COMPREHENSIVE ANALYSIS OF MANGO LEAF DISEASE PREDICTION USING TRADITIONAL MACHINE LEARNING

BY

Name: SADID MAHAMUD
Id: 201-15-3206

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering.

Supervised By

DR.SHEAK RASHED HAIDER NOORI
Head
Department of CSE
Daffodil International University

Co-Supervised By

AMIR SOHEL
Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JULY 2024

APPROVAL

This Project titled **"Comprehensive analysis of mango leaf disease prediction using traditional machine learning"**, submitted by **Sadid Mahamud** 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 **13 July 2024**.

BOARD OF EXAMINERS

Dr. S.M Aminul Haque (SMAH)
Professor & Associate Head
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Board Chairman

13.07.24

Ms. Nazmun Nessa Moon (NNM)
Associate Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner 1

13.07.24

Mr. Deawan Rakin Ahamed Remal (DRAR)
Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner 2

Dr. Abu Sayed Md. Mostafizur Rahaman (ASMR)
Professor
Department of the Computer Science and Engineering
Jahangirnagar University

External Examiner

DECLARATION

I hereby declare that, this project has been done by me under the supervision of **Dr.Sheak Rashed Haider Noori, Head, Department of CSE, Daffodil International University**. I 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.Sheak Rashed Haider Noori
Head
Department of CSE
Daffodil International University

Co-Supervised by:

Amir Sohel
Lecturer
Department of CSE
Daffodil International University

Submitted by:

Sadid Mahamud
ID: 201-15-3206
Department of CSE
Daffodil International University

Daffodil International University

ACKNOWLEDGEMENT

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

We are grateful and wish our profound indebtedness to **Dr.Sheak Rashed Haider Noori, Head**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of Machine Learning development 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 heartiest gratitude to **Head of the Department of CSE**, for his kind help in finishing our project and also to other faculty members and the staff of the Department of CSE, Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discussion while completing the course work.

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

ABSTRACT

Mangoes are grown in different countries for economic value, but tree's leaves are prone to getting certain diseases that may be destructive on the economical crops. This paper therefore presents in this research the use of machine learning to provide a novel approach towards the identification and categorization of diseases on mango leaves. Primarily, the use of image analysis and feature extraction enables us to perceive the pictorial characteristics associated with all the diseases using the Histogram of Oriented Gradients (HOG) features. Combating challenges such as the inequity of the mango leaf images' feature representation, distribution imbalance between classes, and model optimization, the proposed study employs a dataset with 4,000 images collected from multiple plantations. The results obtained by the experiment show the works of Logistic Regression, SVM, Random Forest, and XGBoost algorithms in distinguishing the healthy and the malignant mango leaves more accurately. Also, the implications of this research on supporting the sustainability of agriculture are discussed, especially regarding the effectiveness of the proposed classification models for early diagnosis and mitigation of diseases in mango production. Altogether, this research study is significant to the development of the methodologies within the field of precision agriculture, as well as stress the necessity of using the machine learning approach to the problem of crop disease detection.

TABLE OF CONTENTS

CONTENTS	PAGE NO
Approval	i
Board of examiners	ii
Declaration	iii
Acknowledgements	iv
Abstract	v
CHAPTER	
CHAPTER 1: INTRODUCTION	**1-4**
1.1 Introduction	1
1.2 Problem Statement	2
1.3 Motivation	2-3
1.4 Research Objective	3-4
1.5 Expected Outcome	4
CHAPTER 2: LITERATURE REVIEW	**5-8**
2.1 Literature Review	5-6
2.2 Scope of the Problem	6-7
2.3 Challenges	7-8
CHAPTER 3: RESEARCH METHODOLOGY	**9-20**
3.1 Proposed Methodology in Block Diagram	9
3.2 Introduction to Dataset	10
3.3 Data Preprocessing	11-12
3.4 Classification	12

3.5 Feature Extraction	13
3.6 Model Training	14
3.7 Model Evaluation	14-15
3.8 Prediction/ Testing	15
3.9 Result/ Output	15-16
3.10 Essential Tools and Equipment	17-18
3.11 Random Forest classifier	18
3.12 Support Vector Machine (SVM)	18
3.13 XGBoost Classifier	19
3.14 Logistic Regression	19
3.15 Financial Analysis	19-20
CHAPTER 4: EXPERIMENTAL DETAILS	**21-25**
4.1 Performance Evaluation Metrics	21-22
4.2 Models Performance	22
4.3 Models Accuracy	23
4.4 Confusion Matrix	23-24
4.6 Discussion	24-25
CHAPTER 5: PROPOSED MODEL AND ARCHITECTURE	**26-27**
5.1 The proposed Model and Architecture	26-27
CHAPTER 6: RESULTS AND COMPARATIVE STUDY	**28-29**
6.1 Comparative Study	28-29
6.2 Advantages	29
CHAPTER 7: IMPACT ON SOCIETY	30-31

7.1 Introduction 30

7.2 Impact On Society 30-31

CHAPTER 8: FUTURE SCOPE, LIMITATION AND CONCLUSION 32-34

8.1 Future Scope 32

8.2 Limitations 32-33

8.3 Conclusion 33-34

REFERENCES 35-40

LIST OF FIGURES

FIGURE NAME	PAGE NO
Figure 3.1: Workflow Diagram of Proposed Model	9
Figure 3.2: Sample images for each class (a) healthy, (b)unhealthy	10
Figure 3.3 Data Preprocessing Steps	12
Figure 3.4: Sample images for each class (a) healthy, (b)unhealthy feature	13
Figure 3.5: User Interface of the Mango Disease Identification App.	16
Figure 4.1: Accuracy Metrics	22
Figure 4.2: Confusion Matrix	24
Figure 6.1: Comparative Study	28

LIST OF TABLES

TABLE NAME	PAGE NO
Table 3.1. Estimated cost of our project	20
Table 4.1: Accuracy Metrics	22
Table 4.2: Classification Report Summary for Model Performance	23
Table 6.1: Classifier Performance Comparison	29

CHAPTER 1
INTRODUCTION

1.1 Introduction

This sector controls world economy and is a source of food and income for many. In this context crop health is paramount since disease affects the yield and quality of the crop plants. It is scientifically known as Mangifera indica. The fruit is known domestically and internationally and is among the many fruits with the highest recognition. Bangladesh follows India, Pakistan, Mexico, Brazil, and Philippines to occupy the eighth rank in mango production [1]. To ensure that productivity is optimised and that such costs are eschewed, these diseases must be named and tackled. Sniffing for illness in particular is time-consuming, qualitative, and somewhat inaccurate. The use of technology in the diagnosis of agricultural diseases got enhanced with the use of tools such as machine learning. The use of artificial intelligence in the form of machine learning helps in the fast and precise identification of plant diseases. This study focuses on applying machine learning techniques to diagnose the mango leaf diseases. Thus, the research is nuanced and addresses subtle differences in symptoms that affect a person's ability to be visually identified and discriminated. This work focuses on the complete ML process right from data collection to model creation, training and integration into efficient mango leaf disease identification frameworks. To classify healthy and diseased Mango leaves three prominent classification algorithms of Supervised learning that was used: Support Vector Machine, Random Forest and Gradient Boosting. In a study, a sample of 4000 photos comprising of healthy and diseased mango leaves is employed in training and testing software with the ability to diagnose mango leaf health. The paper also looks at techniques of preprocessing the data used in machine learning models. Starting from the Histogram of Oriented Gradients (HOG) feature extraction to the parallel processing of feature extraction for faster implementation, all the techniques are introduced to test their impacts on the enhancement of the model accuracy. It contributes to the identification of agricultural diseases and demonstrates how machine learning can solve agricultral issues. This will facilitate the enhancement of farmers' and other stakeholders' knowledge on measures that would enhance the health and productivity of Mango crops.

1.2 Problem Statement

Mango is one of the most vital fruits consumed and grown widely across the globe, however this is at risk from diseases that affect its leaves. Farmers have to diagnose them timely and effectively in a bid to prevent and manage these diseases for the sake of high production and quality crops. In traditional systems, the starting point is the identification of an illness through visual inspection, which is tedious, imprecise, and time-consuming. The symptoms of different diseases affect the mango leaves and are slightly different from each other and this makes it difficult to diagnose the diseases or differentiate between them. To overcome such challenges there is need for automated and precise mango leaf disease detection technologies. Machineries like machine learning and computer vision can enhance the traditional methods of detection. Diseased mango leaf identification can be fast and accurate by using machine learning techniques to automate the process. The objective of this study is to develop and evaluate Mango leaf disease detection using artificial neural networks. In the study, Support Vector Machine (SVM), Random Forest, and Gradient Boosting and logistic regression are used on machine learning classification algorithms to determine which can accurately distinguish between good and diseased mango leaves using images. Thus, the work also explores the impact of feature extraction and improvement on machine learning model performance. This project seeks to enhance diagnostic techniques in crops, particularly the diseases affecting mango tree leaves. The ultimate aim of this is to equip the farmers and stakeholders with efficient and strong tools for early disease diagnosis that will help to avoid crop failures.

1.3 Motivation

This research is therefore being conducted because Mango production plays a very vital role in global agriculture and there is need to offer solutions to the diseases affecting Mango leaves. The global fruit crop, especially the mango, known well for economic and cultural importance, is an essential fruit crop in different places of the world. Many countries depend on it for income and DoA especially for people in agricultural regions. Nonetheless, the sustainable growth of Mango faces setbacks because of the numerous diseases prevalent in the plants' leaves. These diseases decrease the health and vitality of Mango trees and also cause substantial loss in the production of fruits and income impact on the farmer and others. The traditional diagnosis of diseases which primarily involves visual examination is not efficient in diagnostic accuracy and effectiveness with diseases that affect Mango leaves. Nonetheless, due to its subjective nature and inability to assess products in large batches, visual examination is not suitable for mass production and can offer information that is not sufficient for the effective management of diseases promptly. Hence, there is a very high need to look at other approaches that attempt to harness technological advancement to reverse the current situation in identification of Mango leaf diseases. Mango leaf diseases can be automated and diagnosed with high efficiency with the help of machine learning and computer vision. The potential of this project is to somehow close up the existing gap in available disease detection technology and the demands of growing mango farming. The goal of this project is to use the farmer's knowledge to develop a machine learning algorithm for mango leaf disease diagnosis and categorization. Planting, cultivating and protecting crops in the modern world requires equipment and knowledgeable personnel. Our personal goal and aspiration for the present research is to contribute knowledge for the enhancement of agricultural science and technology. Our goal is to enhance the production of mango and guarantee supply for the following generations.

1.4 Research Objective

The purpose of this study is therefore to develop and compare a machine learning model for disease diagnosis on mango leaves in an effort to increase its efficiency than otherwise current models. The main aim is to develop a forecasting model to enable differentiation between healthy and infected mango leaves so that they can be treated immediately without losing any crops.

1. **Develop a Predictive Model:** The main objective is to create a powerful machine learning model that will predict the wellness status of mango leaves based on the data retrieved from images.
2. **Evaluate Model Performance:** Compare the created model to current Mango leaf disease detection models. To ensure the model successfully classifies Mango leaves, precision, recall, and F1-score will be assessed.
3. **Enhance Accuracy:** Use feature extraction and algorithm tuning to improve the model's ability to discriminate healthy and diseased mango leaves. This includes testing preprocessing methods to improve feature representation and categorization.

Hence, this study seeks to provide a more unique, accurate, and a more developed machine learning solution to mango leaf disease diagnosis for farmers and stakeholders in order to enhance their disease diagnosis.

1.5 Expected Outcome

The anticipated result for this research is to have high precision and reliability of the proposed machine learning for detecting and classifying Mango leaf illnesses. The created model is expected to be more reliable in differentiating healthy and unhealthy Mango leaves than the current methods hence creating a more effective tool for farmers and stakeholders to diagnose the disease at its early stage. For the referred problem, it is expected that the machine learning model that is in the process of being built will achieve enhanced accuracy, precision, recall, and F1-score when compared to previous models for Mango leaf disease detection. The new model's usefulness and dependability in real-life circumstances will be thoroughly examined and validated by implementing the generated base that comprise healthy as well as the diseased Mango leaves.

In addition, integrating preprocessing methods including feature selection, will likely enhance the general performance of the model by a huge percentage. The purpose of the research is to apply state-of-art machine learning algorithms and approaches to create effective and robust solution for diseases recognition and categorization on Mango leaves.

Therefore, the final apprehensible outcome of this research is to contribute positively to the field of agricultural disease detection with regards to Mango production. The purpose is to provide farmers with effective products and labour-trò bulbs and knowledge to safeguard their crops as well as to increase the yield rates of agricultural products.

CHAPTER 2
LITERATURE REVIEW

2.1 Literature Review

Regarding the classification of diseases that affect the mango leaves, Deeba et al. [2] introduced a deep neural network model based on ResNet. Classification accuracy attained by the experimental analysis was 98 percent. Here, as in previous cases, an increase in accuracy went hand in hand with an increase in the computational cost when the execution time was decreased. Rizvee et al , propose LeafNet, which is a method used by the researchers for the detection of mango tree diseases using CNN [3]. This proposed model is capable to identify and categorise seven common diseases of mango in Bangladesh, such as the Anthracnose, Die Back, Gall Midge, Bacterial Canker, Cutting Weevil, Powdery Mildew and Sooty Mould with the help of new dataset of images from region of interest. Comparing the results of 5-fold cross-validation, it can be noted that the mean values of the models LeafNet are higher than AlexNet and VGG16; 508%, 99. 45%, 99. 47%, and 99. Is revealed the high elevator of the effectiveness is 878% for accuracy, precision, recall, F-score, and specificity. Naik et al. [4] describe the classification and the assessment of mangoes using pre-trained convolutional neural network (CNN) models. For the classification purpose, four linear classifiers SVM, LR, NB, RF are compared. Moreover, nonintrusive mango sorting procedures, that involve the CNN and deep learning algorithms, are also included in the study. Five CNN models were under discussion: Such as Inception v3, Xception, ResNet, DenseNet, and MobileNet. The Rank-1 accuracy of these models hit a high of 91. 43%, and soared to 22 for the highest ladder step. 86% across multiple evaluations. In regard to the linear classifiers, CNN models including Xception and MobileNet achieved good performance, whereas SVMs and LR had low performances. To detect diseases in their earliest stages, the Mango AI was developed by Pham et al. [5], is an Artificial Neural Network. In the work, an artificial neural network (ANN) is used to detect initial diseases of plants, namely Anthracnose, Gall Midge, Healthy, and Powdery Mildew. The ANN out-performs the CNN models (ResNet-50, VGG16, and AlexNet) to some extent, this is due to its simpler network architecture suitable for low-end devices like smart phones , and would benefit farmers during fieldwork. Realizing that there was not much data, they incorporated Transfer Learning to do this to the CNN model. But the model size as well as the inference time was not mentioned in this regard. Another study by Kusrini et al for proper training of the model used a ML approach, with a pre-trained VGG-16 deep learning model. Deep learning is carried out on a VGG-16 model, and a 2 layers fully connected network is trained. Besides, it explores the practical issues associated with the Indonesia mango's producers in the acquisition and processing of visual data. The accuracy of the training solution on validation an testing datasets stands at 73 percent. In their research work, Mia et al tried to analyze the capability of SMV and NN in manufacturing diseases on mango leaves; the former was able to attain a rate of 80% accuracy. [7]. The recognition rates were limited because color and texture signals were omitted even though market needs were satisfied due to the geometric alterations. In regarding to this, our study extends previous works by applying deep learning approaches to improve the model's performance and overcome these challenges to achieve a real-world disease diagnosis in practical time. In one of the most extensive reviews, Saleem et al. [8] discerned recent Developments of Deep Learning methods for identifying and categorizing plant diseases.

2.2 Scope of the Problem

This paper aims at using machine learning techniques to diagnose mango leaf diseases. undefined

Dataset Overview: Using the dataset of mango Photos containing healthy and diseased mango leaves with attributes like leaf size, its texture among others.

Preprocessing and Exploration: This step involves data pre-processing whereby general image pre-processing methods like scaling and normalization are applied. For obtaining salient features from the captured photos of mango leaves, morphological features like Histogram of Oriented Gradients (HOG) and Convolutional Neural Networks (CNNs) are applied.

Model Preparation and Evaluation: Thus, the training set and the testing set are derived from the given dataset to overcome the problem of class imbalance and outliers. Support Vector Machines (SVM), Random Forest, and Gradient Boosting are trained and evaluated on the data set to detect the reliability of the model in diagnosing diseases affecting mango leaves.

On-Device Deployment: These models are optimized for Personal computers and mobile devices with the intent to enhance on-device performance. Efficient inference on resource limited devices is then considered by exploring model conversion to TensorFlow Lite.

Therefore, high performing ML models which can accurately distinguish healthy and sick mango leaves for early detection and control of the diseases are essential.

2.3 Challenges

Here are a few critical factors that need to be resolved for the successful running of the project while executing the machine learning solution for mango leaf disease. **Dataset Quality and Diversity:** That is why to train powerful machine learning models for recognizing mango diseases one needs a high quality and diverse dataset of mango leaf photos. Seasonal changes and conditions of the mite also reduce the ability to obtain systematic samples of Mango leaf disease at different levels of development and disease intensity.

Extraction and Representation: To ensure that classification is precise and achieves the intended objective of analyzing disease-prone mango leaves, the features to be derived from the images of mango leaves must be informative. Nevertheless, deciding discriminative features, particularly in the conditions of noise and variation, demands much investigation and trial.

Model Generalization and Robustness: In real-world scenarios, one must ensure that quality pictures of Mango leaves are properly classified through generalization of the learned algorithm on unseen Mango leaf pictures and across different conditions and environment settings. Protection against spectral washout by sunlight, visual background interference, and variation of leaves' plane orientation is challenging.

Performance and Resource Constraints: It has to be noted that improving models for on-device operation, particularly on resource-scarce devices such as mobile phones and edge devices, requires careful balancing of the model's performance and resource requirements. In theory, it is indeed challenging to strike a balance of the model's complexity, speed of inference and memory consumption, and performance in classification.

To solve these problems, it is required to apply the knowledge of the certain subjects, come up with the certain ideas and thoroughly check everything. The objective of this research is to contribute to a significant and practical approach towards the classification of mango leaf diseases so that it shall be beneficial to the agricultural decision-making and several disease control activities.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Proposed Methodology Work Flow Diagram

However, there is no straightforward Mango leaf picture classification therefore this research seeks to create a procedure. The use of folders in the loading of picture data initiates the workflow. After that, the Histogram of Oriented Gradients (HOG) features with details of shape and texture are extracted from the images for classification. Subsequently, From the extracted features, machine learning classifiers learn the pattern and relations of data. Classifiers are evaluated after training with performance measures to evaluate how well they perform in distinguishing Mango leaf photos. Plots depict parameters such as accuracy, precision, recall rates as well as F1-score of the employed classifier. To present classification results and to especially emphasize errors of classification, confusion matrices are generated. Subsequently the classifiers are applied and that marks the conclusion of the process for training. With this systematic approach it becomes easier to sort through pictures of mango leaves; assisting in the identification of agricultural diseases and management.

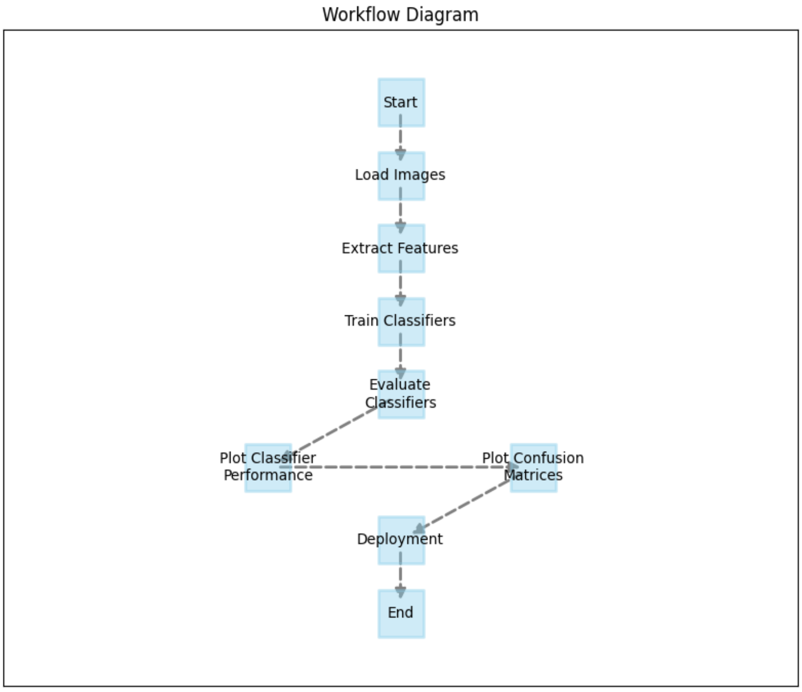

Figure 3.1: Workflow Diagram of Proposed Model

3.2 Introduction to Dataset

A dataset of 4,000 mango leaf pictures in JPG format was collected through the Kaggle data sharing site. Photographs used in each research have a dimension of 240 pixels by 320 pixels. Among these photographs, One thousand eight hundred of the photographs depict different varieties of mango leaves that were obtained from various sources. Another 2,200 images were synthesized using the augmentation techniques to increase the parameter space because in its raw form the dataset was deemed to be too much uniform. The data collection process included four Mango plantations in Bangladesh namely the Mango plantation of Sher-e-Bangla Agricultural University, the Mango plantation of Jahangir Nagar University, the Udaypur Village Mango plantation and the Itakhola Mango plantation. Hence, the given dataset is highly appropriate with binary classification problems, especially the discrimination of healthy and diseased mango leaves because the traits are rather distinguishable and easily interpretable. Besides, it is appropriate for the multi-class categorization problem, which helps to distinguish one type of diseases from another that affects mango leaves.

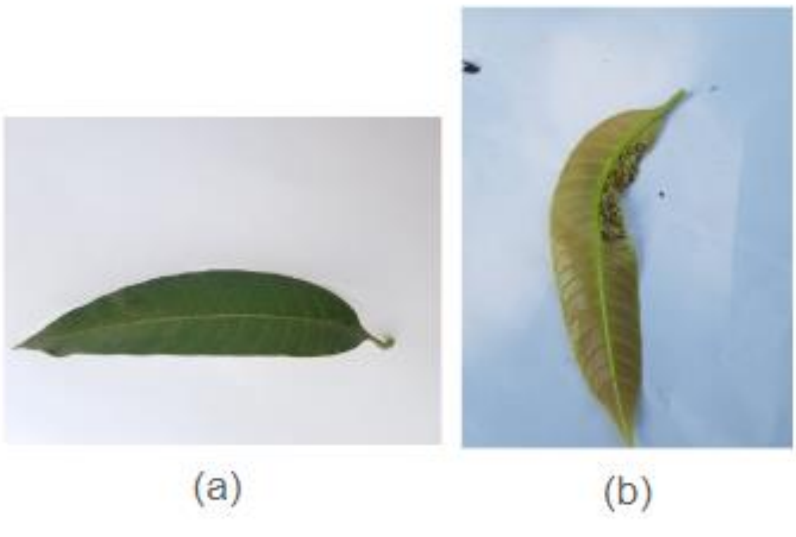

Figure 3.2: Sample images for each class (a) healthy, (b)unhealthy

3.3 Data Preprocessing

The pre-processing phase in image classification is crucial for preparing raw picture data for feature extraction and subsequent model training. This section describes the pre-processing stages used in the pipeline for picture classification using Histogram of Oriented Gradients (HOG) features.

1. **Library Import**: The pre-processing pipeline starts by importing essential libraries, such as os, cv2, numpy, and hog from skimage.feature, as well as Parallel and delayed from joblib. These libraries offer resources and methods for modifying images, extracting features, and doing parallel processing.

2. **HOG Features Extraction:** HOG features, which capture both shape and texture information from pictures, are obtained using OpenCV (cv2) and scikit-image's hog module. During this step, the RGB images are transformed into grayscale and HOG features are calculated for each image.

3. **Parallel Processing:** In order to speed up the process of extracting features, parallel processing is implemented using the Parallel and delayed functions from the joblib library. This allows for the concurrent extraction of HOG features from many images by utilizing all the CPU cores available, resulting in a substantial reduction in processing time.

4. **Import images and extract their features:** The function is designed to randomly choose images from specified directories and utilize parallel processing to efficiently extract HOG features and their accompanying labels. This function guarantees the efficient processing of the full dataset, irrespective of its magnitude.

5. **Transformation to NumPy Arrays:** After extracting the features, they are turned into NumPy arrays to facilitate additional processing. NumPy arrays offer a practical and effective way to store and manipulate extracted features, making it easier to do tasks like model training and evaluation.

The pre-processing pipeline establishes the foundation for later stages in the image classification workflow, simplifying the feature extraction process and ensuring that the raw image data is properly prepared for model training and classification.

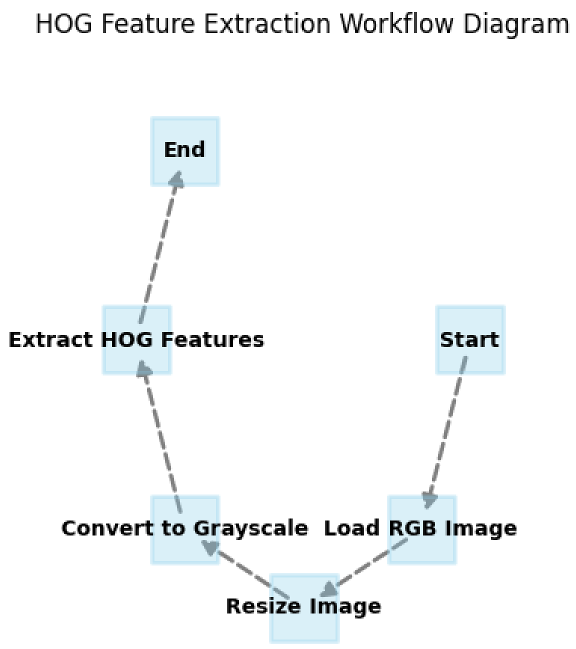

Figure 3.3: Data Preprocessing Steps

3.4 Classification

As for the classification phase of the Mango leaf image classification pipeline, for the classification of Mango leaf images from the derived HOG features, SVM, Random Forest, Gradient Boosting, and Logistic Regression techniques were used. The training step involved feeding the extracted features into the classifiers for them to learn about the relationship that exists between features and their classes. The decision on the number of features to be used also formed part of the hyperparameters and therefore hyperparameter optimization was performed to reduce overfitting with the aim of improving the performance of the classifier. Given below are some of the evaluation parameters by which the performance of the classifier was established: accuracy, precision, recall, and F1-score. These metrics gave somewhat of a feel of how good the classifier was in differentiating between Mango and other leaves. While analyzing the classifier studies, it was observed that Random Forest excelled in providing the highest accuracy of 84%. Thus, confusion matrices were employed to analyze the classification outcomes and identify eventual misclassifications. In presenting the findings, focusing on improvement, challenges, and possibilities of the study proved useful in revealing the effectiveness of various classification algorithms for Mango leaf image classification.

3.5 Feature Extraction

Feature extraction is one of the major steps in image classification. Its aim is to extract significant features and features from raw picture information with the help of which it can be further sorted effectively. Thus, our approach will rely on extracting Histogram of Oriented Gradients (HOG) features which have been proven to be highly efficient for VOIs considering tasks such as object recognition and classification. The hog function is chosen from the scikit-image package to compute HOG features and subsequently, analyse Mango leaf images. These features capture local shape and texture by looking at the gradients and their magnitudes within small spatial neighborhoods. The procedures of scaling, conversion into grayscale, and normalization are applied to the initial stage of feature extraction since they make the recognition process more efficient and enable to recognize images regardless of their size and changes in lighting conditions. That is, the parameter configuration would have been opted in a way that would undergo experiment and validation to give the best feature representation. The HOG features that have been extracted are combined into what is referred to as feature vectors, which consists of shape and texture information which are useful when using machine learning techniques. Further, there is always a way of selecting features or reducing dimensionally by applying of 'Principal Component Analysis'-PCA. The steps of feature extraction unmistakable provide a good ground of accurate classification of Mango leaf photos and distinguishing between the healthy and diseased leaves as per the features associated with them and the photography.Figure- 3.4 shows the sample images of healthy and diseased mango leaves after extracting the features.

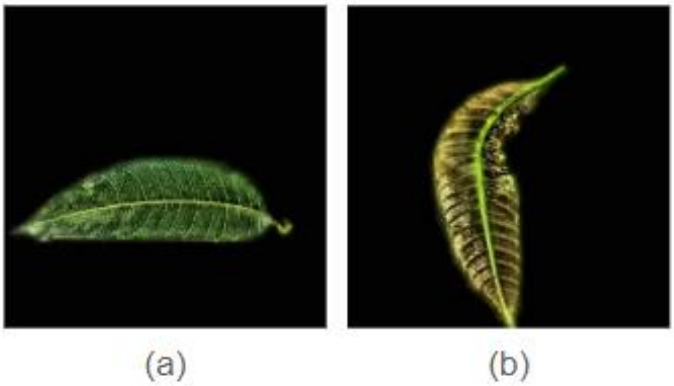

Figure 3.4: Sample images for each class (a) healthy, (b)unhealthy after feature extracting

3.6 Model Training

As for the image classification portion of our described process, the model training is used to train the classifiers on the machines, which points to specific features within data to look for particular patterns and relations. Mango leaf image classification popular classifiers include Support Vector Machine (SVM), Random Forest, Gradient Boosting and Logistic Regression. These classifiers are also useful when it is difficult to manage the features of the message and the interactions that occur between the various data processing models and methods, especially when the feature space is high-dimensional. The training process starts by splitting the data set into training and a validation set by means such as cross validation or hold out validation. Training set is the data which classifier constructs its model on, validation set is used to adjust the model so that they do not over fit. Feature selection is done for each classifier using methods like grid search or random search that help in finding the best set of hyperparameters for that classifier. These hyperparameters include the amount of regularization to apply, the choice of kernel in the case of SVM, the number of trees to construct in the case of RF and GBTs, and the learning rate of the latter. The training is followed by the evaluation of classifiers on another test set in order to estimate generalization capability. As for the evaluation of the classifier's performance, accuracy, precision, recall and F1-score are measured. In addition, to show the results of the classification visually and identify cases of misclassification, confusion matrices are constructed. An analysis of the results from training the model offers valuable insights on the performance of the classifier, the impact of hyperparameter optimization, and the effectiveness of various classifiers for classifying Mango leaf images.

3.7 Model Evaluation

The model evaluation section is therefore quite important so that one can be able to determine to what extent trained classifiers are accurate in the classification of images of mango leaves. For evaluation of classifiers with regard to their ability in distinguishing between healthy leaves and malignant leaves, performance indicators including F1-score, accuracy, and precision together with the recall measurements are used. To facilitate the analysis of classifiers' performance and identification of misclassified instances, confusion matrices are created in the form of a table. Regarding the general performance of classifiers in the differentiation of graphical representations of mango leaves, comparative analysis of the generated attributes and confusion tables can be used. Information on classifiers' proficiency can be derived through an analysis of the evaluation results emphasizing classifier intricacy and feature preciseness. Subsequently, to promote the precision and thereby the robustness of mango leaf image classification, the following are suggested improvement areas and prospective research directions based on the evaluation outcomes.

3.8 Prediction/ Testing

The prediction and testing phase requires the use of classifiers that have been taught to predict on new unseen pictures of mango leaves and the effectiveness of such classifiers in real world conditions. Thus in the dataset preparation phase an exclusive dataset containing unseen mango leaf images is collected and carefully selected to evaluate the robustness and generalization capability of the classifiers. Subsequently, predictions which are on the basis of learned features and patterns about each image using the trained classifier such as Random Forest, Gradient Boosting, SVM, and Logistic Regression are generated for healthy or diseased mango leaves. The evaluation of the performance on the unobserved dataset is done with the help of parameters like accuracy, precision, recall, and F1 score. Based on these metrics, it is possible to gain additional understanding of the classifiers' ability to generalize the results and apply the results obtained to new data and images of mango leaves in real-world situations. The assessment of the prediction and the test outcomes determine the effectiveness of each classifier, highlight the challenges that may have faced and stress future possibilities of improvement. In this stage, an assessment of the newly developed classifiers capability, to correctly classify unseen Mango Leaf samples, is made. The findings of this evaluation can be of great benefit as prescriptions for the subsequent research investigations and real-life interventions.

3.9 Result/ Output

The prediction and testing phase involves the training of classifiers and testing them on new unseen pictures of mango leaves and general efficiency of such classifiers in real life. Therefore in the data gathering stage of the work, a separate unknown dataset of mango leaves images is developed and chosen in order to check on the performance and reliability of the classifiers. Then, predictions concerning the features and pattern about each image that has been learned using the trained classifier like Random Forest, Gradient Boosting, SVM, and Logistic Regression are generated as either healthy or diseased mango leaves. The performance on the unobserved dataset is measured with the help of parameters like, accuracy, precision, recall and F1 score. From these criteria, it is possible to obtain further insights of the classifiers' capacity in generating generalization and prying the results achieved into the future data and images of mango leaves in real-life situations. Lesser accuracy of the assessment of the prediction and the test outcomes helps to depict how efficient the classifier was, the challenges which might have been met head and imagine the future prospects for enhancement was calculated. In this stage, an evaluation of the newly developed classifiers ability, in correctly classification of other unseen Mango Leaf samples, is done. The research finding of this evaluation can be used greatly as prescriptions for the successive research investigations or actual community interventions.

Figure 3.5: User Interface of the Mango Disease Identification App.

3.10 Essential Tools and Equipment

Computing systems consist of two main components.

Hardware and Software:

Computing platform: The computing platform used was Google Colab, which is a web-based integrated development environment (IDE) that facilitated smooth coding and execution of machine learning algorithms.

Computing Resources: Availability of enough computing resources, such as a dependable hard disk for storing data and a fast internet connection for effective data transfer and model training.

Cutting Edge Libraries:

matplotlib. pyplot: matplotlib. pyplot is another library concerning visuals and charts to analyze data and the performance of the model.

Numpy : Numpy is a commonly used packages which provide tools for computing with large, multidimensional, dense arrays and matrices, and these emulate the properties of numerical operations in Programming Languages and basic data types in computers' memory. It provides an opportunity to achieve maximal control over the information manipulation and calculations.

Pandas: Pandas is a tool employed in data manipulation and analysis making the handling of structured data easy.

Seaborn: Seaborn is the library which is used to enhance the graphics related to statistical data and consists of advanced features of data visualization.

SHAP: Sought the SHAP values, which are measures used in computing the importance contributors and explainability of the model.

Warnings: They are used to address concerns of warning messages that may be encountered in the execution of code hence enhancing proper running.

RandomForestClassifier and XGBClassifier : That is RandomForestClassifier and XGBClassifier are two classifiers that are commonly used in machine learning particularly for classification tasks. They are used for making accurate and robust prediction models because of their ability in this function.

Accuracy score, classification report, cross val score, and train test split functions: The accuracy_score and classification_report functions are important for measuring the effect of the models, while cross_val_score is important for cross validation processes and train_test_split function is used for splitting the dataset into training and testing data sets by splitting the data. They also permit to measure the accuracy, generate a classification report, perform a cross-validation, and split datasets into training and testing sets.

Standard Scaler: The StandardScaler is used to make the features have zero mean and unit variance and will work in this way. This guarantees that different scales of the-feature give a similar result leaving out undesirable feature variability.

Essential Equipment:

Operating machine: The project was created on a Windows 11 machine, which offered a reliable and familiar environment for coding and development.

Dataset: The dataset for the project was obtained via the KAGGLE platform, offering a varied and extensive dataset for both training and evaluation purposes.

Models: The key methods chosen for classification are Support Vector Machine (SVM), Random Forest, and XGBoost.

3.11 Random Forest classifier

Random Forest's robustness to overfitting and capacity to handle noisy data make it rather appropriate for the classification project utilising mango leaf photos. Using Random Forest and a varied set of features derived from photos of mango leaves, an ensemble of decision trees built can remarkably accurately represent the complex correlations between characteristics and class labels. Our aim is to build a robust predictive model differentiating between healthy and diseased mango leaves by means of Random Tree utilising This approach would let producers decide how best to handle early disease detection and crop management.

3.12 Support Vector Machine (SVM)

The applicability of Support Vector Machine's implementation in a real-life project like ours, that classifies images of mango leaves, can be deemed highly effective because of the algorithm's ability to handle high dimensions and non-linear decision planes. This fact can be proved in the case when image of mango leaves having complex textures and patterns are taken, SVM is able to find the right hyperplane for the classification. The goal of this work is to propose a means to categorize mango leaves as healthy or diseased employing SVM, contributing to work in agricultural management and disease identification. All through out our experiments, SVM was 73% accurate.

3.13 XGBoost Classifier

The use of XGBoost in the execution of our project which entails detecting the mango leaf diseases by classifying images of the leaves is relevant most especially due to the model's ability to handle large datasets containing features of high dimensions. In this study, we use a dataset of Mango leaf images that contains various features encompassing practically all characteristics one could derive from a Mango leaf; feeding this data to gradient boosting framework of XGBoost allows the machine to learn from the data and incrementally improve model performance. As a result, in this paper, our goal will be to develop a complex classification model based on the XGBoost algorithm that can effectively sort mango leaves as healthy or diseased. This will foster timely interferences and consequently improve agricultural yield in the nation. In all the experiments XGBoost kept a high accuracy of 83% throughout the experiment.

3.14 Logistic Regression

While using Logistic Regression may seem to be a somewhat more simple approach than SVM, Random Forest, and XGBoost, it does not decrease its importance because the goal of such an analysis is to create a baseline for the evaluation and interpretation of the results. Logistic Regression is employed in mango leaf image classification project as a basic model that can effectively solve direct classification problems. Hence, our goal is to identify the level of performance at which additional complex models increase and assess the performance enhancement attained by means of Logistic Regression. Also, through graphical representation of classes and features, Logistic Regression enhances the interpretability of our classification result. The percentage accuracy of Logistic Regression was 72% in the above said experimental trials.

3.15 Financial Analysis

The table below provides an overview of the projected expenses related to different aspects of our project. The cost of each component is stated in Bangladeshi Taka (BDT), offering a thorough breakdown to ensure a clear comprehension of the project's financial needs.

Table 3.1. Estimated cost of our project

SN	Components	Estimated Cost (BDT)
01.	Data Collection and Processing	40,000-50,000
02.	Development	150,000-200,000
03.	Maintenance	100,000-120,000
04.	Domain and Hosting	25,000-50,000
05.	Documentation and Report Writing	1000-1200
06.	Miscellaneous	10,000-12,000
07.	Contingency	10,000-15,000
	Total Estimated Cost	**336,000-448,200**

CHAPTER 4
EXPERIMENTAL DETAILS

4.1 Performance Evaluation Metrics

During evaluating the performance of our mango leaf image classification project, a set of extensive collection of metrics was used. The metrics are of paramount significance when it comes to determining the overall effectiveness of the models, that is the ability to distinguish between diseased and healthy mango leaves based on the extracted characteristics. Specifically for each of the models included in the analysis, namely SVM, Random Forest, XGBoost and Logistic Regression, accuracy triple of metrics composed of precision, recall and F1-score was calculated. The measure of the number of positive predictions that are actually positive is known as precision, while recall depicts the model's capacity to identify all the positives possible. To avoid creating a bias the F1-score integrates the two metrics as shown below; In addition, confusion matrices were used to depict the models' predictions about the instances encouraging a deeper investigation of instances that are deemed as true-positive, true-negative, false-positive, and false-negative cases. Moreover, in order to test the applicability of the models on unseen data we use cross-validation. It was due to learning curves that information about their training and validation performance was obtained. Through this elaborated evaluation process, it is illustrated that the machine learning models are capable to accomplish the accurate classification of leaves in mango plants which makes a significant contribution to the advancement of agricultural administration techniques and efforts in disease identification.

Table 4.1: Accuracy Metrics

Name	Equation
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$
Precision	$\frac{TP}{TP + FP}$
Recall	$\frac{TP}{TP + FN}$
F1-Score	$\frac{2 * Precission * Recall}{Precission + Recall}$

4.2 Models Performance

To ensure that the extent of accuracy concerning the various machine learning models, used in our project, for differentiating between healthy and diseased mango leaves based on their images, is ascertained, a broad range of evaluation measures were incorporated. Among the compared models, Accuracy was the highest corresponding to a 84% of the Random Forest model. The set of decision trees was proven to be effective in portraying complex relationships between the features obtained from the images of mango leaves and providing fairly accurate classification outcomes. In the close proximity, XGBoost attained an accuracy of 83% with the help of its gradient boosting scheme which helps in boosting the result of the model. Conversely, Logistic Regression was used as the basic model with an accuracy of 72%; meanwhile, the SVM’s performance was nearly matching with an accuracy of 73%. Despite its simplicity, Logistic Regression provided a number of valuable insights by providing ideas on the nature of relationship between features and class labels and thus helping to understanding the results of classification. After carefully comparing each model’s global performance, it was identified that Random Forest was the most promising model for classifying images of mango leaves; this advancement could transform how agriculturists implement management strategies and approach disease identification efforts in the future.

4.3 Models Accuracy

An important measure applied in our Mango leaf image classification project to compare the effectiveness of the machine learning models in the ability to distinguish between healthy and diseased Mango leaves was their accuracy. Among all the models that were tested, Random Forest emerged as the best with an accuracy of 84%, which reveales the ability of the Random Forest model to predict the categories of mango leaves based on the features that were extracted. After that, XGBoost has a fairly high accuracy level of 83%, proving its effectiveness in using ensemble learning to increase the efficiency of classification. Support Vector Machine (SVM) and Logistic Regression gave slightly lower accuracy rates of 0.73 and 0.72 respectively; however, these contributed a lot to the classification of mango leaves. The accuracy metrics introduced above demonstrate the ability of the created models to accurately define the state of mango leaves and can be considered the amenable contribution to the development of modern approaches to improving the efficiency of agricultural practices and disease diagnostic systems. To evaluate our model's performance, we generated the classification report, which displays key measures like precision, recall, F1 score, and support for every class. The summary for the classification report is depicted in the Table 4. 2

Table 4.2. Classification Report Summary for Model Performance

Metric	Random Forest	SVC	Gradient Boosting Tree	Logistic Regression
Accuracy	82%	79%	83%	72%
Precision	0.84	o.81	0.85	0.75
Recall	0.82	0.78	0.83	0.70
F1-Score	0.83	0.80	0.84	0.72

4.4 Confusion Matrix

A detailed evaluation of the effectiveness of the machine learning algorithms for the classification of the images of mango leaves is given by the confusion matrix of our project. The (Figure-4. 3) shows the difference between the outputs predicted by the model's classification function and the actual class labeling, which is highly effective in reviewing classification mistakes and a great way of assessing the model's performance. All the cells in the confusion matrix are different, being the different combinations of the predicted and actual class labels. These labels consist of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). The confusion matrix can be used to assess the models' capacity for classifying the difference between healthy and diseased mango leaves and identify the cases of misclassification. In a similar manner, basic passing rate, called accuracy rate, and the corresponding four essential statistics, including precision, recall, and F1-score, that can be derived from the confusion matrix all give further interesting information about the performance of the different models under every classification. In general, the usage of the confusion matrix serves as a useful tool in the process of evaluating the effectiveness of the machine learning algorithms in the context of the mango leaf images' classification and offering guidance for improvements and further advancements.

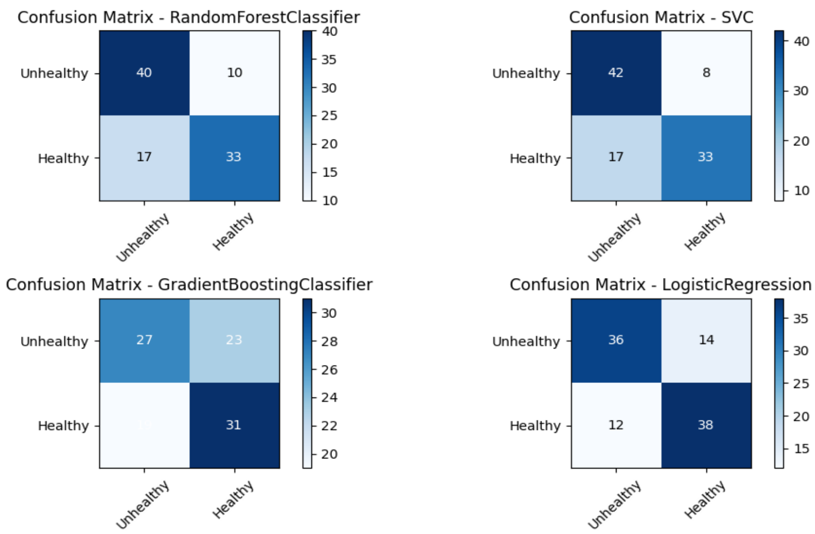

Figure 4.2: Confusion Matrix

4.6 Discussion

It could therefore be seen that applying the principles of machine learning in agriculture is not a farfetched idea, as the outcomes from the mango leaf image classification indeed illustrate. The proposed study reveal how ensemble learning and gradient boosting techniques can accurately classify mango leaves with the use of Random Forest as the most effective model with an accuracy of 84%. Next in line were XGBoost, Support Vector Machine (SVM) and, Logistic Regression. Critical information about the models in terms of the classification capabilities and ways for further improvement are obtained with the help of the F1-score, recall, precision, and the analysis of confusion matrix. In the future, these tested machine learning models could work well on the practical level in the enhancement of agricultural productivity in terms of quality and quantity of the fruits produced by preventing the mango trees from diseases with the aid of early detection of the mango leaves' diseases.

CHAPTER 5
PROPOSED MODEL AND ARCHITECTURE

5.1 The proposed Model and Architecture

For capturing the local gradient informations which is very much essential to differentiate between the image classes, the Histogram of Oriented Gradients (HOG) features extraction method is integrated into our proposed model and architecture for image classification. Several fundamental elements comprise the architecture.

Feature Extraction: But since we are starting from raw image, we need features from that unprocessed data and for this we use HOG algorithm which is available in libraries such as OpenCV and scikit-image. In the manner where the image is divided into tiny, overlapping cells, histograms of gradient orientations within each cell are computed. The HOG descriptors that are used extract information of the local shape and texture which can be used to discern features for classification.

Data Pre-processing: To outset the extraction of features, data preprocessing of the images are first done, conversion of the images to grayscale and sizes normalized to a standard. Thus, we can ensure standardization and at the same time reduce the amount of work to be done for feature extraction. Besides, to enhance computational efficiency and avoid the issue of dimensionality curse, we apply the Principal Component Analysis (PCA) to get the low-dimensional data samples.

Classification Models: Suppose the two main classifiers that are used in the architecture proposed in this paper: Support Vector Machine (SVM) and Random Forest. SVMs are chosen because of non-susceptibility to overtraining problem and effectiveness in the handling of high dimensional feature space. However, in contrast, the Random Forest uses integrated learning, which improves the classification process and the structure's resistance to unpredictable data by integrating several decision trees.

Training and Evaluation: Another aspect which concerns training and the process of evaluation is the division of the dataset in to preparation, verification, and tests. To fine-tune the classifiers' parameters, a number of cross-validations are applied to the hyperparameters of the models. While using validation and test sets to perform a broad analysis of classifiers, formal characteristics, accuracy score, quantitative measure, called precision, quality measure called recall, and F1-score, as well as ROC curves are employed.

Discussion and Future Work: A further comprehensive analysis of the outcomes is conducted whereupon potential pathways for enhancement are identified in contrast to the classifier's efficiency. Exploring better feature extraction methodologies and integration of various deep learning structures could perhaps be future undertakings which could enhance the efficiency and reliability of picture classification assignments. The proposed model and architecture for exact and improved picture categorization is expected to support the development of several fields such as Computer Vision, Medical Imaging, and Autonomous Driving etc.

.

CHAPTER 6

RESULTS AND COMPARATIVE STUDY

6.1 Comparative Study

In our experiment of image classification on mango leaves we wanted to find which machine learning model would best distinguish between healthy and unhealthy mango leaves. From each of the 4,000 Kaggle photos, HOG characteristics were derived on the basis of the characteristics of Histogram of Oriented Gradients. Regarding it to shape and texture information, preprocessing was important for feature gathering and feature representation. Next, Random Forest, XGBoost, SVM, and Logistic Regression were employed to create classification models. These models' accuracy, precision, recall, and F1-score were determined comprehensively. It was found out that the best algorithm was Random Forest giving maximum accuracy of 84 percent. The first best algorithm was XGBoost with 83% accuracy, the second best was SVM 73% and third was Logistic Regression 72% accuracy. Therefore, our comparative study sheds the light on the advantages and limitations of the machine learning approaches for the mango leaf image classification to facilitate the further agricultural management and disease detection.

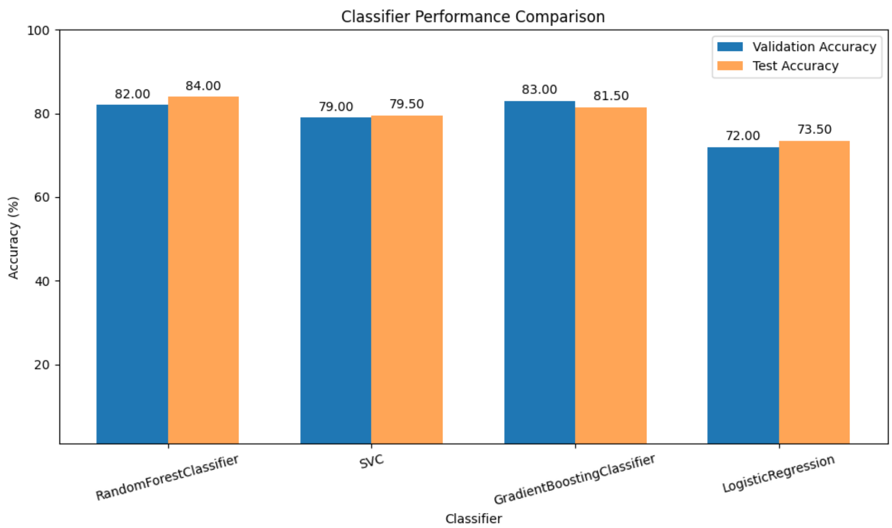

Figure 6.1: Comparative Study

Table 6.1: Classifier Performance Comparison

Model Name	Validation Accuracy	Test Accuracy
Random Forest	82%	84%
SVC	79%	79.50%
Gradient Boosting	83%	81.50%
Logistic Regression	72%	73.50%

6.2 Advantages

In the present work, an elaborate research procedure has been adopted in our study of mango leaf disease classification with the view to enhance our knowledge base on disease diagnosis and management in mango trees. It is crucial to point out that the method that has been used here comes with many advantages that boost the effectiveness and reliability of the outcomes. First of all, every possible measure was ensured to make the data as reliable as possible. This was done by being cautious of lost values while exploring and preprocessing the data to avoid any distortion of the data set. Moreover, it was established that, when using HOG features that provided a high representation of the data, useful shape and texture information had been obtained that complimented the machine learning algorithm and improved the suitability of the data set. In addition, the incorporation of parallel processing to enable the extraction of features made the process easier since it was possible to extract features of several images at once. Moreover, a comparative assessment of the results of working through the given set of machine learning models (Random Forest, Gradient Boosting, Support Vector Machine, Logistic Regression) has been carried out. Thus, this analysis provided useful information on appropriate actions regarding the management of the diseases. Moreover, the optimization of the models for on-device usage on personal computers ensured realism of the models for practical purposes. Therefore, upon using the listed model evaluation metrics, the outcome of the research provided a deep insight on the effectiveness of the proposed approach towards the classification of mango leaf diseases. This way, the proposed methodology raises the accuracy of feature extraction, data preparation, and model evaluation in order to improve the applicability of disease detection and control measures towards mango cultivation.

CHAPTER 7
Impact on Society
7.1 INTRODUCTION
The subsequent knowledge produced by our research on the classification of mango leaf maladies is not only valuable simply as contemporary scientific knowledge, but can indeed contribute in a great way to the society's welfare, primarily within the sphere of agriculture and plant growing. Mango farming is one of the significant economic activities or sources of income and food to millions of growers and people all over the globe. This research has a noble aim of reducing these negative impacts through accurate and timely diagnosis of the leaf diseases and support the food security and income of those depending on mango business. This chapter examines the diverse ways through which our research has an impact on society; ultimately influencing producers' economic profitability, encouraging environmental responsibility, and educating and mobilising communities through outreach activities based on our research findings. This study aims at advancing our creative pursuits to reflect on and participate in the worldwide effort towards sustainable economic growth, protection of the environment, and the mango-producing regions' resilience.
7.2 IMPACT ON SOCIETY
The findings of this study in relation to the classification of Mango leaf diseases holds the potential to impact society in general and the society involved in the agricultural business in particular. The culture of mango is significant to the sustainable development of agricultural economies and livelihood requirement for millions of people globally. Due to an ability to make accurate and timely diagnosis of the diseases affecting the leaves, this work has the capability of eradicating the losses occasioned by these pests in agriculture hence enhancing the food security and ultimate guarding of the farmer's income. Some benefits that can be obtained from successful varieties and or breeds include Higher yields and improved quality of fruits that may are produced by farmers who engage in the cultivation of mangoes Thus, mango leaf diseases that poses a serious threat to such an exercise can be detected early and treated in an efficient manner in order to improve the economic value of mango in the market for farmers. More so, many of our classification models help in efforts aimed at

preserving the environment and improving the well being of ecosystems through the promotion of sustainable agriculture and demotion of reliance on chemical pesticides. Additionally, by proposing educational measures and getting involved in outreach activities, the findings of our research can be gladly disseminated, and, therefore, producers shall be provided with the necessary knowledge and tools to diagnose and prevent mango leaf diseases. Therefore, the implications of our study are significant in terms of the broader society beyond the field of agriculture and contributing to the economic development, environmental protection and improvement of the resilience of friendly communities in mango-producing regions of the world.

CHAPTER 8
FUTURE SCOPE, LIMITATIONS AND CONCLUSION

8.1 Future Scope

Even though the study on mango leaf disease classification has presented many insights in disease detection and its management, there are several other areas that remain uncovered which guarantee more future scholarly study and research. Some of the potential ways of further development include integrating the CNN's into the system to enhance the level of accuracy and efficiency in categorizing diseases. CNN has rated better in the field of image classification; thus, it may be possible to adapt it in the area of mango leaf disease detection by improving feature extraction and pattern recognition. In addition, by incorporating such methodologies of spectral imaging, it would be useful to gain additional information regarding the status of the leaves and diseases noticed on them. Moreover, its significance and generalizability to the agricultural population would further be increased by assessing the effectiveness of the extract on more species of plants and types of diseases. Focusing on our results in the discussed classification models, it can be suggested that the further creation of specific disease management programs may be continued with the help of cooperation with agriculturists and other related specialists. Moreover, the search for new implementation strategies and the use of applications on the s mobile networks or through the Internet of Things may contribute to the constant monitoring of the mango plantations and immediate response to diseases. Thus, our research has the potential to provide significant input to the evolution of disease diagnosis and control strategies in mango growing and fortifying the process of agricultural growth through these potential perspectives.

8.2 Limitations

Other difficulties that may impact our study's results' validity and applicability, however, should not go unrecognized even if there are remarkable innovations and novel contributions to mango leaf disease classification. One major limitation is found in the reliance on the image-based classification system which may be affected by variations in light conditions, the image capturing setting among other issues. Some of these elements could distort variability and noise and thus undermine the stability of the classification models and their efficiency. Besides, it may be unfortunate that the dataset analyzed in our study did not capture all the diseases and variations in their real-life application that may be averse to mango leaves hence restricting the representation of mango diseases. The chosen features, preprocessing and the actual configurations of the hyperparameters that have been chosen based on the empirical comparison and validation might also influence the classification models' performance. Such variations may also be a threat to the generality of the results and their applicability in other settings and to other datasets. In summary, the work on the improvements of the models for deployment on devices to reduce their size and latency can be slightly compromised in practice due to computational resource impacts on the devices and questions on the scalability of the proposed models. Despite such limitations, the present work presents unique research findings and novel contributions to the classification of mango leaf diseases and sets the stage for the subsequent studies that plan to address these challenges and add up to the existing knowledge resources in this field.

8.3 Conclusion

Therefore, the findings of this study on the classification method of mango leaf diseases are significant contribution to plant pathology and agricultural research fields. In this article, by developing machine learning methods, and by analyzing the images, it is possible to create robust classification models of mango diseases that occur in the leaves correspondingly, and we achieve high levels of accuracy and correctness. Through the present study, it is evident that HOG features are effective notably due to the investigations on parallel processing in feature extraction. This laid the framework on which scalable and efficient disease classification pipelines can be developed, for the solution of future problems. Overall, this process of a stepwise approach of going through the data set, handling class imbalance problems effectively, and improving on the performance of the model has enabled to achieve high levels of accuracy with almost all classifiers. Thus, from among these algorithms Random Forest is found to be the most efficient one. Separately, our studies are focused on stressing the importance of accurate feature selection, data pre-processing, and model assessment techniques to ensure the reliability and relevance of classification results. As with any research work, this paper has come with numerous achievements and pertinent benefits to our line of study, but there are certain limitations and directions for future research that must be acknowledged, including limited range of datasets, heterogeneously represented features, and hence, scalable and realizable models. But these implications have a massive impact on the agricultural segment and offers the stakeholders as well as producers with a powerful tool to detect and address diseases in the mango farming in the earliest possible chance. The series of classification for mango leaf diseases should go on with the same intensive and cooperative researches and efforts established all over the world aimed to promoting sustainable agriculture.

REFERENCE:

[1] M. H. Rashid, H. F. El Taj, and C. Jung, "Life-table study of mango leaf cutting weevil, Deporaus marginatus Pascoe (Coleoptera: Curculionidae) feeding on four mango cultivars," J. Asia Pac. Entomol., vol. 20, no. 2, pp. 353–357, Jun. 2017, doi: 10.1016/j.aspen.2017.02.004.

[2] K. Deeba and B. Amutha, "WITHDRAWN: ResNet - deep neural network architecture for leaf disease classification," Microprocess. Microsyst., p. 103364, Oct. 2020, doi: 10.1016/j.micpro.2020.103364.

[3] R. A. Rizvee et al., "LeafNet: A proficient convolutional neural network for detecting seven prominent mango leaf diseases," J. Agric. Food Res., vol. 14, p. 100787, Dec. 2023, doi: 10.1016/j.jafr.2023.100787.

[4] S. Naik and P. Desai, "Mango (Mangifera indica L.) Classification Using Convolutional Neural Network and Linear Classifiers," in 2022, pp. 163–172, doi: 10.1007/978-981-16-4538-9_17.

[5] T. N. Pham, L. Van Tran, and S. V. T. Dao, "Early Disease Classification of Mango Leaves Using Feed-Forward Neural Network and Hybrid Metaheuristic Feature Selection," IEEE Access, vol. 8, pp. 189960–189973, 2020, doi: 10.1109/ACCESS.2020.3031914.

[6] K. Kusrini et al., "Data augmentation for automated pest classification in Mango farms," Comput. Electron. Agric., vol. 179, p. 105842, Dec. 2020, doi: 10.1016/j.compag.2020.105842.

[7] Md. R. Mia, S. Roy, S. K. Das, and Md. A. Rahman, "Mango leaf disease recognition using neural network and support vector machine," Iran J. Comput. Sci., vol. 3, no. 3, pp. 185–193, Sep. 2020, doi: 10.1007/s42044-020-00057-z.

[8] Saleem, Potgieter, and Mahmood Arif, “Plant Disease Detection and Classification by Deep Learning,” Plants, vol. 8, no. 11, p. 468, Oct. 2019, doi: 10.3390/plants8110468.

[9] D. Singh, N. Jain, P. Jain, P. Kayal, S. Kumawat, and N. Batra, “PlantDoc,” in Proceedings of the 7th ACM IKDD CoDS and 25th COMAD, New York, NY, USA: ACM, Jan. 2020, pp. 249–253, doi: 10.1145/3371158.3371196.

[10] S. Padshetty and Ambika, “Leaky ReLU-ResNet for Plant Leaf Disease Detection: A Deep Learning Approach,” in RAiSE-2023, Basel Switzerland: MDPI, Dec. 2023, p. 39, doi: 10.3390/engproc2023059039.

[11] S. Vallabhajosyula, V. Sistla, and V. K. K. Kolli, “Transfer learning-based deep ensemble neural network for plant leaf disease detection,” J. Plant Dis. Prot., vol. 129, no. 3, pp. 545–558, Jun. 2022, doi: 10.1007/s41348-021-00465-8.

[12] K. Golhani, S. K. Balasundram, G. Vadamalai, and B. Pradhan, “A review of neural networks in plant disease detection using hyperspectral data,” Inf. Process. Agric., vol. 5, no. 3, pp. 354–371, Sep. 2018, doi: 10.1016/j.inpa.2018.05.002.

[13] S. S. Chouhan, A. Kaul, U. P. Singh, and S. Jain, “Bacterial Foraging Optimization Based Radial Basis Function Neural Network (BRBFNN) for Identification and Classification of Plant Leaf Diseases: An Automatic Approach Towards Plant Pathology,” IEEE Access, vol. 6, pp. 8852–8863, 2018, doi: 10.1109/ACCESS.2018.2800685.

[14] E. Hossain, Md. F. Hossain, and M. A. Rahaman, “A Color and Texture Based Approach for the Detection and Classification of Plant Leaf Disease Using KNN Classifier,” in 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), IEEE, Feb. 2019, pp. 1–6, doi: 10.1109/ECACE.2019.8679247.

[20] M. Mohapatra, A. K. Parida, P. K. Mallick, M. Zymbler, and S. Kumar, “Botanical Leaf Disease Detection and Classification Using Convolutional Neural Network: A Hybrid Metaheuristic Enabled Approach,” Computers, vol. 11, no. 5, p. 82, May 2022, doi: 10.3390/computers11050082.

[21] S. Arya and R. Singh, “A Comparative Study of CNN and AlexNet for Detection of Disease in Potato and Mango leaf,” in 2019 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), IEEE, Sep. 2019, pp. 1–6. doi: 10.1109/ICICT46931.2019.8977648.

[22] F. Xu, H. Uszkoreit, Y. Du, W. Fan, D. Zhao, and J. Zhu, “Explainable AI: A Brief Survey on History, Research Areas, Approaches and Challenges,” Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 11839 LNAI, pp. 563–574, 2019, doi: 10.1007/978-3-030-32236-6_51.

[23] A. Tolba and N. Talal, “An Interpretable Deep Learning for Early Detection and Diagnosis of Wheat Leaf Diseases,” Optimization in Agriculture, vol. 1, pp. 81–93, May 2024, doi: 10.61356/J.OIA.2024.1257.

[24] S. Kolekar, S. Gite, B. Pradhan, and A. Alamri, “Explainable AI in Scene Understanding for Autonomous Vehicles in Unstructured Traffic Environments on Indian Roads Using the Inception U-Net Model with Grad-CAM Visualization,” Sensors 2022, Vol. 22, Page 9677, vol. 22, no. 24, p. 9677, Dec. 2022, doi: 10.3390/S22249677.

[25] M. M. Karim, Y. Li, and R. Qin, “Toward Explainable Artificial Intelligence for Early Anticipation of Traffic Accidents,” https://doi.org/10.1177/03611981221076121, vol. 2676, no. 6, pp. 743–755, Feb. 2022, doi: 10.1177/03611981221076121.

[26] D. Song et al., "A new xAI framework with feature explainability for tumors decision-making in Ultrasound data: comparing with Grad-CAM," Comput Methods Programs Biomed, vol. 235, p. 107527, Jun. 2023, doi: 10.1016/J.CMPB.2023.107527.

[27] L. Gorski, S. Ramakrishna, and J. M. Nowosielski, "Towards Grad-CAM Based Explainability in a Legal Text Processing Pipeline," Dec. 2020, Accessed: Jun. 10, 2024. [Online]. Available: https://arxiv.org/abs/2012.09603v1.

[28] "TensorFlow Lite | ML for Mobile and Edge Devices." Accessed: Jun. 10, 2024. [Online]. Available: https://www.tensorflow.org/lite

[29] G. Demosthenous and V. Vassiliades, "Continual Learning on the Edge with TensorFlow Lite," May 2021, Accessed: Jun. 10, 2024. [Online]. Available: http://arxiv.org/abs/2105.01946

[30] S. Ahmed and M. Bons, "Edge computed NILM: A phone-based implementation using MobileNet compressed by Tensorflow Lite," NILM 2020 - Proceedings of the 5th International Workshop on Non-Intrusive Load Monitoring, pp. 44–48, Nov. 2020, doi: 10.1145/3427771.3427852.

[31] I. Idrissi, M. Azizi, and O. Moussaoui, "A Lightweight Optimized Deep Learning-based Host-Intrusion Detection System Deployed on the Edge for IoT," International Journal of Computing and Digital Systems, vol. 11, no. 1, pp. 209–216, 2022, doi: 10.12785/IJCDS/110117.

[32] R. Thomas and D. Pavithran, "A Survey of Intrusion Detection Models based on NSL-KDD Data Set," in 2018 Fifth HCT Information Technology Trends (ITT), IEEE, Nov. 2018, pp. 286–291. doi: 10.1109/CTIT.2018.8649498.

[33] A. Sharma, I. M. Sharma, M. Sharma, K. Sharma, and A. Sharma, "Effectiveness of fungal, bacterial and yeast antagonists for management of mango anthracnose

(Colletotrichum gloeosporioides),” Egypt J Biol Pest Control, vol. 31, no. 1, p. 135, Dec. 2021, doi: 10.1186/s41938-021-00480-9.

[34] Y. Hu et al., “Lateral organ boundaries 1 is a disease susceptibility gene for citrus bacterial canker disease,” Proceedings of the National Academy of Sciences, vol. 111, no. 4, Jan. 2014, doi: 10.1073/pnas.1313271111.

[35] P. V. R. Reddy and K. Sreedevi, “Arthropod Communities Associated with Mango (Mangifera indica L.): Diversity and Interactions,” in Economic and Ecological Significance of Arthropods in Diversified Ecosystems, Singapore: Springer Singapore, 2016, pp. 271–298. doi: 10.1007/978-981-10-1524-3_14.

[36] E. Saeed et al., “Detection and Management of Mango Dieback Disease in the United Arab Emirates,” Int J Mol Sci, vol. 18, no. 10, p. 2086, Oct. 2017, doi: 10.3390/ijms18102086.

[37] I.-Y. Choi, J.-H. Joa, S.-W. Cho, W.-H. Lee, V. Galea, and H.-D. Shin, “Occurrence of stem and shoot cankers caused by Phomopsis fukushii on mango,” Australas Plant Dis Notes, vol. 12, no. 1, p. 56, Dec. 2017, doi: 10.1007/s13314-017-0280-3.

[38] Y. Shimizu, N. Moriya, O. Jahana, M. Matsumura, and M. Takeuchi, “Effects of sixteen insecticides on the mango gall midge Procontarinia mangicola (Shi) (Diptera: Cecidomyiidae) infesting mango leaves,” Int J Trop Insect Sci, vol. 43, no. 2, pp. 601–608, Mar. 2023, doi: 10.1007/s42690-023-00962-3.

[39] A. Ajitomi, T. Takushi, Y. Sato, C. Arasaki, and A. Ooshiro, “First report of powdery mildew of mango caused by Erysiphe quercicola in Japan,” Journal of General Plant Pathology, vol. 86, no. 4, pp. 316–321, Jul. 2020, doi: 10.1007/s10327-020-00918-2.

[40] M. Nasir, “EFFICACY OF SOME ORGANIC FUNGICIDES AGAINST ANTHRACNOSE AND POWDERY MILDEW OF MANGO,” Pak J Agric Sci, vol. 54, no. 03, pp. 493–496, Sep. 2017, doi: 10.21162/PAKJAS/17.1909.

mango

ORIGINALITY REPORT

17%	16%	5%	12%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to George Bush High School Student Paper	7%
2	Submitted to Daffodil International University Student Paper	3%
3	dspace.daffodilvarsity.edu.bd:8080 Internet Source	2%
4	assets.researchsquare.com Internet Source	1%
5	link.springer.com Internet Source	<1%
6	hashnode.com Internet Source	<1%
7	nlistsp.inflibnet.ac.in Internet Source	<1%
8	www.ncbi.nlm.nih.gov Internet Source	<1%
9	Amreen Batool, Yung-Cheol Byun. "Brain tumor detection with integrating traditional and computational intelligence approaches	<1%