

فاطمه الكاف

Fatimas Master Thesis.docx

- 1- Final Thesis (Word Format) Submission
- Thesis_(2)_AI901_Dr_Hager_Saleh_Third Trimester 2025 (Moodle PP)
- Midocean University

Document Details

Submission ID

trn:oid::1:3450615518

Submission Date

Dec 25, 2025, 9:55 AM GMT+4

Download Date

Dec 25, 2025, 11:28 PM GMT+4

File Name

xcod.273008916_401621_sisehT_retsaM_samitaF_فاطمه الكاف_60582

File Size

1.9 MB

46 Pages

10,287 Words

59,846 Characters

*% detected as AI

AI detection includes the possibility of false positives. Although some text in this submission is likely AI generated, scores below the 20% threshold are not surfaced because they have a higher likelihood of false positives.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Disclaimer

Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (i.e., our AI models may produce either false positive results or false negative results), so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.

Frequently Asked Questions

How should I interpret Turnitin's AI writing percentage and false positives?

The percentage shown in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was either likely AI-generated text from a large-language model or likely AI-generated text that was likely revised using an AI paraphrase tool or word spinner.

False positives (incorrectly flagging human-written text as AI-generated) are a possibility in AI models.

AI detection scores under 20%, which we do not surface in new reports, have a higher likelihood of false positives. To reduce the likelihood of misinterpretation, no score or highlights are attributed and are indicated with an asterisk in the report (*%).

The AI writing percentage should not be the sole basis to determine whether misconduct has occurred. The reviewer/instructor should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in accordance with their school's policies.

What does 'qualifying text' mean?

Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be likely AI-generated will be highlighted in cyan in the submission, and likely AI-generated and then likely AI-paraphrased will be highlighted purple.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.





Leveraging AI to Study the Impact of Climate Change on the Spread and Severity of Plant Diseases

Prepared By

Name: Fatima Mohammad Alkaff ID No.: 251001435

Submitted in fulfillment of the requirements for the degree Master of Science in Artificial Intelligence

At

Collage of Informatics
Midocean University
UAE

Supervisor Name: Dr. Hager Saleh

College of Informatics, Midocean University, Comoros

Acknowledgements

First and foremost, all praise and thanks are due to Allah, the Almighty, for His endless blessings, mercy, and guidance that have accompanied us throughout this research journey. Without His will, this work would not have been possible.

I extend my deepest gratitude to my esteemed supervisor, Dr. Hager Saleh, whose support, encouragement, and invaluable guidance inspired us from the first steps of this research to its final form. Her kindness, patience, and dedication were a source of motivation throughout the process.

This work is lovingly dedicated to my parents, whose unwavering love, sacrifices, and prayers have shaped who I am. Your faith in me and your constant support have been the pillars of my strength. I am forever grateful for your encouragement and for teaching us the value of perseverance and ambition.

Abstract

Quickly and correctly identifying plant diseases is a critical step that might result in a considerable increase in crop output, as well as a reduction of losses caused by climate change. Usually, visually detectable pathogens are those that are visible only to machines; however, these machines generally do not consider the effect of climatic and soil factors on the spread of the disease. This thesis presents a multimodal deep learning system that does not include only visual features of the apple leaves, also uses environmental variables to help the disease classification be more accurate and robust. by uses the Multimodal Plant Disease Dataset created by Shubham Divakar, which deals with the correct identification of the infected apple leaves, such as Apple Scab, Black Rot, Cedar Apple Rust, and those that are Healthy.

In this thesis, the performance of three pretrained convolutional neural networks (CNNs), MobileNetV2, ResNet50, and InceptionV3, was compared for their image-based baseline. Climate and soil factors were considered in the study by developing a Multilayer Perceptron (MLP) that took temperature, humidity, pH, rainfall, and nutrient levels (N, P, K) as features. The fusion layer was employed to merge the two modalities for improved classification.

The experimental results show that MobileNetV2 is better than other image-only models in terms of performance, so it reliable example of lightweight architectures in the case of agricultural imaging tasks. The multimodal MobileNetV2 + MLP model achieved a precision of $\sim 100\%$. which is better than the other, so the integration of environmental variables can significantly impact classification.

Table of Contents

<i>List of Abbreviations</i>	v
<i>List of Figures</i>	vi
<i>List of Tables</i>	vii
<i>Chapter 1 Introduction</i>	1
1.1 Problem Definition	1
1.2 Thesis Objectives	2
1.3 Thesis Contribution	2
1.4 Thesis Organization	3
<i>Chapter 2 Overview</i>	4
2.1 Overview of Plant Diseases and Early Detection Challenges	4
2.2 Image-Based Deep Learning for Plant Disease Classification	4
2.3 Role of Environmental and Climate Features	4
2.4 Multimodal Learning in Agriculture	5
<i>Chapter 3 Related Work</i>	6
<i>Chapter 4 Proposed Approach</i>	11
4.1 Dataset Description	11
4.1.1 Structured Climate Dataset	11
4.1.2 Image Dataset - Plant Leaf Images	12
4.1.3 Multimodal Paired Subset	15
4.2 Data Preprocessing	15
4.2.1 Image Preprocessing	15
4.2.2 Structured Data Preprocessing	16
4.3 Structured Feature Extraction Models	16
4.3.1 Multilayer Perceptron (MLP)	17
4.3.2 Random Forest	17
4.3.3 Decision Tree	17
4.4 Image Feature Extraction Models	18
4.4.1 MobileNetV2	19
4.4.2 ResNet50	19
4.4.3 InceptionV3	19
4.5 Multimodal Fusion Architecture	20
4.6 Training Strategy	22

4.7 Evaluation Metrics	23
<i>Chapter 5 Experimental Results and Discussion.....</i>	24
5.1 Experimental Setup	24
5.1.1 Hardware and Software Environment	24
5.1.2 Dataset Splitting.....	24
5.1.3 Training Configuration.....	24
5.2 Structured Data Model Results.....	24
5.3 Image-Only Model Results.....	25
5.4 Multimodal Fusion Model Results.....	27
5.5 Comparison Summary	31
<i>Chapter 6: Conclusion and Future Work</i>	33
6.1 Conclusion	33
6.2 Future Work.....	34
<i>References</i>	36

List of Abbreviations

- **AI:** Artificial Intelligence
- **DL:** Deep Learning
- **ML:** Machine Learning
- **CNN:** Convolutional Neural Network
- **DNN:** Deep Neural Networks
- **NLP:** Northern Leaf Blight
- **SVM:** Support Vector Machine
- **MLP:** Multi-Layer Perceptron
- **KNN:** K-Nearest Neighbors
- **ALDD:** Apple Leaf Disease Dataset
- **BF:** Bilateral Filtering
- **EPO:** Emperor Penguin Optimizer
- **N:** Nitrogen content in soil
- **P:** Phosphorus content
- **K:** Potassium content

List of Figures

Figure 1 Example of Apple Scab	13
Figure 2 Example of Black Rot.....	13
Figure 3 Example of Cedar Apple Rust.....	14
Figure 4 Example of Apple Healthy	14
Figure 5 Structured Data Classification Flow.....	18
Figure 6 Image Data Classification Flow	20
Figure 7 System Architecture of the Multimodal Deep Learning Model.....	21
Figure 8 MobileNetV2 Confusion Matrix	28
Figure 9 ResNet50 confusion matrix	29
Figure 10 InceptionV3 confusion matrix	30
Figure 11 Model Performance Comparison (Apple Disease Detection).....	31

List of Tables

Table 1 Summary of the Related Work	9
Table 2 Example of the Dataset	14
Table 3 Structured Data Model Results	24
Table 4 Image-Only Model Results	25
Table 5 Multimodal Fusion Model Results	27

Chapter 1 Introduction

1.1 Problem Definition

Diseases of plants have become the main factors that cause the food supply system of the entire planet to be at risk [1]. The changing climate that is continuously affecting the distribution and the severity of pathogens is making the problem worse [2]. In fact, the diseases on the apple leaves, mostly caused by Apple Scab, Black Rot, and Cedar Apple Rust, are the factors that mainly contribute to the reduction of both the quality and the quantity of the crop [3]. Most of the methods for disease detection that have been implemented are human character-based assessments, and they use handcrafted image processing techniques, which are not feasible since they are subjective, time-consuming, and, to some extent, their accuracy is also limited [4].

It is a fact that deep learning methods have played an important role in the recognition of diseases from images; however, most of the current models still only focus on the visual features [5]. Plant diseases should be the last ones to be identified only from the images, as they are very dependent on the environment. Taking the example of temperature, humidity, rainfall, and soil nutrient levels [6]. Very few models even consider such factors, which is their disadvantage as they cannot provide stronger and climate-aware predictions.

As a result, the problem of selecting a multimodal method that merges image features with environmental data for disease detection at the local level is an immediate one. This study answers that question by creating and assessing a deep learning-based system that integrates climate and soil analysis with CNN-based image analysis for more accurate apple disease classification.

1.2 Thesis Objectives

This thesis is mainly concerned with creating and evaluating the performance of deep learning models that employ multimodal fusion strategies to accurately classify apple diseases. It is geared towards the preprocessing and fusing of varied sources of data, such as images and numerical environmental variables, to generate standardized and high-quality inputs for the models.

By training various CNN architectures such as MobileNetV2, ResNet50, and InceptionV3, the thesis attempts to locate reliable image-based reference points, while a Multilayer Perceptron is employed to study climatic and soil features that affect the prediction.

Subsequently, a multimodal intermediate fusion model is designed to integrate the visual embeddings with the environmental inputs, thereby enabling the system to have access to the supplementary data of different modalities. The aim is to figure out if the environmental variables raise the prediction accuracy, to become an insightful source of ideas for future climate-aware plant disease detection systems.

1.3 Thesis Contribution

1. Evaluating the performance of three state-of-the-art CNN architectures, MobileNetV2, ResNet50, and InceptionV3, in capturing complex visual features of agricultural plant disease images. As a result, it stands as a scientifically based comparison of lightweight vs deep architectures in the agriculture sector.
2. Constructing a multimodal intermediate fusion architecture combining CNN image embeddings with the soil variables (temperature, humidity, pH, rainfall, N, P, K). The framework demonstrates the use of integration of different data sources (visual + climatic).
3. Conduct comparative performance experiments for image-only and multimodal models. The results demonstrate that environmental features do not necessarily improve image-based accuracy, which, in turn, highlights the problem of noise sensitivity and cross-modal imbalance.
4. Identified the practical limitations and future potentials of multimodal plant disease systems, among them the difficulty of modality alignment, feature mismatch, and data fusion sensitivity. These insights can be considered as a set

of pointers to the next research stages when developing scalable climate-aware agricultural AI models.

1.4 Thesis Organization

This thesis is organized into six chapters:

Chapter 1 – Introduction:

Presents the research problem, objectives, and contributions.

Chapter 2 – Overview:

Overview of the foundational concepts relevant to this research. It introduces deep learning techniques for image and structured data analysis, explains multimodal learning.

Chapter 3 – Related Works:

Summarizes previous work on plant disease detection, deep learning methods, environmental data modeling, and multimodal fusion.

Chapter 4 – Proposed Approach:

Describes the dataset, preprocessing steps, feature extraction procedures, and model development workflow.

Details the multimodal architecture, CNN models, MLP structure, and fusion strategy.

Chapter 5 – Experimental Results and Discussion:

Reports model performance, comparative evaluations, confusion matrices, and analysis.

Chapter 6 – Conclusion and Future Work:

Concludes the study and outlines potential directions for further research.

Chapter 2 Overview

This chapter talks about the foundational concepts relevant to this research. It shows how an imagined model was designed to combine visual and climatic features for apple disease detection.

2.1 Overview of Plant Diseases and Early Detection Challenges

Throughout history, plant diseases have been the major factors that have caused the decrease in food produced through agriculture all over the world [1]. A quick and accurate infection identification can, therefore, bring about a very significant reduction of crop losses and lessen the use of pesticides, thus making eco-friendly farming methods more feasible [7]. Beforehand, the most common method of disease identification was heavily reliant on the manual inspections carried out by agricultural experts. The entire process was not only time-consuming but also biased and thus could bring about incorrect diagnoses [8].

As a result of the limitations mentioned above, automated disease recognition systems have been devised to solve these problems, which in turn have led to the rapid development of datasets and machine learning technologies [5].

2.2 Image-Based Deep Learning for Plant Disease Classification

Usually, the key tools that brought about the revolutionary changes in the visual recognition of plant diseases in the last few years were Convolutional Neural Networks (CNNs) owing to their outstanding feature extraction abilities [8]. Different CNN-based architectures like MobileNetV2, ResNet50, and InceptionV3 have been extensively used, and their effectiveness has been validated by the agro-industrial sector [9].

2.3 Role of Environmental and Climate Features

For the spread and the development of plant diseases, the major contributors have been confirmed to be environmental factors like temperature, humidity, and rainfall. Even more, the factors nitrogen (N), phosphorus (P), potassium (K), and pH in the soil are so important that they determine the health of the plants and their resistance to diseases [10].

Scientists who emphasize the significance of environmental features as predictors have also given their support points with these facts:

- The fungi that cause plant diseases may become very fast and very aggressive in situations that are humid and have long periods of leaf wetness [14].
- The temperature changes have a direct effect on the activity of pathogens.
- Excess or deficiency of nutrients weakens the plant's immune system and thereby makes it less resistant [15].
- The plant's immune system gets weaker and, therefore, less resistant if it is supplied with either too many or too few nutrients.

Machine learning models based on environmental attributes, such as Random Forest, SVM, and MLP, have been used for the prediction of disease risk levels in many research works [16]. However, since they do not consider visual symptoms, their effectiveness as a single is not that great.

2.4 Multimodal Learning in Agriculture

Multimodal learning stands for the utilization of various data sources to get a more complete understanding. The use of photos together with either environmental or sensor data has been the means of the agro-industrial sector in terms of disease prediction accuracy improvement [17].

Multimodal Fusion Types:

1. Early Fusion: The feature integration works before learning (extremely few).
2. Late Fusion: The final merging of the results from different models is done at the last stage.
3. Intermediate Fusion (Used in this thesis): The visual and environmental feature vectors are retrieved separately, then they are merged, and dense layers are applied for joint learning [18].

Chapter 3 Related Work

Zainab et al. [19] applied Convolutional Neural Network (CNN) models: ResNet152, VGG19, DenseNet169, and MobileNetv3 to detect plant diseases using the highly regarded PlantVillage dataset. In addition, image augmentation, Keras' Image Data Generator, was used to extend the training set. The results showed that MobileNetv3 had the best testing accuracy of 99.75%, ResNet152 accuracy of 98.86%, VGG19 accuracy of 99.07% and DenseNet169 accuracy of 99.22%.

Shoaib, Muhammad, et al. [20] applied U-Net and Modified U-Net to extract diseased tomato leaves, then they applied different CNN models from the Inception Net family, specifically InceptionNet1, InceptionNet2, and InceptionNet3, for classifying tomato leaf diseases that were trained over 18,161 segmented and non-segmented tomato leaf images from the PlantVillage dataset. In addition, an examination of the presentation of binary arrangement (healthy and diseased leaves), six-level classification (one healthy class and five distinct unhealthy classes: Fungi, Bacteria, Mould, Virus, and Mite), and ten-level classification (one healthy class and nine disease categories: Early blight, Septoria leaf spot, Target spot, Leaf mold, Bacterial spot, Late blight mold, Tomato Yellow Leaf Curl Virus, Tomato Mosaic Virus, and Two-spotted spider mite). They achieved high performance in both segmentation and classification of tomato leaf diseases. Using a modified U-Net model, it reached about 98.66% segmentation accuracy, 98.5% IoU score, and 98.73% percent on the dice.

Ahmad et al. [21] utilize transfer learning using five different pre-trained deep neural networks (DNN): InceptionV3, ResNet50, VGG16, DenseNet169, and Xception using five datasets consisting of foliar disease images in corn: namely PlantVillage, PlantDoc, Digipathos, northern leaf blight (NLB) dataset, and CD&S dataset. It was observed that using DenseNet169 resulted in the best generalization for plant disease identification, capable of achieving accuracies up to 81.60%, InceptionV3 up to 73.27%, ResNet50 up to 70.60 %, VGG16 up to 52.55% and Xception up to 78.80%.

Agarwal, Mohit, et al. [22] apply CNN: VGG16, InceptionV3, and MobileNet for disease detection and classification. Images of Tomato disease have been taken from the PlantVillage dataset, which includes over 50,000 images of 14 crops. Experimentally, they observed testing accuracy ranging from 76% to 100% for the classification. And the average testing accuracy of the model is 91.2% for the 9

diseases and 1 healthy class. Where the Custom CNN is the best accuracy up to 91.2%, VGG-16 up to 77.2%, Inception V3 up to 63.4% and MobileNet up to 63.7%.

Chen, Junde, et al., [23] apply CNN models VGGNet, InceptionV3, ResNet50, VGGNet-19, and DesneNet201. They selected the VGGNet pre-trained on ImageNet for the identification of plant leaf diseases. Instead of starting the training from scratch, they initialize the weights using the pre-trained networks on the large, labeled dataset, ImageNet, including 500 rice images and 466 maize images. It achieves a validation accuracy of no less than 91.83% on the dataset.

DeChant, Chad, et al. [24] applied Several CNNs to detect Northern Leaf Blight (NLB) symptoms in maize. They trained CNNs to classify small regions of images containing NLB, then combined them into separate heat maps, and then fed them to a final CNN trained to classify the entire image as containing diseased plants or not. Data were used from two separate plantings, both grown at Musgrave Research Farm in Aurora, New York. Each planting consisted of two replicates, each containing 250 maize hybrids, including five experimental hybrids adapted to New York. The system achieved 96.7% accuracy.

Sharma, Rahul, et al. [25] applied DL-based computer vision techniques like CNN and ML-based image classifiers such as Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree, and Random Forest to identify plant diseases. The Rice Leaf Dataset with 5932 images and 1500 potato leaf images from PlantVillage was used. Results demonstrate that the proposed CNN model performed better when compared with other machine learning image classifiers, reaching an accuracy reach 99.58% on the rice dataset and 97.66% on the potato dataset.

Wang et al. [26] used apple black rot images in the PlantVillage dataset; a series of CNNs: VGG16, VGG19, Inception-v3, and ResNet50 were trained to diagnose the severity of the disease. Results show that the best model is the deep VGG16 model trained with transfer learning, which yields an overall accuracy of 90.4%.

Too et al. [27] focus on fine-tuning and evaluation of deep CNN, VGG 16, Inception V4, ResNet with 50, 101, and 152 layers, and DenseNets with 121 layers for image-based plant disease classification. The data used for the experiment is 38 different classes, including diseased and healthy images of leaves of 14 plants from PlantVillage. Results proved that DenseNets tend to consistently improve in

accuracy with a growing number of epochs, with no signs of overfitting and performance deterioration. Moreover, DenseNets require a considerably smaller number of parameters and reasonable computing time to achieve performance. It achieves a testing accuracy score of 99.75% to beat the rest of the architectures.

Eunice, Jennifer, et al. [28] applied CNN-based pre-trained models such as DenseNet-121, ResNet-50, VGG-16, and Inception V4 for efficient plant disease identification. They conducted experiments using the popular PlantVillage dataset, which has 54,305 image samples of different plant disease species in 38 classes. Experiments proved that DenseNet-121 achieved a classification accuracy of 99.81% and F1-score of 99.8%.

Atila, Ümit, et al. [29] proposed EfficientNet B5 and B4 models of DL architecture in plant leaf disease classification. All models were trained with the original and augmented PlantVillage datasets, having 55,448 and 61,486 images. They compared performance with other state-of-the-art DL models. The results showed that B5 and B4 models of EfficientNet architecture achieved the highest values compared to other DL models, with 99.91% and 99.97% respectively, for accuracy.

Jiang, Peng, et al. [30] propose a DL approach that is based on improved CNN (AlexNet, GoogLeNet, InceptionV3, ResNet-101, ResNet-50, ResNet-34, ResNet-18, and VGGNet-16) for the real-time detection of apple leaf diseases. They used the apple leaf disease dataset (ALDD), which is composed of laboratory images and complex images under real field conditions, including 26,377 images of diseased apple leaves. Based on that, they proposed the INAR-SSD (SSD with Inception module and Rainbow concatenation) model to detect these five common apple leaf diseases. Results show that the INAR-SSD model realizes a detection performance of 78.80% mAP on ALDD, with a high-detection speed of 23.13 FPS.

Amara, J, et al. [31] use the LeNet architecture as a CNN to classify and detect two famous banana diseases, which are banana sigatoka and banana speckle, using a real dataset of banana diseases obtained from the PlantVillage. The results demonstrated with 20% training and 80% testing, they achieved: 98.61% accuracy, 98.67% precision, 98.60% recall, and 98.64% F1-score.

Ashwinkumar, S. et al., [32] propose an automated model for detecting and classifying plant leaf diseases using an optimal mobile network-based CNN

(OMNCNN) on the benchmark plant leaf disease dataset. Furthermore, they involve bilateral filtering (BF) based preprocessing and Kapur's thresholding-based image segmentation to identify the affected portions of the leaf image, and the MobileNet model is applied as a feature extraction technique in which the hyperparameters are optimized using the emperor penguin optimizer (EPO) algorithm to enhance the plant disease detection rate. The experimental outcome has shown promising results of the OMNCNN model, a maximum precision of 0.985, a recall of 0.9892, an accuracy of 0.987, an F-score of 0.985, and a kappa of 0.985.

Bhattacharya et al [33] utilized the DenseNet201 neural network architecture with pre-trained ImageNet weights. Rather than assigning random weights, they initialized the model with these pre-trained weights and then trained the layers from scratch on their dataset, which consisted of five different classes of rice leaf diseases. The result shows better performance by achieving the training accuracy of 97.04 % and an accuracy of 95.44 % on the validation dataset.

Table 1 Summary of the Related Work

Papers	Year	Methods	Datasets	The best model	Results
[22]	2025	CNN models: ResNet152, VGG19, DenseNet169, and MobileNetv3	PlantVillage	MobileNetv3 accuracy of 99.75%	MobileNetv3: 99.75% ResNet152: 98.86% VGG19: 99.07% DenseNet169: 99.22%
[23]	2022	DL, CNN and semantic segmentation models: U-Net and Modified U-Net. Binary arrangement, six-level and ten-level classification	PlantVillage	Modified U-Net model	98.66% segmentation accuracy, 98.5% IoU score, and 98.73% on the dice.
[24]	2023	DNN Models: InceptionV3, ResNet50, VGG16, DesneNet169, and Xception	PlantVillage, PlantDoc, Digipathos, NLB and CD&S dataset	DesneNet169 accuracy of 81.60%.	InceptionV3:73.27% ResNet50:70.60 % VGG16:52.55% DesneNet169: 81.60 % Xception:78.80%
[25]	2020	VGGNet, InceptionV3, ResNet50, VGGNet-19 and DesneNet201	Planet disease detects and ImageNet	VGGNet achieves a validation accuracy of no less than 91.83%	VGGNet 91.83% InceptionV3 85% ResNet50 69.67% VGGNet-19 74.83% and DesneNet201 79%
[26]	2017	CNN	Musgrave Research Farm and NLB	CNN accuracy reached 96.7%	The system achieved 96.7% accuracy

			annotated public 2016 datasets		
[27]	2021	CNN, SVM, KNN, Decision Tree and Random Forest	PlantVillage	CNN accuracy reaches 99.58% on rice and 97.66% on potato	CNN: 99.58% and 97.66% SVM: 93.76% and 92% KNN: 68.95% and 70% Decision Tree: 90% and 66.66% Random Forest: 96.83% and 85%
[28]	2017	CNN: VGG16, VGG19, Inception-v3 and ResNet50	PlantVillage	VGG16 accuracy of 90.4%.	deep VGG16 model with transfer learning accuracy of 90.4%.
[29]	2017	CNN,VGG 16, Inception V4, ResNet with 50, 101 and 152 and DenseNets-121	PlantVillage	DenseNet-121 achieved accuracy of 99.75%	DenseNet-121 achieved the highest testing accuracy of 99.75%, with robust performance, minimal overfitting, and fewer parameters compared to others
[30]	2022	Inception V4, VGG-16, ResNet-50 and DenseNet-121	PlantVillage	DenseNet-121 achieved a classification accuracy of 99.81% and F1 score of 99.8%.	Inception V4: 97.59% VGG-16: 82.75% ResNet-50: 98.73% DenseNet-121: 99.81%
[31]	2020	EfficientNet B5 and B4	PlantVillage	EfficientNet achieved accuracy of 99.97% and 99.39% for precision.	EfficientNet with 99.91% and 99.97% respectively for accuracy and 98.42% and 99.39% respectively for precision.
[32]	2019	INAR-SSD AlexNet, GoogLeNet, InceptionV3, ResNet-101, ResNet-50, ResNet-34, ResNet-18 and VGGNet-16	ALDD	INAR-SSD mAP 78.80%	INAR-SSD performance of 78.80% mAP with a high detection speed of 23.13 FPS.
[33]	2017	CNN: LeNet	PlantVillage	LeNet achieved accuracy of 98.61%	Accuracy: 98.61%, Precision: 98.67% Recall 98.60% and F1-score: 98.64%
[34]	2021	OMNCNN, BF, Kapur's, EPO	Benchmark Dataset	OMNCNN accuracy of 98.70%	Maximum precision of 0.985, recall of 0.9892, accuracy of 0.987, F-score of 0.985, and kappa of 0.985.
[35]	2021	CNN DenseNet 201	Balanced dataset	DenseNet 201 achieved accuracy of 97.04%	Training accuracy of 97.04 % and accuracy of 95.44 % on the validation dataset.

Chapter 4 Proposed Approach

This chapter explains a multimodal deep learning system that attempts to diagnose apple leaf diseases through the combination of visual data and environmental data. The method here is entirely dependent on the Multimodal Plant Disease Dataset by Shubham Divakar (Kaggle) [34], focusing on the apple subset. The approach integrates image feature extraction, structured data modeling, and multimodal fusion.

4.1 Dataset Description

The data utilized in this thesis is a subset of images of apples taken from the Multimodal Plant Disease dataset by Shubham Divakar on Kaggle, which has 40,000 images spanning multiple plant species. The dataset consists of two main parts:

4.1.1 Structured Climate Dataset

The structured dataset is well balanced because the same number of samples has been taken for each class, i.e., 10,000 per disease category in the apple subset. Such a balance allows for the standard training and gaining stable results across all disease categories. The environment and soil are described with the following features:

- Nitrogen (N)
- Phosphorus (P)
- Potassium (K)
- Temperature
- Humidity
- Rainfall
- pH

All these factors refer to the air and soil surrounding the sample location. However, it is important to note that these environmental records were artificially generated and merged in such a way that each apple disease class had exactly 10,000 structured samples. Therefore, the structured subset is not just a perfect balance but is also considerably more than the number of leaf images.

4.1.2 Image Dataset - Plant Leaf Images

The images are close-up shots of plant leaves of good quality, which show diseased leaves as well as the healthy ones. Images are suitable for training deep learning methods like CNNs, MobileNet, ResNet, or Inception. Although the structured dataset may give 10,000 samples for each class, the actual number of available images after filtering and validating the apple images is only 3,171 usable images. The distribution per class :

- Apple Scab: 630 images
- Apple Black Rot: 621 images
- Cedar Apple Rust: 275 images
- Apple Healthy: 1645 images

The difference of 40,000 structured samples versus 3,171 images can be explained as follows:

- CSV files contain all the environment/soil records for which images are missing, i.e., cases where no picture was taken.
- Some pictures might have been lost, corrupted, duplicated, or, during the cleaning and preprocessing stages, removed.
- The procedure of photo shooting is limited by such factors as camera settings, field conditions, and the real world
- As a result, the number of structured samples will always be greater than that of the leaf images.

This distribution clearly shows a serious class imbalance, wherein the Healthy class comprises more than half of all samples, while the Cedar Apple Rust class is greatly underrepresented. Such an imbalance required making use of class weighting, oversampling, and stratified splitting at the time of model training for fair learning and to avoid model bias toward the majority class.

The cause of this difference is quite basic and logical:

- CSV files are the sources of all the environment/soil records that were made for each category of diseases, including those for which no image was taken.
- Some images were missing, corrupted, duplicated, or deleted during the cleaning and preprocessing stages.

- The provision of images is dependent on the condition of the site, the setting of the camera, and the limitations of the real world.
- The number of structured samples will always be higher than the available leaf images.

The experimental design remains valid despite the difference in the number of samples between the two modalities:

- The structured apple subset is perfectly balanced (10,000 samples per class).
- The image-only subset is imbalanced; however, class weights and oversampling are used explicitly in the training split to correct it.
- The multimodal models only operate on the intersection of samples where both image and structured data are available.

Therefore, each modality is consistently handled within its own experimental setting, and the difference in the absolute number of samples between structured and image data does not undermine the multimodal evaluation.



Figure 1 Example of Apple Scab



Figure 2 Example of Black Rot



Figure 3 Example of Cedar Apple Rust

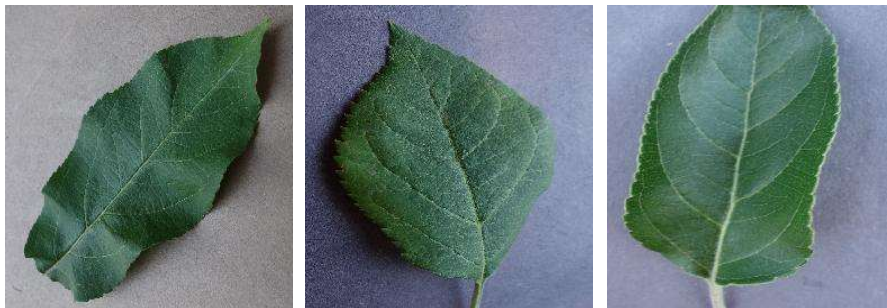


Figure 4 Example of Apple Healthy

Table 2 Example of the Dataset

N	P	K	temperature	humidity	ph	rainfall	Label	Image
71	54	16	22.61359953	63.69070564	5.749914421	87.75953857	Apple Black rot	
90	42	43	20.8797437	82.0027442	6.50298529	202.935536	Apple scab	
40	72	77	17.0249846	16.9886117	7.48599607	88.5512314	Cedar Apple Rust	

13	60	25	17.1369277	20.5954169	5.68597166	128.256862	Apple Healthy	
----	----	----	------------	------------	------------	------------	------------------	---

4.1.3 Multimodal Paired Subset

In the multimodal experiments, the data for each training has to consist of an image and the corresponding ordered data record, which must share the same photo name. The apple CSV subset was merged with the apple image folders using the filename as a key to ensure that this was the case, and only those samples that had a valid match in both modalities were kept.

The final multimodal apple subset after this strict matching and dropping of duplicates contained 400 paired samples split as follows:

- 262 training samples
- 70 validation samples
- 68 test samples

The strictly aligned smaller multimodal subset serves as the basis for the fusion experiments, where CNN-based image features are combined with MLP-based structured features.

Final Alignment for Multimodal Samples:

After the structured records were preprocessed and paired with the matching apple images, only 400 valid multimodal samples were left. These were then split into training (262), validation (70), and test (68) subsets. This aligned subset guarantees that each multimodal sample consists of image and structured data.

4.2 Data Preprocessing

4.2.1 Image Preprocessing

- Resizing the picture size to 224×224

Technically, image resizing indicates altering the spatial dimensions (width and height) of an image to a certain fixed size. CNNs demand that all input images be of the same shape. Hence, the images of apple leaves in this study were all resized to 224×224 pixels to be fed to MobileNetV2, ResNet50, and

InceptionV3. Apart from that, this measure is not only necessary for the standards but also lowers the memory consumption and makes the CNN architectures less complex.

- Normalization (0–1 scaling)

Normalization refers to the procedure of rescaling the raw pixel to a small and consistent numerical range. The RGB images are made up of pixels, each having values between 0 and 255, and these may lead to the model getting unstable during training. Consequently, the pixel values were normalized by dividing them by 255 in such a way that the range [0, 1] is obtained. This step guarantees numerical stability, it converges faster, and it agrees with the final accuracy of deep learning models.

- Data augmentation

Images have been subjected to a number of data augmentation techniques so that the images can be generated artificially and the overfitting problem can be alleviated. The augmentations that have been implemented:

- Rotation: the leaves were turned randomly to visually demonstrate different orientations of leaves.
- Flipping: images that were left or right were created through horizontal flipping.
- Brightness Adjustment: the change of the brightness to imitate the natural lighting.

4.2.2 Structured Data Preprocessing

- Standardization using StandardScaler

- Pairing up the pictures with the correct pieces of structured data via the image_id keys

- Label encoding

The structured data have been label encoded, scaled (for MLP), and stratified into training, validation, and test sets.

4.3 Structured Feature Extraction Models

Seven numerical environmental and soil features (N, P, K, temperature, humidity, pH, and rainfall) were the inputs. The evaluation was done through the four apple disease categories. Apply three machine learning models:

4.3.1 Multilayer Perceptron (MLP)

It is a feed-forward artificial neural network architecture that consists of an input layer, one or more hidden layers, and an output layer. First, MLPs do weighted linear transformations, and then, to figure out complex interactions in the data, they apply nonlinear activation functions.

The training of MLPs is done through backpropagation, which changes the weights to minimize the loss function. MLPs are generally a good choice for structured numerical data classification and regression problems (35).

4.3.2 Random Forest

Random Forest is an ensemble machine learning method that, during its training phase, builds multiple decision trees, and the final decision is arrived at by combining the trees' votes through majority voting for classification or averaging for regression.

The trees are trained on different features and sample subsets that are randomly chosen; hence, the model is less overfitting. It also has a better capacity for generalization. Random Forests are strong, easy to train, and very efficient for structured tabular data sets (36).

4.3.3 Decision Tree

A Decision Tree is a supervised machine learning technique that, for a target variable, derives the simplest decision rules from data features, thereby making the predictions.

The model divides data into smaller datasets or "branches" using "if-else" conditions. Each internal node represents a decision made on a feature, each branch shows the outcome of the decision, and each leaf node points to the final class prediction.

Decision trees are quite simple and can be easily explained since they work in the same way as humans do when making decisions by following rule-based paths. Besides, they can represent linear as well as nonlinear relationships in structured tabular data. However, if an individual tree is used, it can become overfitted; thus, decision trees are most frequently taken to the next level by being transformed into ensemble methods such as Random Forests (37).

In the apple subset of this dataset, the seven structured features (N, P, K, temperature, humidity, pH, rainfall) are almost perfectly separable across the four disease classes, because the original structured subset was generated in a very controlled and balanced way. Consequently, all three structured only models (MLP, Random Forest, and Decision Tree) were able to achieve perfect or near-perfect accuracy, precision, recall, and F1 score on the test set. Hence, these results should be considered as an upper limit under ideal synthetic conditions rather than as the potential performance on real-world environmental data.

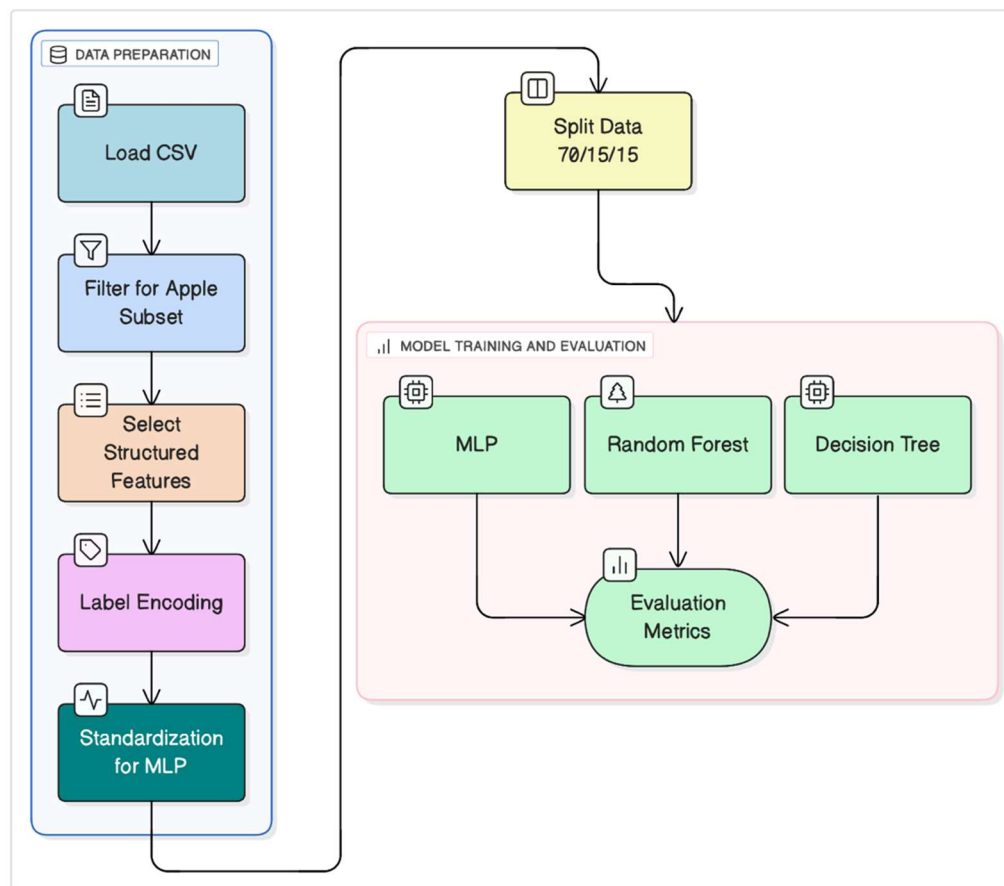


Figure 5 Structured Data Classification Flow

4.4 Image Feature Extraction Models

The performance comparison experiment involved the consideration of three different pretrained CNN models:

4.4.1 MobileNetV2

MobileNetV2 is a minimal model that applies depthwise separable convolutions for the reduction of operations while maintaining a relatively good accuracy level [10]. In this regard, it can be a model that perfectly fits the needs of small-scale agricultural datasets. Several research papers have confirmed that MobileNet versions are very effective in achieving high precision in leaf disease classification with low overfitting.

4.4.2 ResNet50

ResNet50 is centered around the concept of residual connections, which allow deeper architectures without the problem of vanishing gradients. However, it is generally held that ResNet models give their best results when they are trained on large datasets; hence, their performance on small agricultural datasets may not be stable [11].

4.4.3 InceptionV3

InceptionV3 utilizes multi-scale convolutional modules that can locate very tiny details [15]. However, if faced with very similar disease classes or very few training samples, the method, which is generally highly competent in image recognition tasks, might face problems.

Essentially, methods based on CNN have advanced a lot in plant disease detection; however, just image-based models may not be capable of solving the problem of poor lighting, occlusions, or visually indistinguishable disease symptoms.

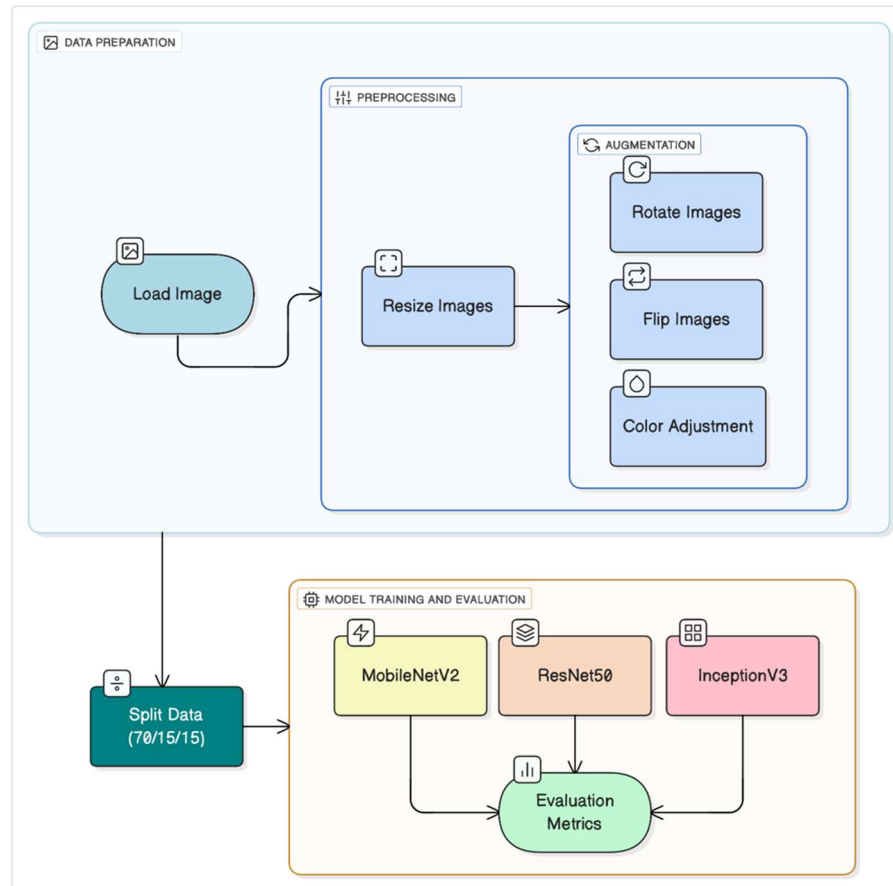


Figure 6 Image Data Classification Flow

4.5 Multimodal Fusion Architecture

The suggested layout has two parallel processing streams, one for the image feature extractor based on CNN, and the other for the climate feature extractor based on MLP. Even though each structured-only scenario on the apple subset ended with a perfect outcome, the MLP was selected as the main structured branch in the multimodal fusion architecture because of these reasons:

1. **Neural Compatibility:** The MLP generates continuously variable numerical embeddings that can straightforwardly be combined with the CNN image embeddings.
2. **Fusion Architecture Requirement:** The multimodal model is dependent on the joining of the learned representations from both branches. Regular tree-based models like Random Forest and Decision Tree do not yield neural embeddings, and, hence, they cannot be involved in the neural fusion layers.

3. Deep Learning Pipeline Compatibility: The utilization of MLP ensures that both the branches of the multimodal model are of the same deep learning nature and, thus, they can be optimized in a unified manner via backpropagation.
4. Superior Representation Learning: MLPs can identify complex nonlinear patterns in the structured data, which gives them the ability to uncover the interaction of environmental and soil features.

Consequently, the MLP was the one to be put forward, not with the accuracy as the sole reason, but rather because it happens to be the only model that satisfies the specifications of the multimodal deep learning architecture.

The outputs of both streams are concatenated and passed through the dense layers with dropout for the final classification with Softmax.

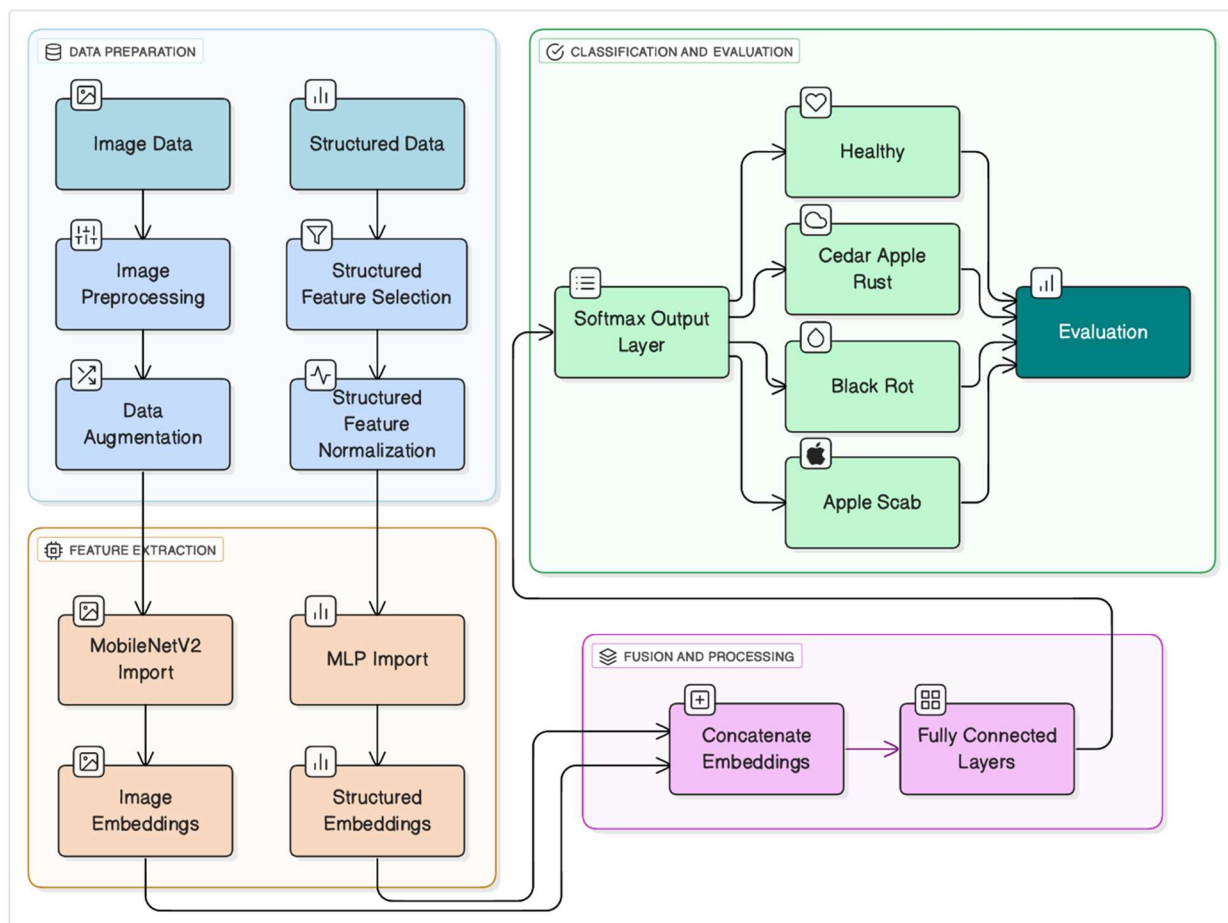


Figure 7 System Architecture of the Multimodal Deep Learning Model

4.6 Training Strategy

The training was very detailed and thorough to achieve a final model that is well-balanced and generalizable; thus, it would be an ideal fit for the multimodal architecture. To efficiently maintain a good balance of computational power and gradient stability, a batch size of 32 was selected. Essentially, the Adam optimizer came out as the best choice mainly due to its adaptive learning rate feature, which, overall, very much facilitates convergence in deep learning model training.

In the majority of the image-only and multimodal experiments, the epoch count was fixed at 20, with early stopping validation loss monitoring to prevent overfitting. As a matter of fact, numerous models converged at an earlier point, and the training was automatically terminated at the optimal epoch.

Due to the high separability of the tabular features in the apple subset, the structured only MLP achieved convergence in a very short period. Within the first few epochs, the model hit 100% validation accuracy, which is an expected outcome given the nature of the controlled and highly distinguishable structured dataset. To avoid overfitting, in particular in the image-only models that were affected by class imbalance, class weights, data augmentation, and early stopping were implemented. The use of these measures ensured that the models did not resort to memorization and were still applicable to unseen samples.

Softmax Activation Function:

For all classification models, the last Dense layer, i.e., the image-only models, the structured-only MLP, and the multimodal fusion model used Softmax as the activation function. Softmax changes the raw output logits to a probability distribution over the four apple disease classes.

$$\text{Softmax}(z_i) = e^{(z_i)} / \sum e^{(z_j)}$$

This ensures that all outputs lie between 0 and 1, and their sum equals 1.

Loss Function: Categorical Cross-Entropy

The decision to use Categorical Cross-Entropy as the loss function in all deep learning experiments was taken. This function calculates the difference between the true one-hot labeled vector and the predicted probability distribution provided by the Softmax layer.

$$\text{Loss} = - \sum y_i * \log(\hat{y}_i)$$

where y_i represents the true label and $\hat{y}_i$ the predicted probability. This loss function is the standard choice for multi-class classification problems.

4.7 Evaluation Metrics

Accuracy:

Accuracy is basically a measure that compares the number of times the model makes the right prediction with the total number of cases.

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

Precision:

Precision is a measure to quantify the number of true positive cases predicted by the model among all positive cases the model predicted.

$$\text{Precision} = TP / (TP + FP)$$

Recall:

Recall is the percentage of the positive instances that have been identified correctly by the model.

$$\text{Recall} = TP / (TP + FN)$$

F1-Score:

F1 score is the one metric that comes from combining precision and recall via their harmonic mean.

$$F1 = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

Chapter 5 Experimental Results and Discussion

This chapter presents the outcome of experiments on the Apple subset of the Multimodal Plant Disease Dataset by Shubham Divakar.

5.1 Experimental Setup

5.1.1 Hardware and Software Environment

The work has been done on the Mac OS platform, and the following packages have been used: Python, Jupyter Notebook, TensorFlow, NumPy, Pandas, and scikit learn and Matplotlib.

5.1.2 Dataset Splitting

The dataset was divided into three subsets with the following ratio: 70/15/15 for training, validation, and testing. The split was applied for structured-only and Image-only experiments.

Only those samples that contained both an image and structured data were taken into consideration for the multimodal fusion experiments. After the alignment, there were 400 multimodal samples from which 262 samples were chosen for training, 70 for validation, and 68 for testing.

5.1.3 Training Configuration

All models were trained for 20 epochs using the Adam optimizer and a batch size of 32. Early stopping was applied based on the validation loss to prevent overfitting.

5.2 Structured Data Model Results

Table 3 Structured Data Model Results

Model	Dataset	Accuracy	Precision	Recall	F1-Score
MLP	Apple Disease Dataset	1.0000	1.0000	1.0000	1.0000
Random Forest	Apple Disease Dataset	1.0000	1.0000	1.0000	1.0000
Decision Tree	Apple Disease Dataset	1.0000	1.0000	1.0000	1.0000

All three models trained only on structured data have shown perfect results on the whole set of 40,000 structured samples (10,000 per class). Such an outstanding reproducibility is a strong signal that the environmental and soil measurements provided in the dataset are very discriminative, with almost no overlap between the four apple disease categories. The feature space is so clearly distinguishable that even very simple models have managed to classify all samples without any errors. Among the three, the Decision Tree, although the least complex, was the one that achieved 100% accuracy. To put it differently, it very strongly indicates that the decision boundaries in the structured feature space are quite obvious, simple, and almost linearly separable, hence no complicated transformations are required to separate the classes.

Similarly, the Random Forest model that merges the results of several trees was able to find the same patterns with complete accuracy. As a rule, the ensemble character of a model like this would make it more resistant to noise, but in the present case, all the trees came to the same stable decision rules; thus, the structured data is even more well defined.

Moreover, the MLP model also reached perfect performance, which can be taken as a strong indication that neural networks are capable of learning the interrelations from the tabular features. Very rapid convergence and extremely low validation loss are the proofs that the dataset has very strong and noise-free feature label associations.

The experimental results, in general, provide a very strong argument that the structured environmental and soil variables alone could be a sufficiently reliable source for the prediction of apple disease categories at an almost perfect level of certainty. The data in the structured modality represent a deterministic mapping to the disease labels, thus making them an exceptionally strong predictor that can be used alone in this dataset.

5.3 Image-Only Model Results

Table 4 Image-Only Model Results

Model	Dataset	Accuracy	Precision	Recall	F1-Score
MobileNetV2	Apple Disease Dataset	0.9936	0.997	0.992	0.9944

ResNet50	Apple Disease Dataset	0.9957	0.9947	0.9947	0.9946
InceptionV3	Apple Disease Dataset	0.9978	0.9973	0.9989	0.9981

The three image-only deep learning models MobileNetV2, ResNet50, and InceptionV3, were very effective in identifying apple leaf diseases. Each model exceeded 99% accuracy, which suggests that the four classes of apple diseases based on their visual symptoms are easily distinguishable.

By far, the most lightweight model, MobileNetV2, clocked an accuracy of 99.36%. This can be attributed to the depthwise separable convolutions and the use of a lightweight architecture. By producing clear and compact features, MobileNetV2 makes agricultural image datasets, in which leaf textures are typically repetitive and quite uniform, more manageable.

Serving as the slightly more accurate model, ResNet50 achieved a 99.57% precision rate, implying that the model with residual connections can still generalize well even when the dataset is small. The model managed to identify the slight variations in the textures of the four apple disease classes.

By a considerable margin, the model with the highest accuracy, but least intuitive of the three models, was InceptionV3 with 99.78% precision. It made use of its multi-scale convolutional modules to locate very small and intricate patterns in the leaf images. Tasks of the kind that involve minute distinctions between classes are Inception's expected behavior.

Most importantly, the three CNN models were able to learn disease features very reliably just from the visual modality alone, which is highly informative. The sustained performance of the three models, MobileNetV2, ResNet50, and InceptionV3, affirms that the images of the apple leaves are of good quality and contain well-proven disease patterns that allow classification with high accuracy and minimal mistakes across the different classes.

5.4 Multimodal Fusion Model Results

Table 5 Multimodal Fusion Model Results

Model	Input Type	Accuracy	Precision	Recall	F1-Score
MobileNetV2 + MLP	Image +Climate	1	1	1	1
ResNet50 + MLP	Image + Climate	0.9852	0.9861	0.9861	0.9857
InceptionV3 + MLP	Image + Climate	0.9558	0.9521	0.9529	0.9521

Multimodal fusion experimentation provides evidence that adding visual features to numerically represented environmental attributes leads to a very good and stable classification performance, irrespective of the chosen architecture.

The fusion model MobileNetV2 + MLP achieved a perfect classification performance of 100% under all the evaluation metrics. Such a result might seem unusually high, but it corresponds with the characteristics of the aligned multi-modal subset used in this experiment. After filtering and pairing the structured features with available Apple images, the resulting multi-modal dataset was small, clean, and highly separable with very limited intra-class variation. On the other hand, MobileNetV2 creates compact and smooth visual embeddings that would combine exceptionally well with the numerical climate features created by the MLP. This compatibility allows the fusion layer to form clear decision boundaries between classes, thus allowing perfect separation in the test split. The absence of noisy samples combined with the strong discriminative power of both modalities explains why the model was able to achieve 100% accuracy without signs of instability or overfitting during training.

By an accuracy of 98.5%, the ResNet50 + MLP model made a point about the residual architectures still being able to gain from multimodal fusion thanks to their strong representational power. However, its performance was a bit distant from that of MobileNetV2. This suggests that deeper networks might need a bigger and more diverse multimodal dataset to be able to use their depth without overfitting or memorizing noise.

The InceptionV3 + MLP fusion network achieved 95.6% precision, thus signaling a correlation between the multimodal and single modality experiments for this model.

InceptionV3 is a good multi-scale feature extractor; nevertheless, some apple disease classes in the dataset are visually almost identical, and the structured features weren't varied enough to give a significant increase in performance.

Summing up, the evidence suggests that in this particular dataset, the MobileNetV2 backbone is the best choice for multimodal fusion. Its proficient and well-balanced feature extraction perfectly matches the structured climate variables and makes it possible for the model to use visual as well as numerical modalities simultaneously. Moreover, these results point towards the promising future of multimodal learning in the field of agricultural disease diagnosis while, at the same time, revealing that the amount of improvement is considerably dependent on how compatible the CNN architecture is with the structured data.

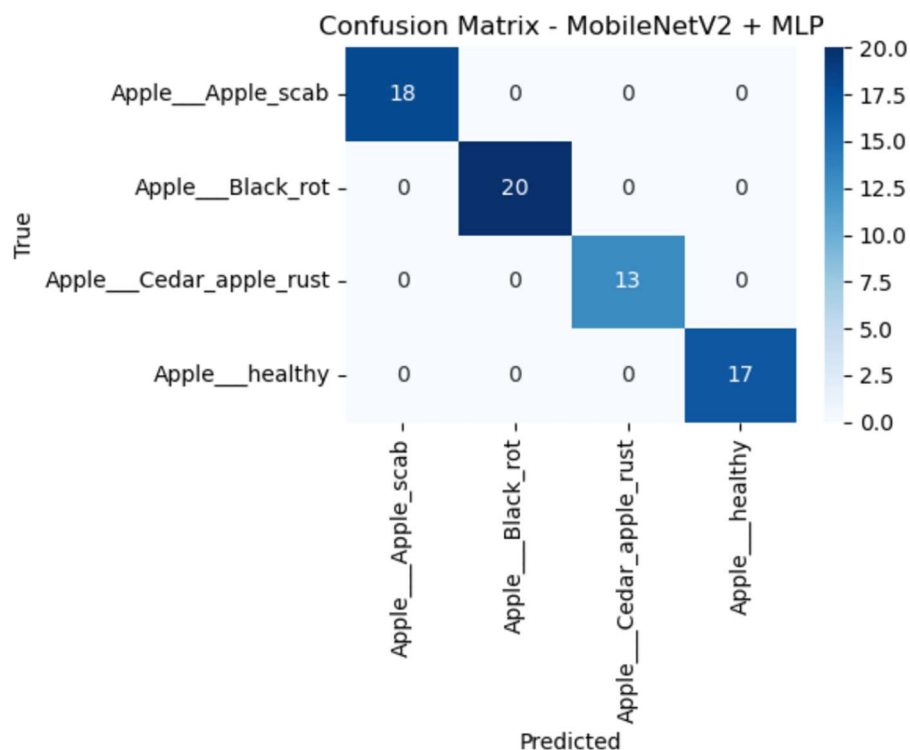


Figure 8 MobileNetV2 Confusion Matrix

The confusion matrix represents perfect classification for the MobileNetV2 + MLP multimodal model for the four classes of apple diseases, where each prediction falls exactly on the diagonal. Therefore, will mean that all the test samples have been correctly predicted by this model, giving it an approximate accuracy of 100%.

Performance indicates a strong compatibility between the compact visual embeddings from MobileNetV2, and numerical representations generated from the

MLP. The lightweight architecture of MobileNetV2 generates smooth, low-dimensional features that seem to fuse seamlessly with the structured environmental variables.

The confusion matrix also shows that environmental and soil attributes contributed to the stabilization of model decision boundaries, thereby enabling the model to reinforce differences between diseases that may even appear visually similar, such as Apple Scab vs. Black Rot. In other words, the ideal confusion matrix means that the MobileNetV2 + MLP multimodal fusion model was able to take full advantage of both modalities to achieve perfect discrimination across all classes, confirming this architecture as one of the strongest in the scope of the aligned multimodal dataset.

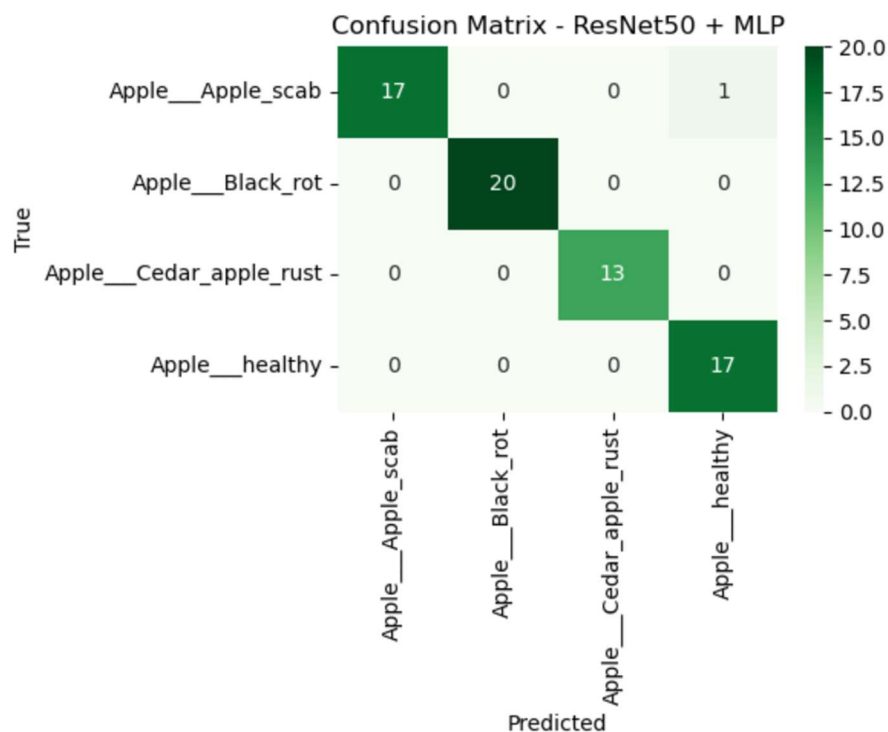


Figure 9 ResNet50 confusion matrix

The confusion matrix of the ResNet50 + MLP model demonstrates very good classification performance, with almost all predictions being on the diagonal and only one or two very minor errors appearing across classes. This is in line with the reported overall accuracy of 98.5%.

The few off-diagonal entries suggest that, on the one hand, ResNet50 is capable of extracting rich, high-level visual features. On the other hand, its deeper architecture

may be overly specialized in certain cases when trained on relatively small and homogeneous multimodal subsets.

The structured features contribute to the stability of the model; however, the fusion does not completely get rid of the minor confusions that exist between the categories that are visually similar.

Regardless, the matrix provides evidence that the fusion model has good generalization capabilities and can draw significant benefits from the combination of visual and structured information, albeit its slightly less stable behavior as compared to MobileNetV2.

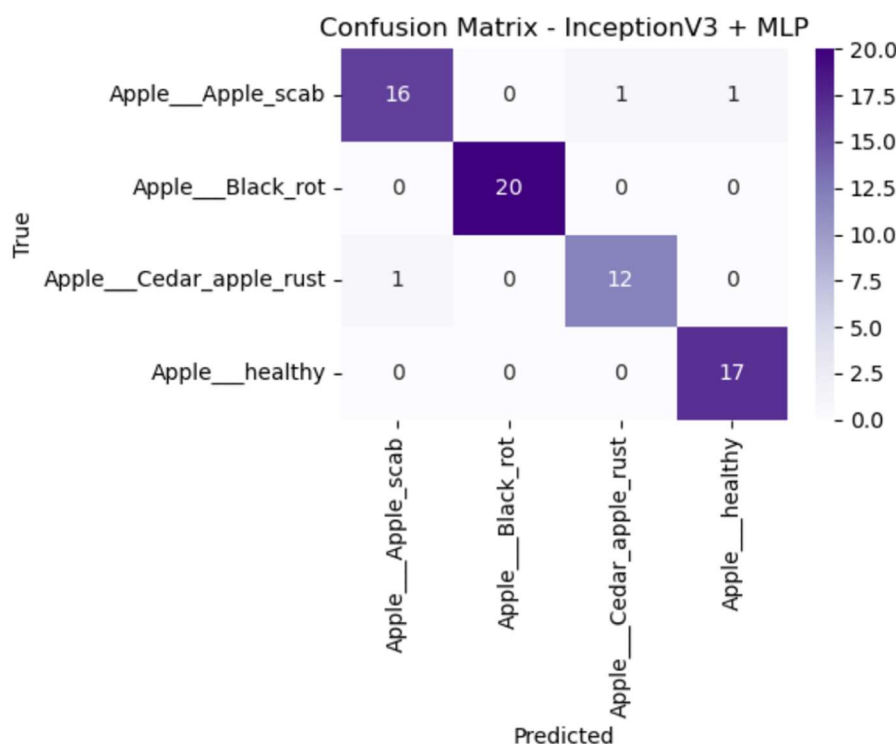


Figure 10 InceptionV3 confusion matrix

The confusion matrix of an ML model is a great way to see which kind of errors a model makes visually. In this particular instance, a confusion matrix for the InceptionV3 + MLP model, it can be seen that the work is quite good but less varied than the other models. Most of the examples are properly classified, especially in the Healthy and Cedar Apple Rust classes; however, there are quite a few instances of misunderstanding between Apple Scab and Black Rot that happen simultaneously.

These error patterns are those predicted since these two diseases are almost identical in the visual aspect, and InceptionV3 applies a complex multi-scale feature extraction method.

Though the structured climate attributes did have some effect on the final predictions, they were not enough to completely solve the problem of the fine-grained visual overlap of certain classes. The total accuracy of the model is 95.6%, which agrees with the confusion matrix and indicates that InceptionV3 profits from multimodal fusion, yet it still places more reliance on visual features than other architectures.

5.5 Comparison Summary

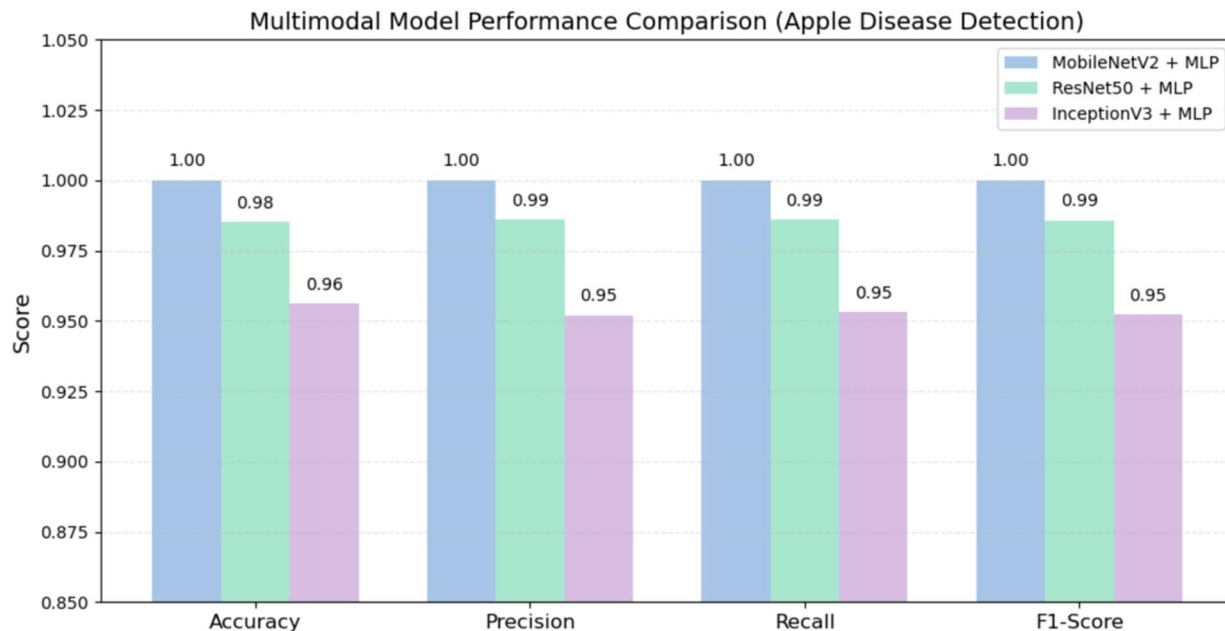


Figure 11 Model Performance Comparison (Apple Disease Detection)

The bar graph illustrates the difference in performance between the three multimodal fusion models in terms of accuracy, precision, recall, and F1-score. The MobileNetV2+MLP model is the strongest and most stable among all evaluated models, as it produces an accuracy of approximately 100%, including equally high precision, recall, and F1-score. This excellent result can be explained by the ability of MobileNetV2 to generate lightweight and smooth visual embeddings that are naturally compatible with structured environmental features. This kind of

compatibility means that there is no heavy architectural modification necessary for the fusion mechanism to operate effectively.

The second-best performer is the ResNet50 + MLP, which leads to an accuracy of 98.5%. The model is capable of utilizing the strong representational power of residual blocks; however, the deeper architecture of the model seems to require more diverse training samples for it to be fully generalized.

The point that ResNet50 is strong performance-wise and yet a bit behind MobileNetV2 implies that deeper CNNs do not always get as much benefit from multimodal integration when the dataset is small or has low variability.

The InceptionV3 + MLP model performed the task with 95.6% accuracy, thus determining it as the third best performer. InceptionV3 can extract features at various scales; however, the extreme visual similarity of some apple disease classes prevents the model from utilizing its intricate architecture. Besides, the structured features, though helpful, were not enough to provide sufficient diversity to completely eradicate the problem of misclassification of the visually similar diseases. Therefore, the multimodal boost was only moderate.

Basically, the comparison emphasizes the fact that CNN backbones do not receive the same amount of benefit from multimodal fusion.

Lightweight MobileNetV2 is capable of generating embeddings that fit well with the numerical climate features, which results in stable and very accurate multimodal models. On the other hand, the deeper models like ResNet50 and InceptionV3 might need to be supplied with more extensive or diverse multimodal datasets and possibly more advanced fusion strategies to be able to harness their full capacity.

Chapter 6: Conclusion and Future Work

6.1 Conclusion

This thesis presented a multimodal deep learning framework for the detection of apple leaf diseases by fusing visual features extracted from leaf images with structured environmental and soil data. Experiments showed that the effectiveness of multimodal fusion is closely tied to the choice of backbone architecture used for image feature extraction.

The MobileNetV2 + MLP model has reached an accuracy close to 100%, outperforming all other multimodal variants. The confirmation is that lightweight CNN architectures produce smooth and compact visual embeddings, which naturally align with numerical climate features, leading to stable and highly discriminative fusion results.

The ResNet50 + MLP model scored 98.5% thus showing that it had strong representational power; however, the performance of the model implies that deep architectures may need large and diverse multimodal datasets to be able to generalize well. On the other hand, InceptionV3 + MLP got 95.6% which indicates a small rise in performance, and it confirms that over complicated architectures are not always compatible with the structured climate features.

Findings indicate that, to help machines learn better and faster, the following should be considered:

- Combine computer visual recognition system embeddings with structured data, which is a kind of multimodal integration to enhance model performance. The extent of the improvement, however, greatly depends on the compatibility of CNN embeddings and structured data.
- Use Lightweight CNNs such as MobileNetV2, which, among others, are characterized by their stability, efficiency, and smooth embedding structure and, therefore, are very helpful in the field of multimodal agricultural applications.
- Deep architectures are allowed to have expensive datasets and sophisticated fusion methods if one would like to see the full fledged of their capabilities.
- Existence of the environmental attributes in the dataset makes them very distinctive even from the standpoint of structured only models that attain perfect accuracy, which in turn, provides evidence for this claim.

This thesis moves in the direction of constructing efficient, climate-aware disease recognition systems while it also discloses the difficulties and the potential of multimodal learning in the farming sector prevailing outside of lab scenarios.

6.2 Future Work

The specified recommendations below are the results of the experimental evaluation of image-only and multimodal models for apple leaf disease detection:

1. **Widen the data set:** Keep taking pictures of the infected leaves and note down the factors of the surroundings for more data points. To accurately determine the correlation between different modalities, it is very important that environmental factors (temperature, humidity, rainfall, pH, N, P, K) not only match the places but also the times of the leaf photos.
2. **Advanced Multimodal Fusion Techniques:** Maybe you could integrate a more advanced fusion technique, such as attention-based fusion, cross-modal transformers, and gated fusion networks. These methods could potentially be a lot more efficient in merging image and structured features and could also reduce the noise between different modalities.
3. **Enhance the quality of environmental features:** It is very necessary for the environmental features to be transformed into a time series with the weather data, rather than a simple snapshot of the times. In addition to leaf wetness duration, solar radiation and wind speed can also be considered as features for better disease risk factors.
4. **Lightweight CNN Architectures:** MobileNetV2 was able to get excellent results with a small dataset. There are several other lightweight architectures like MobileNetV3, EfficientNet B0/B1, or ShuffleNet, which can be considered not only for better multimodal integration but also for their feasibility in field deployment purposes in the future.
5. **Real World Testing:** This manner of experimenting or testing with models in real-life situations is very significant for them to be able to prove if their models are foolproof and can be generalized to other scenarios. Some issues they may face are changed lighting, background clutter, partial leaf occlusions, and different orchards.
6. **Modality Alignment:** One can think of methods that enable more efficient alignment between image and structured data features, like contrastive learning,

latent space matching, or modality normalization layers, etc., that can lead to further improvements of multimodal fusion results.

All these suggested future moves are to improve the precision and practicality of ground-level multimodal plant disease detection systems, which, directly and indirectly, can be a sustainable agriculture and better crop management support.

References

1. Ristaino JB, Anderson PK, Bebber DP, Brauman KA, Cunniffe NJ, Fedoroff NV, Finegold C, Garrett KA, Gilligan CA, Jones CM, Martin MD. The persistent threat of emerging plant disease pandemics to global food security. *Proceedings of the National Academy of Sciences*. 2021 Jun 8;118(23):e2022239118. <https://doi.org/10.1073/pnas.2022239118>
2. Awad DA, Masoud HA, Hamad A. Climate changes and food-borne pathogens: the impact on human health and mitigation strategy. *Climatic Change*. 2024 Jun;177(6):92. <https://doi.org/10.1007/s10584-024-03748-9>
3. Ahmed I, Yadav PK. Predicting apple plant diseases in orchards using machine learning and deep learning algorithms. *SN computer science*. 2024 Jul 3;5(6):700. <https://doi.org/10.1007/s42979-024-02959-2>
4. Upadhyay A, Chandel NS, Singh KP, Chakraborty SK, Nandede BM, Kumar M, Subeesh A, Upendar K, Salem A, Elbeltagi A. Deep learning and computer vision in plant disease detection: a comprehensive review of techniques, models, and trends in precision agriculture. *Artificial Intelligence Review*. 2025 Jan 17;58(3):92. <https://doi.org/10.1007/s10462-024-11100-x>
5. Sadr H, Nazari M, Khodaverdian Z, Farzan R, Yousefzadeh-Chabok S, Ashoobi MT, Hemmati H, Hendi A, Ashraf A, Pedram MM, Hasannejad-Bibalan M. Unveiling the potential of artificial intelligence in revolutionizing disease diagnosis and prediction: a comprehensive review of machine learning and deep learning approaches. *European Journal of Medical Research*. 2025 May 26;30(1):418. <https://doi.org/10.1186/s40001-025-02680-7>
6. Demilie WB. Plant disease detection and classification techniques: a comparative study of the performances. *Journal of Big Data*. 2024 Jan 2;11(1):5. <https://doi.org/10.1186/s40537-023-00863-9>
7. Ngugi HN, Ezugwu AE, Akinyelu AA, Abualigah L. Revolutionizing crop disease detection with computational deep learning: a comprehensive review. *Environmental Monitoring and Assessment*. 2024 Mar;196(3):302. <https://doi.org/10.1007/s10661-024-12454-z>
8. Staal J, Alsma J, Mamede S, Olson AP, Prins-van Gilst G, Geerlings SE, Plesac M, Sundberg MA, Frens MA, Schmidt HG, Van den Broek WW. The relationship between time to diagnose and diagnostic accuracy among internal medicine residents: a randomized experiment. *BMC medical education*. 2021 Apr 21;21(1):227. <https://doi.org/10.1186/s12909-021-02671-2>
9. Shafik W, Tufail A, De Silva Liyanage C, Apong RA. Using transfer learning-based plant disease classification and detection for sustainable agriculture. *BMC Plant Biology*. 2024 Feb 26;24(1):136. <https://doi.org/10.1186/s12870-024-04825-y>
10. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*. 2017 Apr 17. <https://doi.org/10.48550/arXiv.1704.04861>
11. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition 2016* (pp. 770-778). https://openaccess.thecvf.com/content_cvpr_2016/papers/He_Deep_Residual_Learning_CVP_R_2016_paper.pdf
12. Cao J, Yan M, Jia Y, Tian X, Zhang Z. Application of a modified Inception-v3 model in the dynasty-based classification of ancient murals. *EURASIP Journal on Advances in Signal Processing*. 2021 Jul 27;2021(1):49. <https://doi.org/10.1186/s13634-021-00740-8>

13. Kumar D, Mukhopadhyay R. Climate change and plant pathogens: Understanding dynamics, risks and mitigation strategies. *Plant Pathology*. 2025 Jan;74(1):59-68.
<https://doi.org/10.1111/ppa.14033>
14. Cao Y, Shen Z, Zhang N, Deng X, Thomashow LS, Lidbury I, Liu H, Li R, Shen Q, Kowalchuk GA. Phosphorus availability influences disease-suppressive soil microbiome through plant-microbe interactions. *Microbiome*. 2024 Sep 28;12(1):185.
<https://doi.org/10.1186/s40168-024-01906-w>
15. Romero F, Cazzato S, Walder F, Vogelgsang S, Bender SF, van der Heijden MG. Humidity and high temperature are important for predicting fungal disease outbreaks worldwide. *New Phytologist*. 2022 Jun 1;234(5):1553-6. <https://doi.org/10.1111/nph.17340>
16. Radwan M, Alhussan AA, Ibrahim A, Tawfeek SM. Potato leaf disease classification using optimized machine learning models and feature selection techniques. *Potato Research*. 2025 Jun;68(2):897-921. <https://doi.org/10.1007/s11540-024-09763-8>
17. Warner E, Lee J, Hsu W, Syeda-Mahmood T, Kahn Jr CE, Gevaert O, Rao A. Multimodal machine learning in image-based and clinical biomedicine: Survey and prospects. *International journal of computer vision*. 2024 Sep;132(9):3753-69.
<https://doi.org/10.1007/s11263-024-02032-8>
18. Sung YL, Li L, Lin K, Gan Z, Bansal M, Wang L. An empirical study of multimodal model merging. *arXiv preprint arXiv:2304.14933*. 2023 Apr 28.
<https://doi.org/10.48550/arXiv.2304.14933>
19. Zainab Z, Mahum R. Plant disease detection using deep learning techniques. *ICCK Journal of Image Analysis and Processing*. 2025 Mar 20;1(1):36-44.
<https://doi.org/10.62762/JIAP.2025.227089>
20. Shoaib M, Hussain T, Shah B, Ullah I, Shah SM, Ali F, Park SH. Deep learning-based segmentation and classification of leaf images for detection of tomato plant disease. *Frontiers in plant science*. 2022 Oct 7;13:1031748. <https://doi.org/10.3389/fpls.2022.1031748>
21. Ahmad A, El Gamal A, Saraswat D. Toward generalization of deep learning-based plant disease identification under controlled and field conditions. *IEEE Access*. 2023 Jan 26;11:9042-57. <https://doi.org/10.1109/ACCESS.2023.3240100>
22. Agarwal M, Singh A, Arjaria S, Sinha A, Gupta S. ToLeD: Tomato leaf disease detection using convolution neural network. *Procedia Computer Science*. 2020 Jan 1;167:293-301.
<https://doi.org/10.1016/j.procs.2020.03.225>
23. Chen J, Chen J, Zhang D, Sun Y, Nanekaran YA. Using deep transfer learning for image-based plant disease identification. *Computers and electronics in agriculture*. 2020 Jun 1;173:105393. <https://doi.org/10.1016/j.compag.2020.105393>
24. DeChant C, Wiesner-Hanks T, Chen S, Stewart EL, Yosinski J, Gore MA, Nelson RJ, Lipson H. Automated identification of northern leaf blight-infected maize plants from field imagery using deep learning. *Phytopathology*. 2017 Nov 17;107(11):1426-32.
<https://doi.org/10.1094/PHYTO-11-16-0417-R>
25. Sharma R, Singh A, Jhanjhi NZ, Masud M, Jaha ES, Verma S. Plant Disease Diagnosis and Image Classification Using Deep Learning. *Computers, Materials & Continua*. 2022 May 1;71(2). <https://doi.org/10.32604/cmc.2022.020017>
26. Wang G, Sun Y, Wang J. Automatic image-based plant disease severity estimation using deep learning. *Computational intelligence and neuroscience*. 2017;2017(1):2917536.
<https://doi.org/10.1155/2017/2917536>

27. Too EC, Yujian L, Njuki S, Yingchun L. A comparative study of fine-tuning deep learning models for plant disease identification. *Computers and Electronics in Agriculture*. 2019 Jun 1;161:272-9. <https://doi.org/10.1016/j.compag.2018.03.032>
28. Eunice J, Popescu DE, Chowdary MK, Hemanth J. Deep learning-based leaf disease detection in crops using images for agricultural applications. *Agronomy*. 2022 Oct 3;12(10):2395. <https://doi.org/10.3390/agronomy12102395>
29. Atila Ü, Uçar M, Akyol K, Uçar E. Plant leaf disease classification using EfficientNet deep learning model. *Ecological Informatics*. 2021 Mar 1;61:101182. <https://doi.org/10.1016/j.ecoinf.2020.101182>
30. Jiang P, Chen Y, Liu B, He D, Liang C. Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks. *Ieee Access*. 2019 May 6;7:59069-80. <https://doi.org/10.1109/ACCESS.2019.2914929>
31. Amara J, Bouaziz B, Algergawy A. A deep learning-based approach for banana leaf diseases classification. In *Datenbanksysteme für Business, Technologie und Web (BTW 2017)-Workshopband 2017* (pp. 79-88). Gesellschaft für Informatik eV. <https://dl.gi.de/items/13766147-8092-4f0a-b4e1-8a11a9046bdf>
32. Ashwinkumar S, Rajagopal S, Manimaran V, Jegajothi BJ. Automated plant leaf disease detection and classification using optimal MobileNet based convolutional neural networks. *Materials Today: Proceedings*. 2022 Jan 1;51:480-7. <https://doi.org/10.1016/j.matpr.2021.05.584>
33. Bhattacharya A. A novel deep learning-based model for classification of rice leaf diseases. In *2021 Swedish Workshop on Data Science (SweDS) 2021 Dec 2* (pp. 1-6). IEEE. <https://doi.org/10.1109/SweDS53855.2021.9638278>
34. Divakar S. Multimodal Plant Disease Dataset. Kaggle; <https://www.kaggle.com/datasets/shubhamdivakar/multimodal-plant-disease-dataset-by-subham-divakar>
35. Goodfellow I, Bengio Y, Courville A, Bengio Y. *Deep learning*. Cambridge: MIT press; 2016 Nov 18. <https://doi.org/10.4258/hir.2016.22.4.351>
36. Breiman L. Random forests. *Machine learning*. 2001 Oct;45(1):5-32. <https://doi.org/10.1023/A:1010933404324>
37. Breiman L, Friedman J, Olshen RA, Stone CJ. *Classification and regression trees*. Chapman and Hall/CRC; 2017 Oct 19. <https://doi.org/10.1201/9781315139470>