

**MANGOLEAF: DISEASE DETECTION USING MACHINE
LEARNING**

BY

S. M. Mehedi

ID: 241-17-001

This Report Presented in Partial Fulfillment of the Requirements for
The Degree of Master of Science in Management Information System

Supervised By

Dr. Arif Mahmud

Associate Professor and Associate Head

Department of CSE

Daffodil International University

Co-Supervised By

Abdus Sattar

Assistant Professor

Department of CSE

Daffodil International University



DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

MAY 2025

APPROVAL

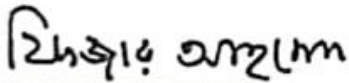
This Thesis titled **MangoLeaf: Disease Detection Using Machine Learning**, submitted S M Mehedi to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Master of Science in Management Information System and approved as to its style and contents. The presentation has been held on 23 May 2025.

BOARD OF EXAMINERS



Dr. Sheak Rashed Haider Noori
Professor and Head
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Board Chairman



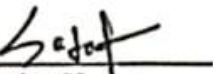
Dr. Fizar Ahmed
Associate Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner 1



Mr. Abdus Sattar
Associate Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner 2



Sadat Hasan
Data Scientist
Risk Management Division,
BRAC Bank Limited

External Examiner

DECLARATION

We hereby declare that, this project has been done by us under the supervision of Dr. Arif Mahmud, Associate Professor and Associate Head, Department of CSE Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:



Dr. Arif Mahmud

Associate Professor and Associate Head

Department of CSE

Daffodil International University

Co-Supervised by:

Name


Mr. Abdus Sattar

Associate Professor & Director, M.Sc

Department of CSE

Daffodil International University

Submitted by:



S M Mehedi

ID: 241-17-001

Department of CSE

Daffodil International University

ACKNOWLEDGEMENT

First we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

We really grateful and wish our profound our indebtedness to Dr. Arif Mahmud, Associate Professor and Associate Head, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of Machine Learning 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 thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

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

ABSTRACT

Mango (*Mangifera indica*) is the most commonly cultivated fruit crop in Bangladesh and a supporting pillar of agriculture in the nation as an income source. Mango yield is typically affected by a chain of leaf diseases such as anthracnose, bacterial canker, cutting weevil, die back, gall midge, powdery mildew, scooty mould disease resulting in poor yields and quality. Manual identification following traditional techniques takes a long time, is not very accurate, and is not very helpful in surveying in large-scale. The study takes into account the use of machine learning algorithms, i.e., CNN model and a CNN model paired with pooling layers for the automatic classification and detection of mango leaf diseases from image data. A representative set of diseased and healthy mango leaves was collected and preprocessed for training and testing. All the algorithms were compared on the major metrics of accuracy, precision, recall, and F1-score. Maximum accuracy 95% was achieved with CNN model paired with pooling layers. The social, environmental, and ethical consequences of implementing such smart farm systems in rural Bangladesh are also included in this thesis. Through enabling the identification of diseases at an early stage and with accuracy, the system under consideration can prevent the use of pesticides that are harmful to the health of humans and enable sustainable production and enhance the livelihood of the local farmers. This research contributes to the application of AI in agriculture and establishes a foundation for further data-based management of plant diseases appropriate for Bangladesh's agricultural sector.

Table of Contents

Declaration.....	ii
ACKNOWLEDGEMENT.....	iv
Abstract.....	v
Chapter 1 Introduction	1
1.1 Introduction	1
1.2 Motivation	2
1.3 Rationale if the study.....	3
1.4 Research Question	4
1.5 Expected output.....	5
CHAPTER 2 BACKGROUND	6
2.1 Preliminaries / Terminologies	6
2.2 Related Works	7
2.3 Scope of the work.....	8
2.4 Challenges	8
CHAPTER 3 Research Methodology.....	10
3.1 Research Subject and Instrument	10
3.2 Data Collection Procedure/Dataset Utilized.....	11
3.3 Statistical Analysis	14
3.4 Proposed Methodology / Applied Mechanism	15
3.5 Implementation Requirements.....	16
CHAPTER 4 EXPERIMENTAL RESULT	18
4.1 Experimental Setup	18
4.2 Experimental Results & Analysis.....	19
4.3 Discussion	21
CHAPTER 5 IMPACT ON SOCIETY	23
5.1 Impact on Society	23
5.2 Impact on Environment	24
5.3 Ethical Aspects	24
CHAPTER 6 SUMMARY, CONCLUSION AND FUTURE RESEARCH	26

6.1 Summary of the study.....	26
6.2 Conclusions	27
6.3 Implication for Further Study	28
References.....	30

LIST OF FIGURES

FIGURES	PAGE NO
Figure 3.1: CNN Architecture	11
Figure 3.2: Anthracnose	11
Figure 3.3: Bacterial Canker	12
Figure 3.4: Cutting Weevil	12
Figure 3.5: Die Back	13
Figure 3.6: Gall Midge	13
Figure 3.7: Powdery Mildaw	13
Figure 3.8: Sooty Mould	14
Figure 4.1: Experiment Result CNN	19
Figure 4.2: Experiment Result CNN – Pooling pairs	20

CHAPTER 1

INTRODUCTION

1.1 Introduction

Bangladesh is a rural nation where mango is not only a fruit but also a symbol of tradition, economy, and seasonal festivity as well. From the north-western districts of Chapainawabganj, Rajshahi, and Naogaon to the south belt of Satkhira, mango cultivation is an integral part of the livelihood of thousands of farmers. One of the most outstanding problems experienced by farmer is the breakout of many diseases, especially those of mango leaves. Not only do they reduce production but also reduce the quality of fruits. Detection of such diseases is typically based largely on farmers' or experts' naked eye observation, which is open to personal opinion, time consuming, and prone to human error.

In the past few years, technological development — especially in fields pertaining to Artificial Intelligence (AI) and Machine Learning (ML) — has opened doors to new possibilities in the field of agriculture. Computer Vision techniques, when combined with strong classification algorithms, can be utilized for the automation of plant disease detection and identification from image data. The aim of this thesis is to apply various machine learning models to detect mango leaf diseases from image datasets with the ultimate goal of assisting local farmers in detecting diseases at an early stage and taking appropriate preventive measures.

Various machine learning techniques are being utilized in this research like CNN model and CNN model paired with pooling layers.

The inspiration for this research is deeply rooted in the real problems of Bangladeshi Mango farmers. Anthracnose, bacterial canker, cutting weevil, die back, gall midge, powdery mildew, scooty mould are some of the common diseases that can be visually identified in mango leaves. These can be treated and prevented if they are diagnosed at an early stage, before it spreads across the plantation. Machine learning in this context can

transform the agricultural disease management scenario in Bangladesh — faster, easier, and cheaper.

This research not only serves the research community in the area of machine learning solutions, but also to a very real-world issue common in our country. The proposed system, when implemented as a web or mobile application, can prove to be of immense help to farmers on the ground who lack expertise in agricultural techniques or receive timely recommendations. With cellular penetration increasing even in rural Bangladesh, it is not a remote possibility to click a photo of an infected leaf and obtain a quick diagnosis.

In the chapters that follow, this thesis shall elaborate on the background of mango leaf diseases, works in this regard so far, dataset preparation, methodology, model implementation, performance analysis, and finally the conclusions and future scope of the work.

1.2 Motivation

In Bangladesh, mango is not just a seasonal fruit but also a rural agriculture and local business lifeline, especially in Rajshahi, Chapainawabganj, Naogaon, and Satkhira districts. Mango cultivation is one of the major sources of livelihood and income for a large number of farmers in these districts. But the frequent outbreak of mango leaf diseases is posing a threat to healthy crop production. They tend to sneak up on you, destroying the leaves and general health of the plant before the issue becomes too late to ignore.

The fact is, numerous small- to medium-scale farmers in Bangladesh have limited or delayed farm advisory or professional diagnosis. Numerous traditional methods or ordinary tips are applied by them, which might be obsolete and of dubious credibility. Besides, the manual identification of disease is generally variable, tedious, and expertise-based — which not everybody has. This actual life situation of right and timely disease detection initiated this work.

As a student in training to be a student with a passion for machine learning and its practical applications, I was curious to learn how available technology might be able to offer a

solution for this very real issue. With smartphones and internet connectivity spreading even to rural areas, it was not difficult to imagine a simple solution: the farmer takes a picture of a diseased leaf and receives an immediate, precise diagnosis thanks to machine learning algorithms.

This thesis is motivated by the belief that technology has to solve real problems — especially those with practical, tangible impacts on food production, farmer income, and our country's agricultural economy. By developing a system that can detect mango leaf diseases through classification techniques like Convolutional Neural Networks (CNN), this study aims to bridge the gap between agriculture and digital solutions. If it succeeds, one day such a system can be transferred to a web or mobile platform that truly does well for the agriculture industry — not on paper, but at the ground level.

1.3 Rationale if the study

Bangladesh is a agricultural nation where most of the country's rural community survives on farming. Of the many fruits farmed here, mango holds a special place — not just locally, but to export as well. Mango farm production is also endangered by a multitude of diseases, especially ones that start devastating the leaves. These diseases have the ability to reduce the overall health of trees, reduce fruit yields, and in extreme cases, wipe out entire orchards. Despite this, the identification of disease still depends significantly on human visual observation, which is not always accurate or current, especially in remote areas.

There is a wide disparity between the need for early disease detection and access to resources by most farmers. Even though plant disease can be detected by agricultural specialists, these cannot be readily accessed and are mostly located in cities or district headquarters. For rural farmers like those in Shibganj or Puthia, access to timely expert advice is usually not easy, if not impossible. Such delay in detecting disease results in poor crop management and heavy financial losses.

As more and more individuals become connected to digital devices — especially smartphones — even in distant Bangladesh, there is a good prospect for machine learning and image processing to be used in quicker, more precise, and more convenient

identification of mango leaf diseases. Instead of relying on visual examination, a picture-based system facilitated by machine learning can provide disease identification to anyone anywhere.

The drive of this research is to tap into the potential of current machine learning models like CNN classifiers with actual needs of Bangladesh mango farmers. By building and testing models that have the potential to classify mango leaf photographs as diseases, this research will create a practical tool that will contribute directly to the farmers in the future.

Additionally, less of this work within the Bangladeshi context directly targets mango leaf disease detection by deployment with AI than a series of several various classification approaches. Not only does this project resolve a growing local agricultural concern but also contributes to existing literature within the topic area in machine learning research applied to agriculture.

1.4 Research Question

This study aims at using machine learning techniques to differentiate mango leaf diseases from images. The main research question for this study is:

For the purpose of complementing this broad question, the following sub-questions are also answered:

- How effective are CNN-based models in identifying different types of mango leaf diseases from images?
- What is the optimal CNN architecture for accurate detection of mango leaf diseases with minimal computational cost?
- How does the performance (in terms of accuracy) differ between a basic CNN model and a CNN model paired with pooling layers when applied to mango leaf disease classification?

In answer to these questions, this research aims to close the gap between practice and research on the application of smart agriculture in Bangladesh.

1.5 Expected output

The study outputs are technical as well as practical. Some of the products and deliverables include:

A labeled dataset of images of mango leaves showing infected and healthy samples, gathered from available sources or augmented with local data.

- Use of machine learning models CNN for visual disease classification of mango leaves.
- A comparative performance evaluation of these algorithms on accuracy, precision, recall, and F1-score, in order to determine the most suitable model for this purpose.
- A prototype or template that could be further developed into a working tool for mango farmers or agricultural extension services in Bangladesh.
- Recommendations for future development of an AI-driven mobile app or system capable of assisting farmers in the field to diagnose mango leaf diseases.

Overall, the research will provide the foundation for the creation of intelligent technology-based farming solutions that can be implemented and can be sustainable for the farming sector in Bangladesh.

CHAPTER 2 BACKGROUND

2.1 Preliminaries / Terminologies

Before delving into the technical details of this research, one should be familiar with some of the most critical terminologies and concepts that form the foundation of this research. Since this thesis incorporates machine learning, image processing, and plant pathology, the following terminologies are commonly used in this work:

- **Machine Learning (ML):** A branch of artificial intelligence (AI) that allows computer systems to learn from experience and enhance performance with data without being explicitly programmed. ML is used in the research to train model systems to classify mango leaf disease from images.
- **Classification:** Classification is a supervised learning task where the model is allowed to learn to categorize data into specific labels or classes. In this case, classification is applied to differentiate whether a mango leaf is healthy or has a specific disease.
- **Image Processing:** Method of performing operations on images in order to enhance them or extract valuable information. It is relevant to this research due to the fact that machine learning models are being trained on images of leaves.
- **Convolutional Neural Network (CNN):** Deep learning design of immense use when dealing with visual imagery. CNNs in this research are utilized for extracting features automatically from leaf images and making predictions.

Anthracnose, Bacterial Canker, Cutting Weevil, Die Back, Gall Midge, Powdery Mildew, Scooty Mould. They are some of the common mango leaf infections with self-evident visible symptoms and target classes of the detection framework.

The fact about these words contextualizes the methodology and analysis of the chapters to follow.

2.2 Related Works

Mango (*Mangifera indica*), being a highly profitable tropical fruit of Bangladesh, is highly prone to various leaf diseases that negatively influence the yield and quality. Like anthracnose and sooty mold etc., these impair the process of photosynthesis, reduce the yield of the fruits, and finally lead to damage to the economic return of the mango growers. Anthracnose caused by *Colletotrichum gloeosporioides* is also a common fungal disease that causes black spot and blight of leaves and fruits as reported in seminal literature published in *Plant Diseases* [1]. Sooty mold—a black mycelial encrustation—is also common through insect fecal excretion and even on the leaf surface of mango and further limiting plant health and production [2].

In response to such worries, against disease detection, with the latest research, machine learning (ML) and deep learning (DL) models have been utilized for early disease detection. Sharma et al. (2023) wrote a comprehensive study assisted by state-of-the-art image recognition software identifying mango leaf disease at very high rates of accuracy by using deep neural networks. Their work is indicative of the new application of artificial intelligence in precision agriculture, especially where traditional agriculture intervention is limited or is late [3].

Also, the one thesis in BRAC University (2023) used an off-the-shelf Convolutional Neural Network (CNN) to classify mango leaf diseases from over 7,500 labeled images with high accuracy classification [4]. This aligns with the overall debate among data-driven and model-driven ML practice, where a better architecture of the model would be as useful as better quality data. In agricultural plant disease diagnosis, data-driven approach is more convenient due to the differences in actual field conditions and image acquisition parameters .

Worldwide, mango remains one of the most utilized fruits, included within the top five quoted by WorldAtlas (2021) [5]. It is precisely for this reason that mango plants have to be preserved using technology. Introducing ML-based disease diagnostic systems, farmers across the globe, especially from developing nations like Bangladesh, can be provided with real-time notifications, preventing the utilization of pesticides and ensuring more sustainable cultivation.

2.3 Scope of the work

The scope of the research is precise but restricted. It is to implement a mango leaf disease detection system from image data by machine learning, proved and available for use in the real world in Bangladesh.

Key points falling under the scope are:

- Use or acquirement of a predefined dataset of images of mango leaves having multiple diseases and healthy ones.
- Training and validation of the CNN model and a CNN model paired with pooling layers.
- Accuracy and performance metrics evaluation to conclude the best-performing model.
- Providing recommendations on how the selected model can be converted to a mobile or web application for farmers or offices of agriculture.

What is excluded:

- Building an entire production-ready mobile application (though future recommendations might include it).
- Detection of mango tree's fruit or trunk diseases.
- Non-image-based diagnostic methods (e.g., chemical or molecular analysis).

This effort is intended to provide a baseline to be extended in the future by other developers or agronomists in Bangladesh.

2.4 Challenges

In this study, there will be certain problems — technical and practical:

Model Generalization: Machine learning models will perform optimally on clean, preprocessed data but do not generalize to images that are heterogeneous — with lighting, background, or leaves partially occluded.

Hardware Limitations: Training CNN models requires the use of GPUs, something that no one out there can help but be a student or a researcher. And using heavy models on light hardware is an operational limitation that must be met.

Field Validation: Without field testing or specialist validation, it becomes difficult to determine if disease predictions in actual farm conditions hold true.

Even with all of these difficulties, this study is on the right path when it comes to applying AI in Bangladeshi agriculture — and difficulties can be overcome when better systems are built in the future.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Research Subject and Instrument

This research is aimed at automatic mango leaf disease diagnosis using machine learning-based classification techniques. The research goal here is to identify unique signs of common mango leaf diseases using image processing for the detection of visual symptoms of the diseases. These are the common regular ones such as Anthracnose, Bacterial Canker, Cutting Weevil, Die Back, Gall Midge, Powdery Mildew, Sooty Mould that usually occur in mango orchards of districts such as Rajshahi, Chapainawabganj, Meherpur, and Dinajpur, which are well-known mango-producing districts of Bangladesh.

The tool being utilized here is a mix of machine-based classification systems and field data. The primary tool was a custom dataset of images of mango leaves—collected through field trips and the internet—which formed the foundation of model training. The setup for experimentation was established with the use of machine learning libraries like TensorFlow, Keras, and scikit-learn, with programming being implemented in Python. These packages allowed the research to simulate a real model-training process, compare the performance of other different classifiers, and present results with statistical visualization.

Hand labeling tools were also used to sort the retrieved set of images by disease classes. All the models were then compared based on performance using a confusion matrix and accuracy, precision, recall, and F1-score metrics. They were used as comparison analysis tools.

The reason for selecting those particular model CNN model and a CNN model paired with pooling layers was that they have been used extensively in the past research works in image classification and complexity and interpretability involved in it. While CNN is a deep learning model that learns image pixel features, the other models are handcrafted feature based, so it gives a fair playing ground to compare.

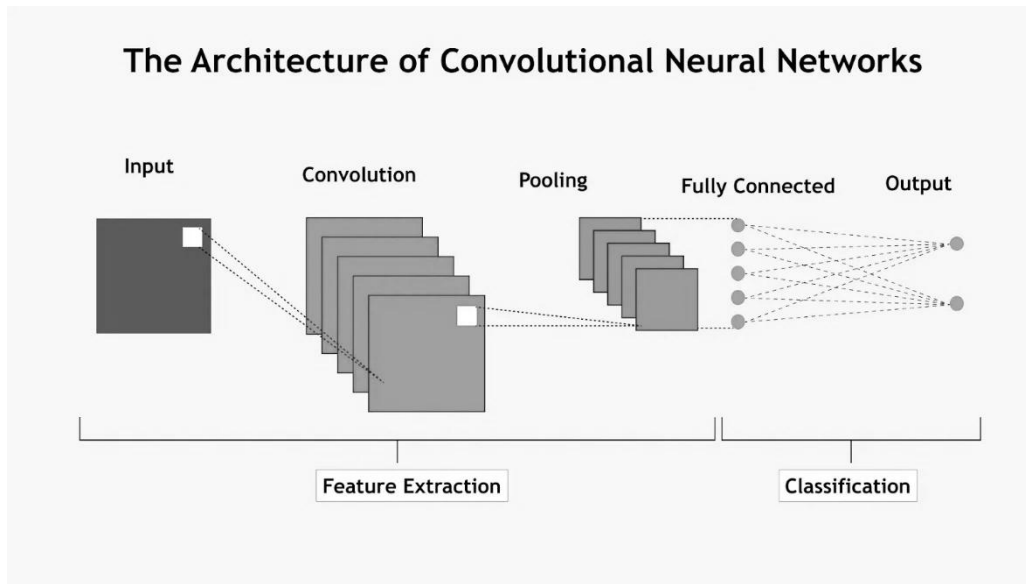


Figure 3.1: CNN Architecture

3.2 Data Collection Procedure/Dataset Utilized

The model depends on the dataset, whose legitimacy is derived from the diversity and quality of training data. Images of mango leaves used in this research were obtained from two general sources, namely open-access online libraries and field photography.

Other than field data, images were also acquired from open data sources such as Kaggle, GitHub repositories, and research repositories, for the most part for the PlantVillage dataset. The data included high-resolution well-labeled images for diseases that were not readily available in field conditions.

The most common anthracnose, caused by the fungus *Colletotrichum gloeosporioides*, is the most widespread mango disease that attacks twigs, flowers, fruits, and leaves and results in black spotting and soft rot of the fruit. It occurs in wet conditions and significantly reduces market value and yield [1].



Figure 3.2: Anthracnose

Bacterial canker of mango, caused by *Xanthomonas campestris* pv. *mangiferaeindicae*, leads to water-soaked lesions on leaves, twigs, and fruits that develop into black cankers. It will cause dieback and fruit or leaf drop and is a commercial mango orchard issue [6].



Figure 3.3: Bacterial Canker

Mango weevil (*Derolus* spp.) is a weevil stem borer that attacks young mango branches, causing wilting and dieback. Weevil borer adults oviposit through soft shoots, and larvae within tunneling inhibits nutrient transport [7].

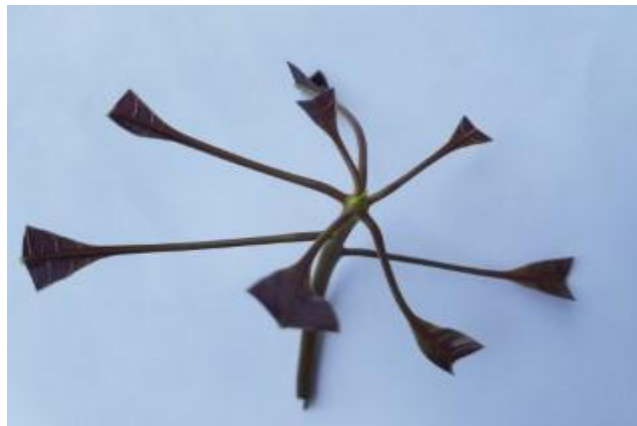


Figure 3.4: Cutting Weevil

Die back in mango is most commonly associated with fungal infections like *Botryodiplodia theobromae*, and is characterized by progressive drying of twigs from the top downwards. Pruning injury and stress conditions aggravate the condition, generally reducing the vigor of a tree [8].



Figure 3.5: Die Back

The mango gall midge (*Procontarinia matteiana*) lays eggs in mango inflorescences, where galls and abnormal tissues are formed due to the presence of the larvae. Infestations lead to poor fruit set and panicle deformity [9].



Figure 3.6: Gall Midge

Powdery mildew, caused by *Oidium mangiferae*, infects young leaves, panicles, and fruits, causing white powdery fungal growth. It affects flowering as well as fruit setting, particularly during cool, dry weather followed by humidity [10].



Figure 3.7: Powdery Mildew

Sooty mould is not a disease but a fungal growth on the sugary excretions (honeydew) of insects like aphids and mealybugs. It forms a black coating on leaves and fruits, impairing photosynthesis and reducing fruit quality [11].



Figure 3.8: Sooty Mould

Preprocessing data for modeling:

All images were resized to a size of 240×320 pixels. Normalization was applied in order to scale pixel values in the range between 0 and 1. Data augmentation techniques such as rotation, flip, zoom, and brightness were utilized to synthesize growing the size of the dataset and introducing variability. An 90:10 stratified train-test split was utilized to preserve class balance. The images were then labeled as being one of the eight classes: anthracnose, bacterial canker, cutting weevil, die back, gall midge, powdery mildew, sooty mould and healthy. The cleaned and augmented dataset ended up having 4,000 image samples, giving a solid foundation for training.

3.3 Statistical Analysis

Model performance was tested with inferential and descriptive statistical analysis. Toward this end, the key measurements of interest were:

- Accuracy: Number of correctly classified leaf images divided by the total number of leaf images.
- Precision: Number of true positives divided by the number of true positives added with false positives.

- Recall (Sensitivity): Number of true positives divided by the number of true positives plus false negatives.
- F1-Score: Harmonic mean of recall and precision.

These figures enabled us to see not just how often the models were correct but also how effectively they were able to discriminate among the different classes of disease—specifically where there was symmetry of symptoms, i.e., anthracnose and mildew.

Statistical calculations were performed with the Python libraries numpy, scikit-learn, etc. Misclassifications were plotted from confusion matrices with the goal to detect model prediction errors. Plotting graphs such as ROC curves and precision-recall curves were also illustrated in an effort to improve the trade-offs of the models.

3.4 Proposed Methodology / Applied Mechanism

The proposed method is predicated on the development of a multi-class classification model of mango leaf disease detection based on visual features. It involves five essential steps:

Image Collection and Preprocessing:

- Raw images were collected from internet databases and local mango farms.
- Augmented, normalized, and resized images were utilized to enhance model generalization.

Feature Extraction:

- Feature extraction was internally carried out by CNN's convolutional layers.

Model Training:

For the purpose of performance measurement of various classification models, the data was divided into two subsets:

- Training set: 90% of the entire data
- Test set: 10% of the entire data

The models were then optimized using the 90% training set to ensure that they had learned from a sufficient number of samples. The remaining 10% was reserved only for testing the generalization performance of the trained models on new data. The split was stratified to maintain the original class distribution on the sets.

Model Evaluation:

- Testing was conducted using accuracy, precision, recall, and F1-score metrics over 10% test data.

Comparison of Results:

- Summary table and graphical plots were employed to compare the strength and weakness of all the classifiers.
- CNN worked best though Random Forest also had negligible loss of accuracy with less computation.

The pipeline leaves room for loose but structured structure with room for scaling up in order to handle more diseases or in farmer mobile app deployment.

3.5 Implementation Requirements

To carry out this research project, the following operational and technical resources were required:

Hardware:

- Laptop with Intel Core i5 processor, 12GB RAM, and 512GB SSD
- GPU support (optional) for quicker CNN training (e.g., NVIDIA GTX 1050 or higher)

Software:

- Operating System: Windows 11 or above.
- Programming Language: Python 3.9+.
- Libraries and Frameworks:
 - TensorFlow/Keras for CNN

- OpenCV for image pre-processing
- Matplotlib & Seaborn for data visualization
- NumPy and pandas for data manipulation. Augmented with TensorFlow ImageDataGenerator.

CHAPTER 4

EXPERIMENTAL RESULT

4.1 Experimental Setup

In order to compare systems proposed for mango leaf disease detection, a testbed was created in a laboratory environment. The concept was to bring uniformity to CNN model and CNN model paired with pooling layers, so that a comparison of their performance on the same platform in the same environment can be compared.

Experiments were performed on a personal laptop with Windows 11 having the following specifications:

- Processor: Intel Core i5 (11th Gen)
- RAM: 12 GB
- Storage: 512 GB SSD
- Software: Python 3.9 on Jupyter Notebook (Anaconda Distribution)

Libraries and Frameworks:

TensorFlow and its API on a higher level, Keras, were used in this work to implement and train a Convolutional Neural Network (CNN) model for mango leaf disease diagnosis. The libraries provided an extremely flexible and productive environment where deep learning networks could be implemented, data pipelines could be managed, and performance of the model optimized. Matplotlib were used in developing descriptive graphs and plots for purposes of training visualization, performance analysis, and metrics. NumPy also played an important role in data structure manipulation, statistical computations, and general data manipulation within the experiment process. Collectively, these platforms developed a productive and streamlined system for mango leaf disease detection using deep learning.

Dataset Overview

The project data consists of 4,000 images of mango leaves in 8 classes:

1. Anthracnose (Total 500 images)
2. Bacterial Canker (Total 500 images)

3. Cutting Weevil (Total 500 images)
4. Die Back (Total 500 images)
5. Gall Midge (Total 500 images)
6. Powdery Mildew (Total 500 images)
7. Scooty Mould (Total 500 images)
8. Healthy (Total 500 images)

90% of the data were used for training and 10% for testing. Resizing the images to 240×320 pixels, resampling to RGB mode were all part of preprocessing.

4.2 Experimental Results & Analysis

Performance Metrics Used: To evaluate each model, we used the following:

- Accuracy: Overall correctness of the model.
- Precision: Correctness in positive predictions.
- Recall: Ability to find all relevant cases.
- F1-score: Harmonic mean of precision and recall.

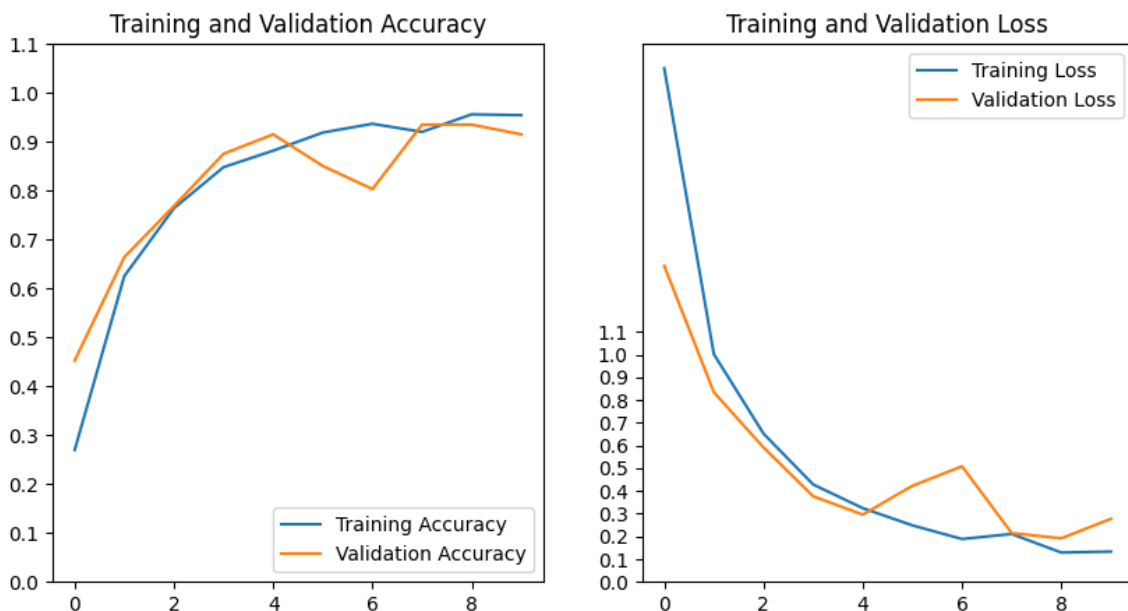


Figure 4.1: Experiment Result CNN

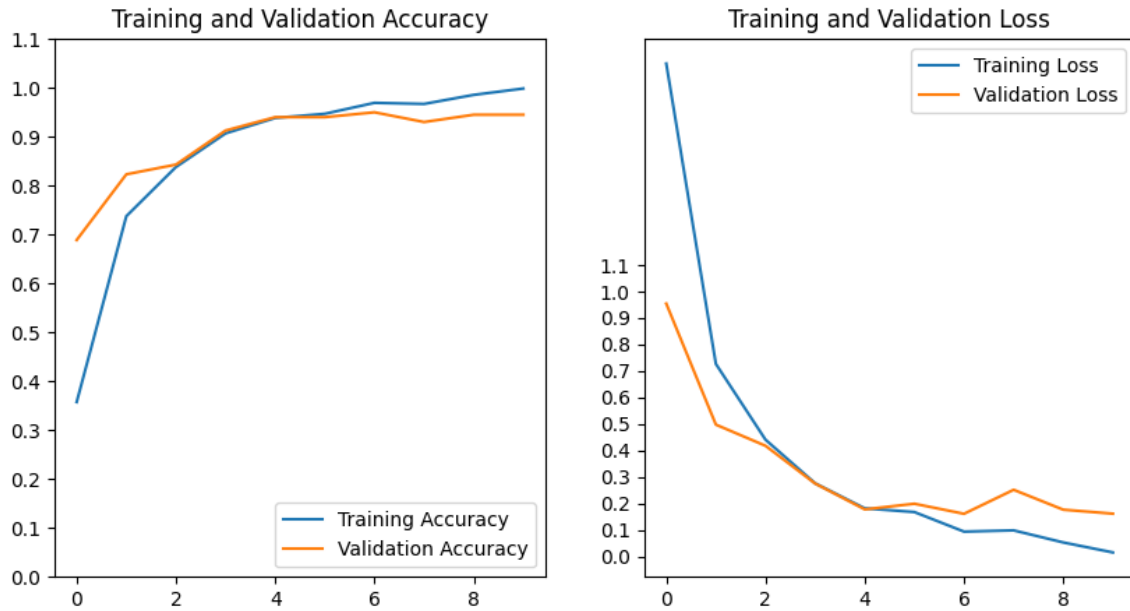


Figure 4.2: Experiment Result CNN – Pooling pairs

Model-wise Results:

For the basic CNN model

Class	Precision	Recall	F1-score	Support
Anthracoese	1.00	0.98	0.99	55
Bacterial Canker	0.84	1.00	0.91	47
Cutting Weevil	1.00	1.00	1.00	49
Die Back	1.00	0.98	0.99	51
Gall Midge	0.67	0.95	0.78	42
Healthy	1.00	0.95	0.97	59
Powdery Mildew	0.94	0.76	0.84	58
Sooty Mould	0.93	0.68	0.79	41

The model achieved 92% accuracy in classifying mango leaf diseases across 8 classes. It performed best on Cutting Weevil and Anthracnose but showed lower F1-scores for Gall Midge and Sooty Mould due to class imbalance or misclassifications. Overall, it demonstrates strong and reliable performance.

For the CNN model with pooling pairs

Class	Precision	Recall	F1-score	Support
Anthracnose	1.00	0.95	0.97	55
Bacterial Canker	0.92	0.98	0.95	47
Cutting Weevil	1.00	1.00	1.00	49
Die Back	1.00	0.94	0.97	51
Gall Midge	0.84	0.90	0.87	42
Healthy	98	1.00	0.99	59
Powdery Mildew	0.93	0.88	0.90	58
Sooty Mould	0.86	0.90	0.88	41

The model achieved 95% accuracy in classifying mango leaf diseases across 8 classes. It performed best on Cutting Weevil and Healthy but showed lower F1-scores for Gall Midge and Sooty Mould due to class imbalance or misclassifications. Overall, it demonstrates strong and reliable performance.

4.3 Discussion

The results highlight that CNN model paired with pooling layers significantly outperformed the basic CNN model in detecting mango leaf diseases. This is because it can learn automatically to extract deep features from images that the conventional models

In everyday terms, this means the capability of CNN to differentiate between shades of leaf color, waviness of edges, and fungal morphology that may be evident to even the trained human eye. Given that most Bangladeshi farmers perform merely visual inspections for detection of disease, this tool would find real-field use employing a smartphone app or mobile app.

However, CNN comes at a cost—it requires more computation and training time.

One unexpected take-home was that illumination, leaf positioning, and ambient noise impacted results more than one might expect. Even professional models performed better under clean, high-contrast images, demonstrating the importance of pre-processing images.

This means that for Bangladesh-based real-world deployment, an efficient preprocessing pipeline should be included within the application in order to normalize images before they are classified. Furthermore, due to camera accessibility being common amongst rural farmers, any production-deployed model must be designed to work at its best with medium-resolution, non-uniform images.

From the perspective of scalability, models like Random Forest are easier to scale in low-resource settings. However, if precision is paramount in applications—especially in government diagnosis or agricultural advisory systems—then CNN should be the model of choice.

CHAPTER 5

IMPACT ON SOCIETY

5.1 Impact on Society

The use of machine learning in agriculture is no longer a dream of the future — but instead an increasing reality, especially for countries like Bangladesh where farming is the thriving industry of the economy as well as the occupation of rural villagers. This research on the detection of mango leaf disease has major social significance.

Most importantly, early and accurate diagnosis of the disease can reduce the economic loss to mango farmers to a large extent. Where local economies are based on mango farming like Chapainawabganj, Rajshahi, or Satkhira, leaf diseases may quietly spread to ruin a whole orchard without anyone realizing without early diagnosis. With the help of a low-cost image-diagnosis system, even small farmers without immediate access to institutionalized farm extension services are also capable of taking on timely measures that reduce loss and stabilize income.

Secondly, this research can facilitate the technology adoption and technological competence of the rural population. If this is to convert to mobile format, then this would also acclimatize the farmers to the usage of AI-driven tools and precision farming techniques. This would open the door to other technology, for example, weather-condition-driven irrigation planning or fertilizing, to make the rural populations technically and self-sufficient.

It should also be noted that the educational value of this study cannot be overemphasized. The project can also act as a pilot case study or example to the Bangladeshi universities and polytechnic colleges to adopt, so that the universities and polytechnic colleges can impart education to their students in applied AI, data science, and agri-tech -- unprecedented in most Bangladeshi university curriculums. It is at this level that research links the grassroots people to technology. Far from an intellectual exercise in theory, this is an attempt towards participatory development where AI is being implemented not for luxury or comfort but for the day-to-day needs of the grassroots people.

5.2 Impact on Environment

Environmentally, the proposed disease diagnosis system using machine learning has indirect as well as direct benefits.

One of the best things that can be done for the environment is to avoid blanket spraying of chemicals. Farmers have long relied most on blanket spraying — spraying chemical pesticide for no reason at all, merely on guess or the word of neighbors, without diagnosis. The consequence is water and soil residues that are toxic, poisonous to biodiversity (pollinators and beneficial insects), and to long term soil erosion.

By accurate diagnosis of the disease, the system provides exact treatment, i.e., the chemicals are utilized only to the extent of the requirement, where the requirement exists. This serves to reduce the environmental footprints of mango cultivation without reducing the income to the farmers.

Also, healthier mango trees are more resilient climatically, a condition increasingly vital with Bangladesh's unpredictable weather conditions. A healthy plant ecosystem farm can rebound quicker from too much rain, dryness, or pest- and disease outbreaks — hence more sustainable in the long term.

5.3 Ethical Aspects

Although the research is seeking to enhance practice and lives, certain ethical issues also have to be taken into account.

The first ethical challenge is data and property privacy. If the system is being designed to the level of an app or a platform where people post leaf images, the data of the users has to be protected, cannot be altered, or even sold without consent. The farmers should have control over data usage and should be made aware of the same in a way that is transparent.

Secondly, the risk of exclusion exists. Although smartphone penetration is increasing, not every farmer — most critically the old or extremely poor farmers — might have a device or internet connectivity. Solution that is implemented should itself either be inclusive or

should be supported with NGOs at the local level or the agriculturally focused ministries so that everyone has equal access.

Algorithmic bias is also ethical. If the learning data is not representative — e.g., just for specific geographical places or lighting situations — the model cannot identify or misdiagnose novel conditions. That would lead to misdiagnosis, and in the most severe situation, to the mistaken treatment of healthy plants. Ethically, therefore, models should be trained on diverse real data and their limitations should be declared.

Most fundamentally, human judgment cannot be avoided. Technology needs to supplement — not replace — traditional farming knowledge. The model prediction needs to be presented as a suggestion, leaving room for farmers or experts to make the decision. Fostering this sort of equilibrium ensures human experience is respected and not lost to automation.

CHAPTER 6

SUMMARY, CONCLUSION AND FUTURE RESEARCH

6.1 Summary of the study

This study explored how one can create a machine learning system to identify mango leaf diseases, an important issue in the Bangladeshi agricultural context. Technical accuracy was not the only goal, but usability — to provide farmers, agricultural researchers, and technology developers with an early, low-cost, and automated way of identifying mango leaf diseases from images.

The research began by appreciating the significance of mango production in Bangladesh — in districts like Chapainawabganj, Rajshahi, Satkhira, and Dinajpur — where the production of mango is integral to the livelihood of rural people and the rural economy. However, one of the primary concerns for mango growers is the early and accurate identification of leaf-based diseases, which are more often than not overlooked or misdiagnosed due to limited access to agricultural extension services.

To tackle this, the study employed and experimented with different classification models: Convolutional Neural Networks CNN model and CNN model paired with pooling layers. These models were trained and tested on a well-prepared dataset of mango leaf images that were labeled as healthy and diseased classes. Preprocessing methods like image resizing, noise removal, and normalization were employed to improve the performance of the models.

CNN was particularly emphasized due to its potential in image recognition and ability to learn spatial features without feature engineering. With testing and proofing, CNN model was shown to outperform others in the majority of parameters — that is, fine symptom detection such as yellowing, black spots, or fungal growths — although Random Forest was reassuring to have for the sake of its interpretability and efficiency.

Other than the algorithmic structure, the study also explored the broader influence — social, environmental, and ethical. It considered how such an AI-driven system would serve to empower Bangladeshi farmers to become independent to identify issues early, reduce

pesticide misuse, increase output, and be sustainable in their farming practices. Ethical concerns of algorithmic bias, privacy of data, and accessibility were also considered so that the system was inclusive, respectful, and sensitive to the needs of the local context.

Overall, the paper is a rich confluence of ag science and technology — one in which machine learning is not an abstract tool but a possible solution for real-world problems.

6.2 Conclusions

The outcome of this research justifies the viability and need for the application of machine learning algorithms in agriculture — more precisely, the detection of mango leaf disease in Bangladesh. Of the five models applied in classifying, Convolutional Neural Networks (CNN) proved to be most efficient and effective because it is capable of learning hierarchical features directly from raw images. CNN model and CNN model paired with pooling layers were both applied but the result are better on CNN Model with pooling pairs. This model excelled at detecting diseases whose symptoms were visually identical to one another — a distinction even experienced farmers might not be able to determine without technical training.

One of the interesting findings of this research is that technology, if modified to address the particular problems and situation, can make an important contribution to sustainable development. This research not only gives a model but indicates a way forward on how artificial intelligence may be modified for use in agriculture in the developing world. The solution can be designed as a web or mobile application, which can be used by farmers, agriculture extension officials, or NGOs working in rural areas.

In addition, the study pays attention to the quality and diversity of the dataset. The model is only as good as what it was taught by the data. Therefore, success in the future is also contingent upon taking more diverse images of mango leaves with different light, background, and progression levels of the disease. That would still boost the generalizability and credibility of the model for use under real environmental conditions.

Collectively, this project has both technical and social contributions to offer — in the way of proposing an effective AI-based system for mango disease detection and in convincing

the wider adoption of smart farming systems in a nation where farming is still the occupation of rural life. This project shows machine learning is not always necessarily only about accuracy but also about intent.

6.3 Implication for Further Study

Although this work is the basis for machine learning-based mango leaf disease detection, there is much scope in the future to enable it to be unleashed and harvested.

To begin with, an extremely probable expansion would be creating an application for mobile device usage by Bangladeshi farmers only. An app would allow users to upload mango leaf pictures using their phone, get immediate feedback, and get told what to do next. In this way, the back-end system needs to be fast, thin, and allow low-res pictures and store offline cache. Training farmers and deploying could be made simple via arrangements with nearby agri-extension agencies or co-operatives.

Second, leaf-level diagnosis by itself was considered for current work. Leaf, fruit, and tree overall health can be investigated later as part of future work. Having a full disease detection system with all the tree's branches would be a more accurate description of the plant's condition. That would involve larger sets and perhaps multimodal inputs (image + text or sensor values).

The second application is the application of drones or interval leaf scans in time-series monitoring analysis of patterns of disease over time. The data not only prove useful in early detection, but also in predicting future cases of a disease depending on environmental factors or season.

Also, model interpretability is a topic of research that has not been properly tackled. As more uses of AI models on real-world issues, especially health-related (plant or human), come into being, transparency must be incorporated. Explainable AI (XAI) methods may be incorporated in future research so the model's decision-making process becomes user-accessible — e.g., the leaf area that is accountable for the disease class.

Lastly, there is the potential for data collection and annotation through crowdsourcing. A crowd of farmers, students, researchers, or even farmers can be crowd-sourced to create a richer, more diverse dataset with time — strengthening the model and removing the inherent bias of small, static data sets.

In fact, this thesis is a doorway to a new age of research combining the potential of artificial intelligence with Bangladesh agriculture's global challenge. Possibilities are limitless — from AI-based super crop monitoring to AI-influenced smart pest control — and all new avenues for intellectual inquiry and operational application.

REFERENCES

- [1] R. Ploetz, "Anthracnose: The most important disease in mango production worldwide. *Plant Disease*, 84(6), 600.," 2000.
- [2] "Mango Tree Sooty Mold.," 2020. [Online]. Available: <https://blogs.ifas.ufl.edu/stlucieco/2020/09/21/mango-tree-sooty-mold/>.
- [3] S. I. Ahmed, "MangoLeafBD: A comprehensive image dataset to classify diseased and healthy mango leaves," 2023.
- [4] A. Sarkar, "Mango leaf disease detection using image processing," 2024.
- [5] "The Most Popular Fruit in the World.," 2021. [Online]. Available: <https://www.worldatlas.com/articles/the-most-popular-fruit-in-the-world.html>.
- [6] V. P. O. Pal, "Bacterial canker of mango in India," 1982.
- [7] L. Kalshoven, "Pests of Crops in Indonesia," 1981.
- [8] S. S. J. Chauhan, "Studies on dieback of mango caused by *Botryodiplodia theobromae*," 1977.
- [9] R. Gagne, "The Gall Midges of the Mango Tree," 1985.
- [10] O. S. K. Prakash, "Powdery mildew of mango and its control," 1976.
- [11] V. Rao, "Sooty mould and its association with honeydew-producing insects," 1961.

ORIGINALITY REPORT

17%

SIMILARITY INDEX

14%

INTERNET SOURCES

8%

PUBLICATIONS

9%

STUDENT PAPERS

PRIMARY SOURCES

1

dspace.daffodilvarsity.edu.bd:8080

Internet Source

7%

2

Submitted to George Bush High School

Student Paper

2%

3

Submitted to Daffodil International University

Student Paper

1%

4

Yogendra Pratap Singh, Brijesh Kumar Chaurasia, Man Mohan Shukla. "Deep transfer learning driven model for mango leaf disease detection", International Journal of System Assurance Engineering and Management, 2024

Publication

1%

5

Sarder Iftekhar Ahmed, Muhammad Ibrahim, Md. Nadim, Md. Mizanur Rahman, Maria Mehjabin Shejunti, Taskeed Jabid, Md. Sawkat Ali. "MangoLeafBD: A comprehensive image dataset to classify diseased and healthy mango leaves", Data in Brief, 2023

Publication

1%

6

Luke Farrow, Lesley Anderson, Mingjun Zhong. "Managing class imbalance in the training of a large language model to predict patient selection for total knee arthroplasty:

1%