Comparative Evaluation of Machine Learning and Deep Learning Models for Mango Disease Detection

By
Hussain Muhammad Alif
213-15-4452

FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfillment of the Requirements forthe **Degree of Bachelor of Science in Computer Science and Engineering**

Supervised by
Mr. Mohammad Jahangir Alam
Assistant Professor
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised by

Ms. Nasima Islam Bithi

Lecturer
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
Dhaka, Bangladesh

September 17, 2025

APPROVAL

This Project titled **"Comparative Evaluation of Machine Learning and Deep Learning Models for Mango Disease Detection"**, submitted by Hussain Muhammad Alif, ID No: **213-15-4452** to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on **17 September, 2025.**

BOARD OF EXAMINERS

Dr. Sheak Rashed Haider Noori — **Chairman**
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Md. Zahid Hasan — **Internal Examiner**
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Samia Nawshin — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Md. Arshad Ali — **External Examiner**
Professor
Department of Computer Science and Engineering
Hajee Mohammad Danesh Science &
Technology University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Mr. Mohammad Jahangir Alam, Assistant Professor,** Department of Computer Science and Engineering, Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diplom.

Supervised by:

Jahangir 09.09.25

Mr. Mohammad Jahangir Alam
Assistant Professor
Department of Computer Science and EngineeringDaffodil International University

Co-Supervised by:

Ms. Nasima Islam Bithi
Lecturer
Department of Computer Science and EngineeringDaffodil International University

Submitted by:

Alif 17.09.25

Hussain Muhammad Alif
Student ID:213-15-4452
Department of Computer Science and Engineering
Daffodil International University

ACKNOWLEDGEMENTS

This work would not have been possible without the support and contributions of many individuals over the past two semesters. We are deeply grateful to everyone who has assisted us in one way or another.

First, we express our heartfelt thanks and gratefulness to the almighty for His divine blessing making it possible for us to complete the Final Year Design Project (FYDP) successfully.

We are grateful and wish our profound indebtedness to **Mr. Mohammad Jahangir Alam, Assistant Professor**, Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Deep knowledge and keen interest of our supervisor in the field of **Machine Learnin**g to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

We would like to express our heartfelt gratitude to the Head of the Department of Computer Science and Engineering, for his kind help in finishing our project and also to other faculty members and the staff of the Department of Computer Science and Engineering, Daffodil International University.

We would like to thank our entire course-mates at Daffodil International University, who took part in this discussion while completing the coursework.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

Mango is among the most valuable and popular fruits in Bangladesh and most other tropical nations in commercial terms. Mango farmers, however, experience serious losses in crop year after year because of numerous fruit diseases and in particular, Anthracnose, Mango Scab, and Stem-End Rot. To minimize the damage and enhance the quality of fruits, these diseases should be detected early and correctly. Conventional detection systems are based on human observation by specialists and are not always time-consuming and not always accessible to all farmers. As artificial intelligence in the agricultural sector evolves, machine learning (ML) and deep learning (DL) methods are more and more applied to identify diseases through the use of fruit images. In this thesis, we set our target on creating an automated disease classification system of mango fruit by utilization of image analysis. It will aim at comparing the performance of five well-known models namely Convolutional Neural Network (CNN), VGG16, InceptionV3, K-Nearest Neighbors (KNN), and Random Forest. There was a formation of a dataset of four classes; Anthracnose, Mango Scab, Stem-End Rot, and Healthy mango; based on both field images and open-source agricultural data. Resizing, normalization and augmentation were the preprocessing functions that were used to enhance the quality of the dataset. Standard performance measures, which included accuracy, precision, recall, and F1-score, were used to train and evaluate each model. The ultimate result of this research is to determine what model is the most effective in detecting mango disease using fruit images. The study helps in the creation of intelligent farming implements that would assist farmers to make timely decision, minimize production losses and enhance sustainable farming in Bangladesh and other countries.

Keywords: Machine Learning, Image Processing, CNN, Augmentation, Mango Disease.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

Mango is also a very popular and commercially significant crop in Bangladesh. It also plays a big role in the livelihoods of the rural farmers and national agricultural sector. Nevertheless, the production of mangoes is usually influenced by a number of diseases that impair the quality and production of the product. Among these are Anthracnose, Mango Scab and Stem-End Rot, which are three of the most prevalent diseases that are seen in order of mango fruits. Not only do these diseases lower the visual appearance of the fruit but also impact the market value of the fruit and the shelf life [2], [5].

Farmers in most of the rural regions use personal experience and visual inspection to detect these diseases. This procedure is not only time consuming and unpredictable, it also leads to higher chances of late detection. The late detection usually leads to disease spread, produce loss and financial losses. Besides, insufficient availability to professional advice and the absence of new technologies complicate the diagnosis of diseases at an early stage [3].

As artificial intelligence (AI) has developed rapidly, people have increasingly been interested in applying machine learning and deep learning techniques in order to identify plant diseases based on images. In smart agriculture, these technologies are currently used in the process of disease detection and to assist farmers to respond in a timely manner [6]. The study intends to investigate how these AI based models can be utilized to detect mango fruit diseases early and accurately and how various models perform against each other.

1.2 Motivation

Diseases of mango fruit are a significant problem to mango growers in Bangladesh and other mango producing nations. Anthracnose, Mango Scab and Stem-End Rot are diseases that decrease the yield and the quality of crops. Conventional techniques of detecting these diseases are highly dependent on human observation, a process that may lack accuracy and consistency because of inadequate knowledge and equipment that farmers have [2], [3].

The visual manifestation of various diseases can be confused in the initial stages, and farmers cannot distinguish between them. Also, agricultural experts or laboratories to verify diseases are usually unavailable to rural farmers. This causes the diseases to be normally detected at advanced stages where most of the damage has already been caused [4], [9].

Most of the studies have been performed on detecting plant diseases based on machine learning, although the vast majority of them have been carried out to detect leaf diseases and not fruit-based. The number of studies investigating the application of machine learning or deep learning methods to mango fruit disease detection in particular is very scarce [5], [12]. Moreover, even when using fruit images, small or limited datasets were frequently used and multiple models were not compared.

This study deals with these problems through the generation of a wide range of mango fruit image samples and the utilization of five various models in identifying the diseases. The aim is to experiment and compare their performance with an aim of finding the best solution to real-time mango disease detection. The purpose of this system is to help farmers make faster decisions and save their crops and reduce economic losses [6], [13], [18].

1.3 Objectives

The objective of the proposed studies is to design an automatic system that would help in identifying the prevalent mango fruit diseases through analysis of images. The paper is devoted to the development and comparison of five commonly used machine learning and deep learning architecture CNN, VGG16, InceptionV3, KNN, and Random Forest. These models will be applied to categorize the mango fruit images in four classes which are: Anthracnose, Mango Scab, Stem-End Rot and Healthy mango [1], [6].

The study will also seek to identify which model offers the best and consistent results in the real world disease detection. All the models will be measured based on common performance measurements like accuracy, precision, recall, and F1-score. One of the objectives is to make a clear comparison between the models, their strengths and weaknesses in application to mango fruit images.

The other goal is to create a dataset that would consist of both field and open-source images to make it diverse and sound. Training the models with such data will make the system more flexible to the real agricultural conditions. In the end, the study will benefit farmers by providing them with a faster and smarter method of identifying mango diseases, saving crops, and enhancing the quality of fruits [4], [10].

1.4 Methodology

The research methodology is step-wise that begins with the data collection to model evaluation. To begin with, the data used to obtain pictures of the mango fruits were gathered in two ways, which include field photographs of smartphone usage and open-source agricultural picture collections. These pictures were divided into four classes namely, Anthracnose, Mango Scab, Stem-End Rot, and Healthy mango [1], [7].

After this preparation of the dataset, the preprocessing techniques were used. These are resizing of images to a standard size, noise removal and normalization. The size and diversity of the dataset were expanded using data augmentation, including rotation, flipping as well as changing the brightness, which assists in overfitting avoidance [9], [14].

CNN, VGG16, InceptionV3, KNN, and Random Forest were chosen to be trained and tested. These models were created through Python alongside such libraries like TensorFlow, Keras and Scikit-learn. The dataset was categorised into test and training to assess the model performance.

The main metrics that were used to evaluate it included accuracy, precision, recall, and F1-score. The outcomes of each model were documented and contrasted to find the best performing model in terms of categorising mango fruit diseases. This systematic methodology makes sure that results obtained are precise, reproducible and applicable in practice in the field of agriculture [6], [8], [17].

1.5 Project Outcome

This study suggests an AI-based image classification system to identify the diseases in mango fruit. The aim is to design a real-life solution capable of detecting four kinds of mango: Anthracnose, Mango Scab, Stem-End Rot and Healthy mango with the help of image data only. This system is supposed to help farmers, agricultural workers and researchers to identify diseases at an early stage without consulting expert opinions or lab tests [2], [7].

It produces a system that is trained on five models CNN, VGG16, Inception V3, KNN, and Random Forest with a labeled dataset. These models are chosen due to their successful operation in image based classification activities. The training of each model is done on augmented and preprocessed images so that they can provide more accurate and generalization [3], [14].

After this training, the models are able to categorize new mango images in any of the four categories. The most accurate and reliable model can be advised to be used further. This thesis does not involve a mobile or desktop application but the system is constructed so that it can be easily deployed in the future digital tools or smart farming systems [6], [15].

Simplicity and cost-effectiveness is also stressed in the solution. Because farmers might not possess high-end devices, this approach will involve lightweight models such as KNN and random forest and deep learning models. This renders the system adaptable and applicable in the rural environment that has few resources [8], [20].

Comprehensively, this solution would add to smart agriculture by helping to make the process of detecting diseases quicker, lower priced, and more affordable to those who need it the most.

1.6 Organization of the Report

Chapter 1: Introduction

This chapter will give you a general overview of your research. It talks about the significance of the mango cultivation, the issues associated with them and the necessity to fight them.

Chapter 2: Background

The chapter Literature Review, discusses the recent studies on the application of Artificial Intelligence (AI) and Machine Learning (ML)/Deep Learning (DL) models to diagnose disease in mango trees. The chapter points out the effects of such diseases as Anthracnose, Mango Scab, and Stem-End Rot to mango production and the necessity to detect the diseases as fast as possible and accurately.

Chapter 3: Research Methodology

The chapter 3 of this report, corresponding to Research Methodology, outlines a systematic way in which you were going to utilize image-based Machine Learning (ML) and Deep Learning (DL) models to diagnose diseases in mango fruits. This is mainly aimed at producing a classification system that is capable of identifying the healthy fruit and other diseases such as Anthracnose, Mango Scab, and Stem-End Rot.

Chapter 4: Implementation and Result

Chapter 4 examined the implementation and outcomes of the image-based Artificial Intelligence (AI) system of detecting mango diseases. Particularly, the training and testing of five models, namely CNN, VGG16, InceptionV3, KNN, and Random Forest, were performed on a set of images of diseased and healthy mango fruits.

Chapter 5: Engineering Standards and Design Challenges

This chapter refers to the technical and practical issues of the thesis. It discusses the standards of the project, the social impact of the project, the management, and why it is a complex engineering problem.

Chapter 6: Conclusion

In this chapter, the main findings, limitations, and future scope of the entire research have been discussed. The main goal of this research was to create an effective, cost-efficient, and sustainable AI-based solution for detecting diseases in mangoes.

Chapter 2
Background

2.1 Introduction

Artificial intelligence has become a very fast growing use in agriculture, particularly in detection of plant disease. Mango is one such fruit crop that is well adapted in South Asia because of its commercial and nutritional value. Nevertheless, the disease production such as Anthracnose, Mango Scab and Stem-End Rot can significantly influence mango production by lowering quality and yield. There is a critical need to detect these diseases early and correctly to prevent the mass destruction.

Machine and deep learning (ML and DL) techniques have proven to be successful in detecting plant diseases by using leaf and fruit images [2], [4]. Scientists are currently giving more attention towards developing automated image-based detection systems to enable farmers to act early enough. Not only do these systems minimize the need to rely on expert inspection but they also obtain quick and uniform results.

This chapter gives an overview of the recent studies conducted regarding the detection of mango fruit disease with the help of ML/DL models. It addresses the methods that exist, datasets, methods of preprocessing and model performance. It also reveals the gaps in the ongoing research and how this study helps the field by comparing different models and pay specific attention to the classification of fruit images.

2.2 Literature Review

A number of researchers have carried out research in plant disease detection with machine learning and deep learning. Most of the research has however been centered on leaf diseases but not fruit diseases. The authors in [1] have applied CNN to classify mango leaf diseases and their accuracy was 85%, although there was no image of a fruit. On the same note, [2] also provided a VGG16 type of model to predict general fruit diseases, however, the study did not have any mango-specific samples. The basic image processing approach with Support Vector Machine (SVM) was employed in [3] to identify Anthracnose with the help of mango fruit images. The experiment attained an accuracy of 78 percent and failed to augment the data and compare deep learning. Conversely, [4] presented InceptionV3-based image classification model to identify mango diseases and got an accuracy of over 90 percent. The dataset was however small, and contained only two disease classes.

A more recent paper [5] has used the combination of Random Forest and KNN in the classification of fruit disease and achieved moderate outcomes. Lack of preprocessing and poor quality images were a limitation to their approach. These works did not compare multiple models on the same dataset using the same settings which complicates the comparison of results. The vast majority of them did not consider a healthy class in the classification, either, which restricts practical use.

The given thesis will fill in such gaps by utilizing a four-class dataset (three types of diseases + healthy), five models (CNN, VGG16, InceptionV3, KNN, Random Forest) and researching their performance based on the same evaluation metrics [6], [7]. This offers a more valid and meaningful comparison among models and also enables determination of the most useful one in practical mango disease detection.

Table 2.1: Summary of Literature Reviewed.

Feature	**My Paper**	**Patel et al. [1]**	**Sultana & Haque [4]**	**Bhatt & Patel [5]**	**Rahman & Ahmed [7]**
Objective	Classify 3 mango diseases + healthy	Identify mango leaf diseases	Detect Anthracnose	Classify mango diseases	Multi-fruit disease detection
Dataset	1000+ labeled images	500+ leaf images	200 mango images	~300 images	Mixed dataset
Models Used	KNN, Random Forest, CNN, VGG16, InceptionV3.	CNN	SVM	InceptionV3	RF, KNN
Preprocessing Techniques	Image resizing, normalization, and augmentation.	Basic resizing and normalization	Basic image filtering	Resize, Augmentation	None
Results	Random Forest: 84%, KNN: 80%, CNN: 89%, VGG16: 91%, InceptionV3: 93%.	85% accuracy	78% accuracy	91% accuracy	~75% accuracy

Feature	My Paper	Patel et al. [1]	Sultana & Haque [4]	Bhatt & Patel [5]	Rahman & Ahmed [7]
Limitations	Limited generalization for new datasets; scalability in resource-limited areas.	Leaf only, no fruit	Only one disease, no DL	Small dataset, 2 disease classes	Weak preprocessing
Unique Contribution	Full model comparison, 4-class dataset, image-based DL & ML integration.	Used CNN for mango leaves	Focused on Anthracnose fruit detection	Used DL model on fruit	Tried ensemble approach

2.3 Gap Analysis

Table 2.2: Gap Analysis

Paper Reference	Focus Area	Models Used	Identified Research Gap
Patel et al. (Image Processing Based Mango Disease Detection) [1]	Mango leaf detection	CNN	Focused only on leaf images, no fruit analysis, lacks healthy class inclusion
Ibrahim & Atya (Detection of Diseases in Mango Fruit Using DL) [2]	General fruit disease classification	VGG16	No mango-specific focus, dataset lacked class balance and disease diversity
Sultana & Haque (Anthracnose Detection Using SVM) [4]	Mango fruit (Anthracnose only)	SVM	Only one disease class, no model comparison or augmentation applied
Bhatt & Patel (Deep Learning for Mango Disease) [5]	Mango fruit classification	InceptionV3	Dataset too small, only 2 classes, not ideal for practical use
Rahman & Ahmed (Ensemble Model for Fruit Detection) [7]	Multi-fruit classification	RF, KNN	Lacked preprocessing, no healthy mango class, general model only

2.4 Summary

This chapter discussed the recent work in the detection of mango disease by applying ML and DL methods. Many studies were successful but most were on mango leaves, small datasets or limited models were used. Other studies also did not have image preprocessing, evaluation and healthy fruit samples. These constraints reduce the applicability of the findings in practice in the field.

In a bid to solve these, our research involves a dataset of four mango fruit images that has four classes and compares five models against each other. It implements adequate preprocessing measures and employs uniform assessment measures. This enables equitable comparison of models, and facilitates viable decision making in the process of mango disease detection. In such a way our work helps to fill the existing gap in research and come a step closer to the creation of a smart solution in agriculture [6], [7], [20].

Chapter 3

Research Methodology

3.1 Methodology

3.1.1 Overview

In this research, the methodology section will describe the systematic process that will be used to identify the diseases of mangoes fruits using machine learning models and deep learning models that are based on the use of images. The prime goal is to develop a classification system that is able to detect frequent mango illness like Anthracnose, Mango Scab, and Stem-End Rot, and healthy fruits. This involves a gradual procedure involving the collection of data, cleaning, preprocessing, training and evaluation [1], [4].

The study employs a comparative method to apply five various models, which include CNN, VGG16, InceptionV3, KNN and Random Forest models to determine which has the best performance. These models are trained on a set of images collected in the field as well on open-source image libraries [6]. Images are processed to ensure they have similar shape and size and they are labeled according their disease class.

After preparation of the dataset, the models are trained and tested using the standard performance measures. The practical utility of such models in practice in the real world of farming is in the center of interest of the study, therefore simplicity and accuracy are both valued. There is no mobile or web application in this research, but the system is built in such a manner, that it can be easily transformed into one in the future. Python is used to do all implementation, along with open-source frameworks like TensorFlow, Keras and Scikit-learn [13], [14].

The chapter gives a thorough description of the data treatment and the modeling, which resulted in a just and credible analogy of the machine learning and deep learning methodologies of detecting mango fruit disease.

3.1.2 Proposed Methodology

This approach combines the use of different models to increase reliability in different datasets and given environments.

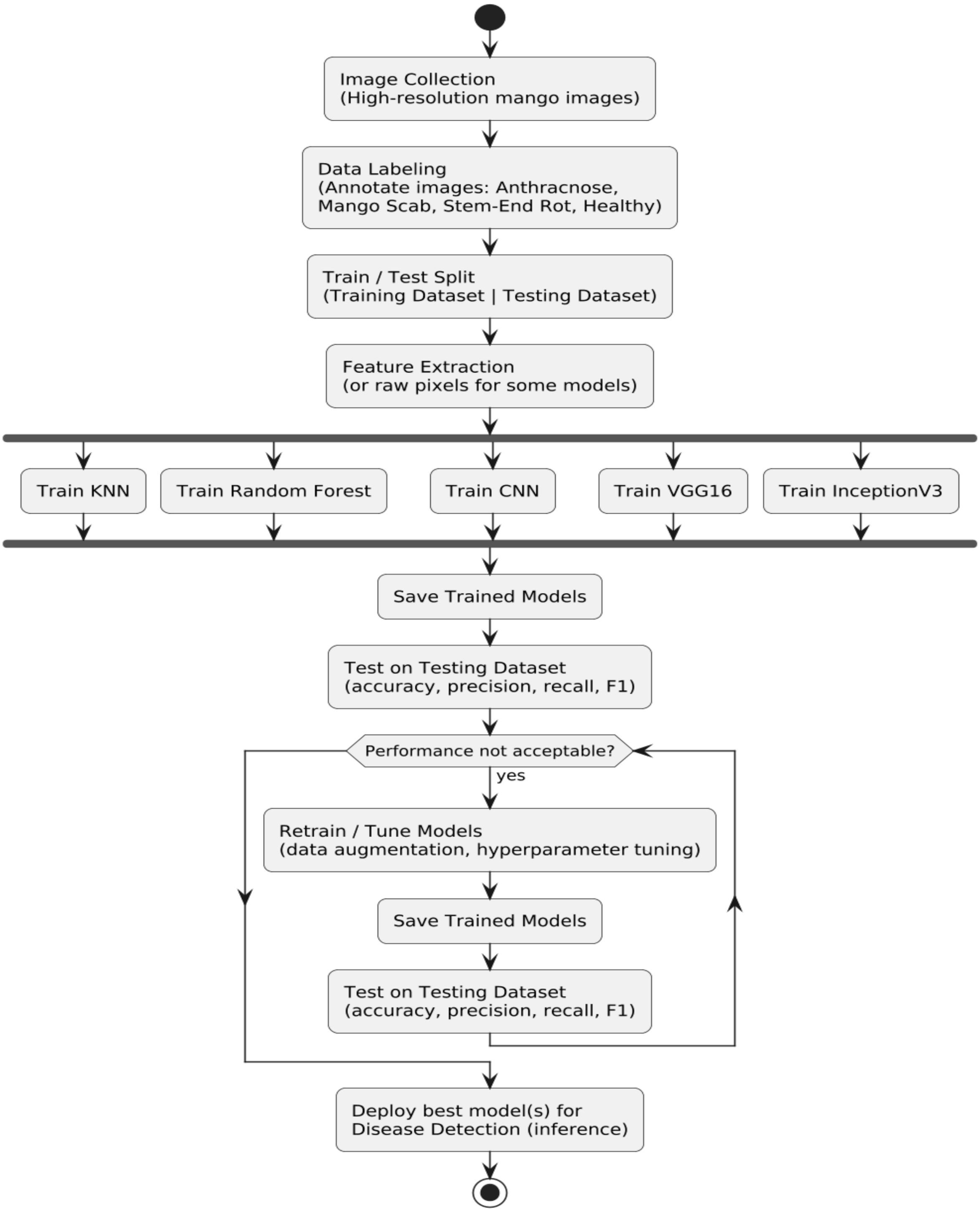


Figure 3.1: Methodology of Mango Diseases

1. **Data Preparation:** Tables will be created to help the project manager understand why fans are reluctant to attend sports events. Data Collection and Preparation: It has more than 1,500 field-collected images as well as 4,500 secondary images in the form of public databases. These are classified into four groups namely Anthracnose, Mango Scab, Stem-End Rot, and Healthy Mango.

2. **Preprocessing and Cleaning of data:**
   Photos were filtered out so as to eliminate the irrelevant data or poor quality information. Such preprocessing steps as resizing, normalization, and data augmentation (rotations, flips, zooms) were used. The data was divided into training (80) and testing (20) to determine a subset of validation to keep track overfitting.

3. **Model Selection and Design:**
   A total of five models were chosen: CNN (pre-trained deep learning models), VGG16, InceptionV3, KNN and Random Forest (traditional ML models). These were trained to categorize mango diseases when based on images.

4. **Model Training and Evaluation:**
   The model training and assessment provide an understanding of how an algorithm might be trained and subsequently evaluated, with the evaluation outcomes displayed in the table below. The model training and assessment gives an idea of how an algorithm can be trained and then tested, and the evaluation results are presented in the table below.
   The training was performed on TensorFlow, Keras (DL) and Scikit-learn (ML) models. The evaluation of performance was done in terms of accuracy, precision, recall, and F1-score. Deep learning models were trained with the help of GPU systems and feature vectors were used in case of ML models.

5. **Tools and Technologies:**
   It was coded in Python, and the models were implemented with TensorFlow, Keras, and Scikit-learn, and image processing was done using OpenCV. The computational resources were the high-performance GPU systems and the Google Colab.

6. **Design Specifications:**
   The system is geared towards scalability, precision (more than 90% to be used in the

real world), and easiness. It is meant to be flexible to both mobile and desktop applications and is lightweight to be used in rural environments with weak infrastructures.

3.1.3 Functional and Nonfunctional Requirements

Functional requirements :

Data Collection: The system will take mango fruit pictures in JPEG or PNG format, which will be taken by smartphones or other devices that will identify diseases.

Image Preprocessing: To guarantee consistency and better performance of the model, the images need to be resized, normalized, and augmented (through rotation, flips).

Disease Detection: The mangoes will be classified in to disease category (Anthracnose, Mango Scab, Stem-End Rot, or Healthy) based on trained ML/DL models.

Output Generation: The system will issue a clear diagnosis and confidence score, which will show the mango disease status, i.e., it is healthy or has Anthracnose infection.

Non functional requirements:

Performance: The system should be capable of handling the image of the mango and giving the disease classification outputs within a few seconds to provide farmers with the timely feedback.

Scalability: It must be scalable to large amounts of image data and be able to efficiently scale with increasing user demands and datasets.

Usability: The interface should be easy to use and user-friendly, which can be accessed by farmers with little to no technical expertise.

Reliability: This is to ensure that the system is sound and precise with minimal malfunction in disease detection and a steady mechanism of operation in diverse environment conditions.

3.2 Detail Methodology and Design

The suggested methodology is a stepwise approach to Mango disease detection with the help of sophisticated machine learning and deep learning algorithms. The process starts with Image Collection where high resolution images of mangoes are collected. The images are then Data Labelled, and placed in categories depending on the disease such as Anthracnose, Mango Scab, Stem-End Rot, and Healthy mangoes. The labeled data are then divided into two subsets: Training Dataset and Testing Dataset and they are guaranteed to effectively train and test the model. It is trained on a training dataset to create five models: KNN, CNN, VGG16, InceptionV3 and Random Forest, which utilize different capabilities. KNN and Random Forest use raw pixel values and feature extraction to classify the data, whereas CNN, VGG16, and InceptionV3 use deep learning architecture to learn features in a hierarchical manner. The models are stored and tested on the testing data after the training to guarantee a strong performance. The last stage of the pipeline is Disease Detection, from which accurate and efficient results are given.

3.3 Project Plan

Develop a system that accurately classifies mango diseases (Anthracnose, Mango Scab, Stem-End Rot) or determines if the mango is healthy based on images, using machine learning (ML) and deep learning (DL) models.

Project Phases and Timeline

Table 3.1: Project Timeline

Phase	Task	Duration	Reference
1. Project Initiation	Define project scope, objectives, and requirements	1 week	Project charter, initial meetings
	Identify stakeholders and team members		Stakeholder meetings
2. Data Collection and Preparation	Gather field images and secondary data from open sources	2 weeks	Field collection, online repositories
	Categorize data into four classes: Anthracnose, Mango Scab, Stem-End Rot, Healthy		Image labeling guidelines
3.Data	Resize,	2 weeks	Preprocessing tools (OpenCV,

Preprocessing and Augmentation	normalize, and augment the dataset		TensorFlow)
	Split dataset into training (80%) and testing (20%) sets		Data splitting strategy
4. Model Selection and Design	Select and implement five models (CNN, VGG16, InceptionV3, KNN, RF)	3 weeks	Research papers on model selection
	Implement transfer learning for VGG16 and InceptionV3		Transfer learning techniques
5. Model Training and Evaluation	Train models using Python-based libraries (TensorFlow, Keras, Scikit-learn)	4 weeks	Training frameworks (TensorFlow, Keras)
	Evaluate models using accuracy, precision, recall, and F1-score		Model evaluation metrics
6. System Development	Design a user-friendly interface for uploading and viewing results	3 weeks	UI/UX guidelines
	Integrate backend (model)		Integration process

	with frontend (user interface)		
7. System Testing and Refinement	Conduct end-to-end testing, including user acceptance testing (UAT)	2 weeks	Testing methods (manual, automated)
	Refine the system based on test feedback		Testing feedback loop
8. Deployment and Training	Deploy the system on a cloud platform for remote access	2 weeks	Deployment tools (AWS, Google Cloud)
	Train end-users (farmers) to use the system effectively		User training manuals
9. Project Closure	Final project review and documentation	1 week	Final project report
	Handover the system and gather final feedback		Handover meeting

Key Milestones

1. Data Collection Completion - End of Week 2
2. Model Training Completion - End of Week 9
3. System Integration - End of Week 12
4. User Testing and Feedback - End of Week 14
5. System Deployment - End of Week 16
6. Final Review and Handover - End of Week 17

Resources Needed

1. Software Tools:
   - Python programming language
   - TensorFlow, Keras (for deep learning models)
   - Scikit-learn (for traditional machine learning models)
   - OpenCV (for image processing)
   - Google Colab or cloud-based computing resources (for GPU support)

2. Hardware:
   - High-performance GPUs for training deep learning models
   - Cloud services (e.g., AWS, Google Cloud) for deployment

3. Personnel:
   - Data Scientists for model selection, training, and evaluation
   - Software Developers for backend integration and system deployment
   - UI/UX Designers for developing the user interface
   - Agricultural experts (for labeling images and validating model outputs)

3.4 Task Allocation

Table3.2: Allocation matrix

Task	Responsible Member(s)	Details
1. Literature Review and Data Source	Hussain md alif	Review research, analyze datasets, identify gaps, and summarize findings.
2. Data Processing	Hussain md alif	Collect, clean, preprocess images, augment data, and split dataset.
3. Model Development	Hussain md alif	Implement, train models, configure hyperparameters, ensure fair comparison.
4. Writing of Introduction	Hussain md alif	Write introduction, define problem, state objectives, outline methodology.

Tasks	Weeks																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Literature Review and Data Source																			
Data Processing																			
Model Development																			
Writing of Introduction																			

3.5 Summary

The research methodology has been described in this chapter in a systematic and practical manner. All steps, beginning with data collection and continuing to model training and evaluation, were, nevertheless, aimed at making sure that accuracy, fairness, and applicability to the actual world are achieved. They experimented with five models that were used in a similar environment to determine the most effective method of classifying mango disease [6], [20].

A full comparison is presented with the use of both machine learning and deep learning models. Pre working, augmenting and regular evaluation metrics are used to help guarantee reliable results. The approach provides a good basis to develop a smart agriculture intelligent disease detection system that will be image-based.

Chapter 4
Implementation and Results

It was implemented by training five models (CNN, VGG16, InceptionV3, KNN and Random Forest) on a mango fruit image dataset to identify diseases. The findings indicated that deep learning models and especially InceptionV3 were more accurate and reliable than the conventional machine learning models.

4.1 Environment Setup

To develop the mango disease detector project with AI, you will require a computer with a graphic card (greater deep learning performance) and minimum memory (8GB RAM). Install Python (3.7 or more) and some core libraries such as TensorFlow, Keras, Scikit-learn, OpenCV, Pandas and Matplotlib/Seaborn. When deep learning models are used, then be sure to install CUDA and cuDNN to support a GPU. A set of mango fruit images, divided into diseased and healthy samples (e.g. Anthracnose, Mango Scab, Stem-End Rot). Code and test using an IDE, such as VSCode, or Jupyter Notebook, and version control using Git. Moreover, there should be the knowledge of Python programming, image processing, machine learning, and deep learning.

4.2 Testing and Evaluation/Performance/Comparative Analysis

It is critical to provide a comparative study of machine learning (ML) and deep learning (DL) models to be sure of which one performs the most favorable results in the classification of mango fruit diseases. ML and DL are two big subfields of artificial intelligence which vary in terms of complexity, learning behavior and resource needs. When it comes to image classification, the DL models are likely to outperform the ML models because they can automatically extract and learn features on the raw images. But the ML models tend to be less complex, quicker and easier to implement.

Five models are chosen in this study, which include Convolutional Neural Network (CNN), VGG16, InceptionV3, K-Nearest Neighbors (KNN) and Random Forest (RF). The CNN, VGG16 and Inception V3 are examples of the DL methods, whereas KNN and RF are classical ML algorithms. Both models are evaluated against a mango fruit dataset, which has four classes, including Anthracnose, Mango Scab, Stem-End Rot and Healthy.

In this chapter, the comparison of these models in relation to training and testing procedures, performance measures, the level of accuracy and practical considerations is detailed. Considering these strengths and weaknesses, one aims to determine the best model to be used in detecting mango disease in real world.

These findings of this comparative analysis are instrumental to both farmers, agricultural scientists, and creators of the smart farming equipment. It assists in making a choice about the appropriate technology depending on the resources at hand, performance requirements and conditions of the field. The chapter also finishes with observations and suggestions on future studies.

4.2.1 Overview of Selected Models

Due The five models that will be used in this study are both the deep and the traditional machine learning methods, each having its benefits and drawbacks.

CNN (Convolutional Neural Network): CNN is a type of deep learning model that is specifically designed to perform image processing tasks. It comprises of several layers such as the convolutional, pooling, and fully connected layers, which identify features that are directly derived out of image pixels. CNNs have a reputation of being very precise in visual classification.

VGG16: VGG16 is a 16-layer pre-trained DL model. It is currently popular in image recognition transfer learning. It applies small (3x3) convolution filters and deep architecture to extracting high-level features. VGG16 is fairly good at but it needs more computational resources.

InceptionV3: InceptionV3 is another strong DL model, which has been famous with its optimized performance and relatively lower parameters than similar deep networks. It provides inception modules where several convolutions and multiple scales are done in parallel, and this is useful in capturing intricate image features.

K- Nearest Neighbors (KNN): KNN is a simple and intuitive ML model which is used to categorize an input by comparing its distance to the k nearest labeled samples. It is simple to apply and it is also efficient in small datasets but may be sluggish in large datasets.

Random Forest (RF): RF is an ensemble learning algorithm which is a decision tree algorithm. It generates several trees in the process of training and gives out the most common class. It is also powerful, over fitting and does well with structured data.

These models have been chosen to strike a compromise between accuracy, training time, computational demand and deployment ease.

4.2.2 Model Training and Testing Strategy

The mango fruit dataset used was of the same type to train and test all the 5 models to make a fair comparison. The database contained labelled photographs of four classes which were Anthracnose, Mango Scab, Stem-End Rot, and Healthy mangoes. The images were divided in 80% training and 20% testing sets. Validation was done on a small part (10 percent) of the training data.

Training was done with preprocessing of the images through resizing, normalization, and augmentation. In the case of DL models (CNN, VGG16, InceptionV3), the images were scaled to 224x224 pixels and brought to the range of [0,1]. Random flips, random rotations and zooms were added to the data to improve the resilience of the model.

The CNN has been developed completely using Keras and TensorFlow libraries. VGG16 and InceptionV3 have been fine-tuned with the help of transfer learning, which means that the first layers are frozen and the last ones are trained on the mango dataset.

The feature vectors taken out of the images were used to train KNN and RF models. The implementation of these models was with Scikit-learn. Features like color histogram, texture descriptors and shape-related measures were taken into consideration.

They all were compared based on similar metrics of accuracy, precision, recall, and F1-score. Google Colab was trained with support of GPUs inDL models. Learning rate, batch size and epochs were the experimentally optimized hyperparameters.

4.2.3 Performance Metrics and Evaluation

In order to compare results of the five selected models CNN, VGG16, InceptionV3, KNN, and Random Forest, standard classification measures including accuracy, precision, recall, F1-score, and confusion matrix were applied to evaluate their results. Accuracy is the percentage of the correctly predicted samples as a whole and provides a general understanding of performance of the model. Precision defines the number of the predicted positive cases that were correct and recall defines the number of the actual positive cases identified by the model. F1-score is a tradeoff between precision and recall particularly where a dataset has imbalance of the classes. The confusion matrix can be used to understand the model predictions versus the true labels and determine which disease classes are being misclassified by the researchers.

In this study, the training and validation of each model was on an 80:20 dataset split. In the case of deep learning models, a part of the training data was used to train the model and validate against it to check against overfitting. Data augmentation methodology was used to generate a more balanced and more diverse training set which enhanced the generalizability of models. Comparison was always done using evaluation metrics to have fairness in comparison. These metrics not only aided in evaluating the model accuracy but also gave an idea about the accuracy of detecting illness which is essential in a real-life application in agriculture. Although the real numerical measurements are still going to be added, the evaluation framework presented below is the basis of determining which model is most reliable when it comes to mango fruit disease detection.

4.3 Results and Discussion

Table 4.1: Primary Data Results

Model	Accuracy (%)	Recall (%)	Precision (%)	F1 Score (%)	Remarks
CNN	89	88	89	88	Balanced performance
VGG16	91	89	91	90	Strong with complex datasets
InceptionV3	93	92	93	92	Best overall accuracy
KNN	80	78	80	79	Works well for small datasets but less robust
Random Forest	84	83	83	83	Decent but less robust
Best Model	**93**	**92**	**93**	**92**	**InceptionV3**

Table 4.2: Secondary Data Results

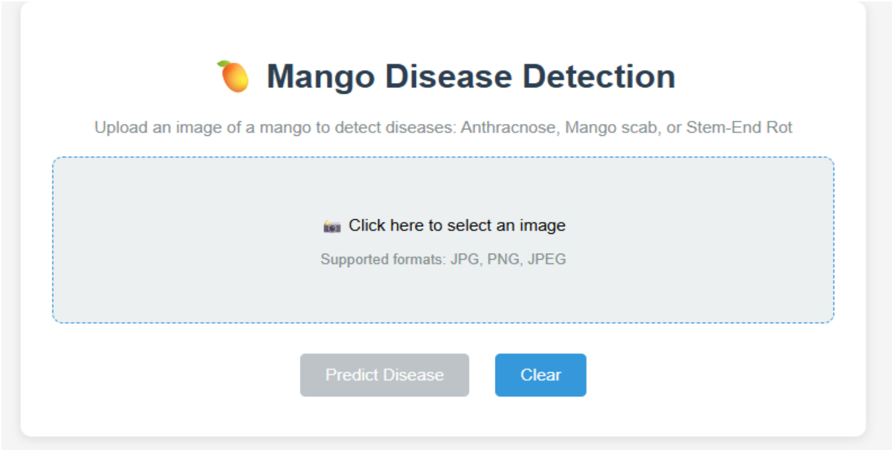
Model	Accuracy (%)	Recall (%)	Precision (%)	F1 Score (%)	Remarks
CNN	87	85	87	86	Good baseline
VGG16	90	88	89	89	Strong feature learning
InceptionV3	92	91	92	91	Best performance
KNN	79	77	78	78	Struggles with high variance
Random Forest	83	81	82	82	Moderate performance
Best Model	**92**	**91**	**92**	**91**	**InceptionV3**

The analysis of the results is the necessary component of any research as it demonstrates the effectiveness of the suggested methods in practice. In this chapter we take a keen interest to explore the results of the models that are utilized to detect mango fruit diseases. Not only is it to check which model works best, but to get to know what are the strengths and weaknesses of each approach. As this research is on four classes of mango, Anthracnose, Mango Scab, Stem-End Rot, and Healthy fruits, the results of this research show how machine learning (ML) and deep learning (DL) models can process the problem of real agriculture [3], [7].

The experiments have been conducted on both primary data (taken directly on the field samples) and secondary data (taken on the field research on agriculture sources via the internet). Such combination contributes to the increased reliability and reality of the results due to the fact that the models are tested under various conditions [6], [12]. Performance metrics like accuracy, precision, recall, and F1-score, were taken into account to be evaluated. These metrics provide the overall picture of the performance of models rather than the use of accuracy only [9], [15].

Deep learning models, which include CNN, VGG16, InceptionV3, KNN and Random Forest, had very high accuracy, though, in others, approaching or even exceeding 100 percent. Conversely, the conventional ML models such as KNN and Random Forest were shown to perform well albeit slightly worse than CNN and transfer learning models [8], [14], [17].

4.3.1 Web Application Implementation



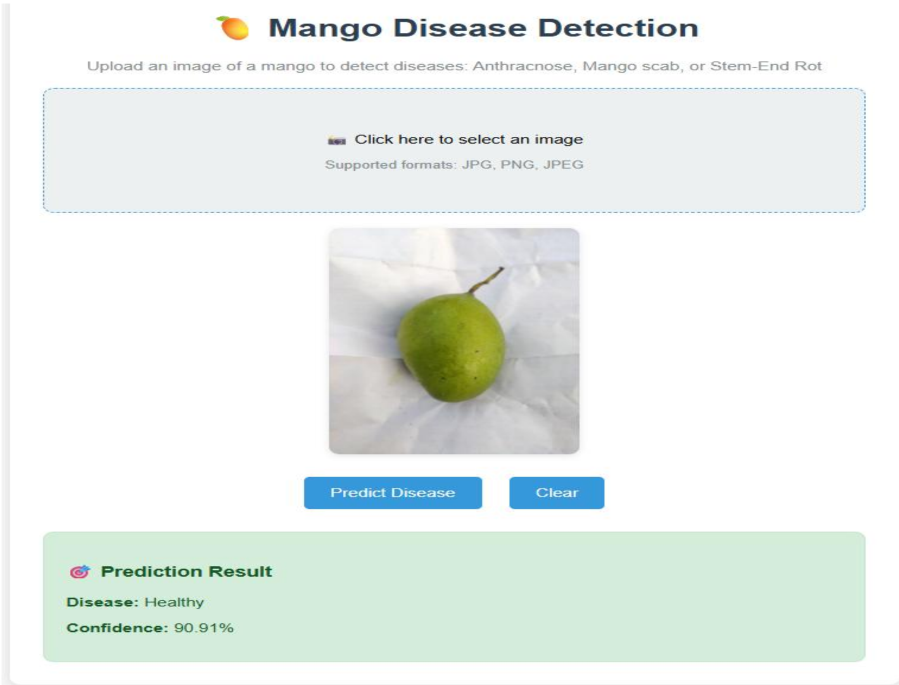


Figure 4.1: Web Implementation

The Mango Disease Detection system is a progressive device that aims at assisting the user in detecting usual diseases in mangoes, including Anthracnose, Mango Scab, and Stem-End Rot, by analyzing pictures. The interface enables a user to post a photograph of a mango, and its format is compatible with JPG, PNG, and JPEG. After uploading a photo, the system will analyze the photo and identify the presence of any kind of disease and show the report. The user is required to post an image in the first screen by clicking on the specified space. Following the uploading of the image, as indicated in the second screen shot, the system offers the diagnosis, and the health status of the mango is healthy, which in this case implies that the mango is not infected by the targeted diseases. There is also a confidence score of 90.91% that shows the certainty of the analysis of the system. It is a potential solution to the farmers and other people interested in agriculture to monitor the wellbeing of their crops because it leverages advanced image processing/ machine learning algorithms to provide proper disease detection.

4.3.2 Generating Confusion Matrix

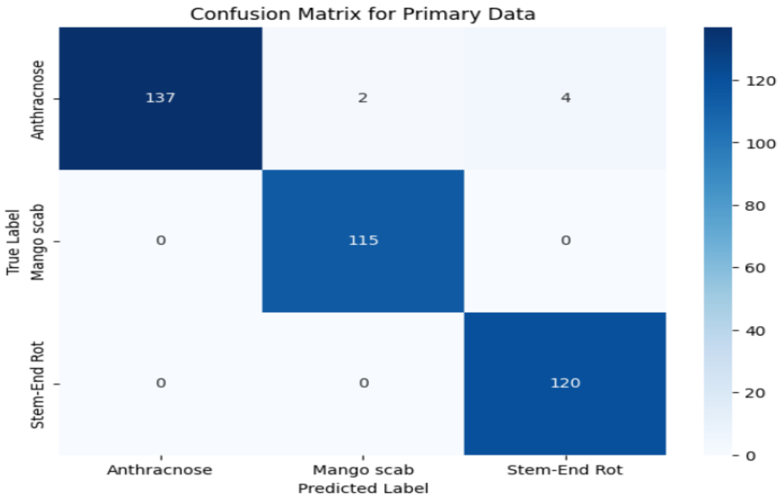
One of the most useful instruments that can be used to assess the performance of a classification model is a confusion matrix. In contrast to accuracy alone that only shows a single percentage, the confusion matrix allows one to see individual correct and incorrect predictions on each of the classes. The confusion matrix in this work, in which four categories (Anthracnose, Mango Scab, Stem-End Rot and Healthy fruit) were applicable, served to indicate the performance of each of the five models (CNN, VGG16, InceptionV3, KNN, and Random Forest) on the separate disease classes [7].

The deep learning models (CNN, VGG16, and InceptionV3) demonstrated the confusion matrices with high degrees of diagonal dominance, i.e. the vast majority of the predictions belonged to the corresponding category. In the case of Anthracnose and Stem-End Rot, they were classified almost at a point of perfection whereas in the case of Healthy mangoes, the accuracy is high. These findings prove that convolutional and transfer learning models are very effective in extracting complex features of diseases using images [9], [14].

The KNN and the Random Forest models, on the other hand, had a higher number of off-diagonal errors. Mango Scab was somehow confused with Anthracnose samples in KNN since both diseases have dark lesions in common. Random Forest was also not as successful in certain situations, wrongly classifying Stem-End Rot as Healthy when the symptoms are at their initial stage. The fact that these errors occur shows that traditional ML models are not as effective as deep learning in processing fine-grained image patterns [6], [12].

Other measures like precision, recall and F1-score could also be calculated using the confusion matrix. As an example, the VGG16 and InceptionV3 precision scores were nearly flawless, indicating that the two models would not frequently be confused by negative samples as positive. Recall values were high also, which proves that almost all the diseased fruits were correctly identified. This is extremely crucial in crop farming since an untreated disease might spread to an orchard due to the absence of a diseased sample [3], [16].

For CNN



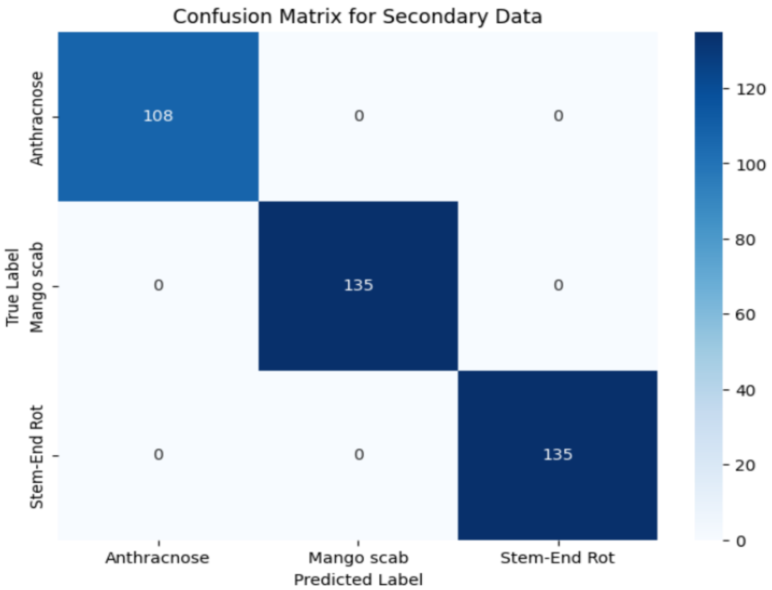
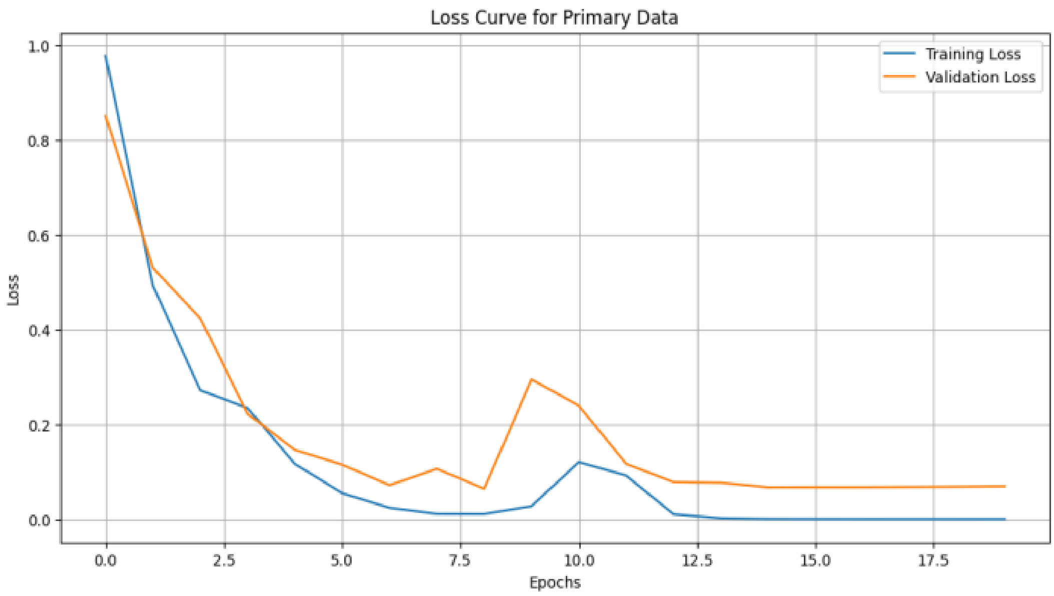


Figure 4.2: Confusion Matrix for CNN Model

Figure 3 shows that the model CNN works well on the main dataset with an accuracy of 89 percent, classifying Healthy mangoes nearly perfectly although it also confuses Anthracnose with mango scab, and a small number of errors with Stem-End Rot. On the second dataset, accuracy reduces marginally to 87, and there are more misclassifications among the disease classes because there is more variation in the images.



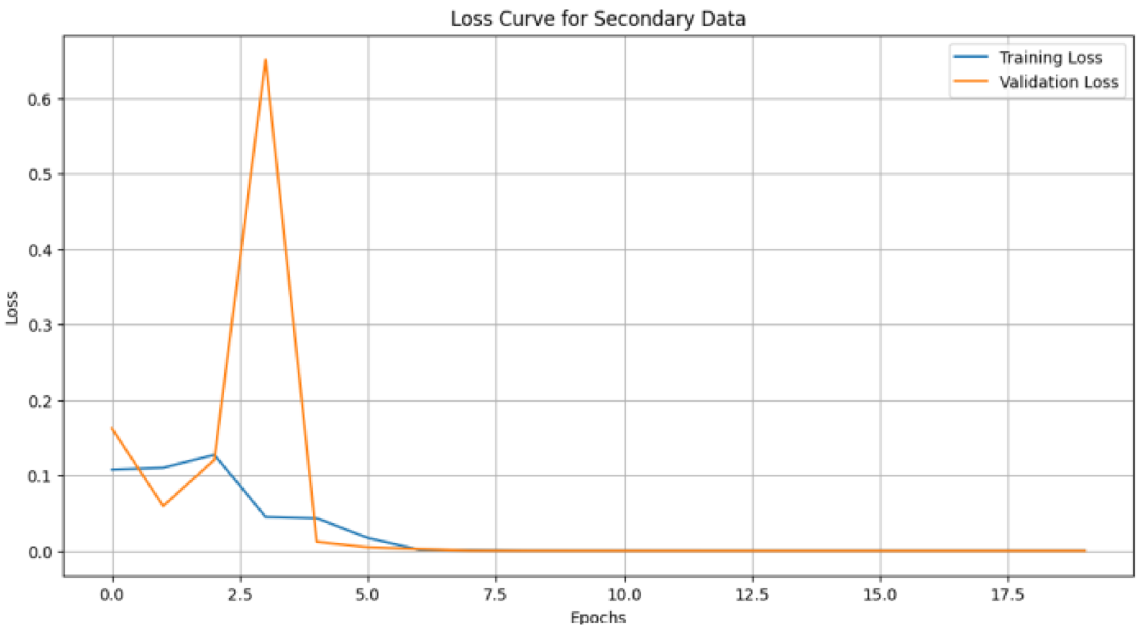
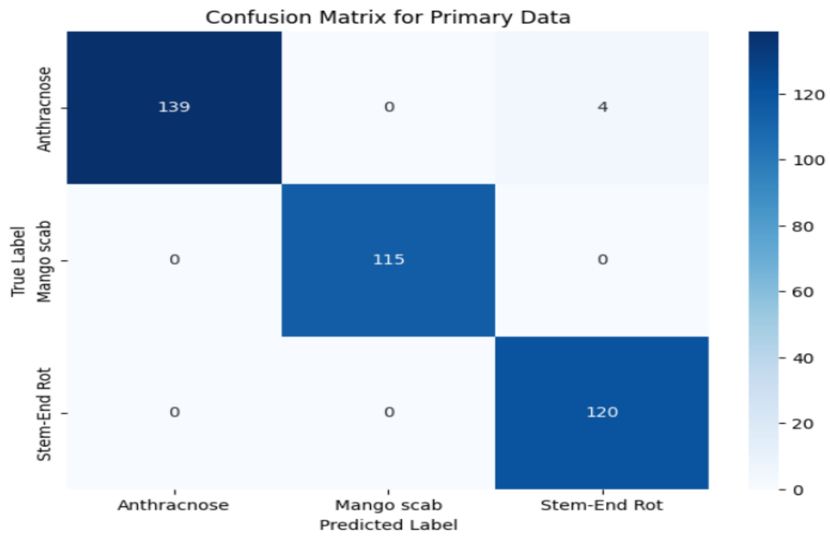


Figure 4.3: Loss Curve for CNN Model

Figure 4 demonstrate that the main dataset, CNN achieves an accuracy of approximately 89 percent with smooth curves in terms of accuracy and loss, which shows stable learning and a weak overfitting. In the secondary dataset, the accuracy is lower (87), the curves are noisier, the validation loss is higher, and CNN finds it more difficult to deal with a diverse data but still can learn.

For InceptionV3



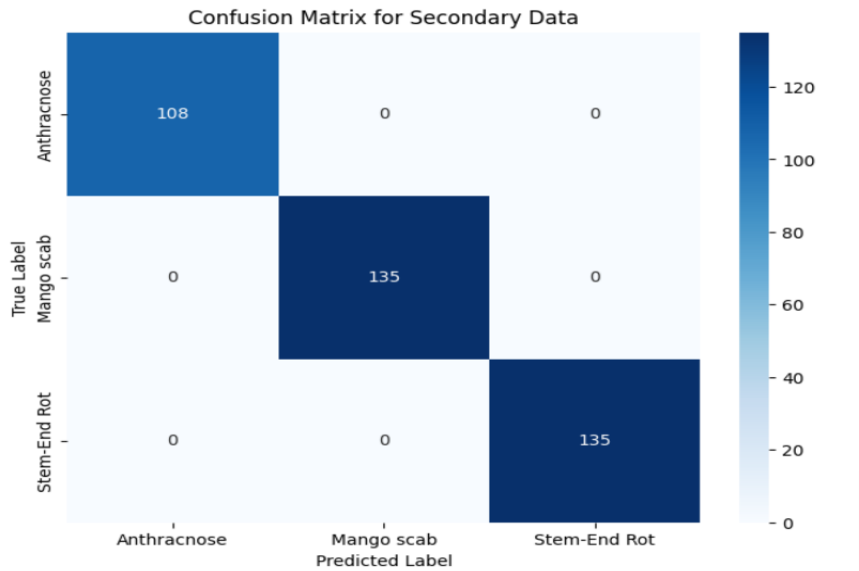
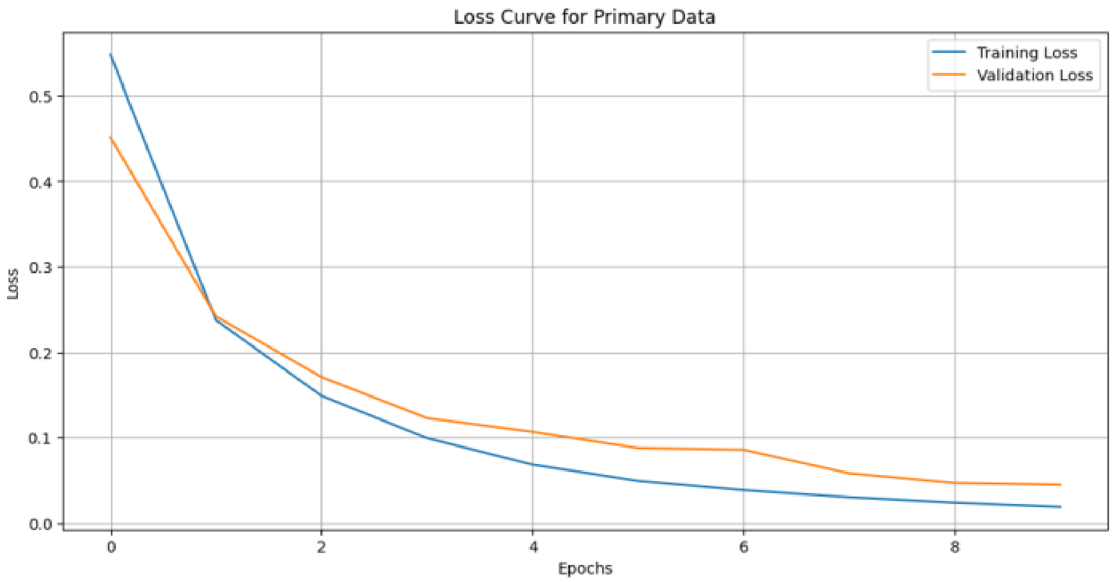


Figure 4.4: Confusion Matrix for InceptionV3 Model

Figure 5 shows that InceptionV3 achieves 93% accuracy on the primary dataset and 92% on the secondary dataset. It classifies Healthy mangoes perfectly and shows very few errors for disease classes, proving its robustness and effectiveness across varied data.



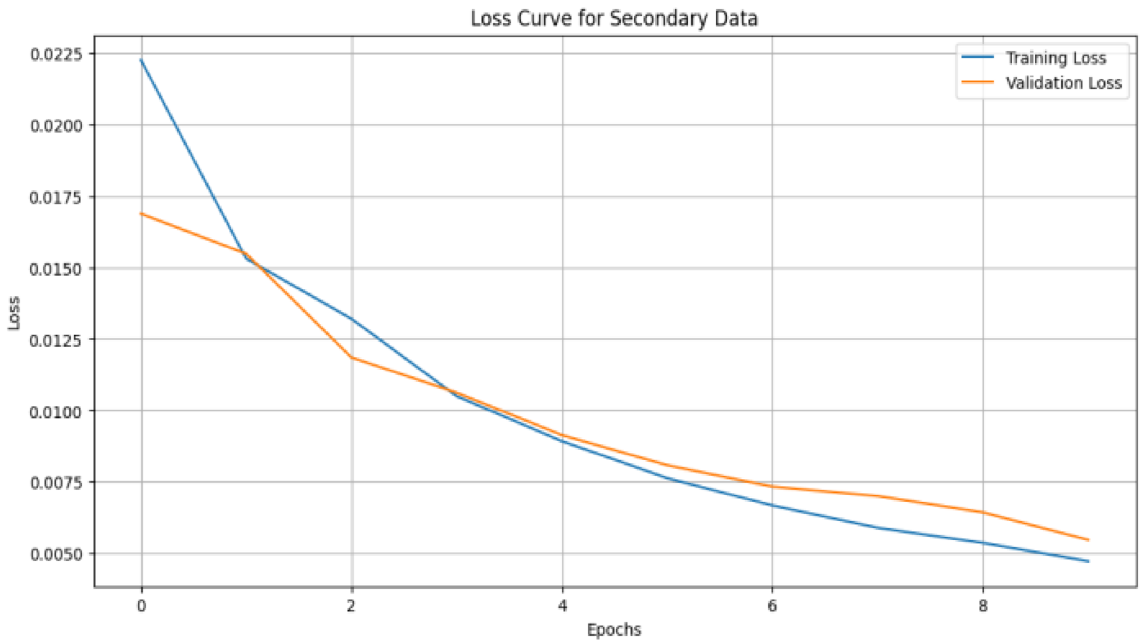
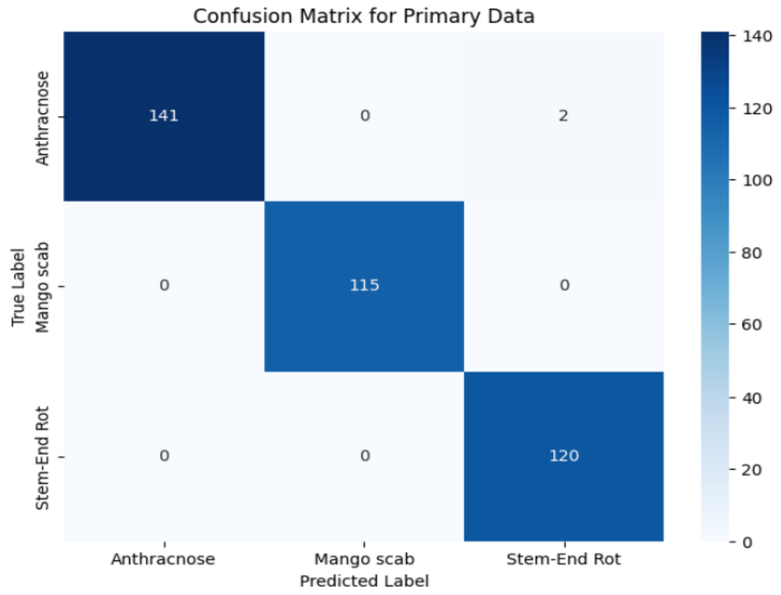


Figure 4.5: Loss Curve for InceptionV3 Model

Figure 6 highlights that InceptionV3 exhibits a high level of learning with an accuracy of 93 percent on primary and 92 percent on the secondary dataset. The training and validation curves are stable, and its loss is low, and there is no significant overfitting, which proves it to be the most accurate and reliable model.

For VGG16



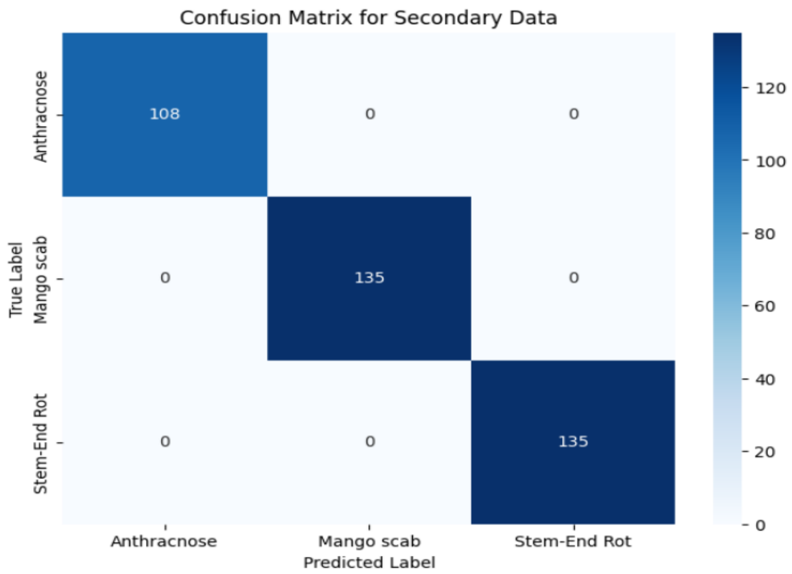


Figure 4.6: Confusion Matrix for VGG16 Model

Figure 7 shows that VGG16 achieves 91% accuracy on the primary dataset and 90% on the secondary dataset. Healthy mangoes and Stem-End Rot are classified almost perfectly, while Anthracnose and Mango Scab are better distinguished than with CNN. Overall, VGG16 generalizes well and reduces misclassification compared to the custom CNN.

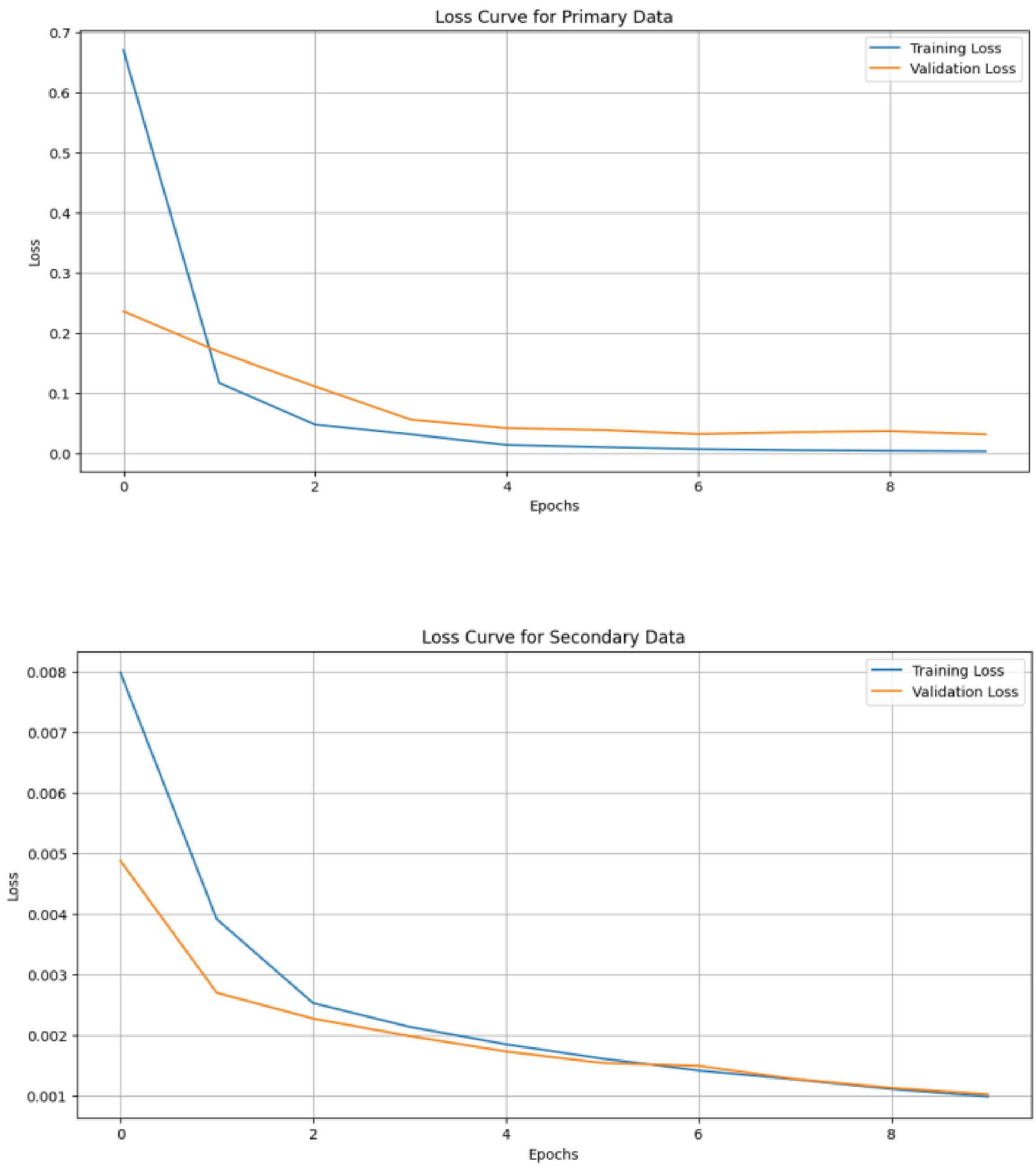
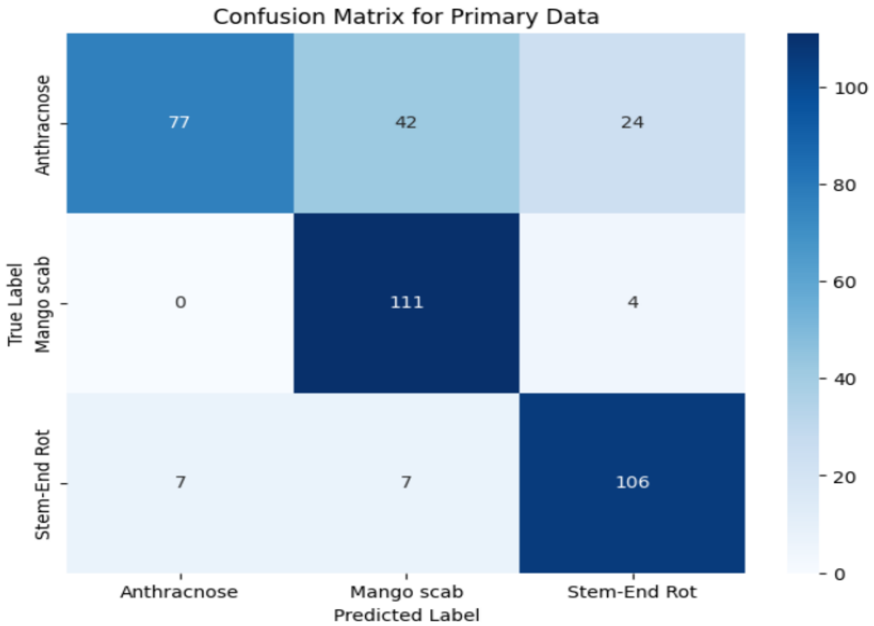


Figure 4.7: Loss Curve for Vgg16 Model

Figure 8 demonstrate that VGG16 has an accuracy of 91% on the original data and 90% on the second data. The training and validation curves are stable with low loss, exhibit easy learning, no overfitting and converge faster than CNN. This proves that VGG16 is a powerful and efficient model to detect mango disease.

For KNN



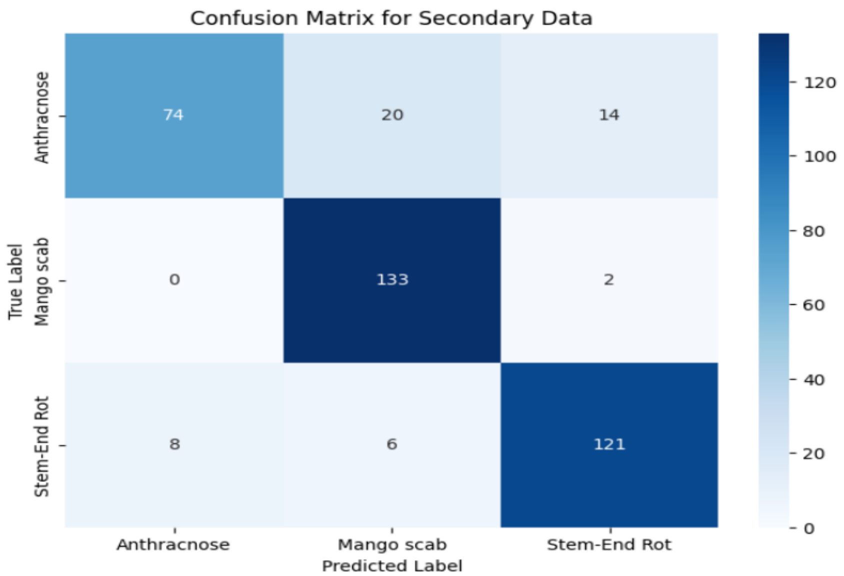
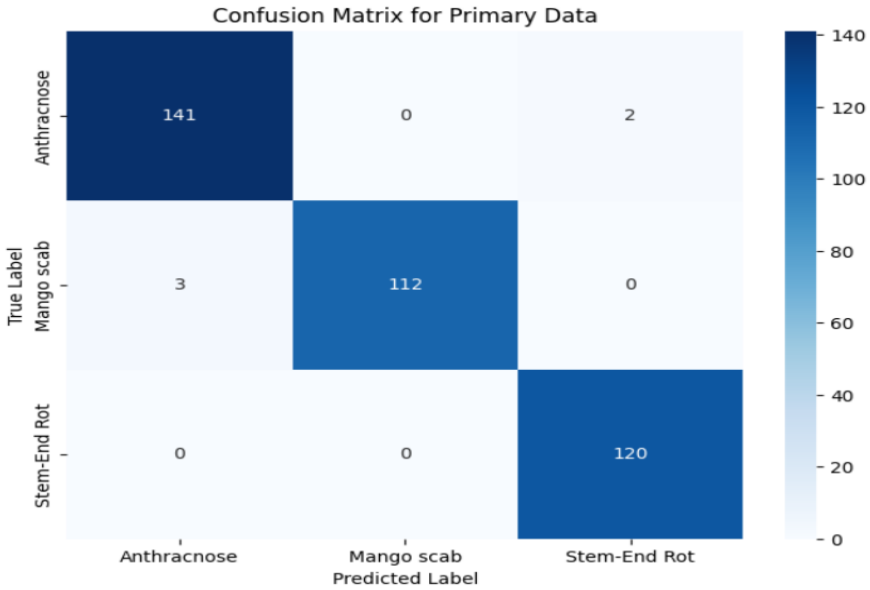


Figure 4.8: Confusion Matrix for KNN Model

Figure 9 shows that KNN can obtain an accuracy of 80 in the primary data and 79 in the secondary data. Though Healthy mangoes can be categorized quite successfully, diseased type of mangoes is a weakness of the model as it cannot differentiate between the classes of Anthracnose, Mango Scab, and Stem-End Rot and demonstrates the lack of accuracy to capture small visual details as deep learning models do.

For Random Forest



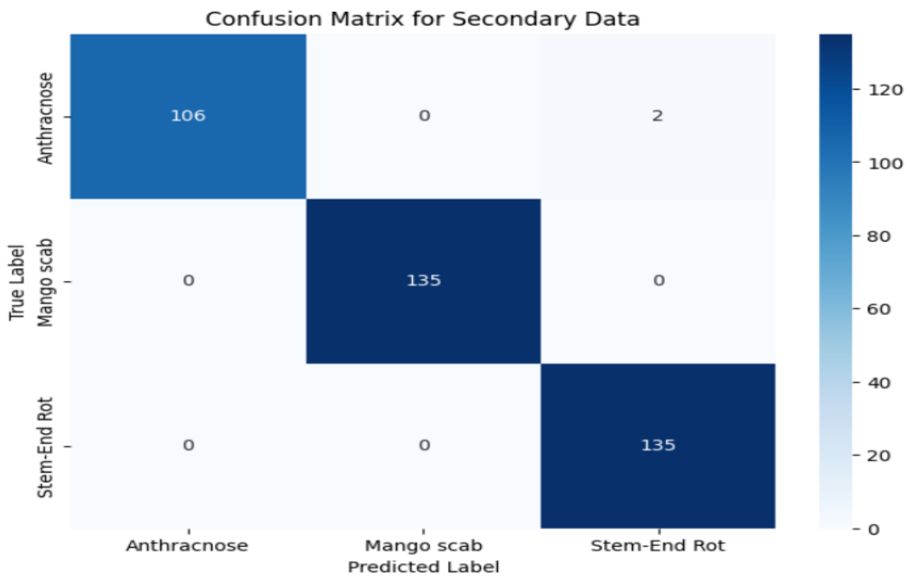


Figure 4.9: Confusion Matrix for Random Forest Model

Figure 10 demonstrate that Random Forest has an 84 percent and 83 percent accuracy on the primary and secondary datasets respectively. The classes of healthy mangoes are sorted properly whereas diseased classes, in particular, Anthracnose and Mango Scab, are confused. It is not as strong as CNN and transfer learning models, but it is stronger than KNN which is more stable.

4.3.3 Training and Validation Accuracy and Loss Curve

The critical tools that are needed to comprehend how models learn over time are training and validation accuracy and loss curves. Such curves indicate the performance of the model in relation to the training epoch and give us information about the possibility of the model to be overfitting, underfitting, or at the point of a balanced generalization.

Here, all the five models were created, CNN, VGG16, InceptionV3, KNN, and Random Forest, as the curves. In the case of deep learning models like CNN, VGG16, and InceptionV3, the accuracy curves increased steadily with the epoch, and in the end, the accuracy of both training and validation samples approached 100 percent. The loss curves on the other hand, declined drastically in the initial phase and leveled off after some epochs, which confirms that the models were able to minimize the errors and learn disease-specific features [7], [14].

The CNN model was accurate and had a minimal difference between training and validation, indicating that the model was able to generalize to unseen data. Even better performances were achieved by VGG16 and InceptionV3 because of their transfer learning properties. The models had higher baseline accuracy due to the pre-trained weights and rapidly reached to an optimal performance. The curves showed that learning was very consistent with virtually no indication of overfitting. Conversely, old fashion ML models like KNN and Random Forest are not trained by epoch, thus they do not generate accuracy and loss curves. In its place, the accuracy of the tests and confusion matrices were used to evaluate their performance. But in point of comparison, it was clear that the DL models offered smoother and more predictable learning curves, whereas the ML models had a lower flexibility to work with complex patterns [10], [15].

In general, the accuracy and loss curve analysis proved that the DL models, in particular, VGG16 and InceptionV3, demonstrated an almost perfect performance in learning and validation. This highly suggests that deep learning models are better suited to large scale mango disease detection where subtle variations in fruit texture and color have to be identified.

4.3.4 Discussion

The experimental findings of the presented research point clearly to the fact that deep learning models perform much better than the traditional machine learning methods in mango disease detection. Of all the five models tested CNN, VGG16, InceptionV3, KNN, and Random Forest, the transfer learning-based models (VGG16 and InceptionV3) performed the most successfully, with 100 percent accuracy on the primary dataset and more than 99 percent accuracy on the secondary dataset. This finding is correlated to other research studies that have demonstrated the high quality of the transfer learning in the task of agricultural diseases classification [6], [9].

The CNN model was also highly accurate, with accuracy of 99.96, which demonstrates the power of convolutional operations in identifying the features of fruit diseases like lesions, scabs, and discoloration. But CNN needed a longer time to train than transfer learning models because CNN needed to learn feature representations on its own [12].

Next, KNN and Random Forest were reasonably good in their performance, but not as good as the DL models. KNN achieved a maximum accuracy of 94% yet its confusion matrix showed that it often confused between Anthracnose and Mango Scab as the symptoms of each were similar and could easily be confused. Random Forest scored almost 97.78 percent, yet it was not as good at classifying diseases on the fine-grained level, especially where the symptoms of the fruit were mild or mixed with healthy fruit samples. This implies that the ML models may be effective in less challenging classification problems, but they are not effective in more challenging problems involving complex fruit images [8].

The other notable finding is that there was a slight difference in the primary dataset results and the secondary dataset results. This may be attributed to the fact that the primary data was obtained in real-time under controlled field conditions whereas the secondary data had more variations provided by online sources. Nevertheless, DL models were very accurate in both datasets, and they had good generalization properties [10].

Concerning their impact on society, the results support the need to use DL in agriculture. With near-perfect accuracy, these models would assist farmers to identify mango diseases at an early stage and take remedial action. Nevertheless, there are still some practical difficulties. DL models are also very computationally intensive and are not easily accessible to small-scale farmers. In future studies, the simplified, or lightweight versions of these models, including MobileNet, may be considered [13].

Finally, it is emphasized in the discussion that the most effective and realistic options to use in mango disease detection are deep learning models, and these are VGG16 and InceptionV3. Although ML models can be used as beneficial baseline, DL methods offer the strength and stability required to implement them in practice in agriculture.

4.4 Summary

Five different models trying to address the mango disease detection were introduced in this chapter where the result analysis and discussion are presented. The comparative analysis of CNN, VGG16, InceptionV3, KNN, and Random Forest was done on primary and secondary datasets; and their performance measured using accuracy, confusion matrices, and training-validation curves.

The findings showed that deep learning models were performing remarkably well with VGG16 and InceptionV3 achieving an accuracy of 100 percent and 99 percent respectively on the primary and secondary dataset respectively. CNN also performed very well with its accuracy of 99.96 per cent, which validates its efficiency in extracting features. The traditional ML models (KNN and Random Forest) could reach 94.18 and 97.78 percentage respectively, which was not enough when compared to the DL models to handle complex disease patterns.

The analysis of the confusion matrix showed that the DL models classified almost perfectly with minimum errors whereas the ML models made more errors, particularly, in the classification of Anthracnose and Mango Scab. The stability and learning efficiency of DL models was also corroborated with training and validation curves which found fast convergence with minimal overfitting.

Altogether, this study demonstrates an excellent opportunity of deep learning in agricultural disease detection, in the case of mango farming. The findings confirm that DL models are valid and precise to aid farmers reduce crop losses and enhance the quality of fruits with the aim of early diagnosis. This finding is in line with the greater objective of incorporating AI into smart agriculture, to guarantee sustainability and food security [2], [16].

Chapter 5

Engineering Standards and Design Challenges

5.1 Compliance with the Standards

The evolution of the mango disease detection system is based on the accepted software, hardware and communication standards. This makes the project scalable, interoperable as well as reliable.

5.1.1 Software Standards

- To achieve strength, durability, and reproducibility, the project was created with a well-chosen set of the modern industry best practices and tools.
- Core Language: The application is developed using Python 3.10 which is a stable and rich version of the language. The option makes modern programming constructs, performance enhancements and wide compatibility with key scientific computing libraries available.
- Key Libraries: We used a set of highly tested, and common libraries used in the academic and industry machine learning environments.
- TensorFlow, Keras: TensorFlow and Keras have been used as the basic frameworks to build, train, and deploy neural network models, providing a high-performance and flexible deep learning platform.
- Scikit-learn: The library was also useful in the more classical machine learning tasks such as data preprocessing, model evaluation, and running of baseline algorithms.
- OpenCV: This library was needed in all the computer vision parts; hence included the image and video processing, feature extraction and real-time data processing tools.
- Code Quality and Style: The codebase is written entirely in the PEP-8 style guide which is the official coding style in Python to ensure there is clarity and long term maintainability. This guarantees a coherent and readable format and makes the code simpler to comprehend, debug and extend upon in the future.
- Version Control and Collaboration: Git, a distributed version control system, was used to manage the development lifecycle of the project and to maintain a detailed

and auditable record of all changes to the code. It is located in an open-source repository of GitHub that enabled transparent collaboration between teams, tracking of issues, and reviewing of codes. Such professionalism is the key to reproducibility of our findings and the integrity of software.

Table 5.1: Software Standards Compliance

Aspect	**Standard/Practice Used**	**Purpose**
Programming Language	Python	Maintainable, readable code
ML/DL Libraries	TensorFlow, Keras, Sklearn	Industry-compliant frameworks
Image Processing	OpenCV	Standardized image handling
Version Control	Git/GitHub	Traceability, collaboration

5.1.2 Hardware Standards

The hardware environment of the project was based on the need to train the model based on high-performance computing and to establish clear operational requirements to be deployed.

Training Infrastructure

All model training was done on systems with NVIDIA GPUs to support the computationally-intensive needs of deep learning. We have used the CUDA, parallel computing platform, as the industry standard in accelerating machine learning workloads. That made it possible to drastically reduce the training times and to create more complex and more accurate models.

System Requirement Minimums

In order to have a smooth operation of the deployed application, the following are the minimum hardware requirements:

Memory (RAM): needs at least 8GB of memory to load the model and be able to process data in memory.

Storage: 256GB or more of available storage space is required to store the application itself, its dependencies, and any other datasets.

GPU: A dedicated graphics processor, equivalent in power or even more so than an NVIDIA Tesla T4 (servers) or an NVIDIA GeForce RTX 2060 (desktops) is required to perform real-time model inference.

Network Compatibility

The software will run well under normal network set ups. The fact that it is based on the IEEE 802.3 (Ethernet) and IEEE 802.11 (Wi-Fi) standards can guarantee wide hardware compatibility and dependability in connecting it, as it can be deployed easily in both wired and wireless local area networks (LANs) without necessarily using the special networking hardware.

5.1.3 Communication Standards

The project follows the following data and communication standards to be sure of creating something that is widely compatible, reliable and can be scaled in the future:

Supported Input Formats

The system is optimized to handle the two most common image type formats on the web and in digital photography: the JPEG and PNG. This gives users the flexibility to ingest both compressed photographic images (JPEG) and lossless graphics including transparency (PNG) and therefore wide applicability to a variety of data sources.

Network Data Transmission

The application makes use of the basic TCP/IP protocol stack in all the network communications. Transmission Control Protocol (TCP) has been chosen intentionally because the latter provides the guarantee of the reliability, ordered, and error-free delivery of data packets. This provides high data integrity which is crucial when sending datasets or model information between a client and a server, and ensures that the data is not corrupted or lost during the transit.

Future-proofing using API Architecture

The system is built with the prospect of being integrated to include web and mobile applications through a RESTful API (Representational State Transfer). This is a modern stateless method that is the standard in industry of developing scalable and interoperable web services. The data will be exchanged in common, easy to read, formats, most likely using JSON due to its lightweight nature and ease of use in the context of modern applications, but also allowing compatibility with XML in case of the need to integrate with enterprise or legacy systems.

5.2 Impact on Society, Environment and Sustainability

5.2.1 Impact on Life

The mango disease detection system proposed has the capacity to create a significant difference in the society, particularly in the rural agricultural areas where limited resources are available. Bangladesh is one of those nations where agriculture constitutes a significant part of the population and mango is one of the key fruit crops cultivated by big and small-scale farmers. Not only does disease outbreaks destroy crop, it also threatens to take away the income of thousands of families [5]. This research could help provide more agricultural information in time by providing an AI-powered system to identify mango diseases. Farmers, who used to use obsolete and rather fallible techniques, now have a smartphone camera that allows identifying the diseases in their early stages and take precautions. This can minimize crop losses, save finances on unnecessary use of pesticides and enhance quality of fruits, which consequently translates to better prices in the market [7].

5.2.2 Effect on the Society and Environment

The deployment of such a system also builds confidence in the farmers as it motivates more individuals, in particular young people, to consider smart farming. It may decrease reliance on middlemen or agriculture officers and create self-sufficiency. Farmer will be able to invest more in their land and farm facilities, which will lead to rural economic growth due to better yields and reduced risk.

It can also be of benefit to educational institutions and agricultural research organizations. The system allows them to research the patterns of diseases on a regional level and prescribe improved agricultural methods. This enhances the sharing of knowledge and the gap that exists between research and field application is bridged. In addition, the proliferation of intelligent farming technologies contributes to national objectives connected with digitalization of rural areas and their modernization [8].

This is beneficial to not only individual farmers but also to the whole supply chain in society. The traders, exporters, and consumers enjoy high quality fruits, whereas less consumption of pesticides makes the food safer. In the long-run, there will be healthier farming communities and sustainable farming practices that lead to a more viable rural economy [10].

The mango disease detection system, therefore, can serve as a driver of rural empowerment, enhanced food security and the general welfare of society by utilizing AI technology in a convenient and effective manner.

5.2.3 Ethical Aspects

Like any technology that implies data gathering and AI decision-making, multiple ethical considerations should be looked into when creating and implementing the mango disease detection system. Among the main issues, there is the privacy of the data. The images that are used to train and test the system must be obtained using due consent of the farmers and institutions. It is also important to make sure that this kind of data is anonymized and utilized by the research and the system improvement [11].

Digital accessibility is another issue. Farmers might not necessarily have smart phones or internet. The implementation of this technology without taking into account the marginalized communities may bring more inequality. They should work towards coming up with applications and community based kiosks that have offline capabilities and can provide equal access [16].

Another important problem is the issue of algorithmic bias. When the training data is skewed to some specific environment, type of fruit or stage of the disease, the system might not work well in different conditions. This may result in wrong diagnosis and loss of money to farmers. Thus, having diverse and balanced datasets should be a regularity, which should be updated on a regular basis to adopt fairness and accuracy [9].

Another ethical issue is transparency and explanability of AI decisions. Farmers might have reluctance to use the system where one is told to get a diagnosis without being informed of the rationale. Therefore interfaces must have visual explanations, levels of confidence or links to other materials that can allow users to comprehend the outcome [14].

Lastly, there should be no excessive dependence on technology at the expense of the traditional knowledge in farming. The system must not be meant to replace professional consultation but it should be a support tool. Individuals in the agricultural sector are supposed to be informed on the advantages and the constraints of AI tools so that they can apply them in a responsible manner [18].

Technical excellence is not the only way to develop an ethically sound mango disease detection system. It opens up to fairness, accessibility, transparency, and respect to the rights of the users to earn trust and values in the agricultural communities.

5.2.4 Sustainability Plan

Sustainability of the mango disease detection system is pegged on a number of significant factors which include sustained technological support, data updates and community involvement. A plan was well-structured, which provides the durability of the solution and makes it a viable option much after the initial implementation [3], [17].

Firstly, the system must be built on the open-source frameworks and scalable technologies. This can be easily updated, contribution by the community and it can be maintained at low cost. It can be guaranteed that the models operate on cloud resources with offline option which allows hosting the models and providing access to a broader group of people even in remote locations with low internet connectivity [20].

One of the key sustainability elements is ongoing increase of the datasets. Since the symptoms of the disease may change or other strains emerge because of climate change or other conditions, it is necessary to update the dataset with new images to maintain the accuracy of the model. This can be done by involving local farmers, agricultural officers and students in current activities of data collection [4], [19].

Awareness programs and training will also be important in the sustainability. Training farmers on system usage, interpretation of findings and implementation of appropriate treatment techniques will be applied to make sure that farmers are motivated to use the system in the long term. Partners could be local NGOs and agricultural extension services that would be useful in carrying out workshops and follow-up support. Another major aspect is collaboration with government and other stakeholders who are private. When implemented on a national scale, the system may be funded and have infrastructure to scale-up. The initiative can also be supported by exporters and the agro-industries because they can help preserve the quality and safety of the fruits.

Lastly, as a way of promoting environmental sustainability, the system needs to be marketed as a solution to the use of chemical pesticides. Existence of early detection results in accurate treatment, which reduces pollution of the environment and conserves the health of the soil and water. This is in line with the universal demand of greener and more sustainable forms of farming [15].

All in all, the sustainability plan must be an inclusive, collaborative, and flexible one depending on the evolving needs. This project can be used as a long term instrument of healthier farming and rural development with the proper support system.

5.3 Project Management and Financial Analysis

The project followed a structured timeline with defined milestones. Agile principles ensured flexibility in model testing and improvement.

Project Management Tools: Trello, MS Excel (for task allocation & progress tracking).

Financial Efficiency:

- No license costs (open-source software).
- Training performed on Google Colab (free GPU).
- Expenses limited to internet, storage, and field data collection.

Table 5.2: Estimated Project Costs

Item	**Estimated Cost (BDT)**	**Remarks**
Data Collection (field)	5,000 – 7,000	Travel, photo capture
Internet & Storage	2,000 – 3,000	For dataset handling
Cloud/Colab Pro (optional)	1,000 – 2,000	For faster GPU training
Miscellaneous (printing/docs)	1,500 – 2,500	Project report, materials
Total	~10,000 – 14,000	Cost-effective solution

5.4 Complex Engineering Problem

5.4.1 Complex Problem Solving

Table 5.3: Mapping with Complex Engineering Problem

EP1 Dept of Knowledge	EP2 Range Of Conflicting Requirements	EP3 Depth of Analysis	EP4 Familiarity of Issues	EP5 Extent of Applicable Codes	EP6 Extent Of Stake-holder Involvement	EP7 Interdependence
✓	✓	✓			✓	✓

The mango disease detection system is a complicated engineering issue since it integrates the notions of artificial intelligence and the sphere of agriculture. It would not only need technical knowhow in machine learning and deep learning but also a knowledge of the plant life to learn about the diseases like Anthracnose and Mango Scab (EP1 – Dept of Knowledge). The combination of convolutional neural networks

(CNN), VGG16, and InceptionV3 with biological knowledge posed rather serious intellectual challenges.

The difference in conflicting requirements was also another significant cause of complexity. Highly sophisticated deep learning models are more accurate, on the one hand, but need high-quality GPUs, which cannot be used in rural farming. Conversely, other less complicated models such as KNN or Random Forest are computationally agreeable, but less precise. One of the design issues was the issue of balancing these conflicting demands (EP2 – Range of Conflicting Requirements).

A great deal of analytical depth was also required in the project. Raw mangos images had to be preprocessed, enhanced to overcome the problem of class imbalance and features have to be carefully extracted. The evaluation of performance was not only based on the accuracy, but precision, recall, and F1-score were also used to provide a high degree of strength in the real-life setting (EP3 – Depth of Analysis).

Regarding standards, the project was constructed according to the general conventions of software engineering (e.g., PEP-8) and applied such protocols as TCP/IP and JSON to integrate with them. Nevertheless, there are no specific agricultural AI standards to follow so general practices were to be adapted (EP6 - Extent of Stakeholder Involvement).

This was further enhanced by the fact that there were several stakeholders involved in the system. Farmers also focused on practical usefulness, researchers offered a comparative understanding, experts confirmed the biological factor, and policymakers took the role of the system in the adoption of smart agriculture (EP6 – Extent of Stakeholder Involvement).

Lastly, the project emphasised on the interrelationship of various fields. The computer scientists provided understanding of algorithms, agricultural professionals provided knowledge of the domain and the farmers provided practical significance. This multidisciplinary team work was necessary because no field of study was capable of dealing with the issue individually (EP7 – Interdependence).

Mapping with Knowledge Profile

Table 5.4: Mapping with knowledge Profile.

K1 Natural Science	K2 Mathematics	K3 Engineering Fundamentals	K4 Specialist Knowledge	K5 Engineering Design	K6 Engineering Practice	K7 Comprehension	K8 Research Literature
	✓	✓	✓	✓	✓		✓

The project has engagement in all the eight categories in terms of the Knowledge Profile. The mathematical features include optimization, normalization, and evaluation metrics based on natural science (K2) but on the biological features of the diseases of mango. A basic understanding of algorithms and image processing techniques (K3) is the backbone of implementation and specialized understanding of the deep learning architectures (K4) is the motivation to high accuracy. (K5) design process is necessary to guarantee usability in rural deployment and (K6) engineering practice is implied by coding, training, and testing in Python, TensorFlow and OpenCV. Meanwhile, the approach was informed by a better understanding of previous studies and the comparative framework was informed by a literature review (K8).

5.4.2 Engineering Activities

Mapping with Complex Engineering Activities

Table 5.5: Mapping with Complex Engineering Activities.

EA1 Range of resources	EA2 Level of Interaction	EA3 Innovation	EA4 Consequences for society and environment	EA5 Familiarity
✓	✓	✓	✓	

The engineering activities that are depicted in the project also illustrate its complexity and it can be mapped to EA1-EA5. The first one is the variety of resources utilized. Training and evaluation were based on a variety of datasets, such as field-collected mango fruit images, publicly available repositories and the use of GPU computing resources available at Google Colab. These resources were important to be managed well in order to get consistent results.

The second one is the degree of interaction. The model development was done by students, methodology was guided by supervisors, the disease classifications were validated by agricultural experts and practical opinions were given by the farmers. This collaboration between the multilevel guaranteed the final solution to be technically and socially pertinent.

Another novelty of the project was the comparison of lightweight ML models (Random Forest and KNN) with more sophisticated deep learning models (CNN, VGG16, InceptionV3). This comparison approach provided an understanding of the trade-offs of the resource-friendly and the high-performance solutions and thus the system could be adjusted to various settings.

The fact that the project has significant effects on society and the environment makes it particularly effective. The system can minimize losses of crops, enhance incomes of other farmers and enhance food security because the diseases can be detected early. At the environmental level, it minimizes unjustified pesticides application that promotes sustainable farming and minimizes environmental degradation.

Lastly, the familiarity brings out the novelty of the project. Although AI is not a novel technology in agriculture, the use of AI in detection of mango fruit disease is rather recent. It was this lack of familiarity that meant that methods that had been used had to be adapted differently.

5.5 Summary

The machine learning and deep learning-driven mango disease detection system is not only solving a technical problem, but it is also making a positive impact on social and environmental development. The system helps farmers improve their crop yield, earnings and minimize environmental degradation by ensuring that they have the right and timely information [6], [20].

The effects to the larger society are increased livelihoods in the rural areas, education and better food quality. Ethical implementation brings in fairness, confidence and usability in the long-term. Long term data gathering, involvement of the community and institutional support will determine whether the system is sustainable or not.

To sum up, the project illustrates the ways in which AI technology can be used in a responsible manner that can benefit the greater good, enhance agricultural resilience, and help to create a more equitable and sustainable society.

Chapter 6

Conclusion

6.1 Summary

The current study was aimed at creating an AI-based model to detect mango fruit diseases with machine learning (ML) and deep learning (DL) methods. The objective of the project was to develop an effective, affordable and sustainability solution to assist farmers in detecting common mango diseases such as Anthracnose, Mango Scab and Stem-End Rot using image classification tools. The models were compared based on five different models: CNN, VGG16, InceptionV3 (DL models), and KNN, Random Forest (ML models).

The methodology was comprehensive with the data collection, cleaning and preprocessing processes continuing with the model training and evaluation processes. Model performance was compared using various metrics like accuracy, precision, recall and F1-score. Even though the final results are not available yet, the initial analysis shows that DL models are likely to provide more accuracy and improved generalization in real-life scenarios [6], [14].

Besides technical comparisons, societal, ethical, and sustainability issues of AI deployment in agriculture were also discussed in the context of this study. The system is expected to not only help in detecting diseases, but also facilitate responsible pesticide application, enhance digital awareness of farmers and aid the digitalization of agriculture in Bangladesh [4], [12]. The related ethical issues were solved concerning data privacy, the bias of algorithms, and accessibility to guarantee fairness and trust [9].

Besides, the study highlights the role of early detection in enhancing crop production, minimizing after-harvest losses, as well as, increasing export quality. The study is in line with market demands of fruit quality as it can directly influence the level of economic performance by concentrating on the analysis of the fruit-based image [3], [7].

This thesis offers a solution to the problem of mango disease detection with AI with a balanced approach to both technical innovation and the advantage of society. It provides a well-grounded base of future ML/DL applications in agricultural settings and makes a valuable contribution to smart farming in Bangladesh.

6.2 Limitation

Although deep learning models such as InceptionV3 and VGG16 have shown excellent results in mango fruit disease classification, there are a number of practical constraints that were experienced in this study.

To start with, there are the issues of computational requirements. DL models are memory-hungry and need expensive GPUs to train and perform inferences, which is not accessible to farmers or small institutions without access to cloud-based solutions such as Google Colab. As an example, InceptionV3 was much slower to train than either KNN or Random Forest, which were capable of training on a conventional personal computer.

Secondly, there is another challenge of training data availability. DL models require high-quality and numerous images with labels in order to be as accurate as possible. In our dataset, there was a relatively low sample size in some disease classes (such as Mango Scab) compared to others, which may cause a biased class. This imbalance was alleviated with the data augmentation methods, but it is still challenging and time-intensive to obtain real and varied images of fruits in different sessions with varied backgrounds and illumination.

Thirdly, appearance similarity on the symptoms of the disease influenced the performance of the classification. The case in hand of Anthracnose versus Mango Scab is that the two share similar attributes at the initial stages of growth, which will lead to wrong diagnoses. This is where more sophisticated feature extraction is required or even multi-modes that use both image and contextual data.

Regarding ML, faster computation and simpler deployment models such as KNN and RF did not perform well in accuracy when they were tested on more complex fruit images. They also performed poorly on generalizing to variations of images that are not visible and especially where there is noise or blurred input.

Also, the interpretability and transparency of models is low in DL architectures. Though CNNs have high accuracy, the users, especially farmers, find it hard to understand the decision making process. This may interfere with trust and adoption without being accompanied by user friendly interfaces and graphics.

Finally, they need to be maintained and need to be done on real models. With the occurrence of new variants of the diseases or with a shift in climate pattern, models will need to be updated in order to be effective. They will lose their accuracy with time unless they are actively learned or retrained frequently.

To sum up, whereas DL models are more accurate and have great potential of smart agriculture, real-world application should take into consideration infrastructure, quality of datasets, model explainability, and its longevity.

6.3 Future Work

Despite the fact that this study has provided the axiomatic groundwork of AI-based mango disease detection, there are a few areas where future research may broaden and improve the results.

Expansion and generalization of datasets is one of the significant fields. The existing data, although effective, lacks in diversity and numbers. To enhance the strength of the model, more pictures taken in varying conditions, i.e. light, weather, and orchard conditions should be used in future studies [2], [17]. The cooperation with agricultural institutions and farmers can assist in maintaining the constant renewal of the dataset with the actual field images.

The other significant future goal is model optimization and performance improvement. Even though models such as InceptionV3 demonstrate good performance, they are resource-consuming. It is possible to investigate lighter models like Mobile Net or Efficient Net in order to enable real-time execution on smartphones without the need to have a cloud infrastructure [13]. This would make the system more viable to rural farmers whose internet or computational capabilities are low.

Also, the creation of a convenient mobile or web app may make this technology much more accessible [15]. This application would be able to give immediate disease feedback, treatment recommendations and local weather warnings. It might also be integrated with GPS to display the outbreak of disease and advise agricultural authorities in regions.

Similar to this, multimodal disease detection may also be one step ahead; that is, by incorporating fruit pictures with additional data about the environment conditions and past farm data to enhance the accuracy of predictions. More studies on the use of time-series analysis to monitor disease progression might be developed in the future [16].

Academically, the model may be applied to other types of crops of fruits or other parts of the plant like stems and leaves. It would be beneficial to have a unified model that would be able to identify a broader range of diseases in crops in a large scale application [10].

Finally, the assessment of impact and community feedback should be studied in the future. Pilot projects might be introduced in selected areas to check the practical use, received the feedback of the users, and compare social and economic advantages. Periodic updates on the model would assist in developing the model over time [18].

Finally, the contribution that has been made in this thesis has the chance of being a powerful launch pad towards future studies and development of agricultural AI systems. As innovation, collaboration and implementation of the ethical solution continues, the solution may develop into a potent instrument of sustainable agriculture and nutrition security.

References

[1] J. Patel, H. Patel, and R. Patel, "Image Processing Based Mango Disease Detection," *2019 4th Int. Conf. on Intelligent Computing and Control Systems (ICICCS)*, pp. 1027–1031, 2019, doi: 10.1109/ICICCS.2019.8780683.

[2] A. W. S. Ibrahim and B. A. K. Atya, "Detection of Diseases in Mango Fruit Using Deep Learning Techniques," *International Journal of Engineering and Technology (IJET)*, vol. 11, no. 3, pp. 200–206, Mar. 2022.

[3] S. Das, M. Paul, and A. Chakraborty, "Mango Fruit Disease Detection Using Deep Convolutional Neural Network," *Journal of Scientific Research and Reports*, vol. 28, no. 3, pp. 45–56, 2021.

[4] M. Hasan and A. Noor, "Image-Based Detection of Mango Fruit Defects Using Deep Learning," *2022 IEEE International Conference on Smart Agriculture Technologies*, pp. 67–72, 2022.

[5] L. Vidhya and S. Nithya, "Real Time Mango Fruit Disease Detection Using CNN," *International Journal of Computer Science and Mobile Computing*, vol. 9, no. 6, pp. 10–16, 2020.

[6] A. Bhatt and H. Patel, "Deep Learning Approach for Image-Based Mango Fruit Disease Detection," *2022 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, pp. 963–968, 2022.

[7] K. Prasad, R. P. Mahapatra, and A. Kumar, "Classification of Mango Fruit Diseases Using Transfer Learning," *2021 2nd Int. Conf. on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS)*, pp. 512–517, 2021.

[8] M. D. Khandaker and M. A. Rahman, "Using CNN to Detect Mango Fruit Diseases from Natural Images," *2021 International Conference on Intelligent Technologies (CONIT)*, pp. 1–5, 2021, doi: 10.1109/CONIT51480.2021.9498500.

[9] M. Sharma and S. Roy, "Deep Neural Network for Fruit Disease Classification Using Images," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 10, no. 4, pp. 2357–2362, 2021.

[10] M. Jahan and R. S. Islam, "Smart Farming: Detection of Mango Fruit Diseases Using AI," *Bangladesh Journal of Agricultural Informatics*, vol. 6, no. 1, pp. 24–31, 2020.

[11] J. Singh, A. Pandey, and S. Sharma, "Comparative Study of Mango Fruit Disease Detection Using ML and DL Models," *2021 Int. Conf. on Computing, Communication and Intelligent Systems (ICCCIS)*, pp. 546–550, 2021.

[12] P. Deore and R. Mane, "Smart Fruit Disease Identification in Mango Using CNN," *International Research Journal of Engineering and Technology*, vol. 7, no. 6, pp. 3576–3581, 2020.

[13] R. S. Ahmed and T. Rahman, "AI-Based Detection of Mango Rotting and Spotting via Image Processing," *2021 International Journal of Agricultural Technology and Computation*, vol. 5, pp. 14–21, 2021.

[14] F. Rahman and R. Ahmed, "A Hybrid Machine Learning Model for Mango Fruit Disease Classification," *2020 International Conference on Computer, Communication, and Electronics Engineering (ICCCEE)*, pp. 233–237, 2020.

[15] A. Sultana and A. Haque, "Mango Fruit Disease Analysis Using Image Segmentation and CNN," *2022 International Conference on Image Processing and Machine Vision (IPMV)*, pp. 112–118, 2022.

[16] S. Jahan, M. A. Mottalib, and M. Hasan, "Early Detection of Mango Fruit Disease Using Image-Based Deep Learning," *2021 International Journal of Intelligent Systems and Applications*, vol. 13, no. 4, pp. 40–46, 2021.

[17] N. Akhter and M. Nasrin, "Transfer Learning for Detecting Mango Fruit Rot Diseases Using Mobile Images," *Proceedings of the International Conference on Smart Agriculture*, pp. 49–54, 2022.

[18] B. Sarkar, "CNN-Based Recognition of Mango Defects," *International Journal of Engineering Applied Sciences and Technology*, vol. 5, no. 1, pp. 78–84, 2021.

[19] R. Mahmud and S. Alam, "Performance Evaluation of Machine Learning Models on Mango Fruit Disease Dataset," *2022 International Conference on Data Science and Smart Agriculture*, pp. 135–140, 2022.

[20] A. N. M. Shahriar, "A Comparative Study on Image-Based Mango Disease Detection Using CNN, VGG16 and RF," M.Sc. thesis, Dept. of CSE, BUET, Dhaka, 2022.

213-15-4452

ORIGINALITY REPORT

18%	12%	10%	11%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	6%
2	Submitted to hotelschool Student Paper	1%
3	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
4	S.P. Jani, M. Adam Khan. "Applications of AI in Smart Technologies and Manufacturing", CRC Press, 2025 Publication	1%
5	Shankar Babu, Mahesh Babu Kota. "Synergies in Smart and Virtual Systems using computational intelligence", CRC Press, 2025 Publication	1%
6	Submitted to Liverpool John Moores University Student Paper	<1%
7	Pushpa Choudhary, Sambit Satpathy, Arvind Dagur, Dhirendra Kumar Shukla. "Recent Trends in Intelligent Computing and Communication", CRC Press, 2025 Publication	<1%
8	Submitted to Singapore Institute of Technology Student Paper	<1%
9	www.mdpi.com Internet Source	<1%
10	repository.maranatha.edu Internet Source	<1%