

T.C.
ERZİNCAN BİNALİ YILDIRIM ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
YAPAY ZEKA VE ROBOTİK ANABİLİM DALI

DENGESİZ GÖRÜNTÜ VERİ KÜMELERİNDE NESNE TANIMA İÇİN GÖRSEL
TRANSFORMATÖR TABANLI ÖZNİTELİK ÇIKARMA VE MAKİNE
ÖĞRENMESİ SINIFLANDIRICILARINI DAYALI HİBRİT BİR YAKLAŞIM

Ali Khudhair Abbas ALI

Danışman: Dr. Öğr. Üyesi Yıldız AYDIN

TEZ JÜRİ ÜYELERİ

Dr. Öğr. Üyesi Işıl Karabey Aksakallı
Dr. Öğr. Üyesi Yıldız AYDIN
Doç. Dr. VOLKAN KAYA

YÜKSEK LİSANS TEZİ
ERZİNCAN, 2025

© 2025 [Ali Khudhair Abbas ALI]. Tüm hakları saklıdır.

Kabul ve Onay Sayfası

Dr. Öğr. Üyesi Yıldız AYDIN danışmanlığında, Ali Khudhair Abbas ALI tarafından hazırlanan bu çalışma 23 / 12 / 2025. tarihinde aşağıdaki jüri tarafından Yapay Zeka ve Robotik Yüksek Lisans Tezi olarak kabul oybirliği (3/3) ile kabul edilmiştir.

Başkan: Dr. Öğr. Üyesi Işıl Karabey Aksakallı

İmza:

Üye: Dr. Öğr. Üyesi Yıldız AYDIN

İmza:

Üye: Doç. Dr. VOLKAN KAYA

İmza:

Yukarıdaki Yüksek Lisans Tezi Enstitü Yönetim Kurulunun / / 2025 tarih ve/..... sayılı kararı ile onaylanmıştır.

Doç. Dr. Kemal Volkan ÖZDOKUR

Enstitü Müdür V.

Not: Bu tezde kullanılan özgün ve başka kaynaklardan yapılan bildirişlerin, şekil ve tabloların kaynak olarak kullanımı, 5846 sayılı Fikir ve Sanat Eserleri Kanunundaki hükümlere tabidir.

Bilimsel Etięe Uygunluk Sayfası

“Dengesiz Görüntü Veri Kümelerinde Nesne Tanıma İçin Görsel Transformatör Tabanlı Öznitelik Çıkarma Ve Makine Öğrenmesi Sınıflandırıcılarını Dayalı Hibrit Bir Yaklaşım” isimli “Yüksek Lisans” tezım tarafımca intihal tespit programı ile incelenmiştir. Buna göre tezimde bilimsel etik ihlali ve intihal olarak nitelendirilebilecek herhangi bir durum olmadığını taahhüt ederim.

Bu çalışmadaki tüm bilgilerin, akademik ve etik kurallara uygun bir biçimde elde edildiğini; aynı zamanda bu kural ve davranışların gerektirdiğı gibi, bu çalışmanın özünde olmayan tüm materyal ve sonuçları tam olarak aktardığımı ve referans gösterdiğimi beyan ederim. 23 / 12 / 2025 .

(İmza)

Ali Khudhair Abbas ALI

ÖZET

DENGESİZ GÖRÜNTÜ VERİ KÜMELERİNDE NESNE TANIMA İÇİN GÖRSEL TRANSFORMATÖR TABANLI ÖZNİTELİK ÇIKARMA VE MAKİNE ÖĞRENMESİ SINIFLANDIRICILARINI DAYALI HİBRİT BİR YAKLAŞIM

Ali Khudhair Abbas ALI

**Yüksek Lisans Tezi, Erzincan Binali Yıldırım Üniversitesi, Fen Bilimleri Enstitüsü,
Yapay Zeka ve Robotik Anabilim Dalı
Danışman: Dr. Öğr. Üyesi Yıldız AYDIN**

2025, 55 sayfa

Bu çalışma, Görsel Transformatör (Vision Transformer, ViT) modelinden elde edilen özniteliklerin klasik makine öğrenmesi sınıflandırıcıları olan LightGBM, AdaBoost, ExtraTrees ve Lojistik Regresyon ile birleştirilerek nesne tanıma performansının artırılmasını amaçlamaktadır. Araştırmada, Caltech-101 veri seti üzerinde gerçekleştirilen deneyler sonucunda farklı öznitelik çıkarıcı-sınıflandırıcı kombinasyonları karşılaştırılmıştır. Özellikle ViT ve Lojistik Regresyon kombinasyonu, önerilen yöntem olarak öne çıkmış ve %95,5 doğruluk ile %89,7 duyarlılık değerlerine ulaşmıştır. Bu sonuçlar, önerilen yaklaşımın mevcut son teknoloji yöntemlere kıyasla önemli ölçüde daha yüksek performans sergilediğini ortaya koymaktadır. Elde edilen bulgular, ViT modelinden çıkarılan özniteliklerin klasik sınıflandırıcılarla sentezlenmesinin nesne tanıma görevlerinde kayda değer ilerlemelere yol açabileceğini göstermektedir. Önerilen hibrit yöntem, derin öğrenme tabanlı otomatik öznitelik çıkarımının yüksek temsil gücünü; geleneksel makine öğrenimi sınıflayıcılarının anlaşılabilirliği ve düşük hesaplama maliyeti gibi avantajlarıyla birleştirerek doğruluk, işlem performansı ve genel sistem verimliliği açısından önemli iyileştirmeler sağlama potansiyeline sahiptir. Ayrıca, gerçek zamanlı analizler ve sınırlı donanıma sahip akıllı sistem uygulamaları için etkin bir çözüm alternatifi olarak öne çıkmaktadır.

Anahtar Kelimeler: Nesne tanıma, görsel transformatörler, öznitelik çıkarımı, Caltech-101.

ABSTRACT

A HYBRID APPROACH BASED ON VISUAL TRANSFORMER-BASED FEATURE EXTRACTION AND MACHINE LEARNING CLASSIFIERS FOR OBJECT RECOGNITION IN UNBALANCED IMAGE DATASETS

Ali Khudhair Abbas ALI

**Master's Thesis, Erzincan Binali Yıldırım University, Institute of Science and
Technology,
Department of Artificial Intelligence and Robotics**

Advisor: Asst. Prof. Dr. Yıldız AYDIN

2025, 55 pages

This study aims to improve object recognition performance by combining features obtained from the Vision Transformer (ViT) model with classical machine learning classifiers such as LightGBM, AdaBoost, ExtraTrees, and Logistic Regression. In the research, different feature extractor-classifier combinations were compared as a result of experiments conducted on the Caltech-101 dataset. In particular, the combination of ViT and Logistic Regression stood out as the recommended method, achieving 95.5% accuracy and 89.7% sensitivity. These results demonstrate that the proposed approach exhibits significantly higher performance compared to existing cutting-edge methods. The findings show that synthesizing features extracted from the ViT model with classical classifiers can lead to significant improvements in object recognition tasks. The proposed hybrid method offers high representational power for deep learning-based automated feature extraction; By combining the advantages of traditional machine learning classifiers, such as intelligibility and low computational cost, it has the potential to deliver significant improvements in accuracy, processing performance, and overall system efficiency. Furthermore, it stands out as an effective alternative solution for real-time analytics and intelligent system applications with limited hardware.

Keywords: Object recognition, visual transformers, feature extraction, Caltech-101.

TEŞEKKÜR

Öncelikle, tüm akademik çabalarımnda başarılı olmamı sağlayan sağlık ve esenlik nimetleri için Yüce Allah’a şükrediyorum. Ardından, başından sonuna kadar doğru rehberliği, yapıcı yorumları ve değerli katkılarıyla bana yol gösteren tez danışmanım Dr. Öğretim Üyesi Yıldız AYDIN’a en derin minnettarlığımı sunarım. Süreç boyunca sürekli cesaretlendirmesi ve her aşamada desteğini esirgememesi, bu tezin tamamlanmasında büyük rol oynamıştır.

Bana her zaman parlak bir gelecek dileyen sevgili anneme ve en zor zamanlarımda bana ışık olan babama teşekkür ederim. Özellikle eşim, kızım ve kız kardeşlerim, gösterdikleri sevgi ve sabırla bu sürecin en büyük motivasyon kaynağı olmuşlardır.

Bu araştırmayı tamamlamamda büyük destekleri olan arkadaşlarıma ve hocalarıma da teşekkür ederim. Ayrıca, sınıf arkadaşlarıma ve meslektaşlarıma bilgilerini benimle paylaştıkları için minnettarım. Onların tavsiyeleri ve teşvikleri, doğru yolda ilerlememde hayati önem taşımıştır.

Ali Khudhair Abbas ALI

Aralık, 2025

İÇİNDEKİLER

ÖZET	i
ABSTRACT	ii
TEŞEKKÜR	iii
İÇİNDEKİLER	iv
TABLolar DİZİNİ	vi
ŞEKİLLER DİZİNİ	vii
SİMGELER VE KISALTMALAR DİZİNİ	viii
1. GİRİŞ.....	1
2. KAVRAMSAL ÇERÇEVE VE İLGİLİ ÇALIŞMALAR	4
2.1. Geleneksel ve Elle Tasarlanmış Öznitelikler	4
2.1.1. Ölçekten bağımsız öznitelik dönüşümü (SIFT)	4
2.1.2. Hızlandırılmış sağlam öznitelikler (SURF)	5
2.1.3. Yönlendirilmiş FAST ve döndürülmüş BRIEF (ORB)	5
2.1.4. İkili sağlam ve ölçeklenebilir anahtar noktalar (BRISK)	6
2.1.5. Maksimum kararlı uç bölgeler (MSER)	7
2.1.6. KAZE.....	8
2.1.7. Kelimelerin çantası yöntemi (BoVW).....	8
2.2. Klasik Makine Öğrenmesi	9
2.2.1. Makine öğrenmesi sınıflandırıcıları	10
2.2.1.1. Lojistik regresyon (LR)	10
2.2.1.2. Hafif gradyan artırma makinesi (LightGBM)	11
2.2.1.3. AdaBoost.....	12
2.2.1.4. Aşırı rastgeleştirilmiş ağaçlar (Extra Trees).....	13
2.3. Modern Derin Öğrenmeye Dayalı Yöntemler	14
2.3.1. Evrişimli sinir ağları (CNN)	15
2.3.1.1. Evrişimsel sinir ağı katmanları.....	15
2.3.1.1.1. Giriş katmanı	16
2.3.1.1.2. Evrişim katmanı	16
2.3.1.1.3. Havuzlama katmanı	17
2.3.1.1.4. Aktivasyon katmanı	18
2.3.1.1.5. Düzleştirme katmanı	18
2.3.1.1.6. Tam bağlı katman	19
2.3.1.1.7. Sınıflandırma katmanı.....	19

2.3.2. Görsel transformatörler (ViT).....	19
2.3.3. VGG16.....	21
2.3.4. EfficientNetB0	21
2.4. Literatür Taraması.....	23
3. YÖNTEM.....	30
3.1. ViT Modelinin Matematiksel Yapısı.....	32
3.2. LR Optimizasyon Prensipleri.....	33
4. BULGULAR.....	35
4.1. Veri Seti (Caltech-101).....	35
4.2. Performans Metrikleri	36
4.2.1. Doğruluk (Accuracy).....	37
4.2.2. Kesinlik (Precision)	38
4.2.3. Duyarlılık (Recall).....	38
4.2.4. F1 değeri (F1 Score).....	38
4.3. Deneysel Sonuçlar	38
5. SONUÇLAR.....	46
KAYNAKÇA	48
ÖZGEÇMİŞ.....	Error! Bookmark not defined.

TABLÖLAR DİZİNİ

Tablo 1. Literatür özeti.....	27
Tablo 1. (Devamı).....	28
Tablo 2. Tüm yöntemlere ait Doğruluk (Accuracy) değerleri.....	40
Tablo 3. Tüm yöntemlere ait Kesinlik (Precision) değerleri.....	41
Tablo 4. Tüm yöntemlere ait Duyarlılık (Recall) değerleri	42
Tablo 5. Tüm yöntemlere ait F1- değeri değerleri.....	42
Tablo 6. Önerilen sistemin bazı yeni deneylerle karşılaştırmalı analizi	45



ŞEKİLLER DİZİNİ

Şekil 1. Yapay zeka, makine öğrenmesi, derin öğrenme kronolojisi	9
Şekil 2. Lojistik regresyonun ikili sınıflandırma modeli	11
Şekil 3. Hafif gradyan artırma makinesi	12
Şekil 4. AdaBoost'un çalışma prensibi	13
Şekil 5. Aşırı rastgeleleştirilmiş ağaçlar sınıflandırıcısı	14
Şekil 6. CNN mimarisi	15
Şekil 7. Filtreleme işlemine ait örnek	17
Şekil 8. Maksimum havuzlama tekniği örnek	18
Şekil 9. Düzleştirme katmanı örnek	19
Şekil 10. Görsel transformatörler (ViT) mimarisi	20
Şekil 11. VGG16 mimarisi	21
Şekil 12. EfficientNetB0 mimarisinin genel yapısı	22
Şekil 13. Sistemin genel yapısı	30
Şekil 14. Deneysel Çalışmalara Ait Akış Diyagramı	31
Şekil 15. Caltech-101 veri setine ait örnek görüntüler	36
Şekil 16. Tez kapsamında gerçekleştirilen deneyler	39
Şekil 17. Klasik ve derin öğrenme tabanlı yöntemlerin Doğruluk (Accuracy) performansı ...	40
Şekil 18. Yöntemlerin kesinlik, Duyarlılık, F1 değerleri	43

SİMGELER VE KISALTMALAR DİZİNİ

$\bar{X}$	Ortalama
%	Yüzde
α	Güvenirlilik katsayısı
β	Regresyon katsayısı
B	Regresyon sabiti
t	t-değeri
T^*	Optimal karar ağacıdır
f	Model fonksiyonu
W	Ağırlık matrisi
$\hat{y}$	Tahmin edilen sınıf (0 veya 1)
x_i	Giriş özellikleridir
σ	Sigmoid activation fonksiyonu
YZ	Yapay zeka(YZ) (Artificial intelligence)
ML	Makine öğrenmesi (Machine learning)
DL	Derin öğrenme (Deep learning)
DT	Karar ağacı (Decision tree)
LR	Lojistik regresyon (Logistic regression)
BoVW	Kelimelerin Çantası Yöntemi
YSA	Yapay sinir ağı (Artificial Neural Networks)
SVM	Destek vektör makineleri (Support vector machine)
K-NN	K-En yakın komşu algoritması (K-nearest neighbors)
CNN	Konvolüsyonel sinir ağı (Convolutional neural network)
ViT	Görsel transformatör (Vision transformer)

1. GİRİŞ

Nesne tanıma, yapay zeka ve görüntü işleme alanında sıklıkla araştırılan bir konudur. Görsellerde bulunan nesneleri tanımlayan uygulamalar nesne tanıma uygulamaları olarak adlandırılmaktadır. Nesne tanıma uygulamaları, akıllı robot üretiminde önemli bir alt bileşen olmasının yanı sıra güvenlik sistemlerinde, otomatik sürüş teknolojilerinde ve sağlık alanlarında da kullanılmaktadır (Dosovitskiy vd., 2021; Krizhevsky vd., 2012). Teknolojinin hızlı gelişimiyle birlikte akıllı cihazların günlük yaşamda kullanımı her geçen gün artmaktadır. Bu nedenle, döndürme, ölçek ve aydınlatma değişiklikleri gibi değişken faktörlere karşı etkili ve verimli bir nesne tanıma uygulaması geliştirmek büyük önem taşımaktadır. Gerçek dünya uygulamalarında kullanılan görüntülerde nesneler her zaman net olmayabilir ve arka plan karmaşası gibi sorunlarla karşılaşılabilir. Nesne tanıma uygulamalarında bu tür sorunların üstesinden gelebilmek için etkili ve verimli çözümler geliştirilmesi gerekmektedir. Yapay sinir ağları ve derin öğrenme tabanlı yaklaşımlar yüksek doğruluk sağlasa da, sınıf dengesizliği ve yüksek hesaplama maliyeti gibi zorluklar da ortaya çıkarmaktadır (Fei-Fei vd., 2007). Son dönemde Görsel transformatörler (ViT) modelleri nesne tanıma görevlerinde önemli başarılar göstermiş ve geleneksel CNN tabanlı yöntemlere alternatif olarak öne çıkmıştır (Touvron vd., 2021). Ancak ViT modellerinin yüksek hesaplama gereksinimleri ve dengesiz veri setlerinde performans düşüşü, yeni hibrit yaklaşımların geliştirilmesini gerekli kılmıştır (Naseer, Alzahrani, vd., 2024). Bu çalışma, ViT tabanlı öznitelik çıkarımı ile klasik makine öğrenmesi sınıflandırıcılarını entegre ederek nesne tanıma performansını artırmayı amaçlamaktadır. Nesne tanıma uygulamaları temelde iki adımdan oluşmaktadır: öznitelik çıkarımı ve sınıflandırma. Bu adımlarda kullanılacak yöntemler doğrudan uygulamanın başarısını etkilemektedir. Nesne tanıma uygulamaları klasik makine öğrenmesi sınıflandırıcıları veya derin öğrenme yöntemleriyle geliştirilebilmektedir. Öznitelik çıkarımı, nesne tanıma ve görüntü işleme uygulamalarında kritik bir adımdır. Bu çalışmada kullanılan öznitelik çıkarım yöntemleri farklı yapısal ve algoritmik özniteliklere sahiptir. SIFT (Scale-Invariant Feature Transform), ölçek ve döndürme değişikliklerine karşı dayanıklılığı nedeniyle nesne tanımada yaygın olarak kullanılmaktadır. SURF (Speeded-Up Robust Features), SIFT'e benzer ancak daha yüksek hesaplama verimliliği sunar. ORB (Oriented FAST and Rotated BRIEF), hızlı ve döndürmeye karşı dayanıklı bir öznitelik çıkarım tekniğidir. KAZE, diferansiyel operatörler kullanarak kenar ve dokuları tespit ederken, MSER (Maximally Stable Extremal Regions) değişen parlaklık koşullarında kararlı kalan bölgeleri belirler. BRISK (Binary Robust Invariant Scalable Keypoints) ise hızlı ve ölçeklenebilir bir öznitelik çıkarım yöntemidir. Bu teknikler,

görüntüdeki kilit noktaları belirleyerek sınıflandırma algoritmalarının performansını artırmaktadır. Klasik makine öğrenmesi yöntemlerinde SIFT, SURF, KAZE ve ORB gibi yerel veya küresel öznitelikler kullanılırken, derin öğrenme yöntemleri öznitelikleri kendileri çıkarmaktadır.

Nesne tanıma uygulamalarında karşılaşılabilecek temel sorunlar şunlardır:

1. Görüntüdeki nesnenin arka planla karışması.
2. Düşük çözünürlüklü veya bulanık görüntüler.
3. Aydınlatma koşullarının değişkenliği.
4. Nesnenin döndürülmesi, ölçeklenmesi.

Literatürde nesne tanıma uygulamalarında karşılaşılan bu sorunları çözmek için çeşitli hibrit yaklaşımlar önerilmektedir (Aysha Naseer Nouf Abdullah Almujaally, 2024; Naseer, Mudawi, vd., 2024; Sikder vd., 2024). Son yıllarda hibrit yöntemlerin kullanımıyla elde edilen başarılı sonuçlar, hibrit yöntemlere yönelik çalışmaların artmasına yol açmıştır. Hibrit sistemler, derin öğrenme yöntemleri, klasik öznitelikler veya klasik makine öğrenmesi sınıflandırıcılarının farklı varyasyonlarla birleştirilmesiyle elde edilmektedir. Bu hibrit sistemler sayesinde, özellikle klasik makine öğrenmesi veya derin öğrenme yöntemleriyle geliştirilen ve düşük başarı oranına sahip uygulamalarda daha başarılı sonuçlar elde edilmiştir. Bu uygulamalara örnek olarak dengesiz veri setleriyle gerçekleştirilen çalışmalar verilebilir. Dengesiz veri setlerinde, bir sınıftaki örnek sayısı oldukça düşükken diğer sınıflarda örnek sayısı oldukça yüksek olabilir. Günlük yaşamda sıkça karşılaşılan bu sorunun çözülmesi büyük önem taşımaktadır. Bu çalışmada dengesiz bir veri seti olan Caltech-101 kullanılmıştır. Literatürdeki çalışmalarda, dengesiz veri setleriyle klasik yöntemler kullanılarak yapılan uygulamalarda, örnek sayısı fazla olan sınıfa doğru bir eğilim olduğu gözlemlenmiştir. Bu çalışmada, dengesiz veri setlerinin sorunlarını ortadan kaldırarak daha etkili ve verimli bir nesne tanıma uygulaması geliştirmek amacıyla hibrit bir yöntem önerilmektedir. Önerilen hibrit yöntemde öznitelik çıkarım adımında Görsel transformatörler (ViT) yöntemi, sınıflandırma adımında ise LightGBM, AdaBoost, ExtraTrees ve Lojistik Regresyon sınıflandırıcıları kullanılmıştır. Bu yaklaşımın benimsenmesinin temel motivasyonu iki nedene dayanmaktadır:

- Dengesiz veri setlerinin sınırlamalarını aşmak için derin öğrenme tabanlı öznitelikler sağlam sınıflandırma modelleriyle entegre edilerek, özellikle az temsil edilen sınıflar için daha iyi genelleme sağlanabilir.
- Gürültülü ve düşük çözünürlüklü veri setlerinde, büyük ölçekli hesaplama kaynaklarına ihtiyaç duymadan tanıma doğruluğunu artırmak, yöntemi daha verimli ve gerçek dünya uygulamalarında uygulanabilir hale getirmektedir.

Klasik makine öğrenmesi ve derin öğrenme yaklaşımlarının güçlü yönlerini birleştirmek amacıyla önerilen yöntem, Görsel transformatörler (ViT) modelini klasik makine öğrenmesi sınıflandırıcılarıyla entegre etmektedir. ViT, öz-dikkat mekanizmasıyla derin öznitelikler çıkarırken, bu öznitelikler LightGBM, AdaBoost, ExtraTrees ve Lojistik Regresyon gibi güçlü makine öğrenmesi algoritmalarıyla sınıflandırılmaktadır. Bu hibrit yaklaşım, yüksek hesaplama gereksinimleri ve sınıf dengesizliğine duyarlılık gibi derin öğrenmenin sınırlamalarını aşarken sınıflandırma doğruluğunu artırmaktadır. Yalnızca derin öğrenme tabanlı sınıflandırma kullanmak yerine, önerilen yöntem çıkarılan öznitelikleri makine öğrenmesi sınıflandırıcılarıyla daha verimli işleyerek özellikle az temsil edilen sınıfların genellemesini iyileştirmektedir. ViT yöntemi, daha az parametreyle daha başarılı sonuçlar elde edebildiği için öznitelik çıkarım adımında tercih edilmiştir (Venugopal vd., 2022). Sonuç olarak, önerilen hibrit yöntem ile daha etkili ve verimli bir nesne tanıma uygulaması geliştirilmiştir. Bu yaklaşımın literatüre katkıları üç başlık altında özetlenebilir:

- Dengesiz bir veri seti olan Caltech-101 veri setindeki dengesizlik probleminin etkisinin azaltılması.
- Derin öğrenme yöntemleriyle çıkarılan özniteliklerin klasik makine öğrenmesi sınıflandırıcılarıyla birlikte kullanılmasıyla daha başarılı bir hibrit yaklaşım önerilmesi.
- Yöntemin kapsamlı deneysel doğrulama yoluyla nesne tanıma görevlerinde mevcut en güncel yöntemlere kıyasla üstün performans göstermesi.

Önerilen yöntemin temel yeniliği, bazı nesne kategorilerinin yeterince temsil edilmediği dengesiz veri setlerinin dezavantajlarını ortadan kaldırmasıdır. Mevcut çalışmaların çoğu bu dezavantajları, sınıf dengesizliğini gidermek için veri artırma teknikleri veya birleşik öğrenme (ensemble learning) stratejileri uygulayarak ortadan kaldırmaktadır. Ancak önerilen yöntem, derin öznitelik çıkarımı yöntemini klasik makine öğrenmesi sınıflandırıcılarıyla entegre ederek daha etkili bir çözüm sunmaktadır. Bu yaklaşımda, literatürde yaygın olarak kullanılan dengesiz bir veri seti olan Caltech-101 kullanılmıştır. Deneysel sonuçlar, önerilen yaklaşımın kesinlik (%92,3), duyarlılık (%89,7) ve doğruluk (%95,5) açısından en güncel yöntemlerden daha iyi performans gösterdiğini ortaya koymaktadır. Çalışmada ikinci bölümde ilgili literatür ayrıntılı olarak ele alınmakta, üçüncü bölümde yöntem ayrıntılı olarak açıklanmaktadır. Dördüncü bölümde deneylerden elde edilen sonuçlar sunulmakta, beşinci ve son bölümde ise genel sonuçlar ile geleceğe yönelik öneriler yer almaktadır.

2. KAVRAMSAL ÇERÇEVE VE İLGİLİ ÇALIŞMALAR

Nesne tanıma, yapay zeka ve görüntü işleme alanında sıklıkla araştırılan bir konudur ve güvenlik, otomatik sürüş, sağlık gibi birçok uygulama alanında kritik rol oynamaktadır. Bu alanda kullanılan yöntemler temel olarak klasik makine sınıflandırıcıları ve derin öğrenme tabanlı yöntemler olmak üzere 2 kategoriye ayrılmaktadır. Klasik makine sınıflandırıcıları belirli koşullarda başarılı olsa da, dengesiz veri setlerinde sınırlı performans göstermektedir. Derin öğrenme tabanlı yaklaşımlar (VGG16, EfficientNetB0, ViT) daha yüksek doğruluk sağlamış, ancak sınıf dengesizliği ve yüksek hesaplama maliyeti gibi sorunlarla karşılaşmıştır. Literatürde hibrit yöntemlerin bu sorunları azaltmada etkili olduğu gösterilmiş, özellikle ViT tabanlı öznetelik çıkarımı ile klasik sınıflandırıcıların entegrasyonu başarılı sonuçlar vermiştir (Aysha Naseer Nouf Abdullah Almujaally, 2024; Touvron vd., 2021). Bu bağlamda, önerilen hibrit yaklaşım literatürdeki mevcut yöntemlere kıyasla daha yüksek performans ortaya koymaktadır.

2.1. Geleneksel ve Elle Tasarlanmış Öznetelikler

Derin öğrenme kullanımının yaygınlaşmasından önceki dönemde, nesne tanıma ve bilgisayarlı görü sistemleri genellikle el yapımı öznetelik (hand-crafted features) çıkarım yöntemlerine dayanarak görüntüleri analiz etmekteydi (Bay vd., 2006; Dalal & Triggs, 2005; Lowe, 2004). Bu yöntemler, görüntülerdeki belirgin noktaları tespit ederek dokuların ve diğer görsel yapıların tanınmasına olanak sağlamaktadır. Her ne kadar derin öğrenme yöntemleri tarafından büyük ölçüde geride bırakılmış olsalar da, geleneksel yöntemler hâlâ gerçek zamanlı sistemlerde, düşük güç tüketimine sahip aygıtlarda ve sınırlı eğitim verisiyle çalışılması gereken durumlarda kullanılmaktadır. Bu bağlamda hibrit işleme hatlarında önemli bir rol üstlenmekte ve birçok modern yöntemde bulunmayan açıklanabilirlik avantajı sağlamaktadırlar (Cheng vd., 2025).

2.1.1. Ölçekten bağımsız öznetelik dönüşümü (SIFT)

SIFT algoritması, ölçekten bağımsız anahtar noktaların belirlenmesi için Gauss fonksiyonlarıyla farklı ölçeklerde bulanıklaştırılmış görüntülerden elde edilen gauss filtresi farkı (DoG) uzayını kullanmaktadır (Lowe, 2004). Bu uzayda seçilen anahtar noktaların yönelimleri gradyan bilgisiyle atanarak ölçek ve dönme değişimlerine karşı dayanıklılık sağlanır. Her anahtar nokta etrafında 16×16 piksel boyutunda bir alan tanımlanır ve bu alan 4×4 hücreye ayrılarak toplam 16 alt bölge oluşturulur. Her alt bölgede 8 yönelim histogramı

hesaplanarak 128 boyutlu bir öznitelik vektörü elde edilir. Histogram değerleri Gauss ağırlıklarıyla yumuşatılır ve doğrusal enterpolasyonla komşu bölmelere dağıtılarak sınır etkileri azaltılmaktadır (Lowe, 2004). Literatürde SIFT'in, alternatif yerel tanımlayıcılara göre daha yüksek performans sergilediği rapor edilmiştir (Mikolajczyk & Schmid, 2005). Denklem 1'de ise farklı ölçeklerde hesaplanan iki Gauss fonksiyonunun farkı gösterilmektedir (Tareen & Saleem, 2018).

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \quad (1)$$

Denklem 1'de $G(x, y, \sigma)$ Gauss fonksiyonunu, $I(x, y)$ giriş görüntüsünü, $D(x, y, \sigma)$ ise iki ölçek arasındaki farkı ifade etmektedir.

2.1.2. Hızlandırılmış sağlam öznitelikler (SURF)

SURF algoritması, SIFT algoritmasına kıyasla daha hızlı bir seçenek olup benzer bir sağlamlık seviyesini korurken işlem süresini önemli ölçüde azaltmaktadır. Anahtar nokta belirleme için integral görüntü ve Hessian matrisinin yaklaşık hesaplamasını kullanır ve öznitelik tanımlama aşamasında Haar dalgacık tepki dağılımlarını temel alır. SURF özellikle robotik görme veya artırılmış gerçeklik gibi gerçek zamanlı uygulamalarda hızın ve doğruluğun önemli olduğu durumlarda oldukça etkilidir. Ancak SIFT gibi SURF de patentli bir algoritmaya sahiptir ve bu durum bazı ticari projelerde kullanımını sınırlandırabilir (Anandhasilambarasan vd., 2024; Tareen & Saleem, 2018). Denklem 2'de σ ölçeğinde $x(x, \sigma)$ noktasındaki Hessian Matrisini temsil eden formül verilmiştir (Aydın & Türkeş, 2025).

$$H(x, \sigma) = f(x) = \begin{cases} L_{xx}(x, \sigma), & L_{xy}(x, \sigma) \\ L_{xy}(x, \sigma), & L_{yy}(x, \sigma) \end{cases} \quad (2)$$

2.1.3. Yönlendirilmiş FAST ve döndürülmüş BRIEF (ORB)

ORB algoritması, SIFT ve SURF algoritmalarına alternatif olarak 2011 yılında Rublee ve arkadaşları tarafından oluşturulmuştur (Rublee vd., 2011). Gerçek zamanlı uygulamalar için iyi çalışan hızlı ve etkili bir çözümdür. Ayrıca patent kısıtlamalarından etkilenmemesi ile de öne çıkmaktadır. ORB temel noktaları bulmak için köşe tespit algoritması (FAST), öznitelikleri tanımlamak için öznitelik tanımlayıcı algoritması (BRIEF) yöntemini birleştirir. FAST algoritması köşeleri hızlı ve verimli bir şekilde bulmasına rağmen yönelim değişikliklerine karşı dayanıklı değildir. ORB algoritması yoğunluğun merkezinden her bir anahtar noktanın

yönünü hesaplayarak bu sınırlamayı ortadan kaldırır ve dönmeye karşı dirençli hale getirir. Ayrıca, BRIEF tanımlayıcıları dönme değişikliklerine duyarlı olduğundan, ORB algoritması yönlendirilmiş BRIEF tanımlayıcıları kullanarak dönmeye karşı sağlam (robust) bir sürüm oluşturur. ORB algoritması daha iyi öznitelik eşleştirmesi için öğrenme tabanlı bir yaklaşım kullanarak ilişkisiz ikili testleri seçmektedir. Düşük hesaplama maliyetine sahip olan ORB dönüş ve gürültüye karşı dayanıklılığı ile öne çıkmakta yanı sıra gömülü sistemler ve mobil platformlar için uygunluğuyla da dikkat çekmektedir. Performans ve hız arasındaki dengeyi başarılı bir şekilde sağlayan ORB algoritması genellikle robotik sistemlerde ve görsel Eş Zamanlı Konumlandırma ve Haritalama (SLAM) ile artırılmış gerçekliği içeren uygulamalarda kullanılmaktadır (Rublee vd., 2011). Denklem 3'te, doğrusal olmayan difüzyon sürecine ait formül sunulmaktadır.

$$\theta = \tan^{-1} \left(\frac{m_{01}}{m_{10}} \right) \quad (3)$$

2.1.4. İkili sağlam ve ölçeklenebilir anahtar noktalar (BRISK)

BRISK adlı algoritma 2011 yılında Leutenegger ve arkadaşları tarafından (Leutenegger vd., 2011) tanıtıldı. BRISK, öne çıkma ve tanımlama işlemlerinde hem öznitelik değişimlerine karşı dayanıklılığı hem de hesaplama verimliliğini artırmayı amaçlar. Her anahtar nokta etrafında çevresel bir örnekleme düzeni kullanır ve hızlı bir şekilde anahtar noktalara ulaşmak için AGAST dedektörünü kullanır. Tanımlayıcı iki bileşenli bir yapıya sahiptir ve tanımlanmış örnekleme düzeni üzerinde gerçekleştirilen ikili yoğunluk karşılaştırmalarına dayanmaktadır. Bu yaklaşım BRIEF'e benzer ancak daha yüksek ölçeklenebilirlik ve sağlamlık sunacak şekilde uyarlanır. BRISK algoritması, hızlı çalışması ve düşük bellek gereksinimi sayesinde gerçek zamanlı bilgisayarlı görü uygulamaları için uygun bir öznitelik çıkarma ve eşleştirme yöntemidir. Özellikle nesne tanıma, nesne takibi ve mobil görme sistemlerinde yaygın olarak kullanılmaktadır. SIFT veya SURF kadar yüksek doğrulukta eşleşme üretmese de, hız ve doğruluk arasında dengeli bir performans sunması nedeniyle sınırlı hesaplama kaynaklarına sahip gömülü platformlar için ideal bir alternatif olarak değerlendirilmektedir (Leutenegger vd., 2011). Denklem 4'te BRISK ikili tanımlayıcılarının oluşturulma süreci gösterilmektedir.

$$d_k = \begin{cases} 1 & \text{eğer } I(p_i) < I(p_j) \\ 0 & \text{aksi takdirde} \end{cases} \text{ her bir } (p_i, p_j) \text{ örnekleme çifti için} \quad (4)$$

Burada $I(p_j)$ ve $I(p_i)$, tanımlayıcı oluşturulurken karşılaştırılan iki noktanın parlaklık (yoğunluk) değerlerini ifade etmektedir. Elde edilen tüm d_k değerleri bir araya getirilerek tanımlayıcı bit vektörü oluşturulur.

2.1.5. Maksimum kararlı uç bölgeler (MSER)

Maksimum kararlı uç bölgeler (MSER) adı verilen öznitelik belirleme algoritması, Ning ve arkadaşları tarafından 2011 yılında tanıtılmış olup görüntüdeki kararlı bağlı bileşenleri belirlemek için kullanılmaktadır (Ning vd., 2011).

Diğer anahtar nokta tabanlı yöntemlerden farklı olarak MSER, görüntüdeki aşırı bölgeleri (extremal regions) tespit etmeye odaklanır.

MSER bölgeleri, görüntüde çevresine göre daha parlak veya daha koyu piksellerden oluşan bağlı kümeler olarak tanımlanır. Bu bölgeler, yoğunluğun farklı eşik değerleri boyunca şekil ve büyüklük açısından belirgin bir değişim göstermemesi durumunda ‘maksimum derecede kararlı’ kabul edilir. Bu özellik, değişken görüntüleme koşulları altında yüksek tekrarlı oranı sağlamaktadır.

MSER’in en önemli avantajlarından biri, aydınlatma değişiklikleri ve bakış açısı farklılıklarına karşı yüksek dayanıklılık göstermesidir. Özellikle metin tanıma uygulamalarında, metin ile arka plan arasındaki yüksek kontrast sayesinde karakterlerin veya kelimelerin başarılı bir şekilde ayrıştırılmasına olanak tanımaktadır.

Bununla birlikte MSER, gürültüye duyarlı olabilmekte ve gereksiz ya da çakışan bölgeler üretebilmektedir; bu durum ek filtreleme ve düzeltme işlemlerini gerekli kılmaktadır. Yüksek verimliliği ve dönüşüm öznitelikleri sayesinde MSER, birçok gerçek dünya bilgisayarlı görüş uygulamasında yaygın olarak kullanılmaktadır.

Bir bölgenin kararlılığı ise, eşik değerindeki küçük değişikliklerin bölge alanında ne kadar farklılığa yol açtığına bağlı olarak tanımlanmaktadır.

Bu kararlılık ölçütü Denklem 5’te matematiksel olarak ifade edilmektedir.

$$\Delta(R_i) = \frac{|R_{i+\Delta} - R_{i-\Delta}|}{|R_i|} \quad (5)$$

Burada R_i , i eşik seviyesindeki bölgeyi; Δ , eşik değişim aralığını; $|R_i|$ ise bölgenin piksel sayısını (alanını) ifade etmektedir.

2.1.6. KAZE

KAZE, Alcantarilla ve arkadaşları tarafından 2012 yılında önerilmiştir. Bu algoritma doğrusal olmayan ölçek uzayında çalışan bir öznitelik tespit ve tanımlama yöntemidir (Alcantarilla vd., 2012). Geleneksel yöntemlerde (SIFT, SURF) ölçek uzayı Gauss bulanıklaştırma ile oluşturulduğundan, kenarlar ve ince yapılar belirgin şekilde yumuşar. KAZE'nin temel farkı, ölçek uzayını doğrusal olmayan difüzyon filtreleme ile oluşturarak kenarları koruyan, yapısal bilgiyi daha iyi temsil eden bir çok ölçekli görüntü üretmesidir. Bu doğrusal olmayan ölçek uzayı üzerinde her bir ölçekte Hessian determinantı kullanılarak anahtar noktalar tespit edilir. Bu yaklaşım, özellikle kenar bölgelerinde daha doğru konumlandırılmış ve daha ayırt edici anahtar noktalar elde edilmesini sağlar. Tanımlayıcı oluşturma aşaması da yine bu ölçek uzayında gerçekleştirilir; böylece KAZE tanımlayıcıları gürültüye karşı daha dayanıklı ve yapısal olarak daha tutarlı hale gelir. KAZE'nin en önemli avantajı, kenar koruma ve yüksek ayırt edicilik sağlamasıdır. Ancak doğrusal olmayan difüzyonun hesaplama maliyeti yüksek olduğundan, yöntem klasik yöntemlere göre daha yavaştır. Bu nedenle daha sonra aynı araştırmacılar tarafından hızlandırılmış bir sürüm olan AKAZE geliştirilmiştir. Difüzyon sürecinin tanımı Denklem 6'da verilmiştir.

$$\frac{\partial L}{\partial t} = \text{div}(c(x, y, t) \cdot \nabla L) \quad (6)$$

2.1.7. Kelimelerin çantası yöntemi (BoVW)

Kelimelerin çantası yöntemi (BoW) ilk olarak metin sınıflandırma problemlerinde kullanılmıştır (Quinlan, 1993). Bu yaklaşımda her metin, kelimelerin sıklıklarına dayalı bir histogram ile temsil edilir; kelimelerin sırası dikkate alınmaz. Daha sonra benzer bir yöntem görsel sınıflandırmada öznitelik vektörleri üzerinden uygulanmıştır (Aydin, 2023). Görsel kelimelerin çantası yöntemi (BoVW) iki temel aşamadan oluşur: kodlama ve havuzlama. Kodlama aşamasında, görüntü yamalarından elde edilen öznitelik vektörleri üzerinde k-ortalamlar gibi bir kümeleme algoritması uygulanarak görsel sözlük oluşturulur. Bu süreç sonucunda k adet temsili 'görsel kelime' elde edilir. Havuzlama aşamasında ise bir görüntü, bu görsel kelimelerden oluşan bir çanta ile temsil edilir; bu temsil bölgesel konum bilgisi içermez (Bosch vd., 2007). Yöntem, nesnelerin konum ve yönelimleri değişse bile sınıflandırmada yüksek performans sağlamaktadır. Bununla birlikte, konumsal bilginin öznitelikler arasında bulunmaması bir dezavantaj olarak değerlendirilmektedir (Marsh vd., 2003).

2.2. Klasik Makine Öğrenmesi

Yapay zeka, görsel girdilere dayalı karar alma süreçlerinde kapsamlı bir çerçeve sunmaktadır. Yapay zekanın bir alt alanı olan makine öğrenmesi, verilerden öğrenebilen algoritmaların geliştirilmesine odaklanmaktadır. Programlanmak yerine örneklerle dayalı olarak çalışır; verilerden desenler çıkarır ve bu temelde tahminlerde bulunur veya kararlar alır (Alpaydin, 2014). Nesne tanıma görevlerinde makine öğrenmesi sayesinde sistemler, daha önce görülmemiş verilere genelleme yapabilmektedir. Makine öğrenmesi içerisinde yer alan yapay sinir ağları, öznitelik çıkarma sürecine dayanmakta ve anormallik tespiti, yüz tanıma, nesne takibi gibi çeşitli uygulamalarda insan seviyesinde veya daha yüksek doğruluk sağlayabilmektedir. Bu ağların çok katmanlı versiyonları olan derin öğrenme yöntemleri ise modern bilgisayarlı görme sistemlerinde nesne tanıma, sahne analizi ve ortam algısı gibi alanlarda önemli ilerlemelere öncülük etmiştir (LeCun vd., 2015). Son yıllarda yüksek performanslı bilgi işleme kaynaklarının (örneğin GPU tabanlı paralel işlemciler) ve büyük ölçekli veri setlerinin kullanımı, bilgisayarlı görme sistemlerinin doğruluk ve genelleme başarısını artırmıştır (He vd., 2016; Krizhevsky vd., 2017). Bilgisayarlı görme amacıyla makine öğrenmesinin etiketlenmiş görüntü verilerinden öğrenmeyi sağlayan yöntemleri kullanması önemli bir yaklaşım olarak değerlendirilmektedir (Gollapudi, 2019). Şekil 1’de yapay zekâ, makine öğrenmesi ve derin öğrenme arasındaki hiyerarşik ilişki kronolojik olarak sunulmaktadır. Şekilden görüldüğü üzere, bilgisayarlı görme hem yapay zekânın hem de makine öğrenmesinin bir alt alanı olarak konumlanmakta ve derin öğrenme yöntemleriyle desteklenerek pratik uygulamalarda gelişimini sürdürmektedir.



Şekil 1. Yapay zeka, makine öğrenmesi, derin öğrenme kronolojisi (TAN vd., 2021)

Bu yapılanma, bilgisayarlı görmenin hem yapay zekanın hem de makine öğreniminin bir parçası olduğunu ve pratik bir uygulama alanı olarak gelişimini sürdürdüğünü ortaya koymaktadır.

2.2.1. Makine öğrenmesi sınıflandırıcıları

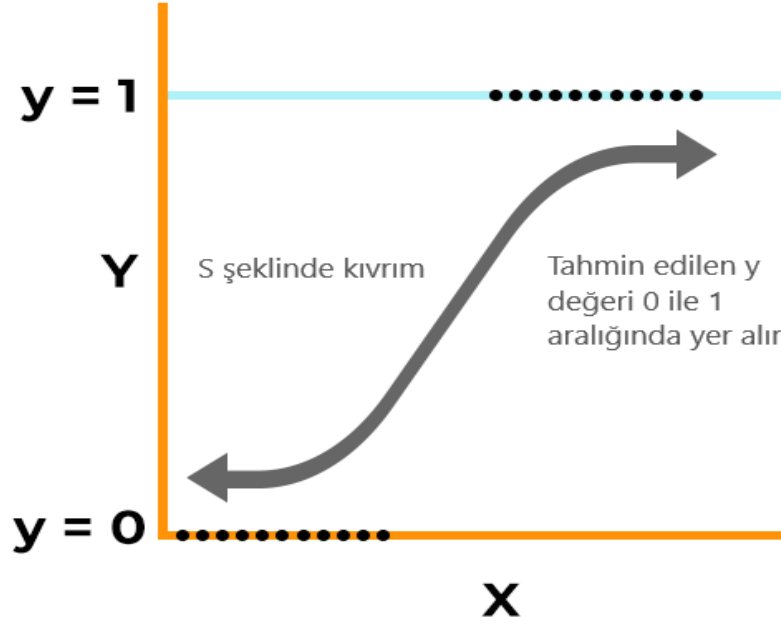
Denetimli sınıflandırma algoritmaları genellikle etiketlenmiş eğitim verilerinin mevcut olduğunda nesne tanıma sistemlerinin temelini oluşturmakta ve birçok nesne tanıma sistemi için önemlidir. Bu algoritmalar genellikle görüntülerden elde edilen giriş özniteliklerini kategorik çıktı etiketleriyle eşleştirecek şekilde derin öğrenme veya elle çıkarılan yöntemler kullanarak öğrenilir. Derin sinir ağları genellikle yüksek boyutlu görüntü verilerinde üstün performans sergilerken, bazı geleneksel denetimli öğrenme algoritmaları ve topluluk temelli yöntemler de belirli koşullarda rekabetçi sonuçlar verebilmektedir. Özellikle tablo biçimli özniteliklerin (örneklerin satır-sütun yapısında düzenlediği sayısal özellikler) kullanıldığı, veri miktarının sınırlı olduğu veya yorumlanabilirliğin öncelikli olduğu senaryolarda bu yöntemler tercih edilmektedir. Bu bölümde, nesne tanıma sürecinde en yaygın olarak kullanılan denetimli sınıflandırıcılar özetlenmiş ve performans ölçütleri ile pratik kullanım alanları ele alınmaktadır.

2.2.1.1. Lojistik regresyon (LR)

LR genellikle ikili sınıflandırma görevleri için kullanılan doğrusal bir modeldir. Ancak bir-sınıfa-karşı-diğerleri gibi teknikler ile çoklu sınıflı problemlere genişletilebilir bir modeldir. Bu yöntemde verilen bir girdinin belirli bir sınıfa katılma olasılığı sigmoid fonksiyonu kullanılarak tahmin edilmektedir (Kravets vd., 2024). Sigmoid fonksiyonu denklem 7’de tanımlanmaktadır.

$$Z = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 \chi_1 + \beta_2 \chi_2 + \dots + \beta_n \chi_n \quad (7)$$

Burada P Girdinin sınıf 1'e ait olma olasılığını göstermektedir. β_0 modelin kesim noktası (bias), β_n ise bağımsız değişkenlerin katsayılarıdır. $X_1, \dots, X_n$: giriş özniteliklerini ifade etmektedir. Şekil 2, LRnin temel kavramını göstermektedir. Bu tür uygulamalarda amaç, giriş özniteliklerine dayanarak ikili bir sonuç (0 veya 1) tahmin etmektir. Grafikte, giriş değişkeni X ile tahmin edilen olasılık Y arasındaki ilişki, karakteristik S şeklindeki bir sigmoid eğrisi ile gösterilmektedir.

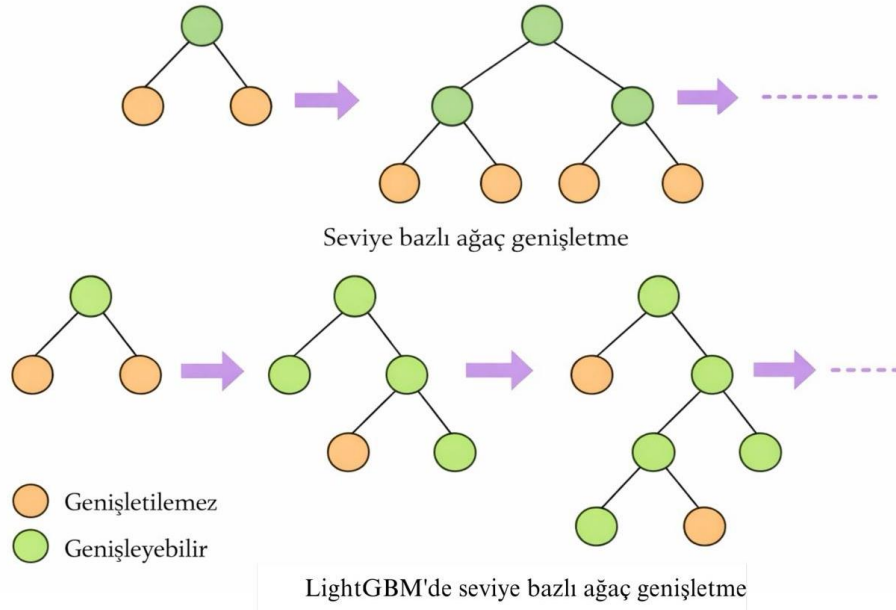


Şekil 2. Lojistik regresyonun ikili sınıflandırma modeli (Rajawat & Srivastava, 2024)

LR, basit yapısı, yüksek yorumlanabilirliği ve düşük hesaplama maliyeti nedeniyle görsel görevlerde hızlı baseline (temel model) geliştirmek ve açıklanabilir sonuçlar üretmek için etkili bir yöntem olarak kabul edilmektedir.

2.2.1.2. Hafif gradyan artırma makinesi (LightGBM)

LightGBM, Microsoft tarafından geliştirilen yüksek performanslı bir gradyan artırma algoritmasıdır. Bu algoritma özellikle büyük ölçekli sınıflandırma görevlerinde etkili olup, eğitim süresini ve bellek kullanımını verimli hale getirmektedir (Ke vd., 2017). LightGBM'nin yaprağa dayalı ağaç büyüme stratejisi, karmaşık etkileşimleri daha iyi yakalayabilmekte ve bu nedenle büyük ölçekli sınıflandırma ve regresyon görevlerinde tercih edilen bir yöntem haline gelmiştir. Şekil 3'ten görüldüğü üzere, geleneksel seviye bazlı ağaç büyüme stratejilerinin aksine LightGBM, yalnızca en yüksek bilgi kazancı sağlayan yaprağı genişleterek daha derin ve dengesiz ağaç yapıları oluşturmakta; bu da hem doğruluk hem de işlem verimliliği açısından avantaj sağlamaktadır.



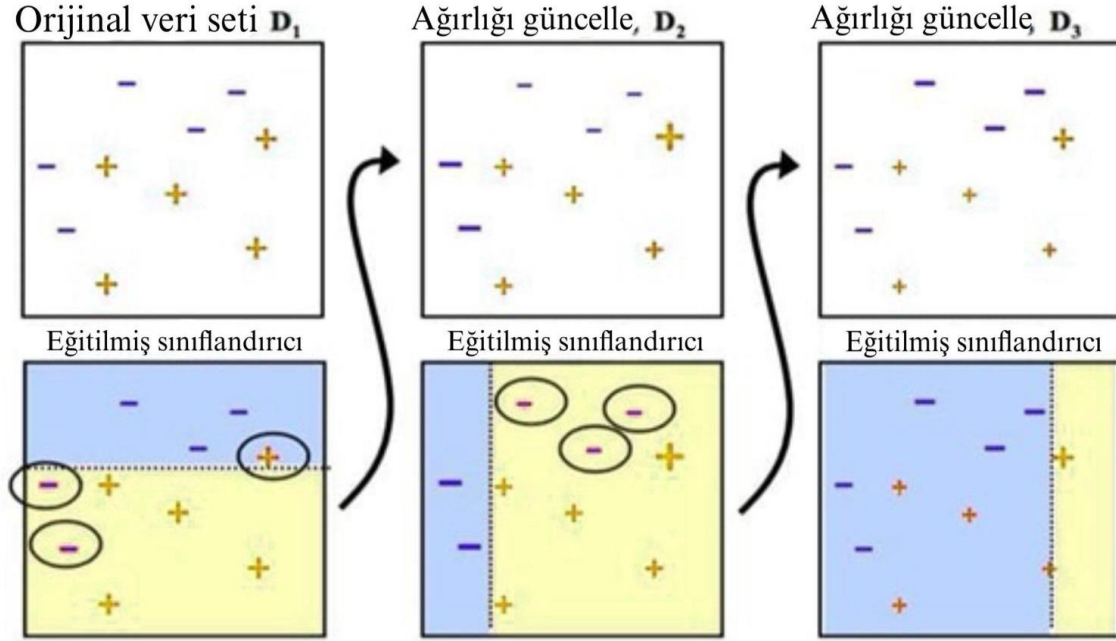
Şekil 3. Hafif gradyan artırma makinesi (Dong vd., 2022)

LightGBM, kategorik özniteliklerin ve eksik verilerin işlenmesini de desteklemektedir. Bu yetenek, özellikle yapılandırılmış meta veriler veya sensör füzyonu içeren görsel görevlerde önemli avantajlar sağlayabilir.

2.2.1.3. AdaBoost

AdaBoost, zayıf sınıflandırıcıları bir araya getirerek güçlü bir sınıflandırıcı oluşturan bir topluluk yöntemidir. Bu yaklaşımda, modelin doğru tahmin yapma yeteneğini artırmak için yanlış sınıflandırılan örnekler odaklanılır ve bu hatalar tekrar tekrar düzeltilir. Özellikle yüz tanıma gibi uygulamalarda (örneğin Viola-Jones detektörü), AdaBoost algoritmasında çok sayıda zayıf öğrenme algoritması ardışık biçimde birleştirilerek güçlü ve yüksek doğruluklu sınıflandırıcılar elde edilmektedir (Ding vd., 2022). Şekil 4.'te gösterildiği üzere AdaBoost, zayıf öğrencilerin ardışık biçimde eğitilmesiyle çalışır. Her yeni model önceki modelin hatalarına odaklanarak yanlış sınıflandırmaları düzeltir. Başlangıçta sınıflandırıcı orijinal veri setindeki örnekler eşit önem vererek eğitilir. Eğitim sonrasında yanlış sınıflandırılmış örnekler belirlenir. Sonra algoritma bu yanlış sınıflandırma yapılmış örnekler daha fazla odaklanarak bu verilerin bir sonrasındaki sınıflayıcıya daha etkin bir şekilde eğitilebilmesini sağlar. AdaBoost algoritmasında her yeni model, önceki modellerin yanlış sınıflandırdığı örnekler daha fazla ağırlık vererek yeniden eğitilmekte ve bu süreçte öğrenme performansı aşamalı olarak geliştirmektedir. Son aşamada ise daha doğru sınıflandırıcıların çıktıya daha fazla etki

ettiği bir yöntem olan “çoğunluğun oyu” prensibiyle en son sınıflandırma sonucu elde edilmektedir.

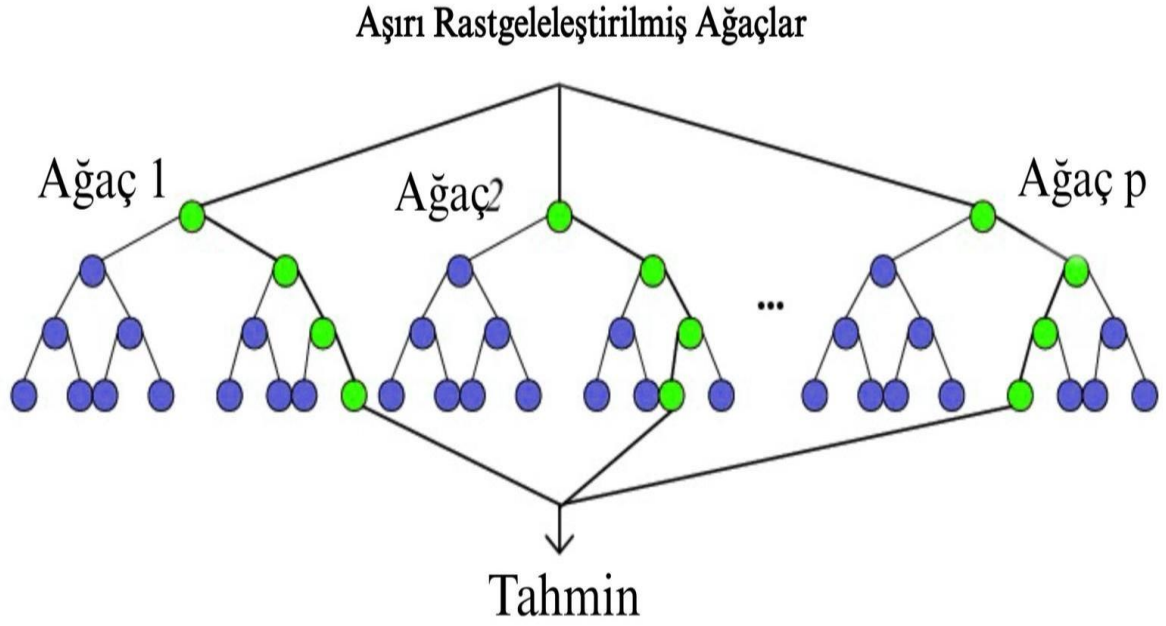


Şekil 4. AdaBoost'un çalışma prensibi (L. Wang vd., 2020)

Bu tekrarlayan yeniden ağırlıklandırma mekanizması sayesinde AdaBoost'un önyargı etkisini azaltarak ve özellikle karmaşık veriler üzerinde genel tahmin doğruluğunu artırarak tekil öğrencilerin zorlandığı konularda iyileşme sağlamaktadır.

2.2.1.4. Aşırı rastgeleştirilmiş ağaçlar (Extra Trees)

Aşırı Rastgeleştirilmiş Ağaçlar (Extra Trees) sınıflandırıcısı, rastgele orman'a (Random Forest) benzer bir topluluk öğrenmesi yöntemidir ancak bazı temel farklılıkları sahiptir. Bu yöntem, her düğümdeki kullanılacak bölme noktalarını (split point) rastgele seçmekte ve rastgele orman'dan farklı olarak tüm veri seti kullanılmaktadır. Bu yaklaşım genellikle daha hızlı bir eğitim süreci sunar ve varyansı artırır. Bu durum özellikle yüksek boyutlu görsel özniteliklerin bulunduğu görevlerde avantaj sağlayabilmektedir (Geurts vd., 2006). Şekil 5'te, Aşırı rastgeleştirilmiş ağaçlar algoritmasının temel çalışma prensibini göstermektedir.



Şekil 5. Aşırı rastgeleleştirilmiş ağaçlar sınıflandırıcısı (Akbulut vd., 2023)

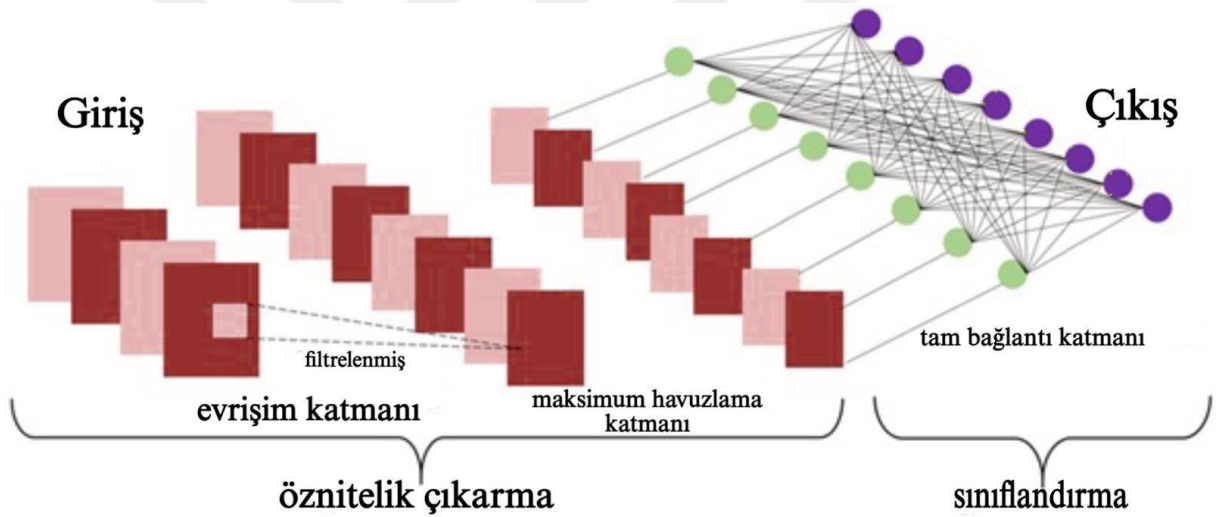
Şekil 5'ten görüldüğü üzere, Extra Trees algoritması birden fazla karar ağacını rastgele parametrelerle eğitmekte ve bu ağaçların çıktıları toplu biçimde birleştirilerek nihai tahmin elde edilmektedir. Her bir ağaç, farklı veri alt kümeleri ve bölme noktalarıyla eğitildiği için modelin çeşitliliği artmakta ve genelleme başarısı güçlenmektedir. Aşırı rastgeleleştirilmiş ağaçlar algoritması, öznelilik uzayının farklı yönlerini yakalayabilen çeşitli bir topluluk modeli oluşturma hedeflendiğinde özellikle yararlıdır. Bu yöntem, özellikle öznelilik fazlalığının yüksek olduğu ve performans ayarlamasının kritik önem taşıdığı nesne sınıflandırma sistemlerinde yaygın olarak kullanılmaktadır.

2.3. Modern Derin Öğrenmeye Dayalı Yöntemler

Derin öğrenmenin kullanımının yaygınlaşması ile bilgisayarla görme alanında büyük bir devrim yaşanmış, nesne tanıma sistemlerinin doğruluğu ve ölçeklenebilirliği önemli ölçüde artırılmıştır. Geleneksel yöntemlerde manuel olarak tasarlanan öznelilikler yerine modern derin öğrenme yaklaşımları, ilgili desenleri ham görüntü verilerinden otomatik olarak öğrenmektedir. Bu durum, çeşitli görsel görevlerde sağlamlık ve genelleme yeteneği sayesinde performansta önemli iyileştirmelere yol açmaktadır. Günümüzde evrişimli sinir ağları (CNN) ve ViT gibi mimariler en etkin yapılar arasında öne çıkmaktadır (J. Chen vd., 2024).

2.3.1. Evrişimli sinir ağları (CNN)

Görsel tanıma alanında derin öğrenme tabanlı sistemlerde öne çıkan mimariler arasında CNN önemli bir yer tutmaktadır. CNN özel olarak tasarlanmış olup, görüntü verilerini işlemek için konvolüsyon katmanlarıyla birlikte havuzlama (pooling) katmanları ve tam bağlı (fully connected) katmanlar kullanmaktadır. Bu ağlar, düşük seviyede kenar ve doku gibi özniteliklerden başlayarak derin katmanlarda nesne yamaları ve kavramsal temsiller gibi yüksek seviyeli özelliklere doğru hiyerarşik gösterimler öğrenir (Nimma & Uddagiri, 2024). Şekil 6.'da Evrişimli sinir ağları mimarisinin yapısı ayrıntılı olarak gösterilmektedir. Şekilden görüldüğü üzere, giriş görüntüsü önce evrişim ve havuzlama katmanlarından geçerek öznitelik çıkarımı gerçekleştirilmekte, ardından tam bağlantılı katmanlar aracılığıyla sınıflandırma işlemi yapılmaktadır. Bu yapı, hem düşük seviyeli hem de yüksek seviyeli özniteliklerin katmanlar boyunca aşamalı olarak öğrenilmesini sağlamaktadır.



Şekil 6. CNN mimarisi (Hashemi vd., 2023)

2.3.1.1. Evrişimsel sinir ağı katmanları

Bir evrişimsel sinir ağı mimarisi, giriş, evrişim, havuzlama, düzleştirme, tam bağlantı ve sınıflandırma olmak üzere altı temel katmandan oluşmaktadır. Bu katmanların her biri farklı işlevlere sahiptir ve birbirine ardışık biçimde bağlanarak CNN yapısını meydana getirir (Karadağ & Özdemir, 2022). Evrişim ve havuzlama katmanları özniteliklerin çıkarılmasını sağlarken; düzleştirme, tam bağlantı ve sınıflandırma katmanları ise sınıflandırma sürecini gerçekleştirmektedir. Ağın başlangıç katmanları daha çok düşük seviyeli öznitelikleri yakalarken, ilerleyen katmanlarda yüksek seviyeli özniteliklerin çıkarılması mümkün

olmaktadır. Katmanlar derinleştikçe yapılan işlemlerin karmaşıklığı artmakta ve bu sayede daha keskin, daha belirgin görüntü temsilleri elde edilmektedir (Buyukarican & Ülker, 2020).

2.3.1.1.1. Giriş katmanı

Evrişimsel sinir ağlarında ilk katman giriş katmanı olarak tanımlanmaktadır. Bu katman, ham görüntüyü alarak onu girdi vektörüne dönüştürür. Eğitim süresinin uzunluğu, doğrudan girdi verisinin boyutuna bağlıdır. Tipik bir giriş görüntüsü ise Genişlik $\times$ Uzunluk $\times$ Renk kanalı biçiminde temsil edilmektedir (Tekerek & Yapici, 2022).

2.3.1.1.2. Evrişim katmanı

Evrişimsel sinir ağlarında öznitelik çıkarımı sürecinde evrişim katmanı kritik bir rol oynamaktadır. Bu katman, girdi görüntüsü üzerinde uygulanan filtreler aracılığıyla öznitelikleri belirlemektedir. Evrişim işlemi sırasında kullanılan filtreler genellikle çok boyutlu ve piksel tabanlı değerler içerir. Girdi verisinin filtrelenmesi sonucunda daha küçük boyutlara sahip bir matris elde edilir. Görüntü üzerinde dolaşan bu matris, belirli bölgelerin tanımlanmasını sağlar (Ciresan vd., 2011). Tanımlama işlemi sonucunda ortaya çıkan yeni matris öznitelik haritası olarak adlandırılmaktadır. Çıkış matrisinin boyutu ise üç temel parametreye bağlıdır: derinlik (depth), adım (stride) ve sıfır dolgu (padding). Bu parametreler, evrişim işlemi sonucunda oluşan matrisin boyutlarını belirlemektedir. Çıkış matrisine ait boyut hesaplaması aşağıdaki formül ile yapılmaktadır.

$$\frac{n + 2p - f}{s} + 1 * \frac{n + 2p - f}{s} + 1 \quad (8)$$

Denklem 8’ de kullanılan parameterlerden; p = sıfır dolgu değerini, s = adım uzunluğunu ifade etmektedir.

7	2	3	3	8
4	5	3	8	4
3	3	2	8	4
2	8	7	2	7
5	4	4	5	4

*

1	0	-1
1	0	-1
1	0	-1

=

6		

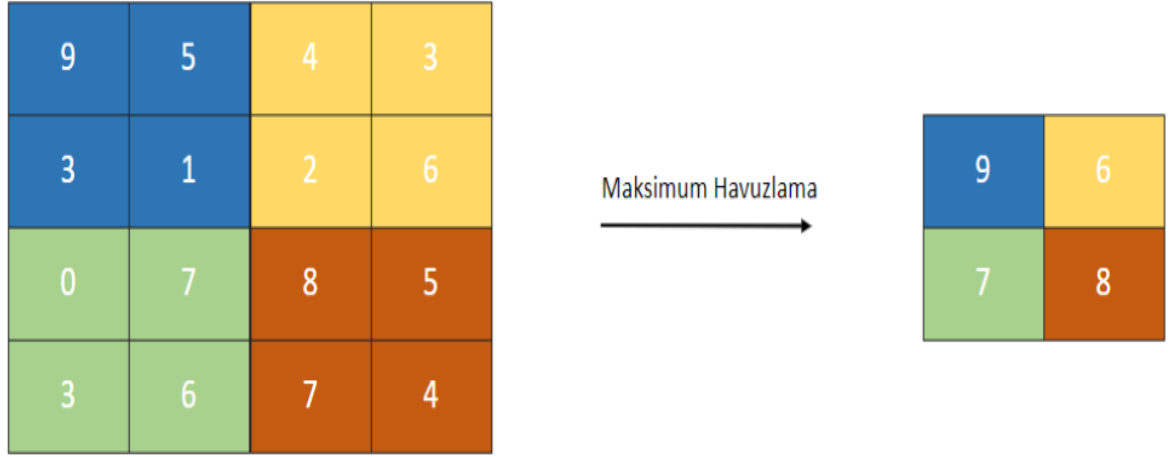
$$\begin{aligned}
&7 \times 1 + 4 \times 1 + 3 \times 1 + \\
&2 \times 0 + 5 \times 0 + 3 \times 0 + \\
&3 \times -1 + 3 \times -1 + 2 \times -1 \\
&= 6
\end{aligned}$$

Şekil 7. Filtreleme işlemine ait örnek (Dahiya vd., 2021)

Evrişimsel sinir ağlarında ilk değişken, giriş katmanına aktarılan görüntüdür. Bu görüntüden öznitelik çıkarımını sağlayan ikinci değişken ise çekirdek (kernel) olarak adlandırılır. Çekirdek, girdi görüntüsündeki değerlerle çarpma ve toplama işlemleri gerçekleştirilerek evrişim katmanını oluşturur.

2.3.1.1.3. Havuzlama katmanı

Evrişim katmanından elde edilen çıktı, havuzlama katmanına aktarılmaktadır. Bu katmanın temel amacı, sonraki evrişim katmanına hazırlık olarak görüntünün boyutlarını küçültmektir. Boyut küçültme işlemi genellikle maksimum havuzlama (max pooling) yöntemiyle gerçekleştirilir. Maksimum havuzlama, belirlenen dikdörtgen bölgeler içerisindeki en yüksek değeri seçerek görüntünün boyutlarını azaltır (Ian Goodfellow, 2016). Havuzlama katmanı sayesinde ağıdaki parametre sayısı azalır; buna bağlı olarak işlem hacmi ve hesaplama maliyeti de düşer. Bu azalma, modelin ezberleme eğilimini en aza indirerek hem hesaplama hızını artırır hem de anlamlı verilerin oranını yükseltir (Tekerek & Yapıcı, 2022). Şekil 8’de havuzlama katmanına ilişkin bir örnek sunulmuştur.



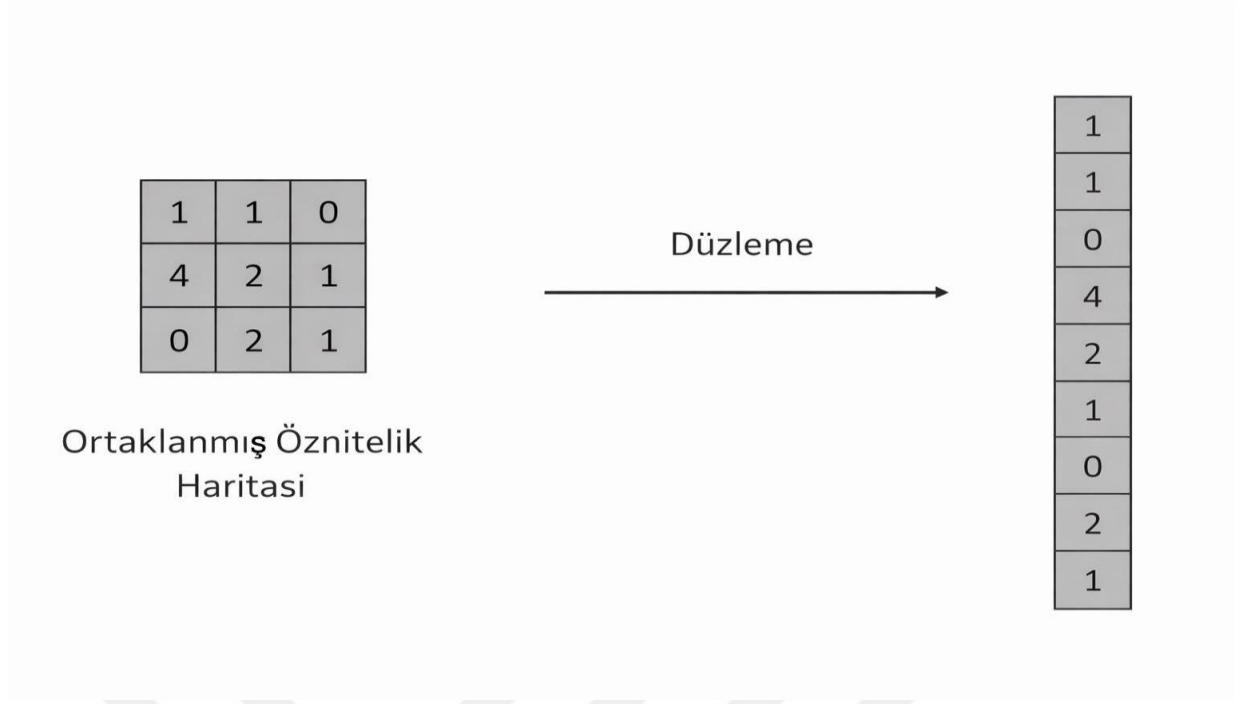
Şekil 8. Maksimum havuzlama tekniği örnek (Ajit vd., 2020)

2.3.1.1.4. Aktivasyon katmanı

Aktivasyon fonksiyonları, girdi sinyali lineer olmayan dönüşümler uygulayarak modelin daha karmaşık ilişkileri öğrenmesini sağlar. Bu dönüşüm sonucunda elde edilen çıktılar, bir sonraki katmana girdi verisi olarak aktarılır. Aktivasyon fonksiyonları, girdi değerleri ile ağırlık ve sapma parametrelerini birleştirerek çıktının hesaplanmasına olanak tanır. Hesaplanan bu değer, verinin sonraki aşamalara iletilip ileilmeyeceğini belirler (Rachapudi & Lavanya Devi, 2021). Lineer olmayan fonksiyonların kullanımı, modelin karmaşıklığını ve öğrenme kapasitesini artırmaktadır. Aktivasyon fonksiyonları genellikle gizli katmanlarda ve çıktı katmanında kullanılmakta olup, seçilecek fonksiyon türü gerçekleştirilecek tahmin yöntemine bağlıdır.

2.3.1.1.5. Düzleştirme katmanı

gerçekleştirilen işlemlerden elde edilen çıktılar, ağırlık derinliğini artırmak amacıyla tekrar evrişim ve havuzlama süreçlerine tabi tutulabilmektedir. Bu yaklaşım sayesinde çok sayıda ardışık katman oluşturulmakta ve bu katmanlar, çok boyutlu görüntüler üzerinde işlem yaparak yeni özneliklerin çıkarılmasına olanak tanımaktadır (Ian Goodfellow, 2016). Şekil 9.'da düzleştirme katmanına ilişkin bir örnek sunulmuştur.



Şekil 9. Düzleştirme katmanı örnek (Apoorva vd., 2021)

2.3.1.1.6. Tam bağlı katman

Evrişimsel sinir ağlarında diğer katmanlar doğrudan birbirine bağlı değildir; ancak tam bağlı katman, kendisinden önceki katmana doğrudan bağlanmaktadır. Bu katmanda, evrişim ve havuzlama katmanlarında çıkarılan öznitelikler kullanılarak sınıflandırma işlemi gerçekleştirilir. Ayrıca, sınıflandırma sürecinde her sınıfa ait olasılık değerleri hesaplanarak çıktı elde edilir. Evrişim ve havuzlama katmanlarında.

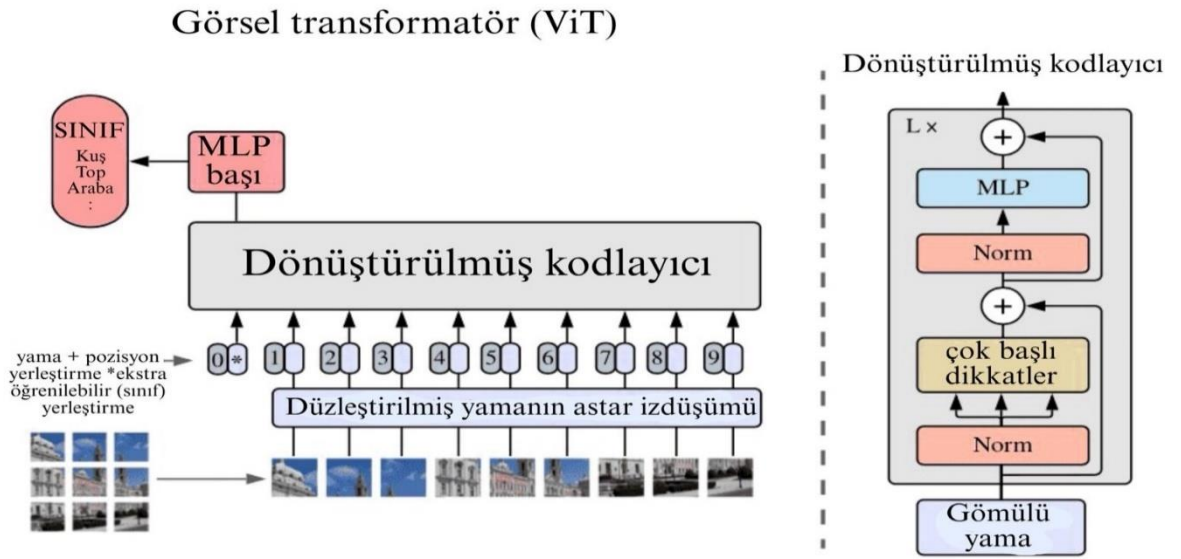
2.3.1.1.7. Sınıflandırma katmanı

Sınıflandırma katmanı, işlenen veri setindeki sınıf sayısı kadar çıktı üretmektedir. Bu çıktılar, farklı sınıflara karşılık gelmektedir. Ağın son katmanı olan sınıflandırma katmanında, verilerin sınıflara ayrılması için genellikle Softmax fonksiyonu kullanılmaktadır. Softmax, elde edilen çıktılar üzerinde olasılık hesaplamaları yaparak her sınıfa ait olasılık değerini belirler. Böylece sınıflandırma katmanının çıktısı ile sınıf sayısı kesin olarak tanımlanmış olur.

2.3.2. Görsel transformatörler (ViT)

Görsel transformatörler (ViT) modeli, derin öğrenme alanında yeni bir yaklaşım sunmakta olup bilgisayarlı görmede transformatör mimarisinden esinlenerek geliştirilmiştir (Dosovitskiy vd.,

2021; Palanisamy vd., 2025). ViT mimarisinde görüntüler sabit boyutlu yamalara ayrılır, bu yamalar düzleştirilerek bir diziye yerleştirilir ve konvolüsyon yerine öz-dikkat (self-attention) mekanizmaları kullanılarak işlenir. Sürecin ilk adımı giriş görüntüsünün sabit boyutlu yamalara bölünmesidir; ardından yamalar düzleştirilir, konumsal gömmeler (Positional Embeddings) eklenir ve doğrusal projeksiyon katmanından geçirilerek sınıflandırma belirteci (class token) eklenir. Elde edilen dizi, öz-dikkat mekanizmalarını kullanan bir transformer mimarisi kodlayıcısına aktarılır ve yamalar arasındaki ilişkiler modellenir. Son aşamada, elde edilen çıktı sınıfa özgü bir belirleyici ile alınır ve sınıflandırma tahmini için çok katmanlı bir algılayıcı (MLP) başlığına iletilir. Bu sürecin genel akışı Şekil 10'da gösterilmektedir. Şekilden görüldüğü üzere, giriş görüntüsü sabit boyutlu yamalara bölünmekte, bu yamalar konumsal yerleştirme ile birlikte düzleştirilerek transformer kodlayıcısına aktarılmakta ve çok başlı dikkat mekanizmaları ile işlenmektedir. Kodlayıcıdan elde edilen çıktı, MLP başlığı aracılığıyla sınıflandırma sonucuna dönüştürülmektedir.



Şekil 10. Görsel transformatörler (ViT) mimarisi (Dosovitskiy vd., 2021)

ViT, Tespite Dayalı Dönüştürücü (DETR) ve Kaydırmalı Pencere Dönüştürücü gibi modeller, büyük ölçekli veri setleri üzerinde eğitildiklerinde üstün performans sergileyebilmektedirler. Ayrıca bu modeller, çok modlu (multimodal) ve yapısal girdilerle çalışabilme yetenekleri sayesinde özdenetimli öğrenme (self-supervised learning) yaklaşımlarında daha yüksek esneklik ve uyarlanabilirlik sunmaktadır.

2.3.3. VGG16

Darin öğrenme mimarilerin araştırılmasıyla ortaya çıkan VGG16, (Simonyan & Zisserman, 2015) tarafından geliştirilmiştir, 3×3 boyutunda evrişimsel çekirdekler kullanan 16 katmandan oluşan derin bir evrişimsel sinir ağıdır. Küçük çekirdek boyutlarının tutarlı biçimde uygulanması, modelin farklı ölçeklerde karmaşık öznitelikleri yakalamasına ve görüntülerden ayrıntılı desenler çıkarmasına olanak tanımaktadır (Jiang vd., 2021). Ağ, $224 \times 224 \times 3$ boyutlarındaki giriş görüntülerini işleyerek evrişim ve havuzlama katmanlarının yapılandırılmış bir ilerleyişini takip etmekte, ardından ReLU aktivasyon fonksiyonlarıyla donatılmış tam bağlı katmanlara geçmektedir. Evrişimsel katmanlarda kullanılan filtre sayısı kademeli olarak artırılmakta, bu sayede hiyerarşik ve ayrıntılı öznitelik çıkarımı gerçekleştirilmektedir. Şekil 11’de VGG16 mimarisine ilişkin genel yapı sunulmaktadır. Şekilden görüldüğü üzere, ağın ilk katmanlarında düşük seviyeli öznitelikler yakalanırken, derin katmanlarda daha soyut ve kavramsal temsiller oluşturulmakta; son aşamada ise tam bağlı katmanlar aracılığıyla sınıflandırma işlemi gerçekleştirilmektedir.

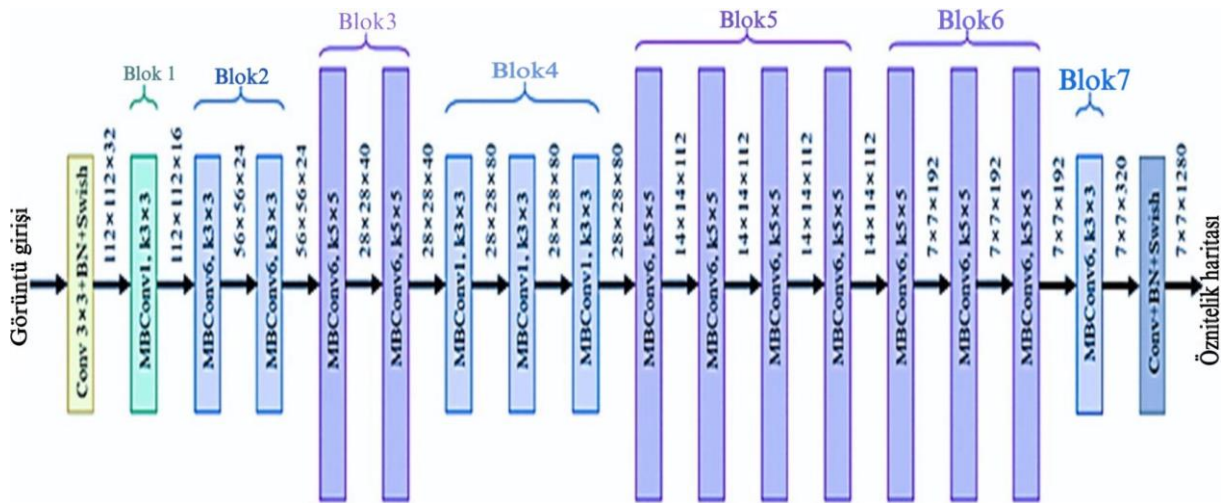


Şekil 11. VGG16 mimarisi (Rangari vd., 2023)

2.3.4. EfficientNetB0

EfficientNet mimarisi, derin sinir ağlarının ölçeklendirilmesinde bileşik ölçekleme yöntemini kullanır. Bu yöntem, ağın giriş çözünürlüğünü, genişliğini ve derinliğini doğru bir şekilde artırarak doğruluk ve hesaplama maliyeti arasında ideal bir denge sağlar (Tan & Le, 2020). Bu ailenin temel modeli olan EfficientNetB0, B1'den B7'ye kadar olan diğer varyantların

geliştirilmesi için temeldir. EfficientNetB0 mimarisinde giriş görüntüsünü işleyen ilk bölüm stem katmanı olarak adlandırılmaktadır. Bu katman, adım uzunluğu (stride) 2 olarak tanımlanmış bir evrimsel katman (Conv2D) ile başlamaktadır. Ardından; yığın Normalizasyon ve Swish aktivasyon fonksiyonu kullanılmaktadır. Giriş görüntüsü bu yapı sayesinde temel kenar, köşe ve doku gibi düşük seviyeli özniteliklere ayrıştırılır. Toplu Normalizasyon (Batch Normalization), her mini-batch üzerinde aktivasyonları normalleştirerek eğitim sürecini hızlandırır ve daha istikrarlı hale gelir. Doğrusal olmayan, yumuşak geçişli yapısıyla Swish aktivasyonu, modelin daha karmaşık ilişkileri öğrenmesine yardımcı olur. MBConv blokları, EfficientNetB0'un ana bileşenleridir. Bu bloklar, daraltma-genişletme (bottleneck) yapısı ve Sıkıştırma-Uyarma Katmanı (SE) mekanizması ile birlikte çalışarak parametre verimliliğini artırmakta ve kanal bazlı dikkat (channel attention) sayesinde önemli özelliklere odaklanmaktadır. Modelin giriş boyutu $224 \times 224 \times 3$ 'tür. Çıkış katmanında ise tam bağlantılı katman sınıflandırmayı gerçekleştirir. EfficientNetB0, önceden eğitilmiş ağırlıkların kullanılabilmesi için transfer öğrenme uygulamalarında tercih edilmektedir. EfficientNetB0 mimarisi Şekil 12'de şematik olarak gösterilmiştir. Şekilden görüldüğü üzere, model giriş görüntüsünü Conv + BN + Swish yapısıyla işlerken, ardından MBConv blokları aracılığıyla öznitelik haritasını kademeli olarak daraltmakta ve derinleştirmektedir. Her blokta kullanılan farklı çekirdek boyutları ve kanal sayıları, modelin hem düşük hem de yüksek seviyeli öznitelikleri etkili biçimde öğrenmesini sağlamaktadır. Bu yapı, transfer öğrenme uygulamalarında yüksek doğruluk ve parametre verimliliği sunarak yaygın biçimde tercih edilmektedir.



Şekil 12. EfficientNetB0 mimarisinin genel yapısı (Yang vd., 2021)

2.4. Literatür Taraması

Görüntü tanıma, bilgisayar görüşünde önemli bir araştırma konusudur ve son yıllarda evrişimli yapay sinir ağları (Convolutional Neural Network-CNN) modellerinin yaygın kullanımıyla yeni bir boyut kazanmıştır (Karadağ & Özdemir, 2022; Keerthana vd., 2023; Telçeken, 2022). CNN'ler, geleneksel makine öğrenimi yöntemlerine kıyasla görüntü tanıma sorunlarında üstün performans göstermiştir (R. Zhang vd., 2023). Örneğin, Krizhevsky ve arkadaşları tarafından yapılan AlexNet çalışması (Krizhevsky vd., 2012), ImageNet veri setinde üstün performanslı sonuçlarla derin öğrenmenin gücünü kanıtlamıştır.

Görsel transformatörler (ViT) modeli, görsel verilerdeki korelasyonları anlamak için öz dikkat mekanizmasını kullanan yeni bir yaklaşımdır. Dosovitskiy ve arkadaşları (Dosovitskiy vd., 2021) ViT modelini tanıttıklarında, bu modelin yüksek performans elde etmek için büyük ölçekli veri setleri ve yüksek hesaplama gücü gerektirdiğini belirtmişlerdir. ViT, özellikle görüntü tanıma görevlerinde başarılı olmuştur, ancak yüksek hesaplama maliyetleri uygulama alanlarını sınırlamaktadır.

Daha hesaplanabilir ve verimli modeller geliştirmek için, ViT ve diğer makine öğrenimi yöntemleriyle birleştiren hibrit yaklaşımlar önerilmiştir. Örneğin, Touvron ve arkadaşları (Touvron vd., 2021) tarafından geliştirilen Veri Verimli Görüntü Dönüştürücüler (Data-efficient Image Transformer-DeiT) modeli, ViT'in eğitim sürecini hızlandırmak ve daha az veriyle yüksek performans elde etmek için saflaştırma (distillation) yöntemini kullanır.

Aşırı rastgeleleştirilmiş ağaçlar sınıflandırıcılarını kullanarak çeşitli nesne türlerini tanımlamaya yönelik bir teknik Bosch ve arkadaşları tarafından önerilmiştir (Bosch vd., 2007). Caltech-101 ve Caltech-256 veri setleri üzerinde yaptıkları çalışmada, ROI içeren yaklaşımlarının performansı %5 oranında artırdığını göstermişlerdir.

Liu ve arkadaşları (Liu vd., 2017) tarafından yürütülen çalışmada, renk histogramının Yerel İkili Desenler (Local Binary Patterns-LBP) ile birleştirilmesiyle daha başarılı bir sistem önerilmiştir. LBP özelliği, görüntülerin renk bilgisini etkili bir şekilde kullanmadığından, renk özelliklerinden yararlanmak için Renk Bilgi Özelliği (Computational Image Formation-CIF) ile birleştirilmiştir. Deneysel sonuçlar, bu hibrit özelliğin daha etkili olduğunu göstermiştir.

Bansal ve arkadaşları (Bansal, Kumar, & Kumar, 2021) tarafından gerçekleştirilen çalışmada, nesne tanıma uygulamalarında SIFT, SURF ve ORB özellik tanımlayıcılarını kullanmış ve bu üç yöntemin performansını karşılaştırmalı olarak analiz etmişlerdir. Ölçekten Bağımsız Öznitelik Dönüşümü (SIFT), Hızlandırılmış Sağlam Öznitelikler (SURF) ve Yönlendirilmiş FAST ve Döndürülmüş BRIEF (ORB) özelliklerinin performansını ayrı ayrı ve bu

tanımlayıcıların farklı şekillerde birleştirilmesiyle elde edilen hibrit özelliklerin performansını incelemişlerdir. SIFT, SURF ve ORB özelliklerini birleştiren yöntemin en iyi performansı sağladığını bildirmişlerdir.

Anandhasilambarasan ve arkadaşları (Anandhasilambarasan vd., 2024) tarafından gerçekleştirilen çalışmada, bilgisayarlı görme alanında sıkça kullanılan iki öznelik çıkarım yöntemi olan SIFT ve SURF yöntemlerini karşılaştırmıştır. Çalışmada doğruluk ve hız arasındaki performans dengesi incelenmiş; deneysel bulgular SIFT yönteminin %95 doğruluk oranı ile daha yüksek tanıma başarısı sunduğunu, SURF yönteminin ise biraz daha düşük doğruluk oranına (%90) sahip olmasına rağmen önemli ölçüde daha hızlı çalıştığını göstermiştir. Bu sonuçlar, SIFT'in hassasiyet gerektiren görevler için daha uygun olduğunu, SURF'ün ise gerçek zamanlı uygulamalarla daha uyumlu bir yapı sunduğunu ortaya koymaktadır. Bununla birlikte, araştırmanın dar kapsamlı görüntü eşleştirme veri setleri üzerinde gerçekleştirilmiş olması, bu tekniklerin farklı ve gürültülü ortamlardaki performansı hakkında belirsizlik yaratmaktadır.

Naseer ve arkadaşları (Aysha Naseer Nouf Abdullah Almujaally, 2024) tarafından gerçekleştirilen çalışmada, Yapay Sinir Ağları (YSA) ile birlikte bir segmentasyon yaklaşımı kullanmışlardır. İlk olarak, giriş görüntülerini RGB renk uzayını kullanarak segmentlere ayırmışlardır. Daha sonra bu segmentlere ayrılmış görüntüler YSA'lar kullanılarak sınıflandırılmış ve MSRC, MS COCO ve Caltech-101 veri kümelerinde sırasıyla %89, %83 ve %90 doğruluk oranları elde edilmiştir.

El-Kenawy ve arkadaşları (El-Kenawy vd., 2020) tarafından gerçekleştirilen çalışmada, COVID-19 vakalarının CT görüntülerinden sınıflandırılmasına yönelik olarak oylama tabanlı sınıflayıcılarla entegre edilmiş bir öznelik seçimi stratejisi önerilmiştir. Çalışmada sunulan bu karma yaklaşım, modelin karmaşıklığını azaltırken %96 doğruluk oranına ulaşarak yüksek bir sınıflandırma performansı sergilemiştir. Birleşik stratejinin hem performansı hem de modelin anlaşılabilirliğini artırdığı belirtilmiş ve bu yönüyle yüksek riskli medikal görüntüleme uygulamalarında etkili bir yöntem olduğu ortaya konmuştur. Bununla birlikte, araştırmanın yalnızca tek bir uygulama alanı üzerinde değerlendirilmiş olması ve farklı veri setleri ya da medikal durumlarla çapraz doğrulama yapılmamış olması, modelin genellenebilirlik kapasitesi hakkında belirsizlik yaratmaktadır.

Hussein ve arkadaşları (Yemi Hussain, 2024) tarafından gerçekleştirilen çalışmada, nesne tanıma için üç adım önerilmiştir. İlk aşama, Caltech-101 veri setindeki dengesizlik sorununu ele almak için veri artırmayı içerir. İkinci aşama, Inception V3 derin öğrenme, Piramit HOG (PHOG) ve Merkezi Simetrik LBP (CS-LBP) özelliklerini birleştirir. Son adım, optimum

özellikleri belirlemek için Ortak Entropi ve KNN (Joint Evidential K-Nearest Neighbor-JEKNN) tekniğini kullanır. Yöntemleri, Caltech-101 veri setinde %90,4'lük bir doğruluk oranına ulaşmıştır.

Kate ve arkadaşları (Kate vd., 2023) tarafından gerçekleştirilen çalışmada, görüntü sınıflandırması için CNN, SVM ve diğer geleneksel makine öğrenimi modellerini karşılaştırmalı olarak incelemiştir. Çalışmada derin sahtecilik görüntülerinden oluşturulan hibrit veri setlerinin yanı sıra SoyNet, Soybean Leaf Disease Prediction ve Roboflow gibi tarımsal görüntü veri setleri kullanılmıştır. Deneysel sonuçlar, %94 doğruluk oranı ile CNN modelinin en yüksek performansı sunduğunu, SVM'nin ise %89 doğruluk oranı ile ikinci sırada yer aldığını göstermektedir. Bulgular, CNN'lerin karmaşık görsel verilerle etkin bir şekilde baş edebildiğini ortaya koymaktadır. Bununla birlikte, çalışmanın doğruluk performansı ile modelin yorumlanabilirliği arasındaki dengeyi ayrıntılı biçimde ele almaması eleştirilmiş; bu durum, özellikle kritik uygulamalarda özniteliklerin önemini ve performans-şeffaflık dengesi üzerine daha kapsamlı analizlerin gerekliliğini vurgulamaktadır.

Nandeshwar ve arkadaşları (Nandeshwar vd., 2025) tarafından gerçekleştirilen çalışmada, geleneksel görüntü tanımlayıcılarıyla birleştirilmiş CNN'lerden oluşan hibrit bir gerçek zamanlı algılama modeli geliştirmiştir. Önerilen model, %93 doğruluk oranına ulaşarak hem derin öğrenme tabanlı soyut temsillerden hem de geleneksel tanımlayıcıların sağladığı ayrıntılı yerel bilgilerden yararlanmıştır. Bu bütünlük yapısı, özellikle gerçek zamanlı izleme ve robotik uygulamalarda hızdan ödün vermeden daha yüksek tespit performansı sunmuştur. Bununla birlikte, modelin artan karmaşıklığı ve yüksek bellek gereksinimi, gömülü veya mobil sistemlerde uygulanabilirliği sınırlandıran önemli bir dezavantaj olarak belirtilmiştir. Ayrıca çalışmada model budama veya optimizasyon tekniklerinin kullanılmamış olması, kaynak kısıtlı aygıtlarda hibrit yöntemlerin pratik uygulanabilirliği açısından bir eksiklik olarak değerlendirilmektedir.

Zhang ve Pu (L. Zhang & Pu, 2024) 2024 yılında yapmış oldukları çalışmada, nesne tanıma görevleri için Hu momentlerini ve Hızlandırılmış Sağlam Öznitelikler (Speeded-Up Robust Features-SURF) yöntemlerini birleştiren hibrit bir model geliştirmiştir. İki farklı yöntem, çeşitli ağırlıklandırma faktörleri kullanılarak etkili bir şekilde birleştirilmiştir. Önerdikleri yöntemin ölçek, bakış açısı, aydınlatma ve gürültüdeki değişikliklere dayanıklı olduğu ve diğer temsili yöntemlerden daha etkili ve sağlam olduğu gösterilmiştir.

Paul ve arkadaşları (Paul vd., 2021) tarafından gerçekleştirilen çalışmada, kilit nokta belirleyicilerin performansını değerlendirerek bu yöntemleri sağlamlık ve verimlilik açısından karşılaştırılmıştır. Çalışmada Yönlendirilmiş FAST ve Döndürülmüş BRIEF (ORB)

yönteminin %87 başarı oranı ile hız ve dayanıklılık bakımından öne çıktığı belirtilmiştir. ORB'nin, özniteliklerin hızlı biçimde değişebildiği gerçek zamanlı sistemler için uygun bir yaklaşım sunduğu gösterilmiştir. Bununla birlikte, yöntemin daha yeni derin öğrenme tabanlı kilit nokta belirleyicilerle karşılaştırılmamış olması, bulguların güncel araştırmalar açısından genellenebilirliğini sınırlamaktadır. Bu nedenle gelecekte ORB'nin performansının sinir ağı tabanlı yöntemlerle gerçek dünya koşullarında karşılaştırmalı olarak değerlendirilmesi önerilmektedir.

Salari ve arkadaşları (Salari vd., 2022) tarafından gerçekleştirilen çalışmada, nesne tanıma verilerini eleştirel bir bakış açısıyla inceleyerek model performansını önemli ölçüde etkileyebilecek yapısal önyargıların varlığını ortaya koymuştur. Araştırma bulguları, özellikle temsil edilmeyen sınıflar veya değişken çevresel koşullar altında veri setindeki önyargıların doğruluk oranında %15'e kadar düşüşe neden olabileceğini göstermektedir. Çalışma, yapay zekâ sistemlerinde eşitlik, genelleme ve adil değerlendirme konularına önemli bir katkı sunmakta; ancak önyargıyı azaltmaya yönelik tekniklerin deneysel olarak test edilmemiş olması, önerilen çıkarımların pratik uygulamalara aktarılabilirliğini sınırlamaktadır. Bu durum, gelecekte veri seti önyargılarının giderilmesine yönelik yöntemlerin sistematik biçimde değerlendirilmesi gerektiğini vurgulamaktadır.

Suresan ve arkadaşları (Suresan vd., 2021) tarafından gerçekleştirilen çalışmada, otomatik plaka tanıma amacıyla çeşitli makine öğrenimi algoritmalarını değerlendirerek bu yöntemlerin performansını karşılaştırılmıştır. Çalışmada CNN modelinin %94 sınıflandırma doğruluğu ile en etkili yöntem olduğu belirlenmiş; CNN'lerin düşük ışık, hareket bulanıklığı ve değişken çevresel koşullar altında dayanıklılık gösterdiği ortaya konmuştur. Araştırma, CNN tabanlı yaklaşımların akıllı ulaşım sistemleri ve otomatik trafik denetim uygulamaları için uygun olduğunu vurgulamaktadır. Bununla birlikte, çalışmanın donanım kısıtlamaları ve gerçek dünya dağıtımlarındaki zorlukları dikkate almamış olması, CNN tabanlı çözümlerin gerçek zamanlı yol ortamlarındaki uygulanabilirliği konusunda önemli bir sınırlılık oluşturmaktadır.

Wang ve arkadaşları (C. Wang vd., 2020) tarafından gerçekleştirilen çalışmada, yüz tanıma algoritmalarını sinir ağları (NN), destek vektör makineleri (SVM) ve en yakın komşular (KNN) kullanarak karşılaştırılmıştır. Çalışmada sinir ağlarının %94 doğruluk oranı ile en yüksek performansı sunduğu; SVM'in %90 ve KNN'in ise %86 doğruluk oranı sağladığı belirlenmiştir. Bulgular, sinir ağlarının karmaşık ve büyük boyutlu yüz verilerini işleme konusunda belirgin bir üstünlük sergilediğini göstermektedir. Bununla birlikte, araştırmada demografik performans farklılıkları ve algoritmik önyargıların ele alınmamış olması önemli bir sınırlılık olarak değerlendirilmektedir. Özellikle etnik gruplar veya yaş grupları arasında gerçekleştirilen

testlerin eksikliği, adalet ve eşitlik tartışmalarının yoğunlaştığı yüz tanıma sistemlerinde değerlendirme kapsamını daraltmakta ve yöntemin etik ile hukuki boyutlarına ilişkin belirsizlik yaratmaktadır.

Yukarıda özetlenen çalışmaların çoğu, değerlendirmelerinde geleneksel modelleri kapsamamakta ve dengesizlik problemini çözmek için çok aşamalı yöntemler önermektedir. Bu sorun çeşitli topluluk (ensemble) yöntemleriyle ele alınabilse de, daha sağlam öznitelikler de çözüm sağlayabilmektedir.

Bu bağlamda önerilen ViT ve lojistik regresyon tabanlı yöntem, söz konusu problemi aşmakta ve mevcut en güncel yöntemlerden daha iyi performans göstererek nesne tanıma ve nesne sınıflandırma alanında literatüre önemli bir katkı sunmaktadır. Tablo 1’de literatürde yer alan çalışmaların yöntemleri, elde edilen sonuçları ve performans ölçütleri karşılaştırmalı olarak özetlenmiştir.

Tablo 1. Literatür özeti

Çalışma	Yöntem	Öznitelik	Performans Metrikleri
Karadağ & Özdemir (2022); Keerthana vd. (2023); Telçeken & Kutlu (t.y.)	CNN modelleri	Nesne tanıma	Accuracy: %93, Precision: %91, Recall: %90, F1: %90
R. Zhang vd. (2023)	CNN vs geleneksel ML	CNN	Accuracy: %92, F1: %89
Krizhevsky vd. (2017)	AlexNet – ImageNet	Derin öğrenme	Top-5 Accuracy: %84.7
Dosovitskiy vd. (2021)	Vision Transformer (ViT)	Transformer tabanlı	Accuracy: %88, Precision: %87, Recall: %86
Touvron vd. (2021)	DeiT	Data-efficient ViT	Accuracy: %85, F1: %83
Bosch vd. (2007)	ExtraTrees	Caltech-101/256	Accuracy: %78
Liu vd. (2017)	LBP + CIF hibrit öznitelikleri	Hibrit öznitelikler	Accuracy: %82, Precision: %80

Çalışma	Yöntem	Öznitelik	Performans Metrikleri
Bansal vd. (2023)	SIFT, SURF, ORB karşılaştırması	Hibrit kullanım	Accuracy: %89, Recall: %87
Anandhasilambarasan vd. (2024)	SIFT vs SURF	Karşılaştırma	SIFT Accuracy: %95, SURF Accuracy: %90
Naseer vd. (2024)	RGB bölütleme + YSA	Çeşitli veri kümeleri	MSRC Accuracy: %89, MS COCO Accuracy: %83, Caltech-101 Accuracy: %90
El-Kenawy vd. (2020)	Oylama tabanlı sınıflayıcı	COVID-19 CT	Accuracy: %96
Yemi Hussain (2024)	Inception V3 + PHOG + CS-LBP + JEKNN	Çok aşamalı yöntem	Accuracy: %90.4, F1: %88
Kate vd. (2023)	CNN vs SVM	Karşılaştırma	CNN Accuracy: %94, SVM Accuracy: %89
Nandeshwar vd. (2025)	Hibrit CNN + geleneksel tanımlayıcılar	Gerçek zamanlı izleme	Accuracy: %93
L. Zhang & Pu (2024)	Hu momentleri + SURF hibrit model	Hibrit öznitelikler	Accuracy: %88
Paul vd. (2021)	ORB kilit nokta belirleyici	Gerçek zaman sistemleri	Accuracy: %87
Salari vd. (2022)	Veri seti önyargı analizi	Yapısal önyargılar	Accuracy düşüşü: %15
Suresan vd. (2021)	CNN ve diğer ML algoritmaları (plaka tanıma)	CNN	Accuracy: %94
Wang vd. (2020)	NN, SVM, KNN karşılaştırması	Karşılaştırma	NN Accuracy: %94, SVM Accuracy: %90, KNN Accuracy: %86

Tablo 1. (Devamı)

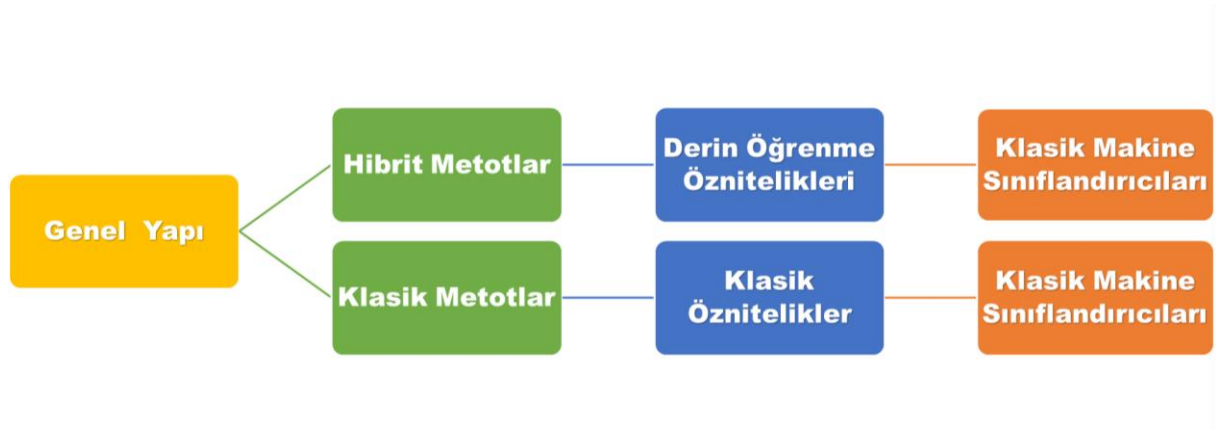
Tablo 1’de sunulan literatür özeti, nesne tanıma alanında kullanılan yöntemlerin çeşitliliğini ve performans farklılıklarını açıkça ortaya koymaktadır. Görüldüğü üzere, derin öğrenme tabanlı modeller özellikle doğruluk açısından güçlü sonuçlar üretirken, geleneksel yöntemler belirli koşullarda daha düşük maliyetli ve daha hızlı çözümler sunabilmektedir. Bununla birlikte, birçok çalışma ya yalnızca tek bir öznitelik çıkarma yaklaşımına odaklanmış ya da dengesizlik

problemini çözmek için çok aşamalı ve karmaşık yöntemler önermiştir. Ayrıca ViT, CNN ve geleneksel öznitelik çıkarma yöntemlerini doğrudan karşılaştıran kapsamlı çalışmaların sınırlı olduğu dikkat çekmektedir. Bu noktada, önerilen ViT–lojistik regresyon tabanlı yaklaşım, hem literatürdeki bu boşluğu doldurmakta hem de farklı yöntemlerin güçlü yönlerini bir araya getirerek daha dengeli ve yüksek performanslı bir çözüm sunmaktadır.



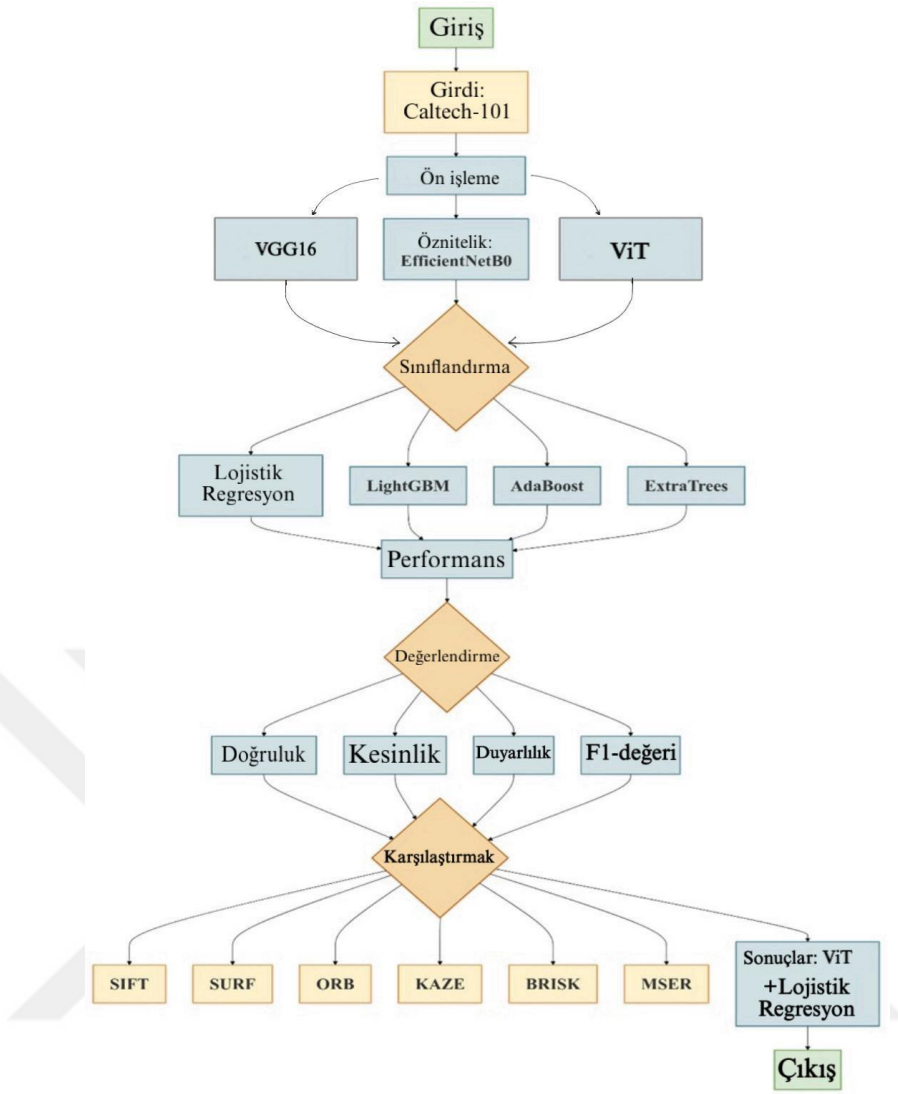
3. YÖNTEM

Bu tez çalışması kapsamında, sınıf dengesizliği problemi içeren Caltech-101 veri seti üzerinde nesne tanıma başarımını artırmaya yönelik hibrit bir sınıflandırma yaklaşımı geliştirilmiştir. Önerilen yöntemde içerik temelli öznitelik temsil gücüyle öne çıkan Görsel transformatörler (ViT) modeli, klasik makine öğrenmesi algoritmalarının istikrarlı sınıflandırma performansı ile birleştirilerek hibrit bir sistem tasarlanmıştır. Çalışmanın temel amacı, veri setisindeki sınıf dağılımı dengesizliğinden kaynaklanan sınırlamaları en aza indirerek, farklı öznitelik çıkarma ve sınıflandırma stratejilerinin başarımını artırmayı hedeflemektir. Önerilen hibrit sistemin genel yapısı Şekil 13.'te şematik olarak sunulmaktadır.



Şekil 13. Sistemin genel yapısı.

Geliştirilen sistem kapsamında öznitelik çıkarma sürecinde, geleneksel yöntemler olan SIFT, SURF, ORB, KAZE, BRISK ve MSER algoritmalarının yanı sıra, ViT, VGG16 ve EfficientNetB0 gibi önceden eğitilmiş derin öğrenme tabanlı modellerden yararlanılmıştır. Elde edilen tüm öznitelikler, LightGBM, AdaBoost, Extra Trees ve Logistic regression sınıflandırıcıları ile değerlendirilmiştir. Sisteme giriş olarak verilen görüntüler, ilgili öznitelik çıkarma yöntemi doğrultusunda işlenmiş; sabit boyutlu öznitelik vektörleri elde edilmiştir. Bu öznitelikler kullanılarak sınıflandırıcılar eğitilmiş ve eğitim sürecinin tamamlanmasının ardından test verilerine ait öznitelikler çıkarılarak sınıflandırma işlemi gerçekleştirilmiştir. Bu kapsamda kullanılan yöntemler ve sınıflandırıcılar aşağıda özetlenmiş olup, tüm deneysel sürecin genel yapısı ve uygulama adımları Şekil 14.'de görsel olarak sunulmaktadır.



Şekil 14. Deneysel Çalışmalara Ait Akış Diyagramı.(Aydın & Ali, 2025)

Şekil 14.'de görüldüğü üzere, tez kapsamında kullanılan yöntemler iki alt başlık altında toplanmıştır. İlk başlık altında, klasik öznitelik çıkarma algoritmaları olan SIFT, SURF, ORB, KAZE, BRISK ve MSER yer almakta olup bu öznitelikler Görsel Kelimeler Torbası (Bag of Visual Words, BoVW) yöntemi ile sabit boyutlu hale getirilmiştir. (BoVW) yöntemi, başlangıçta metin sınıflandırma amacıyla geliştirilmiş olup daha sonra görsel özniteliklerin temsilinde de yaygın olarak kullanılmaya başlanmıştır. (BoVW) yöntemi iki temel adımdan oluşmaktadır: kodlama ve havuzlama. Kodlama aşamasında öznitelik vektörleri üzerinde kümeleme algoritmaları (ör. k-ortalamlar) çalıştırılarak, görsel kelimelerden oluşan bir sözlük elde edilmektedir. Havuzlama aşamasında ise her görüntü, bu görsel kelimelerin sıklık dağılımına göre sabit boyutlu bir öznitelik vektörü ile temsil edilmektedir. Bu yöntem nesnelerin konum ve yön değişimlerine karşı dayanıklı olmakla birlikte öznitelikler arasında

konumsal bilginin bulunmaması temel dezavantajdır (Lazebnik vd., 2006). Elde edilen sabit boyutlu öznitelik vektörler, LR, ExtraTrees, LightGBM ve AdaBoost sınıflandırıcılarına giriş verisi olarak kullanılmıştır. İkinci başlık altında ViT, VGG16 ve EfficientNetB0 gibi derin öğrenme tabanlı modeller Caltech-101 veri seti üzerinde yeniden eğitilerek öznitelik çıkarılmıştır. klasik yöntemler düşük seviyeli özniteliklerin çıkarılmasında avantaj sağlarken, derin öğrenme tabanlı modeller daha karmaşık ve anlamsal temsillerin elde edilmesine olanak tanımaktadır. Deneysel sürecin genel akışı “girdi → öznitelik çıkarımı → sınıflandırma → çıktı” şeklinde olup, Şekil 14’te şematik olarak sunulmuştur. Bu akışta, Caltech-101 veri setinden alınan görüntüler önce normalize edilmekte, ardından ViT veya diğer öznitelik çıkarma yöntemleri ile sabit boyutlu vektörlere dönüştürülmekte ve son aşamada sınıflandırıcılar tarafından işlenerek tahmin sonuçları elde edilmektedir. Tüm sınıflandırıcıların performansı doğruluk (accuracy), kesinlik (precision), duyarlılık (recall) ve F1-değeri ölçütleri açısından karşılaştırmalı olarak değerlendirilmiş, en başarılı yöntemlerden elde edilen Görsel transformatörler (ViT) ve Lojistik Regresyon (LR) bu yöntemlerin matematiksel ifadesi aşağıdaki alt birimlerde 3.1,3.2’de anlatılmıştır.

3.1. ViT Modelinin Matematiksel Yapısı

Özellikle büyük veri setlerinde iyi performans gösteren ViT modeli, görüntüdeki genel bağlamı etkili bir şekilde modelleyebilir (Dosovitskiy vd., 2021). ViT modelinde, girdi görüntüleri sabit boyutlu küçük yamalara bölünerek bir dizin (indeks) olarak işlenir. Denklem 9, girdi görüntüsünden elde edilen P boyutlu yamaların toplam sayısını verir.

$$N = \frac{H \times W}{p^2} \quad (9)$$

Buradaki H , görüntünün yüksekliği ve W ise genişliğidir. Her yama, x_i boyutlu bir vektör olarak düzenlendikten sonra, doğrusal bir katman aracılığıyla daha düşük boyutlu bir alana yerleştirilir, bu doğrusal dönüşümün matematiksel ifadesi denklem 10’da verilmiştir.

$$Z_0^i = E \cdot x_i + e_{POS}^i \text{ for } i = 1, \dots, N \quad (10)$$

Denklem 10'da, E öğrenilebilir bir gömme matrisi, Z_0^i ikinci yamanın gömülü vektörü ve e_{pos}^i konumsal kodlamadır. Her gömülü vektör, Transformer Encoder tarafından işlenir. Kodlayıcı katmanı, çoklu başlı dikkat ve ileri besleme ağ ögelerinden oluşmakta olup, ilgili matematiksel ifade denklem 11'de verilmektedir.

$$Attention(Q, K, V) = softmax\left(\frac{QK^t}{\sqrt{d_k}}\right)V \quad (11)$$

Burada Q, K ve Z sırasıyla sorgu, anahtar ve değer matrisleridir. dk (baş) boyutudur ve her kodlayıcı katmanı denklem 12'de verilen işlemi üretir. MLP bir ileri beslemeli sinir ağıdır ve Z_l l 'inci katmandan gelen giriş vektörüdür.

$$Z_{l+1} = MLP(Attention(Z_l)) + Z_l \quad (12)$$

Tüm Transformer Encoder katmanlarının nihai çıktısı, tam bağlantılı bir ağdan oluşan sınıflandırma katmanına iletilir ve sınıf etiketleri, denklem 13'te verilen matematiksel ifadeyle elde edilir.

$$\hat{Y} = Softmax(W_{fc} \cdot Z_l) \quad (13)$$

Denklem 13'te, W_{fc} sınıflandırma için kullanılan ağırlık matrisi ve Z_l son Kodlayıcı katmanından gelen çıktı vektörüdür. Çapraz entropi kayıp fonksiyonu modeli eğitmek için kullanılır ve bu fonksiyon, tahmin edilen sınıf dağılımı ile gerçek sınıf dağılımı arasındaki farkı en aza indirecek şekilde optimize edilir.

3.2. LR Optimizasyon Prensipleri

Çok değişkenli regresyon analizi yöntemi olan lojistik regresyon, tıp, pazarlama, finans ve daha birçok alanda yaygın olarak kullanılmaktadır (Jin vd., 2022). Lojistik regresyonun temel amacı, bağımlı değişkenin belirli bir sınıfa ait olma olasılığını tahmin etmektir. Bu model, bu olasılıkları tahmin etmek için lojistik fonksiyon (sigmoid fonksiyon olarak da bilinir) kullanır. Lojistik regresyonun temel formülü denklem 14 ile verilmiştir.

$$P = \frac{1}{1 + e^{-z}} \quad (14)$$

Burada P, olayın meydana gelme olasılığı olup, z ise P 'nin odds oranının doğal logaritması olarak denklem 15 ile tanımlanır.

$$Z = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 \chi_1 + \beta_2 \chi_2 + \dots + \beta_n \chi_n \quad (15)$$

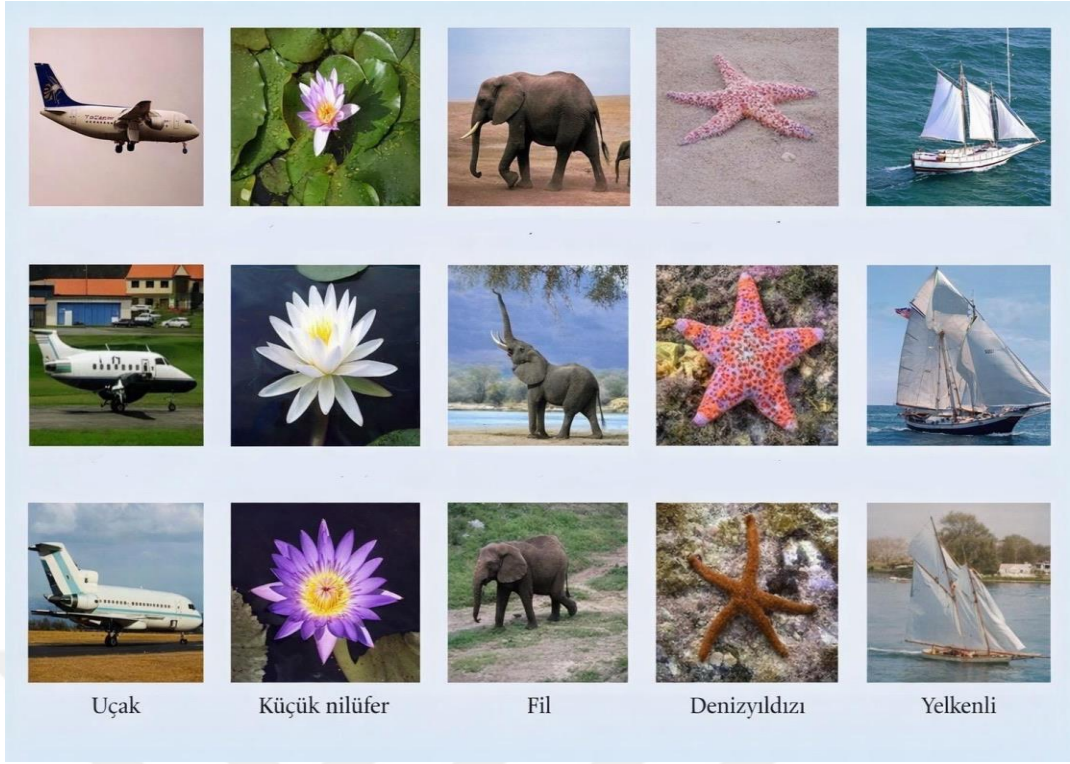
β_0 modelin kesim noktası (önyargı) ve β_n bağımsız değişkenlerin katsayılarıdır. Lojistik fonksiyon, bir olayın gerçekleşme olasılığını hesaplayarak çıktının 0 ile 1 arasında olmasını sağlar. Lojistik regresyondaki sigmoid fonksiyon, doğrusal bir denklemi 0 ile 1 arasındaki bir olasılığa dönüştürür. Bu dönüşüm, modelin bağımsız değişkenlerdeki değişikliklere duyarlılığını artırır ve doğrusal olmayan ilişkileri yakalamasına olanak tanır.

4. BULGULAR

Bu çalışma kapsamında, farklı öznitelik çıkarma yöntemleri ve sınıflandırıcılar kullanılarak nesne tanıma performansı değerlendirilmiştir. Çalışmada hem klasik öznitelik çıkarma yöntemleri SIFT, SURF, ORB, KAZE, BRISK, MSER hem de derin öğrenme tabanlı modeller VGG16, EfficientNetB0, ViT kullanılmıştır. Çalışmada kullanılan Caltech-101 veri seti, LightGBM, AdaBoost, ExtraTrees ve Logistic regression ile test edilmiş; performans değerlendirmesi için doğruluk, kesinlik, duyarlılık ve F1-değeri metrikleri kullanılmıştır. Son olarak, gerçekleştirilen deneylerden elde edilen bulgular aşağıda ayrıntılı olarak sunulmaktadır.

4.1. Veri Seti (Caltech-101)

Nesne sınıflandırma problemlerinde yaygın olarak kullanılan Caltech-101 veri seti (Fei-Fei vd., 2004; Mutch & Lowe, 2006), çoklu sınıf yapısı sunmaktadır. Veri seti 102 kategoride toplam 9.146 görüntü içermektedir. Görüntü sayıları sınıflar arasında oldukça dengesizdir; örneğin Airplanes (Uçaklar) sınıfında 800, Motorbikes (Motosikletler) sınıfında 798, Background_Google (Arka plan sahneleri) sınıfında 468, Faces_easy (Kolay yüzler) sınıfında 435 ve Faces (Yüzler) sınıfında 435 örnek bulunurken, Inline_skate (Paten) sınıfında yalnızca 31, Metronome (Metronom) sınıfında 32, Binocular (Dümbün) sınıfında 33, Wild_cat (Yaban kedisi) sınıfında 34 ve Platypus (Ornitorenk) sınıfında 34 örnek yer almaktadır. Bu çeşitlilik nedeniyle bazı çalışmalarda hariç tutulan Background_Google (Arka plan sahneleri) kategorisi de veri setine dahil edilmiştir. Kategori başına görüntü sayısı en az 31 ile en fazla 800 arasında değişmektedir. Görüntüler düşük çözünürlüklü ve gürültülüdür. Bu dengesiz yapı, sınıflandırma algoritmalarının baskın sınıflara yönelmesine ve az temsil edilen sınıfların performansının düşmesine yol açabilmektedir. Bu çalışmada veri seti standart olarak eğitim ve test kümelerine ayrılmadığından, 5-katlı çapraz doğrulama (cross-validation) yöntemi uygulanmıştır. Her katmanda veri seti eğitim ve test alt kümelerine ayrılmış, tüm örnekler hem eğitim hem de test için kullanılmış ve sonuçlar ortalama alınarak raporlanmıştır. Caltech-101 veri setine ait örnek görüntüler Şekil 15.'te sunulmaktadır.



Şekil 15. Caltech-101 veri setine ait örnek görüntüler (Cao vd., 2024)

4.2. Performans Metrikleri

Sınıf dengesizliği, özellikle nesne tanıma ve sınıflandırma problemlerinde model performansını doğrudan etkileyen kritik bir faktördür. Veri setlerinde bazı sınıfların diğerlerine göre daha fazla örneğe sahip olması, öğrenme sürecinde modelin baskın sınıflara yönelmesine ve az temsil edilen sınıfların göz ardı edilmesine yol açabilmektedir. Bu durum, sınıflandırma sonuçlarının güvenilirliğini azaltmakta ve modelin genelleme yeteneğini sınırlamaktadır. Bu nedenle performans değerlendirmesinde yalnızca doğruluk (accuracy) metriğine bağlı kalmak yeterli değildir. Accuracy, tüm örnekler üzerinden genel doğru tahmin oranını verir; ancak dengesiz veri setlerinde baskın sınıfların doğru tahmin edilmesiyle yüksek görünebilir. Örneğin, az temsil edilen sınıflar yanlış tahmin edilse bile accuracy değeri yüksek çıkabilir. Bu durum, modelin gerçek başarısını yansıtmaz.

Bu sorunu azaltmak ve dengeli bir karşılaştırma sağlamak amacıyla tüm metrikler Makro-Ortalama (Macro-Averaging) yöntemiyle hesaplanmıştır. Makro-ortalama yaklaşımı, her sınıf için ayrı ayrı hesaplanan performans değerlerinin ortalamasını alarak sınıflar arası eşit ağırlıklandırma sağlar. Böylece az temsil edilen sınıfların etkisi göz ardı edilmez ve genel performans değerlendirmesi daha tarafsız bir biçimde yapılabilir.

Sınıflandırma performansının değerlendirilmesinde en yaygın kullanılan ölçütler doğruluk (accuracy), kesinlik (precision), duyarlılık (recall) ve F1 değeridir (Ozdemir & Kunduraci, 2022). Bu metrikler, sınıflandırma sürecinde elde edilen temel hata tipleri ve doğru tahminler üzerinden türetilmektedir. Ayrıca, yalnızca genel başarıyı değil, her bir sınıfın ayrı ayrı ne kadar doğru tahmin edildiğini ölçebilme yeteneğine sahiptir. Özellikle duyarlılık ve F1 değeri, az temsil edilen sınıfların başarısını değerlendirmede kritik rol oynamaktadır.

Doğru Pozitif (TP – True Positive): Pozitif sınıfa ait örneklerin doğru biçimde pozitif olarak sınıflandırılmasıdır. Modelin hedef sınıfı doğru şekilde tanıdığı örnekleri ifade eder.

Doğru Negatif (TN – True Negative): Negatif sınıfa ait örneklerin doğru biçimde negatif olarak sınıflandırılmasıdır. Modelin yanlış sınıfa ait örnekleri doğru şekilde dışladığını gösterir.

Yanlış Pozitif (FP – False Positive): Negatif sınıfa ait örneklerin hatalı biçimde pozitif olarak sınıflandırılmasıdır. İstatistiksel olarak Tip I hata olarak adlandırılır ve modelin yanlış alarm üretme eğilimini gösterir.

Yanlış Negatif (FN – False Negative): Pozitif sınıfa ait örneklerin hatalı biçimde negatif olarak sınıflandırılmasıdır. İstatistiksel olarak Tip II hata olarak bilinir ve modelin pozitif örnekleri gözden kaçırma eğilimini ortaya koyar.

4.2.1. Doğruluk (Accuracy)

Sınıflandırma problemlerinde en temel performans ölçütlerinden biridir ve modelin tüm veri seti üzerinde ne kadar doğru tahmin yaptığını gösterir. Bu ölçüt, doğru tahmin edilen örneklerin toplam tahmin sayısına oranlanmasıyla elde edilmektedir. Matematiksel olarak doğruluk, doğru pozitif (TP) ve doğru negatif (TN) değerlerinin toplamının, tüm gözlemlerin (TP + TN + FP + FN) toplamına bölünmesiyle hesaplanır.

$$\text{Doğruluk} = \frac{TP + TN}{TP + TN + FP + FN} \quad (16)$$

Bu değer 0 ile 1 arasında değişmekte olup, 1'e yaklaştıkça modelin genel başarısının arttığı kabul edilmektedir. Ancak doğruluk tek başına modelin performansını tam olarak yansıtmayabilir. Özellikle sınıf dengesizliği bulunan veri kümelerinde doğruluk değeri yüksek çıkmasına rağmen, kesinlik (precision) ve duyarlılık (recall) gibi diğer metrikler düşük kalabilmektedir. Bu durum, modelin baskın sınıfları doğru tahmin ederken az temsil edilen sınıfları göz ardı ettiğini göstermektedir (Amidi & Amidi, 2018).

4.2.2. Kesinlik (Precision)

Kesinlik kavramı, bir veri seti üzerinden elde edilen doğru tahmin değerlerinin ne kadarının doğru sınıfa ait olduğunu belirleyen ölçümdür. Kavrama ait hesaplama işlemi, doğru pozitiflerin (TP), toplam pozitiflere (TP + FP) oranlanmasıyla gerçekleştirilmektedir.

$$Kesinlik = \frac{TP}{TP + FP} \quad (17)$$

4.2.3. Duyarlılık (Recall)

Duyarlılık, sınıflandırma problemlerinde gerçek pozitif örneklerin ne kadarının doğru tahmin edildiğini gösteren bir performans ölçütüdür. Başka bir ifadeyle, modelin pozitif sınıfları ne ölçüde yakalayabildiğini ortaya koyar. Özellikle dengesiz veri setlerinde, az sayıda pozitif örneğin gözden kaçırılması bile sonuçların güvenilirliğini ciddi şekilde etkileyebilir.

$$Duyarlılık = \frac{TP}{TP + FN} \quad (18)$$

4.2.4. F1 değeri (F1 Score)

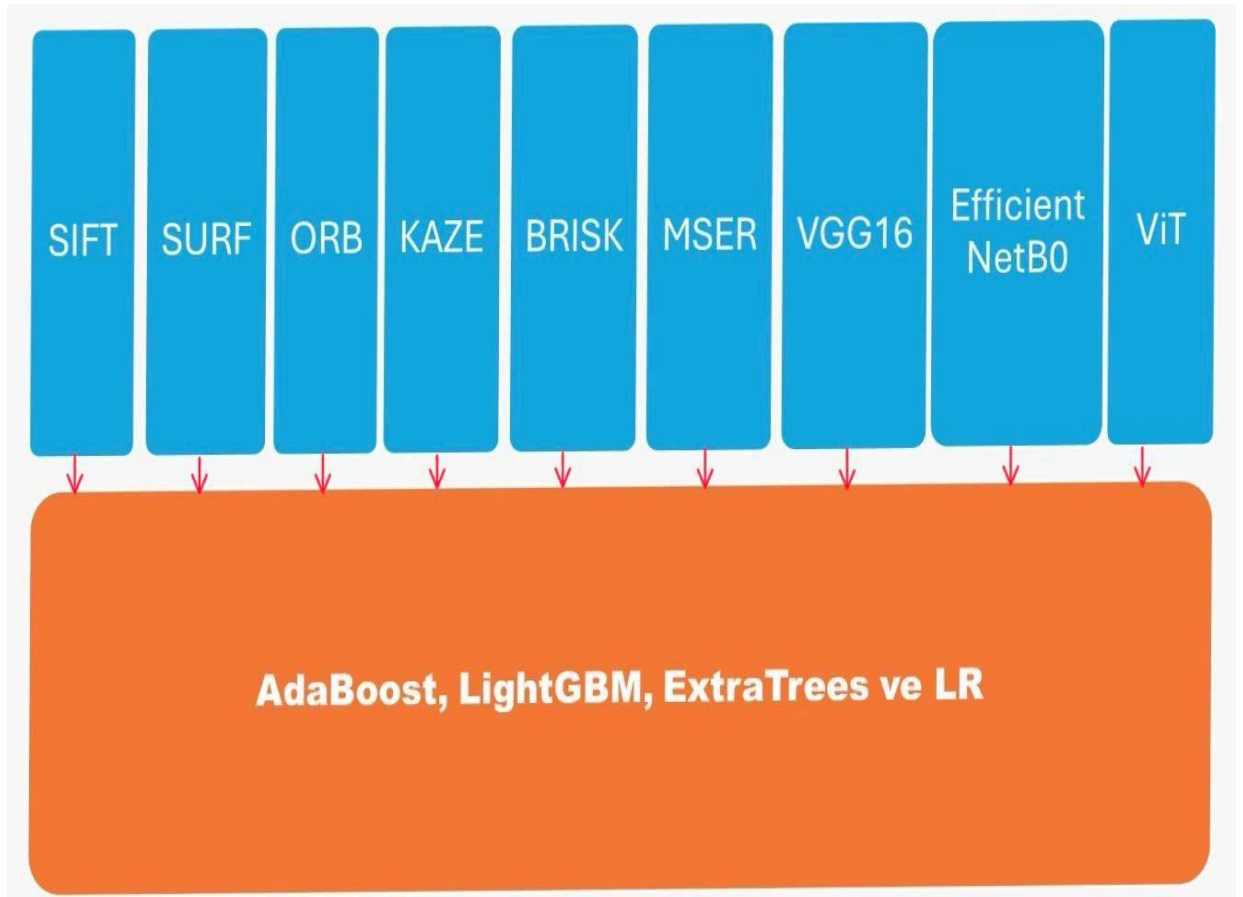
Performans değerlendirilmesinde doğru bir karşılaştırma yapabilmek için yalnızca duyarlılık ve kesinlik kriterleri yeterli değildir. F1 değeri, belirtilen bu iki kriteri birlikte değerlendirerek daha anlamlı veriler sağlar. çünkü kesinlik ve duyarlılık değerlerinin dengesiz olduğu durumlarda, harmonik ortalama daha güvenilir bir değerlendirme sunmaktadır.

$$F1 \text{ değeri} = \frac{Duyarlılık \times kesinlik}{Duyarlılık + kesinlik} \times 2 \quad (19)$$

4.3. DeneySEL Sonuçlar

Tez kapsamında öznetelik çıkarma için derin öğrenme tabanlı modeller VGG16, EfficientNetB0 ve ViT kullanılarak gerçekleştirilirken, sınıflandırma LightGBM, AdaBoost, ExtraTrees ve Lojistik Regresyon gibi gelişmiş yöntemlerle yapılmaktadır.

Çeşitli performans metrikleri, SIFT, SURF, ORB, KAZE, MSER ve BRISK gibi öznelilik çıkarma yöntemlerini karşılaştırmak için kullanılmıştır. Veri setinin çok sınıflı yapısı göz önünde bulundurularak, tüm sınıflar için tahmin sonuçlarının ortalamasını almak amacıyla makro-ortalama (macro-averaging) yöntemi tercih edilmiştir. Caltech-101 veri seti eğitim ve test setlerine ayrılmadığından, standart bir değerlendirme yöntemi olan çapraz doğrulama (cross-validation) kullanılmıştır. Deneyler ayrıca LightGBM, AdaBoost, ExtraTrees ve Lojistik Regresyon gibi çeşitli gelişmiş sınıflandırıcıların karşılaştırmalı analizini de içermektedir. Tüm deneyler, Intel Core i7 işlemci ve 8 GB RAM ile Microsoft Windows 10 (orijinal) çalıştıran bir bilgisayarda gerçekleştirilmiştir. Tez kapsamında yürütülen deneysel çalışmaların detayları Şekil 16’da gösterilmektedir.

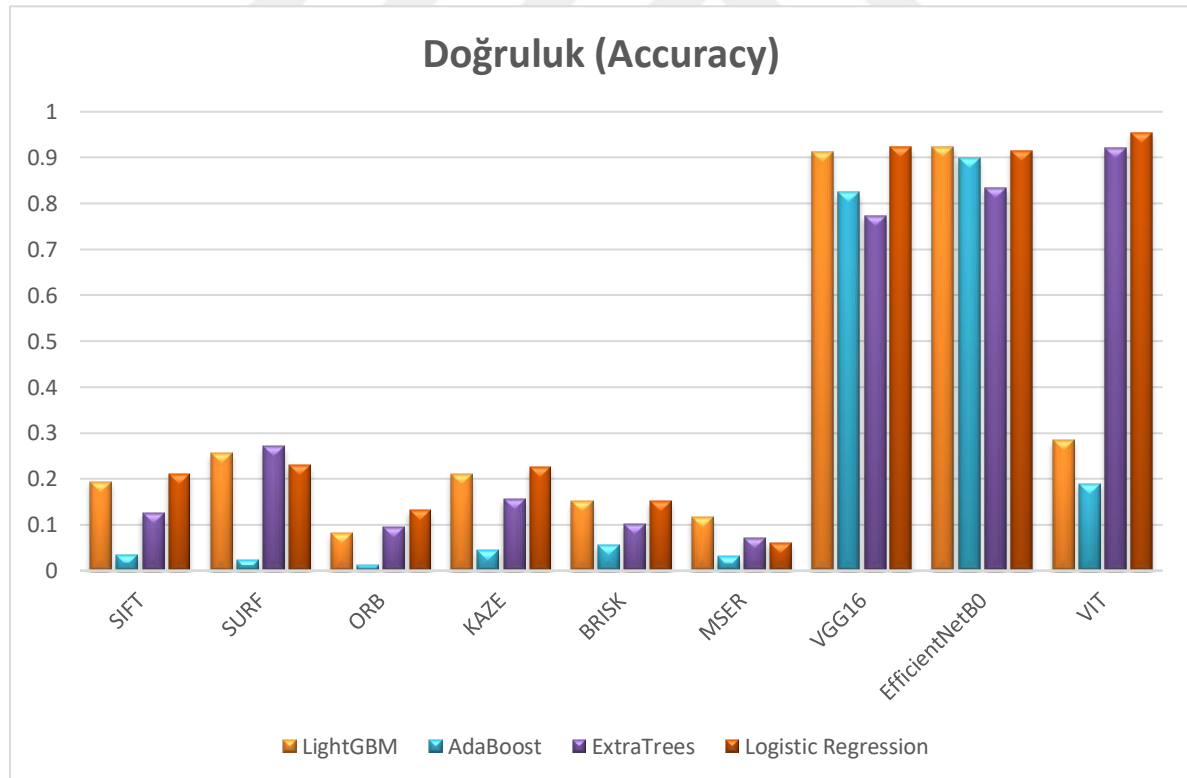


Şekil 16. Tez kapsamında gerçekleştirilen deneyler

Tablo 2’de farklı öznelilik çıkarma yöntemlerinin tanıma doğruluğu açısından karşılaştırması sunulurken, Şekil 17 bu karşılaştırmayı grafiksel olarak göstermektedir.

Tablo 2. Tüm yöntemlere ait Doğruluk (Accuracy) değerleri

Öznitelikler	LightGBM	AdaBoost	ExtraTrees	LR
SIFT	0.1933	0.0352	0.1251	0.2101
SURF	0.2572	0.0235	0.2703	0.2302
ORB	0.0814	0.0114	0.0951	0.1313
KAZE	0.2104	0.0442	0.1552	0.2251
BRISK	0.1506	0.0561	0.1011	0.1505
MSER	0.1153	0.0301	0.0712	0.0601
VGG16	0.9131	0.8241	0.7715	0.9422
EfficientNetB0	0.9443	0.8996	0.8336	0.9154
ViT	0.2851	0.1891	0.9207	0.9550



Şekil 17. Klasik ve derin öğrenme tabanlı yöntemlerin Doğruluk (Accuracy) performansı

Tablo 3, Tablo 4 ve Tablo 5, kesinlik (precision), duyarlılık (recall) ve F1-değeri değerlerine dayalı karşılaştırmayı sunmaktadır. Bu karşılaştırmalı sonuçlar, önerilen sistemin performansını açık bir şekilde gösteren Şekil 18’de grafiksel olarak ortaya konulmuştur. Deneysel sonuçlardan, önerilen yöntemde ExtraTrees ve LR kullanıldığında önemli iyileştirmeler elde edildiği gözlemlenmiştir. VİT tarafından çıkarılan güçlü öznitelikler bu sınıflandırıcılar tarafından etkili bir şekilde kullanılmış ve kesinlik, duyarlılık ve F1-değeri değerlerinde belirgin artışlar sağlanmıştır. Buna karşılık, LightGBM ve AdaBoost için optimal olmayan sonuçlar kaydedildi; bunların sınıf dengesizliğine duyarlı olmaları ve derin öznitelikleri daha az verimli bir şekilde işlemeleri nedeniyle olumsuz etkilendikleri düşünülmektedir. Dolayısıyla, önerilen yöntemin üstünlüğü, derin özellik çıkarma ve klasik sınıflandırma tekniklerinin optimal entegrasyonuna atfedilebilir.

Dolayısıyla, önerilen yöntemin üstünlüğü, derin öznitelik çıkarımı ile klasik sınıflandırma tekniklerinin optimal entegrasyonuna dayandırılmaktadır. Tabloların tümünde görüldüğü üzere, önerilen yaklaşım ile elde edilen sonuçlar tüm performans metriklerinde iyileşmeler göstermektedir. Şekil 17 ve Şekil 18, farklı sınıflandırıcıların tüm performans parametreleri üzerindeki ayrıntılı karşılaştırmasını sunmakta ve önerilen hibrit yaklaşımın klasik yöntemlere kıyasla daha avantajlı olduğunu ortaya koymaktadır.

Tablo 3. Tüm yöntemlere ait Kesinlik (Precision) değerleri

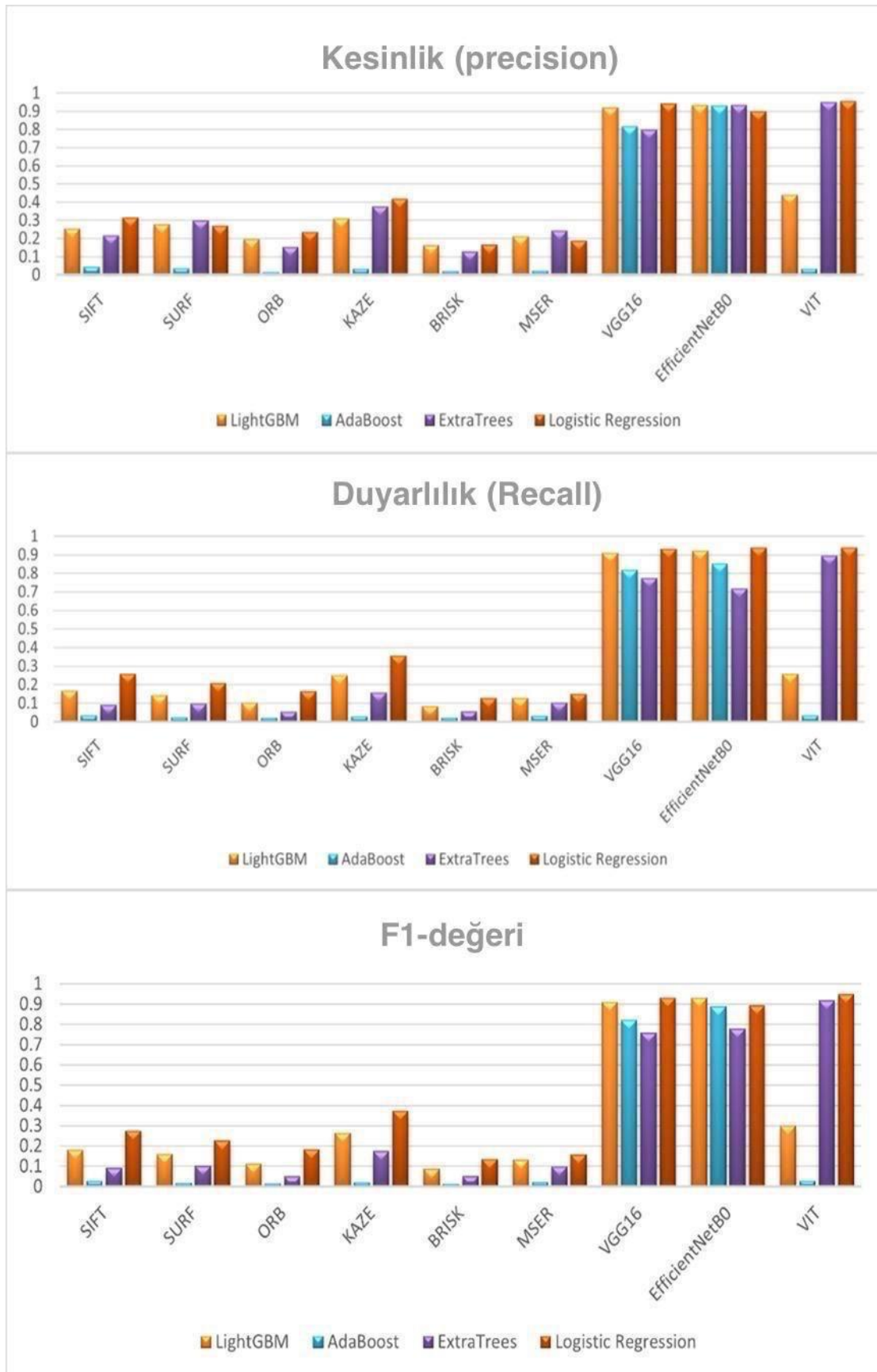
Öznitelikler	LightGBM	AdaBoost	ExtraTrees	LR
SIFT	0.2533	0.0438	0.2154	0.3138
SURF	0.2785	0.0366	0.2975	0.2691
ORB	0.1958	0.0150	0.1531	0.2355
KAZE	0.3092	0.0327	0.3763	0.4187
BRISK	0.1625	0.0190	0.1274	0.1651
MSER	0.2115	0.0228	0.2439	0.1864
VGG16	0.9202	0.8166	0.7976	0.9416
EfficientNetB0	0.9343	0.9317	0.9347	0.9001
ViT	0.4399	0.0331	0.9506	0.9566

Tablo 4. Tüm yöntemlere ait Duyarlılık (Recall) değerleri

Öznitelikler	LightGBM	AdaBoost	ExtraTrees	LR
SIFT	0.1669	0.0347	0.0897	0.2588
SURF	0.1418	0.0248	0.0970	0.2086
ORB	0.1022	0.0208	0.0547	0.1648
KAZE	0.2498	0.0302	0.1580	0.3554
BRISK	0.0842	0.0217	0.0557	0.1276
MSER	0.1274	0.0327	0.1032	0.1491
VGG16	0.9102	0.8191	0.7734	0.9306
EfficientNetB0	0.9208	0.8514	0.7175	0.9331
ViT	0.2578	0.0347	0.8946	0.9393

Tablo 5. Tüm yöntemlere ait F1- değeri değerleri

Öznitelikler	LightGBM	AdaBoost	ExtraTrees	LR
SIFT	0.1792	0.0272	0.0912	0.2749
SURF	0.1592	0.0175	0.1021	0.2255
ORB	0.1119	0.0139	0.0501	0.1824
KAZE	0.2631	0.0203	0.1741	0.3747
BRISK	0.0867	0.0134	0.0515	0.1345
MSER	0.1318	0.0219	0.0995	0.1575
VGG16	0.9114	0.8213	0.7581	0.9311
EfficientNetB0	0.9302	0.8865	0.7782	0.8931
ViT	0.2988	0.0266	0.9181	0.9474



Şekil 18. Yöntemlerin kesinlik, duyarlılık, F1-değeri değerleri

Tablo 2, farklı öznitelik çıkarma yöntemleri ile çeşitli sınıflandırıcılar kullanılarak elde edilen doğruluk değerlerini göstermektedir. Tablo 3, Tablo 4 ve Tablo 5 ise kesinlik (precision), duyarlılık (recall) ve F1-değeri açısından ayrıntılı karşılaştırmaları sunmaktadır. Özellikle Görsel transformatörler (ViT) tabanlı özniteliklerin, diğer geleneksel yöntemlere kıyasla daha tutarlı ve yüksek başarı sağladığı görülmektedir.

Şekil 17 ve Şekil 18’de önerilen yöntemin farklı metrikler açısından sağladığı iyileştirmeler grafiksel olarak sunulmuş ve bu yöntemin geleneksel yöntemlere göre belirgin şekilde üstün olduğu ortaya konulmuştur.

Bu başarı, ViT’in güçlü öznitelik çıkarma yeteneği ile klasik makine öğrenmesi sınıflandırıcılarıyla optimize edilmiş sınıflandırma süreciyle açıklanabilir. Karşılaştırmalar, önerilen yöntemin düşük kaliteli ve dengesiz veri setlerinde daha güvenilir ve başarılı sonuçlar sağladığını açıkça ortaya koymaktadır.

Önerilen yöntemin geleneksel yaklaşımlara kıyasla üstün performansı, sağlam öznitelik çıkarımı ve verimli sınıflandırma stratejisine dayandırılmaktadır. SIFT, SURF ve ORB gibi elle tasarlanmış öznitelik çıkarma teknikleri ölçek, döndürme ve gürültü değişimlerine karşı kesinlik gösterirken, Görsel transformatörler (ViT) öz-dikkat mekanizması kullanarak küresel bağlamsal bilgiyi yakalayabilmektedir.

Bu durum, özellikle Caltech-101 gibi dengesiz ve gürültülü veri setlerinde daha ayırt edici öznitelikler elde edilmesini sağlamaktadır.

Sınıflandırma için yalnızca derin öğrenmeye güvenmek yerine, Görsel transformatörler (ViT) ile çıkarılan öznitelikler LR, LightGBM, AdaBoost ve ExtraTrees gibi klasik makine öğrenmesi sınıflandırıcılarıyla birleştirilmiş ve böylece doğruluktan ödün vermeden hesaplama verimliliği sağlanmıştır.

Geleneksel sınıflandırıcılar çoğu zaman sınıf dengesizliği nedeniyle önyargılı tahminler üretmek olup, Görsel transformatörler’in güçlü öznitelik temsilleri bu sorunu azaltarak tüm sınıflar için daha iyi genelleme sağlamaktadır.

Şekil 17 ve Şekil 18’in analizi, geleneksel yöntemlerin farklı sınıflandırıcılar arasında tutarsız performans sergilediğini; buna karşılık önerilen yaklaşımın üstün doğruluk, kesinlik ve duyarlılık değerlerini koruduğunu ve nesne tanıma görevleri için son derece etkili bir çözüm sunduğunu doğrulamaktadır.

Son dönemde, çeşitli araştırmacılar görüntü sınıflandırmada doğruluk oranlarını artırmak amacıyla farklı topluluk (ensemble) yaklaşımlarını incelemişlerdir.

Tablo 6’da, önerilen sistemin Caltech-101 veri seti üzerinde gerçekleştirilen bazı güncel çalışmalarla karşılaştırmalı analizi sunulmaktadır.

Tablo 6. Önerilen sistemin bazı yeni deneylerle karşılaştırmalı analizi

Yöntem	Doğruluk (Accuracy)
Jalal vd. (Jalal vd., 2021)	0.8926
Naseer vd. (Aysha Naseer Nouf Almujaally, 2024)	0.9030
Rafique vd. (Rafique vd., 2023)	0.8860
Hussain vd. (Yemi Hussain, 2024)	0.9040
Gupta vd. (Gupta vd., 2019)	0.8560
Bansal vd. (Bansal, Kumar, Kumar, vd., 2021)	0.8640
Bansal vd. (Bansal, Kumar, & Kumar, 2021)	0.8330
Önerilen Yöntem (ViT + LR)	0.9550

Bu çalışma, Caltech-101 veri seti kullanılarak gerçekleştirilen kapsamlı analiz ve karşılaştırmalarla literatüre önemli bir katkı sağlamaktadır. Elde edilen sonuçlar, güçlü öznitelik çıkarımı ve gelişmiş sınıflandırma tekniklerinin nesne sınıflandırma problemlerinin çözümünde etkinliğini vurgulamaktadır. Bu bağlamda, çalışma gelecekteki araştırmalar için sağlam bir temel oluşturmakta ve sonraki çalışmalara değerli içgörüler sunmaktadır.

5. SONUÇLAR

Bu çalışma, dengesiz veri setlerinde güçlü öznitelik çıkarma ve sınıflandırma yöntemlerini kullanarak dengesizlik sorununu ele almayı amaçlamaktadır. Bu amaç doğrultusunda, ViT modeli kullanılarak çıkarılan öznitelikler, dengesiz veri setleri için yaygın olarak kullanılan sınıflandırıcılar olan Lojistik Regresyon, LightGBM, ExtraTrees ve AdaBoost yöntemleri ile sınıflandırılmıştır. Deneysel sonuçlar, ViT modeli ile çıkarılan özniteliklerin Lojistik Regresyon sınıflandırıcısı ile birleştirildiğinde, dengesiz veri setlerinde yüksek doğruluk oranları elde edildiğini göstermektedir.

Doğruluk, kesinlik, duyarlılık ve F1-değeri ölçütlerinin analizi, Lojistik Regresyon ve ExtraTrees sınıflandırıcılarının dengesizlik probleminden daha az etkilendiğini ve etkili çözümler sunduğunu ortaya koymaktadır. Özellikle Lojistik Regresyon yüksek başarı oranları elde ederken, LightGBM ve AdaBoost sınıflandırıcılarının dengesizlikten olumsuz etkilendiği ve daha düşük başarı oranları sağladığı gözlemlenmiştir.

Caltech-101 veri seti üzerinde gerçekleştirilen deneyler, önerilen hibrit yöntemin hem klasik hem de en güncel yöntemleri aştığını göstermektedir. SIFT, SURF, KAZE, ORB, MSER ve BRISK gibi öznitelik çıkarma yöntemleriyle yapılan karşılaştırmalar, ViT ve Lojistik Regresyon tabanlı yaklaşımın doğruluk, kesinlik, duyarlılık ve F1-değeri metriklerinde üstün başarı sağladığı tespit edilmiştir. Önerilen yöntem, derin öğrenme ile çıkarılan öznitelikleri klasik makine öğrenmesi sınıflandırıcılarıyla birleştirerek dengesizlik sorununu etkili bir şekilde azaltmakta ve mevcut gelişmiş yöntemlerden daha iyi performans göstermektedir.

Her ne kadar önerilen yaklaşım nesne tanımda üstün performans sergilese de, bazı sınırlamaların kabul edilmesi gerekir. Ana kısıtlamalardan biri, ViT'in öznitelik çıkarma sürecinin getirdiği hesaplama karmaşıklığıdır; bu durum kaynakların sınırlı olduğu ortamlarda uygulanabilirliği kısıtlayabilir. Ayrıca, bazı sınıflandırıcıların, özellikle LightGBM ve AdaBoost'un, sınıf dengesizliğine duyarlı olduğu gözlemlenmiştir ve bu durum dengesiz veri setlerinde önyargılı tahminlere yol açabilmektedir.

Önerilen yöntem, dengesiz veri setlerinin dezavantajlarını ortadan kaldırma yeteneği sayesinde sağlık ve güvenlik uygulamalarında nadir veya anormal olayların tespitini iyileştirerek sistemlerin etkinliğini artırabilir. Ayrıca, aydınlatma, ölçek ve gürültü değişikliklerine karşı sağlamlığı sayesinde, önerilen yöntem sürücüsüz araçlar ve robotik navigasyon gibi otonom sistemlerde gerçek zamanlı nesne tanımda uygulanabilir.

Sonuç olarak, bu çalışma dengesiz veri seti sorununa etkili ve verimli bir çözüm sunmakta ve literatürdeki güncel çalışmalara kıyasla daha yüksek başarı performansına sahiptir. Gelecekteki

alıřmalarda, farklı veri setleri üzerinde uygulamalar geliřtirmek ve daha karmařık hibrit yöntemleri keřfetmek üzerine odaklanacaktır.



KAYNAKÇA

- Ajit, A., Acharya, K., & Samanta, A. (2020). *A Review of Convolutional Neural Networks*. <https://doi.org/10.1109/ic-ETITE47903.2020.049>
- Akbulut, U., Cifci, M. A., & Aslan, Z. (2023). Hybrid Modeling for Stream Flow Estimation: Integrating Machine Learning and Federated Learning. *Applied Sciences (Switzerland)*, 13(18). <https://doi.org/10.3390/app131810203>
- Alcantarilla, P. F., Bartoli, A., & Davison, A. J. (2012). KAZE Features. İçinde A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, & C. Schmid (Ed.), *Computer Vision – ECCV 2012* (ss. 214-227). Springer Berlin Heidelberg.
- Alpaydin, Ethem. (2014). *Introduction to machine learning*. The MIT Press.
- Amidi, A., & Amidi, S. (2018). CS 229-Machine Learning VIP Cheatsheet: Machine Learning Tips. <https://stanford.edu/~shervine>
- Anandhasilambarasan, D., Bhushan, B., Ganga, S., Jain, R., Varma, P., & Mokashi, M. K. (2024). SIFT and SURF: A Comparative Analysis of Feature Extraction Methods for Image Matching. *2024 15th International Conference on Computing Communication and Networking Technologies, ICCCNT 2024*. <https://doi.org/10.1109/ICCCNT61001.2024.10726049>
- Apoorva, V., H K, Y., & M L, C. (2021). *Breast Cancer Prediction Using Machine Learning Techniques*. <https://doi.org/10.2991/ahis.k.210913.043>
- Aydın, Y., & Ali, A. K. A. A. (2025). Vision Transformer-Based Approach: A Novel Method for Object Recognition. *Karadeniz Fen Bilimleri Dergisi*, 15(1), 560-576. <https://doi.org/10.31466/kfbd.1620640>
- Aydın, Y., & Türkeş, M. K. (2025). Enhanced Emotion Recognition through Hybrid Deep Learning and SVM Integration. *Bitlis Eren Üniversitesi Fen Bilimleri Dergisi*, 14(1), 348-360. <https://doi.org/10.17798/bitlisfen.1588046>
- Aydın, Y. (2023). A Comparative Analysis of Skin Cancer Detection Applications Using Histogram-Based Local Descriptors. *Diagnostics*, 13(19). <https://doi.org/10.3390/diagnostics13193142>
- Aysha Naseer Nouf Abdullah Almujaally, S. S. A. A. J. P. (2024). Efficient Object Segmentation and Recognition Using Multi-Layer Perceptron Networks. *Computers, Materials & Continua*, 78(1), 1381-1398. <https://doi.org/10.32604/cmc.2023.042963>
- Bansal, M., Kumar, M., & Kumar, M. (2021). 2D object recognition: a comparative analysis of SIFT, SURF and ORB feature descriptors. *Multimedia Tools and Applications*, 80(12), 18839-18857. <https://doi.org/10.1007/s11042-021-10646-0>
- Bansal, M., Kumar, M., Kumar, M., & Kumar, K. (2021). An efficient technique for object recognition using Shi-Tomasi corner detection algorithm. *Soft Computing*, 25(6), 4423-4432. <https://doi.org/10.1007/s00500-020-05453-y>
- Bay, H., Tuytelaars, T., & Van Gool, L. (2006). SURF: Speeded Up Robust Features. İçinde A. Leonardis, H. Bischof, & A. Pinz (Ed.), *Computer Vision – ECCV 2006* (ss. 404-417). Springer Berlin Heidelberg.

- Bosch, A., Zisserman, A., & Munoz, X. (2007). Image Classification using Random Forests and Ferns. *2007 IEEE 11th International Conference on Computer Vision*, 1-8. <https://doi.org/10.1109/ICCV.2007.4409066>
- Buyukarikan, B., & Ülker, E. (2020). Aydınlatma özneteliği kullanılarak Evrişimsel Sinir Ağı modelleri ile meyve sınıflandırma. *Uludağ University Journal of The Faculty of Engineering*, 25, 81-100. <https://doi.org/10.17482/uumfd.628166>
- Cao, J., Jin, M., Tian, Y., Cao, Z., & Peng, C. (2024). Ancient mural dynasty recognition algorithm based on a neural network architecture search. *Heritage Science*, 12(1), 152. <https://doi.org/10.1186/s40494-024-01274-6>
- Chen, J., Wu, P., Zhang, X., Xu, R., & Liang, J. (2024). Add-Vit: CNN-Transformer Hybrid Architecture for Small Data Paradigm Processing. *Neural Processing Letters*, 56(3), 198. <https://doi.org/10.1007/s11063-024-11643-8>
- Cheng, Z., Wu, Y., Li, Y., Cai, L., & Ihnaini, B. (2025). A Comprehensive Review of Explainable Artificial Intelligence (XAI) in Computer Vision. İçinde *Sensors* (C. 25, Sayı 13). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/s25134166>
- Ciresan, D. C., Meier, U., Gambardella, L. M., & Schmidhuber, J. (2011). Convolutional Neural Network Committees for Handwritten Character Classification. *2011 International Conference on Document Analysis and Recognition*, 1135-1139. <https://doi.org/10.1109/ICDAR.2011.229>
- Dahiya, N., Gupta, S., & Garg, M. (2021). Microsatellite Instability in Gastrointestinal Cancer Using Deep Learning: A Review. *IOP Conference Series: Materials Science and Engineering*, 1022, 012025. <https://doi.org/10.1088/1757-899X/1022/1/012025>
- Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, 1, 886-893 c. 1. <https://doi.org/10.1109/CVPR.2005.177>
- Ding, Y., Zhu, H., Chen, R., & Li, R. (2022). An Efficient AdaBoost Algorithm with the Multiple Thresholds Classification. *Applied Sciences*, 12(12). <https://doi.org/10.3390/app12125872>
- Dong, S., Khattak, A., Ullah, I., Zhou, J., & Hussain, A. (2022). Predicting and Analyzing Road Traffic Injury Severity Using Boosting-Based Ensemble Learning Models with SHAPley Additive exPlanations. *International Journal of Environmental Research and Public Health*, 19, 2925. <https://doi.org/10.3390/ijerph19052925>
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*. <http://arxiv.org/abs/2010.11929>
- El-Kenawy, E. S. M., Ibrahim, A., Mirjalili, S., Eid, M. M., & Hussein, S. E. (2020). Novel feature selection and voting classifier algorithms for COVID-19 classification in CT images. *IEEE Access*, 8. <https://doi.org/10.1109/ACCESS.2020.3028012>
- Fei-Fei, L., Fergus, R., & Perona, P. (2004). Learning Generative Visual Models from Few Training Examples: An Incremental Bayesian Approach Tested on 101 Object

- Categories. *2004 Conference on Computer Vision and Pattern Recognition Workshop*, 178. <https://doi.org/10.1109/CVPR.2004.383>
- Fei-Fei, L., Fergus, R., & Perona, P. (2007). Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories. *Computer Vision and Image Understanding*, 106(1), 59-70. <https://doi.org/10.1016/j.cviu.2005.09.012>
- Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, 63(1), 3-42. <https://doi.org/10.1007/s10994-006-6226-1>
- Gollapudi, S. (2019). Artificial Intelligence and Computer Vision. İçinde S. Gollapudi (Ed.), *Learn Computer Vision Using OpenCV: With Deep Learning CNNs and RNNs* (ss. 1-29). Apress. https://doi.org/10.1007/978-1-4842-4261-2_1
- Gupta, S., Kumar, M., & Garg, A. (2019). Improved object recognition results using SIFT and ORB feature detector. *Multimedia Tools and Applications*, 78(23), 34157-34171. <https://doi.org/10.1007/s11042-019-08232-6>
- Hashemi, A., Orzechowski, G., Mikkola, A., & McPhee, J. (2023). Multibody dynamics and control using machine learning. *Multibody System Dynamics*, 58, 1-35. <https://doi.org/10.1007/s11044-023-09884-x>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- Ian Goodfellow, Y. B. A. C. (2016). Deep Learning. *MIT Press*.
- Jalal, A., Ahmed, A., Rafique, A. A., & Kim, K. (2021). Scene Semantic Recognition Based on Modified Fuzzy C-Mean and Maximum Entropy Using Object-to-Object Relations. *IEEE Access*, 9, 27758-27772. <https://doi.org/10.1109/ACCESS.2021.3058986>
- Jiang, Z. P., Liu, Y. Y., Shao, Z. E., & Huang, K. W. (2021). An improved VGG16 model for pneumonia image classification. *Applied Sciences (Switzerland)*, 11(23). <https://doi.org/10.3390/app112311185>
- Jin, J., Chen, G., Meng, X., Zhang, Y., Shi, W., Li, Y., Yang, Y., & Jiang, W. (2022). Prediction of river damming susceptibility by landslides based on a logistic regression model and InSAR techniques: A case study of the Bailong River Basin, China. *Engineering Geology*, 299, 106562. <https://doi.org/10.1016/j.enggeo.2022.106562>
- Karadağ, C., & Özdemir, D. (2022). BÖBREK TÜMÖRÜ TESPİTİ İÇİN DERİN ÖĞRENME YÖNTEMLERİNİN KARŞILAŞTIRMALI ANALİZİ. *Journal of Scientific Reports-B*, 006, 10-23. <https://dergipark.org.tr/en/pub/jsrb/issue/1179613>
- Kate, C., Kalpana, C., Sharma, A., Yadav, A. S., Kumar, A., & Kumar, S. S. (2023). Investigation of Machine Learning Algorithms for Pattern Recognition in Image Processing. *Proceedings of the 5th International Conference on Inventive Research in Computing Applications, ICIRCA 2023*, 898-904. <https://doi.org/10.1109/ICIRCA57980.2023.10220656>
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Neural Information Processing Systems*. <https://api.semanticscholar.org/CorpusID:3815895>

- Keerthana, D., Venugopal, V., Nath, M. K., & Mishra, M. (2023). Hybrid convolutional neural networks with SVM classifier for classification of skin cancer. *Biomedical Engineering Advances*, 5, 100069.
<https://doi.org/https://doi.org/10.1016/j.bea.2022.100069>
- Kravets, P., Pasichnyk, V., & Prodaniuk, M. (2024). Mathematical Model of Logistic Regression for Binary Classification. Part 1. Regression Models of Data Generalization. *Visnik Naciònal'nogo unièrsitetu "L'vivs'ka politehnika". Serià Informaciònì sistemi ta merežì*, 15, 290-321.
<https://doi.org/10.23939/sisn2024.15.290>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Içinde F. Pereira, C. J. Burges, L. Bottou, & K. Q. Weinberger (Ed.), *Advances in Neural Information Processing Systems* (C. 25). Curran Associates, Inc.
https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Commun. ACM*, 60(6), 84-90.
<https://doi.org/10.1145/3065386>
- Lazebnik, S., Schmid, C., & Ponce, J. (2006). Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories. *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, 2, 2169-2178. <https://doi.org/10.1109/CVPR.2006.68>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444. <https://doi.org/10.1038/nature14539>
- Leutenegger, S., Chli, M., & Siegwart, R. Y. (2011). BRISK: Binary Robust invariant scalable keypoints. *2011 International Conference on Computer Vision*, 2548-2555.
<https://doi.org/10.1109/ICCV.2011.6126542>
- Liu, P., Guo, J.-M., Chamnongthai, K., & Prasetyo, H. (2017). Fusion of color histogram and LBP-based features for texture image retrieval and classification. *Information Sciences*, 390, 95-111. <https://doi.org/https://doi.org/10.1016/j.ins.2017.01.025>
- Lowe, D. G. (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>
- Marsh, Abigail A, Elfenbein, Hillary Anger, & Ambady, Nalini. (2003). Nonverbal "Accents": Cultural Differences in Facial Expressions of Emotion. *Psychological Science*, 14(4), 373-376. <https://doi.org/10.1111/1467-9280.24461>
- Mikolajczyk, K., & Schmid, C. (2005). A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10), 1615-1630. <https://doi.org/10.1109/TPAMI.2005.188>
- Mutch, J., & Lowe, D. G. (2006). Multiclass Object Recognition with Sparse, Localized Features. *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, 1, 11-18. <https://doi.org/10.1109/CVPR.2006.200>
- Nandeshwar, V. J., Bhatlawande, S., Solanke, A., Sathe, H., Satao, S., Satpute, S., & Saste, A. (2025). Comparative analysis of feature descriptors and classifiers for

- real-time object detection. *International Journal of Reconfigurable and Embedded Systems (IJRES)*, 14(1), 89. <https://doi.org/10.11591/ijres.v14.i1.pp89-99>
- Naseer, A., Alzahrani, H. A., Almujaally, N. A., Nowaiser, K. Al, Mudawi, N. Al, Algarni, A., & Park, J. (2024). Efficient Multi-Object Recognition Using GMM Segmentation Feature Fusion Approach. *IEEE Access*, 12, 37165-37178. <https://doi.org/10.1109/ACCESS.2024.3372190>
- Naseer, A., Mudawi, N. Al, Abdelhaq, M., Alonazi, M., Alazeb, A., Algarni, A., & Jalal, A. (2024). CNN-Based Object Detection via Segmentation Capabilities in Outdoor Natural Scenes. *IEEE Access*, 12, 84984-85000. <https://doi.org/10.1109/ACCESS.2024.3413848>
- Nimma, D., & Uddagiri, A. (2024). Advancements in Deep Learning Architectures for Image Recognition and Semantic Segmentation. İçinde *IJACSA) International Journal of Advanced Computer Science and Applications* (C. 15, Sayı 8). www.ijacsa.thesai.org
- Ning, Y., Chen, R., & Xu, P. (2011). Wide baseline image mosaicing by integrating MSER and Hessian-Affine. *2011 4th International Congress on Image and Signal Processing*, 4, 2034-2037. <https://doi.org/10.1109/CISP.2011.6100566>
- Ozdemir, D., & Kunduraci, M. S. (2022). Comparison of Deep Learning Techniques for Classification of the Insects in Order Level With Mobile Software Application. *IEEE Access*, 10, 35675-35684. <https://doi.org/10.1109/ACCESS.2022.3163380>
- Palanisamy, B., Hassija, V., Chatterjee, A., Mandal, A., Chakraborty, D., Pandey, A., Chalapathi, G. S. S., & Kumar, D. (2025). Transformers for Vision: A Survey on Innovative Methods for Computer Vision. *IEEE Access*, 13, 95496-95523. <https://doi.org/10.1109/ACCESS.2025.3571735>
- Paul, S. K., Hoseini, P., Nicolescu, M., & Nicolescu, M. (2021). Performance analysis of keypoint detectors and binary descriptors under varying degrees of photometric and geometric transformations. *Journal of Image and Graphics(United Kingdom)*, 9(1), 1-8. <https://doi.org/10.18178/joig.9.1.1-8>
- Quinlan, J. R. (1993). *C4.5: programs for machine learning*. Morgan Kaufmann Publishers Inc.
- Rachapudi, V., & Lavanya Devi, G. (2021). Improved convolutional neural network based histopathological image classification. *Evolutionary Intelligence*, 14(3), 1337-1343. <https://doi.org/10.1007/s12065-020-00367-y>
- Rafique, A. A., Gochoo, M., Jalal, A., & Kim, K. (2023). Maximum entropy scaled super pixels segmentation for multi-object detection and scene recognition via deep belief network. *Multimedia Tools and Applications*, 82(9), 13401-13430. <https://doi.org/10.1007/s11042-022-13717-y>
- Rajawat, A. S., & Srivastava, A. (2024). Recognition of Parkinson's ailment by using various machine learning procedures. *Current Psychology*, 43(45), 34579-34600. <https://doi.org/10.1007/s12144-024-06891-9>
- Rangari, A., Das, S., & Devarajan, R. (2023). *Cursive Handwriting Recognition Using CNN with VGG-16*. <https://doi.org/10.1109/ICECONF57129.2023.10083561>
- Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011). ORB: An efficient alternative to SIFT or SURF. *2011 International Conference on Computer Vision*, 2564-2571. <https://doi.org/10.1109/ICCV.2011.6126544>

- Salari, A., Djavadifar, A., Liu, X., & Najjaran, H. (2022). Object recognition datasets and challenges: A review. *Neurocomputing*, 495, 129-152. <https://doi.org/10.1016/j.neucom.2022.01.022>
- Sikder, J., Islam, M. K., & Jahan, F. (2024). Object segmentation for image indexing in large database. *Journal of King Saud University - Computer and Information Sciences*, 36(2), 101937. <https://doi.org/https://doi.org/10.1016/j.jksuci.2024.101937>
- Simonyan, K., & Zisserman, A. (2015). *Very Deep Convolutional Networks for Large-Scale Image Recognition*. <http://arxiv.org/abs/1409.1556>
- Suresan, A., G, D. M., Venkatraman, M., Suresh, S., & P, S. (2021). Comparison of Machine Learning Algorithms for Smart License Number Plate Detection System. İçinde J. I.-Z. Chen, J. M. R. S. Tavares, S. Shakya, & A. M. Iliyasu (Ed.), *Image Processing and Capsule Networks* (ss. 63-75). Springer International Publishing.
- TAN, F. G., YÜKSEL, A. S., AYDEMİR, E., & ERSOY, M. (2021). DERİN ÖĞRENME TEKNİKLERİ İLE NESNE TESPİTİ VE TAKİBİ ÜZERİNE BİR İNCELEME. *European Journal of Science and Technology*. <https://doi.org/10.31590/ejosat.878552>
- Tan, M., & Le, Q. V. (2020). *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*. <http://arxiv.org/abs/1905.11946>
- Tareen, S. A. K., & Saleem, Z. (2018). A comparative analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK. *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, 1-10. <https://doi.org/10.1109/ICOMET.2018.8346440>
- Tekerek, A., & Yapici, M. M. (2022). A novel malware classification and augmentation model based on convolutional neural network. *Computers & Security*, 112, 102515. <https://doi.org/https://doi.org/10.1016/j.cose.2021.102515>
- Telçeken, M. (2022). *Derin Öğrenme Tabanlı Meyve Ayırıştırma Sistemi*.
- Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A., & Jegou, H. (2021). Training data-efficient image transformers & distillation through attention. İçinde M. Meila & T. Zhang (Ed.), *Proceedings of the 38th International Conference on Machine Learning* (C. 139, ss. 10347-10357). PMLR. <https://proceedings.mlr.press/v139/touvron21a.html>
- Venugopal, V., Joseph, J., Vipin Das, M., & Kumar Nath, M. (2022). An EfficientNet-based modified sigmoid transform for enhancing dermatological macro-images of melanoma and nevi skin lesions. *Computer Methods and Programs in Biomedicine*, 222, 106935. <https://doi.org/https://doi.org/10.1016/j.cmpb.2022.106935>
- Wang, C., Li, G., Xue, P., & Wu, Q. (2020). A Comparative Study of Face Recognition Classification Algorithms. *International Journal of Advanced Network, Monitoring and Controls*, 5(3), 23-29. <https://doi.org/10.21307/ijanmc-2020-024>
- Wang, L., Yan, J., Mu, L., & Huang, L. (2020). Knowledge discovery from remote sensing images: A review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10, e1371. <https://doi.org/10.1002/widm.1371>
- Yang, C.-H., Ren, J.-H., Huang, H.-C., Chuang, L.-Y., & Chang, P.-Y. (2021). Deep Hybrid Convolutional Neural Network for Segmentation of Melanoma Skin Lesion.

Computational Intelligence and Neuroscience, 2021, 1-15.
<https://doi.org/10.1155/2021/9409508>

Yemi Hussain, N. (2024). Deep Learning Architectures Enabling Sophisticated Feature Extraction and Representation for Complex Data Analysis. *International Journal of Innovative Science and Research Technology (IJISRT)*, 2290-2300.
<https://doi.org/10.38124/ijisrt/ijisrt24oct1521>

Zhang, L., & Pu, J. (2024). Object recognition based on shape interest points descriptor. *Electronics Letters*, 60(9), e13198.
<https://doi.org/https://doi.org/10.1049/ell2.13198>

Zhang, R., Wang, L., Cheng, S., & Song, S. (2023). MLP-based classification of COVID-19 and skin diseases. *Expert Systems with Applications*, 228, 120389.
<https://doi.org/https://doi.org/10.1016/j.eswa.2023.120389>

