



Görme bozukluğu olan bireyler için derin öğrenme tabanlı ürün sınıflandırma sistemi

Deep learning based product classification for individuals with visual impairment system

Fatma Betül Keskin¹ , Nursena Baygın² , Işıl Karabey Aksakallı^{3,*} , Özlem Sökmen⁴ 

^{1,2,3} Erzurum Teknik Üniversitesi, Bilgisayar Mühendisliği Bölümü, 25040, Erzurum, Türkiye

⁴ Erzurum Teknik Üniversitesi, Endüstri Mühendisliği Bölümü, 25040, Erzurum, Türkiye

Öz

Görme bozukluğuna sahip bireylerin sosyal yaşama adapte olma düzeylerini ve bağımsızlık duygularını artırmaya yönelik yapılan araştırmalar toplumsal katkı açısından oldukça önemlidir. Söz konusu bireylerin alışveriş deneyimlerini kolaylaştırmak yapılan çalışmalar arasındadır. Bu alanda yapılan araştırmalar, alışveriş ürünlerinin sınıflandırılması ve rafların tanıtılması için genellikle geleneksel yöntemlerin kullanıldığını göstermektedir. Önerilen bu çalışmada, market raflarındaki ürünler geleneksel yöntemlerden farklı olarak derin yapay sinir ağları kullanılarak sınıflandırılmıştır. Ayrıca, görme bozukluğuna sahip bireylere kolaylık sağlamak amacıyla çalışmanın ikinci fazında geliştirilecek olan günlük hayatta kullanılan akıllı cihazlara entegre edilebilen ve tüm mobil işletim sistemlerini destekleyen Flutter altyapısı tercih edilecektir. Uygulama Erzurum ilinde bulunan bir marketten elde edilen 2222 görüntü, 14 farklı kategorideki veri setini oluşturularak gerçekleştirilmiştir. Kategori bazlı verinin dengeli sayıda olması için internet kaynağından alınan görüntüler ile birlikte toplam 4585 adet görüntü, YOLOv5, YOLOv8 ve EfficientDet D7TF2 modelleriyle eğitilerek sınıflandırılmıştır. Yapılan deneylerde YOLOv8 modeli, %92,8 doğruluk, %98,6 hassasiyet, %95 duyarlılık ve %96,8 F1 skoru ile diğer yöntemlere oranla daha yüksek performans göstermiştir.

Anahtar kelimeler: Görme bozukluğu, Yapay zekâ tabanlı ürün tanıma, Görüntü işleme, Derin öğrenme

1 Giriş

Görme bozukluğuna sahip bireyler, günlük yaşamlarında çeşitli zorlukla karşılaşmaktadır. Bu zorluklar arasında konum belirleme, nesnelere algılama, para birimlerini tanıma, otobüs numaralarını görme, alışverişte ürün raflarını tanıma gibi basit aktiviteler bulunmaktadır. Bu tür zorlukların yanı sıra, engellerle dolu veya güvensiz kaldırımlar gibi durumlar, görme bozukluğuna sahip bireylerin hareketlilik durumunu olumsuz yönde etkileyebilmektedir. Bu durumların sonucunda, bireylerde

Abstract

Investigations to increase the level of adaptation to social life and the sense of independence of individuals with visual impairment are very important in terms of social contribution. Facilitating the shopping experiences of these individuals can be among these studies. Research in this field shows that traditional methods are generally used to classify shopping products and to introduce the shelves. In this proposed study, products on supermarket shelves are classified using deep artificial neural networks different from traditional methods. In addition, in order to provide convenience to individuals with visual impairments, the Flutter infrastructure, which can be integrated into smart devices used in daily life and supports all mobile operating systems, will be preferred in the second phase of the study. The application was realized by creating a data set of 2222 images obtained from a market in Erzurum province and 14 different categories. In order to have a balanced number of category-based data, a total of 4585 images were trained and classified with YOLOv5, YOLOv8 and EfficientDet D7TF2 models. In the experiments, YOLOv8 model outperformed the other methods with 92.8% accuracy, 98.6% precision, 95% sensitivity and 96.8% F1 score.

Keywords: Visual impairment, Artificial intelligence-based product recognition, Image processing, Deep learning

düşme, çarpma veya geç kalma korkusu gibi psikolojik tepkiler gelişebilmektedir. Ayrıca bağımlı davranışlar sergileme ve toplumda kabul görmede zorlanma gibi durumlarla da karşılaşabilmektedir [1].

Dünya Sağlık Örgütü'ne [2] göre, dünya genelinde yaklaşık 1,3 milyar bireyin görme bozukluğundan etkilendiği tahmin edilmektedir. Bu bireylerin 188,5 milyonu hafif, 217 milyonu orta ila şiddetli ve 36 milyonu ise tamamen görme engelli durumundadır [3]. Medical Express'e [4] göre, nüfusun artması ve yaşlanması ile birlikte, 2050 yılında görme engelli birey sayısının 36

* Sorumlu yazar / Corresponding author, e-posta / e-mail: isil.karabey@erzurum.edu.tr (I. Karabey Aksakallı)

Geliş / Received: 14.10.2023 Kabul / Accepted: 17.0.2024 Yayınlanma / Published: 15.10.2024

doi: 10.28948/ngumuh.1375785

milyondan 115 milyona çıkacağı tahmin edilmektedir. Bu artış, görme engelli bireyler için daha efektif yardımlara yönelik acil ihtiyaçları da beraberinde getirmektedir. Görme bozukluğu olan insanların çoğu, günlük görevleri için bir tür destekleyici teknolojiye ihtiyaç duymaktadır. Destekleyici Teknoloji (DT), herhangi bir yetersizliğe ve/veya bozukluğa sahip bireyler tarafından işlevsel yeteneklerini geliştirmek için kullanılan bir cihaz, ürün, öge veya yazılım programı olarak tanımlanabilmektedir [5].

Görme bozukluğuna sahip bireylerin yaşadığı sorunları azaltmak veya en aza indirmek amacıyla bilim insanları ve firmalar tarafından farklı teknolojik cihaz ve uygulamalar geliştirilmektedir. Bu bağlamda, görme bozukluğuna sahip bireylerin çevrelerindeki nesnelere algılayabilmesi için çeşitli çalışmalar yapılmıştır [6-8]. Sensörler ve lazer ışınlarını kullanan akıllı bastonlar gibi çözümler geliştirilerek görme bozukluğuna sahip bireylere yardımcı olunması amaçlanmıştır [6]. Gbenga ve ark. [7] tarafından görme bozukluğuna sahip bireyler için ultrasonik sensörler ve Arduino üzerinde geliştirilen akıllı bastonlar bu cihazlara örnek olarak sunulabilir. Ayrıca Kaur ve Bhattacharya [8], görme engelli bireylere sesli geri bildirim yoluyla çevre algısı sağlamak amacıyla nesnelere tespiti, sınıflandırılması ve konum tahmini üzerine bir sistem geliştirmişlerdir. Geliştirilen düşük maliyetli, hafif ve kolayca taşınabilir özelliklere sahip olan sistem Odroid kartı, USB kamera ve USB lazer entegrasyonu ile oluşturulmuştur. Kullanıcı geri bildirimlerine dayanarak optimize edilmiş olan uygulamada nesne tespiti, çok modlu özellik füzyonunu kullanan derin öğrenme mimarisi ile gerçekleştirilmiş olup, deneysel sonuçlar çeşitli veri setleri ile doğrulanmıştır.

Microsoft iOS uygulaması ile metni tanımlamak ve seslendirmek için akıllı telefon kamerasını kullanmaktadır [9]. Bu uygulama barkod okuyabilme, ürün bilgilerini sağlayabilme, sahneleri veya nesnelere tanımlayabilme özelliklerine sahiptir. Ayrıca bu uygulama insanları tanıyabilme ve resimlerden metin okuyabilme özelliğine de sahiptir [10]. Benzer şekilde, IPPLEX LookTel Money Reader uygulaması, görme bozukluğuna sahip kullanıcıların farklı para birimlerini tanımasına olanak tanımaktadır [11]. Bu uygulama, iOS cihazının kamerasıyla bir faturanın fotoğrafını çekerek para biriminin değerini sesli olarak bildirmekte ve kullanıcının parasını yardım almadan saymasına olanak tanımaktadır. VizWiz by ROCHCI uygulaması ise kullanıcıların fotoğraf çekmesine ve görüntü hakkında sorular sormasına olanak tanımaktadır [12]. Bu uygulama, soruların gönüllü bireylere e-posta ya da Twitter aracılığıyla gönderebileceği ve kullanıcıların genellikle dakikalar hatta saniyeler içinde yanıtlarını alabileceği bir platform sunmaktadır. Bir başka uygulama olan SayText by Norfello Oy, görüntü içindeki metni taramakta ve yüksek sesle okumaktadır [13]. Bu uygulama metin analizi için Optik Karakter Tanıma (OCR) teknolojisini kullanmakta ve kullanıcıların ekranı kaydırarak belgeyi dinlemelerine olanak tanımaktadır. TetraMail uygulaması ise görme bozukluğuna sahip bireylerin sesli komutları kullanarak e-postaları kontrol etmelerine ve yanıtlamalarına yardımcı olmak için tasarlanmış bir e-posta istemcisidir [14]. Bu uygulama dokunmatik ve ses tabanlı etkileşim sunarak

görme bozukluğuna sahip bireyler için kullanıcı dostu bir uygulama haline gelmiştir. Literatürde ve endüstride yer alan bu cihazlar çevreden gelen bilgileri algılayarak ses veya titreşim gibi geri bildirimler sağlamaktadır.

Tayland Mahidol Üniversitesi tarafından yapılan bir çalışmada [15] görme bozukluğuna sahip bireylerin alışveriş sırasında karşılaştığı yaygın problemleri incelenmiştir. Bu çalışmaya göre, bu bireylerin özellikle ürün fiyatlarına, porsiyon büyüklüklerine, içerik bilgilerine ve besin değerlerine erişimde zorluk yaşadıkları gözlemlenmiştir. Benzer şekilde literatürde yer alan bir başka çalışmada [16], alışveriş merkezlerindeki yoğunlukların ve tatil veya indirim günleri gibi durumlarda artan kuyrukların, insanlardaki stres düzeyini artırdığı gözlemlenmiştir. Bu durumdan yaşam kalitesi düşük olan görme bozukluğuna sahip bireyler daha fazla etkilenmekte ve bekleme süreleri ile orantılı olarak kaygı düzeylerinde de artış durumu ortaya çıkmaktadır. Bu durum ise görme bozukluğuna sahip bireylerin sosyal hayata daha zor entegre olmalarına neden olmaktadır. Normal müşterilerin bağımsız olarak alışveriş yapmaları ve bu süreçte mağaza içi ipuçlarından yararlanmaları beklenmektedir. Ancak ileri görme bozukluğuna sahip bireyler için bu durum olası değildir. Bir mağazada gezinme ve satış görevlilerinin yardımına güvenme çabası bu tüketicilerin yaşadığı normallik derecesini azaltmaktadır [17].

Son zamanlarda yapay zekâ teknolojisindeki gelişmelerle birlikte, görme bozukluğuna sahip bireylere yönelik görsel tanıma yetenekleri önemli ölçüde artış göstermiştir. Ancak bu uygulama ve teknolojilerin günlük hayatta kullanılabilirliği hala beklenen düzeye ulaşamamıştır ve mevcut yazılımlar yaygın olarak kullanılamamaktadır. Örneğin, görüntü tanıma motorları [18], net bir görüntü ve bağlamsal anlayış gerektiren bilgilerin sağlanmasını gerektirmektedir. Görme bozukluğuna sahip bireylerin yaşadığı zorlukları azaltmak ve günlük hayatlarını kolaylaştırmak için teknolojinin daha fazla kullanılması oldukça önemlidir. Geliştirilen projeler ve uygulamalar, görme bozukluğuna sahip bireylerin yaşam kalitesini artırabilir ve onlara bağımsızlık sağlayabilir. Ancak, bu alanda yapılan çalışmalar günlük hayatta ek donanım gerektirdiğinden sıklıkla tercih edilmemektedir.

Önerilen bu çalışmada, mobil cihaz dışında herhangi bir ek donanıma ihtiyaç duyulmadan bireylerin nesnelere algılaması ve alışveriş sırasında bağımsız hareket etmeleri amaçlanmıştır. Araştırma kapsamında, market raflarındaki ürünlerin bir mobil kameradan okutulmuş yapay zekâ algoritmaları ile kullanıcıya sesli geri bildirim ve büyük boyutlu görsel olarak sunulması hedeflenmiştir.

Gerçekleştirilen bu çalışmanın temel katkıları ve motivasyon kaynakları aşağıda maddeler halinde sıralanmıştır.

- Akıllı mobil cihaz aracılığıyla gerçek hayattan veri toplama, ürün-rafları tespit etme.
- Elde edilen görüntülerin, görme bozukluğuna sahip bireylerin günlük yaşam aktivitelerini kolaylaştırmak adına market ürünlerini tanıma konusunda yardımcı olma

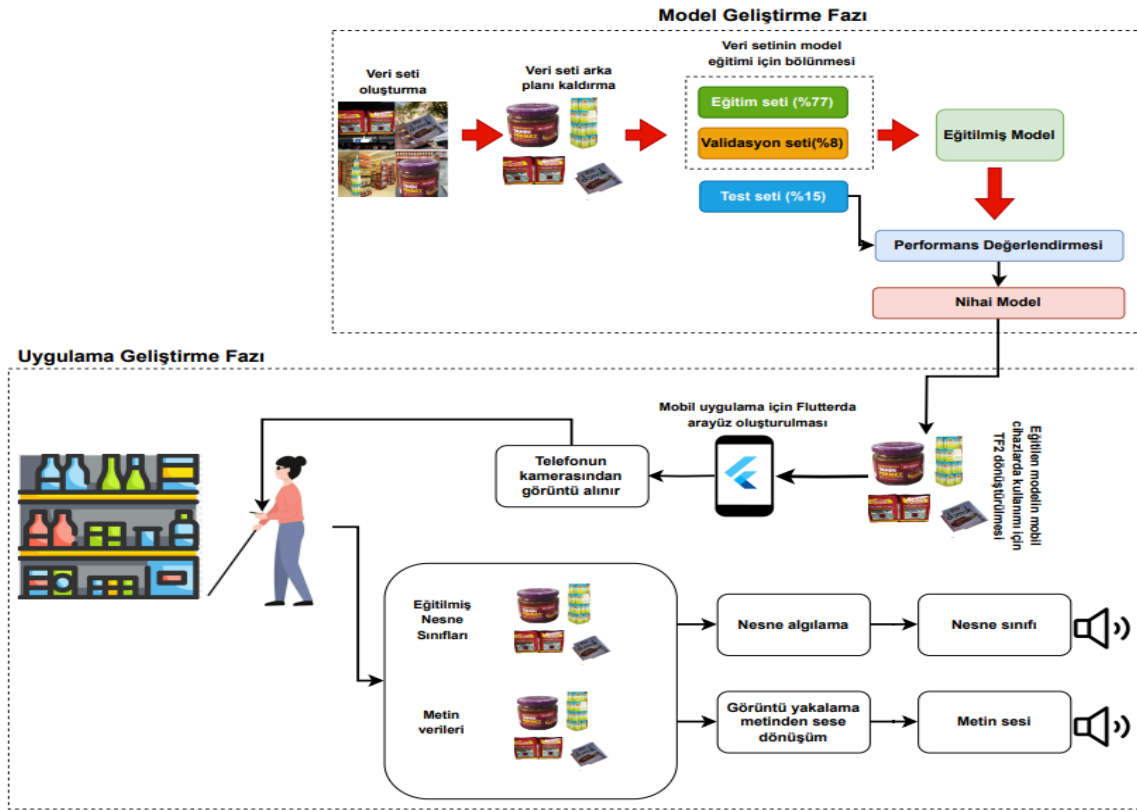
- Yapay zekâ tabanlı nesne tespiti ile görme bozukluğuna sahip bireylerin sosyal hayata bağımsız bir şekilde adapte olmalarını sağlama

Bu çalışmanın sonraki bölümleri şu şekilde düzenlenmiştir: Bölüm 2’de, önerilen çalışmanın yöntemi ve kullanılan materyaller detaylandırılmıştır. Bölüm 3’te, veri setinin oluşturulmasına dair bilgiler sunulmuştur. Bölüm 4’te, önerilen yöntemin deneysel sonuçlarına yer verilmiştir. Bölüm 5’te önerilen yöntemin uygulanma aşamaları incelenmiş ve elde edilen sonuçlar üzerinde değerlendirme yapılmıştır. Son olarak Bölüm 6’da ise çalışmadan elde edilen genel sonuçlar özetlenmiştir.

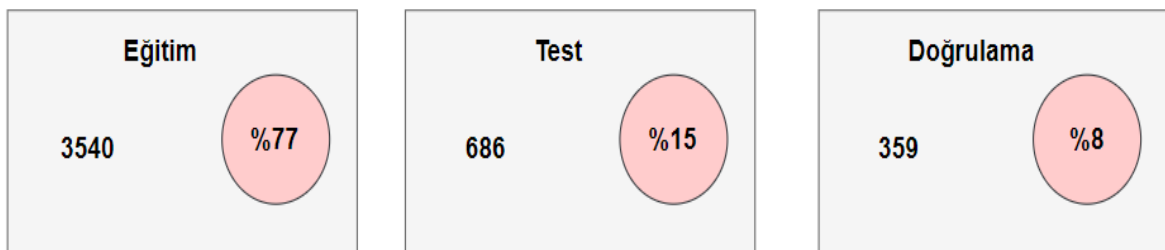
2 Materyal ve metot

Bilgisayarlı görü alanında önemli bir konu olan nesne algılama, hareketli nesne takibi [19], güvenlik kameralarında otomatik silah ve bıçak tespit sistemi [20], su ürünleri

yetiştiriciliği için balık davranışlarının bilgisayarlı görüntü işleme yöntemleriyle izlenmesi [21] gibi birçok uygulama alanında kullanılmaktadır. Bu çalışmada ise market raflarındaki ürünleri algılamak için görüntü işleme yöntemleri ile derin öğrenme tabanlı sinir ağıları kullanılmıştır. İlk olarak marketlerden gerçek görseller elde edilerek veri toplama süreci gerçekleştirilmiştir. Ardından, eğitim aşaması için bu veriler kullanılarak bir yapay zekâ modeli oluşturulmuştur. Model, market raflarındaki ürünleri algılayıp tanımak için eğitilmiştir. Önerilen yöntem, geliştirilen bir mobil uygulama aracılığıyla görme bozukluğuna sahip kullanıcıların kameralarını ürünlere doğrultarak ürünleri tanıyabilmelerine ve ürün detaylarına sesli olarak erişebilmelerine olanak tanımaktadır. Araştırma kapsamında önerilen yöntemi özetleyen bir blok diyagram Şekil 1’de sunulmaktadır.



Şekil 1. Önerilen yöntemin akış diyagramı



Şekil 2. Elde edilen veri setinin eğitim, test ve doğrulama kümelerine ayrılma oranları

Şekil 1’de görüldüğü üzere önerilen yöntem model geliştirme fazı ve uygulama geliştirme fazı olmak üzere iki aşamadan meydana gelmektedir. Model geliştirme fazında veri seti oluşturma, eğitim-test işlemleri ve performans değerlendirme süreçleri yer almaktadır. Uygulama geliştirme fazında ise geliştirilen mobil uygulama üzerinden nesne sınıflandırılması yapılacaktır.

2.1 Eğitim modelleri

Araştırmanın eğitim aşaması için “YOLOv5”, “YOLOv8” ve “EfficientDet D7 Tensorflow 2” olmak üzere üç farklı model kullanılmıştır. Bu modeller alt bölümlerde detaylı bir şekilde sunulmuştur.

2.1.1 YOLOv5

YOLOv5 [22] nesne algılama ve sınıflandırma için kullanılan bir derin öğrenme modelidir. Görsel verilerde nesnelere belirlemek için hızlı ve etkili bir yaklaşım sunmaktadır. YOLOv5 PyTorch çerçevesi üzerine kurulmuş bir mimariye sahip olup, GPU hızlandırmasını kullanarak gerçek zamanlı nesne algılama yetenekleri sunmaktadır. YOLOv5 geliştiricilere kolaylık sağlamak için önceden eğitilmiş ağırlıklar ve örnek veri kümelerini birlikte sunmaktadır. Ayrıca bu algoritma ile kişiselleştirilmiş veri kümeleri kullanılarak da özelleştirilmiş nesne tespiti modelleri eğitebilmektedir. YOLOv5, farklı boyutlarda (tiny, small, medium, large) modeller sunmakta ve çeşitli veri seti boyutlarına uygulanabilmektedir.

2.1.2 YOLOv8

YOLOv8 [23], YOLO serisinin bir versiyonudur ve önceki sürümlerin gelişmiş versiyonudur. Bu algoritma, önceden eğitilmiş evrişimli bir sinir ağı modeli kullanmakta ve görüntüyü farklı ölçeklerde analiz etmektedir. Ardından her ölçekteki tespitleri birleştirmekte ve sonuç olarak nesnelere konumlarını ve sınıflarını vermektedir. YOLOv8, yüksek doğruluk ve hızlı işleme performansı sağlayarak doğru sonuçlar elde etmek için farklı ölçeklerdeki özellik haritalarını birleştirmektedir. Önerilen bu çalışmada, “YOLOv8” modellerinden “YOLOv8s” modeli kullanılmıştır. Bu modelin parametreleri Tablo 1’de karşılaştırmalı olarak sunulmaktadır.

2.1.3 EfficientDet (D7) Tensorflow 2

EfficientDet [25] mimarisi Google Brain Team tarafından yazılmış olup sınıflandırma için kullanılan ImageNet görüntü veri tabanında önceden eğitilmiş evrişimli bir sinir ağı olan EfficientNet tabanlıdır. EfficientDet, belirli ayrıntı düzeylerinde görüntünün bölümlerini havuzlayarak ve karıştırarak, NAS-FPN özellikli bir füzyon katmanından geçirilmiş özellikleri oluşturmaktadır. NAS-FPN çeşitli özellikleri, değişen ayrıntı düzeylerinde birleştirmekte ve daha sonra bu özellikleri sınırlayıcı kutuların ve sınıf etiketlerinin tahmin edildiği algılama bölümüne iletmektedir. EfficientDet-D7 [26] modeli, EfficientNet ana yapısıyla birlikte, önceden

eğitilmiş ImageNet veri seti üzerinde eğitilmiş olup birçok günlük nesneyi algılayabilmektedir.

3 Veri toplama

Bu çalışma için oluşturulan veri seti, Erzurum ilinde yer alan Düzgün Marketler zincirinin mağaza raflarında bulunan ürünlerin fotoğraflarından meydana gelmektedir. Bu fotoğraflar mağaza rafında bulunan ürünlerin buldukları yerlerde, daha sonra ise beyaz bir karton üzerinde çekilerek elde edilmiştir. Fotoğraf çekimi Xiaomi Mi 10T Pro ve Oppo A72 cihazları ile gerçekleştirilmiştir. Daha sonra fotoğraflar roboflow ortamına jpeg formatında yüklenmiş ve etiketlemeler yapılmıştır. Çalışma öncesinde marketler zinciri yöneticisinden yazılı izin alınmıştır. Bu veri setinde toplam 14 farklı kategoride 2222 görüntü yer almaktadır. Her kategoriden verinin nispeten dengeli olması için veri seti internet üzerinden Google lens kullanılarak örnek bir resim verilip benzer resimler çekilerek genişletilmiştir. Bu işlem sonucu toplamda 4585 görüntü elde edilmiş ve deneysel çalışmalar bu görüntüler üzerinde gerçekleştirilmiştir. Şekil 2’de görüldüğü üzere verilerin %77’si eğitim, %15’i test ve %8’i doğrulama için kullanılmıştır.

Tablo 1. YOLOv8 modellerinin karşılaştırılması [24]

Model	Boyut	mAP	Hız (ms)	Param (M)	Düşme (B)
YOLOv8n	640	37.3	80.4	3.2	8.7
YOLOv8s	640	44.9	128.4	11.2	28.6
YOLOv8m	640	50.2	234.7	25.9	78.9
YOLOv8l	640	52.9	375.2	43.7	165.2
YOLOv8x	640	53.9	479.1	68.2	257.8

Tablo 2’de veri setindeki verilerin genişletilme öncesindeki sayı ve kategorik durumu verilmiştir. Ürün fotoğrafları market sahibinin izni ile çekilmiş daha sonra arka planlar temizlenerek önışleme yapılmıştır. Bazı görüntülerde ise arka plan temizlenmeden market görünümü olarak kullanılmıştır.

Tablo 2. Oluşturulan veri küme kategorileri ve adetleri

Kategori	Adet
Bal	123
Çay	104
Çikolata	245
Helva	265
Makarna	158
Peynir	185
Poşet	315
Reçel	112
Su	128
Süt	120
Şehriye	157
Tahin	103
Yağ	126
Zeytin	111

Eğitim Veri Seti



Şekil 3. Eğitim verilerinin bir bölümü

Test Veri Seti



Şekil 4. Test verilerinin bir bölümü

Validasyon Veri Seti



Şekil 5. Doğrulama verilerinin bir bölümü

Veri setini çoğaltma işlemlerinden sonra eğitimde kullanılan verilerin bir kısmı Şekil 3'te sunulmuştur. Veri setinin %77'lik bir bölümü eğitim modeline ayrılmıştır. Toplanan veri setinin genişletilmesi işlemlerinden sonra test ve doğrulama için kullanılan verilerin bir bölümü sırasıyla Şekil 4 ve Şekil 5'te sunulmuştur. Test aşaması için toplam veri setinin %15'i, doğrulama için ise veri setinin %8'i kullanılmıştır.

4 Deneysel Sonuçlar

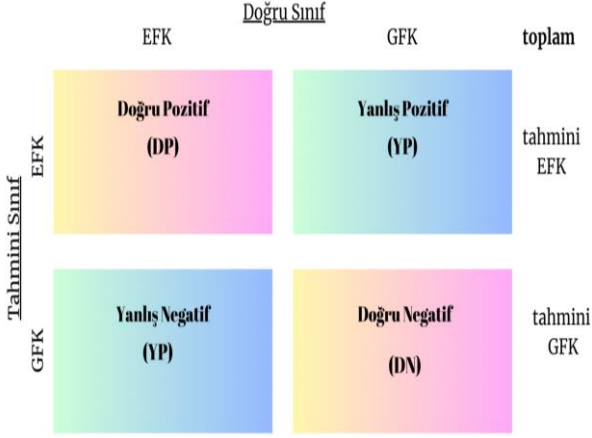
Toplanan veri setinde çoğaltma sonrası elde edilen toplam 4585 görüntünün %77'si eğitim, %15'i test ve kalan %8'i ise doğrulama için kullanılmıştır. Veri setinin eğitim, test ve doğrulama kümelerine ayrılma sürecinde Roboflow kullanılmıştır. Roboflow uygulamasında veri seti

bölünürken her kategoriden dengeli sayıda veri otomatik olarak alınmaktadır. Elde edilen hassasiyet-güven eğrileri incelendiğinde kategori bazlı sınıflandırma doğruluğunun artışı veri setinin ortalama olarak eşit bir şekilde dağıtıldığını göstermektedir. Elde edilen hassasiyet-güven eğrileri incelendiğinde kategori bazlı sınıflandırma doğruluğunun arttığı görülmekte; bu da veri setinin dağılımının ortalama olarak eşit olduğunu göstermektedir. Kullanılan derin sinir ağlarının performanslarını karşılaştırmak amacıyla doğruluk, hassasiyet, duyarlılık ve F1 skor değerleri aşağıda alt başlıklar halinde açıklanmıştır.

4.1 Karışıklık matrisi

Karışıklık matrisi, sınıflandırma modellerinin performansını değerlendirmek için oluşturulmaktadır. Bir

sınıflandırma modeli, gerçek değerler ile tahmin edilen değerler arasındaki ilişkiyi görselleştirmek için kullanılan tablodur. Karışıklık matrisinin hesaplanmasında yer alan metrikler Şekil 6'da gösterilmiştir. Karışıklık matrisi DP, DN, YP ve YN olmak üzere dört temel terimi içermektedir.



Şekil 6. Performans değerlendirme metrikleri

Karışıklık matrisi; modelin doğruluk, hassasiyet, özgünlük ve F1 skoru gibi diğer performans metriklerini hesaplamak için kullanılmaktadır. Şekil 6'da verilen metrikler kullanılarak karışıklık matrisi hesaplanmış ve Şekil 7 ile sunulan sınıflandırma sonuçları her kategori için ayrı ayrı gösterilmiştir. Değerlendirme metriklerinin yüzdeleri ise Tablo 3'te verilmiştir.

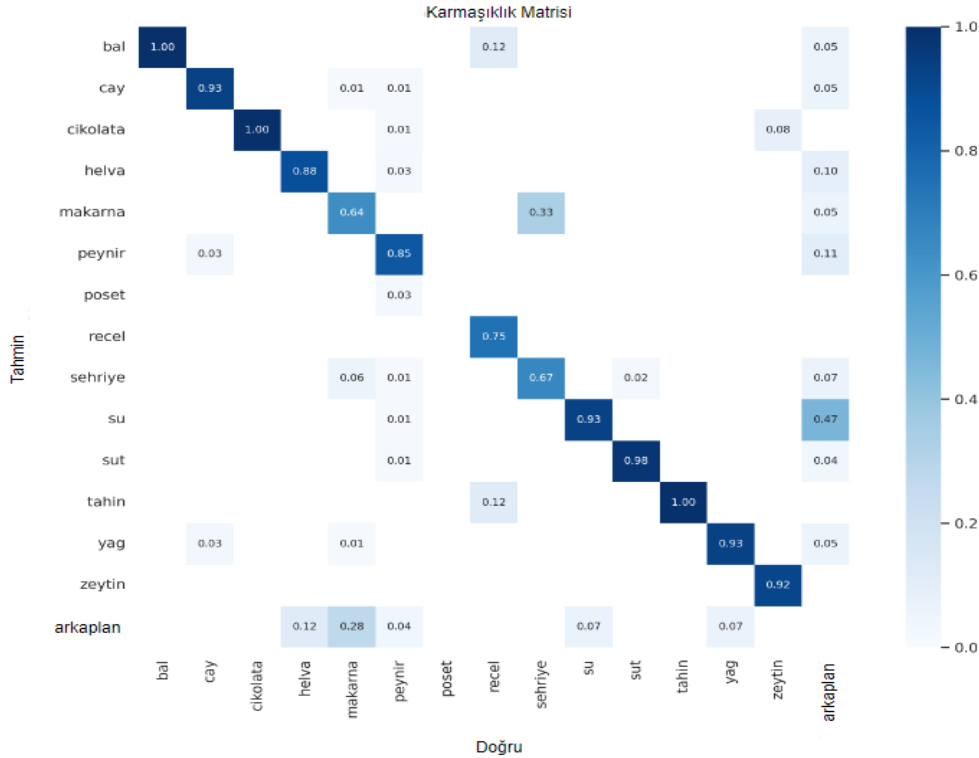
Tablo 3. Performans metriklerinin modellere göre karşılaştırılması

Metrik	YOLOv5	EfficientDet D7TF2	YOLOv8
Hassasiyet	%91.7	%45.2	%98.6
Duyarlılık	%97	%68.3	%95
Doğruluk	%78.3	%55.9	%92.8
F1 Skor	%73	%54.5	%96.79
mAP	%69	%87	%98
Epoch sayısı	250	250	250
Batch sayısı	16	16	16

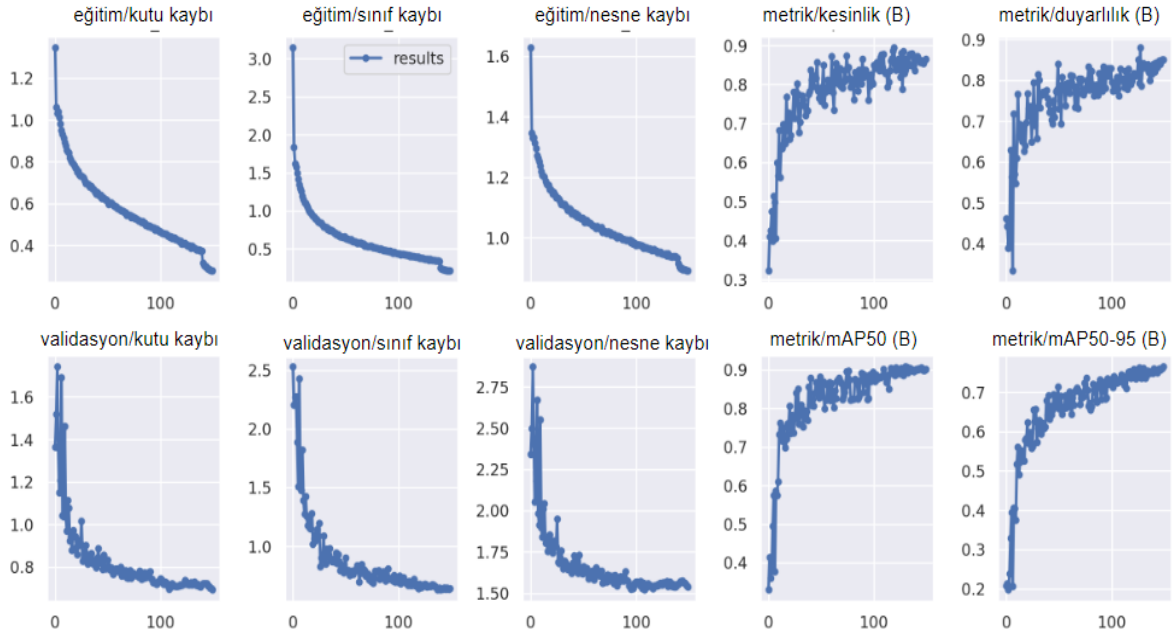
Yapılan çalışmada, Tablo 4'te görüldüğü üzere adım (epoch) sayısının artmasına rağmen, en iyi sonuçlara 250 adım sayısında ulaşıldığı tespit edilmiştir. Ayrıca pencere sayısı 8, 16, 32 ve 64 olarak tüm adımlarda çalıştırılmış olup batch sayıları içerisinde pencere sayısı artsa da en iyi sonucu veren YOLOv8 algoritmasında 16 batch boyutundan sonra iyileştirici değişimler elde edilmediği için pencere sayısı olarak 16'ya karar verilmiştir. Diğer yöntemlerde ise 16 batch ve 250 adımdan sonraki 32 ve 64 batch değerlerinde düşüş gözlenmiştir.

Tablo 4. Farklı epoch ve batch değerlerine göre eğitim modellerinin doğruluk oranları

	Epoch/Batch	8	16	32	64
YOLOv5	50	%56.2	%65.7	%68.1	%60.7
	150	%63.8	%73.2	%73.9	%66.3
	250	%68.5	%78.3	%74.8	%72.4
Efficient Det (D7)	50	%31.2	%48	%49	%37
	150	%39	%52	%51	%40.7
	250	%37	%55.9	%54	%43
YOLOv8	50	%72.9	%74.7	%78.7	%80
	150	%82	%84.7	%82.1	%80
	250	%84	%85	%85	%85



Şekil 7. YOLOv8 model eğitimi-örnek bir karışıklık matrisi



Şekil 8. mAP değerleri görüntüsü

Tablo 3'te eğitim modellerinin hassasiyet, doğruluk, duyarlılık, F1 skor değeri, map değerleri, epoch ve batch sayıları karşılaştırılmıştır. Tablodaki sonuçlara göre en iyi sonuç YOLOv8 ile elde edildiği için eğitim modeli olarak v8 seçilmiştir. YOLOv8 modelinden elde edilen ortalama hata performansı; kesinlik ve duyarlılık metriklerine göre incelenmiş ve Şekil 8'de sunulmuştur.

4.2 Doğruluk (Accuracy)

Doğruluk, bir sınıflandırma modelinin doğru tahminlerinin toplam verilere oranını ifade eden bir performans metriğidir. Doğruluk, modelin genel tahmin yeteneğini değerlendirmek için kullanılmaktadır. Doğruluk metriği, Denklem 1'deki gibi hesaplanmaktadır.

$$\text{Doğruluk} = (\text{Doğru Tahminler}) / (\text{Toplam Veri Sayısı})$$

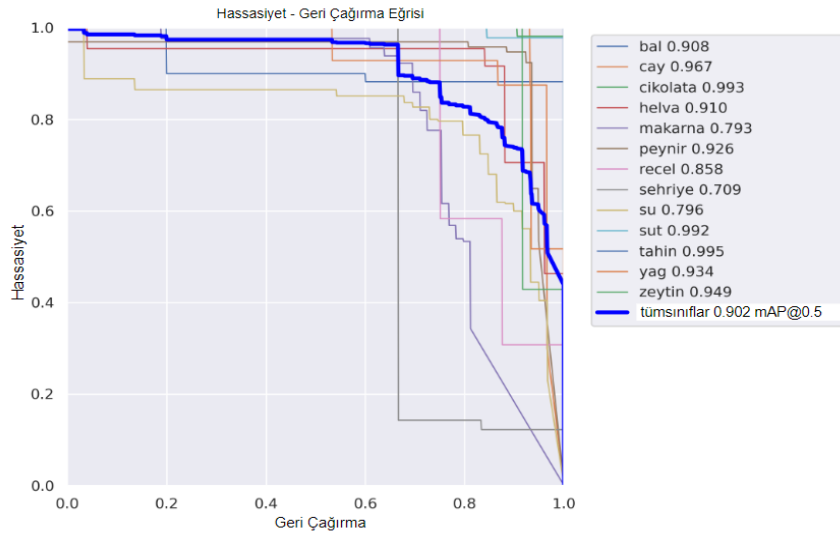
$$\text{Doğruluk} = \frac{DP + DN}{DP + DN + YP + YN} \quad (1)$$

4.3 Hassasiyet (Precision)

Hassasiyet, bir sınıflandırma modelinin pozitif olarak tahmin ettiği örneklerin, gerçek pozitiflere oranını ifade eden bir performans metriğidir. Hassasiyet, yanlış pozitifliği minimize etmeyi hedefler. Hassasiyet, Denklem 2'deki gibi hesaplanmaktadır:

$$\text{Hassasiyet} = (\text{Gerçek Pozitifler}) / (\text{Gerçek Pozitifler} + \text{Yanlış Pozitifler})$$

$$\text{Hassasiyet} = \frac{DP}{DP + YP} \quad (2)$$



Şekil 9. YOLOv8 eğitim modelinden elde edilen hassasiyet-güven eğrileri (Kategoriler bazında)

Önerilen bu çalışmada eğitim modelinden elde edilen hassasiyet ve geri çağırma eğrisi Şekil 9'da verilmektedir. Verilen hassasiyet-geri çağırma eğrisi altındaki alan (AUC), modelin performansını değerlendirmek için kullanılan bir ölçüttür. Bu alan ne kadar büyükse, modelin hassasiyet ve geri çağırma dengesine göre performansı o kadar iyi kabul edilmektedir. Burada bahsedilen eğri altında kalan alanın genişliği, modelimizin doğruluğunun yüksek olduğunu göstermektedir.

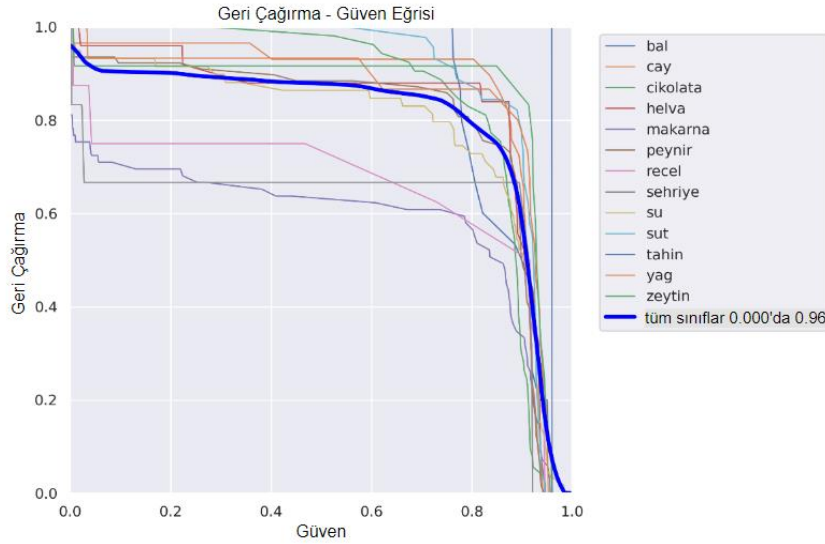
4.4 Duyarlılık (Recall)

Duyarlılık, bir sınıflandırma modelinin gerçek pozitiflerin ne kadarını doğru bir şekilde tespit ettiğini ifade eden bir performans metriğidir. Duyarlılık ile yanlış negatifleri minimize etme hedeflemektedir. Duyarlılık metriği, Denklem 3'teki gibi hesaplanmaktadır:

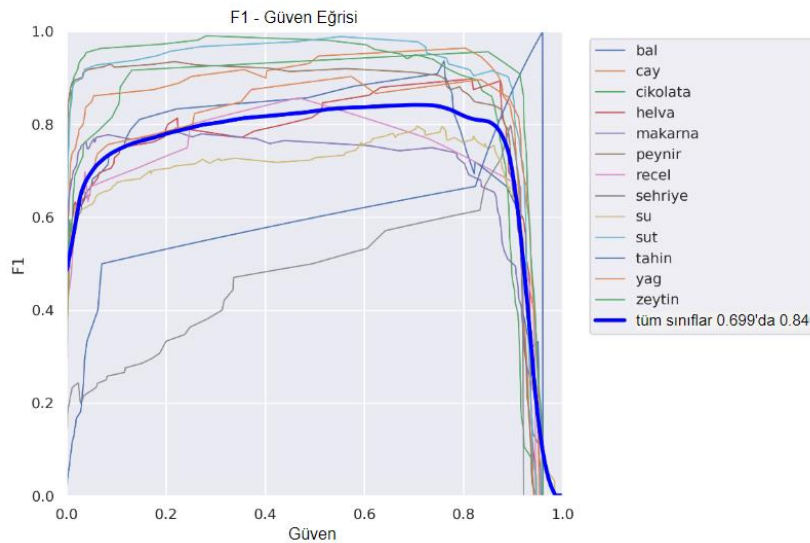
$$\text{Duyarlılık} = (\text{Doğru Pozitifler}) / (\text{Doğru Pozitifler} + \text{Yanlış Negatifler})$$

$$\text{Duyarlılık} = \frac{DP}{DP + YN} \quad (3)$$

Önerilen çalışmada eğitim modelinden elde edilen duyarlılık ve güven eğrisi görüntüsü Şekil 10'da verilmektedir. Şekil 10'da elde edilen grafikte yüksek duyarlılıkla yanlış pozitiflerin en aza indirildiği, yüksek hassasiyetle de doğru tahminlerin ne kadar yüksek olduğunu göstermektedir. İki taraftan da eğrinin 1 değerine yaklaşması ve eğri altında kalan alan modelin doğruluğunun yüksek olduğunu göstermektedir.



Şekil 10. YOLOv8 eğitim modelinden elde edilen duyarlılık ve güven eğrisi



Şekil 11. YOLOv8 eğitim modelinden elde edilen F1 skor ve güven eğrisi

4.5 F1 Skor

F1 skoru bir sınıflandırma modelinin performansını ölçmek için kullanılan bir metrik olup, hassasiyet ile duyarlılık değerlerinin birleşimini temsil etmektedir. F1 skoru, [Denklem 4](#)' teki gibi hesaplanmaktadır.

F1 Skor = $2 * (\text{Hassasiyet} * \text{Duyarlılık}) / (\text{Hassasiyet} + \text{Duyarlılık})$

$$F1 \text{ Skor} = \frac{2DP}{2DP + YP + YN} \quad (4)$$

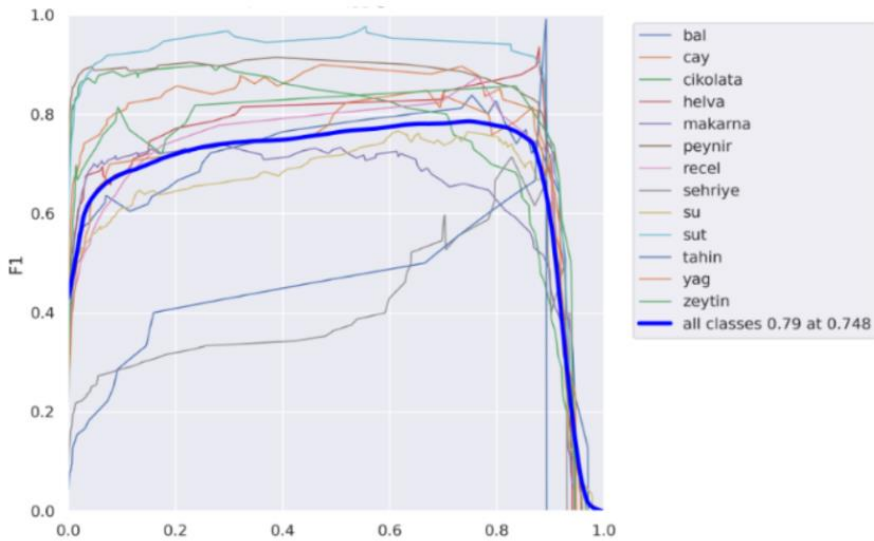
Bu çalışmada eğitim fazında elde edilen F1 skor ve güven eğrisi [Şekil 11](#)'de verilmektedir. Elde edilen grafik incelendiğinde, modelin performansının harmonik ortalamaya (hassasiyet ve geri çağırma değerlerinin ekstrem durumlarını da içerecek şekilde hesaplanan bir ortalama türü) dayalı olarak değerlendirilmesi, seçilen modelin mutlak anlamda en iyi model olmamakla birlikte, incelenen diğer modellere kıyasla tercih edilebilir bir seçenek olduğunu göstermektedir. Mevcut analizler çerçevesinde YOLOv8 modeli, karşılaştırıldığı diğer modellerden daha yüksek oranda metrikler elde etmiştir.

5 Tartışmalar

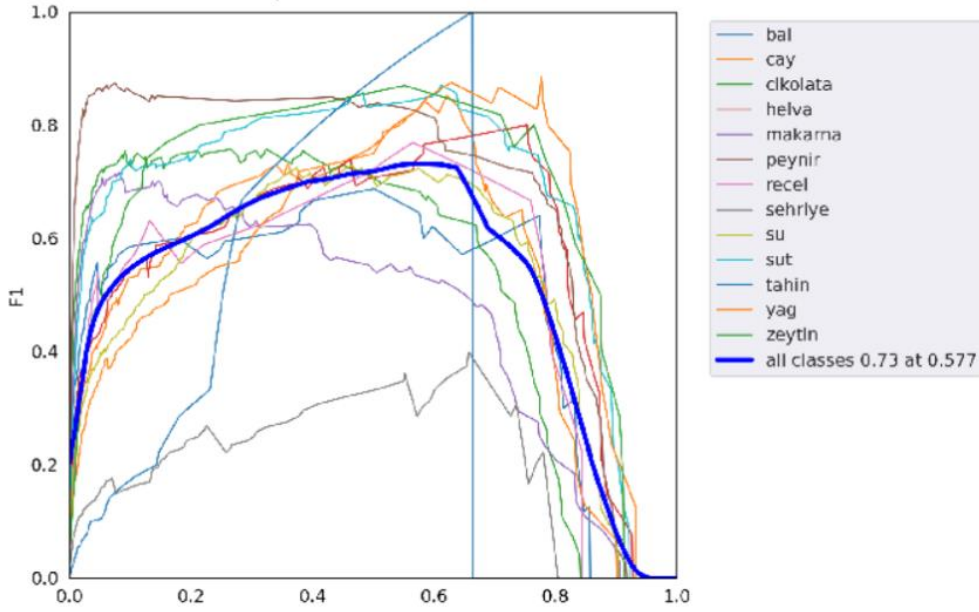
Bu çalışmada, market raflarındaki ürünlerin tanımlanması için YOLOv8, YOLOv5 ve EfficientDet D7TF2 nesne tespiti modelleri kullanılmıştır. İlk olarak EfficientDet D7 modeli üzerinde çalışılmış ve [Şekil 12](#)'de görüldüğü üzere F1 skoru ortalama olarak %79 civarında bulunmuştur. Güven düzeyi ise genellikle F1 skoruna yakın bir seviyede seyretmiştir. Ancak, belirli testlerde modelin istenilen performansı göstermediği ve güven düzeyinin

düşük olduğu gözlenmiştir. Bu durum, modelin belirli koşullarda güvenilir olmayabileceğini ve performansının değişkenlik gösterebileceğini işaret etmektedir. Elde edilen sonuçlar, EfficientDet D7 modelinin genel performansını ölçmek için bir referans noktası sağlamıştır. Ancak, modelin belirli durumlarda güvenilir olmadığı ve daha fazla iyileştirme gerektirebileceği göz önünde bulundurularak YOLOv5 modeli üzerinde çalışılmıştır. [Şekil 13](#)'te kategorize edilmiş ürünlerin çoğunun ortalamasının üzerinde performans gösterdiği, ancak bir kısmının ise ortalamasının altında kaldığı görülmektedir. Bu durum, modelin bazı alanlarda yeterli performansı sergileyemediğini ve geliştirilmesi gerektiğini işaret etmektedir. Modelin daha iyi sonuçlar elde etmesi için, farklı adım ve pencere sayıları gibi değişkenlerle tekrar çalıştırılması veya geliştirilmesi gerekmektedir. Ancak modelin denemesi sırasında adımların 50'nin katları üzerinden değiştirilmesiyle elde edilen en yüksek F1 skoru da bu tabloda görülmüştür.

YOLOv5 modelinin F1 skoru ortalama olarak %73 civarındadır ve güven düzeyi genellikle F1 skoruna yakın bir seviyededir. Ancak, belirli testlerde modelin beklenen performansı göstermediği ve güven düzeyinin düşük olduğu gözlenmiştir. Bu sonuçlar, modelin güvenilirliği ve performansı üzerinde daha fazla çalışma ve iyileştirme gerektirebileceğini vurgulamaktadır. Bu nedenle, modelin belirli durumlarda güvenilir olmadığını ve daha fazla geliştirme işlemi gerektiğini göstermektedir. [Tablo 3](#)'te elde edilen sonuçlar YOLOv5 modelinin genel performansını değerlendirmek için önemli bir referans noktası sağlamaktadır, ancak daha fazla çalışma ve iyileştirme gerektirebileceğini göstermektedir. Model performans metrikleri yeterli sonucu vermediğinden EfficientDet D7 ve YOLOv5 modelleri sonrasında YOLOv8 modeli kullanılmıştır.



Şekil 12. EfficientDet D7 eğitim modelinden elde edilen F1 skor ve güven eğrisi



Şekil 13. YOLOv5 eğitim modelinden elde edilen F1 skor ve güven eğrisi

Önerilen bu çalışma doğrultusunda deneysel sonuçlar bölümünde belirtildiği gibi en yüksek oranların YOLOv8 modelinde elde edildiği gözlemlenmiştir.

6 Sonuçlar

Bu çalışmada görme bozukluğuna sahip bireylerin hayatlarını kolaylaştırmak amacıyla market raflarında bulunan ürünlerin sınıflandırılmasında derin sinir ağları kullanılmıştır. Çalışma kapsamında YOLOv8, YOLOv5 ve EfficientDet D7 modelleri kullanılmış ve market veri seti üzerinde eğitilmiştir. Eğitim sürecinde önceden toplanan bir veri seti kullanılmış ve eğitim aşamasında elde edilen ağırlıklar kaydedilmiştir. Ardından test veri seti üzerinde performans değerlendirmesi yapılmış ve sınıflandırma sonuçları analiz edilmiştir. Performans metriği olarak hassasiyet, duyarlılık, F1 skor, mAp değerleri ölçülmüştür. Model eğitimi için 250 epoch sayısı ve 16 batch değeri kullanılmıştır. Yapılan deneyler sonucunda 8, 16, 32, 64 batch değerleri ve 50, 150, 250 adım değerleri karşılaştırıldığında modellerin 250 adımda en iyi performansı verdiği gözlemlenmiştir. Her üç modelin hassasiyet, duyarlılık, F1 skor ve mAp metrikleri karşılaştırıldığında EfficientDet D7TF2 modelinin diğer modellere göre en az %22.4 daha az doğruluk oranına sahip olduğu görülmüştür. Tüm metriklerde en yüksek performansı YOLOv8 modeli göstermiştir. Bu model üzerinde yapılan test sonucunda; doğruluk, hassasiyet, duyarlılık ve F1 skor oranları sırasıyla %92.75, %98.6, %95, %96.79 olarak hesaplanmıştır. Performansı en yüksek olan model olarak belirlendiği için mobil uygulama üzerinde yapılan eğitim için kullanılmak üzere YOLOv8 modeli TFLite'a çevrilmiştir.

Gelecek çalışmalarda ikinci faz olarak adlandırdığımız uygulama geliştirme fazının gerçekleştirilmesi hedeflenmektedir. Ek olarak, veri setinin kategori bazlı genişletilmesi planlanmaktadır. Bu veri seti üzerinde

modelin eğitilerek geliştirilen mobil uygulama üzerinden görme bozukluğu olan bireyin anlık olarak cep telefonu kamerasını kullanması ile nesne sınıflandırılmasının sesli ve görsel olarak yapılması amaçlanmaktadır.

Teşekkür

Bu çalışma, Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK) tarafından sağlanan 1139B412202533 nolu proje fonu tarafından desteklenmektedir.

Çıkar çatışması

Yazarlarda herhangi bir çıkar çatışması bulunmamaktadır.

Benzerlik oranı (iThenticate): %9

Kaynaklar

- [1] İ. Çevik, H. Çakmak, Ö. Çelik ve P. Okyay, Yaşam boyu göz sağlığı: "2020 vizyonu: görme hakkı". ESTÜDAM Halk Sağlığı Dergisi, 6 (3), 310-321, 2021. <https://doi.org/10.35232/estudamhsd.891156>.
- [2] C. M. Elzean and E. M. Sakr, Proposed three-dimensional designs for the color wheel to help blind persons understand matching colors of their clothes. International Design Journal, 11 (2), 417-423, 2021. <https://dx.doi.org/10.21608/ijj.2021.153624>.
- [3] S. R. Flaxman et al., Global causes of blindness and distance vision impairment 1990–2020: a systematic review and meta-analysis. The Lancet Global Health, 5 (12), e1221-e1234, 2017. [https://doi.org/10.1016/S2214-109X\(17\)30393-5](https://doi.org/10.1016/S2214-109X(17)30393-5).
- [4] MedicalExpress, World's blind population to soar: study. <https://medicalxpress.com/news/2017-08-world-population-soar.html>, Accessed 10 Feb 2024.
- [5] B. S. Lin, C. C. Lee and P. Y. Chiang, Simple smartphone-based guiding system for visually

- impaired people. *Sensors*, 17 (6), 1371, 2017. <https://doi.org/10.3390/s17061371>.
- [6] M. Çakır, A. Çelik, İ. Özyalçın ve A. Uzun, Engelli insanlar için akıllı baston ve akıllı şapka tasarımı. 4th International Vocational Schools Symposium, pp. 1445-1454, Yalova, Turkey, 21-23 May 2015.
- [7] D. E. Gbenga, A. I. Shani and A. L. Adekunle, Smart walking stick for visually impaired people using ultrasonic sensors and Arduino. *International Journal of Engineering and Technology*, 9 (5), 3435-3447, 2017. <https://dx.doi.org/10.21817/ijet/2017/v9i5/170905302>.
- [8] B. Kaur and J. Bhattacharya, Scene perception system for visually impaired based on object detection and classification using multimodal deep convolutional neural network. *Journal of Electronic Imaging*, 28 (1), 013031-013031, 2019. <https://doi.org/10.48550/arXiv.1805.08798>.
- [9] C. Granquist, S. Y. Sun, S. R. Montezuma, T. M. Tran, R. Gage and G. E. Legge, Evaluation and comparison of artificial intelligence vision aids: Orcam myeye 1 and seeing ai. *Journal of Visual Impairment & Blindness*, 115 (4), 277-285, 2021. <https://doi.org/10.1177/0145482X211027492>.
- [10] K. Jolly, *Hands-On Data Visualization with Bokeh: Interactive Web Plotting for Python Using Bokeh*. Packt Publishing Ltd., 2018.
- [11] J. Sudol, O. Dialameh, C. Blanchard, T. Dorcey, LookTel—A comprehensive platform for computer-aided visual assistance. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, pp. 73-80, San Francisco, CA, USA, 2010.
- [12] J. Bigham, VizWiz. Rochester Human Computer Interaction (ROC HCI). <http://itunes.apple.com/us/app/vizwiz/id439686043?mt=8>, Accessed 5 May 2024.
- [13] E. Biknevicius, Say Text Apps for blind and visually impaired people, <http://etaling.com/en/say-text-apps-for-blind-and-visually-impaired-people/>, Accessed 6 May 2024.
- [14] A. Khan, S. Khusro, B. Niazi, J. Ahmad, I. Alam and I. Khan, TetraMail: a usable email client for blind people, *Universal Access in the Information Society*, 19, 113-132, 2020. <https://doi.org/10.1007/s10209-018-0633-5>.
- [15] A. Hoonlor, S. P. N. Ayudhya, S. Harnmetta, S. Kitpanon and K. Khlprasit, UCap: A crowdsourcing application for the visually impaired and blind persons on Android smartphone. 2015 International Computer Science and Engineering Conference (ICSEC), pp. 1-6, Chiang Mai, Thailand, 2015.
- [16] D. Shukla and M. Shah, Smart trolley shopping for automatic billing & assistance for visually impaired, *J. Eng. Sci.*, 14 (04), 2023.
- [17] C. Kaufman-Scarborough and T. L. Childers, Understanding markets as online public places: Insights from consumers with visual impairments. *Journal of Public Policy & Marketing*, 28 (1), 16-28, 2009. <https://doi.org/10.1509/jppm.28.1.16>.
- [18] M. George and C. Floerkemeier, Recognizing products: A per-exemplar multi-label image classification approach. *Computer Vision—ECCV 2014*, 13th European Conference, pp. 440–455, Zurich, Switzerland, 2014.
- [19] S. Öncü, Bilgisayarlı görü ve ses algılama tekniği ile hareketli silah ve bıçak tespit sistemi: karşılaştırmalı yolo modelleri. *Avrupa Bilim ve Teknoloji Dergisi*, (41), 16-22, 2022. <https://doi.org/10.31590/ejosat.1163675>.
- [20] M. T. Ağdaş ve S. Gülşen, Güvenlik kameralarında otomatik silah ve bıçak tespit sistemi: karşılaştırmalı yolo modelleri. *Avrupa Bilim ve Teknoloji Dergisi*, (41), 16-22, 2022. <https://doi.org/10.31590/ejosat.1163675>.
- [21] G. Tonguç, B. A. Balcı ve M. N. Arslan, Su ürünleri yetiştiriciliği için balık davranışlarının bilgisayarlı görüntü işleme yöntemleriyle izlenmesi, *Journal of Anatolian Environmental and Animal Sciences*, 7 (4), 568-581, 2022. <https://doi.org/10.35229/jaes.1197703>.
- [22] G. Jocher et al., Ultralytics/yolov5: v7. 0-yolov5 sota realtime instance segmentation, Zenodo, 2022.
- [23] J. Terven, D. M. Cordova-Esparza and J. A. Romero-Gonzalez, A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas. *Machine Learning and Knowledge Extraction*, 5 (4), 1680-1716, 2023. <https://doi.org/10.3390/make5040083>.
- [24] R. Munawar and G. Jocher, Ultralytics. <https://github.com/ultralytics/ultralytics>, Accessed 5 February 2024.
- [25] M. Tan, R. Pang, and Q. V. Le, Efficientdet: Scalable and efficient object detection. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10781-10790, 2020.
- [26] A. Srikanth, A. Srinivasan, H. Indrajit and N. Venkateswaran, Contactless object identification algorithm for the visually impaired using efficientdet. *Sixth International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, pp. 417-420, Chennai, India, 2021.

