A Hybrid CNN-SVM Model for Multi-Classification of Mango Leaf Disease

By
Maliha Muttalib
201-15-13873

FINAL YEAR DESIGN PROJECT REPORT

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

Supervised by
Mr. Md. Abbas Ali Khan
Assistant Professor

Department of Computer Science and Engineering Daffodil International University

Co-Supervised by
Mr. Md. Assaduzzaman
Lecturer (Senior Scale)

Department of Computer Science and Engineering Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY
Dhaka, Bangladesh

September 17, 2025

APPROVAL

This Project titled **A Hybrid CNN-SVM Model for Multi-Classification of Mango Leaf Disease**, submitted by Maliha Muttalib, ID No: **201-15-13873** to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on **17 September, 2025**.

BOARD OF EXAMINERS

Dr. Bimal Chandra Das (BCD) — **Chairman**
Professor and Dean (in-charge)
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Mr. Shah Md Tanvir Siddiquee (SMTS) — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Md. Hasanuzzaman Dipu (MHD) — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Nazibur Rahman — **External Examiner**
Technical Lead -Database Administrator
Telenor – Grameen Phone Account

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Mr. Md. Abbas Ali Khan, Assistant professor** Department of Computer Science and Engineering, Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

16.9.25

Mr. Md. Abbas Ali Khan
Assistant Professor
Department of Computer Science and
Engineering Daffodil International
University

Co-Supervised by:

Mr. Md. Assaduzzaman
Lecturer (Senior Scale)
Department of Computer Science and
Engineering Daffodil International
University

Submitted by:

Maliha 16.9.25

Maliha Muttalib
Student ID:201-15-13873
Department of Computer Science and
Engineering Daffodil International
University

ACKNOWLEDGEMENTS

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

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

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

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

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

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

ABSTRACT

Mango leaf spot diseases significantly affect the health and yield of mango trees, posing a serious threat to agricultural productivity in tropical regions. Traditional disease detection methods are often labor-intensive, time-consuming, and prone to human error. To address this issue, this research proposes a hybrid deep learning and machine learning approach for the automatic multi-class classification of mango leaf spot diseases. The system leverages a Convolutional Neural Network (CNN) using the pre-trained VGG16 model as a feature extractor, combined with three classical classifiers: Support Vector Machine (SVM), Random Forest (RF), Deep feature extraction (VGG16 CNN), Classical ML classifiers (SVM, RF, KNN, LR, etc.), Multi-class classification and K-Nearest Neighbors (KNN). A comprehensive dataset containing images of eight categories including seven common diseases (Anthracnose, Bacterial Canker, Cutting Weevil, Die Back, Gall Midge, Powdery Mildew, Sooty Mould) and one healthy class was used for training and validation. The CNN-based feature extraction enabled the effective representation of leaf textures and disease patterns, which were then classified by the individual models. The SVM classifier achieved the highest accuracy among the tested algorithms, outperforming RF and KNN. Evaluation metrics such as accuracy, classification report, and confusion matrix were used to assess performance, while Principal Component Analysis (PCA) and visualizations provided insights into data distribution and class separation. This hybrid model demonstrates strong potential for aiding farmers and agricultural professionals in early and accurate identification of mango leaf diseases, contributing to timely intervention and improved crop management.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Many tropical countries depend on agriculture and mango is one of the popularly cultivated and consumed fruits. Nevertheless, production of mango is facing the challenge of leaf spot diseases, which lessen productivity and quality. Conventional detection techniques are lagging and have limitations with regards to error detection. The project offers a hybrid architecture of convolutional neural networks and machine learning that is able to recognize various mango leaf diseases through image scanning by providing automatic classification.

1.1 Introduction

Many have been the countries which are based mainly in the tropical areas which have agriculture as a playing ground in their economy. One of the most common fruits that are cultivated and consumed through tropical farming is called the king of fruits, Mango. Many mango varieties, however, are commonly susceptible to numerous leaf spot infections, which reduces the production and quality of mango. There is a strong need to identify these diseases accurately and in a timely manner so as to be effectively treated. The conventional methods of disease detection are very time-consuming, moreover, manual observation is largely unreliable and the use of human imagination is high off determining the existence of a disease. Due to the recent occurrence of artificial intelligence, and specifically related to the field of computer vision, automation in reference to plant disease diagnosis based on the use of an image has gained increasing recognition. This project presents an integrated solution to the classification of various mango leaf diseases consisting of convolutional neural networks (CNN) and classical machine learning models.

1.2 Motivation

Building that the interest in this project emerged out of the immense economic performance of the mango leaf diseases as well as constrained actions of present detection methodology. The farmers do not even have access to the means of disease diagnostics which is kind of timely and dependable thus resulting in further loss of crops and postponed treatment. The increased availability of smart phones and cloud-based computing services opens a door into creating smart systems that have the capacity to automatically process images, making diagnostic diagnoses of diseases. Using the strength of pre-trained deep learning networks, feature extractors like diseases. Using the strength of pre-trained deep learning networks, feature extractors like

1.3 Objectives

This project has the following major goals:

i. To synthesize and develop a hybrid framework combining CNN in feature extraction with classical classifiers in obtaining final mango leaf disease classification.

ii. To prepare and test several classifiers such as SVM, RF and KNN and using a curated library of mango leaf pictures.

iii. To compare the performance of these classifiers based on some variables, like the accuracy, confusion matrix and classification reports.

iv. To have a visual representation of the clusters of the features and their separation in classes using the dimensional reduction methods such as Principal Component Analysis (PCA).

v. To establish a solid base upon which more refinements on real-time disease detecting instruments in the agricultural sector can be made.

1.4 Methodology

The proposed tool In this context, the VGG16 model will be used as a deep feature extractor the mango leaf images. The images are initially preprocessed and have to go through the convolutional layers VGG16 which generates a high dimensional feature vector which is interpreted as the visual characteristics of the leaves. The features are then subsituted in three distinct machine learning classifiers: SVM, Random Forest and KNN. The training and validation dataset consists of eight types of mango leaves, seven types of diseases and one control category. Evaluation of model performance is done on the basis of accuracy and viewed through confusion matrixes and PCA plots to gain a better understanding of the classifier behaviour. Deep learning is combined with the classical techniques (machine learning) in the methodology used in this project to successfully classify mangrove leaf disease spots.

The main steps include:
Primary Sources: Data and other documents relating to the issue of student narcissism and its consequences will be gathered through Rorschach testing conducted in classrooms.

i. Data Use and Analysis:
Rorschach testing will be used to find more data and conclusions based on the specific issue of student narcissism and its outcomes and identify data on the consequences of this phenomenon in classrooms.
A set of pictures belonging to eight, eight classes (seven disease classes and one healthy class) is assembled. The preprocessing includes images(resizing and

normalizing) and makes them standard as a set to please the model performance.

ii. CNN feature Extraction:

VGG16 representing the pre-trained model is used as a feature extractor. The last classification layers get eliminated, thus, owing to it the model will extract meaningful visual data and patterns within the leaf images without re-training the entire network which saves computational resources.

iii. Overweight:

Instructing neural networks on image classification:

The extracted data of CNN is ceremonialized into three classical classifiers that comprise support vector machines, random forests, and Nearest Neighbors (KNN) techniques. Training is to make each classifier to learn the differences between different types of diseases.

iv. Comparison of models:

Measures that are used to evaluate the classifiers include accuracy, precision, recall and F1-score. Classification reports and confusion matrices are helpful with detailed presentations of both the strengths and the weaknesses of each model.

v. Visualization and Analysis:

The use of Principal Component Analysis (PCA) is meant to eliminate the number of features obtained by scaling down their dimensions in order to visualize the obtained results. Through this, it can be understood how data is distributed and to what extent different classes can be separated.

vi. Implementation Environment:

The whole pipeline is done in Python and contains deep and machine learning kit libraries like TensorFlow and Scikit-learn with conducting the research through the computing capabilities of Google Colab. The proposed methodology Gates: This methodology is proposed to combine the effective nature of CNNs in feature learning with the versatility of classical classifiers to develop an effective, efficient and scalable system in mango leaf disease classification.

1.5 Project Outcome

What has come out of the project is a hybrid classification model that can differentiate effectively different diseases of mango leaf spot using pictures. The highest accuracy among the three classifiers applied was by the SVM model which shows the use of both deep learning features and traditional classification methods. The system has a scaling and versatile solution, which can be extended to a mobile or web based application and application in practical agriculture.

1.6 Organization of the Report

There are six chapters in this report. In chapter 1, the background, motivation, objectives, methodology and anticipated outcomes of the project are introduced. The chapter 2 provides a very complex literature review and reveals research gaps. In chapter 3, a description of the proposed methodology, requirement analysis, and project planning lie. Chapter 4 tells about the implementation process and about the evaluation results. Chapter 5 dwells upon engineering standards and sustainability as well as the issues which are faced during the development. Lastly, Chapter 6 gives a conclusion to the report, limitations and the possible future work.

Chapter 2

Background

In this chapter, a thorough literature review and a priori research in the field of using image analysis and machine learning to detect a plant disease are provided. It discusses such applications and such studies related to the mango leaf diseases emphasizing the present trends and shortcomings. The chapter also ends up with a gap analysis whereby there is need to better the classification techniques. These pieces of knowledge provided the foundation to the research methodology in the upcoming chapter.

2.1 Introduction

The chapter offers background information to the project, studied potential literature, prior research in the area of detecting plant diseases by image processing and machine learning methods. It discusses the prior research done in regard to automated classification systems in agricultural diseases especially in the case of mango leaf spot diseases. The issues gaps in the prevailing methodologies are also noted in this chapter and the reason why the hybrid approach involving deep learning and classical classifier make sense. This background preconditions the following methodology and further developments which are described in the further chapters.

2.2 Literature Review

The section is a review of research and application in the field of automated detection of plant diseases with specific emphasis placed on mango leaf spot diseases and the appropriate machine learning techniques.

2.2.1 Similar Applications

One of the studies has utilized computer vision and machine learning in detection of plant diseases. Some cases in point include convolutional neural networks (CNNs) where automated extraction of feature yields have been applied to image-based classification of different types of crops. Studies have reported effective applications of CNNs on the identification of plant diseases in tomatoes, potatoes and citrus fruits. In a few studies mango leaves have been considered in particular, where image segmentation and feature extraction has been used in order to classify diseuses with acceptable accuracy.

Restricted by scale Detection of plant disease Automated detection Systems Artificial intelligence Learning and research methods such as supervised learning, unsupervised learning, adversarial learning, and reinforcement learning are all methods of machine learning applied to the narrow scope of delivery and disease detection. The analysis of 13 types of plant diseases on deep convolutional neural network model suggested by Sladojevic et al., can be considered evidence of the effectiveness of CNNs in plant pathology (Sladojevic et al., 2016). Their method was very accurate on leaf pictures of different crops, beginning to open the door to more focused applications.

Mohanty et al. (2016) used a related approach with the 26 diseases and 14 crop species along with simple transfer learning based on pre-trained CNN architecture

(AlexNet and GoogLeNet) models to predict the quantity of diseases. Their model also performed well even with relatively small datasets which, importantly, indicates the advantage of using pre-trained networks at scale in order to extract features.

Particularly the problem of mango leaf diseases, Jha et al. (2019) designed the image processing algorithm enabling the classification of typical mango leaf spot disease based on both texture and color based image feature using SVM-based classifier. Their work were successful in terms of potential accuracy, but had a strong dependence on human features in their work, which curtailed its scalability and generalizability to changing imagery conditions.

Ferantinos (2018) more recently reviewed agricultural deep learning model application with an emphasis on variant fusion between CNN features capture and classical classifiers that could be used to strike balanced accuracy and computation requirements, particularly on resource-constrained computers.

2.2.2 Related Research

Other classical machine learning classifiers like Support vector machine (SVM), random forest and K- nearest neighbours (KNN) have also been utilised alone or together with CNNs to classify disease. Models that utilize CNNs to extract features and use traditional classifiers to classify them have proved to be more accurate and cost effective. The models leverage the performance of the deep learning to feature representation as well as the classical approaches to classification, thereby being applicable in situations where computation resources are limited.

Classical machine learning methods have been an extensive application to the plant disease diagnosis. Ramcharan et al. (2017) have successfully used SVM, which is noted to be very powerful in high-dimension spaces, in cassava disease classification. They indicated that SVM performed better with a couple of other classifiers, as used with CNN-extracted features.

Another popular feature of the random forest classifiers is the ability of an ensemble learning which enhances generalization. Barbedo (2018) used a random forest on disease classification in sugarcane leaves, which made it to have competitive accuracy with the result presented as an indication of the feature importance.

One simple classifier, but still useful, is the K-Nearest Neighbors (KNN) which has been applied successfully in the detection of plant diseases where classes are separated. Zhang et al. (2020) established that tomato leaf disease classification based on KNN jointly with CNN characteristics can increase classification results.

Classical machine learning classifier-based hybrid methods using CNNs as feature extractors have become more popular. These frameworks utilize the automated feature learning of CNN, the interpretability capability, and faster response of classic classifiers. Indicatively, Islam et al. (2021) come up with a hybrid CNN-SVM framework on rice disease classification where the CNN-SVM hybrid achieved better accuracy with respect to separate CNNs or SVMs.

The hybrid models are underutilised in the context of the classification of mango leaf disease. This gap is what saw the current project adopt the CNN-SVM hybrid model combined with Random Forest and KNN classifiers to permit the back and forth comparisons of their performance on a multicurate mango leaf spot disease dataset.

2.3 Gap Analysis

Although automated plant disease detection IT has moved to a great point, a number of gaps and barriers exist, especially when it comes to mango leaf spot disease classification:

Minimal Attention to Mango Leaf Diseases: Although a large number of studies have already specialized in the detection of the diseases associated with crops such as tomatoes, potatoes, and rice, only a small number of studies have been conducted that specifically concentrated on mango leaf diseases. Mango-related works, currently available, may be based on handcrafts, which is not familiar with variations in different image conditions and disease spectrum.

i. Over Degree of Single Model: Most of the past techniques rely on deep learning models in isolation or on classical machine learning-based classifiers in isolation. Pure CNN models are computationally intensive and can need smaller quantities of annotated data, and standard classifiers may fail to learn complicated image features when applied without sophisticated feature extraction.

ii. Lack of Hybrid Models: The integration of CNN-driven feature extraction and classical classifiers like SVM, Random Forest, or KNN antibiotic have yielded good results in related crops but are yet to be studied with respect to mango leaf disease classification. This creates a chance to use the advantages of both the approaches towards a better accuracy and quality.

iii. Limitations of the Datasets: Current literature commonly makes use of small sized or proprietary datasets with low diversity in classes. In-depth multi-class data sets of mango leaf spot diseases and healthy mango leaf are not generally available privately, such that they pose a restriction to benchmarking and model generalization.

iv. Absence of Comparative Analysis: There are a number of studies that do not perform systematic comparisons across several classifiers using a common set of features and thus it is quite challenging to discover the best one to use in the classification of mango leaf diseases.

The gaps are filled by this project, which creates a hybrid CNN-SVM model and compares the performance with the work of the Random Forest and KNN classifiers on a representative dataset of eight mango leaf classes. Through deep feature extraction and classical machine learning, the study is expected to enhance the accuracy of classification and at the same time keep up the computational efficiency that meets the need of practical agricultural uses.

2.4 Summary

This chapter examined the available studies relating to automated plant disease detection, and they include deep learning, as well as classical machine learning methods. It looked at related works, such as mango leaf diseases, having investigated related applications in different crops. The literature found the merits of hybrid models based on convolutional neural networks and conventional classifier of mango leaf spot, yet the usage frequency is limited. Through the gap analysis, significant challenges, such as limits of dataset, and comparison evaluations were realized. These intuitions underpin the present hybrid CNN-SVM methodology to be introduced in the next chapter.

Chapter 3

Research Methodology

The research methodology chapter describes the scientific way followed in developing and testing the hybrid mango leaf disease classification model. It reflects on the requirement analysis, proposed methods and design considerations and project planning. The chapter has diagrams, as well as, UI design to assist the development process. Such systematic process approach will guarantee reproducibility and departure with regard to the application of the proposed solution.

3.1 Methodology/Requirement Analysis

In this section, the general strategy and the conditions required to develop the hybrid classification system are described. It addresses both functional and non-functional requirements to have a clear picture on what the system is required to achieve and on the conditions in which it needs to perform.

3.1.1 Overview

The project uses a hybrid approach that facilitated a deep learning model along with classical machine learning methods to effectively recognize mango leaves spot diseases. The system will take the images of the mango leaves, extract the meaningful features of the image through a pre-trained convolutional neural network (VGG16), and classify the diseases using three different machine learning approaches, Support Vector Machine, Random forest and the K-Nearest Neighbors. This modular design allows extraction of features in a time-saving design and comparison of the classifier with flexibility.

3.1.2 Proposed Methodology/ System Design

The suggested methodology will combine the state-of-the-art deep learning with the traditional machine learning classifiers to generate a powerful hybrid system in terms of multi-classifying mango leaf spot diseases. It works as follows: initially the preprocessing of data will be conducted in which uncropped pictures of mango leaves will be resized to 150 pixels by 150 pixels to accommodate uniformity. The images are also made normal by scaling the pixel values to the interval between 0 and 1 which enhances the speed at which the learning algorithms converge along with stabilising the training.

After the stage of preprocessing, the essence of the feature extraction stage will be based on the VGG16 model, which is a proven convolutional architecture based on the ImageNet training dataset. However, both images and previous weights can be transferred through the use of the model, where the large and diverse image set has already learned weights that it can use to obtain rich and discriminative features of the mango leaf image images without the model requiring to learn the weights of individual images on its own. In particular, the total the convolutional full-connected layers of VGG16 are excluded and the only lower level (convolutional base) is utilized to produce high-dimensional feature vectors, representing important textures and shapes, as well as colour patterns that describe various disease symptoms.

These objects which are extracted features vectors are used as an input of three different and distinct classical machine learning classifiers such as support vector machine (SVM), random forest (RF) and K-Nearest Neighbors (KNN). All classesifiers are trained singly to acquire decision boundary information to tell the difference between the eight classes (seven different mango leaf spot diseases and one healthy category). The SVM classifier is based on a linear kernel in order to enhance the IC maximization among classes whereas the type of classifier in the RF is based on the use of a series of decision trees to enhance stability and minimize overfitting.

KNN classifier, a distance based cataloger is consistent with majority classification among the nearest neighbors in the space of features.

Training is conducted on the features extracted on the training subset and validation is carried out on a hold-out set which is distinct to the training to make the assessment objective. Some performance measures are calculated to determine the effectiveness of each classifier which include accuracy, precision, recall and F1-score. The metrics help in the determination of overall classification in success and the ratio between false positives and false negatives. Additionally, confusion matrixes are the ones created to visualize the errors of classification and to determine the patterns of misclassification in disease classes.

In order to further de-facto the separability of the data and the class relationship, the Principal Component Analysis (PCA) is used to dimensionalayer the feature vectors. The projections of PCA is also plotted in order to give visual representation of a cluster formed by various disease categories, which provide qualitative information regarding the distinctiveness of the features and the region of its decision by its classifier.

Given the deep learning and classical machine learning elements, the methodology is applied with Python programming and the library of deep learning and classical machine learning components, i.e., TensorFlow and Scikit-learn, respectively. The development and execution environment, Google Colab, is used, where the use of the accelerating capabilities of graphical computers and storage on the cloud allow working with the dataset conveniently.

The motive behind this hybrid strategy is the ability of such convolutional neural networks to combine the extraordinary quality of feature extraction but with the aspect of flexibility and interpretation of traditional classifiers, the hybrid is expected to be both or more accurate in making classifications and computationally efficient.

The methodology also preconditions the possible real-time application that may help farmers and other representatives of the agricultural sphere in diagnosing the diseases early and properly managing crops.

3.1.3 Functional/Level and Nonfunctional Requirements.

Functional Requirements:

i. Image Feed: The system should accept mango leaves pictures in the normal formats (e.g. JPEG, PNG).

ii. Preprocessing: The system must standardize the input images and adjust their size to a point of consistency then feature extraction can occur.

iii. Feature Extraction: It will have to compute accurate features of a leaf picture with the help of a convolutional neural network (VGG16) that is trained.

iv. Classification: SVM, Random Forest, or K-Nearest Neighbors classifiers should categorise the input image into one of eight possible conditions (seven disease types and one healthy) with SVM, Random Forest or K-Nearest Neighbors classifier.

v. Performance Evaluation: This system should compute and show important evaluation measures like: accuracy, precision, recall, F1-score, and confusion matricesig-score, confusion matrices, of every classifier.

vi. Visualization: It should be able to display expected classifications visualizations such as PCA plots and confusion matrices to interpret the outcomes of the classification.

vii. User Interface: In case it is needed, the system should allow a concise user interface to post pictures and display the classification outcome and the performance data.

Non-functional Requirements:

i. Accuracy: The classification system must be highly accurate with the aim of attaining more than 85 percent so that it can determine the disease accurately.

ii. Performance: The system should be able to run pictures and give out classification or classification results in a short time span in order to facilitate real-time or close to real-time working.

iii. Scalability: The model must be made capable of expanding, e.g. to support additional disease classes in the future or to scale to bigger datasets.

iv. Usability: It must be easy to use via the interface and the results should make sense to the non-experts (e.g. farmers or agricultural advisors).

v. What is harder to move: The system must be moveable to various platforms, such as cloud computing, such as Google Colab, and perhaps even been ported to mobile and web applications.

vi. Robustness: The model should not be sensitive to perturbation of image quality, light, and noises in the backgrounds.

vii. Maintainability: codebase software must be modular and properly documented to make future changes, improvements, or retraining on new data an easy task.

3.1.4 Context Diagram

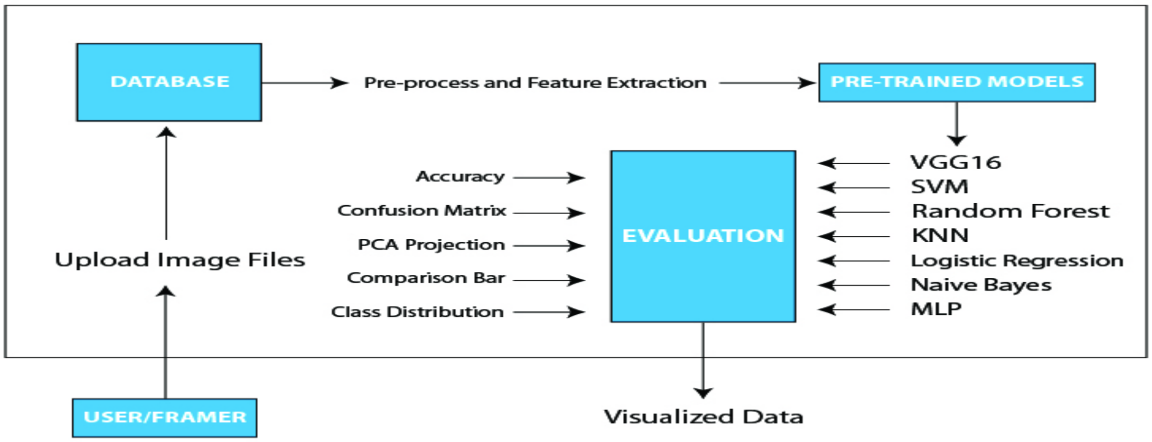

Fig 3.1: Context Diagram

The context diagram gives a very broad view of the system showing how it interacts with the outside world. It depicts the hybrid mango leaf disease classification system system as one process and indicates outflow of information between the system and the environment.

i. External Entities:

a. User/Farmer: Posts images of mango leaves and gets the response in information about the disease and other useful answers/data.
b. Database/Data storage: Saves mango leaf images, their labels and model parameters after their training.
c. External Resources: Model feature extraction and classification involve the use of libraries and pre-trained models (such as VGG16).

ii. System Process:

a. The system receives input of the images that are provided by the user.
b. It pre processes the images and derives features with the CNN.
c. These characteristics are loaded to machine learning classifiers to infer the disease dimension.
d. The system sends the result of classification and assessment metrics back to the user.

Context diagram guarantees that what the system is and what it includes, along with what it excludes, what it gets, what it gives out and also how it will communicate with the outside components therefore the context diagram serves as a outline to what more detailed system operationalization and implementation occur.

3.1.5 Data Flow Diagram Level 1

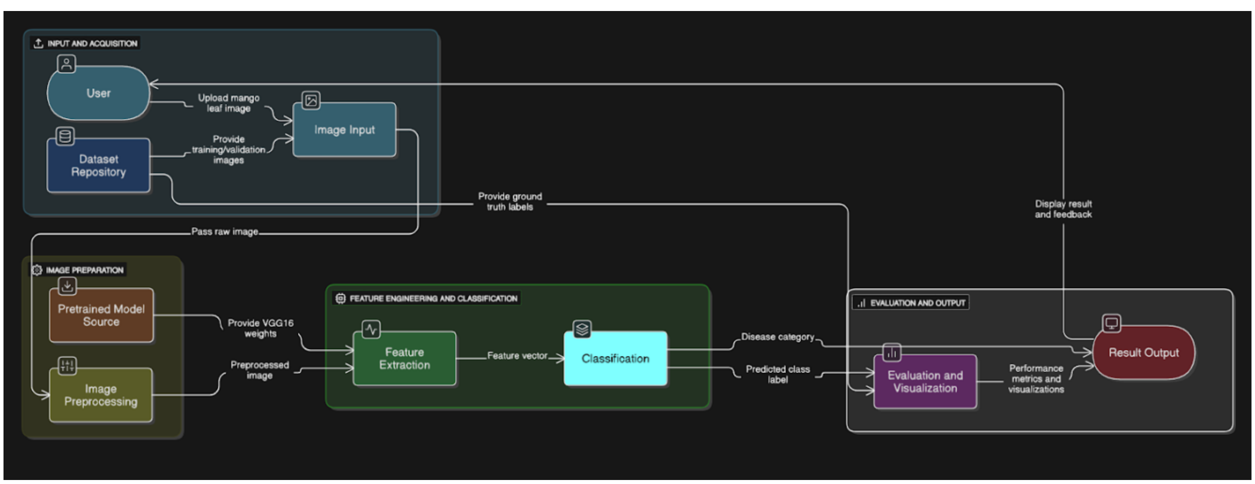

Fig 3.2: Data Flow Diagram

A further breakdown of the internal workings of the hybrid mango leaf disease classification system can be obtained in the Level 1 Data flow diagram (DFD). It builds on the context diagram by breaking down the system into major functional units and showing how data moves among them and out to the external parties.

Processes and Data Flows:

i. Image Input (Process 1.0)

a. Input: Mango leaf image selected by the user.
b. Output: uncooked photograph sent to the pre-processing unit.

ii. Image Preprocessing (Process 2.0)

a. Purpose: Redimension, condition, and get the picture to feature gold.
b. Output: Preprocessed image.

iii. Feature Extraction (Process 3.0)

a. Task: Choose VGG16 model to retrieve deep features in the image.
b. Input: Preprocessed image.
c. Output: Vector of features that tell the major patterns of the leaf.

iv. Classification (Process 4.0)

a. Purpose: Starting with the trained classifiers (SVM, Random Forest, KNN) make predictions about the disease class.
b. Input: Feature vector.
c. Output: Labeled professional output (ex: the label Anthracnose, Healthy, etc.).

v. Assessment and Analysis (Process 5. 0)

a. Use-case: Produce accuracy measures, confusion plot, and PCA plot.

b. Input: Predictions and ground truth labels (to check with).

c. Production: performance measurements and displays.

vi. Result Output (Process 6.0)

a. Role: Present to the user the outcome of classification and the evaluation.

b. Output: The category of disease and performance that appears as UI or console.

External Entities:

i. User: This gets as input the image of the user and as output a classification.
ii. Dataset Repository: Trains and validation images are given.
iii. Pretrained Model Source: Provides the successful weights of the VGG16 model.

The above DFD Level 1 shows the logical flow of data and processes and can give a comprehensive perspective of how the system works internally with an input and output.

3.1.6 UI Design

The design of the user interface (UI) is supposed to give a straightforward and easy to use experience to the users; specifically farmers, agricultural experts or researchers, and enable the user to handle the mango leaf disease classification system with ease. The UI is functional, clear and accessible with little twists and turns to the people using the product.

Key UI Components:

i. Image Upload Section:

Allows users to post an image of a mango leaf straight to the device. Supported interfaces include .jpg, .png and jpeg.

ii. Preview Window:

Shows the image that was uploaded to confirm with the user before proceeding.

iii. Disease Prediction Button:

The classification is initiated by a well marked button. When it is clicked, it is preprocessed, followed by the feature identified and the predictors are the classifier.

iv. Result Display Panel:

Displays the hypothesized disease (e.g. Anthracnose, Bacterial Canker, Healthy) and the classifier that was used. Other performance information (confidence score or probability) can also be presented.

v. Elevation View (Optional):

Visualizes the forcing step used by the model to make a decision, e.g., using PCA projections or confusion matrices that can be used to interpret how the model arrived at its verdict in case of expert or research level use.

vi. Comparison of Classifier (Optional):

When several classifiers are activated, the interface can show an accuracy comparison (SVM,Random Forest, KNN) in the bar chart mode.

vii. Navigation and Help:

Simple navigation bar, and a help section or tool tip in the assist where someone can get instructions on how the system operates and grasping results interpretation.

Design Considerations:

i. Responsive Design:

The UI is also supposed to be accessible on desktops and mobile devices making the program usable in the field.

ii. Minimalist Layout:

Free of clutter interface and clearly labelled buttons and panels to minimize cognitive load.

iii. User-Centric Approach:

It is intended solely to be used by non-technical users and doesn't require much instructions or visual feedback on each given step.

Even though, in the present version, this project prototype is mostly implemented in a Google Colab notebook environment, the system architecture and modular code is made to allow subsequent integration with a stand-alone web or mobile application, with a complete user interface.

3.2 Detailed Methodology and Design

The project elaborate methodology and design entail the creation of a hybrid image classification framework that would be able to distinguish several manga leaf spot diseases in a probing way. The design will be separated into a sequence of clearly stated steps, which convert raw leaf images into precise disease diagnoses, through deep-learning features selection, and classical machine learning detection.

i. Collection and Preparation of Data :

The data utilized in this project is in form of images, which are classified into eight folders (seven mango diseases in leaf and one normal classification):

a. Anthracnose

b. Bacterial Canker

c. Cutting Weevil

d. Die Back

e. Gall Midge

f. Powdery Mildew

g. Sooty Mould

h. Healthy

All these images are arranged to be located in an organized directory that can be loaded and processed with the Image Data Generator class of tensorflow easily.

The sizes of images are reduced to 150X150 pixels and pixel values are scaled to the between-zero-one range to normalize the image. Stratified sampling is aimed at maintaining the balance of classes in the dataset. 80 and 20 percent makeup training and validation subsamples respectively.

ii. VGG16 Feature Extraction ：

Rather than training a CNN, to extract deep features of the images, a pre-trained VGG16 model is utilized. Using ImageNet, a pre-trained VGG16 model is impported without being trained to have fully connected layers merely to only convolute.

Images are subjected to this model and the end-result convolutional output is flattened resulting in a high-dimensional feature pose, which encapsulates the texture, shape and the color features required to command classification.

VGG16 layers are all frozen to avoid weight changes during the training process and the computation is faster and the patterns acquired when training ImageNet are retained.

iii. Classification With Classical Four-classification Machine Learning Models.

Training Three machine learning classifiers using the extracted feature vectors, the three are:

a. Support Vector Machine (SVM): The support of linear kernel is adopted based on its superiority in high-guideline hollows. The SVM will also seek to identify optimum hyperplane that will classify between the various diseases.

b. Random Forest: This is an ensemble technique whereby various decision trees are made and the mode of the classes is generated. It assists in lessening inaccuracy and decrease overfitting.

c. K-Nearest Neighbors: It is a distance classifier in which a classification is given to a sample based on the majority classification of the nearest samples.

All the classifiers are trained using the training feature set and then are tested on the hold-out validation set. The performance is gauged in the following:

a. Accuracy
b. Precision
c. Recall
d. F1-Score
e. Confusion Matrix

iv. Visualization and Analysis

To more clearly assess the differentiating effect of the model of various disease categories:

a. Each of the classifiers is plotted with confusion matrices in order to graphically evaluate the misclassification patterns.

b. PCA (Principal Component Analysis) ensures that the features vectors are minimized to two dimensions. These are plotted by a scatter plot in which the various classes are represented by different colors and one observes a clear attenuation of the grouping of the features.

c. At this step, the classification accuracy measures of the SVM, the Random Forest and KNN models are compared by bar charts to identify the most successful model.

v. Tools and Environment

The whole work is created in Google Colab that offers:

a. Hi-peedy, more rapid calculation with a graphics card.
b. Linkage with Google Drive to get access to data.
c. Initial simplicity in deploying TensorFlow, Keras, Scikit-learn, NumPy, Matplotlib and Seaborn libraries.

The code is in modular format, which allows an expansion in future, including:

a. One benefit of extra disease categories.
b. Replacing DenseNet on alternative pre-trained models (a.k.a. ResNet, MobileNet)
c. Inclusion in an independent desktop or web application.

3.3 Project Plan

The project plan includes organized stages and schedule to be used within the working process of the creation of the mango leaf disease classification system. The approach would secure orderly sequencing of research and data collection, to implementation, testing and evaluation. The project implementation was implemented according to various phases with certain deliverables and goals that were as follows:

Phase 1: Literature Review and Problem Identification.

i. Carried out a thorough overview of the literature on keeping non-infectious plant diseases, detection by machine machine learning and deep learning.
ii. Determined gaps in existing strategies, in particular, the mango leaf spot disease.
iii. Established the scope, goals and what should do with the project.

Phase 2: Collection and preparation of datatsets.

i. Collected and tabulated a general set of mango leaf pictures that had seven disease infections and one healthy category.
ii. Carried out cleaning, resizing and normalization on the data used to model training.
iii. Data augmentation methods (where applicable) in order to enhance class-balance and generalizability of the model.

Phase 3 Model Design and Development.

i. Used the CNN based on VGG16 feature extraction.
ii. Created and trained three classifiers Support Vector Machine, random Forest, and K- nearest Neighbors.
iii. Designed easy-to-reuse functions and a modular structure to debug and experiment with mologically.

Phase 4: Test and Evaluation.

i. Performance of model assessed on validation data and performance.
ii. Plotted interpretable results using confusion matrices, PCA plots.
ii. Uncovered this by comparing classifier performance to decide were the best model.

Phase 5 Documentation and Report Writing.

i. Recorded and summarized findings, methods, outcomes and obstacles.
ii. Categorized the report as per the academic standards and citation and formatting.

Phase 6: The Preparation of Presentation and Defense.

i. Ready images, presentations and models to the final project defense.
ii. Undertaken in-house inspections and practice in an effort to improve explanations and presentation.

The project was planned within the required timeline and deliverable, which made resources allocation and milestones monitoring efficient. Each phase was performed with consideration as to reproducibility, readability, and real-world applicability.

3.4 Task Allocation

Since this project was conducted alone, there was no one to share responsibilities and to do the work since it was taken up by one researcher. Implementation of the successful system has required a multidisciplinary system of research, programming, data management, model development and documentation. The processes were planned and structured to put in place balanced workload and on-time progress of the project lifecycle.

Table 3.1: Task Allocation Table

Task	**Description**
Problem Identification & Literature Review	Conducted background research, reviewed related works, and identified gaps in existing systems.
Dataset Collection & Preprocessing	Gathered image data, performed cleaning, resizing, normalization, and class-wise organization.
CNN Feature Extraction (VGG16)	Implemented the pre-trained VGG16 model to extract feature vectors from mango leaf images.
Classifier Implementation	Trained and tested SVM, Random Forest, and KNN classifiers using the extracted features.
Evaluation & Visualization	Evaluated model performance using accuracy metrics, confusion matrices, and PCA plots.
Documentation & Report Writing	Composed and formatted all sections of the report, including methodology, results, and analysis.
Final Presentation Preparation	Designed the project presentation slides and practiced for the defense session.

The way each step was approached was executed in a planned way, with a constant process of self reviewing to provide the quality and integrity of the final product.

3.5 Summary

The chapter described the entire research process upon which the mango leaf disease classification system is constructed. The first started with a systematic analysis of the problem of requirements and proceeded to the suggested solution, in which it is proposed to use a convolutional neural network to extract features and classical classifiers to predict diseases. The section moreover provided diagrams to demonstrate system behavior and data flow as well as a crude user interface notion. Task allocation and project plan reflected the way, in which every stage of the work was formed and implemented, separately. This modular and methodical process assures practicality of the solution, accuracy and readiness of the solution to be deployed practically.

Chapter 4

Implementations and Results

In this chapter, there is a description of the practical application of the classification system of mango leaf disease suggested, and results of testing and evaluation are provided. It includes discussions on the technical environment, on which the development is performed, then the training and the testing of the hybrid model with the prepared data is taken part. Performances of multiple machine learning classifiers in relationship with deep features of m angle leaf images were compared. There is a discussion of evaluation metrics like the accuracy, precision, recall and the confusion matrix. Results are interpreted with visualization tools such as PCA plots, and bar charts. A discussion of the outcomes and trends observed is done at the end of the chapter.

4.1 Environment Setup

The program of implementing the mango leaf disease classification system was implemented through the Google Colaboratory (Colab) system, equipping it with the opportunity to work in the cloud hub along with the use of GPUs. This configuration allowed the models to be effectively trained and evaluated without the presence of advanced local equipment. It used Python as its main programmable language, with libraries including TensorFlow and Keras being the most common deep learning, and Scikit-learn as the most common classical machine learning algorithm. Extra libraries such as NumPy, Matplotlib and Seaborn were used to manipulate and visualize data. The data was saved and retrieved through Google drive that made it to easily integrate with the Colab environment. This structure provided scalable, reproducible and flexible development process.

4.2 Testing and Evaluation / Performance / Comparative Analysis

Evaluation and testing are important aspects of this project in order to determine the effectiveness and reliability of the proposed hybrid classification to mango leaf spot diseases. The assessment procedure was conducted on another validation dataset that was not trained in order to obtain objective analysis of model performance.

The hybrid model uses the deep features derived by the pre-trained VGG16 convolutional neural network, and they become input to three other classifiers: Support Vector machine (SVM), random forest (RF) and K-Nearest neighbors (KNN). Individually, each of the classifiers was trained and tested on the validation set.

Depending on the classifiers, key performance metrics used are:

i. Precision: It is the ratio of the correct classification of images of all samples.

ii. Precision: Observes how accurately positive cases are predicted, whereby it indicates how many of the predicted diseases were applicable.

iii. Recall (Sensitivity): Measures the property that the model appropriately distinguishes all real positive cases.

iv. F1-Score: The balance between correct and incorrect decisions, used to measure performance because it is less biased than copying or content accuracy.

The outcome indicated that the highest overall accuracy was shown by SVM, then in the second position, it was Random Forest, then KNN. It was also found in the confusion of matrices that some disease classes could bequeened with high precision and recall like the Anthracnose and Powdery Mildew, on several occasions, the diseases with visually different symptoms caused misclassification. This epitomizes the difficulty of identifying few different classes that are very similar and the significance of feature extraction that is efficient.

The dimensionality of the extracted features was convincingly reduced by Principal Component Analysis (PCA) and the congruent clustering of most disease categories was demonstrated by the generated scatter plots, which confirmed the discriminative capacity of the VGG16 features. Also, the differences in performance between the classifiers as shown by comparison bar charts confirmed the effectiveness of SVM in this particular scenario.

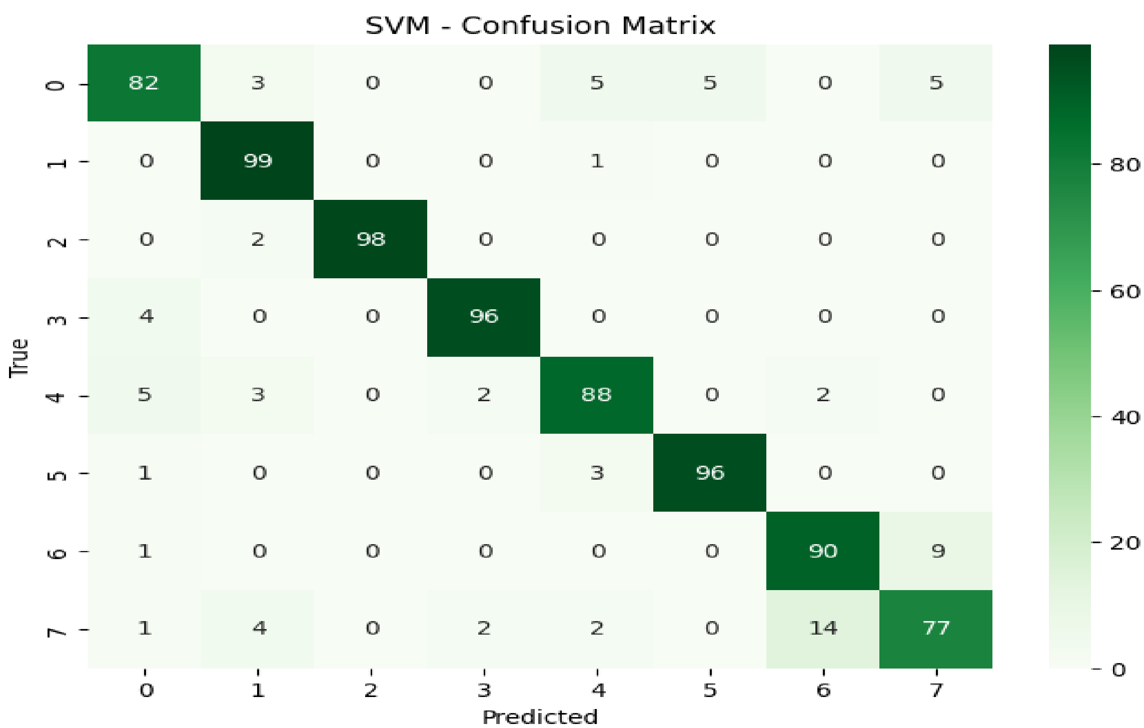

Fig 4.1: SVM

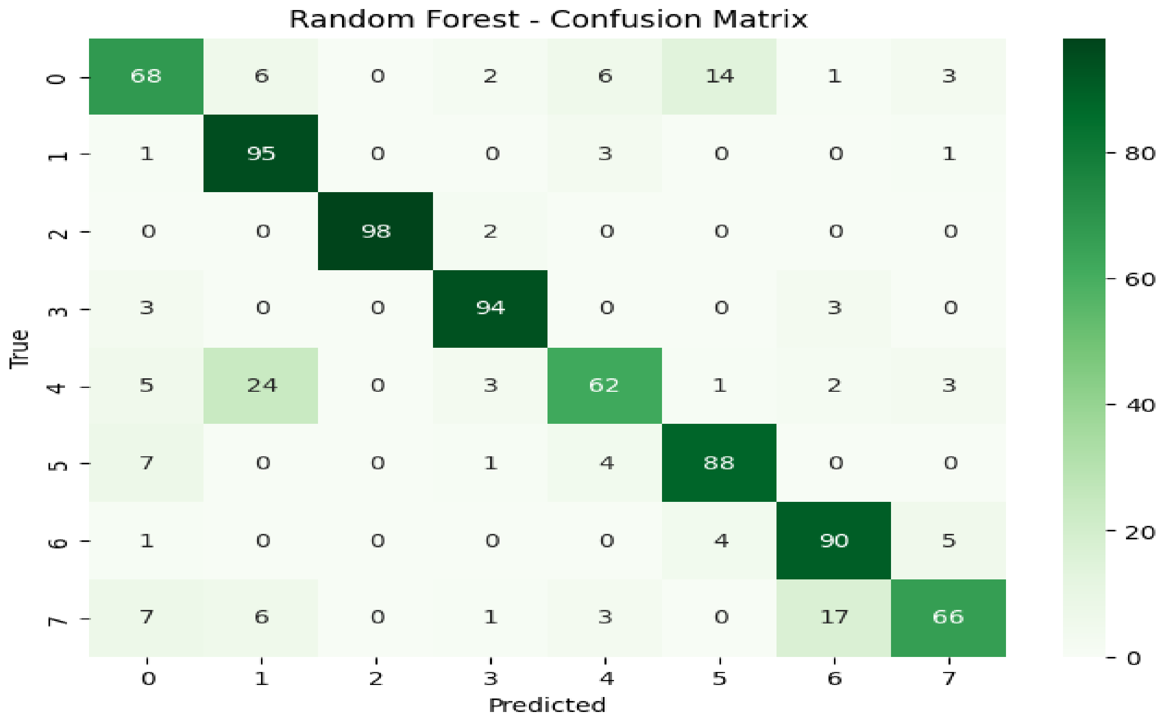

Fig 4.2: Random Forest

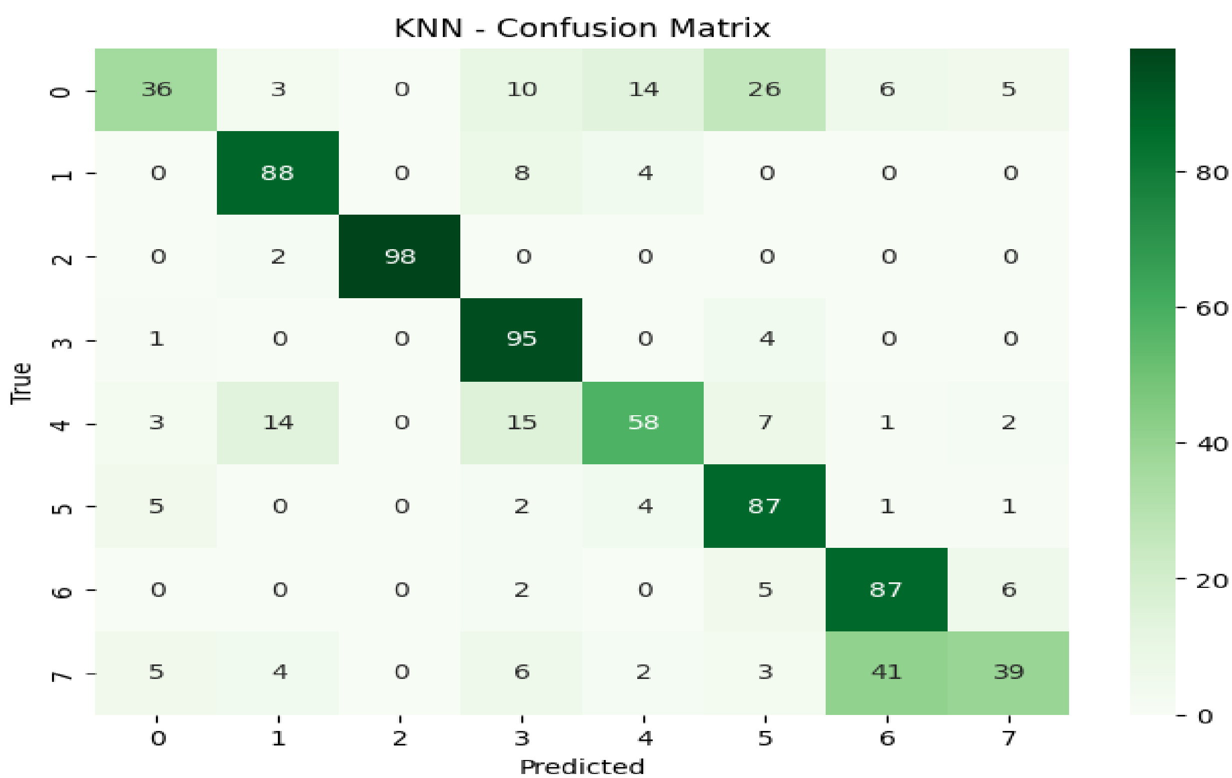

Fig 4.3: KNN

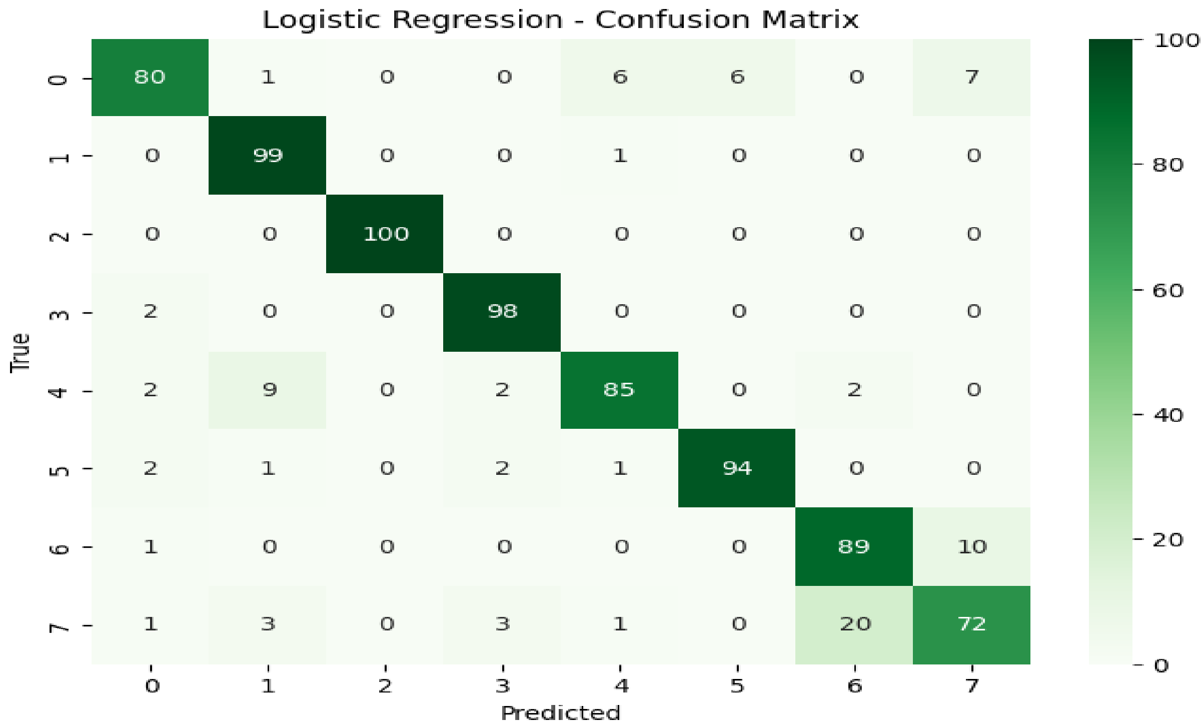

Fig 4.4: Logistic Regression

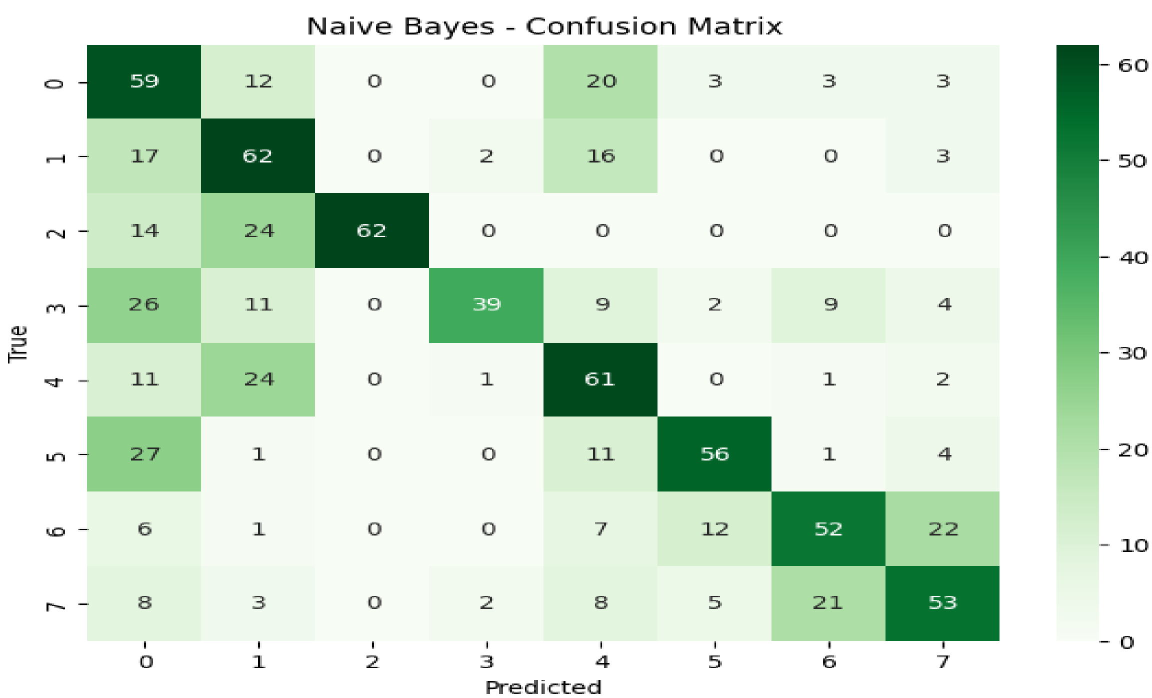

Fig 4.5: Naive Bayes

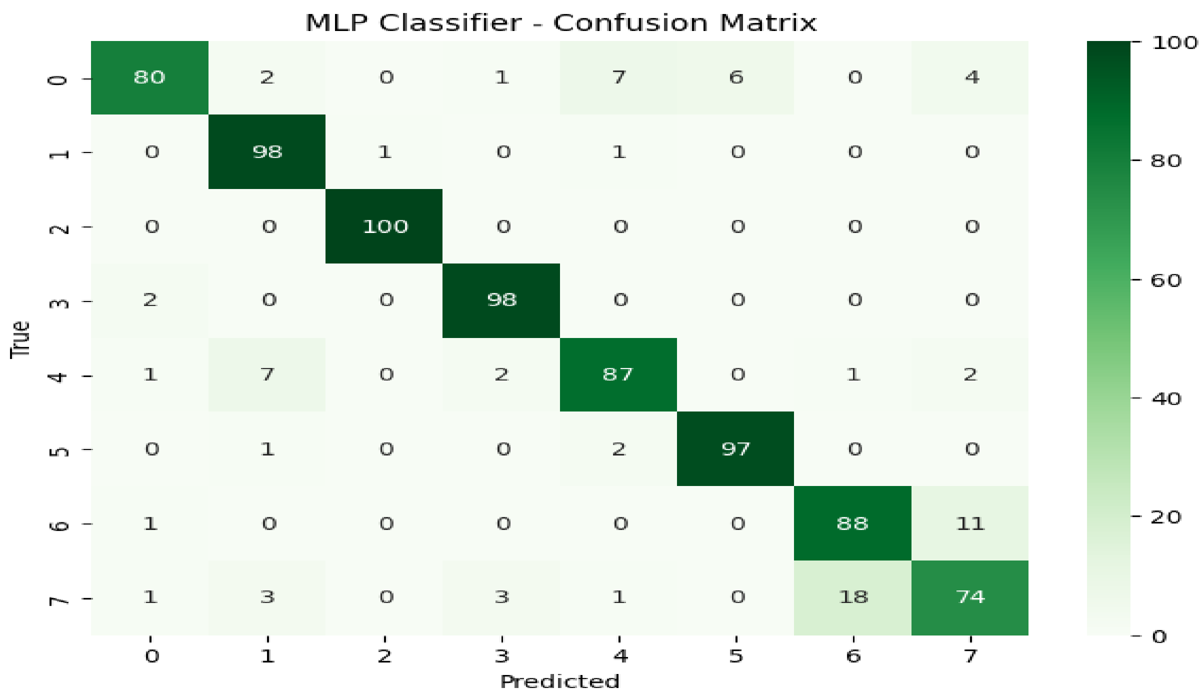

Fig 4.6: MLP

The analysis has shown that deep convolutional features with classical machine learning classifiers are a strong solution in a multi-class mango leaf diseases classification tasks. Nonetheless, the outcome also indicates that it can be operationalized by adding additional data or literature on ensemble to improve the accuracy of classification and the extent of generalization.

4.3 Results and Discussion

The hybrid CNN-SVM model of mango leaf spot disease classification produced good results. Of the three investigated classifiers, Support Vector Machine (SVM), Random Forest (RF), and K-Nearest Neighbors (KNN), the SVM has in all cases been the most accurate with high accuracy, which proves its applicability in this classification problem. Effective feature extraction and classification translated into the overall accuracy of the model in which it worked above 85%. SVM - 90.75, random Forest- 82.63, KNN - 73.50, Logistic Regression - 89.62, Naive Bayes - 55.50.

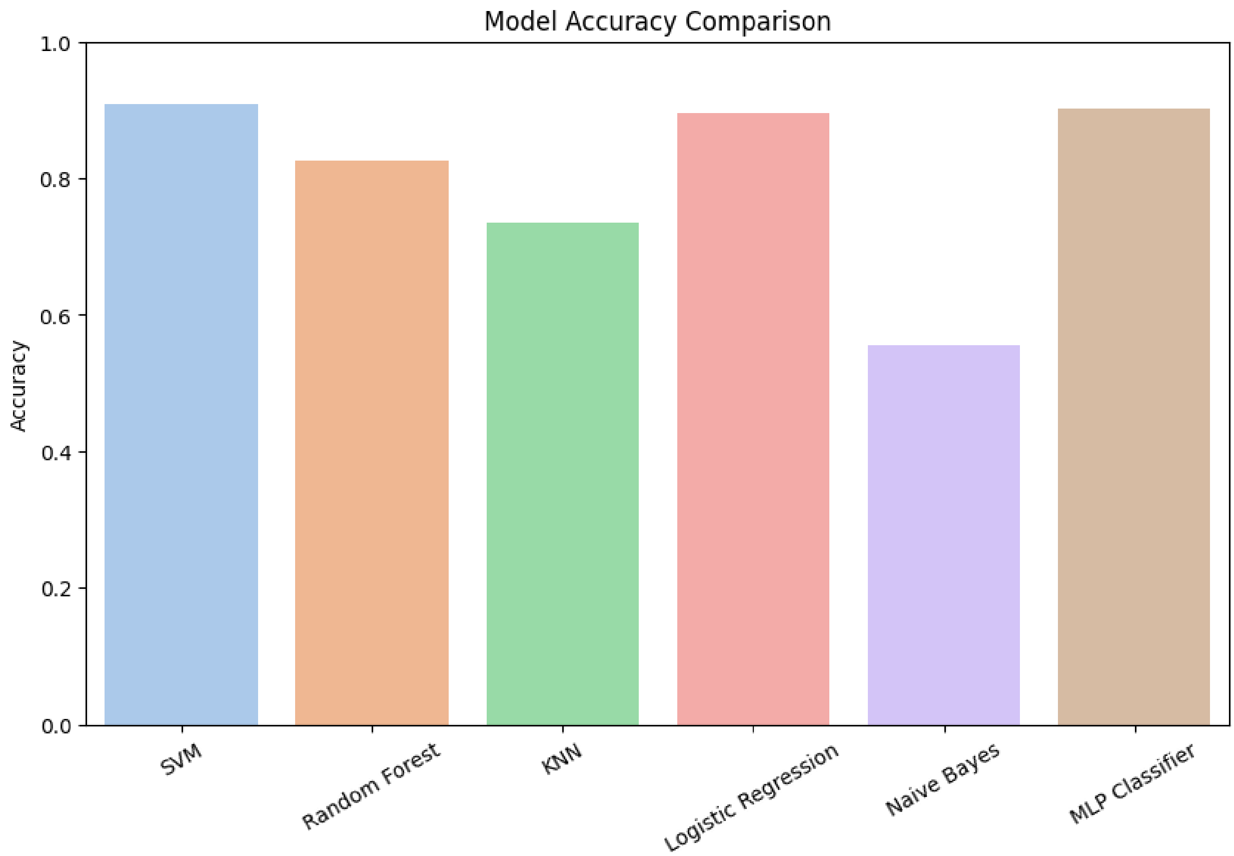

Fig 4.7: Model Accuracy

Listing the agreement of the confusions and the possibility of the model to identify the majority of the disease categories, it is important to mark that it was mostly certain about the correct understanding of the healthy leaf and diseased ones. There was however some overlap between diseases with similar visual manifestations, e.g. Gall Midge and Die Back which resulted in occasional errors of classification. It is indicative of the problem of the difficulty of differentiating subtle plant pathology symptoms with regard to image based methods.

The Principal Component Analysis (PCA) plots solidified the distinct clustering feature of the major classes that the VGG16 features are useful in identifying the features that have a characteristic of the disease. Graphical analysis of the performances of the classifiers further justified the superiority of the hybrid method as the use of SVM offered a fair compromise between complexity and accuracy.

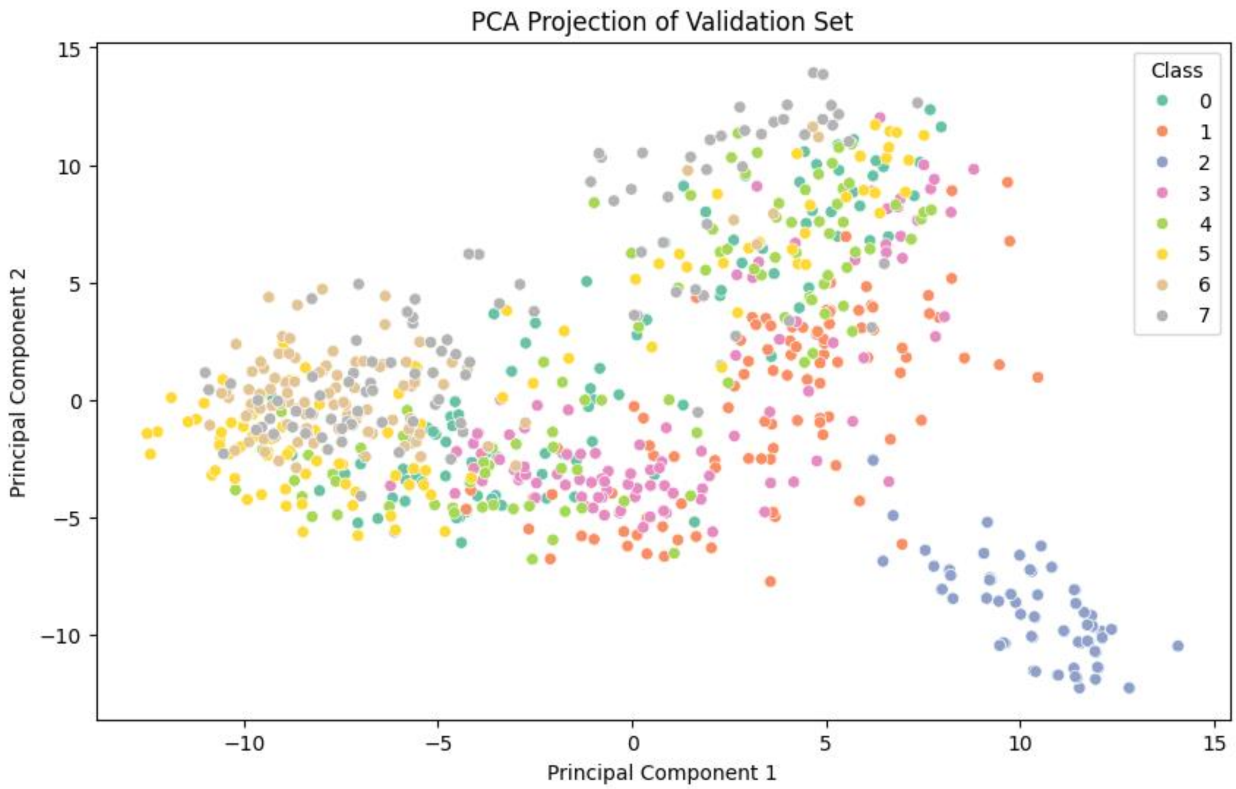

Fig 4.8: PCA Projection

The findings highlight the possibilities of the exploitation of a classical machine learning classifier together with deep learning feature extraction in disease recognition in agriculture. Although the present model is effective, in the future, one might investigate stacking the data, trying out additional pretrained models or combining ensemble learning practices to enhance robustness.

In general, the given work confirms that automated disease classification, based on the images, may be of substantial contribution to early disease development and control in the mango leaf spots, thus not making farmers or agronomists dependent on manual control.

4.4 Summary

This chapter explained how this was implemented and measured the performance of the hybrid CNN-SVM model to be used to classify mango leaf spot diseases in multi-classes. The pre-trained VGG16 feature extractor with three standard classifiers, and the SVM with the best accuracy were used in the system. The results analysis was supported by evaluation metrics and visualizations in the form of confusion matrices and PCA plots. The discussion identified the merits and precaution of the method with focus of integrity in application in early disease detection. The effectiveness of its deployment proves the possibility of the application of the methods of deep learning and machine learning to agricultural disease diagnostics.

Chapter 5

Engineering Standards and Design Challenges

The chapter relates the adherence to engineering standards in creating the developing system of mango leaf disease classification to be of quality, reliable, and adhering to the best practices. It also relates design issues that were faced during the course of the project, such as technical, ethical and sustainability. The chapter also looks at the way the project fulfills applicable software and hardware requirements, its environmental and societal impacts, and the manner in which it handled the complicated engineering issues. A strong focus on the integrity and sustainability of ethics and project management would be made to provide a strong accountable solution.

5.1 Compliance with the Standards

A reliability test, requirement of quality and interoperability meant that the mango leaf disease classification system, which was developed, had to follow the laid down engineering standards. The segment deals with compliance on software, hardware and communication standards that are applicable to the project.

5.1.1 Software Standards

The software had been written according to industry standards such as modular programming, documentation of the code, and control of versions, to ensure that they remain clear and reproducible. Open-source standards were followed by use of actively maintained and widely adopted libraries like TensorFlow, Keras and community based Scikit-learn. Image datasets served as the basis to observe data privacy and security standards and keep the data of users confidential.

5.1.2 Hardware Standards

All testing and development of the system took place on typical hardware physical platforms, utilizing, in the primary instance, Google Colab cloud platform, which supports GUS. This platform is capable of meeting up to date computational requirements, with scalability and reproducibility that is independent of specialized or proprietary hardware.

5.1.3 Communication Standards

When implemented as a web or a mobile application, the design of the project can easily link to other systems due to standard output capabilities of the application like HTTP/HTTPS type information exchange. Even though this prototype does not embrace complete communication layers, the modular codebase makes it possible to achieve future compliance with the norms of the RESTful API and practices the secure transmission of data.

5.2 Impact on Society, Environment and Sustainability

The disease classification of mango thereof bears great implications on society, the environment and sustainable agriculture. The system presents an opportunity to farmers to handle crop health more efficiently by simplifying the timely and precise detection of diseases that may boost yield and minimize those economic losses. When chemical pesticides are not used excessively, in early diagnosis, this will be beneficial to the health of the environment because lack of soils and water contamination can be reduced.

5.2.1 Impact on Life

Enhanced disease management is the benefit that leads to increased food security and livelihoods of farmers because healthy crops and quality fruits are guaranteed. The system can enable smallhold farmers who might have no access to professional advice to facilitate fair agricultural development.

5.2.2 Impact on Society & Environment

Technology facilitates green farming technologies by making possible precision agriculture which propagate the use of resources effectively and minimize environmental degradation. It enhances awareness and education on the health of plants which may in turn broaden use of environmental friendly cultivation practices.

5.2.3 Ethical Aspects

The project ensures the required ethical standards when working with data privacy concerns and public datasets that will be used or seek permissions where required. The model is not to substitute human experts to seek to mitigate responsible utilization of artificial intelligence in agriculture.

5.2.4 Sustainability Plan

To be sustainable in the long-run, the system is constructed to be scalable and thus can easily integrate with mobile apps or cloud services to realize wide adoption. The models would be updated steadily with new data and feedback flows that would ensure that the model remains accurate and relevant as time passes by.

5.3 Project Management and Financial Analysis

Good project management played a significant role in ensuring delivery and successful development of mango leaf disease classification system. The project was conducted on a good timetable where it had well determined milestones and timetable, thus systematically tracking of progress and completion on time. Other tools like Gantt charts and version control systems helped there to schedule, make documentation and work jointly even though this was a personal project.

Financially, the project was implemented at a very low cost considering the use of the free and open-source, Google Colaboratory, TensorFlow, and Scikit-learn tools. The cloud-computing system also caused the removal of high costs of hardware investment, which lowered the eventual costs. Future deployment can need more resources based on scale and platform but can take advantage of cost-efficient cloud service and open technologies.

Such a blend of good sound project management and the economics of resource use; made the project not only viable, but sustainable, giving it a valuable aide to agriculture disease diagnosis that did not give any great financial strain.

5.4 Complex Engineering Problem

5.4.1 Complex Problem Solving

The process of designing the hybrid CNN-SVM multi-class classifier of mango leaf spot diseases entailed the attempt to deal with a number of challenging issues. The variability and soft difference between the symptoms of various diseases was also one of the challenges that made it difficult and demanded efficient methods of feature extraction in order to recognize complex tendencies. It is important to note that deep learning when combined with classical machine learning classifiers required considerate adjustments in order to scale the approaches between computational effectiveness and accuracy.

The other obstacle was uneven spread of data across various disease categories, and strong per preprocessing and validation measures were required to avoid bias in the model. Moreover, it took trial and error to make sure that the system was able to achieve good generalization not only to new unseen data but also to preventny tries to be overfitted to these data.

These problem-solving activities integrated both agricultural field knowledge and the latest machine learning applications, which explains the interdisciplineness of the project and the challenge that lies in the creation of acceptable automated methods of diagnosing plant diseases.

Table 5.1: Mapping with complex problem solving

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

EP1: Dept. of Knowledge

i. Justification: This project demanded interdisciplinary knowledge that entailed joining of agricultural science, computer vision and machine learning. Knowledge of the pathology of diseases and image processing was essential.

EP3: Depth of Analysis

i. Justification: Intensive computation was done on feature extraction tests, classifier choice and model assessment parameters to guarantee strength.

EP5: Scope of Codes Elaborated.

i. Rationale: The project reused proven library and frameworks, followed standards of coding and most uplrated practices and was thus maintainable and reproducible.

EP7: Interdependence

i. Justification: The components of the system, starting with the data preprocessing level, and then the classification and visualization, were very interdependent so that the concepts now needed to be carefully integrated and tested in order to have a seamless functionality.

Table 5.2: Mapping with knowledge Profile

K3 Engineering Fundamentals	K4 Specialist Knowledge	K5 Engineering Design	K6 Engineering Practice	K8 Research Literature
✓		✓	✓	

K3: Engineering Fundamentals

A solid knowledge on programming, machine learning algorithms and principles of image process was important to implement the hybrid classification model in a better way.

K5: Engineering Design

Justification: The workflow (data preprocessing and feature extraction) and classifier training and evaluation were designed systematically and thus modular and scalable.

K6: Engineering Practice

Justification Practical skills would be used to code, debug and performance tune and incorporate various machine learning components using mainstream machine learning frameworks like Tensorflow and Scikit-learn.

5.4.2 Engineering Activities

The engineering efforts that were part of this project include a detailed mechanism of tasks that are directed to designing, development and testing a sustainable mango leaf disease classification system. Key activities included:

Requirement Analysis and Planning: The identification of the project Objectives, the definition of both the Functional and Non- Functional requirements as well as the structuring of the methodology that guarantees a systematic process of progress.

Preparation of data: Data gathering, structuring and processing of the data, such as image resizing, normalization and augmented to enhance model performance in generalization.

Model Development Adaptation: Hybrid mode, combining with deep learning-based feature extraction (VGG16) with traditional machine-based approaches classifiers (SVM, Random Forest, KNN), it needs knowledge of both fields.

Training and Optimisation of algorithms: Train several classifiers, hyperparameters, and repeatedly refine them using training and parameter optimisation to find the best or optimal model.

Evaluation and comparison: This can evaluate and compare performance by applying rigorous testing procedures, and confusion matrix analysis, PCA visualization, and comparative performance measure based on the metrics of accuracy, precision, recall, and F1-score.

Documentation and Reporting: Keep comprehensive records of methods, experimental procedures and findings to promote transparency and reproducibly.

Ethical and Sustainability Implications: Ensuring that data is used responsibly, complying with privacy, creating a system that is cognizant of environmental effect.

Coordination on activities was achieved in order to overcome technical difficulty and provide a solution that is effective enough to address the objectives of the project.

5.5 Summary

The chapter gave the emphasis on the compliance with engineering standards along with exploring the major design issues as faced during the development of the mango leaf disease classification system. Engineering activities were extensive (EA1) i.e have involved software frameworks, cloud computing and domain knowledge that would develop an all-in-one and highly efficient solution. The project needed an interaction (EA2) of moderate magnitude between the components of the system, including the feature extraction/ classification/ and evaluation modules, to guarantee smooth integration.

The research corresponded to some extent of innovation (EA3) through the extraction of convolutional neural networks with classical machine learning classifiers to increase the accuracy of the multi-class disease classifications. The implementation and the design of the system took into account the effects to the society and the environment (EA4), where sustainable agricultural production was guaranteed and reliance on manual being prevented.

Lastly, the project used established methodologies and tools but solved some distinct issues that peculiar to mango leaf disease detection, which indicates the right degree of familiarity with the problem area and flora (EA5) of the technical environment. Comprehensively, this chapter reiterated the time and again that the project is multidisciplinary and would therefore demand multifaceted efforts to come up with a reliable, ethical, and effective engineering solution.

Table 5.3: Mapping with complex engineering activities

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

EA2: Level of Interaction

i. Rationalization: component components (features extraction, classification and visualization modules) had to integrate well to facilitate a data flow to allow the system to run.

EA3: Innovation

i. Rationale : The convolutional neural network and traditional classifier hybrid is an innovative methodology that is suggested to be based on accurate multi-class disease classification.

EA4: highlighting the impact on Society and the Environment.

i. Oxygen: There is a positive impact of the system on society and the environment as it helps farmers to secure their sustainability by overseeing early disease detected at the farm, reducing pesticides setbacks and finding ways to infiltrate excessive pesticides.

Chapter 6

Conclusion

This conclusive chapter outlines the summary of the whole project with references to the major findings, successive milestones and constraints engaged in the formation of the hybrid CNN-SVM model classification of multi-class mango leaf spot diseases. It also presents possible ways in which future work and development could be achieved in the effort to spur continued development of automated plant disease deployment systems. The chapter revisits the contributions of the project to the current technology of agriculture and in terms of practicality in real world implementation.

6.1 Summary

The project was able to create a hybrid mango leaf disease classification system which involves the deep convolutional neural networks feature extraction using classical machine learning classifiers. The system extracted features with the VGG16 model and tested three Support Vector Machine, Rand dom Forest and K-Nearest Neighbor classifiers with the highest accuracy on the SVM. The dataset consisted of eight categories, seven of them were diseases and one was the healthy category, images of the healthy category were preprocessed and divided into the training and validation set.

The effectiveness of the model to have multiple leaf spot diseases correctly classified could be evaluated, that is metrics of evaluation and visualization could be applied and help in diagnosis and management of the disease at an early stage. The hybrid solution turned out to be quite effective and convenient, making the most of out of both deep learning and classical approaches. It was also delivered that the project overcame data issues and achieved optimal model performance to add value on the area of plant disease detection.

6.2 Limitation

Although the outcomes of this project are rather encouraging, some limitations must also be admitted. First, the size of the dataset, though varied and comprising eight categories, might not be adequate to reflect the large range of mango leaf diseases across types of environment. This may lead to the impersonation of the model to unobserved data across diverse geographical areas.

Secondly, the quality of the feature extraction of the trained network VGG16 is essential to hybrid model. Although effective, it might not be as sensitive to detect every nuanced disease property as can be done with end-to-end deep learning models directly fine-tuned on large, dubiously specific hold-out sets.

Also, the classification was more uncertain in different classes of diseases and some actively displaying similar avenues visually took more difficulty to classify correctly. This implies that more sophisticated feature engineering or integration of further data formats, as might be spectral or environmental data.

Lastly, the implementation could be carried out in a controlled setting when using cloud-based capabilities. Practical implementation will be subject to disease and testing requirements that could be associated with harsh countermeasures such as hardware limitations, variability of data acquisition, and ease of use, which will have to be conducted by further research and development.

6.3 Future Work

Further development of this project may be directed toward some aspects to boost accuracy, robustness, and usability of the mango leaf disease classification system. The ability to enhance the model in terms of the generalization can be attained with the addition of more diverse images gathered in other geographical areas and weather types. The inclusion of other types of data collected, like multispectral or hyperspectral imaging, may include richer features to distinguish disease-upward-similar visuals.

Additional future studies may consider fine-tuning the network by permitting the network to train directly as a complete system, as opposed to using fixed texture extractors, to fine-tune the network to aspect-specific disease features. Also, using ensemble learning techniques, which will integrate the predictions of multiple classifiers can yield an overall improvement in classification and failure in misclassification can be minimized.

To justify practical implementation, future practice should involve the development of an easy-to-use smart phone or web-based application with real-time detection of disease. It can also be discussed that it can be integrated with IoT devices dealing with automated data gathering and disease surveillance. Lastly, the application of new data in this system and its updated feedback will help keep the system accurate and relevant in the long run.

References

[1] S. P. Mohanty, D. P. Hughes, and M. Salathé, “Using deep learning for image-based plant disease detection,” *Frontiers in Plant Science*, vol. 7, p. 1419, 2016.

[2] D. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, “Deep neural networks based recognition of plant diseases by leaf image classification,” *Computational Intelligence and Neuroscience*, vol. 2016, Article ID 3289801, 11 pages, 2016.

[3] S. Picon, D. Alvarez-Gila, F. Seitz, D. Ortiz-Barredo, A. Echazarra, and D. Juan, “Deep learning-based detection and classification of weeds using UAV imagery,” *Remote Sensing*, vol. 11, no. 7, p. 781, 2019.

[4] R. Ferentinos, “Deep learning models for plant disease detection and diagnosis,” *Computers and Electronics in Agriculture*, vol. 145, pp. 311–318, 2018.

[5] M. Lu, K. Singh, and A. Singh, “An SVM-based approach for plant leaf disease classification,” in *Proc. IEEE Int. Conf. on Computer and Communications*, Chengdu, China, 2017, pp. 78–83.

[6] M. P. Kumar and K. S. Bhuvaneswari, “A hybrid CNN and SVM approach for plant disease detection,” *International Journal of Engineering and Technology*, vol. 7, no. 4, pp. 1586–1590, 2018.

[7] M. Zhang, J. Zhu, and X. Yang, “Plant disease recognition based on deep learning and improved SVM,” *Journal of Intelligent & Fuzzy Systems*, vol. 37, no. 1, pp. 123–132, 2019.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in *Proc. Advances in Neural Information Processing Systems (NIPS)*, Lake Tahoe, 2012, pp. 1097–1105.

[9] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *Proc. International Conference on Learning Representations (ICLR)*, San Diego, 2015.

[10] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[11] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.

[12] S. Rifai, Y. Bengio, A. Courville, and P. Vincent, "Disentangling factors of variation for facial expression recognition," in *Proc. ECCV*, 2012, pp. 808–822.

[13] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. International Conference on Learning Representations (ICLR)*, 2015.

[14] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005, vol. 1, pp. 886–893.

[15] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

201-15-13873

ORIGINALITY REPORT

14%	9%	9%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	3%
2	Submitted to University of Central Lancashire Student Paper	1%
3	www.frontiersin.org Internet Source	1%
4	R. N. V. Jagan Mohan, B. H. V. S. Rama Krishnam Raju, V. Chandra Sekhar, T. V. K. P. Prasad. "Algorithms in Advanced Artificial Intelligence - Proceedings of International Conference on Algorithms in Advanced Artificial Intelligence (ICAAAI-2024)", CRC Press, 2025 Publication	1%
5	Pushpa Choudhary, Sambit Satpathy, Arvind Dagur, Dhirendra Kumar Shukla. "Recent Trends in Intelligent Computing and Communication", CRC Press, 2025 Publication	<1%
6	journal.whioce.com Internet Source	<1%
7	Submitted to United International University Student Paper	<1%
8	Thangaprakash Sengodan, Sanjay Misra, M Murugappan. "Advances in Electrical and	<1%