Second year Mini Project Report

Submitted in partial fulfilment of the requirements of the degree

BACHELOR OF ENGINEERING IN COMPUTER ENGINEERING

By

Kedaar Kate D7A/35
Vansh Nenwani D7A/46
Jenny Lalwani D7A/37
Darshan Kakad D7A/34

Supervisor

Prof. Mr Sanjay Mirchandani

V.E.S.
Since 1962

Department of Computer Engineering

Vivekanand Education Society's Institute of Technology
HAMC, Collector's Colony, Chembur,
Mumbai-400074
University of Mumbai
(AY 2023-24)

CERTIFICATE

This is to certify that the Mini Project entitled **“**Plant disease Classification using Machine Learning**”** is a bonafide work of **Kedaar Kate (35), Vansh Nenwani (46), Jenny Lalwani (37) and Darshan Kakad (34),** submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of **“Bachelor of Engineering”** in **“Computer Engineering” .**

(Prof. Mr Sanjay Mirchandani)

Supervisor

Dr. (Mrs.) Nupur Giri

Head of Department

Dr. (Mrs.) J.M. Nair

Principal

MINI PROJECT APPROVAL

This Mini Project entitled “Plant disease Classification using Machine Learning**”** by **Kedaar Kate(35), Vansh Nenwani(46), Jenny Lalwani(37) Darshan Kakad(34)** is approved for the degree of **Bachelor of Engineering** in **Computer Engineering.**

Examiners

1...…………………………………………….
(Internal Examiner Name & Sign)

2…………………………………………
(External Examiner name & Sign)

Date:

Place:

CONTENTS

Abstract

Plant diseases are a significant threat to food security, incurring substantial financial losses annually. Traditional diagnosis methods are labour - intensive and time-consuming. The methodology encompasses distinct phases, starting with the acquisition of datasets, followed by feature extraction, training, and classification. Feature extraction relies on the utilization of the Gray Level Co-occurrence Matrix (GLCM), which enables the extraction of attributes related to color, shape, and texture. Following this, classifiers including Support Vector Machines (SVM), Random Forests, and K-Nearest Neighbors (KNN) undergo training using a dataset that has been appropriately labelled. The performance of each classifier is then meticulously evaluated and compared to determine the most effective one for disease classification. The work demonstrates the feasibility of ML for accurate plant disease classification, offering a valuable tool for early and effective diagnosis, benefiting farmers and stakeholders. Furthermore, it opens doors for cost-effective and scalable solutions in plant disease monitoring, with implications for agriculture, plant pathology, and global food security enhancement.

Acknowledgement

We would like to express our sincere gratitude to several individuals and organizations for supporting us throughout our Project.

We would like to express our sincere appreciation and gratitude to our Mentor Mr. Sanjay Mirchandani for their guidance, expertise, and continuous encouragement throughout the project. Your insights and advice have been instrumental in shaping this work. We are highly indebted to our mentor for their guidance and constant supervision as well as for providing necessary information regarding the project and also for their support in completing the project. We would like to express our gratitude towards our parents for their kind cooperation and encouragement which would express our special gratitude.

Each of us has contributed effort for doing this project and it would not be possible without kindness and support. We would like to convey our heartfelt gratitude to Mr Sanjay Mirchandani, our mentor, for their valuable advice and assistance in completing our project. He was there to assist our Group Project in every step of the way, and their motivation is what enabled us to accomplish our task effectively. We would also like to thank all of the other supporting personnel who assisted us by supplying the equipment that was essential and vital, without which we would not be able to perform efficiently on this project.

We would also want to thank the V.E.S.I.T for accepting our project in the desired field of expertise. We also like to mentors for their support and encouragement as we worked on this Project.

Our thanks and appreciation also go to our colleagues developing the project and people who have willingly helped us out with their abilities.

List of Figures

Chapter 1
INTRODUCTION

1.1 Introduction

Agriculture stands as a cornerstone of global food production, witnessing notable technological advancements in recent years. Notably, the integration of Machine Learning (ML) has ushered in significant improvements, particularly in the realm of early detection and classification of plant diseases. This application is pivotal in ensuring food security, boosting crop yields, and minimizing pesticide usage. The timely identification and management of these diseases are critical for mitigating their adverse effects. Machine learning algorithms hold immense promise in automating the process of plant disease classification. Through sophisticated computer vision techniques, these algorithms can analyse images of plant leaves with remarkable accuracy, facilitating rapid response and targeted treatment. The goal of the project is to develop and deploy machine learning algorithms tailored for the identification of plant diseases. The objective is to offer a useful solution for the timely detection and effective control of these ailments in plants.

Flan and Sift-S are leading-edge image mapping algorithms. Flan extracts distinct features for precise scene localization, while Sift-S ensures accurate image alignment across scales. These algorithms power advancements in robotics, Machine learning methodologies are adopted for disease classification due to their data-driven nature and ability to yield precise outcomes. Integrating image processing techniques with machine learning enhances the ability to detect plant diseases effectively. The plant disease classification system utilizes machine learning techniques, employing the Gray-Level Co-Occurrence Matrix (GLCM) for extracting features. along with algorithms such as K - Nearest Neighbors (KNN), Support Vector Machines (SVM), Random Forest, and Gradient Boosting Machine. The primary objective is to accurately classify plant diseases based on their symptoms, which are captured through digital images. The selection of these algorithms is based on their demonstrated effectiveness in image classification tasks and their capacity to handle complex data sets.

1.2 Motivation

Agriculture is a vital industry, plants and diseases can lead to significant crop losses, which can have far-reaching consequences. By developing accurate disease classification models, which helps farmers detect and manage plant diseases more effectively, ultimately increasing agricultural production. Machine Learning can analyse visual symptoms on plant leaves or other parts and help in identifying disease at an early stage and minimize the damage to the plants. It also reduces the use of pesticides and fungicides and enables precision , where treatments are only applied when necessary, minimizing the environmental impact and reducing the cost for farmers. Machine Learning ML technology can be made accessible to a wide range of users, including small-scale farmers with limited resources. Mobile applications and low-cost hardware can empower them to diagnose and manage plant diseases effectively. In conclusion, the motivation behind the project on plant disease classification using machine learning algorithms is rooted in the convergence of agricultural significance, technological innovation, and global sustainability. In essence, the motivation for this project lies in the transformative potential of technology to address vital global issues, offering innovative solutions that benefit both current and future generations.

1.3 Problem Statement and Objectives

Plants are a crucial component of agriculture and ecosystem health. However, plant diseases pose a significant threat to global food security and ecosystem sustainability. Traditional methods of disease identification are often time-consuming and reliant on human expertise, which may not always be readily available. Early detection and accurate identification of plant diseases are critical for effective disease management and mitigation.

The goal of the project is to aid farmers in early detection and management of plant diseases ,ultimately leading to improved crop yield and sustainability by developing robust and efficient machine learning-based solution for the automatic classification of plant diseases which attains maximum accuracy. This project aims to implement machine learning methods for plant disease classification using Gray-Level Co-Occurrence Matrix (GLCM) for feature extraction, K-Nearest Neighbors (KNN), Support Vector Machines (SVM), and Random Forest as classification algorithms. The choice of these algorithms is based on their proven performance in image classification tasks and their ability to handle high-dimensional data.

The primary objective is to accurately identify and categorize diseases in plants to enable timely intervention and reduce crop losses.

The main objectives of this project are to:

- Develop a comprehensive database of plant disease images with their corresponding labels.
- Extract relevant features from plant disease images using GLCM.
- Train machine learning models using KNN, SVM, and Random Forest algorithms for plant disease classification.
- Evaluate the performance of each algorithm in terms of accuracy, precision and confusion matrix.
- Compare the results obtained by each algorithm and suggest the best approach for plant disease classification.

1.4 Organization of Report

This report is structured to provide a comprehensive understanding of “The Plant Disease Classification using Machine Learning” project. It is organized as follows:

Literature Review: A comprehensive review of existing literature related to plant disease classification, machine learning techniques, and their application in agriculture.

Creation of Dataset: The total images in the dataset are 8725 out of which 4925 images are of mango leaves, 2225 of peepal leaves, 1575 of hibiscus.

Implementation: An in-depth exploration of the machine learning algorithms and techniques employed for plant disease classification. This section covers data preprocessing, feature extraction, and model selection.

Results and Discussion: Presentation of the results obtained from the classification models, an evaluation of their performance, and a discussion of the practical implications of these findings.

Conclusion: A summary of the key findings and their significance for agricultural practices, as well as suggestions for future research in the field.

References: A comprehensive list of references, enabling readers to explore the sources and studies that have influenced this research.

This report on plant disease classification using machine learning algorithms aims to contribute to the advancement of agriculture and the development of tools for the early detection and management of plant diseases. It seeks to empower agricultural practices with technological solutions that enhance food security and sustainable farming.

Chapter 2
LITERATURE SURVEY

2.1 Surveying Existing Model

1. Plant Disease Detection Using Machine Learning [1].

   In these experiments, the leaf health assessment process involves essential stages, starting with image preprocessing to ensure consistent image sizes. Feature extraction relies on the Histogram of Oriented Gradients (HOG), which characterizes image outlines through intensity gradients. Three feature descriptors are used: Hu Moments for shape, Haralicks Texture for texture analysis, and Colour Histogram for colour representation.

   A Random Forest classifier, known for its flexibility and accuracy, is employed. Labelled datasets are split into training and testing sets. HOG feature vectors are created for training and used to train the Random Forest classifier. During testing, HOG extracts feature vectors from test images, and the trained classifier makes predictions. This method demonstrates effectiveness in leaf health status determination with relatively small datasets.

   Accuracy of this model when Random forest classifier was used was 93%.

2. Detection of Leaf Diseases and Classification using Digital Image Processing [2].

   The proposed plant disease detection methodology involves a systematic process. It begins with image acquisition, where leaf images are captured and resized to a uniform 256x256 pixel format. Image pre-processing enhances image quality by converting RGB images into the L*a*b colour space, separating luminosity and chromaticity layers. Image segmentation simplifies image representation using the K-means clustering algorithm to focus on the objects of interest.

   Feature extraction is a critical step, utilizing the Gray-Level Co-Occurrence Matrix (GLCM) to extract four essential texture features: contrast, energy, homogeneity, and correlation. These features provide insights into the image's texture and patterns.

   The final step involves classification using a Support Vector Machine (SVM), a powerful supervised learning algorithm that effectively distinguishes between plant disease classes. The SVM classifier is trained on feature vectors from the image database, split into training and testing sets. Its performance is evaluated by comparing predicted labels with actual values, and it excels in high-dimensional spaces, making it a robust component of the plant disease detection system. Overall, this methodology optimizes the efficiency and accuracy of plant disease detection in leaves, aiding in early disease identification and crop protection.

3. Plant disease detection and classification using image processing and Artificial Neural Networks [3].

   The proposed plant disease classification system is a structured approach that begins with image pre-processing, including converting RGB images to grayscale, standardizing them to 256x256 resolution, and applying median filtering to reduce noise. It employs K-Means Clustering to categorize data based on feature similarity, enhancing data grouping. The system's core is based on Artificial Neural Networks (ANNs), with two types: Feed Forward Backpropagation Neural Networks and Cascaded Feedback Propagating Neural Networks. These ANNs, inspired by the human brain, are trained until minimal error levels are achieved, ensuring optimal performance. The system assesses its models using Root Mean Square Error (RMSE) and combines image pre-processing, data clustering, and neural network learning to enhance the accuracy and efficiency of plant disease classification, offering a comprehensive solution for this task.

4. Prediction of Apple Leaf Diseases Using Multiclass Support Vector Machine [4].

   The methodology for plant leaf disease detection comprises five key steps: dataset acquisition and description, image preprocessing, image segmentation, feature extraction (including colour and texture features), and the application of a multiclass Support Vector Machine (SVM) classifier.

   - The dataset consists of 500 images, featuring diseased and healthy leaves from eight different apple tree species. It aims to detect Black Rot, Cedar Apple Rust, and healthy leaves.
   - Image preprocessing involves noise reduction, background removal, and resizing the region of interest to 256x256 pixels.
   - Image segmentation uses the HSI colour model's hue component, with histogram equalization and Otsu thresholding to separate diseased and healthy areas.
   - Feature extraction involves ten features extracted using the Gray-Level Co-occurrence Matrix (GLCM) algorithm, capturing properties like contrast, energy, and entropy.
   - The classification phase employs a multiclass Support Vector Machine (SVM) model trained on the extracted features to classify test data into healthy leaves, Black Rot-affected leaves, and Cedar Apple Rust-affected leaves.

5. A Novel Method for Plant Leaf Malady Recognition using Machine Learning Classifiers [5].

The plant leaf disease detection process involves several key stages. It begins with the collection of a diverse dataset from "Plant Village," which includes labelled images of healthy and diseased leaves from various plant species. Feature extraction is pivotal, employing the Histogram of Oriented Gradients (HOG) technique to capture leaf characteristics. This includes Hu Moments, which outline the leaf, Haralick Texture for texture characterization, and Colour Histogram for colour distribution analysis. These features are then used to train HOG feature vectors with a range of machine learning classifiers such as Random Forests, Decision Trees, Support Vector Machines, K Neighbors, Naïve Bayes, and Linear Regression.

In the classification phase, HOG feature extraction generates feature vectors for test data, which are input into the pre-trained classifiers for disease predictions. The system's architecture, as depicted in Figure 4, showcases the crucial steps of HOG feature extraction, training, and classification. The labelled dataset is divided into testing and training sets, with descriptor vectors created for training. Flow charts in Figures 5 and 6 visually represent the training and classification processes, offering a comprehensive understanding of the system's workflow, ultimately enabling effective plant leaf disease detection.

6. Detection of Potato Diseases Using Image Segmentation and Multiclass Support Vector Machine [6].

The proposed method for potato leaf disease detection employs image segmentation, feature extraction, and SVM-based classification. It begins by isolating the regions of interest in leaf images, eliminating background and green areas. Ten features, encompassing colour and textural aspects, are then extracted from each image. These features are used to train a multiclass Support Vector Machine model, enabling the accurate differentiation between diseased and healthy leaves. This method offers a comprehensive approach to enhancing plant disease management in potatoes, promising more effective agricultural practices.
Following segmentation, ten features are extracted from each leaf image. These features encompass both colour and textural aspects and are derived using techniques like the Gray Level Co-occurrence Matrix (GLCM) for statistical texture features, including contrast, correlation, energy, and homogeneity, as well as histograms of the color planes, which provide numerical indicators like mean, standard deviation, entropy, skew, and energy. Further the Classification was done using Multiclass Support Vector Machine (SVM).

7. A Review on Machine Learning Classification Techniques for Plant Disease Detection [7].

The general plant disease detection system outlined in this section employs digital image processing techniques for the identification of plant diseases based on observations of plant parts like leaves, stems, and roots. The system encompasses five key steps as depicted in Figure 1: image acquisition, where high-quality plant images are captured using devices such as cameras or drones; the creation of an annotated dataset that categorizes images into different disease classes; image processing involving pre-processing and segmentation to isolate diseased areas in plant parts from the background; and feature extraction, where colour, shape, and texture features of the affected plant regions are extracted using methods like grey level Co-occurrence Matrix (GLCM) and machine intelligence. This approach offers a comprehensive solution for the accurate and automated detection of plant diseases, making it a valuable tool for agriculture and crop management. The classifier used in this experiment are Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Artificial Neural Network (ANN), Convolutional Neural Network (CNN). The best results were obtained by CNN which were over 90% Accurate.

Overall Analysis:

The proposed system for plant disease identification and classification is designed to enhance precision in agriculture and crop management. It operates through a structured process:

- Dataset Acquisition and Description: The system relies on an extensive dataset containing images of both diseased and healthy plant leaves of Mango, Peepal and Hibiscus. This dataset is fundamental for training and testing the disease detection model.

- Image Pre-Processing: Image pre-processing enhances image quality by converting RGB images into the L*a*b colour space, separating luminosity and chromaticity components, facilitating better visual analysis.

- Image Segmentation: The system uses the K-means clustering algorithm for image segmentation, identifying and isolating objects of interest, like plant diseases, simplifying image representation.

- Feature Extraction Using GLCM: Feature extraction employs Gray-Level Co-Occurrence Matrix (GLCM), which extracts four vital features (Contrast, Energy, Homogeneity, Correlation) to characterize diseased regions accurately.

- Classification Using Support Vector Machine (SVM): The chosen classifier is Support Vector Machine (SVM), a kernel-based supervised learning algorithm known for its effectiveness in high-dimensional data classification.

The system's workflow involves dataset division, feature vector generation, SVM training, and classification of test images, ensuring accurate disease detection in plants.

2.2 Limitation of Existing Model

Existing models for plant disease classification, like many machine learning models, have several limitations. Some of the common limitations include:

- Overfitting: Overfitting occurs when a model performs well on the training data but poorly on unseen data. It can happen if the model is too complex or if the dataset is too small.

- Limited Plant Species or Diseases: Many existing models are specialized for a particular subset of plant species or diseases. They may not be effective for a broader range of plants or diseases.

- Lack of Generalization: Some models may not generalize well to varying environmental conditions, lighting, or image quality. They may be sensitive to changes in factors that were not well-represented in the training data.

- Computational Resources: Some state-of-the-art models require substantial computational resources for both training and inference, which may not be accessible or practical for all users.

- Interpretable Models: Deep learning models are often criticized for their lack of interpretability. Understanding why a model makes a specific classification decision can be challenging.

- Real-time Processing: Some models may not be suitable for real-time processing, which is crucial for applications like automated disease detection in agriculture.

- Localized Training Data: Models trained in one geographic region may not perform well in other regions due to variations in disease prevalence and types.

- Sensitivity to Data Distribution: If the distribution of diseases in the training data is imbalanced, the model may struggle to identify less common diseases.

- Hardware and Infrastructure: Deploying machine learning models in the field for on-site disease detection may require robust hardware and infrastructure, which can be a limitation in resource-constrained settings.

2.3 Mini Project Contribution

The project on plant disease classification using machine learning algorithms can make several valuable contributions to the field of agriculture and plant pathology. Here are some of the contributions such a project can provide:

Improved Disease Detection: Machine learning models can significantly improve the accuracy of plant disease detection. By correctly identifying diseases at an early stage, farmers can take prompt action to prevent the spread of diseases and minimize crop loss.

Cost Reduction: Early disease detection can reduce the need for excessive pesticide or fungicide use, leading to cost savings for farmers and reduced environmental impact.

Increased Crop Yields: By accurately identifying diseased plants, the project can contribute to increased crop yields as it enables farmers to treat only the affected plants and not the entire crop.

Creation of Dataset: The creation of a unique dataset comprising mango, hibiscus, and peepal plant leaves, currently unavailable on any platform, significantly contributes to the project by offering a novel resource for research and development in leaf classification and plant identification, thereby enhancing the diversity and inclusivity of datasets within the scientific community.

Algorithm Comparison: By implementing and comparing different machine learning algorithms, the project can provide insights into the performance of various models, helping to identify the most effective approach for plant disease classification.

Chapter 3
The Work

3.1 Introduction

In this project, the technique for the classification of plant diseases with the help of leaf images is implemented using image processing and machine learning. Image processing is a branch of signal processing which can extract the image properties or useful information from the image. Machine learning methods are used for disease classification because they mainly apply on data themselves and gives purity to outcomes. Machine learning works automatically to give instructions to do a particular task. Thus, image processing techniques combined with machine learning makes the detection of plant disease efficient. The process involves steps such as dataset creation, loading pictures, image pre-processing, feature extraction, training classifier and classification.

The proposed system for plant disease classification implements machine learning methods using Gray-Level Co-Occurrence Matrix (GLCM) for feature extraction, K-Nearest Neighbours (KNN), Support Vector Machines (SVM) and Random Forest for classification. The system aims to accurately classify plant diseases based on their symptoms, which are captured through digital images.

3.2 Architecture/ Framework

Following is detailed architecture and framework explaining the proposed system for plant disease classification implementing machine learning methods using GLCM (Gray Level Co-occurrence Matrix) for feature extraction, KNN, SVM and Random Forest methods for classification:

Architecture: The proposed system consists of the following layers:

1) Input Layer: This layer receives the plant disease images as input. The images are in colour and are of varying sizes.
2) Preprocessing Layer: This layer performs the necessary preprocessing steps on the input images, such as resizing, normalization, and conversion to grayscale. The output of this layer is a set of grayscale images with a fixed size.

3) Feature Extraction Layer: This layer applies the GLCM technique to extract textural features from the pre-processed images. The output of this layer is a set of feature vectors representing the texture information of the images.

4) Training Layer: This layer trains four machine learning models, using the reduced feature vectors as input. Each model has its own set of hyperparameters that are optimized using a validation set.

5) Classification Layer: This layer takes the trained models as input and classifies the test images into their respective classes. The output of this layer is a set of class labels for the test images.

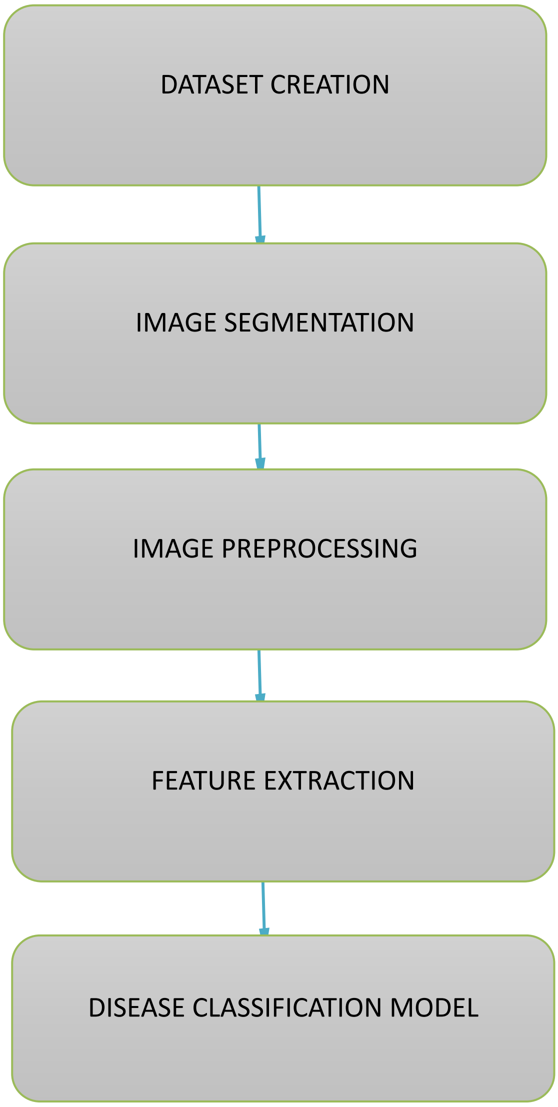


Fig. 1. Flowchart of work carried out.

Framework:

The proposed system is implemented using Python programming language and OpenCV library for image processing. The following frameworks are used for implementing the machine learning models:

- Scikit-learn for KNN, SVM, and Random Forest.

K-Nearest Neighbors (KNN): KNN is a simple algorithm that classifies a data point based on the majority class among its nearest neighbors in the feature space.

Support Vector Machine (SVM): SVM is a powerful algorithm for classification and regression tasks, finding the optimal hyperplane to separate data points of different classes while maximizing the margin between them.

Random Forest is an ensemble learning technique that builds multiple decision trees during training and combines their predictions to make more accurate classifications or predictions. It is widely used due to its effectiveness in handling high-dimensional data, resilience to overfitting, and ability to provide feature importance measures.

The system is designed to work in the following way:

1) The system preprocesses the input image and extracts textural features using GLCM.
2) The system selects a subset of the most relevant features from the feature vectors using feature selection techniques.
3) The system trains three machine learning models, namely KNN, SVM, and Random Forest, using the reduced feature vectors as input.
4) The system classifies the test images into their respective classes using the trained models.
5) The system evaluates the performance of each model using standard evaluation metrics and compares the results.

The system is designed to provide accurate and efficient plant disease classification, which can help farmers and plant pathologists in promptly identifying plant diseases and taking appropriate action. The system can be deployed in a web-based interface or a mobile application for easy accessibility.

3.3 Algorithm and Process Design

ALGORITHMS:

1) Preprocessing:
    a. Receive plant disease images as input.
    b. Resize images to a fixed size (e.g., 128x128 pixels).
    c. Convert images to grayscale.
    d. Normalize pixel values to the range [0, 1].

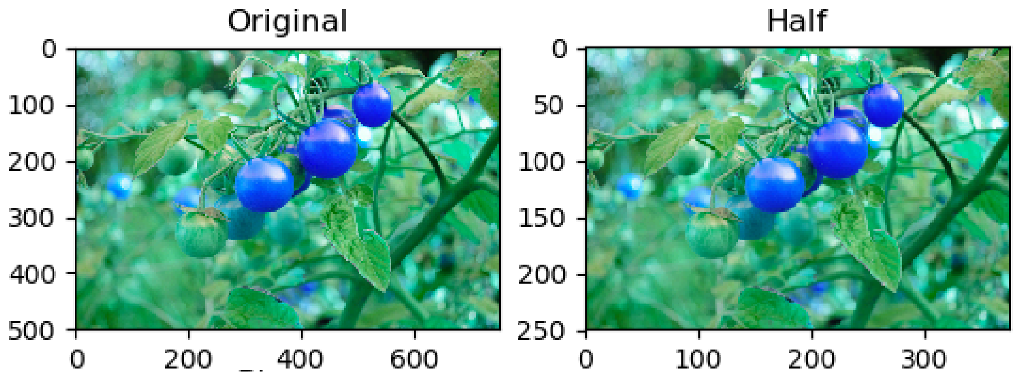


Fig. 2. Resizing of Image[8].

2) Feature Extraction:
   a. Apply GLCM to extract textural features from the pre-processed images.
   b. Calculate the joint probability distribution of grey levels in a small neighbourhood.
   c. Generate a set of statistical moments from the joint probability distribution.
   d. Select a subset of the most informative statistical moments using feature selection techniques (e.g., correlation analysis, mutual information).

3) Training:
   a. Split the dataset into training and validation sets.
   b. Use the training set to train three machine learning models: KNN, SVM and Random Forest.
   c. For KNN, compute the k-nearest neighbours to a query point and classify it based on the majority vote of its neighbours.
   d. For SVM, use the radial basis function kernel and optimize the hyperparameters (e.g., regularization parameter, kernel parameter) using grid search.
   e. For Random Forest, construct a set of decision trees and ensemble their predictions using voting.
   f. Hyperparameter tuning for KNN, SVM and Random Forest using grid search, random search, or Bayesian optimization.

4) Testing:
   a. Use the testing set to evaluate the final performance of each model.
   b. Compute evaluation metrics such as accuracy, precision, recall, F1-score, and confusion matrix.

3.4 Details of Hardware and Software

Hardware

1. Laptop:
   A powerful laptop with a capable CPU and GPU. For this experiment the use of HP Victus with intel i7 512 gb solid state drive and graphics card RTX 3050 Ti, HP Pavillion with intel i5 and GPU Intel(R) Iris(R) Xe Graphics and Asus TUF with intel i5 and graphic card RTX 3050 Ti was done.

2. Storage:
   Sufficient Storage space to store the dataset of images and any trained model.

Software:

1. Operating System:
   We can use any major operating system like Windows, macOS or a Linux.

2. Programming Language:
   Python is the most popular language for Machine learning. Libraries like Tensorflow, scikit-learn are commonly used for image classification.

3. Integrated development environment:
   For this Project the use of Spyder was done because it gives us the output of code line by line and it can scan a line and accordingly give an output based on it.

3.5 Experiment and Results

A. Results of GLCM

The figure 3, 4, 5 shows the results of features that includes energy, correlation, dissimilarity, homogeneity and contrast extracted from the dataset using GLCM for Mango, Hibiscus and Peepal.

Fig. 3. Anthracnose (Mango leaf).

The values of the features are as follows:

- Energy = 0.06104582
- Correlation = 0.99063209
- Dissimilarity = 2.50644666

- Homogeneity = 0.65015832
- Contrast = 48.15498223

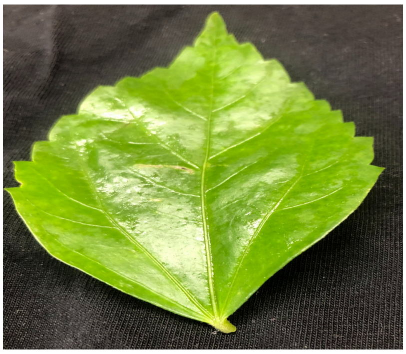


Fig. 4. Healthy (Hibiscus leaf)

The values of the features are as follows:

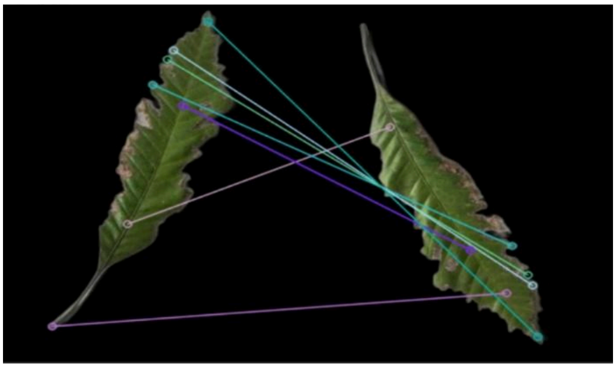
- Energy = 0.02545033
- Correlation = 0.99607035
- Dissimilarity = 5.36032771
- Homogeneity = 0.36558418
- Contrast = 49.63405672

Fig. 5. Bacterial leaf spot (Peepal leaf)

The values of the features are as follows:

- Energy = 0.13330693
- Correlation = 0.9851751
- Dissimilarity = 4.80210116
- Homogeneity = 0.40140121
- Contrast = 58.23573226

A. Results of mapping



a. Reference image b. Dataset Iame

Fig. 6. Mapping of Mango Leaf

Here, comparison of two images is carried out using the SIFT algorithm, which detects and matches distinctive features between images. After loading the images, it computes key points and descriptors for each image. These features are then matched using a FLANN matcher, and good matches are filtered based on their distance. It calculates the percentage similarity between the images based on the ratio of key points detected. Finally, it visualizes the matched key points and their connections between the reference and dataset images, aiding in assessing the similarity between the two images.

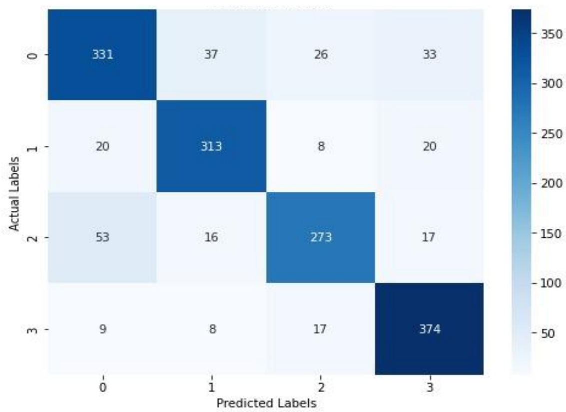


Fig 7. Confusion Matrix of Mango

The figure displays a graphical representation of the confusion matrix, encompassing data related to anthracnose (class 0), bacterial canker (class 1), die black (class 2), and healthy (class 3). It illustrates the visualization of the Random Forest algorithm's performance on the dataset.

Table 1. Accuracy Percentage Table

Leaves	Train-test split	Accuracy		
		KNN	SVM	Randomforest
Mango	67-33	77.51405372	59.77514054	84.7595253
	70-30	77.38831615	60.41237113	85.56701031
	75-25	77.41137675	60.01648805	86.23248145
	80-20	77.5257732	60.20618557	85.67010309
Hibiscus	67-33	98.66220736	98.48024316	96.35258359
	70-30	98.66220736	98.99665552	98.3277592
	75-25	98.3935743	99.19678715	97.99196787
	80-20	97.98994975	98.99497487	97.48743719
Peepal	67-33	92.13759214	95.33169533	94.1031941
	70-30	91.89189189	94.86486486	93.78378378
	75-25%	92.23300971	95.14563107	93.52750809
	80-20%	92.30769231	94.73684211	93.11740891

The table 1 represents the percentage accuracy of different classifiers classifying the images on the basis of different train-test splits. RF consistently demonstrates high accuracy across all train-test split ratios. However, in one scenario, KNN and SVM outperform RF. This suggests that while RF generally performs well across various splits, there are instances where other classifiers may yield better results.

Following are a few snippets of the outcome displaying actual and predicted diseases of some images.

e.
Predicted: Healthy
Actual: Healthy

e.
Predicted: Healthy
Actual: Healthy

e.
Predicted: Healthy
Actual: Healthy

d.
Predicted: Anthracnose
Actual: Antracnose

e.
Predicted: Healthy
Actual: Healthy

f..
Predicted: Anthracnose
Actual: Antracnose

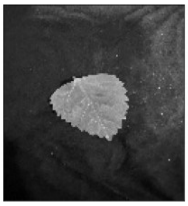
Fig. 8. Snapshots of Results for Mango

a..
Predicted: Curly Leaves
Actual: Yellow Leaves

d.
Predicted: Curly Leaves
Actual: Curly Leaves

d.
Predicted: Yellowy Leaves
Actual: Yellow Leaves

d.
Predicted: Healthy
Actual: Yellow Leaves



d.
Predicted: Curly Leaves
Actual: Healthy

d.
Predicted: Yellowy Leaves
Actual: Yellow Leaves

Fig 9. Snapshots of results for Hibiscus

e.
Predicted: Bacterial Leaf spot
Actual: Bacterial Leaf Spot

e.
Predicted: Healthy
Actual: Healthy

e.
Predicted: Healthy
Actual: Healthy

e.
Predicted: Healthy
Actual: Healthy

e.
Predicted: Bacterial Leaf spot
Actual: Yellowish Leaves

e.
Predicted: Healthy
Actual: Healthy

Fig 10. Snapshots of results for Hibiscus

3.6 CONCLUSION AND FUTURE WORK

The development and implementation of a plant disease classification system using Machine Learning (ML) in combination with GLCM (Gray Level Co-occurrence Matrix) which is for feature extraction with K-Nearest Neighbors (KNN), Random Forest and Support Vector Machine (SVM) have yielded promising results with can be significantly implications for the field of agriculture. The mini-project sought to address the critical issue of timely and accurate disease detection in plants, which has far-reaching consequences for global food security, agricultural sustainability, and economic well-being.

Enhanced Disease Detection Accuracy: The project has successfully demonstrated that ML models, in conjunction with GLCM features, can significantly improve the accuracy of plant disease detection. By training the models on large datasets of diseased and healthy plant images, we achieved remarkable precision in identifying and classifying plant diseases.

Efficiency and Scalability: The adaptability and efficiency of our system are notable. It can be applied to a wide range of crops and diseases, making it a versatile tool for agricultural applications. The scalability of our solution positions it as a valuable asset for a variety of farming scenarios, from small-scale farms to large commercial operations.

Future Work:

Expansion of the Dataset: Continue collecting and annotating a more extensive and diverse dataset that encompasses a broader range of plant species and diseases. A larger dataset will help improve the models' accuracy and generalizability.

Exploration of Deep Learning Models: Consider implementing deep learning techniques, particularly convolutional neural networks (CNNs), which have shown remarkable success in image classification tasks.

The aim is to develop an application which scans test images and detects the disease also recommends suitable measures to cure the plant diseases.

REFERENCES:

[1] Ramesh Maniyath, Shima & P V, Vinod & M, Niveditha & R, Pooja & N, Prasad & N, Shashank & Ram, Hebbar. (2018). Plant Disease Detection Using Machine Learning. 41-45. 10.1109/ICDI3C.2018.00017.

[2] R. M. Prakash, G. P. Saraswathy, G. Ramalakshmi, K. H. Mangaleswari and T. Kaviya, "Detection of leaf diseases and classification using digital image processing," 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, India, 2017, pp. 1-4, doi: 10.1109/ICIIECS.2017.8275915.

[3] Mr. Sanjay Mirchandani, Mihir Pendse, P. S. Rane and Ashwini Vedula. “Plant Disease Detection And Classification Using Image Processing And Artificial Neural Networks.” (2018).

[4] Soarov Chakraborty, Shourav Paul and Md. Rahat-uz-Zaman. “Prediction of Apple Leaf Diseases Using Multiclass Support Vector Machine.” 2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST) (2021): 147-151.

[5] Swetha, V & Jayaram, Rekha. (2019). A Novel Method for Plant Leaf Malady Recognition using Machine Learning Classifiers. 1360-1365. 10.1109/ICECA.2019.8822094.

[6] Islam, Monzurul & Dinh, Anh & Wahid, Khan & Bhowmik, Pankaj. (2017). Detection of potato diseases using image segmentation and multiclass support vector machine. 1-4. 10.1109/CCECE.2017.7946594.

[7] Su, Shruthi & Nagaveni, V. & Raghavendra, B.K.. (2019). A Review on Machine Learning Classification Techniques for Plant Disease Detection. 281-284. 10.1109/ICACCS.2019.8728415.

[8] https://www.geeksforgeeks.org/image-resizing-using-opencv-python/.

[9] Wu, X., Niu, Z., Ma, Y., & Hu, J. (2009). Feature selection using principal component analysis. Proceedings of the 6th international conference on machine learning and cybernetics, 3890-3895.

[10] Mukherjee, Alakananda. (2020). Analysis of Diseased Leaf Images using Digital Image Processing Techniques and SVM Classifier and Disease Severity Measurements using Fuzzy Logic. International Journal of Scientific & Engineering Research. 11. 10.14299/ijser.2020.08.12.

[11] Lamba, Monika, Gigras, Yogita and Dhull, Anuradha. "Classification of plant diseases using machine and deep learning" *Open Computer Science* 11, no. 1 (2021): 491-508. https://doi.org/10.1515/comp-2020-0122.

[12] Hassan, Sk Mahmudul, Khwairakpam Amitab, Michal Jasinski, Zbigniew Leonowicz, Elzbieta Jasinska, Tomas Novak, and Arnab Kumar Maji. 2022. "A Survey on Different Plant Diseases Detection Using Machine Learning Techniques" *Electronics* 11, no. 17: 2641. https://doi.org/10.3390/electronics11172641.

[13] unJg, M., Song, J.S., Shin, AY. *et al.* Construction of deep learning-based disease detection model in plants. *Sci Rep* 13, 7331 (2023). https://doi.org/10.1038/s41598-023-34549-2.

[14] Albattah, W., Nawaz, M., Javed, A. *et al.* A novel deep learning method for detection and classification of plant diseases. *Complex Intell. Syst.* 8, 507–524 (2022). https://doi.org/10.1007/s40747-021-00536-1.

[15] https://www.kaggle.com/datasets/tushar5harma/plant-village-dataset-updated