DOKUZ EYLÜL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

DETECTION AND CLASSIFICATION OF OLIVE QUALITY AND DISEASES BY DEEP LEARNING METHODS

by
Cengiz Mehmet ALBOYACI

February, 2025
İZMİR

DETECTION AND CLASSIFICATION OF OLIVE QUALITY AND DISEASES BY DEEP LEARNING METHODS

A Thesis Submitted to the
Graduate School of Natural And Applied Sciences of Dokuz Eylül University
In Partial Fulfillment of the Requirements for the Degree of Master of Computer Science, Computer Science Program

by
Cengiz Mehmet ALBOYACI

February, 2025
İZMİR

M.Sc. THESIS EXAMINATION RESULT FORM

We have read the thesis entitled "**DETECTION AND CLASSIFICATION OF OLIVE QUALITY AND DISEASES BY DEEP LEARNING METHODS**" completed by **CENGIZ MEHMET ALBOYACI** under supervision of **ASST. PROF. CAN ATILGAN** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

....

Asst. Prof. Can ATILGAN

Supervisor

....

Prof. Dr. Murat Erşen BERBERLER

Jury Member

....

Asst. Prof. Kazım ERDOĞDU

Jury Member

Prof.Dr. Abdullah SEÇGİN
Director
Graduate School of Natural and Applied Sciences

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my thesis advisor, Asst. Prof. Can ATILGAN, for his patient guidance and invaluable support throughout this journey. His insights and encouragement have been instrumental in shaping this work.

I extend my deepest gratitude to my dear Aylin GÖÇOĞLU for her unwavering support and encouragement throughout this journey. Her motivation, patience, and belief in me have made this experience even more meaningful. I would also like to thank Semih MEMİŞ for our technical discussions, which have greatly contributed to my curiosity for knowledge.

Special thanks to my family members, especially my mother, for shaping me into the person I am today.

Cengiz Mehmet ALBOYACI

DETECTION AND CLASSIFICATION OF OLIVE QUALITY AND DISEASES BY DEEP LEARNING METHODS

ABSTRACT

The increasing global demand for food, driven by climate change and pandemics, poses significant risks to food security. Addressing these challenges in agriculture has become more critical than ever. Recent studies have focused on classifying plant diseases using advanced image processing techniques. While these models often achieve high accuracy in controlled laboratory environments, maintaining similar performance in real-world conditions remains a significant challenge. The study introduces an automated classification approach for Peacock Eye Disease in olive leaves by integrating deep learning and anomaly detection techniques. A Convolutional Neural Network (CNN) trained on a well-annotated dataset from a controlled laboratory setting was evaluated using real-world leaf images. However, the model exhibited limitations when classifying field-collected leaves due to the complexities of real-world conditions.

To address this, leaf segmentation was performed using the Segment Anything Model (SAM), followed by anomaly detection through an autoencoder to remove non-leaf elements. The laboratory-based CNN model was then tested on this refined dataset, and its results were analyzed. Finally, a new CNN model was trained using the manually labeled, refined dataset, and its classification performance was compared with previous methods.

The results demonstrated that the dataset filtered by the autoencoder achieved accuracy levels comparable to the dataset used to train the laboratory-based CNN model. These findings indicate that automated segmentation and filtering techniques can significantly reduce the need for manual annotation while maintaining classification accuracy. The study underscores the challenges of applying models trained in controlled laboratory settings to real-world conditions. It further emphasizes the critical role of automated preprocessing techniques in enhancing the

scalability and reliability of AI-driven agricultural disease classification.

Keywords: Olive disease detection and classification, segment anything model, convolutional neural network, peacock eye disease

ZEYTİN KALİTESİ VE HASTALIKLARININ DERİN ÖĞRENME YÖNTEMLERİ İLE BELİRLENMESİ VE SINIFLANDIRILMASI

ÖZ

İklim değişikliği ve pandemilerin tetiklediği artan küresel gıda talebi, gıda güvenliği açısından önemli riskler oluşturmaktadır. Tarımda bu zorluklarla başa çıkmak, her zamankinden daha kritik bir hale gelmiştir. Son çalışmalar, gelişmiş görüntü işleme teknikleri kullanarak bitki hastalıklarının sınıflandırılmasına odaklanmıştır. Bu modeller, kontrollü laboratuvar ortamlarında yüksek doğruluk oranlarına ulaşsa da, gerçek dünya koşullarında benzer performansı sürdürmek önemli bir zorluk olmaya devam etmektedir.

Bu çalışma, derin öğrenme ve anomali tespiti tekniklerini entegre ederek zeytin yapraklarındaki Tavus Gözü Hastalığı'nı otomatik olarak sınıflandırmak için bir yöntem sunmaktadır. Kontrollü bir laboratuvar ortamında iyi bir şekilde etiketlenmiş veri seti üzerinde eğitilmiş bir Evrişimli Sinir Ağı (CNN), gerçek dünya yaprak görüntüleri kullanılarak değerlendirilmiştir. Ancak model, gerçek dünya koşullarının karmaşıklığı nedeniyle tarladan toplanan yaprakları sınıflandırmada sınırlamalarla karşılaşmıştır.

Bu sorunu çözmek için, Segment Anything Model (SAM) kullanılarak yaprak segmentasyonu gerçekleştirilmiş ve ardından bir otokodlayıcı (autoencoder) ile anomali tespiti yapılarak yaprak olmayan unsurlar temizlenmiştir. Laboratuvar temelli CNN modeli, bu iyileştirilmiş veri seti üzerinde test edilmiş ve sonuçlar analiz edilmiştir. Son olarak, manuel olarak etiketlenmiş ve iyileştirilmiş veri seti kullanılarak yeni bir CNN modeli eğitilmiş ve sınıflandırma performansı önceki yöntemlerle karşılaştırılmıştır.

Sonuçlar, otokodlayıcı tarafından filtrelenen veri setinin, laboratuvar temelli CNN modelini eğitmek için kullanılan veri setiyle karşılaştırılabilir doğruluk seviyelerine ulaştığını göstermiştir. Bu bulgular, otomatik segmentasyon ve filtreleme tekniklerinin, sınıflandırma doğruluğunu korurken manuel etiketleme ihtiyacını önemli ölçüde azaltabileceğini ortaya koymaktadır. Bu çalışma, kontrollü laboratuvar ortamlarında eğitilmiş modellerin gerçek dünya koşullarına uygulanmasındaki zorlukları vurgulamaktadır. Ayrıca, otomatik ön işleme tekniklerinin, tarımda yapay zeka destekli hastalık sınıflandırmasının ölçeklenebilirliğini ve güvenilirliğini artırmadaki kritik rolünü vurgulamaktadır.

Anahtar kelimeler: Zeytin hastalıklarının tespiti ve sınıflandırılması, tümünü bölütleme modeli, evrişimli sinir ağları, kuş gözü hastalığı

CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

CHAPTER ONE
INTRODUCTION

The global population is anticipated to surpass 9 billion by 2050, leading to a predicted increase in food demand of over 70%, as the Food and Agriculture Organization (FAO) reported (FAO (2009)). This extraordinary surge in demand exerts significant pressure on global agricultural systems to enhance production while preserving sustainability. Confronting this challenge is essential for attaining food security and mitigating the danger of famine, especially in areas where agriculture underpins the local economy. Nonetheless, other issues, such as climate change, restricted arable land, and the widespread occurrence of plant diseases, provide considerable challenges to sustainable crop production.

Plant diseases, specifically, pose a significant danger to worldwide agricultural productivity. The annual loss of the crop yield due to plant diseases is estimated to be more than %30 globally, worth hundreds of billions of dollars (Savary et al. (2019)). In addition to economic repercussions, plant diseases can undermine food quality and safety, disrupt global supply systems, and exacerbate the difficulties of fulfilling future food requirements. The diseases impacting high-value crops, such as olives, are particularly concerning due to their economic and cultural importance in various places globally.

Peacock Eye Disease, caused by the fungal fungus Spilocaea oleaginea, is a serious ailment that only affects olive plants. Peacock Eye Disease, was described for the first time in Marseille, France by Castagne (1845). The disease is widespread in the Mediterranean regions and in all olive growing areas and continents, where can cause severe yield losses (Buonaurio et al. (2023)). Symptoms on shoots, leaf stalks, fruit pedicels and peduncles, fruits and inflorescences are rare. The first symptoms on upper leaf surface are characterized by circular brown green spots (2-10 mm in diameter), which are barely visible as their colour is similar to that of the healthy surrounding areas (Buonaurio et al. (2023)).

Olive plants, a cornerstone of Mediterranean agriculture and a vital source of olive oil, are particularly susceptible to various diseases. Among them, Peacock Eye Disease (Spilocaea oleaginea) is considered the most important disease affecting olive trees, For instance in Palestine, where olive production contributes approximately 13% to the national income (Salman et al. (2011)). The advancing infection leads to early leaf abscission, diminished tree vitality, decreased fruit yield, and, ultimately, substantial reductions in olive production and quality. In the absence of early detection and prompt therapy, Peacock Eye Disease can inflict significant harm, adversely affecting farmers' livelihoods and destabilizing the worldwide olive oil industry, in which olives serve as a fundamental crop. Plant diseases represent a substantial risk to world agriculture, resulting in large crop losses and economic repercussions. Early disease detection is crucial for executing prompt interventions and maintaining sustainable agricultural practices. Traditional methods of plant disease identification, such as manual inspection, are labor-intensive and unsuitable for extensive operations. Currently available laboratory datasets do not contain images gathered and labeled from real-life situations (Arsenovic et al. (2019)). Therefore, training is conducted with images taken in a controlled environment. Laboratory datasets, generally comprise single-leaf photos obtained under optimal settings, featuring consistent backgrounds and regulated lighting. Although these datasets enable the training of very precise machine learning models, their efficacy frequently declines when utilized on field photos characterized by overlapping objects, noisy backgrounds, and fluctuating lighting conditions. The disparity between laboratory-trained models and real-world applicability underscores the necessity for more robust approaches adapted to field situations. The study introduces an innovative method to address this gap by utilizing the SAM and autoencoders for the detection of Peacock Eye Disease in olive tree foliage. SAM, an advanced segmentation model, delineates objects in field photos, including leaves, branches, and background components. Autoencoders enhance the segmentation output by differentiating leaf objects from non-leaf elements, hence providing precise and targeted inputs for disease categorization. The methodology tackles practical issues by minimizing noise and enhancing model efficacy in intricate field situations. The suggested method provides substantial enhancements in disease detection accuracy, with the capacity to advance precision agriculture via scalable, automated solutions. The research underscores the practical utility of this technology for on-site disease identification, offering a framework that can be modified for various crops and diseases. Our study is organized into five chapters, each addressing critical aspects of the research. Chapter 1, Introduction provides a summary of the issue, the challenges associated with real-world disease diagnosis, and the proposed solution. Chapter 2, Literature Review, examines current research, emphasizing the limitations of laboratory-trained models and the need for approaches tailored to field conditions. Chapter 3, Methodology, outlines the proposed approach, including segmentation with SAM, object differentiation using autoencoders, and disease categorization, along with data preprocessing, which details the dataset used and the preprocessing steps applied to prepare the data for analysis, as well as the implementation, which discusses the technical execution of the methodology, including model training, parameter optimization, and component integration. Chapter 4, Results & Analysis, presents the study's findings, evaluates the methodology's performance, and compares it with baseline approaches. Finally, Chapter 5, Conclusion and Further Work, summarizes the findings, explores their implications for practical agriculture, and outlines potential directions for future research. The study aims to show how the proposed methodology effectively tackles real-world issues and enhances plant disease detection methodologies, assuring both relevance and scalability for practical applications.

CHAPTER TWO
LITERATURE REVIEW

Recent research has focused extensively on the detection of plant diseases using computer-aided methods. Numerous studies on various plant diseases exist in the literature. Numerous disease detection studies on olive leaf disease employ various classification approaches. In these works, researchers employed several machine learning methods and distinct deep learning models for categorization tasks. Below is some research on olive leaf disease found in the literature.

Diker et al. (2024) devised a novel technique for classifying olive leaves afflicted with Peacock Eye Disease, attaining an accuracy of 98.63%. They utilized a new dataset including 954 photos, integrating deep learning architectures (ResNet101 and MobileNet) with machine learning methods, including Random Forest and SVM. Their study is notable for its equilibrium between precision and computational efficiency, exceeding previous techniques such as VGG-based models. The integration of CNN-based feature extraction with machine learning classifiers illustrates the potential for scalable and cost-effective surveillance of agricultural diseases.

Moupojou et al. (2024) introduced an ensemble model that integrates the SAM and Fully Convolutional Data Description (FCDD) to enhance plant disease identification in field pictures. Their approach adeptly addresses issues such as intricate backgrounds and overlapping foliage, enhancing classification accuracy by more than 10% on datasets like PlantDoc. The algorithm utilizes SAM for object segmentation, employs FCDD for leaf identification, and classifies illnesses with a deep learning model trained on PlantVillage data. The introduced model connects controlled laboratory datasets with real-world field circumstances, providing an effective solution for reliable plant disease diagnosis.

Li et al. (2023) presented the Agricultural Segment Anything Model Adapter (ASA) to improve agricultural image segmentation. By modifying the general-purpose SAM with domain-specific enhancements, ASA markedly enhanced segmentation precision for agricultural applications such as crop disease and pest identification. The model attained a 41.48% enhancement in Dice scores for coffee-leaf-disease segmentation and a 9.91% augmentation for pest segmentation relative to the original SAM. The introduced model emphasizes the necessity of modifying fundamental AI models for agricultural use, tackling issues such as intricate backdrops and low-contrast subjects. The incorporation of lightweight adapters into SAM illustrates a scalable approach for zero-shot segmentation in agriculture, representing a significant leap for precision agriculture.

Zhang et al. (2024) modified the SAM for plant identification and automated phenotypic assessment, employing Explainable Contrastive Language–Image Pretraining (ECLIP) for zero-shot segmentation. Their methodology obviated the necessity for annotated datasets while attaining high precision in segmenting plant components across varied situations. An essential innovation involved employing a B-spline curve in the skeletonization process, enhancing measurement robustness, and attaining a mean absolute error (MAE) of under 0.05 for the majority of samples. In contrast to supervised models such as Mask-RCNN, their architecture provided comparable segmentation performance without the necessity for labeled data. The introduced study emphasizes SAM's versatility in agricultural applications, showcasing its capability to connect laboratory models with field-level disease detection systems. The results are pertinent to study, especially in utilizing SAM for the identification of plant diseases in practical environments.

Ksibi et al. (2022) introduced MobiRes-Net, a hybrid deep learning model that integrates ResNet50 and MobileNet for the detection of olive leaf diseases. The program utilized a dataset of 5,400 drone-acquired photos to categorize leaves into four classifications: healthy, Aculus Olearius disease, olive scab, and peacock spot illness. MobiRes-Net attained an accuracy of 97.08%, exceeding the performance of individual models such as ResNet50 (94.86%) and MobileNet (95.63%).

CHAPTER THREE
MATERIALS AND METHODOLOGY

The method utilizes two distinct datasets, the Field Dataset and the Laboratory Dataset, which are employed to develop an automated olive leaf disease detection system. Figure 3.1 illustrates the sequential workflow adopted in the study, detailing the key steps from data preprocessing to final classification.

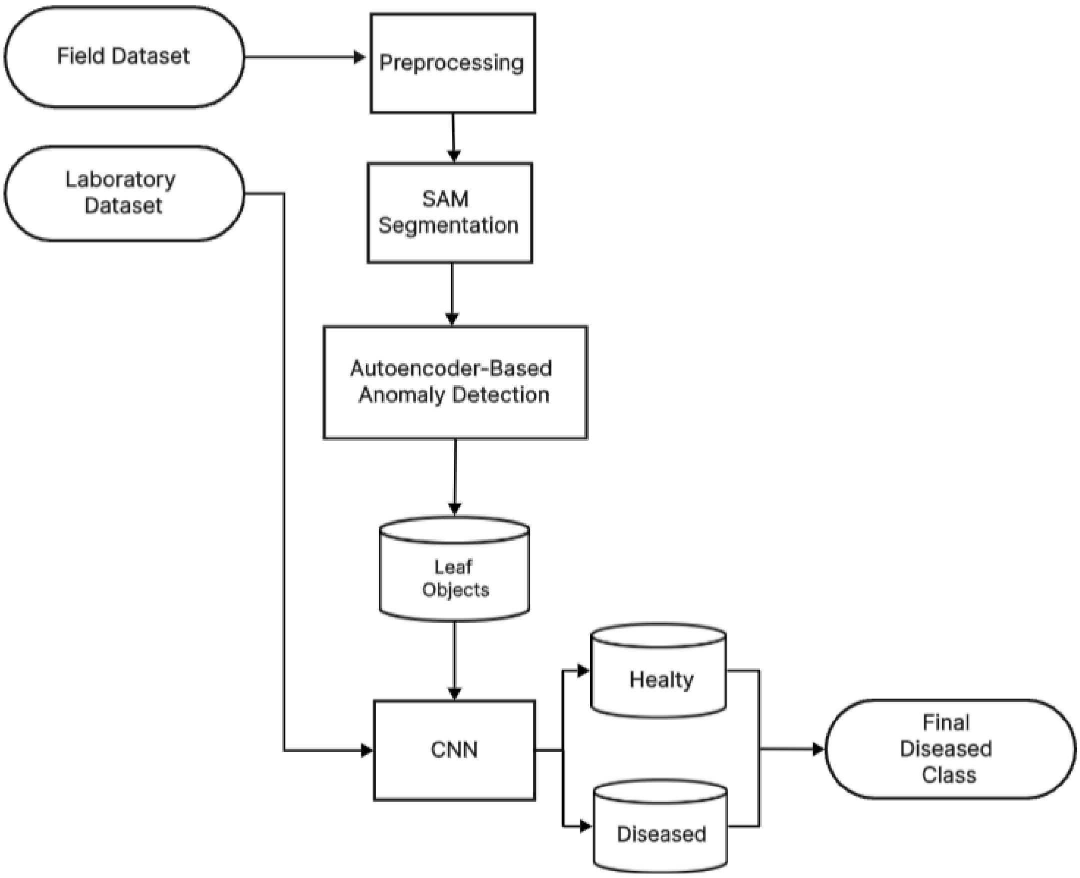

Figure 3.1 Diagram of the process

The proposed system begins by processing input images through the SAM segmentation module, which extracts and segments both leaf objects and background elements. Next, an autoencoder-based anomaly detection model evaluates the segmented objects, identifying actual leaves while filtering out non-leaf components. Among the detected leaves, the Region of Interest (ROI) is selected based on the lowest anomaly score, ensuring that the most representative leaf is chosen for further analysis. Furthermore, the selected ROI is passed to the classifier, which has been trained using the Laboratory Dataset containing healthy and diseased olive leaf samples under controlled conditions. The classifier utilizes these training samples to accurately determine whether the selected ROI corresponds to a healthy or diseased leaf. The final disease class is then determined based on the classification process.

3.1 Materials

3.1.1 Field Dataset

Figure 3.2 Field dataset

The field dataset was constructed by collecting over 1,000 high-resolution images of olive trees from Çobanhasan village, located in the Akhisar district of Manisa, Turkey. These images were captured under diverse field conditions to ensure a representative sample of real-world scenarios. However, during the data collection process, a subset of images was intentionally captured to encompass the entire tree structure, aiming to provide a holistic view of the foliage. While this approach was initially intended to enhance the dataset's comprehensiveness, it introduced significant challenges during the segmentation phase. The primary challenge arose during the application of the SAM, a state-of-the-art segmentation model designed for object delineation in complex visual data. When processing images that captured entire trees, SAM identified thousands of objects per image, including leaves, branches, and background elements. This resulted in excessive computational costs and prohibitive RAM usage, rendering the segmentation process inefficient for large-scale analysis. Additionally, the resolution and spatial distance of the objects in these wide-angle images posed a secondary challenge. While SAM demonstrated robust segmentation capabilities, the low resolution of individual leaves and their proximity to one another hindered the model's ability to accurately isolate and identify distinct leaf structures. To address these limitations, a strategic refinement of the dataset was undertaken. Images captured from a greater distance, which primarily focused on entire trees, were excluded from the final dataset. This decision was made to prioritize computational efficiency and segmentation accuracy. The final curated dataset comprised 100 high-resolution (4032x3024), close-up images of olive tree foliage, ensuring that each image contained sufficient detail for precise segmentation and subsequent disease detection. Figure 3.2 presents the visual representation of the refined dataset, highlighting the quality and focus of the selected images. This refined dataset not only reduced computational overhead but also enhanced the model's ability to accurately identify and analyze individual leaves, thereby improving the overall efficacy of the disease detection pipeline.

3.1.2 Laboratory Dataset

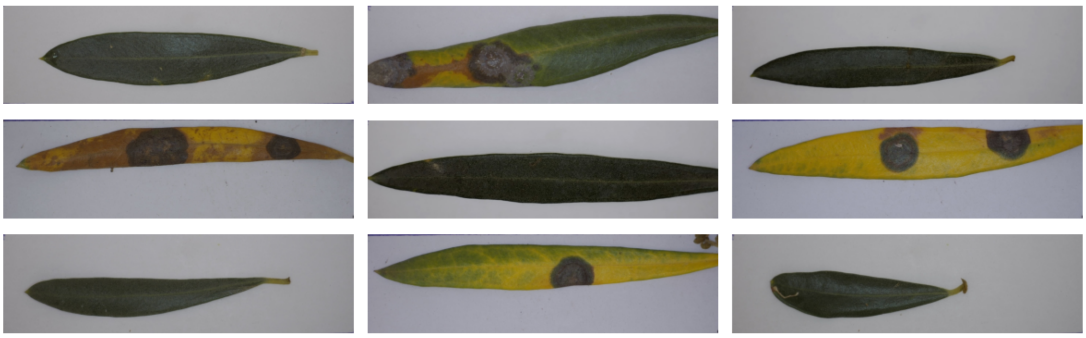

Figure 3.3 Laboratory dataset

The laboratory dataset was obtained from the study by Diker et al. (2024), titled ”An effective feature extraction method for olive peacock eye leaf disease classification.” It comprises 954 olive leaf images captured under controlled conditions, including 572 healthy and 382 infected leaves. Figure 3.3 presents the visual representation of the laboratory dataset. The images were acquired using a 48-megapixel camera within a 42 × 42 cm lighting box equipped with a specialized lighting system designed to eliminate shadow formation in the background. The controlled acquisition environment ensures clean and focused inputs, making this dataset particularly advantageous for training classification models using transfer learning techniques. The consistency in lighting and background conditions allows for enhanced feature extraction and contributes to improved model performance.

3.2 Methodology

3.2.1 Preprocess

The preprocessing phase is crucial for preparing picture data for further analysis and model training. The primary objective of this step was to standardize image dimensions and provide uniformity throughout the dataset, which is essential for optimal model performance. The original high-resolution photos, taken at 4032x3024 pixels, were shrunk to a consistent dimension of 224x224 pixels. The resizing was executed to diminish computational complexity and memory consumption during model training, while ensuring consistency between the field dataset and laboratory dataset, hence facilitating smooth integration and comparison of findings. The aspect ratio of the original photos was not kept during resizing to ensure uniformity throughout the collection. This choice was made to prevent the introduction of distortions or inconsistencies that may result from padding or cropping methods. The resizing procedure employed bilinear interpolation, a prevalent technique for image enlargement that optimizes computational efficiency and resolution. This approach guarantees seamless transitions between pixels, reducing artifacts that may compromise the model's efficacy. Subsequent to resizing, the pixel values of the pictures were normalized to a range of [0, 1] by dividing each pixel value by 255. This normalization step is a conventional procedure in deep learning processes, since it aids in stabilizing and expediting the training process. The field dataset includes photographs captured in real-world environments, and the laboratory dataset contains images acquired in controlled laboratory settings. Both datasets received the same preprocessing procedures to guarantee compatibility and consistency in later analysis. The final preprocessed dataset comprised images of 224x224 pixels, prepared for incorporation into the segmentation and anomaly detection pipelines. This consistent format guaranteed that the models could process the photos effectively and yield dependable results.

3.2.2 Segmentation with SAM

SAM, created by Meta AI Research, serves as a promptable segmentation framework designed for general image segmentation tasks Kirillov et al. (2023). SAM serves as a foundational framework for image segmentation, designed to perform robust zero-shot segmentation across many image domains and tasks without requiring more refinement. It achieves strong generalization by adjusting to various downstream applications, employing a range of prompts—such as dots, boxes, or masks—to dynamically guide segmentation. SAM's design has two main components: an image encoder and a mask decoder. The image encoder derives salient information from the input image. SAM employs a Vision Transformer (ViT) or a comparable deep learning framework for image analysis Dosovitskiy (2020). The encoder converts the input image into a detailed feature map that encompasses spatial and contextual information. This feature map serves as the foundation for the subsequent mask production process. The encoder is pre-trained on a vast dataset of over 11 million images and 1 billion masks, enabling it to generalize across diverse image domains. It concurrently analyzes the entire image, capturing both local and global information, which is crucial for identifying objects of various sizes and shapes. The mask decoder employs the feature map generated by the image encoder to create segmentation masks. Unlike traditional segmentation models that need task-specific training, SAM's decoder is designed to function in a zero-shot manner, enabling it to generate masks for objects it has not explicitly encountered during training. SAM can support several prompt types, such as points, boxes, or text, to guide the segmentation process. In the study, no prompts were used, and SAM was applied in automated segmentation mode. SAM independently generates masks for all identifiable objects in the image without requiring user involvement. This is achieved by employing its pre-trained expertise to recognize and delineate objects based on their visual attributes.

The study utilized SAM in automatic segmentation mode to examine images of field plants. The primary aim was to delineate all observable items, including leaves, stems, fruits, and background elements like as the ground or sky. The SAM model was initialized using a pre-trained checkpoint that includes the weights for the Vision Transformer (ViT) architecture. The input photos were scaled with a scaling factor of 0.5 to diminish computing demand while preserving adequate resolution for precise segmentation. The resized photos were subsequently input into the Sam Automatic Mask Generator, which produced segmentation masks for all discernible items within the images. These masks were created without the utilization of prompts. To improve computational efficiency and focus on relevant elements, pixel-based thresholds were utilized to filter the produced masks. To achieve this, adaptive thresholding was implemented based on the pixel dimensions of the masks. The lower and upper thresholds are established based on the minimum and maximum sizes of the segmented masks, modified by a certain factor. This component established a tolerance range to accommodate variability in object dimensions. These thresholds were employed to differentiate between pertinent objects (e.g., leaves) and extraneous elements (e.g., background or noise) to the greatest extent possible.Masks smaller than a specified pixel size were discarded. This phase was crucial for eliminating noise and irrelevant tiny elements, such as shadows, dust, or minor artifacts, which could hinder the analysis, as illustrated in Figure 3.4. Masks surpassing a specified pixel size were similarly omitted. This phase aimed to remove disproportionately large masks associated with non-target elements, such as the sky, clouds, or tree branches, which often occupy significant portions of the image but are extraneous to this research. The application of pixel-based thresholds was essential for various reasons. Without thresholds, SAM may generate an excessive number of masks, including those for trivial items, leading to over-segmentation and extended

Figure 3.4 Source image and segmented image after applying SAM with pixel-based thresholding

processing time. The analysis concentrated on the primary items of interest (e.g., leaves) by omitting masks that were either overly large or small, hence ensuring more accurate and meaningful results. Minimizing the quantity of processed masks diminished the computing burden, making the technique scalable for large agricultural datasets.

A primary advantage of SAM is its ability to generalize across diverse image domains without requiring task-specific fine-tuning. The study demonstrated that SAM had significant performance in segmenting images of natural plants, even in the absence of explicit cues. The amalgamation of automatic segmentation with pixel-based thresholding enabled swift and accurate object isolation, allowing for subsequent analysis, such as anomaly detection or classification. The implementation of SAM in automatic segmentation mode, coupled with pixel-based thresholding, highlights the model's versatility and adaptability for practical applications. This approach enhances the segmentation process while ensuring computational efficiency and focus on the objects of concern. Future studies may explore the integration of additional filtering techniques to enhance segmentation accuracy and efficiency.

3.2.3 Auto Encoder Based Anomaly Detection

The autoencoder-based anomaly detection technique was utilized to overcome the deficiencies of the SAM in differentiating leaf objects from non-leaf components (e.g., stems, soil, backdrop). Autoencoders, including an encoder and a decoder (Hinton and Salakhutdinov, 2006), were trained solely on leaf masks to acquire the structural and textural characteristics of leaves. During training, the encoder condensed the input data into a latent space, while the decoder rebuilt the original input from this compressed representation. The training goal was to reduce the Mean Squared Error (MSE) between the input and reconstructed masks, allowing the autoencoder to faithfully replicate leaf patterns (Bishop and Nasrabadi (2006)).

To optimize the efficiency and precision of the anomaly detection process, the autoencoder was configured with an input image dimension of (64, 64, 3), wherein images were downsized to 64×64 pixels with three color channels (RGB). The photos were standardized to the range [0, 1] by dividing pixel values by 255. The encoder had two convolutional layers: the initial layer featured 32 filters, while the subsequent layer included 64 filters, both employing 3×3 kernels, ReLU activation, and 'same' padding, succeeded by batch normalization to enhance learning stability. The feature maps were further flattened and processed through a dense layer comprising 128 neurons to provide a compact latent space representation. The decoder employed a symmetric architecture, initiating with a dense layer matching the input size and utilizing a sigmoid activation function, thereafter reshaping the output to restore the input image dimensions. The model was trained for 50 epochs with the Adam optimizer and mean squared error loss function, with a batch size of 32. The dataset was divided into 80% for training and 20% for validation.

An initial optimization step was performed using 100 segmented masks generated by SAM to ascertain the best anomalous threshold. The reconstruction errors for these masks were evaluated, and an ideal threshold value was established by balancing precision and recall, ensuring an effective differentiation between leaf and non-leaf objects. The final threshold was determined through statistical analysis using the equation:

$$Threshold = \mu + \sigma \tag{3.1}$$

Alternatively, an experimentally optimized fixed threshold of 0.165 was employed.

During inference, the autoencoder analyzed segmented masks produced by SAM, encompassing both foliar and non-foliar items. The reconstruction error for each mask was determined as the mean squared error (MSE) between the input and the rebuilt mask. A threshold was set to categorize masks with little reconstruction mistakes as leaves (normal class) and those with significant reconstruction defects as anomalies (non-leaf objects). The threshold was established using a blend of statistical analysis and experimental validation, guaranteeing optimal differentiation between leaf and non-leaf parts.

The amalgamation of SAM with autoencoder-based anomaly detection established a resilient two-stage pipeline. SAM produced masks for all discernible items in the field images, but the autoencoder enhanced these masks by eliminating non-leaf components. This method markedly diminished noise and enhanced the quality of the input data for ensuing classification jobs. The autoencoder's capacity for unsupervised operation rendered it especially appropriate for situations with scarce labeled data, as it could efficiently generalize to unobserved abnormalities.

The autoencoder's performance was assessed using measures including precision, recall, and F1 score, which evidenced its efficacy in differentiating leaf objects from background elements. Visual representations of the segmentation outcomes prior to and after the anomaly identification were presented to demonstrate the method's effectiveness. A histogram of reconstruction errors further demonstrated a distinct separation between leaf and non-leaf objects, hence corroborating the selected threshold.

This methodology is inspired by the research of Shikhar and Sobti (2024), who employed masked autoencoders for anomaly identification in aerial agricultural imagery. Our methodology, however, enhances their research by incorporating SAM for preliminary segmentation and concentrating explicitly on leaf object recognition.

Subsequent studies may investigate the application of more sophisticated anomaly detection models to improve segmentation precision. Such a technique is the incorporation of Variational Autoencoders (VAEs), which provide a probabilistic method for latent space modeling and have demonstrated enhancements in anomaly identification (Kingma (2013)). Moreover, Generative Adversarial Networks (GANs), effectively utilized in unsupervised anomaly detection through the learning of normal sample distributions, may offer a viable approach for enhancing segmentation outcomes (Zenati et al. (2018)). Moreover, the method's relevance to additional agricultural activities, including fruit detection and disease localization, necessitates further examination (Boulent et al. (2019)).

3.2.4 Convolutional Neural Network (CNN)

Convolutional Neural Networks (CNNs) are a category of deep learning architectures tailored for image recognition and processing applications. In contrast to conventional neural networks, CNNs utilize a hierarchical architecture to autonomously and adaptively acquire spatial hierarchies of features from input images (LeCun et al. (1998)). The architecture of a CNN generally comprises three primary components; convolutional layers, pooling layers, and fully linked layers as shown in Figure 3.5.

Convolutional layers utilize a sequence of filters on the input image to extract localized characteristics, including edges, textures, and patterns. Each filter produces a feature map by convolving with the input picture, subsequently using a non-linear activation function (e.g., ReLU) to incorporate non-linearity (Krizhevsky et al. (2012)). Pooling layers diminish the spatial dimensions of feature maps, enhancing the model's computational efficiency and reducing its susceptibility to overfitting (Boureau et al. (2010)). This work utilized max pooling to extract the maximum value from sub-regions of the feature map, thereby preserving dominating characteristics and minimizing computational demands (Scherer et al. (2010)). At the conclusion of

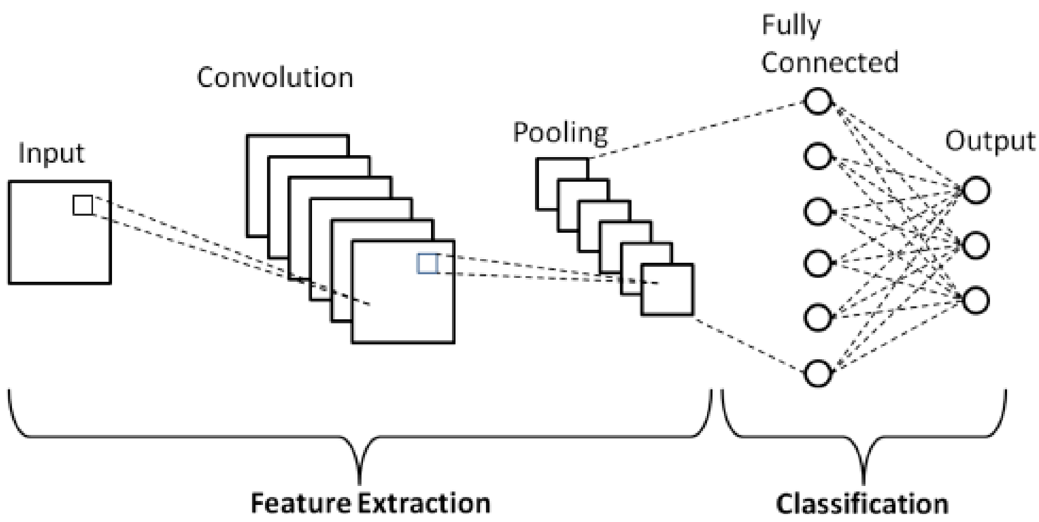

Figure 3.5 The basic architecture of CNN (Phung and Rhee, 2019)

the CNN, fully connected layers convert the feature maps into a one-dimensional vector and execute classification, typically employing a softmax activation function in the last layer to generate probability distributions for each class (Goodfellow (2016)).

This project involved training a CNN model using a laboratory dataset of leaf pictures to differentiate between healthy and sick leaves. The dataset underwent preprocessing by scaling photos to 224 × 224 pixels for uniformity across inputs, and pixel values were normalized to [0, 1] to enhance convergence stability (Ioffe (2015)). Data augmentation techniques, such as random flipping, rotation (±20°), and zooming (±10%), were employed to enhance the dataset's diversity and improve the model's generalization (Shorten and Khoshgoftaar (2019)).

The CNN architecture had five convolutional layers, each intended to extract hierarchical spatial characteristics from leaf pictures. The initial layer, comprising 32 filters, identified low-level features like edges and textures, whereas subsequent layers with 64 filters recognized intricate patterns such as leaf veins and disease spots. Every convolutional layer was succeeded by ReLU activation and max pooling (2 × 2) to downsample the feature maps while preserving critical information (Zeiler and Fergus (2014)).

To alleviate overfitting, dropout regularization was implemented at several stages:

20% dropout following the initial two convolutional layers, 30% after the third and fourth layers, and 40% subsequent to the fully connected layer (Srivastava et al. (2014)). The ultimate categorization utilized a fully linked dense layer including 64 neurons, succeeded by a softmax activation function to provide probability distributions across the potential categories. The model utilized the Sparse Categorical Crossentropy loss function, appropriate for multi-class classification tasks (Murphy (2012)).

Hyperparameters were selected based on previous studies and empirical confirmation. The model was trained with a batch size of 32, chosen to optimize computational efficiency and convergence stability. Studies indicate that smaller batch sizes generally result in more variance in gradient updates, which might enhance generalization, whereas overly large batch sizes may impair generalization (Smith (2018)). The epoch count was established at 50, as initial studies revealed that validation loss stabilized beyond this threshold, indicating the commencement of overfitting (Goodfellow (2016)).

An exponential learning rate decay was employed to guarantee efficient weight updates. The equation is given below.

$$\eta = 1e^{-3}\text{»}10^{(-epoch/20)} \tag{3.2}$$

This method enabled the model to implement substantial initial adjustments, promoting swift convergence during the initial training phases, while maintaining stability in subsequent epochs (Loshchilov and Hutter (2016)). The Adam optimizer was employed for its adjustable learning rate characteristics, which have demonstrated enhanced convergence compared to conventional approaches like Stochastic Gradient Descent (Kingma (2014)).

Subsequent to training, the CNN model was employed to categorize the leaf dataset derived from the autoencoder-based anomaly detection procedure. This dataset exclusively comprised leaf objects, as non-leaf elements (e.g., stems, dirt) were eliminated using the autoencoder. The model’s efficacy was assessed through accuracy, precision, recall, and F1-score, which are standard metrics for evaluating classification models (Sokolova and Lapalme (2009)).

A number of tests were undertaken to compare the performance of the CNN models trained on the laboratory dataset with the autoencoder-filtered dataset. The findings indicated that the model trained on the autoencoder-filtered dataset attained superior accuracy and F1 scores, suggesting that the exclusion of non-leaf elements markedly enhanced the quality of the input data. This comparison underscores the significance of preprocessing and filtering in improving the efficacy of deep learning models for agricultural applications.

The proposed methodology exhibited encouraging findings; nonetheless, numerous drawbacks were recognized. The efficacy of the CNN model may diminish when utilized on more intricate images characterized by fluctuating lighting conditions or obstructions. The dependence on a laboratory dataset for initial training may restrict the model's applicability to real-world field circumstances. Future research may investigate the implementation of more sophisticated CNN designs, such as ResNet or EfficientNet, which have demonstrated enhancements in classification efficacy for large-scale image recognition applications (He et al. (2016); Tan and Le (2019)). Furthermore, the incorporation of transfer learning methodologies and the acquisition of larger, more varied datasets could improve the model's resilience and relevance to various agricultural contexts (Kornblith et al. (2019)).

3.3 Final Classification of Leaf Objects

In the concluding phase of the investigation, the segmented leaf images were categorized into two classifications: healthy and diseased. The main objective of this classification was to assess the efficacy of the autoencoder-based filtering method in eliminating non-leaf items while maintaining the critical characteristics of the leaves for illness categorization. To achieve this objective, we utilized a CNN model that had been pre-trained on a standardized benchmark dataset. This enabled us to evaluate the model's ability to generalize to real-world, field-acquired leaf photos that have undergone automated filtering. The dataset utilized in this classification phase comprised leaf pictures acquired following the autoencoder filtering procedure. The photos were meticulously classified into healthy and diseased groups, establishing a dependable ground truth for assessment. The dataset was subsequently preprocessed to maintain alignment with the training parameters of the benchmark CNN. Each image was shrunk to 256 × 256 pixels and normalized to a range of [0, 1] by min-max scaling. Data augmentation methods, such as random flipping, rotation (±20°), and zooming (±10%), were employed to enhance the diversity of the training dataset and mitigate overfitting.

The classification model employed at this step utilized a fine-tuned CNN architecture, which was originally trained on the benchmark dataset. Transfer learning was utilized by loading the pre-trained CNN weights and subsequently optimizing them with the filtered dataset, rather than building a model from the ground up. In the fine-tuning process, the initial layers of the CNN were frozen to preserve low-level feature representations, while the final four layers were unfrozen to facilitate adaption to the newly introduced dataset. The concluding layers comprised fully connected layers with a dropout rate of 0.3 to enhance generalization, succeeded by a softmax activation function for binary classification.

Class weighting was integrated into the training process to mitigate class imbalance (Bakirarar and Elhan (2023)). The class weights were computed according to the dataset distribution utilizing the formula:

$$W_c = \frac{N}{n_c » C} \tag{3.3}$$

where N represents the total number of samples, nc is the number of samples in class c, and C is the total number of classes. This weighting approach prevented the model from exhibiting bias towards the predominant class (e.g., healthy leaves) and ensured equitable contributions from both categories.

The model was trained with the Adam optimizer with a learning rate of 0.0005, a batch size of 32, and an early stopping criterion with a patience of 15 epochs to mitigate overfitting. The employed loss function was sparse categorical cross-entropy, appropriate for multi-class classification with integer labels. The model was assessed using essential performance indicators, including accuracy, precision, recall, F1-score, and a confusion matrix for class-wise performance analysis.

3.4 Performance Metrics

A confusion matrix is a commonly used tool to assess the performance of a classification model by comparing its predictions with actual outcomes in a test dataset. In this matrix, the rows correspond to the actual class labels, while the columns represent the predicted classifications.

Table 3.1 Performance Metrics

Performance metrics	**Definition**
Accuracy	$\frac{TP+TN}{TP+FN+FP+TN}$ (3.4)
Precision	$\frac{TP}{TP+FP}$ (3.5)
Recall	$\frac{TP}{TP+FN}$ (3.6)
F1 Score	$\frac{2\times(Recall\times Precision)}{Recall+Precision}$ (3.7)

The matrix clarifies four potential outcomes: A true positive (TP) arises when the model accurately identifies the positive class, while a true negative (TN) denotes the proper identification of the negative class. A false positive (FP) occurs when the model erroneously categorizes a negative occurrence as positive, while a false negative (FN) transpires when a positive case is inaccurately classified as negative. Analyzing the confusion matrix allows for the calculation of numerous performance metrics, including accuracy, precision, recall, and F1 score. The equations for these metrics are presented in Table 3.1.

CHAPTER FOUR
RESULTS

The main objective of this classification challenge is to create a strong and efficient system for identifying and diagnosing peacock disease in olive tree leaves through the integration of modern segmentation and classification techniques. This study aims to tackle the difficulties of moving from controlled experimental datasets to practical agricultural applications. The precise objectives of this research encompass the following: Assessing the efficacy of a Convolutional Neural Network (CNN) model trained on the olive peacock disease benchmark dataset for the classification of individual leaf specimens as healthy or diseased. This research seeks to evaluate the model’s capacity to generalize beyond the controlled parameters of the benchmark dataset and to adapt to the intricacies of real-world datasets. Field Image Dataset: This dataset consists of high-resolution photos of olive trees taken in authentic agricultural environments. These photographs encompass diverse climatic circumstances, including varying lighting, shadows, and intricate backgrounds (e.g., clouds, soil, and stems). The dataset represents the actual diversity of olive orchards and functions as a testbed for implementing segmentation and classification pipelines in tough and uncontrolled environments. The laboratory dataset comprises isolated leaf pictures obtained under controlled settings. These photos constitute a benchmark collection specifically designed for the detection of olive peacock illness. Each leaf is meticulously categorized as either healthy or ill, guaranteeing superior annotations. The laboratory dataset was utilized to build a Convolutional Neural Network (CNN) model for binary classification, establishing a dependable baseline for disease diagnosis. Assessing the efficacy of a hybrid preprocessing pipeline that integrates the Segment Anything Model (SAM) and an autoencoder discriminator to extract leaf objects from intricate olive tree photos. This pipeline aims to remove noise from non-leaf entities, like the background, soil, clouds, and stems, which frequently hinder classification efforts. Enabling precise classification of individual leaf objects derived from segmented tree photos, offering accurate and actionable insights about the health status of olive trees. Priority is given on attaining elevated specificity in illness detection at the object level. Improving classification accuracy and reliability by minimizing misclassification due to irrelevant features through the differentiation of leaf and non-leaf objects. This method guarantees that the model concentrates exclusively on pertinent data, enhancing the reliability of disease identification. Evaluating the feasibility and application of a benchmark-trained CNN model in conjunction with SAM-based segmentation for datasets gathered in the field. This stage connects benchmark circumstances with actual agricultural settings, tackling the issues posed by fluctuating environmental and imaging conditions. Exhibiting the scalability and feasibility of the suggested method for extensive applications in olive orchards. The capacity to automate disease detection and diagnosis on a large scale holds considerable promise for enhancing crop management methods and effectively reducing disease burden. This work emphasizes the significance of integrating advanced segmentation approaches, such as SAM, with customized classification models to tackle practical agricultural issues. The results enhance the domain of automated disease identification in precision agriculture, establishing a foundation for subsequent study and practical implementations.

4.1 CNN Training Results

The CNN was trained using a laboratory dataset, with 80% allocated for training and 20% for testing. The model was optimized using the Adam optimizer with a learning rate of 0.001, trained for 50 epochs and a batch size of 32. The training achieved an accuracy of 96% on both the training and testing sets, as illustrated in Table 4.1. The metrics of precision, recall, and F1-score were employed to evaluate the model's overall performance, demonstrating a strong ability to distinguish between healthy and damaged leaves.

Table 4.1 The laboratory dataset CNN training performance

Class	Accuracy	F1-Score	Recall	Precision
Healthy	0.96	0.97	0.99	0.96
Unhealthy	0.96	0.95	0.92	0.98

Figure 4.1 illustrates that both training and validation accuracy exhibit a consistent increase, converging at around 96% and 94%, respectively. The proximity of the curves suggests negligible overfitting, affirming the model's capacity to generalize effectively to novel data. The variations are characteristic of natural picture collections and indicate the difficulty in differentiating between healthy and diseased leaves.

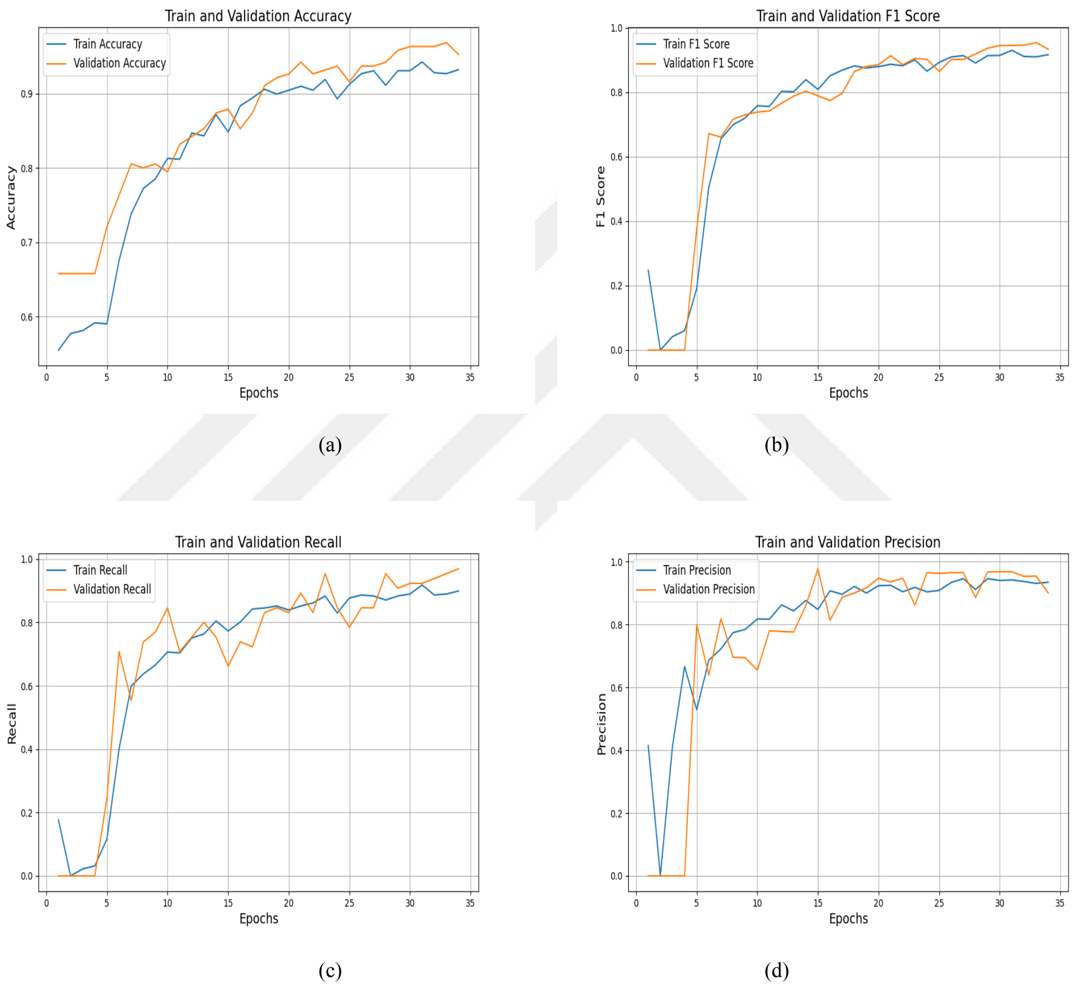

Figure 4.1 Training and validation accuracy (a), F1-score (b), recall (c), precision (d) for laboratory dataset

4.2 Segment Anything Model

The Segment Anything Model (SAM) effectively delineated leaves from photos of trees. Nonetheless, as SAM lacks the intrinsic capability to identify objects, it also extracted non-leaf elements, including branches, sky, and soil.

A pixel-based filtering approach was devised during segmentation to improve computational performance and concentrate on pertinent objects. This method dynamically eliminated masks within a specified pixel threshold, thereby reducing wasteful memory use and enhancing processing speed. The method facilitated the elimination of minor artifacts and background noise, so preserving just the most pertinent segmented items. The filtering enhanced segmentation quality but also risked removing smaller leaves, potentially affecting classification performance.

Figure 4.2 Segmented image with pixel-based thresholding

4.3 Autoencoder Discriminator

The performance of the autoencoder-based anomaly detection method was evaluated by analyzing the reconstruction loss during training, the distribution of reconstruction errors, and the classification of segmented masks into normal and anomalous categories.

4.3.1 Training and Validation Loss

The autoencoder was trained for 50 epochs using the Mean Squared Error (MSE) loss function, with an Adam optimizer and a batch size of 32. The final training loss was 0.0285, while the validation loss converged to 0.0274, indicating that the model achieved stable learning without signs of overfitting. The similarity between training and validation loss values suggests that the model generalizes well to unseen data, effectively capturing the structural and textural patterns of leaf objects.

4.3.2 Reconstruction Error Distribution

The distribution of reconstruction errors across the segmented masks is illustrated in Figure 4.3, where the x-axis represents the reconstruction error, and the y-axis denotes frequency.

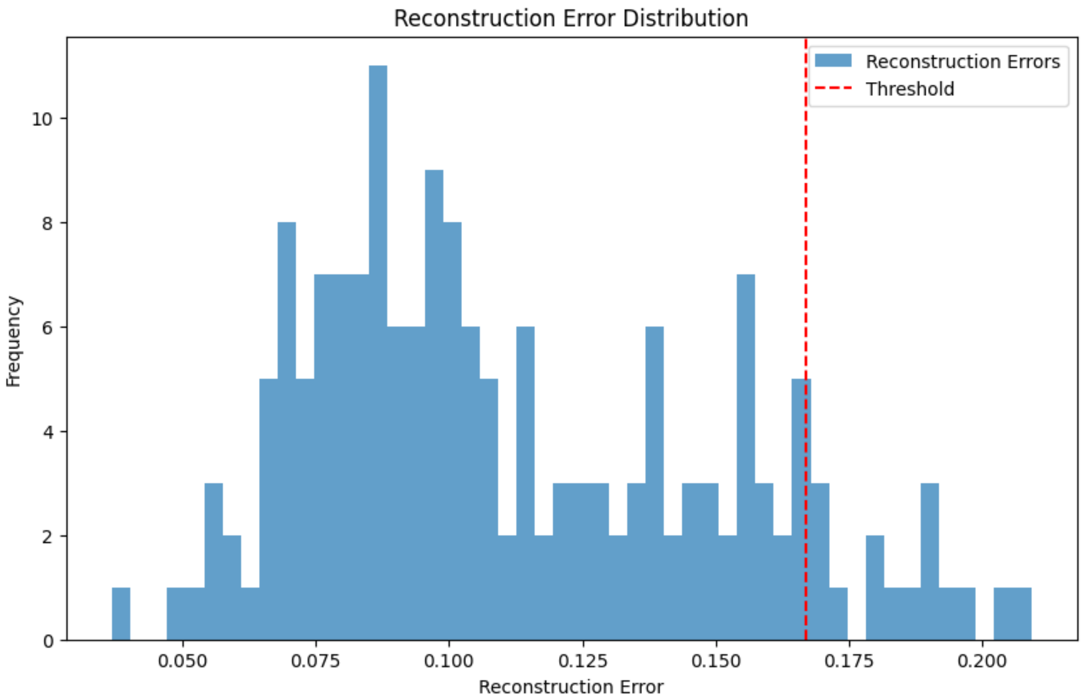

Figure 4.3 Reconstruction error distribution

The histogram has approximately normal distribution, indicating that most of the segmented masks were reconstructed with consistent accuracy. The optimal anomaly threshold was set at 0.165, as determined through a combination of statistical analysis and empirical validation. The red dashed line in Figure 4.3 represents this threshold, where:

- Masks with reconstruction errors below this threshold were classified as normal (leaf objects).
- Masks with reconstruction errors exceeding this threshold were categorized as anomalous (non-leaf objects, e.g., stems, soil, or background elements).

4.3.3 Classification of Normal and Anomalous Masks

The autoencoder-based anomaly detection pipeline processed 24,790 segmented masks produced by SAM. Of these:

Table 4.2 The number of normal and anomalous masks

Category	Number of Masks	Percentage (%)
Normal (Leaf Objects)	18,151	73.2
Anomalous (Non-Leaf Objects)	6,639	26.8
Total	24,790	100

Figure 4.4 Leaf objects and background objects

This result indicates that approximately 27% of the segmented masks contained non-leaf elements, which were effectively filtered out by the autoencoder. The relatively high proportion of anomalies underscores the necessity of a robust filtering mechanism following SAM's segmentation, as it prevents irrelevant components from being forwarded to subsequent classification stages.

4.3.4 Discussion of Findings

The results demonstrate that the autoencoder successfully differentiates between leaf and non-leaf components based on reconstruction errors. The anomaly detection process significantly reduces noise in the dataset, improving the overall quality of the segmented masks before further classification. The consistency in training and validation loss suggests that the model is well-calibrated for the task, and the threshold of 0.165 effectively separates relevant and irrelevant objects, as reflected in the clear distinction observed in the reconstruction error histogram.

4.4 Final Classification of Leaf Objects

4.4.1 Prediction of Leaf Objects with Benchmark Dataset-Trained CNN Model

The performance assessment of the CNN model, trained on the autoencoder-filtered dataset, was executed utilizing standard classification measures. The classification results were evaluated based on accuracy, precision, recall, and F1-score, and further analyzed with a confusion matrix to examine the model's predictions for healthy and diseased leaves.

The acquired metrics are displayed in Table 4.3, illustrating the model's comparative performance across several evaluation criteria. Furthermore, the confusion matrix in Figure Y depicts the allocation of true positive, false positive, false negative, and true negative predictions for each category. These results elucidate the model's categorization inclinations and its efficacy in differentiating between healthy and unhealthy leaf samples.

Table 4.3 Prediction of Leaf Objects with Benchmark Dataset-Trained CNN model

Class	**Accuracy**	**F1-Score**	**Recall**	**Precision**
Healthy	0.87	0.84	0.84	0.85
Unhealthy	0.87	0.89	0.89	0.88

The confusion matrix (Figure 4.5) further illustrates the classification results:

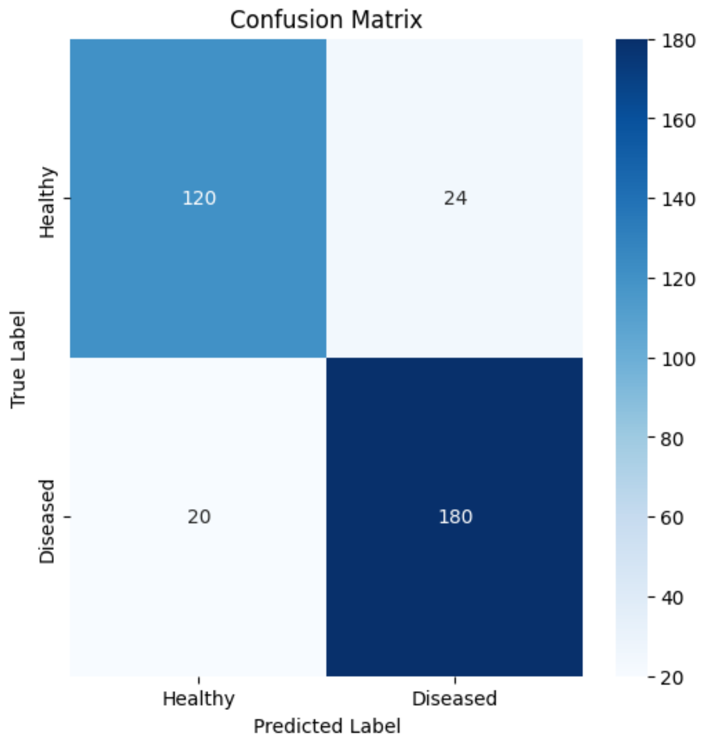

Figure 4.5 Confusion matrix of leaf objects with benchmark dataset-trained CNN Model

- True Positives (Healthy): 120
- False Positives (Healthy misclassified as Diseased): 24

The model, developed under controlled laboratory circumstances, has shown robust performance in illness categorization. Nonetheless, several healthy leaves were inaccurately categorized, possibly due to unpredictability caused by real-world conditions in the segmented dataset.

4.4.2 Classification of Labeled Leaf Objects with CNN Model

This model was trained directly on labeled leaf objects, indicating that the segmented dataset was manually tagged as ”Healthy” or ”Diseased” prior to training.

Table 4.4 Classification of labeled leaf objects with CNN model

Class	**Accuracy**	**F1-Score**	**Recall**	**Precision**
Healthy	0.87	0.84	0.83	0.85
Unhealthy	0.87	0.89	0.89	0.88

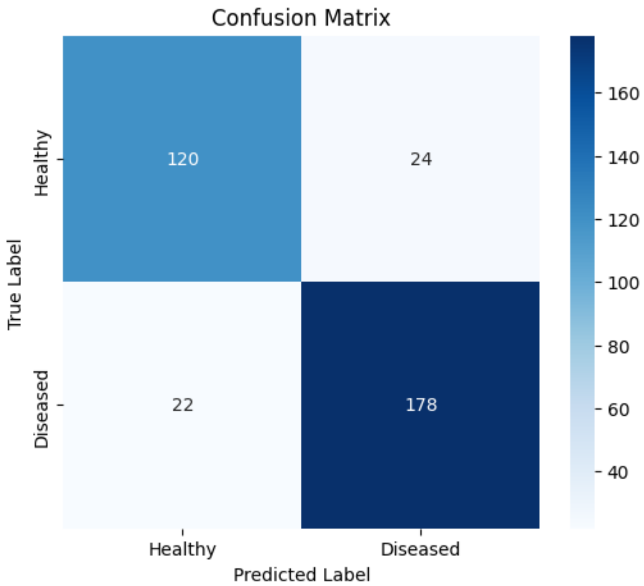

Figure 4.6 Classification of labeled leaf objects with CNN model

- True Positives (Diseased): 178
- False Negatives (Diseased misclassified as Healthy): 22

CHAPTER FIVE
CONCLUSIONS

The current research aimed to address the following two primary obstacles: the restricted applicability of laboratory-controlled datasets in real-world scenarios and the automation of the labor-intensive human annotation procedure for leaf classification. Our findings indicate that both objectives have been satisfactorily accomplished.

5.1 Practical Relevance of Laboratory Datasets

Historically, deep learning models for plant disease identification depend on datasets gathered in controlled laboratory environments. Nevertheless, these models frequently demonstrate diminished performance in real-world contexts when fluctuations in illumination, backdrop intricacy, and occlusions introduce interference. To evaluate this, we trained a CNN model on a benchmark dataset of laboratory-controlled leaf pictures, which were manually annotated as healthy or diseased. Utilized the trained model on actual leaf pictures obtained from comprehensive tree shots, with segmentation executed via the Segment Anything Model (SAM). Noted that performance was markedly diminished owing to background noise (e.g., soil, sky, branches), highlighting the constraints of conventional datasets in practical scenarios.

5.2 Automating the Annotation Procedure for Leaf Classification

The manual annotation of extensive plant datasets is a laborious and expensive endeavor. We presented a two-stage pipeline to automate this process:

Leaf Segmentation Utilizing SAM: The Segment Anything Model (SAM) was employed to isolate individual leaf regions from complete tree pictures.

Anomaly Detection Utilizing Autoencoders: Given that SAM does not categorize objects, we utilized an autoencoder-based anomaly detection technique to eliminate non-leaf entities (including soil, branches, and clouds). This enabled the automatic creation of a more refined dataset without human involvement.

To ascertain the efficacy of this method: We trained a CNN classifier on the automatically segmented and filtered dataset and evaluated its performance against the CNN model trained on the manually labeled benchmark dataset. Our classification findings exhibited significant consistency across both datasets, indicating that our automated annotation method yields classification accuracy comparable to that of manually labeled datasets.

Assessment of Performance: Key performance parameters, including accuracy, precision, recall, and F1-score, were examined for comparison. The benchmark CNN model trained on a laboratory dataset vs the CNN model trained on a real-world dataset acquired by automated segmentation and anomaly filtering.

The final results demonstrate that the SAM + Autoencoder-based dataset attains classification accuracy comparable to the manually labeled benchmark dataset. The automated pipeline effectively diminishes background noise interference and enhances classification performance in real-world scenarios. The necessity for manual annotation is much diminished, rendering the system scalable for extensive agricultural datasets. Our results indicate that benchmark datasets developed in controlled environments fail to generalize to real-world photos, highlighting the necessity for more versatile and resilient data preparation methods. The suggested SAM + Autoencoder framework efficiently automates data annotation and improves classification efficacy in practical applications. The findings indicate that anomaly detection and automated segmentation methods can serve as effective instruments for classifying plant diseases, hence diminishing reliance on expensive manual annotation.

5.3 Future Work

The research conducted has shown encouraging findings; yet, there are various areas for future enhancement. *Ensemble Learning:* Integrating various models (e.g., CNN and Transformer) to improve classification robustness.

Self-Supervised Learning: Investigating unsupervised feature extraction methods to diminish reliance on annotations.

Broadened Plant Species Testing: Evaluating the methodology across a wider array of plant species to determine its generalizability.

Enhanced Anomaly Detection: Exploring sophisticated anomaly detection techniques, including GAN-based filtering, to further optimize the dataset.

Our methodology offers a scalable, automated, and efficient solution for the diagnosis of plant diseases in real-world scenarios, facilitating future progress in AI-driven agriculture.

REFERENCES

Arsenovic, M., Karanovic, M., Sladojevic, S., Anderla, A., and Stefanovic, D. (2019). Solving current limitations of deep learning based approaches for plant disease detection. *Symmetry*, 11(7):939.

Bakirarar, B. and Elhan, A. H. (2023). Class weighting technique to deal with imbalanced class problem in machine learning: Methodological research. *Türkiye Klinikleri Biyoistatistik*, 15(1):19–29.

Bishop, C. M. and Nasrabadi, N. M. (2006). *Pattern recognition and machine learning*, volume 4. Springer.

Boulent, J., Foucher, S., Théau, J., and St-Charles, P.-L. (2019). Convolutional neural networks for the automatic identification of plant diseases. *Frontiers in plant science*, 10:941.

Boureau, Y.-L., Ponce, J., and LeCun, Y. (2010). A theoretical analysis of feature pooling in visual recognition. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 111–118.

Buonaurio, R., Almadi, L., Famiani, F., Moretti, C., Agosteo, G. E., and Schena, L. (2023). Olive leaf spot caused by venturia oleaginea: An updated review. *Frontiers in Plant Science*, 13:1061136.

Castagne, J. L. M. (1845). *Catalogue des plantes qui croissent naturellement aux environs de Marseille*. Nicot et Pardigon.

Diker, A., Elen, A., Közkurt, C., Kılıçarslan, S., Dönmez, E., Arslan, K., and Kuran, E. C. (2024). An effective feature extraction method for olive peacock eye leaf disease classification. *European Food Research and Technology*, 250(1):287–299.

Dosovitskiy, A. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

FAO (2009). The state of food and agriculture: Livestock in the balance. *FAO (Food and Agriculture Organization of the United Nations), Rome, Italy.*

Goodfellow, I. (2016). *Deep learning*, volume 196. MIT press.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Hinton, G. E. and Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *science*, 313(5786):504–507.

Ioffe, S. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*.

Kingma, D. P. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

Kingma, D. P. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., et al. (2023). Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4015–4026.

Kornblith, S., Shlens, J., and Le, Q. V. (2019). Do better imagenet models transfer better? In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2661–2671.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.

Ksibi, A., Ayadi, M., Soufiene, B. O., Jamjoom, M. M., and Ullah, Z. (2022). Mobiresnet: a hybrid deep learning model for detecting and classifying olive leaf diseases. *Applied Sciences*, 12(20):10278.

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.

Li, Y., Wang, D., Yuan, C., Li, H., and Hu, J. (2023). Enhancing agricultural image segmentation with an agricultural segment anything model adapter. *Sensors*, 23(18):7884.

Loshchilov, I. and Hutter, F. (2016). Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*.

Moupojou, E., Retraint, F., Tapamo, H., Nkenlifack, M., Kacfah, C., and Tagne, A. (2024). Segment anything model & fully convolutional data description for plant multi-disease detection on field images. *IEEE Access*.

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.

Phung, V. H. and Rhee, E. J. (2019). A high-accuracy model average ensemble of convolutional neural networks for classification of cloud image patches on small datasets. *Applied Sciences*, 9(21):4500.

Salman, M., Hawamda, A.-A., Amarni, A. A.-A., Rahil, M., Hajjeh, H., Natsheh, B., and Abuamsha, R. (2011). Evaluation of the incidence and severity of olive leaf spot caused by spilocaea oleagina on olive trees in palestine. *American Journal of Plant Sciences*, 2(3):457–460.

Savary, S., Willocquet, L., Pethybridge, S. J., Esker, P., McRoberts, N., and Nelson, A. (2019). The global burden of pathogens and pests on major food crops. *Nature ecology & evolution*, 3(3):430–439.

Scherer, D., Müller, A., and Behnke, S. (2010). Evaluation of pooling operations in convolutional architectures for object recognition. In *International conference on artificial neural networks*, pages 92–101. Springer.

Shikhar, S. and Sobti, A. (2024). Label-free anomaly detection in aerial agricultural images with masked image modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5440–5449.

Shorten, C. and Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of big data*, 6(1):1–48.

Smith, L. N. (2018). A disciplined approach to neural network hyper-parameters: Part 1–learning rate, batch size, momentum, and weight decay. *arXiv preprint arXiv:1803.09820*.

Sokolova, M. and Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information processing & management*, 45(4):427–437.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958.

Tan, M. and Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR.

Zeiler, M. D. and Fergus, R. (2014). Visualizing and understanding convolutional networks. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13*, pages 818–833. Springer.

Zenati, H., Foo, C. S., Lecouat, B., Manek, G., and Chandrasekhar, V. R. (2018). Efficient gan-based anomaly detection. *arXiv preprint arXiv:1802.06222*.

Zhang, W., Dang, L. M., Nguyen, L. Q., Alam, N., Bui, N. D., Park, H. Y., and Moon, H. (2024). Adapting the segment anything model for plant recognition and automated phenotypic parameter measurement. *Horticulturae*, 10(4):398.