteğe bağlı bir ayrıntı iyileştirme modeli (refiner) de bulunur. Bu aşamanın eklenmesi hem üretim düzenini hem de hesaplama maliyetini değiştirdiğinden, temel modelin tek başına kullanıldığı koşuldan ayrı değerlendirilir. Model ailelerini ayrı incelemek, aynı adaptör adı veya rank altında değişen kapasite ve hesaplama gereksinimlerini ayırt etmeyi sağlar.

Kişiselleştirmede Değerlendirme Boyutları

Kişiselleştirme başarısı birkaç farklı özellik üzerinden değerlendirilir. Kimlik sadakati, hedef öznenin ayırt edici özelliklerinin korunmasıdır. İstem uyumu, talep edilen eylem, ortam, görünüş ve ilişkilerin gerçekleşmesini anlatır. Çeşitlilik, aynı koşullarda üretilen örneklerin birbirinden ne ölçüde ayrıldığını

gösterir. Görsel bütünlük değerlendirmesinde bozuk anatomi, tutarsız geometri ve anlamsız ayrıntılar incelenir. Eğitim örneklerine aşırı yakınlık ile hesaplama maliyeti de sonucun kullanılabilirliğini etkiler.

Bu özellikler her zaman birlikte iyileşmez. Referans fotoğraftaki yüzü ve duruşu koruyan bir model, istenen yeni hareketi gerçekleştiremeyebilir. Bir başka model kompozisyonu doğru kurarken nesnenin ayırt edici parçalarını değiştirebilir. Yöntem tercihi bu yüzden kullanım amacına bağlıdır. Sonuçlar ağırlıklı bir ortalamayla özetleniyorsa, ağırlıkların hangi tercihi temsil ettiği ve her bileşenin ayrı sonucu da gösterilmelidir.

CLIP, metin ve görüntüleri ortak bir temsil uzayında ilişkilendirir (Radford et al., 2021). Benzerlik puanı, kategori ve sahne özelliklerine duyarlı olabilir; kişinin yüzündeki veya üründeki ince kimlik farklarını kaçırabilir. Bu nedenle CLIP ile ölçülen genel benzerlik, kimlik değerlendirmesinde ek incelemeyle desteklenir. CLIPScore ise ilk olarak görüntü açıklamalarını değerlendirmek için geliştirilmiştir (Hessel et al., 2021). Kişiselleştirmede kullanıldığında puanın hangi hedef için yorumlandığı açık olmalıdır.

CLIP-I, görüntü gömmeleri arasındaki; CLIP-T ise görüntü ve metin gömmeleri arasındaki benzerliği ölçer. CLIPScore'un puanlama adımları ham kosinüs benzerliğinden ayrılır. Kodlayıcı kontrol noktası, görüntü ön işlemesi, metin şablonları veya puan dönüşümleri değiştiğinde aynı adla raporlanan ölçümler farklı hesaplara dayanabilir. Puanları karşılaştırabilmek için ölçütün adı kadar bu hesaplama tercihlerinin de bilinmesi gerekir.

DINO, özdenetimli görsel temsiller öğrenir (Caron et al., 2021). DreamBooth çalışmasında özne sadakati, üretilen ve gerçek görüntülerin ViT-S/16 DINO gömmeleri arasındaki bütün çiftlerin kosinüs benzerliği ortalanarak değerlendirilmiştir (Ruiz et al., 2023). Bu kullanımda görsel temsilin hangi kimlik ayrıntılarına duyarlı

olduğu önem taşır. LPIPS de görüntüler arasındaki algısal farklılığı değerlendirmek için kullanılabilir (R. Zhang et al., 2018); farklı poz ve arka planların yol açtığı değişim, kimlikteki değişimle birlikte puana yansıyabilir. Araştırılan özelliğe uygun bir ölçüt seçildiğinde bile hesaplama düzeni sonucu etkileyebilir. DINO tabanlı ölçümlerde kodlayıcı sürümü, kırpma, yeniden boyutlandırma, referans-çıktı çiftlerinin seçimi ve ortalama alma düzeyi bu düzenin başlıca öğeleridir. Örneğin yalnızca en benzer referansı seçmekle bütün referansların ortalamasını almak farklı sorulara yanıt verir.

Tablo 2, değerlendirme boyutlarını ölçüm örnekleriyle eşleştirir. Her ölçümün yorum sınırı, aynı görüntü hakkında başka hangi kanıtlara ihtiyaç duyulduğunu gösterir.

Tablo 2. Değerlendirme eksenleri ve yorum sınırları

Eksen	**Ölçüm örneği**	**Kritik sınır**
Özne sadakati	Bağımsız referanslarla CLIP-I ve DINO	Sahne benzerliği kimlik hatasını örtebilir.
İstem ve bileşim	CLIP-T; eylem ve ilişkiler için görev denetimi	Genel uyum puanı ilişki hatalarını kaçırabilir.
Çeşitlilik	Sabit özne ve istemde tohumlar arası LPIPS	Yüksek uzaklık kimlik kaybından da doğabilir.
Eğitim örneğine yakınlık	Yakın komşu taraması ve görsel inceleme	Denetimin kapsamı temsil ve arama kümesiyle sınırlıdır.
Hesaplama maliyeti	Uyarlama ve çıkarım için ayrı süre ve tepe bellek	Kapsam ve donanım farkları sonucu etkiler.

Kaynak: Ruiz et al. (2023), Somepalli et al. (2023) ve R. Zhang et al. (2018) temel alınarak yazar tarafından hazırlanmıştır.

Dağılım Ölçütleri, İnsan Değerlendirmesi ve Belirsizlik

Fréchet Inception Distance (FID), üretilen ve gerçek görüntülerin öznitelik dağılımlarını karşılaştırır (Heusel et al., 2017). Kişiselleştirmede az sayıda özne görüntüsünden hesaplanan FID, örnek sayısına ve seçilen öznitelik çıkarıcıya duyarlıdır. Puanın

kapsamı, karşılaştırılan dağılımlardır; tek bir öznenin kimliğini veya istemdeki ilişkilerin doğruluğunu değerlendirmek için başka ölçümler gerekir. Sonlu örneklem yanlılığı da modele göre değişir. Eşit sayıda örnekle yapılan karşılaştırmalarda bile bu yanlılık yöntem sıralamasını etkileyebilir (Chong & Forsyth, 2020). FID bu yüzden örnek sayısı ve tahmin yöntemiyle birlikte yorumlanır.

Üretim niteliğini ve kapsama alanını ayrı ölçmek, dağılımlar arasındaki farkı açıklamaya yardımcı olur. Kynkäänniemi ve arkadaşlarının yaklaşımında kesinlik (precision), üretilen örneklerin gerçek görüntülerden kestirilen öznitelik bölgesine düşen oranıdır. Kapsama (recall) ise gerçek örneklerin üretilen görüntülerden kestirilen bölgeye düşen oranını verir. Bölgeler, öznitelik uzayındaki k-en yakın komşu uzaklıklarıyla kurulur (Kynkäänniemi et al., 2019). Öznitelik çıkarıcı, örnek sayısı ve k komşuluk parametresi bu nedenle sonuçlarla birlikte belirtilir. Bu puanlar dağılımın niteliğini ve kapsamını anlatır; özne kimliği ve istem uyumu ayrı değerlendirmelerdir. Küçük özne kümelerinde kestirimlerin belirsizliği artabileceğinden, ölçütün araştırma sorusuna ve örneklem büyüklüğüne uygunluğu incelenmelidir.

İnsan değerlendirmesinde kimlik, istem uyumu ve görsel hata için ayrı görevler tanımlanması, genel bir beğeni sorusundan daha ayrıntılı bilgi verir. Yöntem adlarının gizlenmesi, görüntü sırasının rastgeleleştirilmesi ve eşdeğer referans bilgisinin sunulması yanlılığı azaltmaya yardımcı olur. Referans eşliğinde yapılan kimlik değerlendirmesi, referans gösterilmeyen görevden ayrılır. Puan ölçeğinin, yönergenin, eşitlik ve değerlendirilememe seçeneklerinin, katılımcılara atama düzeninin ve eksik puanlara ilişkin kararların değerlendirme başlamadan önce belirlenmesi, sonuçlar görüldükten sonra yapılacak seçimlerin etkisini sınırlar. Aynı kişinin verdiği puanlar ve aynı çıktıya verilen değerlendirmeler birbiriyle ilişkili olabilir; analiz bu bağımlılıkları hesaba katar. Değerlendiriciler arası uyum, uygun bir ölçüt ve belirsizlik aralığıyla özetlenebilir. Ölçülen

görevin özne kimliğini gerçekten yansıtıp yansıtmadığı ise ayrıca incelenir. Yüz görüntülerinde katılımcı ve görüntü sahibi izinleri, saklama süresi ve erişim kapsamı da planlanmalıdır.

Sonuçlardaki belirsizliğe eğitim ve üretim tohumları, özne seçimi, test istemleri ve değerlendiriciler katkıda bulunur. Tek bir özne için daha fazla görüntü üretmek, o özneye ilişkin tahmini iyileştirebilir; yeni öznelere genellemeyi değerlendirmek için özne çeşitliliği gerekir. Analizin hedefi bu ayrımı belirler. Seçilmiş özneler ve sabit istemlerle sınırlı bir karşılaştırmanın belirsizlik hesabı, yeni öznelere veya istemlere genellenen bir karşılaştırmadan farklıdır. Yeni öznelere genellemede özne temel örnekleme birimi olarak alınabilir. Her özneye eşit ağırlık vermekle bütün çıktıları tek ortalamada toplamak, özne başına çıktı sayıları farklıysa farklı sonuçlar verebilir. Bu nedenle özne içi ortalamalarla genel ortalamanın nasıl hesaplandığı, tahminin neyi temsil ettiğini belirler.

Eğitim tekrarları aynı özne ve yöntem içinde kümelenirken ortak istemler birden fazla öznede kullanılarak çapraz bir düzen oluşturabilir. Benzer biçimde, aynı değerlendiricilerin birçok çıktıyı puanlaması değerlendirici ve çıktı etkilerini çaprazlar. Kümeli veya çapraz etmenli analiz seçimi, bu gerçek atama düzenine dayanır. Eşleştirilmiş karşılaştırmalarda yöntem farkları aynı özne ve test koşulu içinde hesaplanır; yeniden örnekleme de eşleşmeyi korur. Eşleştirmenin dayanağı aynı tohum numarasının kullanılması değil, karşılaştırılan koşulların ortak olmasıdır. Etki büyüklükleri ve güven aralıkları, kapsadıkları değişkenlik kaynakları ve sabit tuttukları koşullarla birlikte sunulur.

Yeniden Üretim Paketinin Asgari İçeriği

Yeniden üretim paketi, modelden sonuç tablosuna kadar izlenen işlemleri bir araya getirir. Temel modelin tam kimliği ve mümkünse dosya özeti, adaptörün uygulandığı bileşenler, eğitim ve çıkarım ayarları, veri işleme adımları ve değerlendirme betikleri

pakette yer alır. Sonuçlar ilgili paket sürümüne bağlanır. Model kartları kullanım bağlamını ve sınırları açıklamak için yararlıdır (Mitchell et al., 2019); ayrıntılı yürütme kayıtları da deneyin yeniden çalıştırılmasını sağlar.

Tohum bilgisi, çalışmanın yazılım ve donanım ortamıyla birlikte anlam kazanır. Aygıtlar, çekirdek uygulamaları, paralellik düzenleri ve sayısal duyarlık, işlem sırasını ve yuvarlamayı değiştirebilir. PyTorch'un resmî belgesi, aynı tohumla bile sürümler, platformlar veya CPU ve GPU arasında tam eşleşmenin sağlanamayabileceğini belirtir (PyTorch contributors, 2026). Denetimin amacı buna göre tanımlanır: aynı dosyayı elde etmek, benzer bir ölçüt aralığına ulaşmak veya ana bilimsel çıkarımı yeniden doğrulamak farklı başarı koşullarıdır. Bir sonucun yalnızca belirli bir yazılım ortamında elde edilebilmesi de genellenebilirliğini sınırlar.

Görüntülerin yayımlanamadığı çalışmalarda veri şeması, işleme kodu, izin verilen örnekler ve erişim koşulları paylaşılabilir. Açık, başvuruya bağlı ve kapalı bileşenlerin ayrılması, bağımsız araştırmacının hangi adımları tekrarlayabileceğini gösterir. Başvuruyla erişimde çalışan bir talep kanalı, değerlendirme ölçütleri ve izin verilen kullanım kapsamı gerekir. Ham görüntüler yerine sunulan özet değerlerin, açıklamaların veya sentetik örneklerin nasıl üretildiği ve deneyle ilişkisi bilindiğinde, bağımsız araştırmacı bunlarla hangi adımları denetleyebileceğini değerlendirebilir. Bu malzemeler süreci açıklamaya yardımcı olsa da aynı veriyle yeniden yürütmenin yerini tutmaz.

Tablo 3, kişiselleştirme deneylerini yeniden yürütmek ve sonuçlarını değerlendirmek için gereken başlıca kayıtları toplar. Eğitim ve çıkarımın ayrı ele alınması, hangi kaydın hangi aşamaya ait olduğunu izlemeyi kolaylaştırır. Kayıtların kapsamı da yönteme bağlıdır: özneye özgü optimizasyon içermeyen bir kolda, o özneye ait bir eğitim tohumu bulunmaz.

Tablo 3. Kişiselleştirme deneyleri için önerilen kısa raporlama listesi

Bileşen	**Asgari kayıt**
Veri ve izin	Kaynak, izin kapsamı, bölme birimi, tekrar temizliği
Model ve uyarlama	Tam sürüm, temel ağırlık özeti, hedef modüller, rank
Eğitim	Tohum, duyarlık, mikro-yığın, biriktirme, adım, iyileştirici
Çıkarım	İstemler, negatif istem, örnekleyici, yönlendirme, tohumlar
Değerlendirme	Ölçüt sürümleri, çift/örnek seçimi, bağımsızlık birimi, belirsizlik, insan görevi
Maliyet ve erişim	Donanım, süre kapsamı, tepe bellek, kod ve veri erişimi

Kaynak: Gebru et al. (2021), Mitchell et al. (2019) ve Pineau et al. (2021) temel alınarak yazar tarafından hazırlanmıştır.

Karşılaştırmalı Deney Tasarımı ve Veri Bölmesi

Buraya kadar tartışılan ayrımlar, aynı temel model ve izinli referans verileriyle yürütülen karşılaştırmalı bir deneyde sınanabilir. Böyle bir deneyde temel soru, uyarlama biçiminin kimlik sadakatini, istem uyumunu ve hesaplama maliyetini nasıl etkilediğidir.

Araştırmanın insan yüzüne, nesnelere veya başka bir özne grubuna odaklanması, kimlik ölçümünü ve örneklem seçimini etkiler. Birden fazla alan kapsanıyorsa sonuçlar özne gruplarına göre ayrı incelenir. Özne ve görüntü sayısı seçilirken izinli veri kapsamı ile hedeflenen belirsizlik düzeyi dikkate alınır. Donanım kapasitesi uygulamayı sınırlar; ancak örneklemin yeterliliği, karşılaştırmanın hangi birim üzerinden yapıldığına ve her birimden ne kadar bağımsız bilgi elde edildiğine bağlıdır.

Geliştirme özneleri, hiperparametreleri, görüntü seçme kurallarını ve durdurma politikasını belirlemek için kullanılır; son test özneleri bu seçimlere katılmaz. Şekil 3'teki gibi her test öznesinin görüntüleri de destek kümesine ve bağımsız değerlendirme referanslarına ayrılır. Destek görüntüleri, optimizasyon kullanan kollarda uyarlama verisi; görsel adaptör kolunda çıkarım koşulu olur. Referans sayısı veya görüntüleri birleştirme biçimi yöntemler arasında değiştiğinde, gözlenen farkın bir bölümü sunulan bilginin miktarından ya da kullanımından kaynaklanabilir. Destek ve değerlendirme görüntülerinin çekim oturumu ya da kaynak koleksiyonu üzerinden ayrılması, yakın kopyaların iki kümeye dağılmasını azaltır; kırpılmış veya yeniden sıkıştırılmış kopyalar da denetlenir.

Test öznesine uyarlama, önceden belirlenen adım veya bütçe sınırında durur. Özneye özgü erken durdurma gerekiyorsa destek kümesinden ayrılan doğrulama görüntülerine, geliştirme aşamasında sabitlenmiş karar kuralı uygulanır. Son test çıktıları ve bağımsız değerlendirme referansları seçim sürecinin dışında kalır. Bağımsız referansların bulunmadığı bir çalışmada ise benzerlik ölçümü daha iyimser olabilir ve bu veri bağımlılığı yorumun kapsamını daraltır.

Test istemleri olağan görünüşü, yeni arka planı, alışılmadık görüş açısını, eylemi, nesneler arası ilişkileri ve kısmi örtülmeyi kapsayabilir. Kolay ve zorlayıcı koşulların ayrı gruplarda incelenmesi, yöntemin nerede başarısız olduğunu gösterir. GenEval, sayım, renk ve konum gibi nesne düzeyindeki istem bileşenlerini ayrı değerlendirir (Ghosh et al., 2023). Bu tür görevler bileşimsel uyumu ölçer; özne kimliği ise ayrı bir değerlendirme gerektirir.

Uyarlama Kolları ve Hesaplama Bütçesi

Karşılaştırmaya metin gömmesi öğrenimi, seçilmiş ağırlıkların tam uyarlanması, aynı özne hedefinin LoRA ile eğitilmesi ve önceden eğitilmiş bir görsel adaptörle çıkarım dâhil

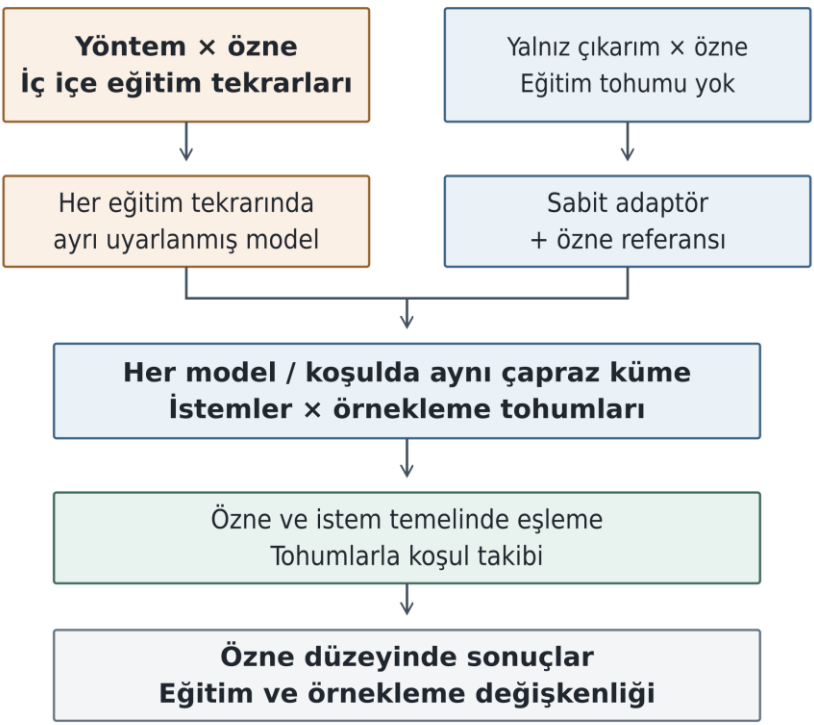
edilebilir. Bu dört kol, öğrenilen temsil ve koşullandırma yolu bakımından farklı bütün işlem hatlarını karşılaştırır. Belirli bir mekanizmanın etkisini incelemek için daha dar bir deney gerekir. Örneğin özne hedefi, önsel koruma ve veri düzeni korunarak tam güncellemeyle LoRA güncellemesi karşılaştırılabilir. Bütçe sınırlı bir deney ise her işlem hattının belirli kaynaklarla ulaşabildiği sonucu araştırır. Bileşen etkisi, işlem hattının başarısı ve bütçe altında erişilen kalite, bu nedenle ayrı araştırma sorularıdır.

Tam uyarlama ile LoRA karşılaştırmasında tanımlayıcıların, açıklamaların, önsel korumanın ve doğrulama istemlerinin ortak tutulması, parametreleştirme farkını incelemeyi kolaylaştırır. Öğrenme oranı gibi ayarların uygun değerleri ise yöntemler arasında değişebilir. Her yöntemin aynı arama hakkıyla geliştirme verisinde ayarlanması, elverişsiz bir yapılandırmayla sınanma riskini azaltır. Arama alanının ve bütçesinin önceden belirlenmesi, yöntemlere ayrılan ayarlama olanağını görünür kılar. Son test çıktılarının ayar seçiminde kullanılması ise bu karşılaştırmanın bağımsızlığını bozar.

Deneyin büyüklüğü yöntem, özne, uyarlama tekrarı, istem ve çıkarım tohumu sayılarına bağlıdır. Temel model, metin ve görsel kodlayıcılar, adaptör ağırlıkları ve yazılım sürümleri her kolda birlikte çalışmalıdır; başka bir temel model gerektiren kol, ortak model karşılaştırmasının dışında tutulur. Geliştirme özneleriyle yürütülen küçük ölçekli bir pilot çalışma, bu uyumu ve süre ile bellek gereksinimini sınamaya yarar. Pilot çalışmadaki ölçümler, ayarlama, son uyarlama, üretim ve insan değerlendirmesi için gereken kaynakların kestirilmesine de yardımcı olur. Kaynaklar bütün koşullara yetmiyorsa ana karşılaştırma korunarak ikincil koşullar azaltılabilir. Bu daraltmanın test sonuçları görülmeden yapılması, kapsamın sonuçlara göre ayarlanmasını önler. Özne sayısını azaltıp aynı özne için daha çok çıktı üretmek ise yeni öznelere genellemeye ilişkin kanıtı zayıflatır.

Şekil 5'te eğitim tekrarları her yöntem ve özne içinde kümelenir; her model veya koşul aynı istem ve çıkarım tohumu kümesiyle sınanır. Özneye özgü optimizasyon içeren kollarda bağımsız eğitim tekrarları, uyarlama sürecinin değişkenliğini ölçmeye olanak verir. Yalnız çıkarım kolundaki tekrarlar ise örnekleme değişkenliğini ölçer. Yöntemler aynı özne ve istem koşulları üzerinden eşleştirilir. Tohumlar deney koşullarını izlemek için kullanılır; başlangıç gürültüsünün de ortak olduğu ancak ayrıca doğrulanmışsa kabul edilir.

Şekil 5. İç içe eğitim tekrarları ve çapraz örnekleme düzeni



Kaynak: Pineau et al. (2021) ve PyTorch contributors (2026) temel alınarak yazar tarafından hazırlanmıştır.

Adım sayısı, süre ve enerji farklı bütçe tanımlarıdır. Adım maliyeti yöntemler arasında değiştiğinden eşit adımlı deneyler farklı

kaynak tüketebilir. Veri yükleme ve doğrulama sürelerinin kapsama alınması, aynı eğitimin bildirilen maliyetini değiştirebilir. Bu yüzden süreye dayalı bir karşılaştırmada ölçümün başlangıç ve bitiş noktaları da önemlidir. Çıkarım tabanlı yöntemde özne başına eğitim maliyeti yoktur; adaptörün ön eğitimi ayrı bir kalem olarak ele alınır. Birden fazla bütçe düzeyinde elde edilen nitelik-maliyet eğrileri, bu farkların uygulamadaki etkisini değerlendirmeyi kolaylaştırır.

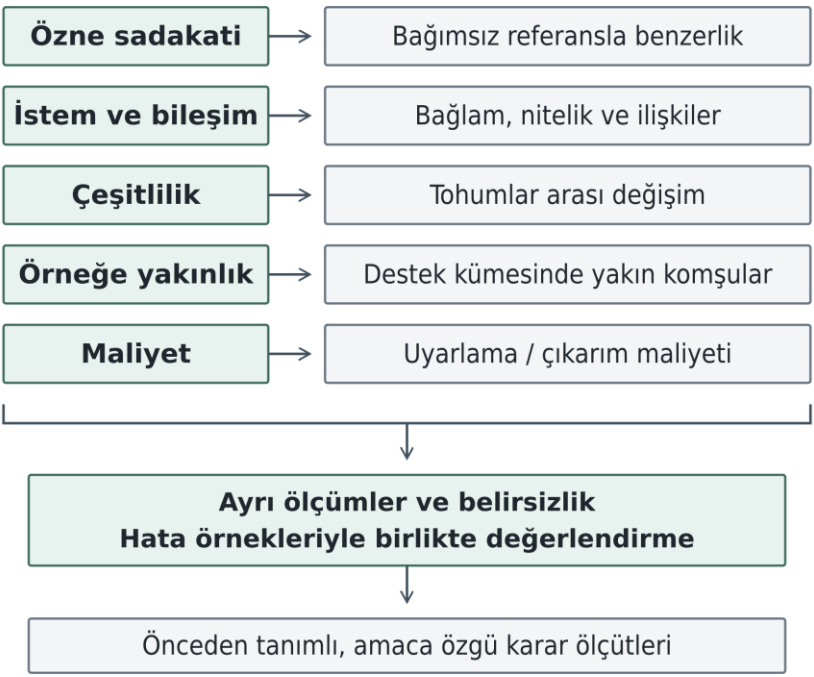
Bellek azaltma tekniklerinin desteği kullanılan yönteme ve uygulamaya göre değişebilir. Etkinleştirmelerin yeniden hesaplanması, bazı ara değerleri saklamak yerine geri yayılımda yeniden üretir (Chen et al., 2016). Sekiz bitlik iyileştiriciler ise iyileştirici durumlarını sıkıştırır; model ağırlıklarının ve etkinleştirmelerin duyarlığı ayrı ayarlara bağlıdır (Dettmers et al., 2022). Yalnızca bir kolun belirli bir bellek azaltma tekniğini desteklemesi, o kola uygulamadan kaynaklanan bir maliyet avantajı sağlayabilir. Kullanılan tekniklerin bilinmesi, bu avantajın uyarlama yönteminin etkisinden ayırt edilmesine yardımcı olur.

Ölçüm ve Sonuçların Yorumlanması

Tekrar düzeni, hangi değişkenlik kaynağının ölçüleceğine göre kurulur. Özne başına optimizasyon içeren kollarda bağımsız uyarlama tekrarları eğitim sürecinin değişkenliğini, her modelden farklı tohumlarla üretim yapılması ise örnekleme değişkenliğini gösterir. Tek bir adaptörden alınan çok sayıda görüntü, eğitim tekrarları hakkında bilgi sağlamaz. Önceden eğitilmiş sabit görsel adaptör kolunda karşılaştırma, aynı özne ve istemlerle yapılan üretim tekrarlarına dayanır. Destek görüntüsü seçiminin etkisi araştırılacaksa yeniden seçim kuralı ve tekrar sayısı bütün ilgili kollarda tanımlanır; bu değişkenlik eğitim rastgeleliğinden ayrı incelenir. Kaynak sınırları nedeniyle yapılamayan tekrarlar, sonucun hangi belirsizlikleri kapsamadığını belirler.

Görsel benzerlik, metin-görüntü uyumu, bileşim, çeşitlilik ve maliyet ayrı sonuçlar olarak değerlendirilir (Şekil 6). Çeşitlilik için aynı özne, istem ve model ayarlarında farklı tohumlarla üretilmiş görüntü çiftlerinin algısal uzaklığı kullanılabilir. Böylece istem değişiminden kaynaklanan fark, aynı istem altındaki örnekleme çeşitliliğinden ayrılır. Çift seçimi, örnek sayısı, ölçüt sürümü ve özne düzeyinde ortalama alma kuralı ölçümü tanımlar. Yüksek uzaklık, kimlik kaybı veya görsel bozulmayla da oluşabileceğinden kimlik ve hata sonuçlarıyla birlikte yorumlanır. Yalnızca başarılı çıktılar değerlendiriliyorsa puanın kapsamı bu alt kümedir ve dışlanan çıktı oranı da verilmelidir.

Şekil 6. Ayrı değerlendirme eksenlerinden amaca özgü karara geçiş



Kaynak: Ruiz et al. (2023), Somepalli et al. (2023) ve R. Zhang et al. (2018) temel alınarak yazar tarafından hazırlanmıştır.

Şekil 6'daki ölçümler birlikte yorumlanırken belirsizlikler ve hata örnekleri de dikkate alınır. Kimlik sadakati, çeşitlilik veya maliyet arasındaki tercih, kullanım amacına bağlıdır.

Eğitim görüntülerine en yakın çıktılar, ayrıntılı inceleme için seçilebilir. Difüzyon modellerinde veri tekrarı, görüntü erişimi ve benzerlik analiziyle de araştırılmıştır (Somepalli et al., 2023). Benzer bir incelemede yakın eşleşmelerin eğitim görüntüsündeki poz, arka plan ve ayrıntıları ne ölçüde yinelediği görsel olarak değerlendirilebilir.

Ana araştırma sorusu, test başlamadan belirlenen birincil ölçütler ve yöntem karşılaştırmalarıyla sınanır; ek ölçümler ve sonradan yapılan alt grup incelemeleri ise keşifsel niteliktedir. Birden fazla doğrulayıcı iddia için karşılaştırma ailesinin ve çoklu sınama yaklaşımının önceden tanımlanması gerekir. Sonuçlar farklı yönlere işaret ettiğinde, yöntemlerin hangi boyutta ne kadar ayrıştığı ve farkın belirsizliği açıklanır.

Toplam başarısızlık oranı, hata türleri arasındaki farkları örtebilir. Kimlik kaybı, istemin bir bölümünün atlanması, eğitim arka planının taşınması, aynı pozun tekrarı ve görsel bozulmalar ayrı incelenebilir. Ana kategorilerin önceden belirlenmesi, yöntemlerin ortak hata tanımlarıyla değerlendirilmesini sağlar; sonradan eklenen kategoriler ise keşifsel bulgulardır. Başarılı ve başarısız örneklerin önceden belirlenmiş bir kuralla seçilip birlikte gösterilmesi, kullanım sınırlarını görmeyi kolaylaştırır.

Tamamlanamayan denemeler de karşılaştırmayı etkiler. Bellek yetersizliği, üretim hatası, güvenlik filtresi veya bozuk dosya nedeniyle değerlendirilemeyen çıktılar, düşük kaliteli görüntülerden ayrı sayılır. Planlanan, tamamlanan, yinelenen ve değerlendirilebilir denemelerin birlikte bildirilmesi, her yöntemdeki çıktı kaybını gösterir. Önceden belirlenen yeniden deneme ve dışlama kuralları da bu kayıpların tutarlı biçimde ele alınmasını sağlar. Yalnızca

değerlendirilebilir çıktılardan hesaplanan kalite puanına üretim başarısızlığı oranı eşlik etmelidir. Eksikler belirli yöntem veya koşullarda yoğunlaşıyorsa dışlamanın karşılaştırmaya etkisi, uygun bir duyarlılık analiziyle incelenir.

Uygulama kararında baskın olmayan seçenekler incelenebilir. Bir yöntem daha düşük maliyetle benzer kimlik sadakati sunarken bir başkası daha yüksek çeşitlilik sağlayabilir. Böyle bir karşılaştırma, tercihleri ölçülen özellikler üzerinden açıklar. Kararın geçerliliği yine kullanılan ölçütlerin amaca uygunluğuna ve farkların belirsizliğine bağlıdır.

Etik, Erişim ve Bulguların Geçerlilik Sınırları

Kişiselleştirilmiş üretim, bir kişinin veya tasarımın ayırt edici özelliklerini model davranışına aktarır. Görüntüleri edinme, eğitimde kullanma, çıktı üretme, model paylaşma ve değerlendiricilere gösterme aşamaları farklı izin kapsamları gerektirebilir. Özellikle yüz görüntülerinde herkese açık erişim, sonraki kullanımların uygunluğunu belirlemek için yeterli bilgi sağlamaz. Veri yönetimi, izinlerin ve kullanım sınırlarının teknik deneyle birlikte planlanmasını gerektirir.

Mahremiyet değerlendirmesinde dağıtılan ağırlıklara kimlerin erişebildiği, referans görüntülerinin ne kadar süre saklandığı ve kişinin örnek çıktılarda nasıl temsil edildiği incelenir. Adaptör dosyasının küçük olması bu riskleri azaltmaz. Kontrollü erişim ve kullanım sınırlamaları, kişisel verinin paylaşımını azaltabilir; aynı zamanda bağımsız yeniden üretimi güçleştirebilir. Açık protokol ve işleme kodu bu güçlüğü kısmen azaltır. Erişilemeyen bileşenlerin hangi iddiaların denetimini sınırladığı, çalışmanın yorumunda açıkça yer almalıdır.

Tartışılan ayrımlar difüzyon tabanlı modellere dayanır. Uyarlama mekanizması ve hesaplama düzeni farklı olan mimarilerde bu ayrımlar yeniden ele alınmalıdır. Veri bölmesi ve ölçüm

tercihlerinin etkisi ise aynı karşılaştırma düzeni farklı veri kümelerinde uygulanarak görülebilir.

Teknik değerlendirmeye rıza, hak sahipliği, temsil hataları ve yanıltıcı içerik üretimi gibi konular eşlik eder. Bunların bir bölümü görsel kalite ölçütlerinin dışında kalır. Model kartları ve veri belgeleri, kullanım amacını, sınırları ve alınan önlemleri açıklamak için kullanılabilir (Gebru et al., 2021; Mitchell et al., 2019). Etik değerlendirme, bu kayıtların yanında verinin nasıl toplandığını ve modelin fiilen nasıl kullanıldığını da kapsar.

Sonuç

Kişiselleştirme yöntemlerini karşılaştırmak, her yöntemin neyi öğrendiğini ve üretime hangi bilgiyi eklediğini ayırt etmeyi gerektirir. Textual Inversion simge gömmesini, DreamBooth özneye yönelik eğitim düzenini, LoRA düşük ranklı ağırlık güncellemesini, görsel adaptörler ise görüntüyle koşullandırmayı öne çıkarır. Bu bileşenlerin aynı sistemde birlikte kullanılabilmesi, karşılaştırmanın kapsamını açık tanımlamayı daha da önemli kılar.

Kimlik sadakati, istem uyumu ve çeşitlilik farklı hata türlerine duyarlıdır. Eğitim görüntülerine yakınlık ve kaynak tüketimi de yöntem tercihini etkiler. Bağımsız veri bölmesi, açık model ve yazılım sürümleri, eğitim ile örnekleme tekrarlarının ayrılması ve başarısız denemelerin hesaba katılması, bu sonuçların denetlenmesini sağlar. Böylece yöntemler arasındaki farkın veriden, uyarlamadan veya çıkarım ayarlarından ne ölçüde kaynaklandığı daha dikkatli incelenebilir.

Ele alınan karşılaştırma düzeniyle yöntem farkları aynı veri ve bütçe koşullarında sınanabilir. Bağımsız yeniden üretimler, ölçütlerin insan değerlendirmeleriyle uyumu ve özne türü ile bütçenin yöntem tercihine etkisi üzerine yapılacak çalışmalar, bu karşılaştırmaların hangi koşullarda geçerli olduğunu netleştirebilir.

Kaynakça/References

Carlini, N., Hayes, J., Nasr, M., Jagielski, M., Sehwag, V., Tramèr, F., Balle, B., Ippolito, D., & Wallace, E. (2023). Extracting training data from diffusion models. In *32nd USENIX Security Symposium* (pp. 5253–5270). USENIX Association. https://www.usenix.org/conference/usenixsecurity23/presentation/carlini

Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., & Joulin, A. (2021). Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 9630–9640). IEEE. https://doi.org/10.1109/iccv48922.2021.00951

Chen, T., Xu, B., Zhang, C., & Guestrin, C. (2016). *Training deep nets with sublinear memory cost*. arXiv. https://doi.org/10.48550/arXiv.1604.06174

Chong, M. J., & Forsyth, D. (2020). Effectively unbiased FID and Inception Score and where to find them. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 6069–6078). IEEE. https://doi.org/10.1109/cvpr42600.2020.00611

Dettmers, T., Lewis, M., Shleifer, S., & Zettlemoyer, L. (2022). 8-bit optimizers via block-wise quantization. In *International Conference on Learning Representations*. OpenReview.net. https://openreview.net/forum?id=shpkpVXzo3h

Gal, R., Alaluf, Y., Atzmon, Y., Patashnik, O., Bermano, A. H., Chechik, G., & Cohen-Or, D. (2023). An image is worth one word: Personalizing text-to-image generation using textual inversion. In *International Conference on Learning Representations*. OpenReview.net. https://openreview.net/forum?id=NAQvF08TcyG

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé, H., III, & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM, 64*(12), 86–92. https://doi.org/10.1145/3458723

Ghosh, D., Hajishirzi, H., & Schmidt, L. (2023). GenEval: An object-focused framework for evaluating text-to-image alignment. *Advances in Neural Information Processing Systems, 36*, 52132–52152. https://doi.org/10.52202/075280-2270

Hessel, J., Holtzman, A., Forbes, M., Le Bras, R., & Choi, Y. (2021). CLIPScore: A reference-free evaluation metric for image captioning. In M.-F. Moens, X. Huang, L. Specia, & S. W.-t. Yih (Eds.), *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing* (pp. 7514–7528). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.emnlp-main.595

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2017). GANs trained by a two time-scale update rule converge to a local Nash equilibrium. *Advances in Neural Information Processing Systems, 30*, 6626–6637. https://proceedings.neurips.cc/paper/2017/hash/8a1d694707eb0fefe65871369074926d-Abstract.html

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems, 33*, 6840–6851. https://proceedings.neurips.cc/paper/2020/hash/4c5bcfec8584af0d967f1ab10179ca4b-Abstract.html

Ho, J., & Salimans, T. (2022). *Classifier-free diffusion guidance*. arXiv. https://doi.org/10.48550/arXiv.2207.12598

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, W. (2022). LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*. OpenReview.net. https://openreview.net/forum?id=nZeVKeeFYf9

Kynkäänniemi, T., Karras, T., Laine, S., Lehtinen, J., & Aila, T. (2019). Improved precision and recall metric for assessing generative models. *Advances in Neural Information Processing Systems, 32*, 3927–3936. https://proceedings.neurips.cc/paper/2019/hash/0234c510bc6d908b28c70ff313743079-Abstract.html

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 220–229). Association for Computing Machinery. https://doi.org/10.1145/3287560.3287596

Pineau, J., Vincent-Lamarre, P., Sinha, K., Larivière, V., Beygelzimer, A., d'Alché-Buc, F., Fox, E., & Larochelle, H. (2021). Improving reproducibility in machine learning research (a report from the NeurIPS 2019 reproducibility program). *Journal of Machine Learning Research, 22*(164), 1–20. https://jmlr.org/papers/v22/20-303.html

Podell, D., English, Z., Lacey, K., Blattmann, A., Dockhorn, T., Müller, J., Penna, J., & Rombach, R. (2024). SDXL: Improving latent diffusion models for high-resolution image synthesis. In *International Conference on Learning Representations*. OpenReview.net. https://openreview.net/forum?id=di52zR8xgf

PyTorch contributors. (2026, May 14). *Reproducibility*. PyTorch 2.14 documentation. https://docs.pytorch.org/docs/2.14/notes/randomness.html

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. In M. Meila & T. Zhang (Eds.), *Proceedings of the 38th International Conference on Machine Learning* (Vol. 139, pp. 8748–8763). PMLR. https://proceedings.mlr.press/v139/radford21a.html

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 10674–10685). IEEE. https://doi.org/10.1109/cvpr52688.2022.01042

Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M., & Aberman, K. (2023). DreamBooth: Fine tuning text-to-image diffusion models for subject-driven generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 22500–22510). IEEE. https://doi.org/10.1109/cvpr52729.2023.02155

Somepalli, G., Singla, V., Goldblum, M., Geiping, J., & Goldstein, T. (2023). Diffusion art or digital forgery? Investigating data replication in diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 6048–6058). IEEE. https://doi.org/10.1109/cvpr52729.2023.00586

Song, J., Meng, C., & Ermon, S. (2021). Denoising diffusion implicit models. In *International Conference on Learning Representations*. OpenReview.net. https://openreview.net/forum?id=St1giarCHLP

Ye, H., Zhang, J., Liu, S., Han, X., & Yang, W. (2023). *IP-Adapter: Text compatible image prompt adapter for text-to-image diffusion models*. arXiv. https://doi.org/10.48550/arXiv.2308.06721

Zhang, L., Rao, A., & Agrawala, M. (2023). Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 3813–3824). IEEE. https://doi.org/10.1109/iccv51070.2023.00355

Zhang, R., Isola, P., Efros, A. A., Shechtman, E., & Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 586–595). IEEE. https://doi.org/10.1109/cvpr.2018.00068

BÖLÜM 7

BÖLÜM I

Çok Kriterli Karar Verme Yöntemlerinden TOPSIS, GRA, VIKOR, WASPAS ve SAW ile Pandemi Dönemi Bist İnşaat Firmalarının Performans Değerlendirmeleri

Gürkan KAPLAN[1]

[1] Arş. Gör. Dr., Eskişehir Osmangazi Üniversitesi, Fen Fakültesi, Matematik ve Bilgisayar Bilimleri, Eskişehir/TÜRKİYE, Orcid: 0000-0002-6393-5546, gkaplan@ogu.edu.tr

Giriş

Küresel ölçekte büyük bir yıkıma yol açan COVID-19 pandemisi, dünya genelinde yürürlüğe giren kısıtlamalar nedeniyle pek çok alanın operasyonlarını askıya almasına ya da yavaşlatmasına sebebiyet vermiştir. Şantiye süreçlerindeki duraksamalar, hammaddeye erişim zorlukları ve makroekonomik belirsizlikler sebebiyle bu krizden en ağır darbeyi alan iş kollarının arasında inşaat sektörü yer almıştır. Uluslararası yazında yer alan çok sayıda ampirik çalışma, pandeminin inşaat firmalarının kârlılık, nakit akışı, borçluluk ve likidite performansları üzerinde yarattığı tahribatı somut verilerle ortaya koymaktadır.

Sektörün finansal göstergeleri incelendiğinde öne çıkan ilk temel dinamik kârlılık oranlarındaki keskin düşüştür. Salami vd. (2023), küresel tedarik zinciri kesintilerinin malzeme maliyetlerini ve teslimat sürelerini artırmasının nakit çıkışlarını tırmandırdığını ve bunun da projelerin bütçe dengelerini bozduğunu vurgulamaktadır. Benzer şekilde, Pamidimukkala ve Kermanshachi (2021) projelere ait ek güvenlik ve hijyen protokollerinin operasyonel maliyetleri tırmandırarak kâr marjlarını daralttığını göstermiştir. Gamil ve Alhagar (2020) ise geciken projeler sebebiyle yüklenici firmaların uğradığı maliyet artışlarının ve nakit akışındaki aksamaların şirketlerin net kârlılığını ciddi şekilde erittiğini saptamıştır.

Kârlılıktaki bu aşınma, inşaat sektörünün likidite ve borç yönetimi dengelerinde ciddi bir kırılmaya yol açmıştır. Assaad ve El-Adaway (2021), kriz sürecinde operasyonel nakit akışlarındaki duraksamanın yüklenici firmaları ek finansman ve borçlanmaya yönelttiğini, bunun da sektör genelinde finansal riskleri ve yükümlülükleri artırdığını vurgulamaktadır. Siraj ve Fayek (2019) ile Assaad ve El-Adaway (2021), kriz dönemlerinde yaşanan bu likidite darlığının özellikle işletme sermayesi yetersiz olan firmalarda nakit tükenmesi (*cash burn*) krizine dönüştüğünü ve sektörel finansal kırılganlığı tırmandırdığını ifade etmektedir.

Krizin borsa ve sermaye piyasalarına yansıması ise piyasa değeri erimeleri ve hisse senedi getirilerindeki dalgalanmalar şeklinde gerçekleşmiştir. Al-Awadhi vd. (2020) ile Zhang vd. (2020), küresel salgın döneminde sermaye piyasalarında artan risk algısı nedeniyle borsa şirketlerinin getirilerinde ciddi düşüşler yaşandığını ve inşaat gibi yüksek riskli sektörlerin finansman maliyetlerinin yükseldiğini ortaya koymuştur.

Sonuç olarak uluslararası akademik literatür, COVID-19 krizinin inşaat sektöründe yalnızca geçici bir gelir kaybı yaratmadığını; aynı zamanda nakit yönetimi, sermaye yapısı ve risk değerlendirme stratejilerini temelden değiştirdiğini doğrulamaktadır.

Bu araştırmada, salgın sürecinde Borsa İstanbul'da (BIST) faaliyet gösteren inşaat şirketlerinin mali yapılarının ve performanslarının analiz edilmesi hedeflenmiştir. Ele alınan dönemi kapsayan 2019-2022 yılları arasında, BIST bünyesindeki önde gelen inşaat işletmelerinin mali tabloları titizlikle taranmış ve performans ölçümüne esas teşkil eden finansal oranlar hesaplanmıştır.

Analiz aşamasında, hem yıllık veriler baz alınarak Çok Kriterli Karar Verme (ÇKKV) tekniklerinden TOPSIS, GRA, VIKOR, WASPAS ve SAW yöntemleriyle sıralamalar gerçekleştirilmiştir. Uygulama neticesinde; pandeminin etkilerinin zirve yaptığı 2020 ve onu takip eden 2021 senelerinde şirketlerin finansal göstergelerinde belirgin bir gerileme yaşandığı saptanmıştır. Çalışmadan elde edilen nihai bulgular ve firma karşılaştırmaları, tablolar ve şekiller aracılığıyla somutlaştırılmıştır.

Bahsedilen 4 dönem boyunca mali verileri Bist'te yer alan ve bu çalışma kapsamında ele alınan 7 inşaat firması incelenmiş olup firmaların sembolleri aşağıdaki tabloda verilmiştir.

Tablo 1: Firma İsimleri ve Sembolleri

Firma İsmi	**Sembolü**

Anel Elektrik Proje Taahhüt ve Ticaret A. Ş.	ANELE
Edip Gayrimenkul Yatırım San. Ve Tic. A. Ş.	EDIP
Enka İnşaat ve Sanayi A. Ş.	ENKAI
Kuyaş Yatırım A. Ş.	KUYAS
Orge Enerji Elektrik Taahhüt A. Ş.	ORGE
Yayla Enerji Üretim Turizm ve İnş. Tic. A. Ş.	YAYLA
Yeşil Yapı Endüstrisi A. Ş.	YYAPI

Materyal ve Metot

Finansal oranlar, inşaat firmalarının performans değerlendirmesinde ve kurumsal analiz süreçlerinde çok yönlü bir yönetsel araç olarak öne çıkmaktadır. Söz konusu rasyolar; yüklenici firmaların kısa ve uzun vadeli borç ödeme kapasitelerinin ölçülmesinde, proje ve şantiye yönetimi başarısının finansal tablolara yansımasında ve sektör paydaşları nezdinde firmanın kredibilitesinin ortaya konulmasında belirleyici temel unsurdur.

Uygulamalı akademik çalışmalar ve finansal analiz literatürü, bu oranların tahmin ve erken uyarı gücünü açıkça ortaya koymaktadır. Yapılan ampirik araştırmalar, finansal darboğaza giren veya iflas riskiyle yüzleşen şirketlerin henüz kriz derinleşmeden önce finansal oranlardaki bozulmalar sayesinde yüksek bir doğruluk payıyla tespit edilebildiğini göstermektedir. Benzer şekilde, bankalar ve finansal kuruluşlar da inşaat firmalarının kredi tahsis ve risk derecelendirme süreçlerinde nihai kararlarını büyük ölçüde bu oranlar üzerinden şekillendirmektedir.

Finansal rasyolar; inşaat sektörünün kendine özgü yapısı, uzun vadeli üretim döngüleri, yüksek stok yükü ve hakediş vadeleri gibi sektörel dinamikler nedeniyle diğer iş kollarından belirleyici

farklılıklar göstermektedir. Buna ek olarak, küresel COVID-19 pandemisi gibi makroekonomik şoklar, tedarik zinciri aksamaları ve dönemsel krizler de inşaat firmalarının finansal tablolarının ve oranlarının doğru yorumlanmasında kritik bir rol oynamaktadır.

Likidite oranları, inşaat şirketlerinin yüksek işletme sermayesi gerektiren şantiye süreçlerinde kısa vadeli yükümlülüklerini yerine getirme kabiliyetini ölçen kritik göstergelerdir. Sektörde yaygın olarak kullanılan cari oran, yüklenici firmanın dönen varlıkları ile kısa vadeli borçlarını karşılaştırarak mevcut borç ödeme gücünü ortaya koyar; bu oranın yüksekliği, firmanın nakit sıkışıklığı yaşamadan projelerini sürdürebileceğini göstererek yatırımcılar ve işveren idareler nezdinde güven yaratır. Stokların çıkarılmasıyla hesaplanan asit-test oranı ise tamamlanmamış konut veya proje stoklarının likiditeye dönüşüm hızını dışarıda bırakarak daha muhafazakâr bir tablo sunar. Bu noktada inşaat sektöründeki alacak devir hızı hayati bir rol oynamaktadır; bilançoda görünen ancak tahsilat kabiliyeti kalmamış taşeron veya işveren alacakları, doğru bir firma değerlendirmesi yapılabilmesi adına analistlerce bilançodan arındırılmalıdır. Firmanın anlık ödeme kapasitesini gösteren nakit oranı değerlendirilirken de sektörel dinamikler göz önünde bulundurulmalıdır. İnşaat sektöründe anlık ham madde alımları veya şantiye giderleri nedeniyle yüksek nakit tutulması gerekebileceği gibi, sermayenin devam eden projelere bağlandığı durumlarda düşük kasa bakiyesi her zaman olumsuz bir durum olarak yorumlanmamalıdır.

İnşaat firmalarında operasyonel etkinliği ve şantiye verimliliğini ölçen faaliyet oranları, varlıkların ne kadar başarılı yönetildiğini ortaya koyar. Net satışların toplam aktiflere oranlanmasıyla bulunan aktif devir hızı, yüklenici firmanın sahip olduğu iş makinesi, ekipman ve gayrimenkul gibi tüm varlıklarıyla yıl içinde ne kadarlık bir hakediş veya satış hacmi yarattığını ölçer ve bu oranın sektör ortalamasının üzerinde olması olumlu değerlendirilir. Dönen varlık devir hızı ise, malzeme stokları ve kısa vadeli alacakların yer aldığı dönen varlıkların yıl içinde kaç kez

satışa ve nakit akışına dönüştüğünü göstererek proje döngüsünün hızını yansıtır.

İşletmenin borç-özkaynak dengesini ve uzun vadeli finansal sürdürülebilirliğini ortaya koyan finansal yapı oranları, özellikle yüksek borçlanma gereksinimi duyan inşaat projelerinde kilit bir role sahiptir. Finansal kaldıraç oranı, inşaat projelerinin ve varlıkların ne kadarlık bir kısmının dış kaynakla finanse edildiğini gösterirken; özkaynak oranı, firmanın bilançosunu kendi kaynaklarıyla fonlama gücünü ifade eder. Özkaynağın aktif fonlamasındaki payının yüksekliği, firmanın dış kredi ihtiyacını ve faiz yükünü azaltır. Benzer şekilde finansman oranı, özkaynakların toplam yabancı kaynaklara oranını göstererek firmanın krizlere karşı dayanıklılığını ölçer. Borçların sağlıklı bir bilanço yapısına işaret etmesi ise doğrudan borçların vade yapısıyla ilgilidir; kısa ve uzun vadeli yükümlülükler oranları, inşaat projelerinin uzun vadeli nakit akışlarına uygun bir borçlanma takvimi oluşturulup oluşturulmadığını doğrular.

Son olarak kârlılık oranları, inşaat faaliyetlerinin ve üstlenilen proje risklerinin finansal başarısını gösterir. Vergi sonrası net kârın toplam aktiflere oranlanmasıyla hesaplanan aktif kârlılığı, firmanın üstlendiği projeler ve kullandığı varlıklar üzerinden ne kadar kâr üretebildiğini gösterir. Firmanın kendi kaynaklarının yıllık kârlılık performansını ölçen özkaynak kârlılığının ise normal bir faaliyet döneminde cari piyasa faiz oranının üzerinde gerçekleşmesi beklenir. Aksi halde, inşaat sektörünün taşıdığı yüksek operasyonel ve finansal riskler karşılığında firmanın özkaynaklarını verimli bir şekilde değerlendiremediği sonucu ortaya çıkar.

Çalışmada kullanılan finansal oranlar ve hesaplamaları Tablo 2'de verilmiştir.

Tablo 2: Finansal Oranlar ve Hesaplamaları

ORANLAR	Fayda/Maliyet

I.	**Likidite Oranları ve Hesaplamaları**		
1	Cari Oran	Dönen Varlıklar / Kısa Vadeli Borçlar	Maksimum
2	Asit-Test Oranı	(Dönen Varlıklar-Stoklar) / Kısa Vadeli Borçlar	Maksimum
3	Nakit Oranı	Hazır Değerler / Kısa Vadeli Borçlar	Maksimum
II.	**Finansal Yapı Oranları ve Hesaplamaları**		
4	Finansal Kaldıraç Oranı	Toplam Borçlar / Toplam Varlıklar	Minimum
5	Kısa Vadeli Yükümlülükler / Top. Varlıklar	Kısa Vadeli Yükümlülükler / Top. Varlıklar	Minimum
6	Uzun Vadeli Yükümlülükle / Top. Varlıklar	Uzun Vadeli Yükümlülükle / Top. Varlıklar	Minimum
7	Özkaynaklar / Top. Varlıklar	Özkaynaklar / Top. Varlıklar	Maksimum
III.	**Verimlilik Oranları ve Hesaplamaları**		
8	Stok Devir Hızı	Satılan Malların Maliyeti / Ortalama Stoklar	Minimum
9	Alacak Devir Hızı	Kredili Satışlar / Ortalama Ticari Alacaklar	Maksimum
10	Aktif Devir Hızı	Net Satışlar / Ortalama Toplam Aktifler	Maksimum
IV.	**Faaliyet Oranları ve Hesaplamaları**		
11	Özsermaye Devir Hızı	Net Satışlar / Ortalama Özkaynaklar	Maksimum

12	Net İşletme Sermayesi Devir Hızı	Net Satışlar / Ortalama Net İşletme Sermayesi	Maksimum
13	Maddi Duran Varlık Devir Hızı	Net Satışlar / Ortalama Net Maddi Duran Varlıklar	Maksimum
14	Hazır Değerler Devir Hızı	Net Satışlar / Ortalama Hazır Değerler	Maksimum
V.	**Karlılık Oranları ve Hesaplamaları**		
15	Özsermaye Karlılık Oranı	Net Dönem Karı / Ortalama Özkaynaklar	Maksimum
16	Aktif Getiri Oranı	Net Dönem Karı / Ortalama Toplam Aktifler	Maksimum
17	Brüt Kar Marjı	Brüt Satış Karı / Net Satışlar	Maksimum
18	Faaliyet Karı Oranı	Faaliyet Karı / Net Satışlar	Maksimum
19	Net Kar Marjı	Net Dönem Karı / Net Satışlar	Maksimum

Çalışmada kullanılan her yılın finansal oranların ağırlıkları her yıl için ayrı ayrı CRITIC metodu ile belirlenmiştir. Sonrasında bu ağırlıklar ile TOPSIS, GRA, VIKOR, WASPAS ve SAW yöntemleriyle elde edilen sonuçlar tablo ve şekiller ile görselleştirilmiştir. Bu bölümde kullanılan tüm yöntemler kısaca açıklanacaktır.

CRITIC (CRiteria Importance Through Intercriteria Correlation) metodu, karar matrisindeki objektif verileri kullanarak kriter ağırlıklarını belirlemek amacıyla Diakoulaki vd. (1995) tarafından geliştirilmiş nesnel bir ağırlıklandırma yöntemidir. Yöntem, uzman görüşüne veya karar vericilerin öznel tercihlerine ihtiyaç duymadan, kriterlerin taşıdığı bilgi miktarını standart sapma ile, kriterler arası çelişkiyi/ilişkiyi ise Pearson korelasyon katsayısı ile analiz eder. Karar matrisi standartlaştırıldıktan sonra, her bir kriterin sahip olduğu sapma miktarı ve diğer kriterlerle gösterdiği

korelasyon bağıntısı birleştirilerek kriterin toplam bilgi yükü hesaplanır; yüksek varyasyona ve diğer kriterlerle düşük korelasyona (yani yüksek bağımsız bilgiye) sahip kriterlere daha fazla bağıl ağırlık atanır.

TOPSIS yönteminin temel felsefesi, değerlendirilen alternatifler arasından pozitif ideal çözüme geometrik olarak en kısa mesafede yer alan, aynı zamanda negatif ideal çözümden ise en uzak konumda bulunan alternatifi belirleme ilkesine dayanmaktadır(Jollyta, 2018). Söz konusu karar verme sürecinin işlem adımları aşağıda sırasıyla sunulmaktadır:

İlk aşamada, farklı birim veya ölçeklere sahip kriter değerlerini ortak bir tabanda karşılaştırılabilir kılmak amacıyla normalleştirilmiş (standartlaştırılmış) karar matrisi matrisi oluşturulur. Denklem (1)'de ifade edilen bu süreçte i alternatifi (i = 1, …, m), j ise değerlendirme kriterini (j = 1,…, n) temsil etmektedir:

$$n_{ij} = \frac{x_{ij}}{\sum_{i=1}^{m} x_{ij}^2} \qquad (1)$$

İkinci aşama olarak ağırlıklandırılmış normalize karar matrisi y_{ij} hesaplanır.

$$y_{ij} = w_i n_{ij}, \; j = 1,...J; i = 1,...m \qquad (2)$$

Burada w_i i. kriterin ağırlığını göstermektedir ve $\sum_{i=1}^{m} w_i = 1$'dir.

Üçüncü aşama Ağırlıklandırılmış karar matrisinin her bir sütunundaki en yüksek ve en düşük performans değerleri tespit edilerek sırasıyla pozitif ideal çözüm (A+) ve negatif ideal çözüm (A-) matrisleri oluşturulur.

$$A^{+} = (y_1^{+}, y_2^{+}, ..., y_n^{+}) \quad (3)$$

$$A^{-} = (y_1^{-}, y_2^{-}, ..., y_n^{-}) \quad (4)$$

Dördüncü aşamada matrisler arası uzaklıklar bulunur.

$$D_i^{+} = \sqrt{\sum_{j=1}^{m} (y_i^{+} - y_{ij})^2} \quad (5)$$

$$D_i^{-} = \sqrt{\sum_{j=1}^{m} (y_{ij} - y_i^{-})^2} \quad (6)$$

Beşinci adımda ise V_i i alternatifinin değeri

$$V_i = \frac{D_i^{-}}{D_i^{-} + D_i^{+}} \quad (7)$$

GRA(Gri İlişkisel Analiz) yöntemi, veri setinin türü ne olursa olsun (küçük örneklem boyutu veya düzensiz veri dağılımları dahil) nesnel bir değerlendirme olanağı sunar (Deng, 1989). m adet alternatif ve n adet kriterden oluşan bir karar probleminde süreç 5 temel adımda yürütülür:

1- Karar matrisinin oluşturulması ve veri normalizasyonu (Kuo et al., 2008).

2- Referans dizisinin ve mutlak fark matrisinin oluşturulması
3- Gri ilişkisel katsayılarının hesaplanması
4- Gri ilişkisel derecenin hesaplanması
5- Alternatiflerin sıralanması

VIKOR yöntemi, birbiriyle çelişen ve farklı ölçü birimlerine sahip kriterlerin varlığı altında, alternatifler arasından grup için en uygun uzlaşık çözümü belirlemeyi amaçlayan bir karar destek tekniğidir (Opricovic & Tzeng, 2007). Yöntemin temelinde, en iyi alternatifin "ideal çözüme en yakın" nokta olması ve karar vericilerin çoğunluğunun faydasını maksimize ederken bireysel pişmanlığı minimize etme felsefesi yer almaktadır (Opricovic & Tzeng, 2004). VIKOR yönteminin uygulama süreci genel olarak dört temel aşamada yürütülmektedir:

1- İdeal ve En Kötü Değerlerin Tespiti: Karar matrisinde yer alan her bir kriterin niteliğine (fayda veya maliyet yönlü olmasına) göre alternatifler içerisinden en iyi (*ideal*) ve en kötü (*aşırı*) değerler belirlenir.
2- Toplu Fayda (S) ve Bireysel Pişmanlık (R) Değerlerinin Hesaplanması: Alternatiflerin ideal çözüme olan bağıl uzaklıkları, kriter ağırlıkları dikkate alınarak analiz edilir. Bu aşamada her bir alternatif için çoğunluk faydasını temsil eden S değeri ile maksimum bireysel pişmanlığı temsil eden R değeri elde edilir (Opricovic & Tzeng, 2004).
3- Uzlaşık İndeks (Q) Skorunun Oluşturulması: Çoğunluk faydası ile bireysel pişmanlık dengesini sağlamak amacıyla, belirlenen bir strateji ağırlığı (genellikle v = 0.5) kullanılarak her bir alternatif için birleşik uzlaşık indeks skoru (Q) hesaplanır.
4- Sıralama ve Uzlaşık Çözümün Belirlenmesi: Alternatifler hesaplanan S, R ve Q parametrelerine göre küçükten büyüğe doğru sıralanır. Q skoru en küçük olan alternatif en başarılı seçenek olarak öne çıkarken; yöntemin sunduğu kabul edilebilir avantaj ve istikrarlı karar koşullarını sağlayan alternatifler nihai uzlaşık çözüm olarak belirlenir.

WASPAS yöntemi, ÇKKV teknikleri arasında karar verme hassasiyetini ve doğruluğunu artırmak amacıyla Zavadskas vd. (2012) tarafından geliştirilmiş karma bir yaklaşımdır. Yöntemin temel felsefesi, iki klasik Çok Kriterli Karar Verme modelinin(Ağırlıklı Toplam Modeli (WSM - Weighted Sum Model) ile Ağırlıklı Çarpım Modeli (WPM - Weighted Product Model)) belirlenen bir kombinasyon katsayısı(λ) üzerinden entegre edilmesine dayanır. WSM ve WPM yöntemlerinin tek başlarına kullanıldığında gösterebileceği sapmaları ve hassasiyet kayıplarını minimuma indirerek daha istikrarlı bir sıralama sunar. WASPAS yöntemi 5 temel aşamadan oluşur:

1- Karar matrisinin oluşturulması ve normalizasyonu
2- WSM'e göre bağıl önem derecesinin hesaplanması
3- WPM'e göre bağıl önem derecesinin hesaplanması
4- Birleşik performans skorunun oluşturulması
5- Alternatiflerin sıralanması

SAW yöntemi, ÇKKV literatüründe en köklü ve yaygın kullanılan doğrusal aditif modellerden biridir (Churchman & Ackoff, 1954). Yöntemin temel ilkesi, farklı birim veya ölçeklere sahip kriter değerlerinin normalleştirildikten sonra kendi önem ağırlıklarıyla çarpılması ve her bir alternatif için elde edilen ağırlıklı puanların toplanarak bileşik bir performans skoru türetilmesi felsefesine dayanır (MacCrimmon, 1968). Sezgisel olarak anlaşılması son derece kolay olan bu yaklaşım, tüm kriterlerin birbiriyle ikame edilebilir olduğunu varsayarak, bir kriterdeki düşük performansın başka bir kriterdeki yüksek performansla telafi edilmesine olanak tanır (Hwang & Yoon, 2012). Süreç; karar matrisinin fayda veya maliyet yönlü dönüştürmelerle ölçeksizleştirilmesi, belirlenen kriter ağırlıklarının matrise aktarılması ve nihayetinde toplam puanı en yüksek olan alternatifin en uygun seçenek olarak belirlenmesi adımlarıyla tamamlanır (Aghajani Bazzazi et al., 2011).

Bulgular

CRITIC yöntemi ile her yıl için bulunan finansal oran ağırlıkları Tablo 3'te verilmiştir.

Tablo 3: CRITIC Yöntemi Sonrası Elde Edilen ve Çalışmada Kullanılan Kriter Ağırlıkları

Oran No	Critic-2019	Critic-2020	Critic-2021	Critic-2022
1	0,048873	0,043842	0,048604	0,041655
2	0,04754	0,045336	0,049947	0,041694
3	0,042109	0,049978	0,069774	0,04523
4	0,054477	0,041983	0,042947	0,056751
5	0,066768	0,06317	0,062147	0,052039
6	0,066973	0,044929	0,043226	0,048366
7	0,054477	0,04196	0,042947	0,056751
8	0,074429	0,046696	0,043867	0,051303
9	0,043532	0,067157	0,075493	0,053353
10	0,046556	0,049535	0,040672	0,047583
11	0,052151	0,062552	0,042822	0,051149
12	0,052041	0,054821	0,057222	0,052004
13	0,072743	0,078172	0,06671	0,055136
14	0,063327	0,081039	0,065199	0,050712
15	0,039526	0,043626	0,037533	0,052251
16	0,036142	0,037204	0,03603	0,061838
17	0,044173	0,05996	0,068134	0,041082
18	0,047317	0,0427	0,062081	0,075914
19	0,046844	0,045342	0,044647	0,065187

TOPSIS metodu 4 kez kullanılarak her yılın karar matrisi için farklı hesaplamalar yapılmış olup sonuçlar Tablo 4'te, firma sıralamaları Tablo 5'te, grafiksel gösterimleri ise Şekil 1'de verilmiştir.

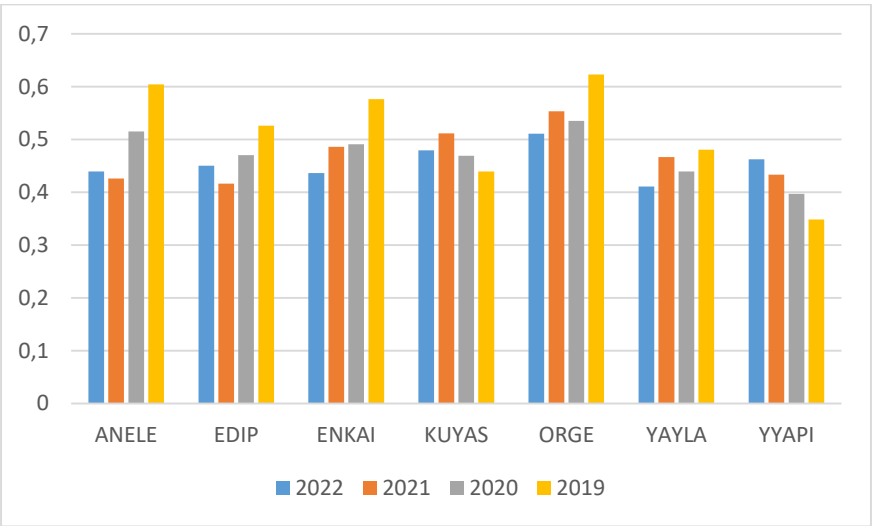
Tablo 4: TOPSIS Yöntemi ile Yıllara Göre Sonuçlar

Bist İnşaat	2019	2020	2021	2022

ANELE	0,6042312	0,5150863	0,4260769	0,4393345
EDIP	0,5260604	0,4702959	0,415994	0,450358
ENKAI	0,5764722	0,4908755	0,4857832	0,4364673
KUYAS	0,4391914	0,4688426	0,5111475	0,4791309
ORGE	0,6228545	0,5349703	0,552996	0,5106747
YAYLA	0,4807138	0,4390193	0,4665601	0,411059
YYAPI	0,3484108	0,3968495	0,4330448	0,4620807

Tablo 5: TOPSIS Yöntemi ile Yıllara Göre Firma Rankları

Bist İnşaat	2019	2020	2021	2022
ANELE	2	2	6	5
EDIP	4	4	7	4
ENKAI	3	3	3	6
KUYAS	6	5	2	2
ORGE	1	1	1	1
YAYLA	5	6	4	7
YYAPI	7	7	5	3



Şekil 1: Yıllara Göre TOPSIS Sonuç Grafiği

GRA metodu 4 kez kullanılarak her yılın karar matrisi için farklı hesaplamalar yapılmıştır. GRA yönteminde, karar matrisinde negatif veri olması herhangi bir sorun çıkarmaz. GRA yöntemi

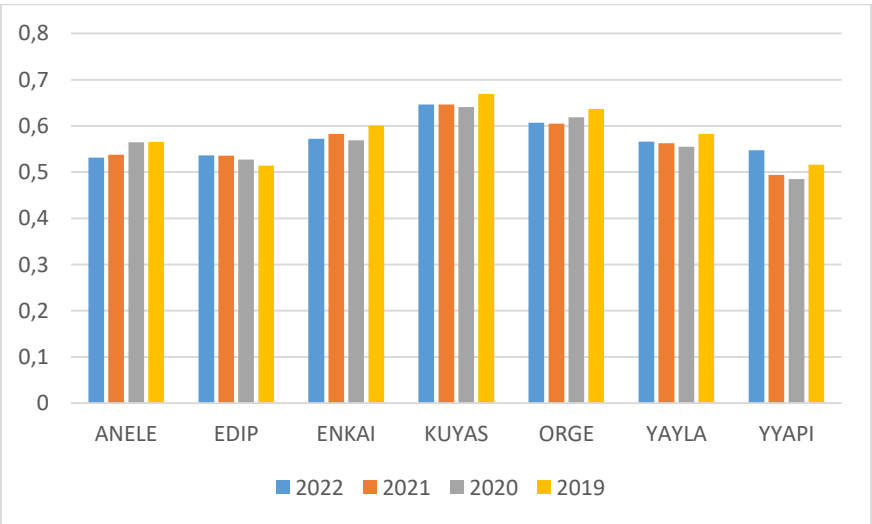
sonuçları Tablo 6'da, firma sıralamaları Tablo 7'de, grafiksel gösterimleri ise Şekil 2'de verilmiştir.

Tablo 6: GRA Yöntemi ile Yıllara Göre Sonuçlar

Bist İnşaat	2019	2020	2021	2022
ANELE	0,5651548	0,5645238	0,537459	0,5313039
EDIP	0,5138421	0,5270307	0,5356788	0,536
ENKAI	0,6008287	0,568461	0,5823431	0,5719544
KUYAS	0,6689072	0,6406705	0,6464718	0,6461278
ORGE	0,6367783	0,6184287	0,6050781	0,6066517
YAYLA	0,582506	0,5550823	0,5626377	0,5663344
YYAPI	0,5163362	0,4851614	0,4942416	0,547335

Tablo 7: GRA Yöntemi ile Yıllara Göre Firma Rankları

Bist İnşaat	2019	2020	2021	2022
ANELE	5	4	5	7
EDIP	7	6	6	6
ENKAI	3	3	3	3
KUYAS	1	1	1	1
ORGE	2	2	2	2
YAYLA	4	5	4	4
YYAPI	6	7	7	5

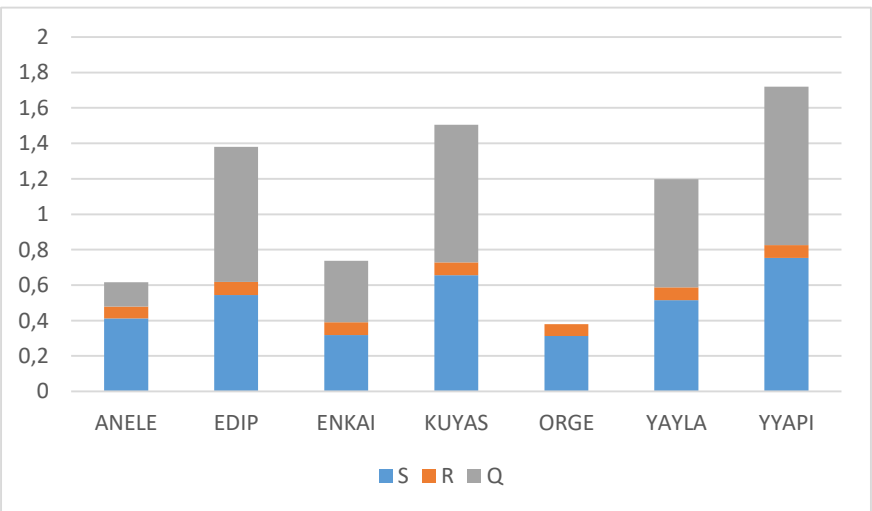


Şekil 2: Yıllara Göre GRA Sonuç Grafiği

VIKOR metodu 4 kez kullanılarak her yılın karar matrisi için farklı hesaplamalar yapılmıştır. VIKOR yönteminde, karar matrisinde negatif veri olması herhangi bir sorun çıkarmaz. Yöntemden üç farklı parametre elde edilir ve birlikte yorumlanır. Düşük olan değerler daha başarılıyı ifade eder. Her yıla ait VIKOR sonuçları(S, R, Q sıralamaları) ve grafiksel gösterimleri tablo ve şekil halinde aşağıda sırasıyla verilmiştir.

Tablo 8: VIKOR Yöntemi 2019 Sonuçları

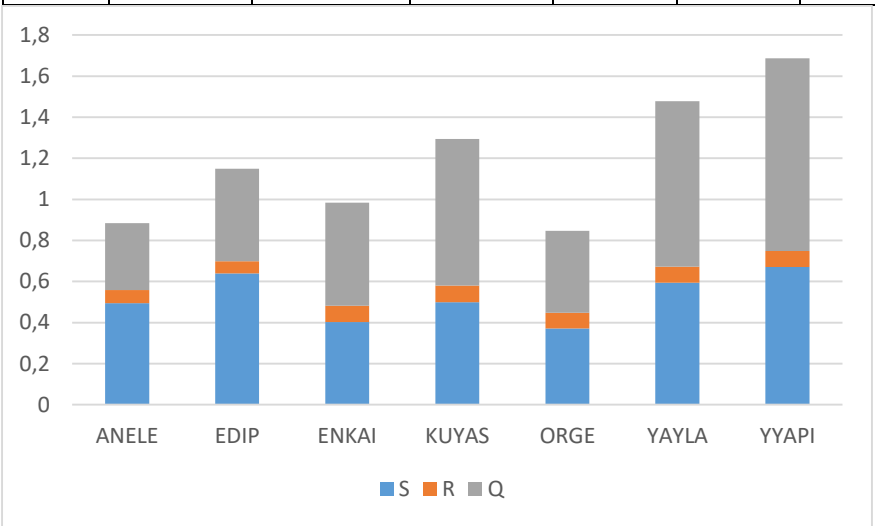
Bist İnşaat 2019	S	R	Q	S Sıralama	R sıralama	Q Sıralama
ANELE	0,4120763	0,06676834	0,1369728	3	2	2
EDIP	0,5431651	0,07442902	0,7621207	5	7	5
ENKAI	0,3180864	0,07187428	0,3475495	2	3	3
KUYAS	0,6556975	0,07261042	0,7769542	6	5	6
ORGE	0,3124972	0,0663852	0	1	1	1
YAYLA	0,5149672	0,07250526	0,6104977	4	4	4
YYAPI	0,7524004	0,07274326	0,8952139	7	6	7



Şekil 3: 2019 VIKOR Sonuç Grafiği

Tablo 9: VIKOR Yöntemi 2020 Sonuçları

Bist İnşaat 2020	S	R	Q	S Sıralama	R sıralama	Q Sıralama
ANELE	0,4945081	0,06388393	0,3255034	3	2	1
EDIP	0,6402115	0,05850354	0,4497789	6	1	3
ENKAI	0,4031198	0,07870681	0,5015627	2	6	4
KUYAS	0,4991844	0,08103865	0,7139454	4	7	5
ORGE	0,3712463	0,076487	0,3990098	1	3	2
YAYLA	0,5941211	0,07803027	0,8059551	5	4	6
YYAPI	0,6702434	0,07828734	0,938955	7	5	7

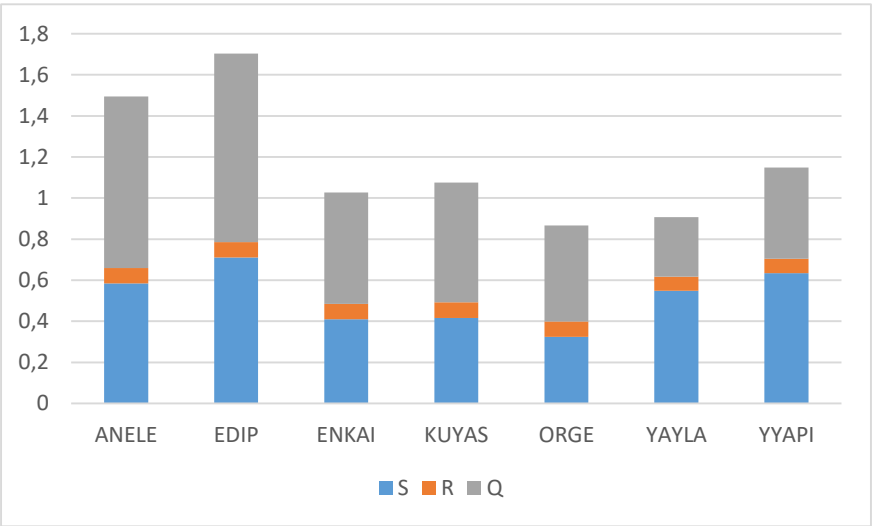


Şekil 4: 2020 VIKOR Sonuç Grafiği

Tablo 10: VIKOR Yöntemi 2021 Sonuçları

Bist İnşaat 2021	S	R	Q	S Sıralama	R sıralama	Q Sıralama
ANELE	0,5839511	0,07549331	0,8358884	5	7	6

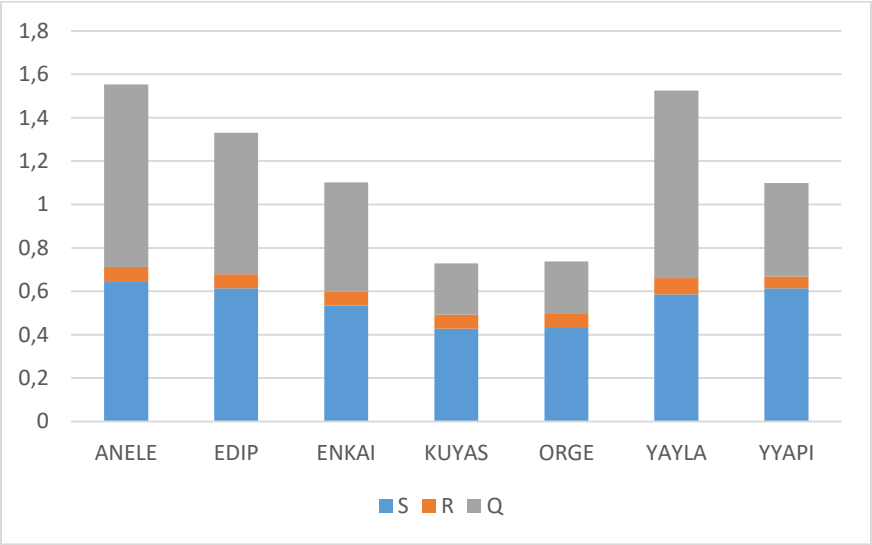
EDIP	0,711283	0,07446255	0,9176821	7	3	7
ENKAI	0,4100935	0,07462791	0,5426998	2	4	4
KUYAS	0,4167779	0,07502867	0,5833205	3	5	5
ORGE	0,3233999	0,07509402	0,4681119	1	6	3
YAYLA	0,5480932	0,06923245	0,289673	4	1	1
YYAPI	0,6343794	0,0697738	0,4441156	6	2	2



Şekil 5: 2021 VIKOR Sonuç Grafiği

Tablo 11: VIKOR Yöntemi 2022 Sonuçları

Bist İnşaat 2022	S	R	Q	S Sıralama	R sıralama	Q Sıralama
ANELE	0,6425811	0,06929266	0,8409906	7	6	6
EDIP	0,6141681	0,06418437	0,6519865	6	2	5
ENKAI	0,5337539	0,06576128	0,5021292	3	5	4
KUYAS	0,4284018	0,06489995	0,235501	1	3	1
ORGE	0,4299175	0,06509296	0,242954	2	4	2
YAYLA	0,5847912	0,07591401	0,8650899	4	7	7
YYAPI	0,6130376	0,0550934	0,4310308	5	1	3



Şekil 6: 2022 VIKOR Sonuç Grafiği

WASPAS metodu 4 kez kullanılarak her yılın karar matrisi için farklı hesaplama yapılmıştır. WASPAS yönteminde, karar matrisinde negatif veri olması sorun çıkarır. Yüksek olan değerler daha başarılıyı ifade eder. Yöntem sonuçları Tablo 12'de, firma sıralamaları Tablo 13'te, sonuç grafikleri Şekil 7'de verilmiştir.

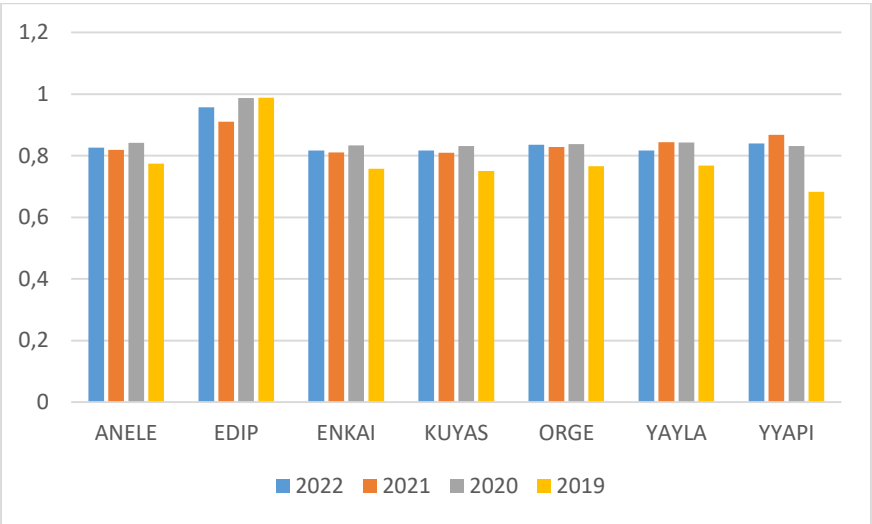
Tablo 12: WASPAS Yöntemi ile Yıllara Göre Sonuçlar

Bist İnşaat	2019	2020	2021	2022
ANELE	0,7742195	0,8420981	0,8187252	0,8259852
EDIP	0,9879871	0,9868127	0,9099433	0,9576609
ENKAI	0,757858	0,8331525	0,8107839	0,8164773
KUYAS	0,7499341	0,8314427	0,8097228	0,8166374
ORGE	0,7663433	0,8380518	0,828288	0,8352293
YAYLA	0,767862	0,8426037	0,8442095	0,817242
YYAPI	0,6829417	0,8309498	0,8677905	0,8393098

Tablo 13: WASPAS Yöntemi ile Yıllara Göre Firma Rankları

Bist İnşaat	2019	2020	2021	2022
ANELE	2	3	5	4
EDIP	1	1	1	1

ENKAI	5	5	6	7
KUYAS	6	6	7	6
ORGE	4	4	4	3
YAYLA	3	2	3	5
YYAPI	7	7	2	2



Şekil 7: Yıllara Göre WASPAS Sonuç Grafiği

SAW metodu 4 kez kullanılarak her yılın karar matrisi için farklı hesaplama yapılmıştır. SAW yönteminde, karar matrisinde negatif veri olması sorun çıkarır. Yüksek olan değerler daha başarılıyı ifade eder. Yöntem sonuçları Tablo 14'de, firma sıralamaları Tablo 15'te, sonuç grafikleri Şekil 8'de verilmiştir.

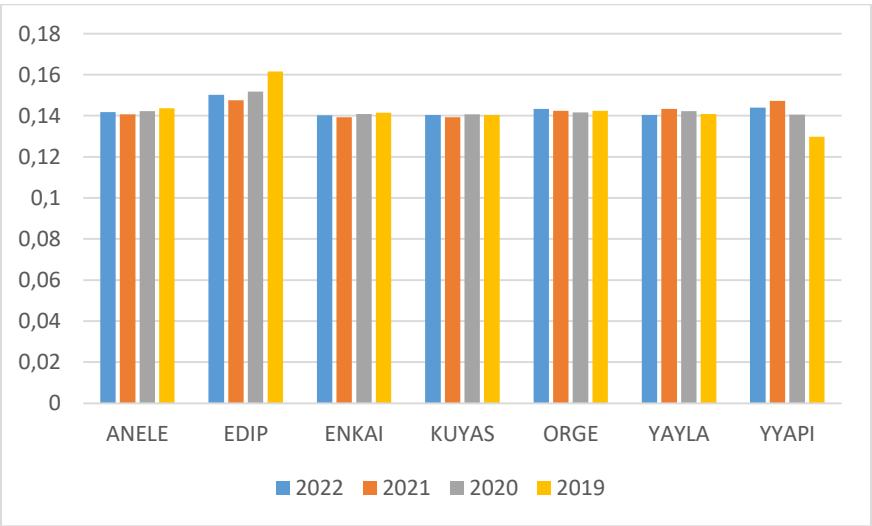
Tablo 14: SAW Yöntemi ile Yıllara Göre Sonuçlar

Bist İnşaat	2019	2020	2021	2022
ANELE	0,1436896	0,1423263	0,1407401	0,1417227
EDIP	0,1615182	0,1516979	0,1475452	0,1501944
ENKAI	0,1414135	0,1409003	0,1393602	0,1402088
KUYAS	0,1403084	0,1406566	0,1392748	0,1403065

ORGE	0,142473	0,1415717	0,1423824	0,143292
YAYLA	0,140846	0,1422763	0,1434022	0,1403457
YYAPI	0,1297513	0,140571	0,1472951	0,1439299

Tablo 15: SAW Yöntemi ile Yıllara Göre Firma Rankları

Bist İnşaat	2019	2020	2021	2022
ANELE	2	2	5	4
EDIP	1	1	1	1
ENKAI	4	5	6	7
KUYAS	6	6	7	6
ORGE	3	4	4	3
YAYLA	5	3	3	5
YYAPI	7	7	2	2



Şekil 8: Yıllara Göre SAW Sonuç Grafiği

Sonuçlar ve Öneriler

Bu çalışmada, COVID-19 pandemisi ve takip eden makroekonomik süreçte Borsa İstanbul'da (BIST) işlem gören 7 inşaat firmasının (ANELE, EDIP, ENKAI, KUYAS, ORGE, YAYLA, YYAPI) 2019-2022 yıllarını kapsayan 4 yıllık finansal performansları Çok Kriterli Karar Verme (ÇKKV) teknikleriyle bütüncül bir şekilde analiz edilmiştir. Kriter ağırlıklarının tespitinde nesnel verileri temel alan CRITIC yöntemi kullanılırken; performans sıralamalarında TOPSIS, GRA, VIKOR, WASPAS ve SAW yöntemlerinden yararlanılmıştır. Elde edilen bulgular aşağıdaki temel sonuçları ortaya koymaktadır:

1- Salgının etkilerinin en yoğun hissettirildiği 2020 ve 2021 yıllarında, incelenen inşaat şirketlerinin genel finansal performans ve likidite rasyolarında belirgin bir gerileme ve dalgalanma gözlemlenmiştir. Bu durum, uluslararası literatürde vurgulanan şantiye kısıtlamaları, tedarik zinciri kırılmaları, artan işletme maliyetleri ve nakit akışı aksamalarıyla örtüşmektedir.
2- TOPSIS ve VIKOR Yöntemleri: İdeal çözüme yakınlığı ve toplam fayda/pişmanlık dengesini esas alan bu yöntemlerde ORGE firması 2019–2021 döneminde tutarlı bir biçimde en yüksek performansı sergileyerek ilk sırada yer almıştır. 2022 yılı itibarıyla ise KUYAS üstün bir ivme yakalayarak TOPSIS ve VIKOR (Q skoru) sıralamalarında 1. sıraya yükselmiştir. GRA Yöntemi: Gri ilişkisel derecelendirme sonuçlarına göre KUYAS tüm analiz yılları (2019-2022) boyunca istikrarlı bir şekilde 1. sırada yer alırken; onu ikincil sırada ORGE ve üçüncül sırada ENKAI takip etmiştir. WASPAS ve SAW Yöntemleri: Doğrusal aditif ve ağırlıklı bileşik modelleri temel alan WASPAS ve SAW yöntemlerinde ise EDIP firması 2019-2022 yılları arasında kesintisiz olarak 1. sırada konumlanmıştır. YYAPI ise 2019-2020 yıllarında son sıralarda yer alırken, 2021 ve 2022 yıllarında performansını hızla artırarak 2. sıraya kadar yükselmiştir.

3- Finansal Yapı ve ÇKKV Hassasiyeti: Farklı ÇKKV yöntemlerinin sıralamalarında görülen varyasyonlar, yöntemlerin arka planında yatan matematiksel algoritmaların (mesafe bazlı, gri sistem teorisi, ağırlıklı toplam/çarpım modelleri) farklı finansal rasyolara (likidite, kârlılık, borçluluk, devir hızları) verdiği hassasiyetten kaynaklanmaktadır. Yüksek özkaynak kârlılığı, güçlü aktif devir hızı ve dengeli borç yapısına sahip firmaların kriz dönemlerinde daha dirençli kaldıkları somutlanmıştır.

Araştırma bulguları ışığında, inşaat sektörü uygulayıcıları, firma yöneticileri, yatırımcılar ve politika yapıcılar için geliştirilen yönetsel ve stratejik öneriler şu şekildedir:

1- Dirençli Nakit ve Likidite Yönetimi: Kriz ve belirsizlik dönemlerinde "nakit akışı yönetimi" kârlılığın önüne geçmektedir. İnşaat firmalarının işletme sermayesi yetersizliğinden kaynaklanan nakit tükenmesi (*cash burn*) riskine karşı muhafazakâr nakit oranı ve asit-test dengesi muhafaza edilmelidir. Tedarik Zinciri ve Maliyet Risklerinin Yönetimi: Pandemi sürecinde yaşanan malzeme ve navlun artışları göz önüne alınarak, projelerde dinamik maliyet ayarlama mekanizmaları, türev araçlar (hedging) ve tedarikçi çeşitlendirme stratejileri benimsenmelidir. Sermaye Yapısının İyileştirilmesi ve Borç Vadesi Uyumlaştırılması: Yüksek finansal kaldıraçla çalışan inşaat firmaları, artan faiz ve finansman maliyetleri karşısında kırılganlaşmaktadır. Borçların vade yapısının projelerin uzun vadeli hakediş ve nakit akış takvimine uygun hale getirilmesi, özkaynak yapısının güçlendirilmesi hayati önem taşımaktadır.
2- Bütüncül ve Çok Kriterli Değerlendirme: Yatırımcıların yalnızca net kâr veya ciro artışı gibi tekil göstergelere odaklanmak yerine; likidite, borçluluk, faaliyet verimliliği ve kârlılık rasyolarını bir arada değerlendiren ÇKKV bazlı analitik yaklaşımları kararlarına entegre

etmeleri önerilmektedir. Sektörel Duyarlılık ve Erken Uyarı İşaretleri: İnşaat sektörünün yüksek stok yükü ve alacak devir hızlarındaki aksamalar krizlerin ilk habercisidir. BIST'te işlem gören inşaat hisselerinin seçiminde, alacak kalitesi ve bilanço dışı yükümlülükler titizlikle incelenmelidir.

3- Yöntem ve Model Genişletmeleri: Gelecek çalışmalarda finansal oranların ağırlıklandırılmasında BWM, SWARA veya AHP gibi yöntemler CRITIC ile karşılaştırılabilir; firmanın sektörel konumlandırması için BWM-TOPSIS veya Interval-Valued (Aralık Değerli) bulanık ÇKKV modelleri uygulanabilir. Kapsam ve Zaman Genişletmesi: Analiz dönemi pandemi sonrası (2023 ve sonrası) yüksek enflasyon ve faiz ortamını da kapsayacak şekilde genişletilerek krizlerin uzun vadeli kalıcı etkileri incelenebilir.

Kaynakça

Salami, B. A., Ajayi, S. O., & Oyegoke, A. S. (2023). Tackling the impacts of Covid-19 on construction projects: an exploration of contractual dispute avoidance measures adopted by construction firms. *International journal of construction management*, *23*(7), 1196-1204.

Pamidimukkala, A., & Kermanshachi, S. (2021). Impact of Covid-19 on field and office workforce in construction industry. *Project leadership and society*, *2*, 100018.

Gamil, Y., & Alhagar, A. (2020). The impact of pandemic crisis on the survival of construction industry: a case of COVID-19. *Mediterranean Journal of Social Sciences*, *11*(4), 122-128.

Assaad, R., & El-Adaway, I. H. (2021). Guidelines for responding to COVID-19 pandemic: Best practices, impacts, and future research directions. *Journal of management in engineering*, *37*(3), 06021001.

Siraj, N. B., & Fayek, A. R. (2019). Risk identification and common risks in construction: Literature review and content analysis. *Journal of construction engineering and management*, *145*(9), 03119004.

Al-Awadhi, A. M., Alsaifi, K., Al-Awadhi, A., & Alhammadi, S. (2020). Death and contagious infectious diseases: Impact of the COVID-19 virus on stock market returns. *Journal of behavioral and experimental finance*, *27*, 100326.

Zhang, D., Hu, M., & Ji, Q. (2020). Financial markets under the global pandemic of COVID-19. *Finance research letters*, *36*, 101528.

Diakoulaki, D., Mavrotas, G., & Papayannakis, L. (1995). Determining objective weights in multiple criteria problems: The critic method. *Computers & operations research*, *22*(7), 763-770.

Jollyta, D. (2018). TOPSIS technique for selecting of property development location. *Softw. Eng*, *6*(1), 20-26.

Kuo, Y., Yang, T., & Huang, G. W. (2008). The use of grey relational analysis in solving multiple attribute decision-making problems. *Computers & industrial engineering*, *55*(1), 80-93.

Julong, D. (1989). Introduction to grey system theory. *The Journal of grey system*, *1*(1), 1-24.

Opricovic, S., Tzeng, G. H. (2004). Compromise solution by MCDM methods: A comparative analysis of VIKOR and

TOPSIS. *European journal of operational research*, *156*(2), 445-455.

Opricovic, S., Tzeng, G. H. (2007). Extended VIKOR method in comparison with outranking methods. *European journal of operational research*, *178*(2), 514-529.

Zavadskas, E. K., Turskis, Z., Antucheviciene, J., & Zakarevicius, A. (2012). Optimization of weighted aggregated sum product assessment. *Elektronika ir elektrotechnika*, *122*(6), 3-6.

Aghajani Bazzazi, A. B. B. A. S., Osanloo, M., & Karimi, B. (2011). A new fuzzy multi criteria decision making model for open pit mines equipment selection. *Asia-Pacific Journal of Operational Research*, *28*(03), 279-300.

Churchman, C. W., & Ackoff, R. L. (1954). An approximate measure of value. *Journal of the Operations Research Society of America*, *2*(2), 172-187.

Hwang, C. L., & Yoon, K. (2012). *Multiple attribute decision making: methods and applications a state-of-the-art survey*. Springer Science & Business Media.

MacCrimmon, K. R. (1968). *Decisionmaking among multiple-attribute alternatives: a survey and consolidated approach* (No. RM4823ARPA).

BÖLÜM 8

GÖMÜLÜ SİSTEMLERDE DERİN ÖĞRENME: DONANIM PLATFORMLARI, MODEL UYARLAMA VE PERFORMANS DEĞERLENDİRMESİ

VOLKAN YAMAÇLI[1]
SEMİH KAHVECİ[2]
RAMAZAN AKKURT[3]

Giriş

Gömülü sistemler, belirli görevleri yerine getiren donanım ve yazılım bileşenleriyle ev otomasyonundan endüstriyel üretim makinelerine ve robotik uygulamalara kadar geniş bir alanda kullanılmaktadır. Bu sistemlerin topladığı ses, görüntü ve sensör verileri, derin öğrenme modelleri aracılığıyla değerlendirilebilmektedir. Çok katmanlı sinir ağlarının verilerden özellik öğrenebilmesi, elle tanımlanması güç örüntülerin

[1] Doç. Dr., Mersin Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, ORCID: 0000-0003-0331-8818
[2] Dr. Arş. Gör., Mersin Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, ORCID: 0000-0002-1495-6295
[3] Arş. Gör., Mersin Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, ORCID: 0000-0003-2319-9887

işlenmesine olanak sağlamaktadır (LeCun et al., 2015). Böylece çevreyi algılama ve elde edilen bilgiye göre karar verme işlevleri aynı cihaz üzerinde bir araya getirilebilmektedir.

Verilerin uzak sunuculara gönderilmesi geniş hesaplama kaynaklarından yararlanmayı sağlasa da ağ bağlantısına bağımlılık, aktarım gecikmesi ve iletişim yükü doğurmaktadır. Hesaplamanın veri kaynağına yaklaştırılması, yerel karar üretimini desteklemekte ve gönderilmesi gereken veri miktarını azaltabilmektedir (Shi et al., 2016). Bunun karşılığında, modelin cihazın işlem kapasitesine, belleğine ve enerji bütçesine uygun olması gerekir. Dolayısıyla gömülü derin öğrenmede yüksek doğrulukla birlikte uygulamanın çalışma koşulları da değerlendirilmelidir.

Bu bölümde gömülü sistemlerde derin öğrenmenin kullanımına ilişkin temel kavramlar, donanım ve yazılım seçenekleri ile model uyarlama yöntemleri ele alınmaktadır. ESP32, Arduino, Raspberry Pi ve Jetson platformları farklı kaynak koşulları bakımından incelenmekte; ardından model küçültme, performans değerlendirmesi ve bakım süreçleri açıklanmaktadır. Amaç, literatürdeki yaklaşımlardan yararlanarak donanım olanakları ile uygulama gereksinimleri arasındaki ilişkiyi ortaya koymaktır.

Gömülü Sistemlerde Derin Öğrenme ve TinyML

Derin öğrenme, makine öğrenmesinin çok katmanlı yapay sinir ağlarından yararlanan bir alt alanıdır. Bu ağlarda giriş verisi, ardışık katmanlarda işlenerek görev için kullanılabilecek bir gösterime dönüştürülmektedir. Katmanların kullandığı ağırlıklar eğitim sırasında belirlenmektedir. Böylece görüntüdeki bir nesnenin ayırt edilmesi veya bir ses kaydındaki sözcüğün tanınması için gerekli ilişkiler verilerden öğrenilebilmektedir. LeCun et al. (2015), derin öğrenmenin görüntü ve konuşma gibi alanlardaki gelişimini, farklı düzeylerdeki veri temsillerinin öğrenilebilmesiyle açıklamaktadır.

Kullanılan ağ yapısı, verinin özelliklerine göre değişmektedir. Evrişimli sinir ağları, komşu veri elemanları arasındaki ilişkilerden yararlandığı için görüntü işlemede yaygın olarak kullanılmaktadır. Bu ağlar bir boyutlu sensör verilerine de uygulanabilmektedir. Yinelemeli sinir ağlarında ise önceki adımlardan gelen bilgi sonraki işlemlere taşınmakta, böylece verinin zaman içindeki sırası değerlendirilmektedir. Farklı ağ yapılarının gelişimi ve kullanım alanları Schmidhuber (2015) tarafından ayrıntılı olarak incelenmiştir.

Modelin ürettiği çıktı türü problemin doğasına bağlıdır. Sınıflandırmada bir girdinin hangi sınıfa ait olduğu belirlenirken, regresyonda sayısal bir değer tahmin edilmektedir. Bir motorun çalışma sesinin normal veya anormal olarak ayrılması sınıflandırma biçiminde kurulabilir; sıcaklığın ilerleyen zamandaki değerinin tahmin edilmesi ise regresyon problemidir. Bu ayrım, eğitimde kullanılacak hedefleri ve başarımın nasıl ölçüleceğini belirlemektedir. Gömülü uygulamalarda ağın karmaşıklığı, çözülmek istenen problemin gerektirdiği düzeyde tutulmalıdır.

Modelin geliştirilmesinde eğitim ve çıkarım işlemleri birbirinden ayrılmalıdır. Eğitimde ağın çıktısı hedeflenen sonuçla karşılaştırılmakta, kayıp fonksiyonu ile hesaplanan farkı azaltmak üzere ağırlıklar güncellenmektedir. Çıkarımda ise eğitilmiş model yeni girdiler için sonuç üretmekte; ağırlıkların yeniden öğrenilmesi gerekmemektedir (Sze et al., 2017). Yaygın uygulama, modeli yeterli hesaplama kaynağı bulunan bir bilgisayarda eğiterek gömülü cihaza aktarmaktır. Eğitim ve kullanım ortamlarının giriş biçimi, örnekleme düzeni ve ön işleme adımları bakımından uyumlu olması beklenmektedir.

Hesaplamanın gerçekleştirildiği yer, sistemin çalışma biçimini etkilemektedir. Bulut tabanlı yapıda veri uzak sunuculara aktarılırken, uç bilişimde işlem verinin üretildiği cihazda veya

yakındaki bir ağ geçidinde yapılmaktadır. Yapay zekâ modellerinin bu noktalarda çalıştırılması uç yapay zekâ kapsamında ele alınmaktadır. Böylece bazı kararlar sunucu yanıtı beklenmeden üretilebilir; ham veri yerine gerekli sonuçların gönderilmesi haberleşme yükünü azaltabilir (Shi et al., 2016). Bulut ve uç birimleri birlikte de kullanılabilir. Örneğin eğitim ve sürüm yönetimi merkezde yürütülürken günlük çıkarım işlemleri cihaz üzerinde gerçekleştirilebilir.

TinyML, bellek, işlem kapasitesi ve enerji kaynakları sınırlı cihazlarda makine öğrenmesi modellerinin kullanılmasına odaklanmaktadır. Özellikle mikrodenetleyici uygulamalarında ağ mimarisi ile çıkarım yazılımının birlikte tasarlanması önem taşımaktadır (Lin et al., 2023). Lin et al. (2020) tarafından geliştirilen MCUNet yaklaşımında bu iki bileşen birlikte ele alınmış ve mikrodenetleyiciler üzerinde görüntü sınıflandırma uygulamaları gerçekleştirilmiştir. Bu tür çalışmalar, modelin yalnızca küçültülmesinin değil, çalıştırılma biçiminin de kaynak kullanımını belirlediğini göstermektedir.

TinyML ile uç yapay zekâ birbiriyle ilişkili olmakla birlikte aynı kapsamı ifade etmemektedir. GPU destekli bir gömülü bilgisayarda büyük bir görüntü modelinin çalıştırılması da uç yapay zekâ uygulamasıdır. TinyML ise daha dar kaynak koşullarına odaklanmaktadır. Ayrıca TinyML kapsamında yalnızca derin sinir ağları değil, daha küçük makine öğrenmesi yöntemleri de kullanılabilir. Bu nedenle kavram, belirli bir model ailesinin veya kartın adı olarak değerlendirilmemelidir (Lin et al., 2023).

Tablo 1, bulut, uç yapay zekâ ve TinyML yaklaşımlarını genel özellikleri bakımından karşılaştırmaktadır. Bu yaklaşımlar birbirini dışlamaz; aynı uygulamada birlikte kullanılabilir. Verinin yerel işlenmesi gizlilik açısından yarar sağlayabilse de saklanan verilerin ve aktarılan çıktıların korunması ayrıca ele alınmalıdır.

Tablo 1. Bulut, uç yapay zekâ ve TinyML yaklaşımlarının karşılaştırılması

Özellik	Bulut tabanlı yapay zekâ	Uç yapay zekâ	TinyML
Hesaplama yeri	Uzak sunucu veya veri merkezi	Veri kaynağında veya yakın bir işlem biriminde	Genellikle mikrodenetleyicinin üzerinde
Temel kaynak koşulu	Daha geniş işlem ve depolama kaynakları	Kullanılan uç cihazın kapasitesine bağlı kaynaklar	Sınırlı bellek, işlem kapasitesi ve enerji bütçesi
Bağlantının rolü	Uzaktaki işlem birimine veri aktarımı gerekir	Yerel çıkarım sürekli bağlantı gerektirmeyebilir	Yerel çıkarım bağlantısız yürütülebilir; güncelleme ayrı ele alınır
Uygulama örneği	Sunucuda görüntü analizi	Robot bilgisayarında nesne tespiti	Küçük sensör cihazında anahtar sözcük algılama

Kaynak: Shi et al. (2016) ve Lin et al. (2023) temelinde hazırlanmıştır.

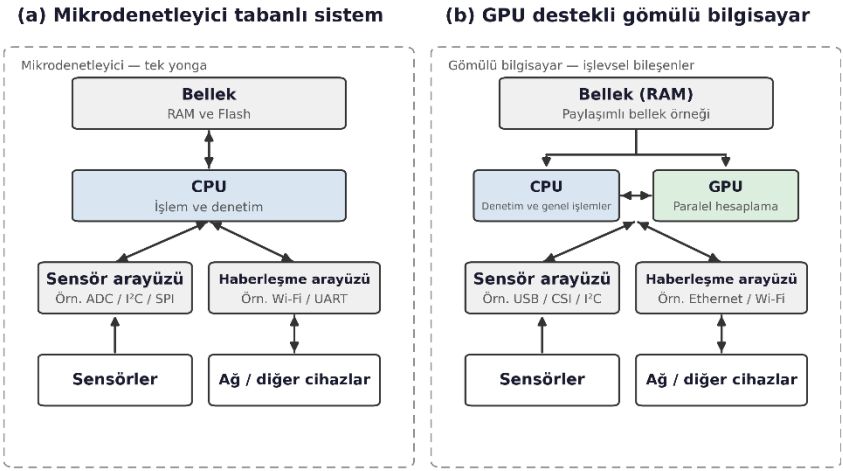
Donanım Platformları ve Yazılım Altyapısı

Bir gömülü sistem, işlem birimi, bellek, giriş-çıkış birimleri ve haberleşme arayüzleriyle belirli görevleri yerine getirmektedir. Sensörlerden alınan bilgiler işlenmekte; gerektiğinde motor veya valf gibi aktüatörler üzerinden fiziksel sürece müdahale edilmektedir (Marwedel, 2021). Mikrodenetleyicilerde işlemci, bellek ve çevre birimlerinin önemli bölümü aynı yongada bulunurken, gömülü bilgisayarlar daha geniş bellek ve farklı hızlandırıcı seçenekleri sunabilmektedir. Bu nedenle küçük bir sensör düğümü ile kamera görüntülerini değerlendiren robot bilgisayarı aynı kaynak koşullarına sahip değildir.

İşlem kapasitesinin değerlendirilmesinde saat frekansı tek başına yeterli değildir. İşlemci mimarisi, desteklenen komutlar ve bellek erişimi de çalışma süresini etkilemektedir. GPU birimleri,

uygun matris ve evrişim hesaplamalarını paralel yürütebilse de veri hazırlama ve haberleşme gibi diğer görevler aynı ölçüde hızlanmayabilir. Sze et al. (2017), sinir ağlarının verimli çalıştırılmasında aritmetik işlemler kadar bellekler arasındaki veri hareketinin de önemli olduğunu belirtmektedir. İşlem biriminin katkısı bu nedenle uygulamanın bütünü içinde değerlendirilmelidir. Mikrodenetleyici ve GPU destekli gömülü bilgisayar yapısına dair şematik Şekil 1 ile sunulmuştur.

Şekil 1. Mikrodenetleyici ve GPU destekli gömülü bilgisayar yapısı



Kaynak: Marwedel (2021) ile Arduino (t.y.), Espressif Systems (t.y.), Raspberry Pi Ltd (t.y.-a, t.y.-b) ve NVIDIA (t.y.-b) temelinde hazırlanmıştır.

Bellek gereksinimi, model dosyasının boyutundan daha geniştir. Ağırlıkların saklandığı alanın yanında ara katman çıktıları, sensör tamponları ve uygulama değişkenleri için çalışma belleği gerekir. Örneğin bir görüntünün ham ve yeniden boyutlandırılmış kopyalarının birlikte tutulması ek alan kullanmaktadır. Çevre birimleri de verinin istenen hızda ve kesintisiz alınmasını sağlamalıdır. Örnekleme aralığının değişmesi veya veri kaybı, eğitimde kullanılan giriş düzenini bozabilir. Özellikle birden fazla

sensörün bulunduğu sistemlerde zamanlama ve ortak kaynak kullanımı başlangıçta planlanmalıdır.

ESP32, farklı işlemci ve çevre birimi özelliklerine sahip yongalardan oluşan bir ailedir. Bu nedenle kullanılan ürünün tam modeli belirtilmelidir. Espressif'in ESP-DL dokümantasyonunda ESP32-S3 ve ESP32-P4, derin öğrenme uygulamaları için önerilen hedefler arasında yer almakta; özgün ESP32 için işleç uygulamalarının farklı olduğu açıklanmaktadır. ESP-IDF üzerinde çalışan ESP-DL ile model kullanılabilmesi, desteklenen işleçlere ve uygun model gösterimine bağlıdır. Nicemleme ve dışa aktarma için sunulan ESP-PPQ aracının ayarları da hedef yongayla eşleşmelidir (Espressif Systems, t.y.).

Arduino da tek bir kartın adı olarak değerlendirilmemelidir. Aynı ekosistem içinde farklı mikrodenetleyiciler, bellek kapasiteleri ve kütüphane olanakları bulunmaktadır. Nano 33 BLE Sense Rev2, nRF52840 mikrodenetleyicisi ve kart üzerindeki sensörleriyle küçük makine öğrenmesi uygulamalarında kullanılabilecek örneklerden biridir (Arduino, t.y.). Ses ve hareket verilerinin kart üzerinde elde edilmesi geliştirmeyi kolaylaştırabilir. Bununla birlikte ölçümlerin etiketlenmesi, eksen sırasının ve örnekleme hızının belirlenmesi, girişlerin modelin beklediği düzene getirilmesi yine uygulama yazılımının görevleri arasındadır.

Raspberry Pi 5, işletim sistemi çalıştırabilen bir tek kart bilgisayar olarak sensör erişimini, veri işlemeyi ve haberleşmeyi aynı ortamda birleştirmektedir (Raspberry Pi Ltd, t.y.-b). Uygun büyüklükteki modeller CPU üzerinde çalıştırılabilir. Kartta GPU bulunması, çıkarım yazılımının bu birimden doğrudan yararlanacağını göstermez. Desteklenen modellerin haricî bir sinir ağı işlem biriminde çalıştırılması için AI HAT seçenekleri de bulunmaktadır (Raspberry Pi Ltd, t.y.-a). Bu durumda ek donanım

ve yazılım uyumu gerekir; ana kartın tek başına kullanımı ile NPU eklenen yapı ayrı değerlendirilmelidir.

Jetson platformları, CPU ve GPU kaynaklarını gömülü bir bilgisayarda bir araya getirmektedir. NVIDIA'nın JetPack paketi, Jetson Linux ile hesaplama kütüphanelerini ve geliştirme araçlarını birlikte sunmaktadır. Jetson Nano ile Orin ailesi arasında donanım ve yazılım nesli bakımından fark bulunmaktadır. Jetson Nano'yu destekleyen JetPack 4 serisinin yaşam döngüsünün Kasım 2024'te sona ermiş olması, mevcut cihazın çalışmasını durdurmaz; ancak yeni uygulamalarda bakım ve sürüm seçimi açısından dikkate alınmalıdır (NVIDIA, t.y.-a). Kullanılan örneklerin hangi kart ve yazılım sürümü için hazırlandığı bu nedenle kontrol edilmelidir.

Güç ve soğutma koşulları da uzun süreli kullanımda değerlendirilmelidir. Kısa bir denemede elde edilen çalışma hızı, cihazın sürekli yük altında gösterdiği davranışı bütünüyle açıklamayabilir. Bu nedenle deneylerde kullanılan güç ayarı, soğutma düzeni ve ölçüm süresi belirtilmelidir. Kartın üzerinde çalıştırılacak diğer görevler de aynı denemeye dâhil edilmelidir. Böylece laboratuvar ortamında çalışan bir modelin uygulamanın tamamıyla birlikte hangi koşulları karşılayabildiği daha açık görülebilir.

Tablo 2, ele alınan platformların genel konumunu özetlemektedir. Tablo ortak bir hız deneyi veya üstünlük sıralaması içermemektedir; aynı ailedeki ürünlerin kaynakları ve yazılım olanakları da değişebilmektedir.

Tablo 2. İncelenen donanım platformlarının genel konumu

Platform	**Donanım yapısı**	**Ele alınabilecek görev**	**Seçimde öne çıkan konu**
ESP32 ailesi; S3 ve P4 hedefleri	Yongaya özgü mikrodenetleyici kaynakları	Sensör verisinde küçük modellerle yerel değerlendirme	İşleç desteği ve kullanılabilir bellek

Platform	Donanım yapısı	Ele alınabilecek görev	Seçimde öne çıkan konu
Arduino Nano 33 BLE Sense Rev2	nRF52840 ve kart üzeri sensörler	Ses ve hareket verileriyle TinyML geliştirme	Giriş verisinin hazırlanması ve model boyutu
Raspberry Pi 5	İşletim sistemi çalıştıran tek kart bilgisayar	CPU üzerinde çıkarım; uygun ek NPU ile hızlandırma	Model yükü, yazılım desteği ve ek hızlandırıcı seçimi
Jetson Nano	CPU ve GPU içeren gömülü bilgisayar	Görüntü işleme ve robotik prototipleri	JetPack 4 desteğinin yaşam döngüsü
Jetson Orin ailesi	GPU destekli farklı kapasitedeki modüller	Daha kapsamlı görsel algılama ve robotik iş yükleri	Tam ürün modeli, güç ve yazılım uyumu

Kaynak: Arduino (t.y.), Espressif Systems (t.y.), Raspberry Pi Ltd (t.y.-a, t.y.-b) ve NVIDIA (t.y.-a) temelinde hazırlanmıştır.

Modelin çalıştırılmasında eğitim ortamı, dönüştürme araçları ve çıkarım yazılımı farklı görevler üstlenmektedir. TensorFlow Lite Micro, sınırlı kaynaklara sahip gömülü cihazlarda model çalıştırmak için geliştirilmiştir. Gerekli işleçlerin seçilmesi ve çalışma alanının ayrılması, sensör ve kontrol yazılımıyla birlikte yürütülen uygulamanın parçalarıdır (David et al., 2020). CMSIS-NN ise Arm Cortex-M işlemcileri için optimize edilmiş sinir ağı işlem işlevlerini sağlamaktadır (Lai et al., 2018). Buradaki çekirdek, ek bir fiziksel işlemciyi değil yazılım işlevini ifade eder. Bu nedenle iki bileşen birbirinin doğrudan alternatifi değildir; uygun çıkarım altyapısı CMSIS-NN işlevlerinden yararlanabilir.

NVIDIA sistemlerinde kullanılan TensorRT, eğitilmiş modellerin GPU üzerinde çıkarım için optimize edilmesine yönelik araçlar sunmaktadır. Modelden hedef ortama uygun bir çalıştırma yapısı oluşturulmakta ve desteklenen sayısal gösterimlerden yararlanılmaktadır (NVIDIA, t.y.-b). Kullanılabilecek özellikler

GPU ve yazılım sürümleriyle ilişkilidir. Özellikle farklı Jetson nesillerinde aynı TensorRT sürümünün veya aynı model dosyasının doğrudan kullanılabileceği varsayılmamalıdır.

Platform seçimine, işlenecek verinin ve beklenen kararın tanımlanmasıyla başlanmalıdır. Girdi boyutu, örnekleme sıklığı, kabul edilebilir hata ve yanıt süresi belirlendikten sonra bellek ve enerji gereksinimleri değerlendirilebilir. Sürekli çalışan bir sensör düğümünde enerji bütçesi öncelikli olabilirken, birden fazla görüntü akışını işleyen sistemde hesaplama ve veri aktarımı öne çıkabilir. Modelin gerektirdiği işleçlerin desteklenmesi ve diğer görevlerle birlikte çalışabilmesi de aranmalıdır. Model, yazılım ve donanımın ortak tasarımı bu kararların birlikte ele alınmasını sağlamaktadır (Lin et al., 2023).

Geliştirme kartının edinilmesi toplam maliyetin yalnızca bir bölümüdür. Sensörler, güç kaynağı, bağlantı birimleri ve gerekiyorsa soğutma çözümü de sisteme dâhil edilmektedir. Yazılımın kurulması, sürümlerin uyumlu tutulması ve uygulamanın güncellenmesi için gereken çalışma da dikkate alınmalıdır. Bu bakımdan uygun platform, yalnızca ilk denemeyi çalıştıran değil, hedeflenen kullanım koşullarında sürdürülebilir bir uygulama kurulmasına olanak veren platformdur.

Model Seçimi ve Donanıma Uyarlama

Bir model üzerinde değişiklik yapılmadan önce hangi kaynağın yetersiz kaldığı belirlenmelidir. Model dosyasının saklama alanına sığmaması ile çıkarım sırasında çalışma belleğinin tükenmesi farklı sorunlardır. Benzer şekilde, belleğe sığan bir ağ istenen sürede karar üretemeyebilir. Lin et al. (2023), küçük cihazlar için model geliştirilirken ağırlıkların yanında ara hesaplama sonuçlarının da önemli bir bellek gereksinimi oluşturduğunu açıklamaktadır. Dolayısıyla yalnızca parametre sayısına bakılarak verilecek bir karar, cihazdaki asıl sınırlamayı gözden kaçırabilir.

Başlangıçta mevcut modelin dosya boyutu, en yüksek çalışma belleği kullanımı ve çıkarım süresi kaydedilebilir. Sonraki değişiklikler bu başlangıç değerleriyle karşılaştırılır. Uygulama açısından kabul edilebilir hata düzeyi de önceden tanımlanmalıdır. Örneğin bir sınıflandırma probleminde toplam doğruluk korunurken belirli bir sınıfın daha sık karıştırılması mümkündür. Bu nedenle küçültülen modelin yalnızca ortalama başarısı değil, uygulamanın önem verdiği durumlarda nasıl davrandığı da incelenmelidir. Böylece optimizasyonun hangi kazanım karşılığında hangi kayba yol açtığı daha açık görülebilir.

Kaynak kullanımını azaltmanın bir yolu, başlangıçtan itibaren daha az hesaplama gerektiren bir ağ seçmektir. Hafif mimarilerde katmanların sayısı, kanalların genişliği ve kullanılan işlemlerin türü bu amaçla düzenlenmektedir. MobileNetV2, kanal bazında evrişim işlemlerini ters artık bağlantılar ve doğrusal darboğaz yapılarıyla bir araya getiren bir örnektir. Bu tasarım, görüntü verisinin işlenmesinde hesaplama yükü ile temsil kapasitesi arasında uygun bir denge kurulmasını hedeflemektedir (Sandler et al., 2018).

Hafif bir mimarinin farklı büyüklükteki sürümleri aynı uygulamada değerlendirilebilir. Kanal sayısının azaltılması ağın işlem yükünü değiştirirken, giriş çözünürlüğünün düşürülmesi işlenen görüntü miktarını azaltmaktadır. Buna karşılık görüntüdeki küçük ayrıntılar kaybolabilir. Örneğin yüzey üzerindeki küçük bir kusurun belirlenmesinde, çözünürlüğü azaltmak modelin ayırt etmesi gereken bilgiyi de ortadan kaldırabilir. Bu yüzden giriş boyutuna ilişkin karar, yalnızca çalışma süresi üzerinden verilmemelidir.

Model seçimi otomatik arama yöntemleriyle de desteklenebilmektedir. Cai et al. (2019) tarafından sunulan Once-for-All yaklaşımında, farklı derinlik, genişlik ve giriş boyutlarına

sahip alt ağların ortak bir eğitim yapısından seçilmesi ele alınmaktadır. Böylece farklı cihaz gereksinimleri için model seçenekleri oluşturulabilmektedir. Bununla birlikte, bir arama yönteminin kullanılması sonucun her karta doğrudan aktarılabileceğini göstermez. Seçilen alt ağın hedef cihazda desteklenen işlemlerle çalışması ve öngörülen kaynak sınırlarını karşılaması yine kontrol edilmelidir.

Modelin çalışma belleği gereksinimi de mimari seçimle birlikte düzenlenebilir. Ara katman sonuçlarının bir bölümüne yalnızca belirli işlemler tamamlanıncaya kadar ihtiyaç duyulur. TensorFlow Lite Micro, bu kullanım sürelerini dikkate alarak bellek alanlarının yeniden kullanılmasına olanak sağlamaktadır (David et al., 2020). Böylece artık gerekli olmayan bir ara sonucun alanı sonraki hesaplamaya ayrılabilir. Model ağırlıklarını küçültmek ile çalışma belleğini planlamak farklı kaynaklara yönelik işlemlerdir; her ikisinin etkisi ayrı ölçülmelidir.

Bir başka yaklaşım, büyük ara verilerin tamamını aynı anda tutmak yerine işlemi parçalara ayırmaktır. Lin et al. (2023), küçük cihazlarda yüksek bellek gerektiren ilk katmanlar için parça tabanlı çıkarımı ele almaktadır. Ancak komşu parçaların hesaplanmasında ortak giriş bölgelerine ihtiyaç duyulabilir. Bu durum bazı işlemlerin tekrarlanmasına yol açtığından, bellek tasarrufu ile ek hesaplama yükü birlikte değerlendirilmelidir. Ağın yapısı ile çalıştırma sırasının birlikte düzenlenmesi burada önem kazanmaktadır.

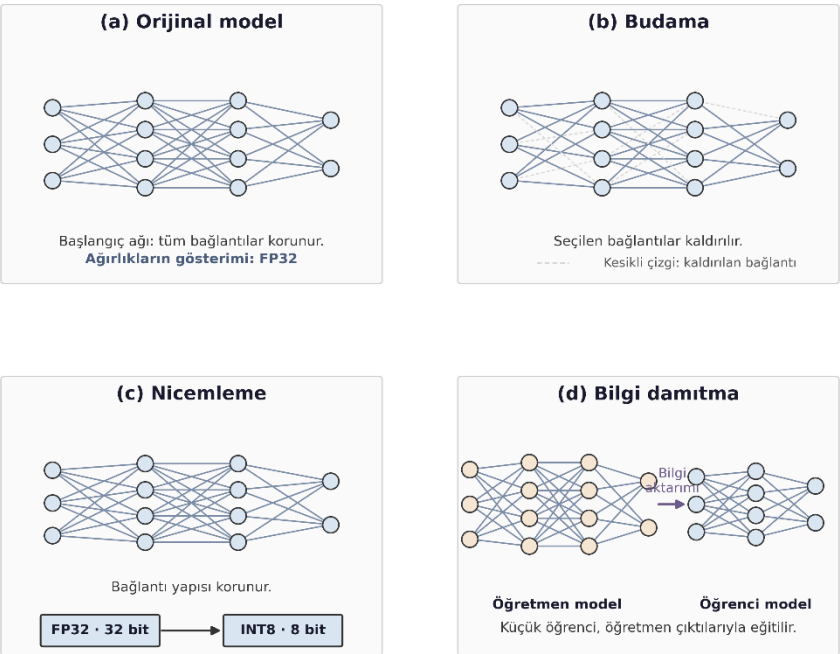
Son aşamada model, hedef çıkarım ortamının desteklediği biçime dönüştürülmelidir. Bilinen girişler için bilgisayarda ve cihazda üretilen sonuçlar karşılaştırılırken veri sırası, normalizasyon, sayısal gösterim ve işleç uyumu kontrol edilir. Ardından gerçek sensör verileriyle, haberleşme ve diğer görevler açıkken değerlendirme yapılır. Böylece dönüştürme hataları ile veri toplama veya kaynak paylaşımı sorunları birbirinden ayrılabilir.

Küçük cihazlarda model ile çalıştırma altyapısının birlikte ele alınması, David et al. (2020) tarafından da vurgulanmaktadır.

Model Küçültme Yöntemleri

Model küçültme yöntemlerinden nicemleme, budama ve bilgi damıtma, Şekil 2'de şematik olarak gösterilmiştir.

Şekil 2. Derin öğrenme modellerinde küçültme yöntemleri



Kaynak: Sandler et al. (2018), Jacob et al. (2018), Han et al. (2015), Hinton et al. (2015) ve David et al. (2020) temelinde hazırlanmıştır.

Nicemleme, modeldeki sayısal değerlerin daha az sayıda bit kullanılarak temsil edilmesidir. Örneğin 32 bit kayan noktalı olarak saklanan ağırlıklar, uygun bir dönüşümle 8 bit tamsayı gösterimine aktarılabilir. Yalnızca ağırlık değerleri dikkate alındığında bu değişiklik, aynı sayıdaki parametrenin saklama ihtiyacını dörtte bire indirir. Model dosyasındaki diğer bilgiler, ölçek değerleri ve farklı duyarlıkta tutulan bileşenler nedeniyle bütün dosyada aynı küçülme

oranı elde edilmeyebilir. Jacob et al. (2018), ağırlıklar ve ara değerler için nicemleme düzenini eğitim ve tamsayı çıkarımıyla birlikte ele almaktadır.

Eğitim sonrası nicemlemede, eğitilmiş model üzerinde sayısal dönüşüm yapılmaktadır. Yaygın uygulamalarda ara değerlerin uygun temsil aralıklarını belirlemek için bir kalibrasyon veri grubu kullanılır. Bu verilerin cihazda karşılaşılacak girişleri temsil etmesi gerekir. Yalnızca tek bir aydınlatma koşulundaki görüntülerle belirlenen aralıklar, farklı koşullarda oluşan değerleri yeterince karşılamayabilir. Nagel et al. (2021), eğitim sonrası nicemlemede aralık seçiminin ve yuvarlama işlemlerinin model başarısına etkisini açıklamaktadır. Kalibrasyon için kullanılacak örnekler, nihai başarının raporlanacağı bağımsız test kümesinden ayrı tutulmalıdır.

Nicemleme farkındalıklı eğitimde ise düşük duyarlığın oluşturacağı sayısal etkiler eğitim sırasında hesaba katılır. Ağın ağırlıkları, bu etkiler altında uygun çıktılar üretecek şekilde yeniden düzenlenebilir. Eğitim sonrası dönüşümde kabul edilemeyecek bir başarı kaybı oluşması durumunda bu yaklaşım değerlendirilebilir. Ancak yeniden eğitim için veri, zaman ve hesaplama kaynağı gerekmektedir. Hangi yöntemin yeterli olacağı, kullanılan model ve hedeflenen sayısal gösterime bağlıdır (Jacob et al., 2018; Nagel et al., 2021).

Nicemlenmiş bir modelde bütün işlemler aynı bit genişliğinde yürütülmeyebilir. Örneğin 8 bit değerlerle yapılan çarpımlar daha geniş bir tamsayı alanında biriktirilebilir; bazı işleçler ise farklı duyarlıkla çalışabilir. Bu nedenle “8 bit model” ifadesi tek başına hız kazanımını açıklamaz. Kullanılan işleçlerin hedef donanımda nasıl yürütüldüğü ve dönüşüm sonrasında görev başarısının korunup korunmadığı değerlendirilmelidir.

Bununla birlikte budama, bir ağdaki bağlantıların veya yapısal bileşenlerin belirli ölçütlere göre kaldırılmasıdır. Amaç, modelin ürettiği sonuçlara katkısı sınırlı olan kısımların kaynak tüketimini azaltmaktır. Han et al. (2015), bağlantıların budanmasını, nicemleme ve kodlama adımlarıyla birleştirerek modelin saklama ihtiyacını azaltan bir yöntem sunmuştur. Bu çalışma, farklı sıkıştırma işlemlerinin birbirini tamamlayabileceğini göstermektedir. Bununla birlikte, belirli ağlarda elde edilen küçülme oranlarının bütün modeller için geçerli olduğu düşünülmemelidir.

Yapısal olmayan budamada tek tek ağırlıklar kaldırılır veya sıfırlanır. Bu işlem sonrasında ağırlık dizisinin boyutları değişmeden kalabilir. Kullanılan yazılım sıfır değerleri de yoğun matris işlemleri içerisinde hesaplıyorsa, dosyadaki seyrekliğin aynı oranda hızlanma sağlaması beklenemez. Seyrek gösterimin saklanması için konum bilgilerine de ihtiyaç duyulabilir. Yapısal budamada ise kanal ya da filtre gibi bütün birimler kaldırılmaktadır.

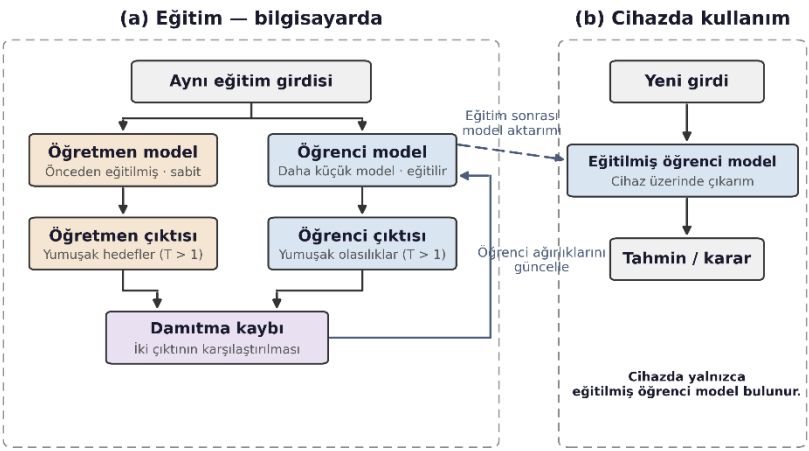
Bu yaklaşım, katmanların boyutlarını küçülttüğü için standart hesaplama yapılarında daha doğrudan bir azalma sağlayabilir; yine de cihazdaki süre ölçümü belirleyicidir (Sze et al., 2017). Budama sonrasında kalan ağın yeniden eğitilmesi, oluşan başarı kaybının bir bölümünü giderebilir (Han et al., 2015). Ancak toplam doğruluk az değişirken seyrek görülen bir arıza sınıfının başarısı düşebilir. Bu nedenle budama oranı belirlenirken dosya boyutuyla birlikte uygulama açısından önemli sınıfların sonuçları da incelenmelidir. Daha küçük bir model elde edilmiş olması, yapılan düzenlemenin tek başına yeterli olduğunu göstermez.

Bilgi damıtma ise daha kapsamlı bir modelin ürettiği bilgiden yararlanılarak daha küçük bir modelin eğitilmesidir. Yaygın öğretmen-öğrenci düzeninde büyük model öğretmen, cihazda çalıştırılması hedeflenen küçük model ise öğrenci olarak ele alınır. Öğrenci, doğru sınıf etiketlerinin yanında öğretmenin sınıflar

arasında kurduğu olasılık dağılımından da yararlanabilir. Hinton et al. (2015), bu dağılımların küçük modelin eğitiminde kullanılmasını açıklamaktadır. Böylece eğitim sürecinde daha zengin bir yönlendirme elde edilmesi amaçlanmaktadır. Bilgi damıtmada öğrencinin öğretmenin bütün katmanlarını taşıması gerekmez; öğrenci ağ hedef cihazın kaynaklarına uygun seçilebilir. Eğitim tamamlandığında cihazda yalnızca öğrenci çalıştırılır ve öğretmene ihtiyaç duyulmaz (Hinton et al., 2015).

Bununla birlikte öğretmenin yüksek başarısı, öğrencinin aynı düzeye ulaşacağını garanti etmez. Öğrenci kapasitesinin yetersizliği veya eğitim verisinin uygulamayı temsil etmemesi kazanımı sınırlayabilir. Bu yaklaşım hem mikrodenetleyiciler için küçük modeller geliştirmede hem de gömülü bilgisayarlarda çıkarım yükünü azaltmada değerlendirilebilir. Bilgi damıtmada bilgisayarda yürütülen eğitim ve eğitilmiş modelin gömülü cihazda kullanımı, Şekil 3'te şematik olarak gösterilmiştir.

Şekil 3. Bilgi damıtmada eğitim ve cihazda kullanım



Kaynak: Hinton et al. (2015) temelinde hazırlanmıştır. (T: sıcaklık parametresi)

Tablo 3, model seçimi ve küçültme yöntemlerini değiştirdikleri bileşenler açısından özetlemektedir. Yöntemler

birlikte kullanılabilir; ancak her düzenlemenin ayrı kaydedilmesi ve ölçülmesi, başarı kaybının veya kaynak kazanımının hangi değişiklikten doğduğunu belirlemeyi kolaylaştırır.

Tablo 3. Gömülü sistemler için başlıca model uyarlama yöntemleri

Yöntem	Değiştirilen bileşen	Beklenen katkı	Kontrol edilmesi gereken husus
Hafif mimari seçimi	Ağın katman ve kanal düzeni	Hesaplama ve parametre yükünün azalması	Görev başarısı ve gerçek çıkarım süresi
Nicemleme	Ağırlık ve ara değerlerin sayısal gösterimi	Saklama ihtiyacının ve uygun donanımda işlem yükünün azalması	Sayısal hata, kalibrasyon ve işleç desteği
Budama	Bağlantılar, kanallar veya filtreler	Ağın saklama ve hesaplama ihtiyacının azalması	Kalan ağın başarısı ve seyrekliğin nasıl işlendiği
Bilgi damıtma	Küçük modelin eğitimde aldığı yönlendirme	Daha küçük bir öğrenci ağın geliştirilmesi	Öğrenci kapasitesi ve öğretmen bilgisinin aktarımı
Bellek planlama	Ara sonuçların tutulduğu alan ve işlem sırası	En yüksek çalışma belleği ihtiyacının azalması	Veri bağımlılıkları ve olası ek hesaplamalar

Kaynak: Sandler et al. (2018), Jacob et al. (2018), Han et al. (2015), Hinton et al. (2015) ve David et al. (2020) temelinde hazırlanmıştır.

Performans Değerlendirmesi

Gömülü bir modelin değerlendirilmesinde görev başarısı, gecikme, bellek ve enerji tüketimi birlikte ele alınmalıdır. Ölçümler, başlangıçta tanımlanan gereksinimlerin karşılanıp karşılanmadığını göstermelidir. Kullanılan kartın tam modeli, çalışma yazılımı, giriş boyutu ve eş zamanlı görevler belirtilmeden verilen sonuçların karşılaştırılması güçleşmektedir. Özellikle model tek başına çalışırken elde edilen değerlerle bütün uygulama devredeyken elde

edilen değerler ayrılmalıdır. Başarı ölçütü, uygulamanın görevine uygun seçilmelidir. Sınıflandırmada doğruluk, bütün örnekler içerisinde doğru sınıflandırılanların oranını göstermektedir. Bununla birlikte, sınıfların dengesiz dağıldığı bir veri kümesinde yalnızca doğruluk değerinin verilmesi yanıltıcı olabilir. Örneğin normal çalışma verisinin çok fazla olduğu bir durum izleme uygulamasında, az sayıdaki arıza örneğinin kaçırılması genel doğruluk içerisinde yeterince görünmeyebilir. Bu durumda sınıf bazında kesinlik ve duyarlılık değerleri ile bunların harmonik ortalaması olan F1 puanı değerlendirmeyi tamamlar. Makro F1, her sınıfın F1 puanına eşit ağırlık verir. Sokolova ve Lapalme (2009), sınıflandırma ölçütlerini görev yapısı ve değerlendirdikleri özellikler açısından incelemektedir. Dolayısıyla sonuçların yorumlanmasında seçilen ölçütün hangi hataları görünür kıldığı açıklanmalıdır.

Zaman ölçümünde ise modelin çıkarım süresi ile uygulamanın toplam yanıt süresi ayrılmalıdır. Sensör verisinin toplanması, girişin hazırlanması, modelin çalıştırılması ve sonucun işlenmesi ayrı zaman bileşenleridir. Zhang et al. (2017), sesli komut uygulamalarındaki kart gösteriminde bellek kopyalama ve ses özelliklerinin çıkarılmasını da süre ölçümüne katmıştır. Buna karşılık Banbury et al. (2021) tarafından sunulan MLPerf Tiny çalışmasının ölçüm aralığı model çıkarımına odaklanmakta, ön ve son işlemeyi kapsamamaktadır. Her iki yaklaşımın amacı farklıdır; bir sonuç başka bir çalışmayla karşılaştırılırken ölçümün başlangıç ve bitiş noktaları bilinmelidir. Uygulama değerlendirmesinde tek bir çalıştırmadan elde edilen süre yerine, tekrarlanan ölçümlerin dağılımı incelenebilir. Ortalama gecikmeye ek olarak yüksek gecikmelerin görülme sıklığı, zaman sınırı bulunan görevlerde yararlı bilgi sağlar. Video uygulamalarında saniyede işlenen kare sayısı da tek bir karenin sistemden geçiş süresiyle aynı anlama gelmemektedir. Birikmiş görüntüleri yüksek hızda işleyen bir

düzenek, kullanıcıya gecikmiş sonuçlar üretebilir. Bu nedenle ölçüm, kararın kullanılacağı noktaya kadar uzatılmalıdır.

Bellek ve enerji değerlendirmesi de ölçüm kapsamının tanımlanmasını gerektirir. Model dosyasının boyutu ile çalışırken gereken en yüksek RAM miktarı ayrı değerlerdir. Benzer biçimde güç, birim zamandaki enerji kullanımını; enerji ise tanımlanan işlem veya çalışma süresindeki toplam tüketimi ifade eder. Pil ömrü değerlendirilirken çıkarımın yanında sensörler, haberleşme ve bekleme süreleri de hesaba katılmalıdır. MLPerf Tiny, çıkarım başarısı, gecikme ve enerjiyi birlikte ele alarak bu tür karşılaştırmalar için ortak bir değerlendirme yaklaşımı sunmaktadır (Banbury et al., 2021). Tablo 4'te temel ölçütler ve ölçüm kapsamları verilmiştir.

Tablo 4. Gömülü derin öğrenme uygulamalarında değerlendirme ölçütleri

Değerlendirme ölçütü	**Nasıl ölçülür?**	**Uygulama açısından önemi**
Görev başarısı	Ayrılmış test verisinde göreve uygun ölçüt; sınıflandırmada sınıf bazında kesinlik, duyarlılık ve F1	Hataların türünü ve sınıflara dağılımını gösterir.
Çıkarım gecikmesi	Modelin çalışmaya başlaması ile çıktısının hazır olması arasındaki süre	Model ve çalışma yazılımının işlem süresini belirler.
Toplam yanıt süresi	Tanımlanan veri toplama başlangıcından kararın kullanılmasına kadar geçen süre	Uygulamanın zaman gereksinimini karşılayıp karşılamadığını gösterir.
Bellek kullanımı	Modelin saklama boyutu ve uygulamanın en yüksek RAM ihtiyacı ayrı ölçülür.	Modelin ve diğer görevlerin birlikte çalışabilirliğini değerlendirir.
Enerji tüketimi	Tanımlanan işlemde veya çalışma aralığında güç, zaman üzerinden bütünleştirilir.	Çalışma sıklığıyla birlikte pil ömrünün değerlendirilmesini sağlar.

Değerlendirme ölçütü	Nasıl ölçülür?	Uygulama açısından önemi
Çalışma sürekliliği	Uzun süreli deneyde veri kaybı, yeniden başlama ve gecikme değişimleri kaydedilir.	Kısa testlerde görülmeyen işletim sorunlarını ortaya çıkarır.

Kaynak: Sokolova ve Lapalme (2009), Zhang et al. (2017), Banbury et al. (2021) ve Lin et al. (2023) temelinde hazırlanmıştır.

Gerçek Çalışma Koşulları ve Model Güncelleme

Laboratuvarda elde edilen sonuçların gerçek kullanımda korunabilmesi, değerlendirme verisinin kullanım ortamını ne ölçüde temsil ettiğine bağlıdır. Kamera açısının değiştirilmesi, mikrofonun farklı bir yere konulması veya sensörün başka bir yüzeye bağlanması, modele ulaşan veriyi değiştirebilir. Bu nedenle geliştirme sırasında yalnızca uygun koşullarda toplanmış kayıtların kullanılması yeterli olmayabilir. Değerlendirme planında, uygulamada karşılaşılması beklenen ortam ve kullanıcı farklılıklarına da yer verilmelidir.

Verinin eğitim ve test kümelerine ayrılması bu noktada dikkat gerektirir. Aynı kaydın birbirine çok benzeyen parçalarının iki kümeye dağılması, modelin yeni bir ortamda göstereceği başarıyı olduğundan yüksek gösterebilir. Örneğin bir titreşim kaydından çıkarılan örtüşen zaman pencereleri, ortak ölçümler içerir. Bu bölüm için önerilen değerlendirme yaklaşımında, ayrımın uygulamaya göre kayıt, kullanıcı, cihaz veya zaman dönemi üzerinden yapılması uygundur. Böylece testin hangi genelleme koşulunu temsil ettiği açık biçimde ifade edilebilir.

Uzun süreli kullanımda başlangıçta uygun olan veri dağılımı da değişebilir. Gama et al. (2014), veri akışındaki değişimler ile girdiler ve hedef arasındaki ilişkinin zamanla değişmesini kavram kayması çerçevesinde ele almaktadır. Bununla birlikte her ölçüm değişikliği, modelin yeniden eğitilmesini gerektirmez. Sensör

bağlantısındaki bir sorun ile çalışma koşullarındaki kalıcı değişim farklı nedenlerdir. Bu nedenle hata artışı görüldüğünde önce veri toplama düzeni incelenmeli; model güncelleme kararı, değişimin niteliği ve elde edilebilen doğrulama verileriyle birlikte verilmelidir.

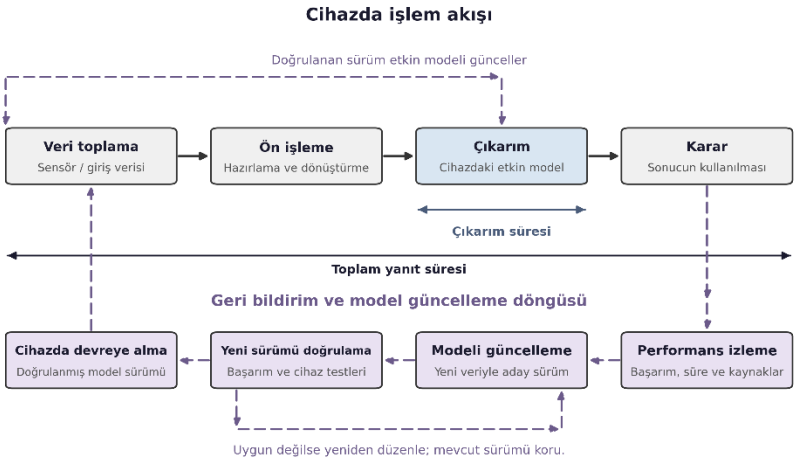
Uzun süreli çalışmada veri kaybı, işlem kuyruğunun büyümesi, yeniden başlatmalar ve kaynak kullanımındaki değişimler izlenmelidir. Kısa bir çıkarım denemesi bu sorunların tamamını ortaya koymayabilir. İnternet bağlantısı kesildiğinde hangi işlevlerin sürdürüleceği de tanımlanmalıdır; modelin cihazda bulunması, ihtiyaç duyulan bütün verilerin ve hizmetlerin yerel olduğu anlamına gelmez. Bağlantı kesintisi, eksik sensör verisi ve yeniden başlama durumları bu nedenle ayrı kullanım senaryoları olarak değerlendirilmelidir.

Bakım sürecinde model, ön işleme ayarları ve çalışma yazılımının sürümleri birlikte kaydedilmelidir. Yeni modelin giriş boyutu veya veri normalleştirme düzeni değişmişse, yalnızca model dosyasının değiştirilmesi yeterli olmayacaktır. Bu bölümde önerilen güncelleme sürecinde, yeni sürüm önce ayrılmış doğrulama verileriyle ve hedef kart üzerinde değerlendirilir. Başarı, gecikme ve bellek kullanımı kabul edilen sınırlar içerisinde kaldığında kullanıma alınır. Önceki çalışan sürümün korunması da güncelleme sonrasında ortaya çıkabilecek sorunların giderilmesini kolaylaştırabilir.

Cihaz üzerinde öğrenme, bu bakım sürecine farklı bir olanak sunmaktadır. Ren et al. (2021), TinyOL yaklaşımında Arduino Nano 33 BLE Sense üzerinde, önceden eğitilmiş yapıya eklenen bir katmanın gelen verilerle güncellenmesini incelemiştir. Bu tür sınırlı uyarlamalar, modelin tamamının yeniden eğitilmesinden farklı bir kaynak gereksinimine sahiptir. Ancak cihazda güncelleme yapılabilmesi, yeni sonuçların kendiliğinden daha doğru olacağını göstermez, kullanılan verilerin uygunluğu ve önceki çalışma

koşullarındaki başarı ayrıca değerlendirilmelidir. Değerlendirme ve güncelleme sürecinin genel şeması Şekil 4'te verilmiştir.

Şekil 4. Değerlendirme ve güncelleme süreci



Kaynak: Zhang et al. (2017), Banbury et al. (2021) ve Ren et al. (2021) temelinde hazırlanmıştır.

Cihaz üzerinde öğrenmenin yaygınlaşması, eğitim sırasında gereken ara değerler ve gradyanlar nedeniyle ek bellek ve hesaplama ihtiyacını gündeme getirmektedir (Lin et al., 2023). Yapılacak uyarlamanın yararı, enerji tüketimi ve bakım yüküyle birlikte değerlendirilmelidir. Bu yükü sınırlamak için genellikle ağın tamamı yerine yalnızca son katmanın veya sınırlı sayıda parametrenin güncellendiği seçici yaklaşımlar tercih edilmektedir, böylece cihaz üzerinde öğrenme, ancak kişiselleştirme kazanımının getirdiği ek maliyeti karşılayabilen uygulamalarda anlamlı bir seçenek hâline gelmektedir.

Gelecekteki çalışmalar açısından modelin hangi kısmının, hangi verilerle ve ne zaman güncelleneceği önemli bir araştırma konusudur. Sonuçların açık deney koşullarıyla raporlanması da geliştirilen yöntemlerin farklı cihaz ve uygulamalara aktarılmasını kolaylaştıracaktır.

Sonuç

Gömülü sistemlerde derin öğrenmenin kullanılması, modelin cihaz üzerinde çalıştırılmasının ötesinde bir tasarım sürecidir. Verinin nasıl toplandığı, hangi işlemlerden geçirildiği ve üretilen kararın nerede kullanıldığı bu sürecin parçalarıdır. Bu bölümde ele alınan yaklaşımlar, model başarısının donanım ve yazılım koşullarıyla birlikte değerlendirilmesi gerektiğini göstermektedir. Geliştirme ortamında yüksek doğruluk veren bir ağ, uygulamanın bellek, zamanlama veya enerji sınırlarını karşılamayabilir.

İncelenen platformlar farklı gereksinimlere karşılık gelmektedir. ESP32 ve mikrodenetleyici tabanlı Arduino kartları küçük modellerle yerel sensör değerlendirmesinde kullanılabilirken, Raspberry Pi daha geniş bir yazılım ortamı sunmaktadır. Jetson platformları ise uygun hesaplamalarda GPU kaynaklarından yararlanılmasına olanak sağlamaktadır. Bu seçenekler arasında bütün uygulamalar için geçerli bir üstünlük sıralaması kurulamaz. Kartın tam modeli, desteklenen işleçler ve diğer görevlerle paylaşılan kaynaklar birlikte incelenmelidir.

Hafif mimariler, nicemleme, budama ve bilgi damıtma, modelin kaynak ihtiyacını farklı yollarla azaltmaktadır. Bellek planlama ise ara sonuçların tutulma biçimini düzenlemektedir. Kullanılacak yöntem, öncelikle uygulamayı sınırlayan kaynak belirlenerek seçilmelidir. Dosyanın küçülmesi, çalışma belleğinin veya çıkarım süresinin aynı oranda azalacağını göstermediğinden, elde edilen kazanımlar hedef cihaz üzerinde ölçülmelidir. Görev başarısındaki değişim de bu değerlendirmenin ayrılmaz bir parçasıdır.

Değerlendirme sürecinde çıkarım süresi ile toplam yanıt süresinin ayrılması ve test verisinin gerçek kullanım koşullarını temsil etmesi önem taşımaktadır. Uzun süreli çalışmada ortaya çıkan veri kaybı, gecikme ve ortam değişimleri izlenmeli; bakım sırasında

model, ön işleme ayarları ve yazılım sürümleri birlikte ele alınmalıdır. Yeni bir modelin kullanıma alınması, doğrulama sonuçlarına ve cihazdaki kaynak ölçümlerine dayanmalıdır.

Bu çerçevede gömülü derin öğrenmenin uygulamadaki değeri, cihazın üstlendiği görevi belirlenen koşullar içinde sürdürebilmesiyle ortaya çıkmaktadır. Model, veri ve donanım arasındaki uyumun geliştirme boyunca korunması, ilk denemeden gerçek kullanıma geçişi desteklemektedir. Cihaz üzerinde öğrenme gibi yeni olanaklar da aynı çerçevede değerlendirilerek, sağlayacakları katkı ile gerektirdikleri kaynaklar arasında uygun bir denge kurulmalıdır.

Kaynakça

Arduino. (t.y.). *Nano 33 BLE Sense Rev2*. 6 Eylül 2026 tarihinde erişildi. https://docs.arduino.cc/hardware/nano-33-ble-sense-rev2/

Banbury, C., Reddi, V. J., Torelli, P., Holleman, J., Jeffries, N., Kiraly, C., Montino, P., Kanter, D., Ahmed, S., Pau, D., Thakker, U., Torrini, A., Warden, P., Cordaro, J., Di Guglielmo, G., Duarte, J., Gibellini, S., Parekh, V., Tran, H., … Xu, X. (2021). *MLPerf Tiny benchmark* [Ön baskı]. arXiv. https://doi.org/10.48550/arXiv.2106.07597

Cai, H., Gan, C., Wang, T., Zhang, Z., & Han, S. (2019). *Once-for-all: Train one network and specialize it for efficient deployment* [Ön baskı]. arXiv. https://doi.org/10.48550/arXiv.1908.09791

David, R., Duke, J., Jain, A., Reddi, V. J., Jeffries, N., Li, J., Kreeger, N., Nappier, I., Natraj, M., Regev, S., Rhodes, R., Wang, T., & Warden, P. (2020). *TensorFlow Lite Micro: Embedded machine learning on TinyML systems* [Ön baskı]. arXiv. https://doi.org/10.48550/arXiv.2010.08678

Espressif Systems. (t.y.). *ESP-DL user guide: Getting started*. 6 Eylül 2026 tarihinde erişildi. https://docs.espressif.com/projects/esp-dl/en/latest/getting_started/readme.html

Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys, 46*(4), Makale 44. https://doi.org/10.1145/2523813

Han, S., Mao, H., & Dally, W. J. (2015). *Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding* [Ön baskı]. arXiv. https://doi.org/10.48550/arXiv.1510.00149

Hinton, G., Vinyals, O., & Dean, J. (2015). *Distilling the knowledge in a neural network* [Ön baskı]. arXiv. https://doi.org/10.48550/arXiv.1503.02531

Jacob, B., Kligys, S., Chen, B., Zhu, M., Tang, M., Howard, A., Adam, H., & Kalenichenko, D. (2018). Quantization and training of neural networks for efficient integer-arithmetic-only inference. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (ss. 2704–2713). IEEE. https://doi.org/10.1109/CVPR.2018.00286

Lai, L., Suda, N., & Chandra, V. (2018). *CMSIS-NN: Efficient neural network kernels for Arm Cortex-M CPUs* [Ön baskı]. arXiv. https://doi.org/10.48550/arXiv.1801.06601

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lin, J., Chen, W.-M., Lin, Y., Cohn, J., Gan, C., & Han, S. (2020). *MCUNet: Tiny deep learning on IoT devices* [Ön baskı]. arXiv. https://doi.org/10.48550/arXiv.2007.10319

Lin, J., Zhu, L., Chen, W.-M., Wang, W.-C., & Han, S. (2023). Tiny machine learning: Progress and futures [Feature]. *IEEE Circuits and Systems Magazine, 23*(3), 8–34. https://doi.org/10.1109/MCAS.2023.3302182

Marwedel, P. (2021). *Embedded system design: Embedded systems foundations of cyber-physical systems, and the Internet of Things* (4. baskı). Springer. https://doi.org/10.1007/978-3-030-60910-8

Nagel, M., Fournarakis, M., Amjad, R. A., Bondarenko, Y., van Baalen, M., & Blankevoort, T. (2021). *A white paper on neural network quantization* [Ön baskı]. arXiv. https://doi.org/10.48550/arXiv.2106.08295

NVIDIA. (t.y.-a). *Jetson FAQ*. 6 Eylül 2026 tarihinde erişildi. https://developer.nvidia.com/embedded/faq

NVIDIA. (t.y.-b). *NVIDIA TensorRT documentation*. 6 Eylül 2026 tarihinde erişildi. https://docs.nvidia.com/deeplearning/tensorrt/latest/index.html

Raspberry Pi Ltd. (t.y.-a). *AI HATs*. 6 Eylül 2026 tarihinde erişildi. https://www.raspberrypi.com/documentation/accessories/ai-hat-plus.html

Raspberry Pi Ltd. (t.y.-b). *Raspberry Pi 5*. 6 Eylül 2026 tarihinde erişildi. https://www.raspberrypi.com/products/raspberry-pi-5/

Ren, H., Anicic, D., & Runkler, T. A. (2021). TinyOL: TinyML with online-learning on microcontrollers. *2021 International Joint Conference on Neural Networks (IJCNN)* (ss. 1–8). IEEE. https://doi.org/10.1109/IJCNN52387.2021.9533927

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (ss. 4510–4520). IEEE. https://doi.org/10.1109/CVPR.2018.00474

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks, 61*, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003

Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. *IEEE Internet of Things Journal, 3*(5), 637–646. https://doi.org/10.1109/JIOT.2016.2579198

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management, 45*(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Sze, V., Chen, Y.-H., Yang, T.-J., & Emer, J. S. (2017). Efficient processing of deep neural networks: A tutorial and survey. *Proceedings of the IEEE, 105*(12), 2295–2329. https://doi.org/10.1109/JPROC.2017.2761740

Zhang, Y., Suda, N., Lai, L., & Chandra, V. (2017). *Hello Edge: Keyword spotting on microcontrollers* [Ön baskı]. arXiv. https://doi.org/10.48550/arXiv.1711.07128


BİDGE