

T.C.
BİLECİK ŞEYH EDEBALI ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
ELEKTRİK ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI

**BİTKİ HASTALIKLARININ TESPİTİNDE GELENEKSEL MAKİNE ÖĞRENMESİ
VE DERİN ÖĞRENME YÖNTEMLERİNİN BAŞARIM KARŞILAŞTIRMASI**

YÜKSEK LİSANS TEZİ

MUSA ÇETİNKAYA

TEZ DANIŞMANI

DOÇ. DR. ÜMİT ÇİĞDEM TURHAL

BİLECİK, 2024

10677600

T.C.
BİLECİK ŞEYH EDEBALI ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
ELEKTRİK ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI

**BİTKİ HASTALIKLARININ TESPİTİNDE GELENEKSEL MAKİNE ÖĞRENMESİ
VE DERİN ÖĞRENME YÖNTEMLERİNİN BAŞARIM KARŞILAŞTIRMASI**

YÜKSEK LİSANS TEZİ

MUSA ÇETİNKAYA

TEZ DANIŞMANI

DOÇ. DR. ÜMİT ÇİĞDEM TURHAL

BİLECİK, 2024

10677600

BEYAN

Bitki Hastalıklarının Tespitinde Geleneksel Makine Öğrenmesi ve Derin Öğrenme Yöntemlerinin Başarım Karşılaştırması adlı yüksek lisans tezi hazırlık ve yazımı sırasında bilimsel araştırma ve etik kurallarına uyduğumu, başkalarının eserlerinden yararlandığım bölümlerde bilimsel kurallara uygun olarak atıfta bulunduğumu, kullandığım verilerde herhangi bir tahrifat yapmadığımı, tezin herhangi bir kısmının Bilecik Şeyh Edebali Üniversitesi veya başka bir üniversitede başka bir tez çalışması olarak sunulmadığını, aksinin tespit edileceği muhtemel durumlarda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Bu çalışmanın, Bilimsel Araştırma Projeleri (BAP), TÜBİTAK veya benzeri kuruluşlarca desteklenmesi durumunda; projenin ve destekleyen kurumun adı proje numarası ile birlikte, ETİKKURUL onayı alınması durumunda ise ETİKKURUL tarih karar ve sayı bilgilerinin beyan edilmesi gerekmektedir.			
DESTEK ALINMIŞTIR	☐	DESTEK ALINMAMIŞTIR	☒
Destek alındı ise;			
Destekleyen kurum;			
Desteğin Türü		Proje Numarası	
1- BAP (Bilimsel Araştırma Projesi)			
2- TÜBİTAK			
Diğer;.....			
ETİK KURUL onayı var ise;			
ETİK KURUL karar tarih/sayı:		/.....	

Musa ÇETİNKAYA

Tarih

.....

İmza

.....

ÖN SÖZ

Yüksek lisans eğitimim ve tez çalışmam boyunca desteğini benden hiçbir zaman esirgemeyen, karşılaştığım bütün zorluklar karşısında bilgisi ve fikirleriyle yolumu aydınlatan, öğrencisi olmaktan gurur duyduğum Doç. Dr. Ümit Çiğdem TURHAL'a, teşekkürlerimi ve saygılarımı sunarım.

Hayatım boyunca bu uzun ve zor yolda maddi, manevi desteğini esirgemeyen babam Hüseyin ÇETİNKAYA'ya, her türlü fedakarlığı ile bütün kararlarımda yanımda olan annem Selma ÇETİNKAYA'ya, zor zamanımda bana moral, motivasyon kaynağı olan kardeşlerim Salih ÇETİNKAYA ve Fatih ÇETİNKAYA'ya sonsuz teşekkürlerimi sunarım.

Musa ÇETİNKAYA

2024

ÖZET

BİTKİ HASTALIKLARININ TESPİTİNDE GELENEKSEL MAKİNE ÖĞRENMESİ VE DERİN ÖĞRENME YÖNTEMLERİNİN BAŞARIM KARŞILAŞTIRMASI

Geleneksel makine öğrenmesi (GMÖ) ve derin öğrenme (DÖ), yapay zeka alanında iki önemli yaklaşımdır. GMÖ, tahmine dayalı modeller oluşturmak için destek vektör makineleri (DVM), k-en yakın komşular, karar ağaçları gibi çeşitli algoritmalar kullanır. Bu yaklaşımda özellik çıkartımı, algoritmanın başarımını etkileyen en önemli unsurdur. GMÖ küçük veri kümelerinde başarılı sonuçlar elde ederken karmaşık veriler ve büyük ölçekli veri kümelerinde aynı başarıyı gösteremezler. Makine öğrenmesinin bir alt kümesi olan DÖ ise ham veriden hiyerarşik özellik temsillerini otomatik olarak öğrenmek için çok katmanlı sinir ağlarını kullanır. Finans, endüstri, tıp gibi alanlarda ve görüntü sınıflandırma, doğal dil işleme ve ses-tanıma işleme gibi çalışmalarda çok iyi sonuçlar elde ederken büyük miktarda veri ve hesaplama kaynağı gerektirir. Bu tez çalışması GMÖ ve DÖ modellerinin tahmin performansını karşılaştırmak amacıyla bitki hastalıklarının bilgisayar destekli teşhisini sunmaktadır. Bitki hastalıklarının tespiti, modern tarımda ürün kayıplarını en aza indirmek ve gıda güvenliğini sağlamak açısından çok önemlidir. Bitki hastalıklarının erken ve doğru bir şekilde belirlenmesi pestisit kullanımını azaltabilir, mahsul verimini artırabilir ve ürün kalitesini iyileştirebilir. Bu amaçla, çalışmada geleneksel yöntemlerden Lineer Regresyon, DVM ve Rastgele Orman ile DÖ yöntemlerinden Evrişimli Sinir Ağı, LSTM, YOLOv8 ve VGG16 algoritmalarının performansları karşılaştırılmıştır. Elde edilen sonuçlar GMÖ'de DVM için %83 ile en yüksek olmak üzere ortalama %77,9 doğruluk elde edilirken DÖ'de CNN için doğruluk en yüksek %99 ve ortalama %91,8 olarak elde edilmiştir.

Anahtar kelimeler: Yapay zeka, makine öğrenmesi algoritmaları, özellik çıkartımı, derin öğrenme.

ABSTRACT

PERFORMANCE COMPARISON OF TRADITIONAL MACHINE LEARNING AND DEEP LEARNING METHODS IN PLANT DISEASE DETECTION

Traditional machine learning (TML) and deep learning (DL) are two important approaches in artificial intelligence. To build predictive models, TML uses algorithm such as support vector machines (SVMs), k-nearest neighbors and decision trees. In this approach, feature extraction is the most important process that affects the algorithm's performance. While TML achieves successful results in small datasets, it struggles to show the same success in complex data and large-scale datasets. DL, a subset of machine learning, uses multilayer neural networks to automatically learn hierarchical feature representations from raw data. While it achieves very good results in fields such as finance, industry, medicine, and in studies such as image classification, natural language processing, and voice recognition-processing, it requires a large amount of data and computational resources. This study presents a computer-aided diagnosis of plant diseases to compare the predictive performance of TML and DL. Plants are a source of food and life of great importance for humans and other living things. Detection of plant diseases is very important in modern agriculture to minimize product losses and ensure food safety. Early and accurate identification of plant diseases can reduce pesticide use, increase crop yields and improve product quality. For this purpose, the performance of traditional methods such as LR, SVM, and RF was compared DL methods such as convolutional neural network (CNN), LSTM, YOLOv8 and VGG16 algorithms. The results were obtained as an average accuracy of 77,9%, with the highest being 83% for SVM in TML, and an average accuracy of 91,8%, with the highest being 99% for CNN in DL.

Keywords: Artificial intelligence, machine learning algorithms, feature extraction, deep learning.

İÇİNDEKİLER

Sayfa

ÖN SÖZ	i
ÖZET	ii
ABSTRACT	iii
İÇİNDEKİLER.....	iv
TABLolar LİSTESİ	v
ŞEKİLLER LİSTESİ	vi
GRAFİKLER LİSTESİ.....	vii
KISALTMALAR VE SİMGELER LİSTESİ	viii
1. GİRİŞ	1
1.1. Literatürdeki Çalışmalar	3
1.2. GMÖ ve DÖ Modellerinin Yöntemsel Olarak Karşılaştırması	5
1.2.1. Özellik Çıkartımı	5
1.2.2. Veri Seti Boyutu.....	6
1.2.3. Çalışma Biçimi ve Kullanılan Birimler	6
1.2.4. Çalışma Zamanı ve Doğruluk.....	6
2. MATERYAL METOT	7
2.1. Materyal	7
2.2. Metot.....	9
2.2.1. Özellik Çıkartımı	9
2.2.2. Sınıflandırma	9
2.2.3. Lojistik Regresyon	9
2.2.4. Destek Vektör Makinesi	10
2.2.5. Rastgele Orman	11
2.2.6. LSTM	12
2.2.7. VGG16.....	13
2.2.8. YOLOv8.....	13
2.2.9. CNN.....	14
2.2.10. Performans Değerlendirme	16
3. DENEYSEL ÇALIŞMALAR VE BULGULAR	18
4. SONUÇLAR VE ÖNERİLER.....	22
KAYNAKÇA	23

TABLÖLAR LİSTESİ

	Sayfa
Tablo 2.1. Eğitim ve doğrulama veri seti	8
Tablo 3.1. Model doğruluk yüzde verileri	19



ŞEKİLLER LİSTESİ

	Sayfa
Şekil 2.1. Veri setinde bulunan örnek bitki görüntüleri	8
Şekil 2.2. Lojistik Sigmoid Fonksiyonu	10
Şekil 2.3 Destek vektörlerinin iki sınıf arasındaki ayrımı yapması	11
Şekil 2.4 Rastgele Orman sınıflandırma modelini yapısı	12
Şekil 2.5. VGG16 mimarisi	13
Şekil 2.6. Genel YOLO mimarisi.....	13
Şekil 2.7. YOLO nesne tespit etme akış şeması	14
Şekil 2.8. ReLU fonksiyonu	15
Şekil 2.9. Havuzlama çeşitleri verilmiştir. Filtre boyutu 2×2 ve adım sayısı 2 verilmiştir.	15
Şekil 2.10. Karışıklık Matrisi.....	16
Şekil 3.1. Akış Şeması.....	18
Şekil 3.2. Çeşitli özellik çıkarımı yöntemleri	19

GRAFİKLER LİSTESİ

	Sayfa
Grafik 2.1. ROC Eğrisi	17
Grafik 3.1. CNN modelinin kayıp (loss) eğrisi	20
Grafik 3.2. CNN modelinin doğruluk (accuracy) eğrisi	20



KISALTMALAR VE SİMGELER LİSTESİ

AUC: Area Under Curve

AÖM: Aşırı Öğrenme Makinesi

CNN: Convolutional Neural Networks

DÖ: Derin Öğrenme

DVM: Destek Vektör Makinesi

DP: Doğru Pozitif (True Positive)

DN: Doğru Negatif (True Negative)

DT: Decision Trees

GLM: Gri Seviye Eş-Oluşum Matrisi (Gray Level Co-Occurrence Matrix)

GPU: Grafik İşlemci Birimi (Graphics Processing Unit)

HOG: Yönlü Gradyanlar Histogramı (Histogram of Oriented Gradients)

HSV: Renk Tonu, Doygunluk, Değer (Hue, Saturation, Value)

IoU: Birleşimin Kesişime Oranı (Intersection over Union)

İHA: İnsansız Hava Aracı

k-NN: k-En Yakın Komşular (k-Nearest Neighbors)

KA: Karar Ağaçları

LR: Lojistik Regresyon (Logistic Regression)

LSTM: Uzun Kısa Süreli Bellek (Long Short-Term Memory)

LB: Yerel İkili Örüntü (Local Binary Pattern)

ROC: Receiver Operating Characteristic

ReLU: Rectified Linear Unit

RF: Random Forest

RNN: Tekrarlayan Sinir Ağı (Recurrent Neural Network)

RO: Rastgele Orman

SVM: Support Vector Machine

TML: Traditional Machine Learning

YOLOv8: You Only Look Once version 8

YP: Yanlış Pozitif (False Positive)

YN: Yanlış Negatif (False Negative)



1. GİRİŞ

Tarım ürünleri hem insan yaşamının sürdürülmesi hem de ekonomik kalkınmanın teşvik edilmesi için olmazsa olmazdır. Bu ürünler gıda ve giyim üretiminde kullanılır, toplumların refahını iyileştirir ve ekonomilerine önemli ölçüde katkıda bulunur. Dünya nüfusunun hızla artmasıyla birlikte tarım ürünlerine olan talep artmıştır. Bu nedenle, tarım ürünlerindeki kayıpları önlemek büyük önem taşımaktadır. Anormal mevsimsel değişiklikler, doğal afetler, insan etkileri ve bitki hastalıkları gibi çeşitli faktörler tarım ürünlerinde kayıplara neden olabilir. Bu kayıpları en aza indirmek, tarımın sürdürülebilirliğini ve dolayısıyla insanlığın geleceğini güvence altına almak için oldukça önemlidir. Bitki hastalıkları, hem bitki gelişimini hem de daha geniş sosyal ve ekonomik manzarayı etkileyerek tarım üzerinde önemli olumsuz etkilere neden olabilir. Bitki yaprak hastalıklarının erken tespiti, bitki sağlığının korunması açısından son derece önemlidir. Her bitki hastalığının yapraklar üzerinde farklı etkileri olabileceğinden, yalnızca insan müdahalesi yeterli değildir. Bu nedenle, hastalıkların kesin ve zamanında tespiti esastır. (Sardogan, M., Tuncer, A., & Ozen, Y. (2018)).

Görüntü işleme ve bilgisayar destekli sistemler, yaprak analizi yoluyla bitki hastalıklarını tespit etmede çok önemlidir. Bitkilerde görülen hastalıkların tespiti, zaman alıcı, emek gerektiren, tespitler yapılırken ve sonucunda insan faktöründen kaynaklanan hatalara neden olabilen ve uzmanlar tarafından yapılması gereken incelemeye dayanır. Ancak, görüntü işleme tekniklerindeki gelişmeler, artık büyük veri kümelerinin yüksek doğruluk ve verimlilikle analiz edilmesine olanak tanımaktadır. Bu sistemler, genellikle hastalığın erken belirtileri olan yaprak rengi, dokusu ve şeklindeki ince değişiklikleri belirleyebilir. Bu, zamanında müdahaleyi mümkün kılarak yaygın ürün hasarı olasılığını azaltır.

Bilgisayar destekli sistemlerin görüntü işlemeyle bütünleştirilmesi yalnızca hastalık tespit hassasiyetini artırmakla kalmaz, aynı zamanda süreci otomatikleştirir. Hastalıklı ve sağlıklı yaprak görüntülerinin geniş veri kümeleri üzerinde eğitilen makine öğrenmesi algoritmaları, hastalıkları doğru bir şekilde belirleyebilir ve sınıflandırabilir. Bu otomasyon, her bitkinin ayrı ayrı izlenmesinin pratik olmadığı büyük ölçekli çiftçilikte özellikle faydalıdır. Çiftçiler bu teknolojileri kullanarak bitki hastalıklarını derhal tespit edebilir ve bu tespitler doğrultusunda gerekli önlemleri alarak, ürün kayıplarını en aza indirebilir ve verimi artırabilir. Ayrıca bu sistemler, hedefli uygulamalara olanak sağlayarak, geniş spektrumlu pestisitlere olan ihtiyacı azaltarak ve sonuç olarak daha sağlıklı mahsul ve ekosistemleri destekleyerek sürdürülebilir tarım uygulamalarına katkıda bulunmaktadır.

Günümüzün bilgi odaklı çağında, değerli görsel bilgilere erişim çok sayıda sorunu çözmede önemli bir rol oynar ve görüntüler önemli miktarda bilgi içerir. Görsel bilgileri çıkarmak, sınıflandırma ve kümeleme gibi görüntü işleme metotlarını ve çeşitli yapay zeka alanındaki uygulamalarını içerir. Görüntü sınıflandırması bu alandaki araştırma konularında önemli bölümünü oluşturmuştur ve çeşitli algoritmalar bu alanda kullanılmaktadır. Makine öğrenmesinin bir alt başlığı olan DÖ, grafik kartlarının gücü sayesinde son yıllarda bu alanda öne çıkmıştır.

Her modelin çeşitli problemler üzerinde farklı performans değerleri vardır. Bu nedenle, GMÖ ile DÖ algoritmaları kullanılırken mevcut probleme göre seçilmesi, karşılaştırılması ve yorumlanması problemin çözülmesi açısından en uygun algoritmaların seçilmesi önemlidir (Wang, P., Fan, E., & Wang, P. (2021)).

Günümüzdeki yapay zeka uygulamalarını daha iyi anlamak için GMÖ ile DÖ arasındaki farkları idrak etmek önemlidir. Makine öğrenmesi modelleri, verilerden öğrenerek belirli görevleri otomatik olarak sonuçlarını doğruluk değerleriyle harmanlayan modellerdir. GMÖ yöntemlerinde, insan uzmanlığı öne çıkmaktadır. Bu modellerde özellik çıkartma işlemleri manuel yapılır. Bu yöntemlerde, lojistik regresyon (LR), karar ağacı (KA), destek vektör makinesi (DVM) gibi algoritmalar kullanılır ve daha küçük yapılandırılmış veri gruplarıyla problemleri çözmede etkilidir.

DÖ, büyük ve karmaşık veri gruplarından önemli verileri otomatik olarak çıkartmak için sinir ağlarını kullanan bir makine öğrenmesi alt kümesidir. DÖ algoritmaları finans, tıp, endüstri gibi alanlarda özellikle görüntü sınıflandırma, bilgisayarlı görü, doğal dil işleme, ses tanıma-işleme uygulamalarında başarılı olmuştur. Bu algoritmalar, yapay sinir ağlarının katmanlarını kullanarak ham verilerden özellikler çıkarabilir ve sınıflandırmada kullanılabilir. DÖ'nün, büyük veri kümelerinde, önemli bilgi içeren kaynaklarda ve uygulamalarda oldukça etkili olduğu kanıtlanmıştır (Yılmaz, A., Demircali, A. A., Kocaman, S., & Uvet, H. (2020)). Bu tez çalışmasında, en uygun görüntü sınıflandırma algoritmasının seçilmesinde yol göstermek amacıyla literatürde olan makine öğrenmesi yöntemlerinden LR, DVM ve rastgele orman (RO) algoritmaları ile literatürde popüler olan DÖ yöntemlerinden LSTM, VGG16, YOLOv8 ve KSA algoritmalarının yaprak görüntüleri üzerinden bitki hastalıklarının teşhisi şeklinde gerçekleştirilen performansları karşılaştırılmalı verilmiştir.

1.1. Literatürdeki Çalışmalar

Literatürde GMÖ ve DÖ arasındaki performans farklarını, kullanılan veri türleri gibi konularda inceleyen çalışmalar bulunmaktadır. Özellikle son yıllarda yaprak hastalıkları üzerine çalışmaların sayısı artmıştır (Jayswal, H. S., & Chaudhari, J. P. (2020)).

DÖ ve GMÖ alanında yazılmış bir makalede bu iki yaklaşıma geniş bir pencereden bakılmıştır. Analitik model oluşturma aşamasında veri girişi, özellik çıkartımı, model oluşturma ve oluşturulan modellerin değerlendirilmesi gibi aşamalara değinilmiştir (Janiesch, C., Zschech, P., & Heinrich, K. (2021)).

GMÖ ve DÖ ile ilgili domates bitkisi üzerinde hastalık tespiti yapılan bir çalışmada, farklı özellik çıkartımı işlemleri yapılmış bunun sonucunda renkli görüntülerden alınan özellikler k-NN sınıflandırıcısında kullanılmış ve renkli görüntüler ile GLCM en iyi sonucu göstermiştir. En iyi sonucu veren özellikler çıkarılmış ve 3 farklı makine öğrenmesi sınıflandırıcısı (k-NN, DVM ve RO) ve 5 farklı derin öğrenme metodu (AlexNet, ResNet34, EfficientNet-b0, MobileNetV2, VGG16) kullanılmıştır. En iyi sonucu sunan ResNet34, %99,7 doğruluk yüzdesiyle başarılı olmuştur (Tan, L., Lu, J., & Jiang, H.(2021)).

Domates bitkisinde görülen hastalıkların incelendiği bir tez çalışmasında (Kapucuoğlu, 2022) birden fazla CNN modeli karşılaştırılmıştır. Çalışmada dört farklı veri seti ve AlexNet, ResNet-50, Inceptionv3, Inception ve ResNetv2 DÖ modelleri kullanılmıştır. Farklı veri setlerinde en yüksek doğruluk oranıyla Inceptionv3 modeli başarılı olmuştur.

Narenciye meyvelerinin olduğu veri seti ile yapılan bir çalışmada, 150 görüntü ile %67 doğruluk elde edilmiştir. Veri setinde veri arttırımı ve veri üzerinde ön işleme adımlarından sonra aynı CNN modeliyle %89,1 doğruluk elde edilmiştir. Yapılan çalışmada, veri seti üzerindeki ön işlemlerin modelin doğruluğunu önemli ölçüde arttırdığı belirtilmiştir (Kukreja, V., & Dhiman, P. (2020)).

GMÖ ve DÖ alanında yapılmış bir yüksek lisans tezinde patates ve üzüm yaprakları üzerine çalışma gerçekleştirilmiştir. Test edilen modellerden CNN olan modelleri daha başarılı sonuçlar elde etmiştir. Ayrıca literatürde aynı veri setleriyle yapılmış diğer çalışmalar ile doğruluk oranları karşılaştırılmıştır. Önerilen modelin literatürde kullanılan modellere kıyasla %8,6 oranında iyileşme sağladığı belirtilmiştir. (Sofuoğlu, C. (2024)).

Derin öğrenme modellerinin karşılaştırıldığı ve hastalıklı pirinç veri setinin kullanıldığı bir çalışmada, 7 farklı hazır model haricinde geliştirilen modelin de performans metrikleri verilmiştir. MobileNet-V2 üzerine şekillendirilen ve ağırlıklarını ImageNet'ten alan

bu model, hastalıklı pirinç veri setinde %98,52 başarı oranı elde etmiştir (Chen, J., Zhang, D., Zeb, A., & Nanekaran, Y. A. (2021)).

Soya fasulyesi bitkisi veri setinin kullanıldığı bir çalışmada, geliştirilen AlexNet ve GoogleNet modellerinin 3 farklı hastalıkta ayrı ayrı ve bütün hastalıklarının olduğu kombine veri setindeki doğruluk sonuçları verilmiştir. AlexNet için %98,75 ve GoogleNet için %96,25 doğruluk elde edilmiştir. Manuel olarak özellik çıkarımı yapılmış ve bir makine öğrenmesi modeli ile sonuçlar karşılaştırılmıştır. Makine öğrenmesi modeli %93 doğruluk elde etmiştir. Derin öğrenme modellerinin makine öğrenmesi modeline göre daha başarılı olduğu görülmüştür (Jadhav, S. B., Udupi, V. R., & Patil, S. B. (2021)).

PlantVillage veri setinin kullanıldığı bir çalışmada hastalıklı bitki üzerine uygulanan GLCM ile 6 farklı renk özelliği ve 22 farklı doku özelliği hesaplanmıştır. Çıkarılan özellik vektörü modellere uygulanarak modelin daha verimli çalışmasını sağlamıştır. 5 model arasında SVM modeli, %98,79 doğruluk oranıyla en başarılı model olmuştur (Ahmad, N., Asif, H. M. S., Saleem, G., Younus, M. U., Anwar, S., & Anjum, M. R. (2021)).

Yolov5 ile bitki hastalıkları veri setinde diğer modeller ile karşılaştırılmış ve geliştirilen model %92,97 oranında başarı göstermiştir. Aşırı öğrenmeyi önlemek için veri seti genişletilmiştir. Ayrıca farklı IASM (Geliştirilmiş Dikkat Alt Modülü) gibi özelleştirilen Yolov5 modelleri de kendi aralarında karşılaştırılmıştır (Wang, H., Shang, S., Wang, D., He, X., Feng, K., & Zhu, H. (2022)).

Aşırı öğrenme makinesi (AÖM) modeli kullanılan bir çalışmada hastalıklı domates yaprağı veri setinde özellikleri çıkarılan görüntüler DVM ve karar ağaçları makine öğrenmesi modelleri ile karşılaştırılmıştır. Özellik çıkarımı olarak Hue-Saturation-Value (HSV), Haralick gibi yöntemler kullanılmıştır. Sınıflandırma sonucunda en başarılı model %91,43 doğruluk oranıyla SVM olmuştur (Xian, T. S., & Ngadiran, R. (2021)).

Bitki hastalıkları veri setinin kullanıldığı Bilecik Şeyh Edebali Üniversitesi'nde yapılan bir yüksek lisans tezinde geliştirilen bir CNN modeli, VGG16, VGG19, ResNet50, GoogleNet ve DenseNet gibi modeller kullanılmış ve modellerin başarımları karşılaştırılmıştır. Veri setinde bulunan çeşitli bitki türlerine ilişkin performansları ayrı ayrı incelenmiştir. Sonuç olarak geliştirilen CNN modeli, tüm veri setinde %92 oranında başarı göstermiştir. Diğer modellerden en başarılı olan VGG19 modeli ise %95 oranında başarı göstermiştir (Benli, N., S. (2024)).

Çeltik bitkisi ile ilgili yapılan bir çalışmada, derin öğrenme modellerinin performansları ele alınmıştır. Bu çalışmada VGG, ResNet, DenseNet, EfficientNet, Inception ve Xception gibi derin öğrenme modelleri kullanılmıştır. Bu modeller, Paddy Doctor veri setinde değerlendirilmiştir. Çalışma kapsamında ayrıca veri setinde ön işleme, veri arttırma, hiper-parametre optimizasyonu gibi teknikler uygulanmıştır. Çalışma sonucunda EfficientNetv2 modeli %98,01 doğruluk oranıyla diğer modelleri geride bırakmıştır (Veziroglu, E., Pacal, I., & Coşkunçay, A. (2023)).

Türkoğlu ve diğerlerinin çalışmasında (2020), kayısı bitkisi veri seti kullanılmıştır. Veriler İHA (İnsansız Hava Aracı) ve fotoğraf makinesiyle toplanmıştır. Çalışmada modelin evrişim filtrelerinin boyutları değiştirilerek incelenmiştir. 3×3, 5×5, 7×7, 9×9 ve 11×11 gibi filtreler kullanılmıştır. En iyi sonuç, 9×9 olan filtrelerde %98,2 doğruluk oranı olarak elde edilmiştir. Ayrıca Gradyanların Histogramı, Yerel İkili Örüntüler, Gri Seviye Eş-Oluşum Matrisleri gibi yöntemler kullanılarak elde edilen niteliklerle DVM kullanılarak performansları incelenmiştir. En yüksek doğruluk oranı (%89,64), Yerel İkili Örüntüden elde edilmiştir (Türkoğlu, M., Hanbay, K., Sivrikaya, I. S., & Hanbay, D. (2020)).

Tez çalışması, literatürdeki çalışmalarda kullanılan GMÖ ve DÖ modellerinin performanslarını incelemek amacıyla bitki hastalıklarının teşhisinde incelenmiştir. Elde edilen sonuçlara bakıldığında DÖ modelleri, GMÖ modellerinden daha iyi sonuçlar göstermiştir. Bitki hastalıklarının tespitinde ve tarımda ürün kayıplarını en aza indirmek ve ürünlerde meydana gelen hastalıkları azaltmak için DÖ modellerinin kullanılması gerektiği vurgulanmıştır.

1.2. GMÖ ve DÖ Modellerinin Yöntemsel Olarak Karşılaştırması

GMÖ ve DÖ modelleri farklı özelliklere ve kullanım alanlarına sahiptir. Görüntülerden elde edilen özellikler, kullanılan veri setinin boyutu, çalışma yapılırken kullanılan birimler ve işlem süresi gibi farklılıkları vardır. Bu karşılaştırmaların detayları aşağıda verilmiştir.

1.2.1. Özellik Çıkartımı

Özellik çıkartımı işlemleri DÖ modellerinde otomatik olarak yürütülür. GMÖ modellerinde ise özellik çıkarma ve çıkarılan özelliklerin seçimi manuel olarak yapılmaktadır. Bu nedenle zaman alıcıdır (Janiesch, C., Zschech, P., & Heinrich, K. (2021)).

1.2.2. Veri Seti Boyutu

DÖ modelleri daha fazla veri ile daha iyi sonuçlar verir. GMÖ modelleri ise küçük ve orta boyutlu verilerde daha başarılı sonuçlar verir (Nasir, V., & Sassani, F. (2021)).

1.2.3. Çalışma Biçimi ve Kullanılan Birimler

Derin öğrenme problemleri uçtan uca çözerken, makine öğrenmesi daha karmaşık problemleri daha küçük görevlere ayırarak çözer. Bu yüzden DÖ modelleri, GPU (Grafik İşlemci Birimi) kullanır (Dargan, S., Kumar, M., Ayyagari, M. R., & Kumar, G. (2020)).

1.2.4. Çalışma Zamanı ve Doğruluk

Derin öğrenme problemlerinin işlem süreleri, makine öğrenmesi modellerinin işlem sürelerine göre daha uzun sürer ama doğruluk yüzdeleri makine öğrenmesi modellerine göre daha yüksektir (Halbouni, A., Gunawan, T. S., Habaebi, M. H., Halbouni, M., Kartiwi, M., & Ahmad, R. (2022)).

2. MATERYAL METOT

Bu çalışmada, geleneksel makine öğrenmesi modelleri ile derin öğrenme modellerinin bitki hastalıkları veri seti üzerinde performans değerlendirmesi gerçekleştirilmiştir. Google Colab üzerinden T4 GPU kullanılarak sonuçlar alınmıştır. CNN modeli için veri seti Kaggle üzerinde bulunan PlantVillage (Hughes, D., & Salathé, M. (2015) veri setinin görsel olarak şekil, boyut, renk ve yön açısından çoğaltma (augmented) işlemi yapılmıştır. Bu bölümde, çalışmayla ilgili materyal ve yöntem hakkında detaylara yer verilmiştir.

2.1. Veri Seti

Kullanılan veri setinde (http-1), 14 farklı bitki türü ve 38 farklı sınıfa ilişkin yaklaşık 87.000 renkli görüntü bulunmaktadır. Görüntüler 256x256 piksel boyutundadır. Veri kümesinde bulunan görüntüler 3 klasöre ayrılmıştır. Bu klasörler eğitim (train), doğrulama (validation), ve test görüntüleridir. Bu ayrılma oranı %80 eğitim görüntüleri, %20 doğrulama verileridir. Test veri setinde, diğer veri setlerinde bulunmayan ve modelin test edilmesi için kullanılan 33 adet görüntü bulunmaktadır. Tablo 2.1’de eğitim seti ve doğrulama setinde bulunan görüntülere ilişkin, sınıf isimleri ve toplam görüntü sayısı verilmiştir. Şekil 2.1’de veri setinde bulunan hastalıklı ve sağlıklı verilerden örnek görüntüler verilmiştir. Görüntüler olduğu gibi kullanılıp sinir ağından geçirildiğinde yüksek sayısal değerler elde edilir ve bu durum hesaplamayı zorlaştırır. Bu yüzden görüntülerdeki piksel değerleri, 0 ile 1 değerleri arasında normalleştirilmiştir. Bu işlem, her bir piksel değerinin 255 değerine bölünmesiyle gerçekleştirilir.

Tablo 2.1. Eğitim ve doğrulama veri seti

Eğitim ve Doğrulama Setindeki Sınıf İsimleri	Görüntü Sayıları	
	Eğitim Seti	Doğrulama Seti
1- Apple__Apple_scab	2016	504
2- Apple__Black_rot	1987	497
3- Apple__Cedar_apple_rust	1760	440
4- Apple__healthy	2008	502
5- Blueberry__healthy	1816	454
6- Cherry_(including_sour)__Powdery_mildew	1683	421
7- Cherry_(including_sour)__healthy	1826	456
8- Corn_(maize)__Cercospora_leaf_spot Gray_leaf_spot	1642	410
9- Corn_(maize)__Common_rust_	1907	477
10-Corn_(maize)__Northern_Leaf_Blight	1908	477
11-Corn_(maize)__healthy	1859	465
12-Grape__Black_rot	1888	472
13-Grape__Esca_(Black_Measles)	1920	480
14-Grape__Leaf_blight_(Isariopsis_Leaf_Spot)	1722	430
15-Grape__healthy	1692	423
16-Orange__Haunglongbing_(Citrus_greening)	2010	503
17-Peach__Bacterial_spot	1838	459
18-Peach__healthy	1728	432
19-Pepper,_bell__Bacterial_spot	1913	478
20-Pepper,_bell__healthy	1988	497
21-Potato__Early_blight	1939	485
22-Potato__Late_blight	1939	485
23-Potato__healthy	1824	456
24-Raspberry__healthy	1781	445
25-Soybean__healthy	2022	505
26-Squash__Powdery_mildew	1736	434
27-Strawberry__Leaf_scorch	1774	444
28-Strawberry__healthy	1824	456
29-Tomato__Bacterial_spot	1702	425
30-Tomato__Early_blight	1920	480
31-Tomato__Late_blight	1851	463
32-Tomato__Leaf_Mold	1882	470
33-Tomato__Septoria_leaf_spot	1745	436
34-Tomato__Spider_mites Two-spotted_spider_mite	1741	435
35-Tomato__Target_Spot	1827	457
36-Tomato__Tomato_Yellow_Leaf_Curl_Virus	1961	490
37-Tomato__Tomato_mosaic_virus	1790	448
38-Tomato__healthy	1926	481



Şekil 2.1. Veri setinde bulunan örnek bitki görüntüleri

2.2. Metot

Bu bölümde, tez çalışmasının asıl amacı olan bitkilerin yapraklarında oluşan hastalıkların makine öğrenmesi ve derin öğrenme modelleri yardımıyla teşhisinde kullanılan metotlar verilmiştir. Çalışmada kullanılan veri setinin elde edilmesi ve verilerin ön işleme süreci, bitki üzerindeki hastalıkların daha yüksek doğrulukla tespiti için özellik çıkarımı işlemleri, sınıflandırma ve elde edilen sonuçların değerlendirilmesi ve karşılaştırılması aşamaları detaylandırılmıştır.

2.2.1. Özellik Çıkartımı

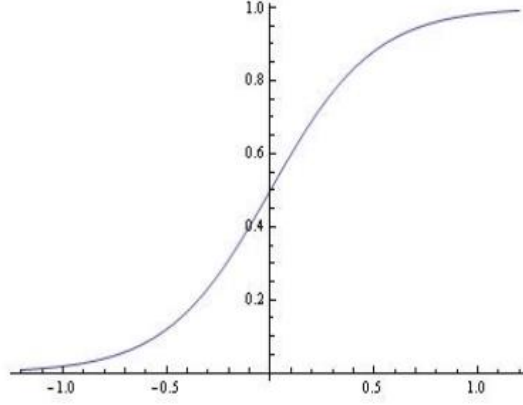
Bu aşamada, görüntü işleme yöntemleri kullanılmıştır. Makine öğrenmesi modellerinde manuel yapılan bu işlem, derin öğrenme modellerinde kendiliğinden gerçekleşir. Çalışmada, HSV, GLCM, HOG ve gri görüntüsünün alınması gibi yöntemler denenmiştir. Bu yöntemler, yaprakta meydana gelen hastalık bölgelerinin daha kolay tespit edilmesine ve sınıflandırılmasına yardımcı olmaktadır.

2.2.2. Sınıflandırma

Sınıflandırma işlemi, eğitim veri seti kullanılarak eğitilen modelin daha önce eğitimde kullanılmayan örneklerin tahmin edilmesidir. Burada tahmin edilen hastalıklar, daha önceden etiketlenmiş verilerden yararlanılarak sınıflandırılır. Makine öğrenmesi modelleri olarak LR, DVM, RO, uzun kısa süreli bellek ve derin öğrenme modelleri olarak VGG16, YOLOv8 ve CNN modelleri tercih edilmiştir. Tez çalışmasında kullanılmış olan GMÖ ve DÖ modellerinin açıklamaları aşağıda verilmiştir.

2.2.3. Lojistik Regresyon

Lojistik Regresyon (LR), genelde ikili sınıf problemlerinde tahmin için kullanılan bir regresyon yöntemidir. Bir karakteristik vektör yardımıyla bir örneğin hangi iki gruba ait olduğunu yinelemeli bir algoritma kullanarak maksimum benzerlik (maximum likelihood) tahminiyle elde eder (Green, G. H., Boze, B. V., Choundhury, A. H., & Power, S. (1998)). Sınıflandırma yapabilmek için bağımsız değişkenleri 0 ile 1 arasında değere eşleyen sigmoid (lojistik) fonksiyonunu kullanır. Bu fonksiyondan gelen çıktı değeri ile belirlenen eşik değerine göre sınıflandırma yapılmaktadır.



Şekil 2.2. Lojistik Sigmoid Fonksiyonu

Kaynak: (Kyurkchiev, N., & Markov, S., 2015)

$$\sigma(z) = \frac{1}{1+\exp(-z)} \quad (2.1)$$

$$z = w.x + b \quad (2.2)$$

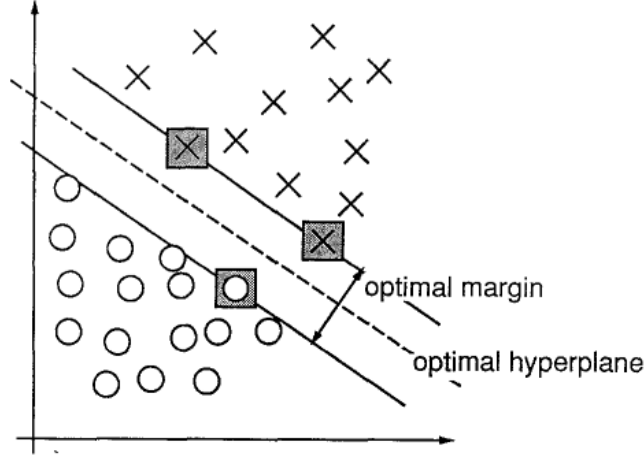
Denklem 2.1 sigmoid fonksiyonudur. Denklem 2.2'deki z değeri eğitim verilerinden gelen x (bağımsız değişken) ağırlık vektörleri, w ağırlıkları (weights) ve b önyargı (bias) teriminden oluşur. İkili sınıflandırma problemde 0 ve 1 çıktı değerleri Y (bağımlı değişken) olasılık değeri için eşitlik Denklem 2.3 ve 2.4 gibi olacaktır.

$$P(Y = 1) = \sigma(w.x + b) \quad (2.3)$$

$$P(Y = 0) = 1 - \sigma(w.x + b) \quad (2.4)$$

2.2.4. Destek Vektör Makinesi

Destek Vektör Makinesi (DVM), sınıflandırma ve regresyon gibi görevlerde kullanılabilen bir makine öğrenmesi algoritmasıdır. İkili sınıflandırma problemlerinde her iki kategoriden işaretlenmiş eğitim verileri verilir ve DVM algoritması destek vektörleri yardımıyla iki sınıf arasındaki en başarılı ayrımı yaparak sınıfları birbirinden ayırır. Bu ayrımı, marjin adı verilen ayrımı yapan vektörlerin geçtiği noktalara denir. Şekil 2.3.'te detaylı gösterimi verilmiştir.



Şekil 2.3. Destek vektörlerinin iki sınıf arasındaki ayrımı yapması

Kaynak: (Cortes, C., & Vapnik, V.,1995)

Hiper düzlemlerin (hyperplane), en uygun olan hiper düzlemden (optimal hyperplane) uzaklıkları Denklem 2.5'te verilen denklemdeki gibi ifade edilir.

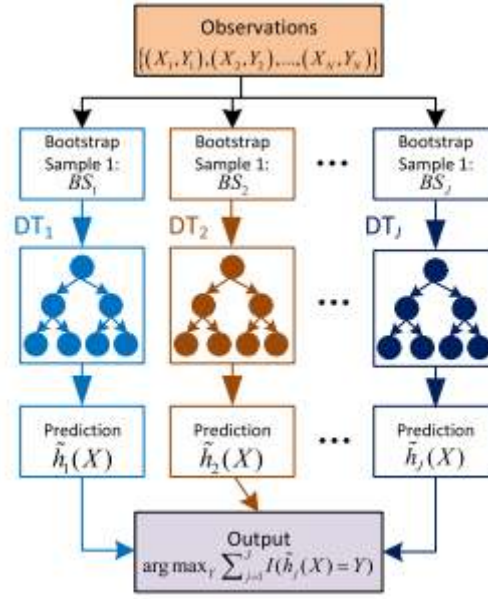
$$\frac{b}{||w||} \quad (2.5)$$

Modelin başarılı olabilmesi için en yakın veriden geçen hiper düzlemler arasında uzaklığın en yükseğe çıkarılması için $||w||$ ifadesinin en aza indirilmesi gerekir. Genel Denklem 2.6'da verilmiştir.

$$y_i (w^T x_i - b) \geq 1, \quad 1 \leq i \leq n \quad (2.6)$$

2.2.5. Rastgele Orman

Rastgele Orman (RO), algoritması birbirinden bağımsız olarak çalışan birden fazla karar ağacının aralarından en çok oy olan ağacın seçilmesi ile sınıflandırma yapar. Şekil 2.4.'te detaylı gösterimi verilmiştir.



Şekil 2.4. Rastgele Orman sınıflandırma modelini yapısı

Kaynak: (Liu vd., 2021)

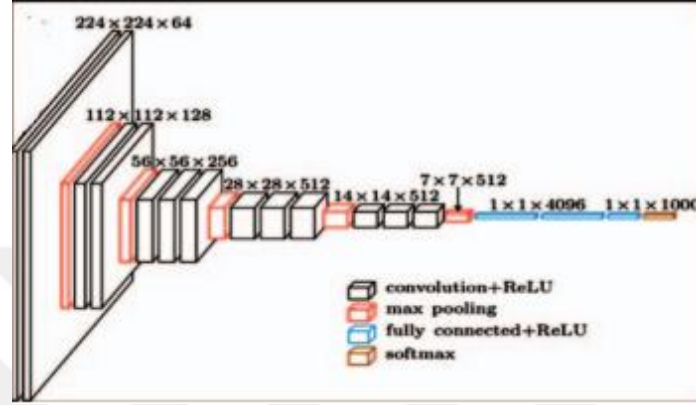
Gözlem (observation) yani kullanılan veri setinden rastgele alınan önyükleme örnekleri (bootstrap sample, BS) ile farklı noktalarına odaklanıp daha iyi öğrenme gerçekleştirecektir. Alınan önyükleme örnekleri her bir karar ağacına verilerek yapılır. Bu yöntem torbalama (bagging) yöntemidir. Rastgele ormanda bir torbalama yöntemidir. Aradaki fark ise her ağacın düğüm noktalarında özellik seçiminin rastgele olmasıdır ve bütün veri setinin kullanılmamasıdır.

2.2.6. LSTM

Uzun Kısa Süreli Bellek anlamına gelen (Long Short Term Memory, LSTM), dizilerdeki bağımlılıkları modellemek için kullanılan bir tür tekrarlayan sinir ağıdır (Recurrent Neural Network, RNN) (Tatsunami, Y. ve Taki, M. (2022)). DÖ modellerinde, hataların kullanılmasını içeren geri yayılım, kaybolan gradyan sorununa yol açmıştır. Bu sorun, önceki katmanların öğrenmesini zorlaştırır. RNN modelleri daha önce görülen girdilerin bilgilerini bir sonraki adıma taşıyabilmektedir. Karmaşık veri setlerinde önceki veriden gelen bilgiyi eksik veya yanlış hatırlama gibi hatalar ile karşılaşmaktadır. RNN modellerindeki bu hatalara çözüm olarak, LSTM modelleri bellek hücreleriyle donatılmıştır. Bellek hücreleri sayesinde standart RNN modelinden farklı olarak kısa vadeli bellek problemine çözüm olacaktır. (Karakuş, B. A. (2022)).

2.2.7. VGG16

VGG16 modeli 13 konvolüsyon 3 tam bağlı katmandan oluşan bir mimaridir. Dropout, katmanı, havuzlama katmanı (pooling), ReLU ve Softmax fonksiyonlarının olduğu katmanlar ile birlikte 41 katmandan oluşmaktadır (Doğan, F., & Türkoğlu, İ.(2018)). Conv1 katmanı giriş verisi olarak 224x224x3 boyutunda renkli bir görüntü alır (Kaur, T., & Gandhi, T. K. (2019)). Şekil 2.5.'te VGG16 mimarisi yer almaktadır.

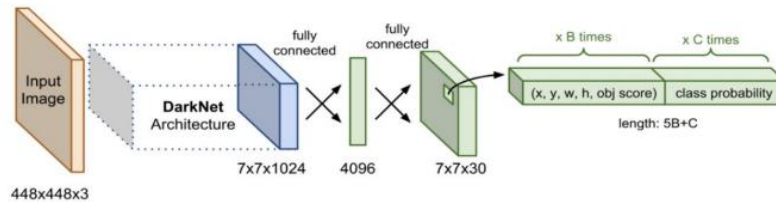


Şekil 2.5. VGG16 mimarisi

Kaynak: (Kaur, T., & Gandhi, T. K., 2019)

2.2.8. YOLOv8

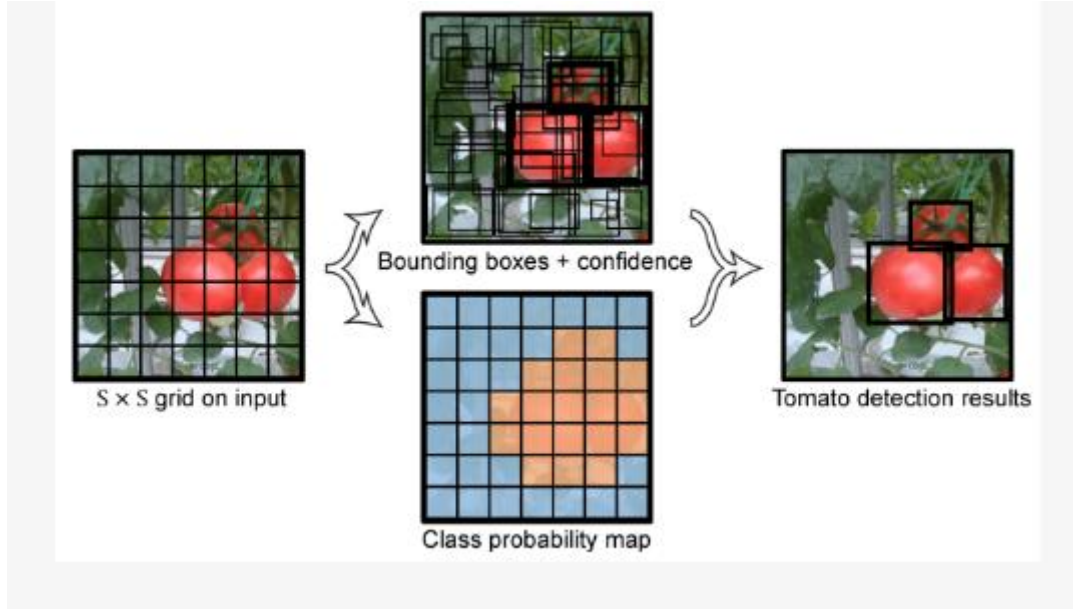
YOLO (You Only Look Once), gerçek zamanlı nesne algılama ve sınıflandırma gerçekleştiren evrişimli sinir ağlarına dayalı bir algoritmadır. Diğer nesne algılama algoritmalarından farklı olarak, YOLO giriş verilerini sinir ağı üzerinden bir kez işler ve her nesne için sınırlayıcı kutuları ve güven puanlarını tahmin etmek üzere ızgaralara böler. Nesnenin sınıfı güven puanına göre belirlenir. (Shinde, S., Kothari, A., & Gupta, V. (2018)). Şekil 2.6.'da YOLO mimarisinin yapısı açıklamalı şekilde gösterilmiştir.



Şekil 2.6. Genel YOLO mimarisi

Kaynak: (Pham, T. L., & Le, V. H., 2024)

Şekil 2.7.'de, her sınırlayıcı kutu bir güven puanına karşılık gelir. Nesne bulunmazsa, bu puan sıfır olur. Nesne bulunursa, sınırlayıcı kutu bulunan nesneyle kesişir ve şu anda 0,5 olan gerçek IoU (Kesişim/Birleşme) değerini gösterir. Bu değeri artırmak daha kesin sonuçlar sağlayabilir. Bu eşik değerinin altındaki veriler dikkate alınmayacaktır. (Wu, J. D., Chen, B. Y., Shyr, W. J., & Shih, F. Y. (2021)).



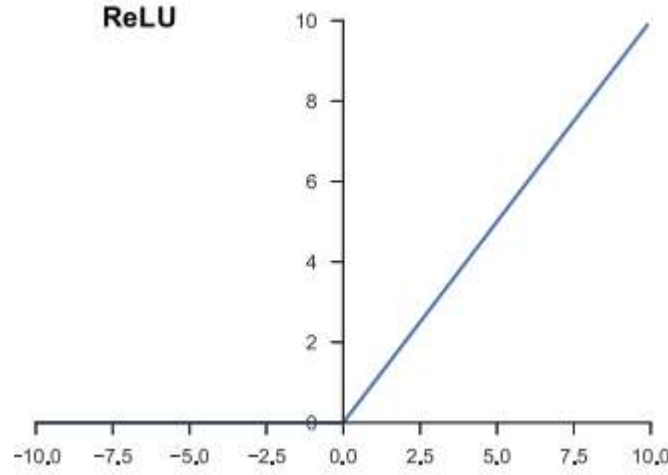
Şekil 2.7. YOLO nesne tespit etme akış şeması

Kaynak: (Liu vd.,2020)

2.2.9. CNN

Evrişimli Sinir Ağı (Convolutional Neural Network, CNN), evrişim katmanı (convolution layer), havuzlama katmanı (pooling layer) ve tam bağlantılı katman (full connected layer) gibi ana bölümlerden oluşan yapay sinir ağı türüdür. Görüntü işleme ve sınıflandırma alanlarında sıklıkla kullanılır.

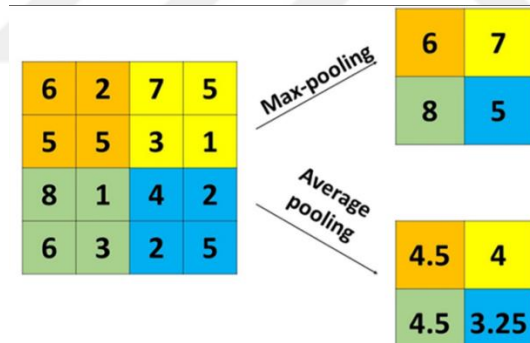
Evrişim Katmanı: Bu katman görüntülerden özellik çıkarımının yapıldığı katmandır. Resimden özelliklerin çıkarılması için görüntüye seçilen boyutlarda (3×3, 5×5, 7×7) filtre uygular. Filtrenin kaç piksel kaydırılacağı ise adım sayısı (stride) ile belirlenir. Bunun sonucunda, özellik haritası (feature map) boyutu girdi görüntüsünün boyutu ile farklı olacaktır. Evrişim işlemi boyunca aynı düzlemin korunması için doldurma (padding) işlemi yapılır. Elde edilen çıktının bir sonraki nörona verilebilmesi için aktifleştirilmesi gerekir, aksi takdirde lineer (doğrusal) bir modelden farkı olmayacaktır. En yaygın olarak ReLU aktivasyon fonksiyonu (Şekil 2.8.) kullanılır (Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018)).



Şekil 2.8. ReLU fonksiyonu

Kaynak: (Yamashita vd.,2018)

Havuzlama Katmanı: Özellik haritalarının boyutunu azaltan temel katmandır. Adım sayısı ve filtre boyutu önemli değişkenleridir. Önemli verilerin alınması ve değersiz verilerin atılmasıyla yapılır. Birçok havuzlama yöntemi vardır. Bunlar maxpooling, average pooling'dir (A. Mohamed, E., Gaber, T., Karam, O., & Rashed, E. A. (2022)) Şekil 2.9.'da havuzlama çeşitleri örneklendirilmiştir.



Şekil 2.9. Havuzlama çeşitleri verilmiştir. Filtre boyutu 2×2 ve adım sayısı 2 verilmiştir.

Kaynak: (Mohamed vd., 2022)

Tam Bağlı Katman: Bu katman evrişim ve havuzlama katmanlarından sonra gelen en son katmandır. Bu katmanda elde edilen özellik haritaları bir boyutlu sayı dizisine dönüştürülür. Giriş verileri, öğrenebilen ağırlıklar ile tam bağlantılı katmana bağlanır. Diğer bir adıyla yoğun (dense) katmanına bağlanır. Bir sınıflandırma problemi için düşünülürse olasılıklar her bir sınıf ile eşlenir. Eşitlenen olasılıklar karışıklık matrisi, ROC eğrisi gibi modelin performansını ölçebilecek ifadeler ile verilir (Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K., 2018)).

2.2.10. Performans Değerlendirme

Karışıklık matrisi, sınıflandırma yapan modelin her bir sınıf için performansının incelendiği bir matristir (Kapucuoğlu, K. (2022)). Şekil 2.10.'da en basit hali ile karışıklık matrisi verilmiştir. Modelin performansı, hataları ve zayıflıkları hakkında detaylı bilgiler içerir.

	Gerçek Pozitif Değerler	Gerçek Negatif Değerler
Tahmin Edilen Pozitif Değerler	DP	YP
Tahmin Edilen Negatif Değerler	YN	DN

Şekil 2.10. Karışıklık Matrisi

Karışıklık matrisindeki değerler yaptığımız çalışma için şu şekilde açıklanır:

DP (True Positive, Doğru Pozitif): Gerçekte hastalıklı olan bitkilerin hastalıklı olarak doğru tahmin edilmesini ifade eder.

DN (True Negative, Doğru Negatif): Gerçekte sağlıklı olan bitkilerin sağlıklı olarak doğru tahmin edilmesini ifade eder.

YP (False Positive, Yanlış Pozitif): Gerçekte hastalıklı olan bitkilerin sağlıklı olarak yanlış tahmin edilmesini ifade eder.

YN (False Negative, Yanlış Negatif): Gerçekte sağlıklı olan bitkilerin hastalıklı olarak yanlış tahmin edilmesini ifade eder.

Performans değerlendirmesi için esas alınan en önemli metriklerden olan doğruluk (accuracy) esas alınır. Doğruluk değeri, toplam doğru tahminlerin sayısının, toplam tahmin sayısına bölünmesiyle bulunur (Paymode, A. S., & Malode, V. B. (2022)). Denklem 2.7'de doğruluk denklemi verilmiştir.

$$\text{Doğruluk} = \frac{(DP+DN)}{(DP+DN+YP+YN)} \quad (2.7)$$

Tek başına doğruluk değeri, modelin performansı için yeterli değildir. Bu sebeple derin öğrenme modellerinin performans değerlendirmesi için ayrıca kesinlik (precision), duyarlılık (recall) ve bu iki değer harmonik ortalaması olan F-1 skor metrikleri de dikkate

alınmıştır (Paymode, A. S., & Malode, V. B. (2022)). Denklem 2.8, 2.9 ve 3.0'te sırasıyla formülleri verilmiştir.

$$Kesinlik = \frac{DP}{DP+YP} \quad (2.8)$$

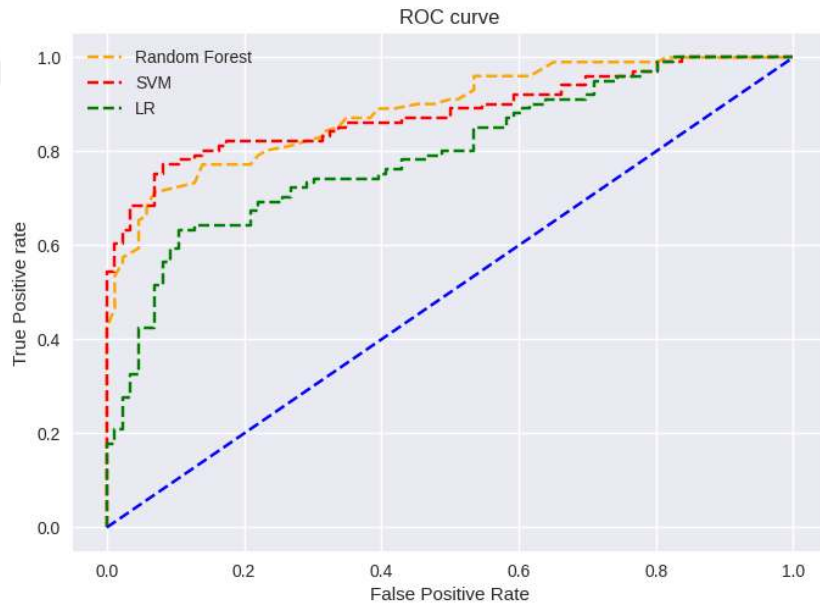
$$Duyarlılık = \frac{DP}{DP+YN} \quad (2.9)$$

$$F1 Skoru = 2 \times \frac{Kesinlik \times Duyarlılık}{Kesinlik + Duyarlılık} \quad (3.0)$$

$$\text{Özgüllük} = \frac{DN}{DN+YP} \quad (3.1)$$

$$1 - \text{Özgüllük}(\text{Specificity}) = \frac{YP}{YP+DN} \quad (3.2)$$

Modellerin performans değerlendirmesini yaparken kullanılan bir diğer yöntem ROC (Receiver Operating Characteristic) eğrisi ve AUC (Area Under Curve) değeridir. AUC ise ROC eğrisinin altında kalan alanı ifade eder. ROC eğrisi, modellerin veri setindeki sınıfları nasıl ayırt edebildiği hakkında bilgi verir.

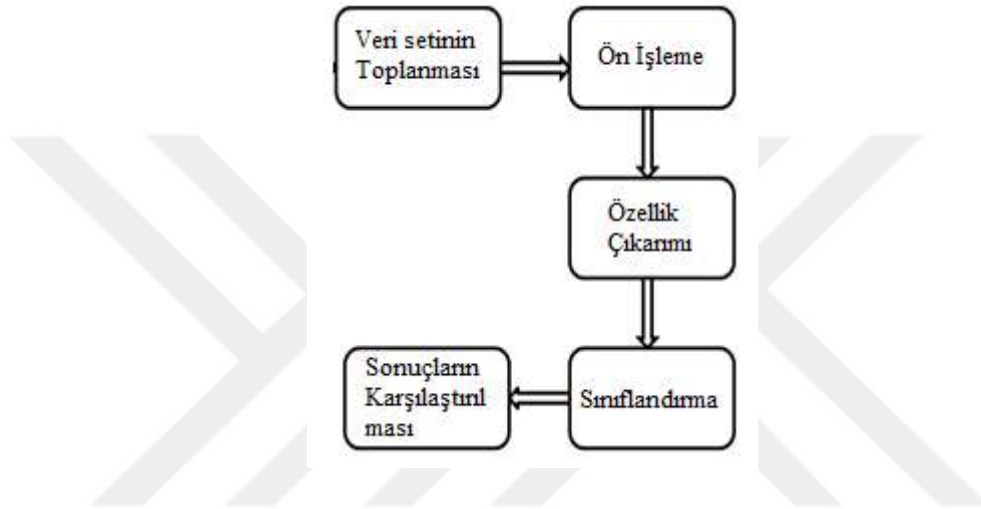


Grafik 2.1 ROC Eğrisi

Grafik 2.1'de GMÖ modellerinden elde edilen grafik verilmiştir. y eksenindeki değerler Denklem 2.9 ile elde edilirken, x eksenindeki değerler Denklem 3.2 ile elde edilmektedir. AUC verisi eğrinin altında kalan alanın hesaplanması ile elde edilir. Grafikte verilen mavi renkli çizgi eşik değeri 0,5 değerindedir. Bu değer, modellerin doğru ya da yanlış veri arasında ayırım yapma yeteneğinin olmadığını belirten sınırdır.

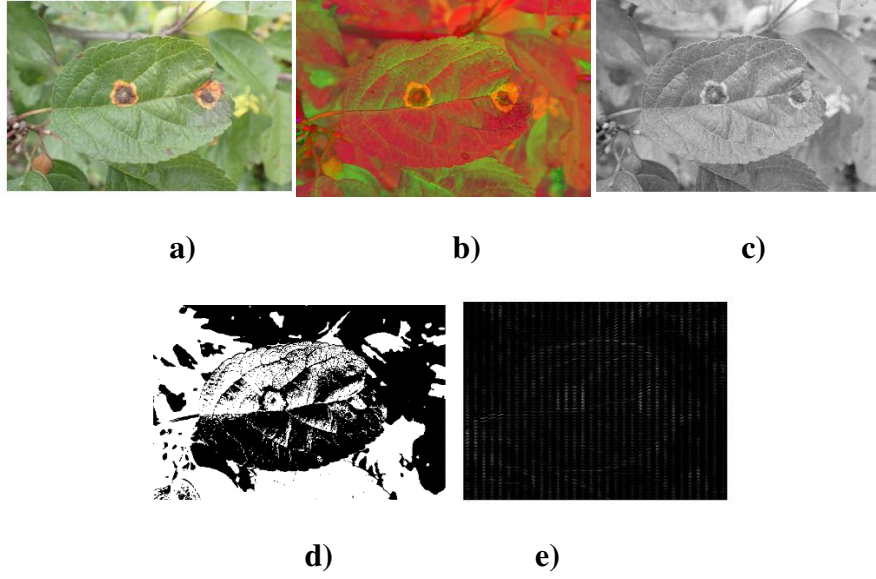
3. DENEYSEL ÇALIŞMALAR VE BULGULAR

Çalışma, Google Colab ortamında kullanılarak gerçekleştirilmiştir. Google Colab'ta bulunan ve güncel olan Python yazılım dili ile gerçekleştirilmiştir. İstatiksel ve sayısal hesaplamalar için numpy ve pandas, verileri görselleştirmek için seaborn ve matplotlib, derin öğrenme modelleri ve katmanları için tensorflow ve scikit-learn kütüphaneleri kullanılmıştır. Geleneksel makine öğrenmesi modellerinden LR, DVM, RO ve DÖ modellerinden LSTM ile keras kütüphanesinden önceden eğitilmiş ImageNet ağırlıkları ile VGG16, YOLOv8 ve CNN'den yararlanılmıştır. Şekil 3.1'de yapılan çalışmanın akış şeması verilmiştir.



Şekil 3.1. Akış Şeması

Görüntülerden elde edilen özelliklerden olan renk, şekil, doku bilgisi gibi veriler eğitim ve test işlemlerinde en iyi performansı elde etmek için kullanılmaktadır. Renk özellikleri olarak kırmızı, yeşil, mavi (Red, Green, Blue, RGB) ve renk tonu, doygunluk, değer (Hue, Saturation, Value, HSV) verilebilir. Bu yöntemlerin kullanılmasının amacı bitki üzerinde hastalıklı bölgelerin bitkinin renginden farklı bir renk ile işaretlemektir. Doku özellikleri olarak gri düzey eş oluşum matrisleri (Gray Level Co-Occurrence Matrix, GLCM) ve yönlü gradyanlar histogramı (Histogram of Oriented Gradients, HOG) verilebilir. Bu yöntemlerin kullanılmasının amacı ise yaprak üzerinde hastalıklı bölgenin yaprağın dokusundan farklı bir dokuyu tespit etmek için yapılmıştır. Yapılan özellik çıkarımı işlemleri Şekil 3.2'de verilmiştir. Bu aşamalardan sonra modellerin performans değerlendirmesi incelenmiştir.



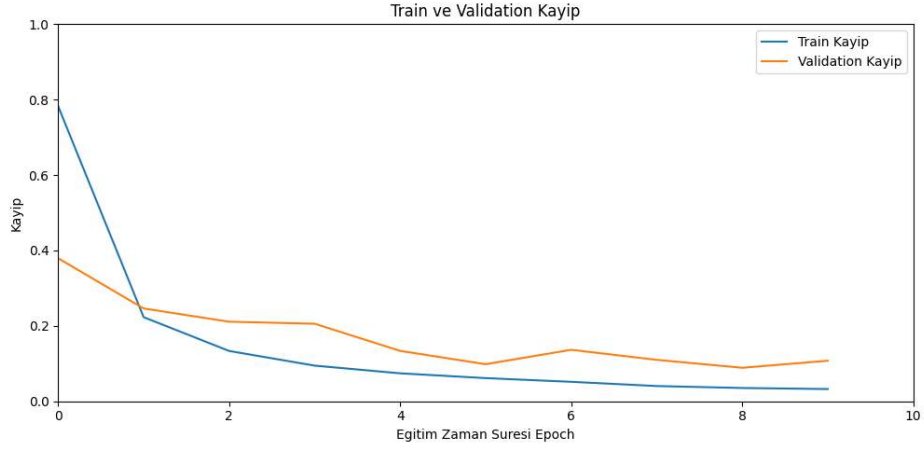
Şekil 3.2. Çeşitli özellik çıkarımı yöntemleri; a)Orijinal Görüntü, b)HSV Görüntüsü, c)Gri Görüntüsü, d)Binary Görüntüsü, e)Yönlü Gradyanlar Histogramı

Yapılan çalışmada GMÖ modelleri olarak LR, DVM, RO ve DÖ modelleri olarak LSTM, VGG16, YOLOv8 ve CNN modelleri kullanılmıştır. Modellerin doğruluk yüzdeleri Tablo 3.1’de verilmiştir.

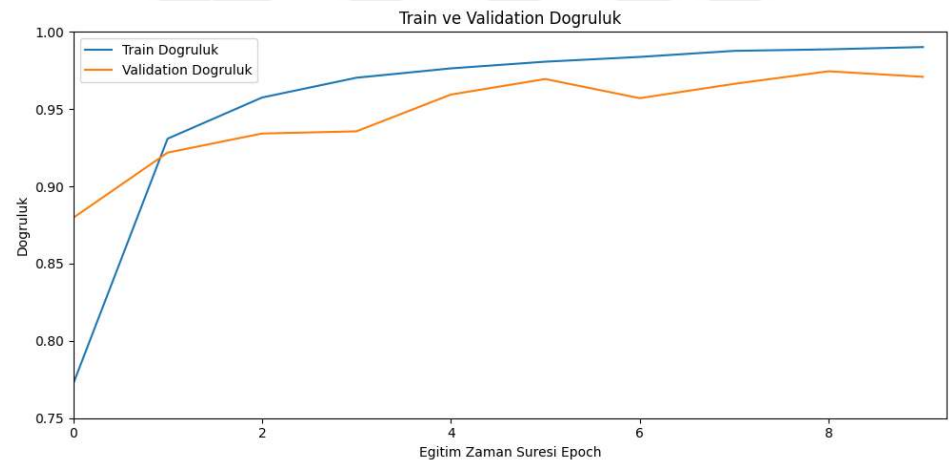
Tablo 3.1. Model doğruluk yüzde verileri

No	Sınıflandırma Çeşidi	Sınıflandırma Modelleri	Ortalama Doğruluk(%)
1	GMÖ	Destek Vektör Makinesi(DVM)	83
2		Rastgele Orman(RO)	78,6
3		Lojistik Regresyon(LR)	72,3
4	DÖ	Uzun-Kısa Vadeli Bellek(LSTM)	81,9
5		VGG16	88
6		YOLOv8	98,3
7		CNN	99

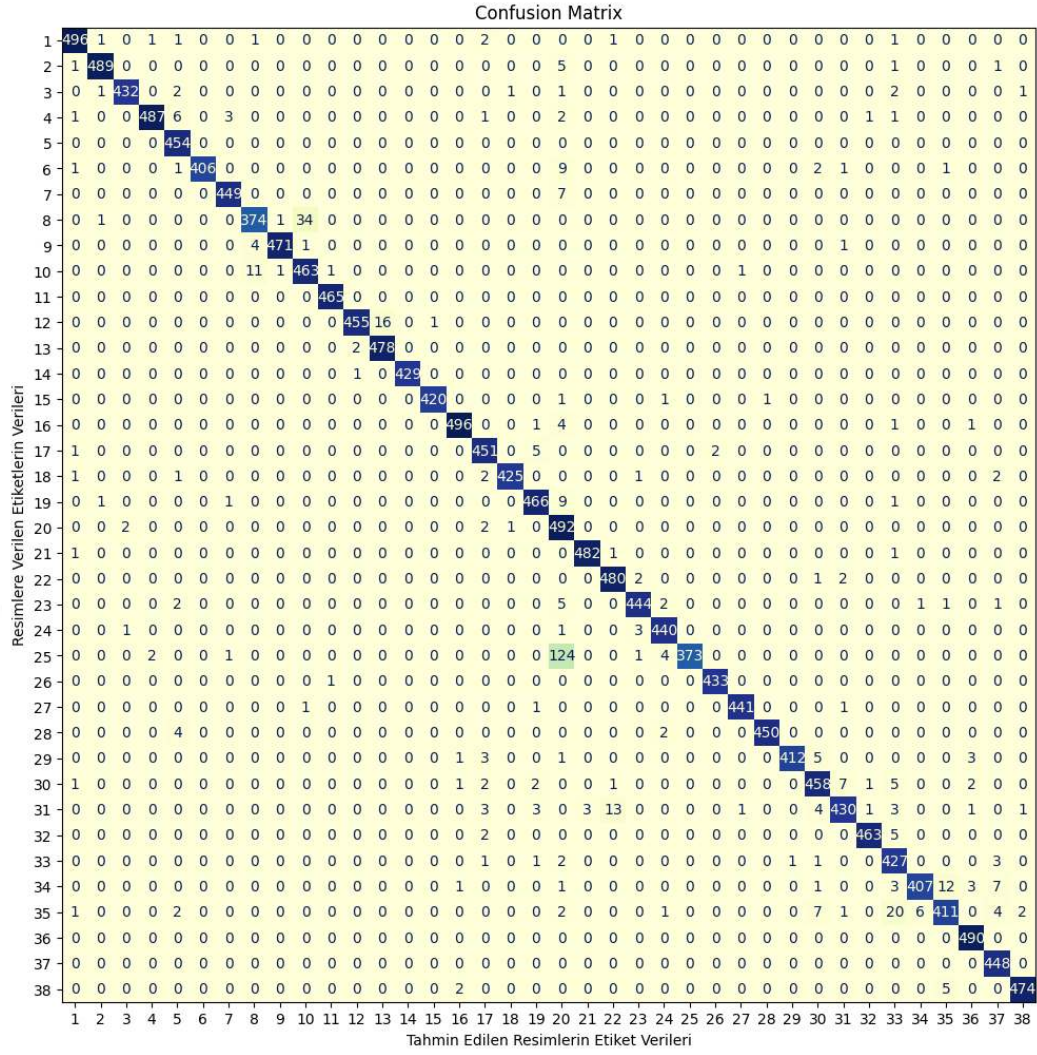
CNN modelinin elde edilen eğitim ve doğrulama işlemlerine ilişkin kayıp ve doğruluk eğrilerinin grafikleri sırasıyla 3.1 ve 3.2’de verilmiştir.



Grafik 3.1. CNN modelinin kayıp (loss) eğrisi



Grafik 3.2. CNN modelinin doğruluk (accuracy) eğrisi



Şekil 3.2. Çalışmada kullanılan modellerin karışıklık matrisi sonuçları

CNN modelinin sınıflandırma sonucunun detaylı incelenebilmesi için karışıklık matrisi Şekil 3.2’de verilmiştir. 14 numaralı sınıf için modelin doğruluk oranı %100, ancak 20 numaralı sınıf için bu oran %75’tir. Bu modelin 14 numaralı sınıfı tahmin etmeden başarılı olduğunu fakat 20 numaralı sınıfı tahmin ederken yanlış tahminde bulunduğunu gösterir. Diğer sınıfların genelinde yüksek doğruluk oranları elde edilmiştir. Tablo 3.1’e bakıldığında geleneksel makine öğrenmesi modellerinin derin öğrenme modellerine göre daha az doğruluk yüzdelerine sahiptir. Bu kullanılan veri setinde bulunan görüntülerin kalitesinin yetersiz olması, özellik çıkarımının manuel olarak yapılması, kullanılan modellerden bazılarının amacına uygun olmayan problemlerde kullanılmaları gibi nedenlerden dolayı doğruluk oranları düşük kalmıştır. Derin öğrenme modellerinde özellikle YOLOv8 ve CNN modelinde istenilen başarılı sonuçlar elde edilmiştir.

4. SONUÇ VE ÖNERİLER

Türkiye'nin tarım alanındaki ithalat verilerinin artması ve son yıllarda artan nüfusun neticesinde beslenme sorunu gibi olumsuz durumlardan kurtulabilmesi için tarımda teknolojiye yararlanmalı ve özellikle de yazılım, makine öğrenmesi, derin öğrenme gibi günümüzün gelişen teknolojilerine ayak uydurması gerekmektedir. İnsan hayatının devamı için tarımda üretilen ürünlerin hastalık belirtileri ve hastalık çeşitlerinin belirlenmesinde derin öğrenme modellerinin kullanılması hastalığın ilk evrelerinde müdahale edilmesi hızlandıracaktır. Makine öğrenmesi modelleri incelendiğinde sınıflandırma problemlerinde yeteri kadar başarılı olmadığı görülmüştür. Manuel özellik çıkarımı, büyük veri setlerinde düşük doğruluk seviyelerinde kalması bu alanda kullanılmasının verimli olmayacağını gösterir. Derin öğrenme modelleri incelendiğinde YOLOv8 ve CNN modelinin sonuçları oldukça başarılıdır.

Günümüzde akıllı telefonlar ile birçok işlem gerçekleştirilebildiği için bitki hastalıklarının tespitine yönelik bir geliştirilecek bir mobil uygulama ile telefonda tespit yapılabilir ve bu sayede daha kısa sürede gerekli müdahale başlatılabilir. Kullanılan veri setindeki hastalık tiplerinin çeşitliliği artırılabilir. Tarımda kullanılan drone benzeri insansız hava araçları sayesinde tespit edilen hastalıkların YOLOv8 gibi gerçek zamanlı olarak sınıflandırma yapılabilir. Farklı derin öğrenme modelleri ve daha iyi veri seti ile sonuçlar karşılaştırılabilir. Seçilen model ile her yerde erişim için web uygulaması yapılabilir. Tespit edilen hastalık sonrası bitki için uzaktan müdahale, ilaçlama vb. sistem gerçekleştirilebilir.

KAYNAKÇA

A. Mohamed, E., Gaber, T., Karam, O., & Rashed, E. A. (2022). A Novel CNN pooling layer for breast cancer segmentation and classification from thermograms. *Plos one*, 17(10), e0276523.

Ahmad, N., Asif, H. M. S., Saleem, G., Younus, M. U., Anwar, S., & Anjum, M. R. (2021). Leaf image-based plant disease identification using color and texture features. *Wireless Personal Communications*, 121(2), 1139-1168.

Benli, N., S. (2024). *Derin evriřimli yapay sinir ağı kullanarak meyve yaprağı hastalık tespiti (Yüksek Lisans Tezi)*. Bilecik Şeyh Edebali Üniversitesi, Fen Bilimleri Enstitüsü.

Chen, J., Zhang, D., Zeb, A., & Nanekaran, Y. A. (2021). Identification of rice plant diseases using lightweight attention networks. *Expert Systems with Applications*, 169, 114514.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20, 273-297.

Dargan, S., Kumar, M., Ayyagari, M. R., & Kumar, G. (2020). A survey of deep learning and its applications: a new paradigm to machine learning. *Archives of Computational Methods in Engineering*, 27, 1071-1092.

Doğan, F., & Türkoğlu, İ. (2018). Derin öğrenme algoritmalarının yaprak sınıflandırma başarımlarının karşılaştırılması. *Sakarya University Journal of Computer and Information Sciences*, 1(1), 10-21.

Green, G. H., Boze, B. V., Choundhury, A. H., & Power, S. (1998). Using logistic regression in classification. *Marketing Research*, 10(3).

Halbouni, A., Gunawan, T. S., Habaebi, M. H., Halbouni, M., Kartiwi, M., & Ahmad, R. (2022). Machine learning and deep learning approaches for cybersecurity: A review. *IEEE Access*, 10, 19572-19585.

Hughes, D., & Salathé, M. (2015). An open access repository of images on plant health to enable the development of mobile disease diagnostics. *arXiv preprint arXiv:1511.08060*.

http-1 <https://www.kaggle.com/datasets/vipooooool/new-plant-diseases-dataset> [Eriřim: 24.09.2024]

- Jadhav, S. B., Udupi, V. R., & Patil, S. B.** (2021). Identification of plant diseases using convolutional neural networks. *International Journal of Information Technology*, 13(6), 2461-2470.
- Janiesch, C., Zschech, P., & Heinrich, K.** (2021). Machine learning and deep learning. *Electronic Markets*, 31(3), 685-695.
- Jayswal, H. S., & Chaudhari, J. P.** (2020). Plant leaf disease detection and classification using conventional machine learning and deep learning. *International Journal on Emerging Technologies*, 11(3), 1094-1102.
- Kapucuoğlu, K.** (2022). *Derin öğrenme ağları kullanılarak doğal ortamda hastalıklı domateslerin belirlenmesi (Yüksek Lisans Tezi)*. İstanbul Teknik Üniversitesi, Lisansüstü Eğitim Enstitüsü.
- Karakuş, B. A.** (2022). *Derin öğrenme ve büyük veri yaklaşımları ile metin analizi (Doktora Tezi)*. Fırat Üniversitesi, Fen Bilimleri Enstitüsü.
- Kaur, T., & Gandhi, T. K.** (2019). Automated brain image classification based on VGG-16 and transfer learning. In *2019 international conference on information technology (ICIT)* (pp. 94-98). IEEE.
- Kukreja, V., & Dhiman, P.** (2020). A Deep Neural Network based disease detection scheme for Citrus fruits. In *2020 International conference on smart electronics and communication (ICOSEC)* (pp. 97-101). IEEE.
- Kyurkchiev, N., & Markov, S.** (2015). Sigmoid functions: some approximation and modelling aspects. *LAP LAMBERT Academic Publishing, Saarbrücken*, 4.
- Liu, G., Nouaze, J. C., Touko Mbouembe, P. L., & Kim, J. H.** (2020). YOLO-tomato: A robust algorithm for tomato detection based on YOLOv3. *Sensors*, 20(7), 2145.
- Liu, K., Hu, X., Zhou, H., Tong, L., Widanage, W. D., & Marco, J.** (2021). Feature analyses and modeling of lithium-ion battery manufacturing based on random forest classification. *IEEE/ASME Transactions on Mechatronics*, 26(6), 2944-2955.
- Nasir, V., & Sassani, F.** (2021). A review on deep learning in machining and tool monitoring: Methods, opportunities, and challenges. *The International Journal of Advanced Manufacturing Technology*, 115(9), 2683-2709.

- Paymode, A. S., & Malode, V. B.** (2022). Transfer learning for multi-crop leaf disease image classification using convolutional neural network VGG. *Artificial Intelligence in Agriculture*, 6, 23-33.
- Pham, T. L., & Le, V. H.** (2024). Ovarian Tumors Detection and Classification from Ultrasound Images Based on YOLOv8. *Journal of Advances in Information Technology*, 15(2).
- Sardogan, M., Tuncer, A., & Ozen, Y.** (2018). Plant leaf disease detection and classification based on CNN with LVQ algorithm. In *2018 3rd international conference on computer science and engineering (UBMK)* (pp. 382-385). IEEE.
- Shinde, S., Kothari, A., & Gupta, V.** (2018). YOLO based human action recognition and localization. *Procedia computer science*, 133, 831-838.
- Sofuoğlu, C.** (2024). *Plant leaf disease detection and classification using deep learning (Yüksek Lisans Tezi)*. Dokuz Eylül Üniversitesi, Fen Bilimleri Enstitüsü.
- Tan, L., Lu, J., & Jiang, H.** (2021). Tomato leaf diseases classification based on leaf images: a comparison between classical machine learning and deep learning methods. *AgriEngineering*, 3(3), 542-558.
- Tatsunami, Y., & Taki, M.** (2022). Sequencer: Deep lstm for image classification. *Advances in Neural Information Processing Systems*, 35, 38204-38217.
- Türkoğlu, M., Hanbay, K., Sivrikaya, I. S., & Hanbay, D.** (2020). Derin Evrişimsel Sinir Ağı Kullanılarak Kayısı Hastalıklarının Sınıflandırılması. *Bitlis Eren Üniversitesi Fen Bilimleri Dergisi*, 9(1), 334-345.
- Veziroğlu, E., Pacal, I., & Coşkunçay, A.** (2023). Derin Evrişimli Sinir Ağları Kullanılarak Pirinç Hastalıklarının Sınıflandırılması. *Journal of the Institute of Science and Technology*, 13(2), 792-814.
- Wang, H., Shang, S., Wang, D., He, X., Feng, K., & Zhu, H.** (2022). Plant disease detection and classification method based on the optimized lightweight YOLOv5 model. *Agriculture*, 12(7), 931.

Wang, P., Fan, E., & Wang, P. (2021). Comparative analysis of image classification algorithms based on traditional machine learning and deep learning. *Pattern recognition letters*, 141, 61-67.

Wu, J. D., Chen, B. Y., Shyr, W. J., & Shih, F. Y. (2021). Vehicle Classification and Counting System Using YOLO Object Detection Technology. *Traitement du Signal*, 38(4).

Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights into imaging*, 9, 611-629.

Yilmaz, A., Demircali, A. A., Kocaman, S., & Uvet, H. (2020). Comparison of deep learning and traditional machine learning techniques for classification of pap smear images. *arXiv preprint arXiv:2009.06366*.