DEEP LEARNING APPROACH TO DETECT BANANA PLANT DISEASES WITH IMAGE PROCESSING

Romesh Perera

A dissertation submitted in partial fulfillment of the requirement for Bachelor of Science (Honors)degree in Computer Science

Department of Computing

Informatics Institute of Technology, Sri Lanka

in collaboration with

University of Westminster, UK

2021

Declaration

I hereby declare that this project report and all the artefacts associated with it, is my own work and it has not been submitted before nor is currently being submitted for any degree programme.

Full Name of the Student:- Palihawadana Arachchige Romesh Anton Jude Perera

Student Number:- 2017397

University of Westminster ID:- w1698453

Abstract

Banana is a famous fruit that commonly available across the world because it instantly boosts your energy. Bananas are one most consumed fruit in the world. According to modern calculations, Bananas are grown in around 107 countries since it makes a difference to lower blood pressure and to reduce the chance of cancer and asthma.

Banana plant diseases have become a severe problem in Sri Lanka and all over the world. This damages banana cultivation by affecting the banana quality and quantity. The classical and ancient strategy for reconditioning and identifying banana plant diseases is based on bare eye observation. Identifying from the naked eyes is not a good method cause it spends time and it depends on the knowledge. And the main reason is the less availability of experts.

In order to identify banana plant leaf diseases, the author has decided to develop a Convolutional Neural Network (CNN) followed by a residual architecture. There wasn't an open-source dataset so the author had to create a dataset by doing image augmentations. Two models were created with removing the background and without removing the background and was able to get 97.14% and 98.7%.

Keywords: Deep learning, Banana disease detection, Residual networks, Feature extraction

Acknowledgement

I would like to take this opportunity to thank you each and everyone who supported me towards the successful completion of my final year project – Banana Plant Disease Detection System. I would like to extend my gratitude for their great assistance and guidance through out the project.

I thank my supervisor Mr Lakshan Costa, without whom this project would not have been completed. His constant assistance and supervision led to the success of this project. I am grateful for his insightful comments and feedback on my thesis was very approachable and was always open to me when ever I was left with a doubt. He consistently allowed this to be my own work, but guided me whenever he felt that I needed it.

I would also like to express my gratitude to the Dean of Informatics Institute of Technology Mrs. Neomi Krishnarajah, The head of the department of Computer Science Mr. Saman Hettiarachchi, and to administration of Informatics Institute Technology for providing me this opportunity to work on a project that I desire.

I would like to thank each and every one who took part in my online surveys, interviews and contributed their ideas to analyze and evaluate the project.

I would also like to thank my friends who helped me throughout the project. Finally I would like to express my sincere gratitude to my parents and siblings for their unfailing support, words of encouragement, and for bearing up with me during my research project period. This accomplishment would not have been a success without the support of all whom I have mentioned here

Table of Contents

List of Figures

List of Tables

List of Abbreviations

CNN	Convolunioal Neural Network
SVM	Support Vector Machine
AI	Artificial Inteligence
RNN	Recurrent Neural Network
KNN	K-nearest Neighbour
CSS	Cascading Style Sheet
HTML	Hyper Text Markup Language
ML	Machine Learning
SGD	Stochastic Gradient Descent
RAD	Rapid Application Development
OOD	Object Oriented Develpoing
UML	Unified Modelling Language
OS	Operating System

CHAPTER 1: INTRODUCTION

1.1 Chapter Overview

The main reason for doing this chapter is to examine and come to an assertion with the supervisor about the project background, the structure, fundamental objective, aim, scope and the prerequisites of the project. This chapter explains the software and hardware requirements to carry out this project.

1.2 Background to the problem

Banana is a pleasing fruit among Sri Lankans. Most of the small farmers are like to cultivate banana widely. One of the problems that occur with banana cultivation is the number of diseases affecting the crop. In Sri Lanka, a total of 100 farmers from Batticaloa district (25 Farmers from 4 locations) are interviewed to get feedback that they are facing when cultivating banana plants. The major diseases reported were banana streak virus disease, banana sigatoka, panama wilt and bunchy top. There are many diseases like banana speckle, banana freckle, cordana leaf spot, banana crown rot and anthracnose. To control these types of diseases most of the farmers use integrated chemicals and cultural methods. (W.Shanika, 2016)Therefore, if the farmers can identify the correct disease at early stages it would be an advantage to the farmers.

Figure 1-1 Bunchy top virus

Figure 1-2 Banana streak virus

Figure 1-3 Banana sigatoka

Figure 1-4 Banana speckle

Figure 1-5 Cordona leafspot

Figure 1-6 Good leaves

1.3 Problem Definition

Disease, any harmful divergence from the normal structural or functional state of an organism. Generally, it has symptoms and differing in nature. Sigatoka, banana speckle, panama wilt and bunchy top, banana streak etc... are the main diseases that all the researchers are trying to solve from different methods. According to (Jihen Amara, 4, june 2016) has used a LeNet architecture to identify two diseases. Author (Ferentinos, 2018) has used pre-trained models like Alexnet, VGG to identify diseases without implementing a new architecture. Using a pre-trained model will cause overfitting problems. Implementing a new model will reduce the validation loss as well as it will increase the accuracy.

1.4 Problem Statement

The paper "A deep learning-based approach for banana leaf diseases classification". (Jihen Amara, 4, june 2016) has used LeNet architecture that consists with three or four simple convolutional

layers and it's detecting a limited number of diseases. When we summarize all the above researches, they have developed systems to identify maximum two or three diseases and almost all of them were based on India. But none of those systems have provided all the diseases which provides by this system.

1.5 Motivation

In Sri Lankan economy the agricultural industries play a main role to develop the economy. So to identify these diseases at early stages they should test periodically in order to avoid these type of diseases. This will be helpful for the farmers to identify the diseases before affecting the total farm.

1.6 Existing Work

"A deep learning-based approach for banana leaf diseases classification" (Jihen Amara, 4, june 2016) A model has been trained to detect mainly two diseases namely, banana sigatoka and banana speckle. So the author of this article has used LeNet architecture to train the model with a classification to classify the images.

According to the journal has written by Konstantinos (Ferentinos, 2018) has developed a system to detect plant diseases containing 87,848 images with 25 different plants. The main diseases were "Banana sigatoka" and "Banana speckle". This system has used many pre trained architectures like Alexnet, VGG, GoogLeNet.

Following the paper on "Spring nature" (Mary, 2020) has used a novel deep learning technique called Heap Auto Encoders(HAEs) to extract the important features and to reduce the exhausted features. The author has used "HAEs" because it includes the "ReLU" activation functions. According to the author this system has provided an 99.35% accuracy for real datasets.

In 2019 The International conference related to signal processing communication (ICSPC, India) has introduced a banana plant disease detection system to identify the affected plants. In this system for the identification they have used texture pattern techniques like (LBP). The extracted features are sent to (SVM) support vector machine and K-nearest neighbor when classifying the final output. Classification consists of two different diseases "Banana sigatoka" and "Cordona leaf-spot". This has given accuracy of 89.1% and 90.9% as the final output.

In 2020 The faculty of (ECA, India) have published a paper to detect banana leaf and fruit diseases in India. (N.Saranya, 2020) According to this research, it has several steps image acquisition, image pre-processing, feature extraction and disease detection with ANN based classifier to classify the diseases. “Banana sigatoka”,“Anthracnose”, “Freckle affected fruit and leaves” are the main diseases identifying through this system.

Research	**Techniques Used**	**Improvements**	**Limitations**
A deep learning-based approach for banana leaf diseases classification. (Jihen Amara, 4, june 2016)	LeNet Architecture Optimized Algorithm (SGD)	Accuracy Color – 92.8% Gray scale – 85.9%	Limited number of diseases have been detected.
Deep learning models for plant disease detection and diagnosis (Ferentinos, 2018)	Alex Net VGG16, VGG19 GoogLeNet	All techniques have above 90% accuracy	Limited number of diseases have been detected.
Novel HEAP auto encoder (HAE) to detect banana leaf diseases. (Mary, 2020)	Heap Auto Encoders	Accuracy is up to 99.35%	This research is based on India so the diseases that are detecting is limited
Using local binary pattern and support vector machine to detect and classify banana plant diseases.	Used (LBP) to detect texture patterns SVM Classification K-nearest algorithm	Accuracy is about 89.1% and 90.9%	Limited number of diseases have been detected.
Using neural networks to detect	ANN to classify images. MATLAB	Average accuracy is 90%	Limited number of diseases have been

fruit and banana leaf diseases (N.Saranya, 2020)	Fuzzy c-means		detected. Research is based on India.

Table 1-1 Existing Work

1.7 Research Gap

This system has some similarities and differences with other existing systems. Though there are systems to identify a limited number of diseases the main diseases like “Banana Sigatoka” and “Banana speckle”. When comparing with the existing systems the reader can identify that the other researchers have used most of the features in different combinations. But none of those systems have provided most of the diseases which provide by this system. So, the research gap is to identify more banana plant diseases like “Cordana leafspot”, “Banana bunchy top”, “Banana streak”, “Banana sigatoka” and “Banana Speckle” using a researched-based Residual network with a CNN classification. All most, most of the researches were done in India, so this would be a great opportunity for Sri Lanka as well to come with a system like this to prevent such diseases .

1.8 Research Contribution

In India and Bangladesh, there are applications to identify a limited number of banana plant diseases. So, most of the researchers are used pre-trained models to train as well as for the classification. According to the research done in Batticaloa, Sri Lanka (W.Shanika, 2016) the main diseases are “banana streak” and “panama wilt”. In present we cannot find a proper system to identify these two diseases. In Sri Lanka there are no experts to identify banana plant diseases in the field. Most of the Farmers are using old techniques like anti-pesticides etc... According to the author, none of those systems have provided the main five diseases which provide in this system. The dataset is not available for this project so the author needs to take photos and augment them to increase the accuracy.

1.9 Research Challenge

After a deep research, the author has identified some challenges to be carried out through the research.

1. Reviewing the latest CNN and ANN models. (Have to make a decision to select the best matching model.)

2. Identifying the latest banana plant diseases to check whether there some unexplored diseases.
3. Collecting data for the training process.

1.10 Research Question

1. How effective does neural networks and machine learning in identifying banana plant diseases?

2. What factors affect the planter to identify banana plant diseases?

1.11 Research Aim

Banana plant diseases have become a major threat to the universe including Sri Lanka. The aim of this project is to develop a system to identify banana plant diseases and to aware of the solutions to prevent them at the early stages.

1.12 Research Objectives

1. To compare with the existing frameworks that are used to identify banana plant diseases, using different convolutional neural network structures.
2. To identify the limitations of the current system.
3. To evaluate the prototype whether its accuracy and efficiency at a satisfactory level.
4. To Augment the dataset.

1.12.1 Operational Objectives

Objective	Description
Project Initiation Document	The project initiation document should include following concepts like project environment, problem domain, project scope, project aim, objectives, prerequisites and and work break down structure.
Literature Review	Deeply research about excisting banana plant diseases detection systems and to identify a valid reseach gap and to slect the most suitable algorithms.

Software Development Methodology	After analyzing the literature survey stepping into software design and development methodology to develop the system with the slected softwares and tools.
Requirement Gathering	Collecting data from domain experts', technical experts, feedbacks and end-user feedback to evaluate the system.
Software Requirements	Identifying the software requirements specification. This includes functional and non-functional requirements.
Testing and Evaluation	Testing the system according to a testing critera nad serveral benchmarks to validate the system. Developed system needs to evaluate from different evaluators.
Thesis	The research paper and thesis should submit within the given time.

Table 1-2 Operational Objectives

1.13 Project Scope

Following are the focus concepts and non-focus concepts of the research

1.13.1 In-Scope

- Reviewing the latest CNN and ANN models (have to research about the layers and architectures)
- There are types of loss functions and optimizations to test and evaluate so this needs more patience.

1.13.2 Out-Scope

- Loss functions are limited so to explore a better result it needs to test and evaluate the accuracy. There can be miss-testing when evaluating.

1.13.3 Rich Picture

Following Rich picture gives an idea about the project domain.

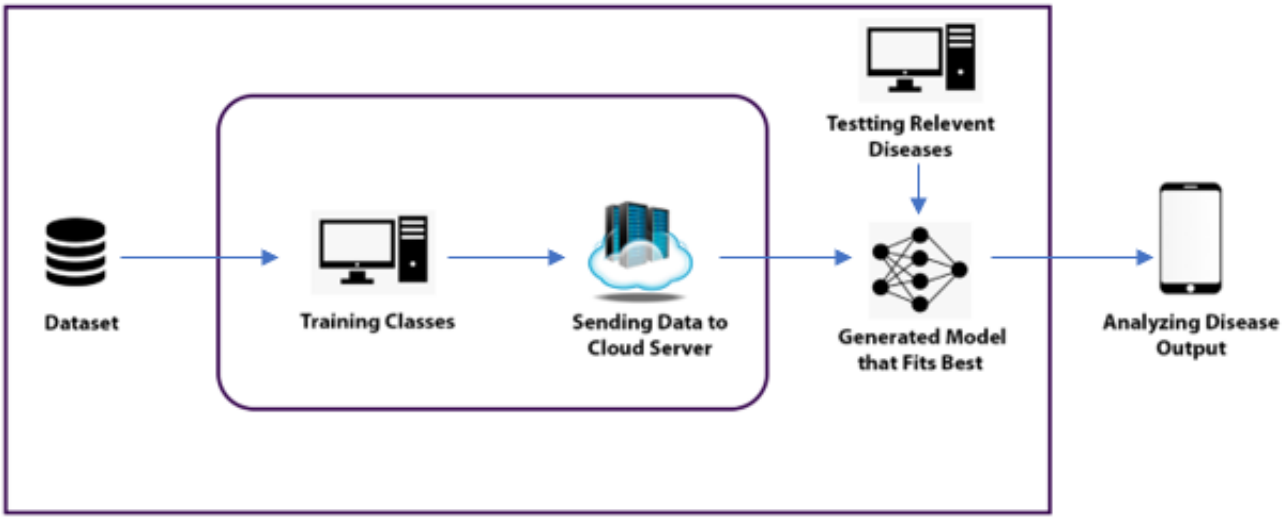


Figure 1-7 Rich picture

1.14 Resource Requirements

According to the research, to implement the system the software, hardware, and data resource requirements are as follows.

1.14.1 Hardware Requirements

- Core i7 processor
- 16GB Ram
- Disk space of 40GB or above
- A GPU contains with 600+ cuda cores

1.14.2 Software Requirements

- OS (Windows 10 / Mac / Ubuntu)
- PyCharm – Python IDE
- TensorFlow-GPU – Developed by Google to run framework computations
- Adobe Lightrooms – To augment the dataset
- Zotero – To manage references
- MS Office/ Google Docs – To create documents and reports
- Google Drive – To backup files and datasets

- Adobe Photoshop – To augment images

1.14.3 Data Requirements

- Datasets – Valid Banana disease images approved by agricultural officials

1.14.4 Skill Requirements

- Fluency in Python libraries (NumPy, Pandas)
- An idea about CNN models like VGG16, VGG19 etc...
- Researched based writing skills
- Needs experience in Tensorflow and Keras
- Needs to learn OpenCV library

1.15 Chapter Summary

The first chapter of the thesis has briefly explained about problem domain. A summary of excisting work has documented with a table format. The research questions, objectives, aim, research motivation were discussed under this chapter.Additionaly the resource requirements were summarized in this chapter. Further a critical analysis of excisting systems, the technologies, technical domain has been documented in next chapter 2 literature review.

CHAPTER 2: LITERATURE REVIEW

2.1 Chapter Overview

Banana plant leaf diseases is a major threat in modern days as well as in the past. Many researchers are still focusing on various approaches to identify banana plant diseases in the early stages. According to previous work, it's not legitimately recorded nor exceptionally clear due to the contrasts within the approaches. This chapter consists with a deep analysis of existing work, model architectures, algorithms, techniques, etc. The objective of this content is to show a summarized assessment and to explore new possible approaches to identify banana plant diseases.

2.2 Concept Graph

The entire literature review including the scope and the algorithms that are used can be displayed in a graphical form. The literature review is divided into subcategories like technologies and existing work. The graph can be found in **Appendix A-3.**

2.3 Literature review of the problem domain

2.3.1 Sri Lankan Agriculture

In Sri Lanka, the agriculture sector plays a major role in contributing to the national economy. Around 38% of the total labor constrain was engaged in agriculture in 1999. (Nations encyclopedia, 2020). The root of bananas is set in Southeast Asia in the jungles of Philippines, Malaysia, Indonesia. Banana is a vital natural fruit widely cultivated in Sri Lanka.

2.3.2 Banana Types in Sri Lanka

Banana has a top place among most cultivated fruit crops across the world and it's the most famous fruit among Sri Lankans. (Perming, n.d.) In Sri Lanka, we can see many varieties of bananas grow freely all year round. According to the department of agriculture records, the farmers cultivated in large, medium, and small-scale orchards and in-home gardens. According to the records, there are almost nearly 1000 varieties across the world. In Sri Lanka gloats around 29 varieties, the more prominent number which shows up to be innate to the nation (even though a few are said to have been brought from South India) (Weerasooriya, 2016).

2.3.3 Banana Plant Diseases

"Anything that anticipates a plant from performing to its greatest potential" can be define as plant disease. Plant diseases can be divided into two subcategories named "**Abiotic or non-infectious disease**" and "**Biotic infectious disease**". There are various types of banana plant diseases in modern days. Such as Banana Sigatoka, banana leaf streak, banana wilt, moko disease, bunchy top, banana freckle, banana speckle, etc. Sigatoka is a famous disease caused by a fungus called *Mycospharella fijiensis* and this can be identified with black leaf spots. Banana leaf streak is a common disease around Asia, Africa, and Central America. The scientific name is *Banana streak badnavirus*. This has symptoms like lines of yellow that run from the midrib of the leaf to the edge. (Autralian Government, 2017) Streak virus is much more vulnerable than banana sigatoka. Banana wilt is a risky disease it begins in the soil and can travel to the root of the system. This virus can travel through water, air, moving soil and cultivating equipment. This has symptoms like yellowing the leaves within a short period. The bunchy top is a virus which conveys via aphids *Pentalonia nigronervosa*. Queensland had to wipe out the commercial banana industry due to this virus. Signs of bunchy top is the leaves become narrow and short with a small margin. New leaves yellow and ended up wavy with dull green "dot and dash".(Grant, 2020) The famous Cordona leafspot is a worldwide disease infected by two *Neocordana organisms* that are regularly found as secondary intruders of leaf lesions caused by other parasites. The most characteristic indications of the disease are on the leaf. It has large pale brown spots with different patterns.

According to the researchers (Shanika and Prasannath, 2016) have done a survey in Batticalao district to investigate number of banana plants that are affected by various diseases. The author can identify the major diseases namely banana streak, panama wilt, banana sigatoka and bunchy top. Among these infections, the foremost predominant infection (64%) in Batticaloa was banana streak disease.

2.3.4 Disease Identifying Methods

There are not sufficient specialists to recognize the infections at the early stages and the time have to spend during the field time is high. As well as there are farmers who don't have a proper knowledge to identify whether it's a disease or not. The classical and ancient strategies for reconditioning and detecting plant diseases are based on bare eye perception and it could be an

exceptionally moderate method. According to (Shanika and Prasannath, 2016) most of the farmers were using cultural methods and chemicals to control the disease.

2.4 Existing Systems and Evaluations

2.4.1 Banana leaf disease classification

According to the past research papers, there is an interest to find a system for plant diseases. Jihen Amara (Amara et al., 2016) was the first researcher to identify a solution for the banana plant diseases. He has used LeNet architecture as a convolutional neural network for the classification to classify images. There were many challenges like illumination, different backgrounds, image sizes to evaluate the accuracy. Author implies that DNA based methods are the techniques that are using to discover such diseases. Farmers do not have diagnostics tools and this cause into real trouble. Author of this project has limited to two diseases namely “Black sigatoka” and “Banana speckle”. This consists with the feature extraction where the network can identify input images with high level features.

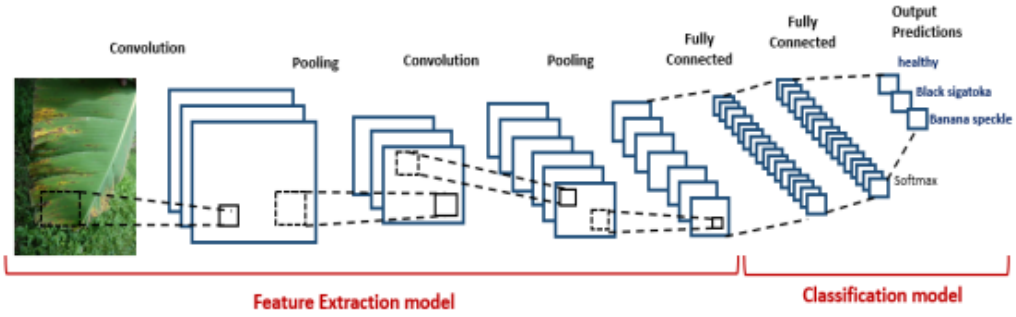


Figure 2-1 Classification model

2.4.2 Plant disease detection and diagnosis

According to the journal has written by Konstantinos (Ferentinos, 2018) has developed a system to detect plant diseases containing 87,848 images with 25 different plants. The dataset was randomly split into two classes including 80% as training set and 20% as the testing set. The main diseases were “Black sigatoka” and “Banana speckle”. This system has used pre trained models like Alexnet, VGG, GoogLeNet to evaluate the accuracy. The final performance for the VGG net was 99.56% as the highest accuracy among the other models.

2.4.3 Image classification using HEAP auto encoder

Following the paper on "Spring nature" (Ani Brown Mary N et al., 2020) has used a novel deep learning procedure called Heap Auto Encoders(HAEs) to extricate the critical highlights and to diminish the depleted features. The author has used "HAEs" because it includes the "ReLU" activation functions. The auto encoder is consists with unsupervised algorithms which includes three layers called input layer, output layer and hidden layer. The encoder is connected for recording the input data into hidden representation. decoder is connected for rebuilding input

data from the hidden up representation (Ani Brown Mary N et al., 2020). According to the author this framework has provide an 99.35% accuracy for real datasets.

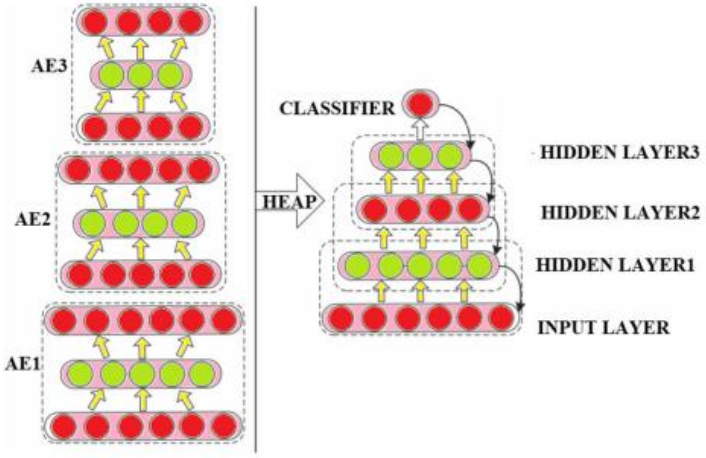


Figure 2-2 HEAP auto encoder

2.4.4 Banana plant classification - SVM

In 2019 The International conference related to signal processing communication (ICSPC, India) has introduced a banana plant disease detection system to identify the affected plants. In this system for the identification they have used texture pattern techniques like (LBP). The extracted features are sent to (SVM) support vector machine and K-nearest neighbor when classifying the final output. (Aruraj et al., 2019a) The database is consisting with 123 images to test the proposed algorithm. According to the research this algorithm needs RGB images as a parameter a value of 3. Classification consists of two different diseases "Black sigatoka" and "Cordona leaf-spot". This has given an accuracy of 89.1% and 90.9% as the final output.

2.4.5 Banana leaf and fruit diseases identification

In 2020 The faculty of (ECA, India) have published a paper to detect banana leaf and fruit diseases in India. (Saranya et al., 2020a) According to this research it has several steps image acquisition, image pre-processing, feature extraction and disease detection with ANN based classifier to classify the diseases. The images are captured with a 12 mega pixels camera for more accuracy. “Black sigatoka”, “Anthracnose”, “Freckle affected fruit and leaves” are the main diseases identifying through this system. Banana freckle is a dangerous fungus disease that yellows the tissue with dark brown spots on the leaves. In Hawaii, Indonesia, Thailand, and Taiwan are the most famous countries for this disease.

2.4.6 Banana plant disease identification through aerial images

In Africa, observing larger landscapes using unmanned airborne vehicles (UAV) are challenging, So using high resolution image datasets consists with satellite images, the researcher (Gomez Selvaraj et al., 2020) has done a pixel based classification. This classification was based on random forest (RF) model combined with vegetation indices (VI) and principal component analysis (PCA). The overall accuracy was 97%. The targeted diseases were “Banana bunchy top” and “Xanthomonas wilt”. When using satellite images there’s a chance to miss out infected plants due to various backgrounds. This system identifies diseases from a sensor monitoring system and it cost money to buy the needful.

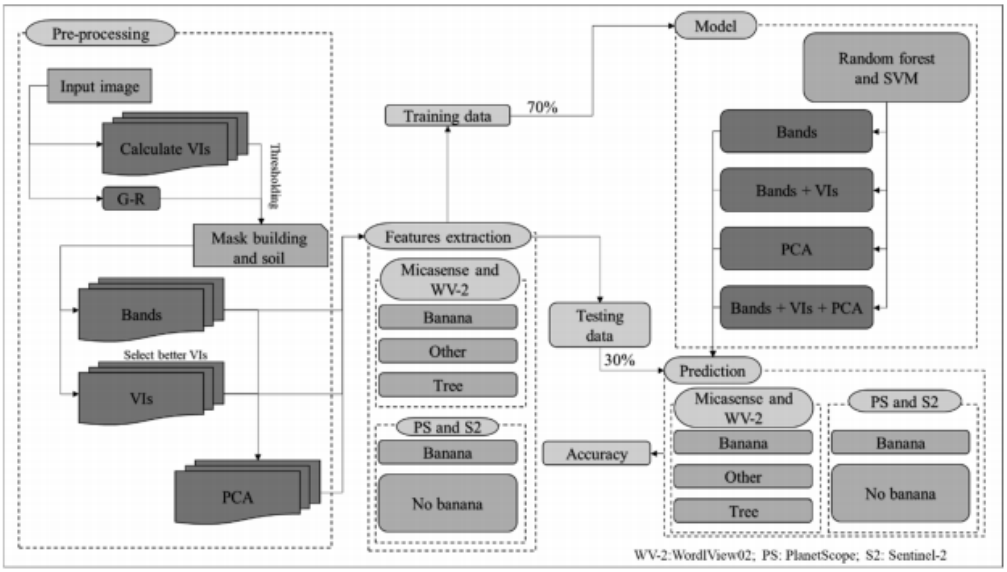


Figure 2-3 Aerial images classification

2.5 Literature Survey of the Technologies

2.5.1 Algorithmic Analysis

When selecting an algorithm, we have to consider its compatibility with the model which is going implement. When selecting the best-suited algorithm, we have to compare some factors like interpretability, data format, training time, memory requirement and etc.

2.5.2 Machine Learning

Machine learning is a part of Artificial Intelligence (AI) that gives systems to learn automatically to improve from the encounter without being programmed. According to Fhel Dimanoo Machine learning is kind of a scientific study about algorithms and statistics which use to perform a particular task. (Dimaano, 2019) The training dataset represents the data which the machine learning model will ingest to illuminate the issue its design to unravel. There are different types of machine learning patterns.

- Supervised Learning
- Unsupervised Learning
- Semi-supervised Learning
- Reinforcement Learning
- Transfer Learning

Machine learning patterns can be changed according to the project and to the final output. Since the banana plant disease detection system is using supervised learning patterns, the author tries to persuade these patterns as well to compare and select the foremost appropriate approach.

2.5.3 Supervised Learning

Supervised learning is a kind of mapping the input variables (x) to the output variable (y) through algorithms. This method is called as supervised learning because the algorithm which process to learn from the training dataset can be consider as a teacher. Finally, the algorithm will make predictions according to the training dataset. Supervised learning can split into two groups namely **Classification** and **Regression.** (Brownlee, 2020) When we compare these two algorithms the author can identify that the regression algorithms are utilized to predict values like price, age, salary and etc. To predict values like True or False, Male or Female are using classification

algorithms. Regression is kind of procedure which finds the relationships between dependent and independent factors.

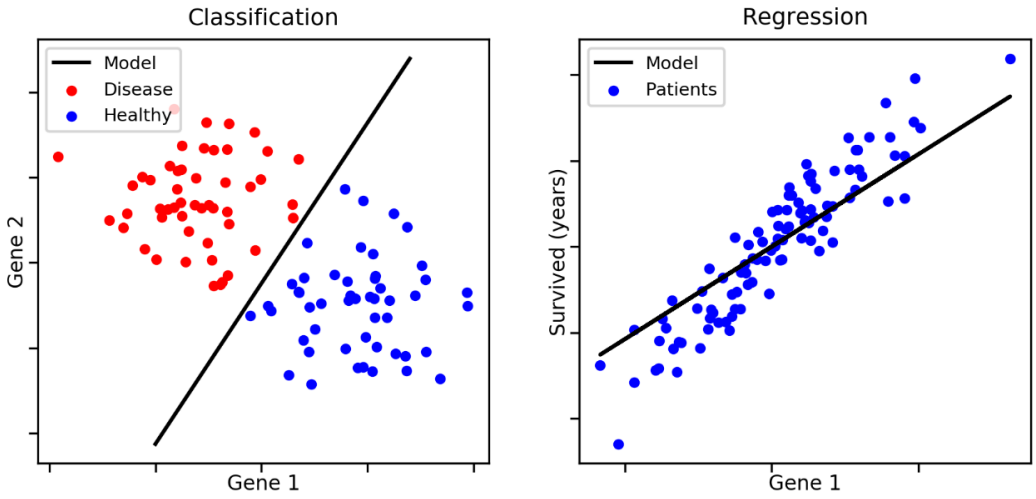


Figure 2-4 Supervised Learning methods

Classification is a process which divides the dataset into classes based on different parameters. There are some classification algorithms like support vector machine (SVM), Logistic regression (LR), K-nearest neighbors (KNN), Random forest (RF), Artificial neural network models (ANN). According to the project author tries to discuss more classification algorithms to choose the most appropriate algorithm to build the model.

2.5.4 Support Vector Machine (SVM)

The main purpose of using this algorithm is to find out hyperplane in N-dimensional space (N - number of highlights) that unmistakably classifies the information points. (Gandhi, 2018) This can be used as a text classification algorithm as well. In 2019 (Aruraj et al., 2019) has used texture patterns techniques (LBP) to extract features. For the classification, the author has used SVM and k-nearest neighbor algorithms to classify banana plant diseases. SVM classifier has gained 90.9% of average accuracy. According to the research, SVM classifier is suitable for high dimensional spaces but for larger datasets, it requires a lot of memory power. The advantages of SVM

- SVM has powers kernel trick.
- Stability
- Compared to ANN it gives a better accuracy

The disadvantages of SVM is as follows

- Choosing a suitable kernel function
- SVM needs a lot of memory power
- For the training, it takes long time
- Feature scaling is required

2.5.5 Logistic Regression (LR)

This is a classification algorithm used to relegate perceptions to a discrete set of classes. This is used to identify whether classification problems like spam email or not, fraud online transaction or not, etc. This algorithm can transform its output by applying a logistic sigmoid function to estimate probability. (Pant, 2019)

What is sigmoid function? In order to outline anticipated values to probabilities, we use sigmoid function. This function can reveal the dataset and map values between 0 and 1. For these types, developers use the sigmoid function to estimate predictions to probabilities. There are advantages and disadvantages when using the sigmoid function. The advantages are

- It's easy to implement and for training
- Doesn’t require high memory power
- It's easy for low dimensional datasets because it won't overfit compared to other algorithms
- This can split into multiple class classification as well

Sigmoid function has few Disadvantages

- It's hard to identify complex relationships
- This works more efficiently with larger data sets

What is softmax function? Softmax is a numerical calculation that changes a vector of functions into a vector of probabilities. For the applied machine learning patterns researchers use softmax function.

2.5.6 K-nearest Neighbor (KNN)

K nearest neighbor belongs to supervised learning domain and this algorithm can be used for pattern recognition and for data mining purposes. This called as **lazy learning algorithm** because for the training phase it's not too much accurate and doesn't accept anything from underlying data. Previous researchers have used KNN classification models to identify banana plant diseases. (Aruraj et al., 2019b) has used this classification for their research and got 89.1% accuracy. When comparing KNN with Support vector machine algorithm SVM classifier has a better accuracy. Advantages of KNN classifier are

- It doesn't need time for training, much faster than SVM and LR
- New data can add without impacting the accuracy of the algorithm
- Easy to develop

The disadvantages of KNN is as follows

- This algorithm is not suitable for large datasets
- Have to manually impute missing data values

2.5.7 Random Forest

This algorithm is suitable for both classification and regression. In 2020 a group of researchers (Gomez Selvaraj et al., 2020) have implemented a system to identify banana plant diseases through aerial images using machine learning methods. In this scenario, the researchers have investigated African landscapes for the research, and the machine learning models were determined from satellite images (Planet Scope, WorldView-2) and UAV platforms. Using Random forest classification algorithm, with a combined feature of vegetation indices (VIs) the overall accuracy was around 97.1%. There are advantages when using this algorithm

- For complex problems, this can be used for both classification and regression
- Algorithm can find out missing data values

The limitations are

- For the training it will take long time period
- There are overfitting problems

2.5.8 Artificial Neural Networks (ANN) with Deep Learning

An artificial Neural Network is a model which consists with mathematical functions and biological neural networks. This contains with neurons and it gathers information employing a connectionist form to computation. ANN's adjust its structure according to the dataset which runs oner neurons during the training process. ANN learning can be either supervised or unsupervised learning. if it's supervised training, the neural network needs a set of test data in conjunction with the anticipated outputs from each of these tests. (Nasser and Abu-Naser, 2019) In India (Saranya et al., 2020b) has implanted a system to identify banana plant diseases using neural network and has used ANN classifier to classify images. ANN can be used for text classification, voice recognition and etc.

What is Deep Learning?

Deep learning is a part of Artificial neural networks and subcategory of machine learning. According *Geoffrey Hinton* He describes the meaning "Deep" as it has so many layered network of Boltzmann machines. (Brownlee, 2019) There are few deep learning types based on neural networks.

- Artificial Neural Networks (ANN)
- Convolutional Neural Networks (CNN)
- Recurrent Neural Networks (RNN)

Among these networks. CNN models are being utilized over different applications and areas, and they're particularly far reaching in video and image processing ventures.

2.5.9 Convolutional Neural Networks (CNN)

Neural networks are a part of machine learning. Neural networks are the heart of deep learning algorithms. Convolutional neural network (CNN) is a commonly using deep learning algorithm which read an input images and assign importance significance (learnable weights and inclinations) to different aspects to differentiate each other. It takes this title from scientific direct operation between matrixes called convolution. (Albawi et al., 2017)

Architecture of CNN

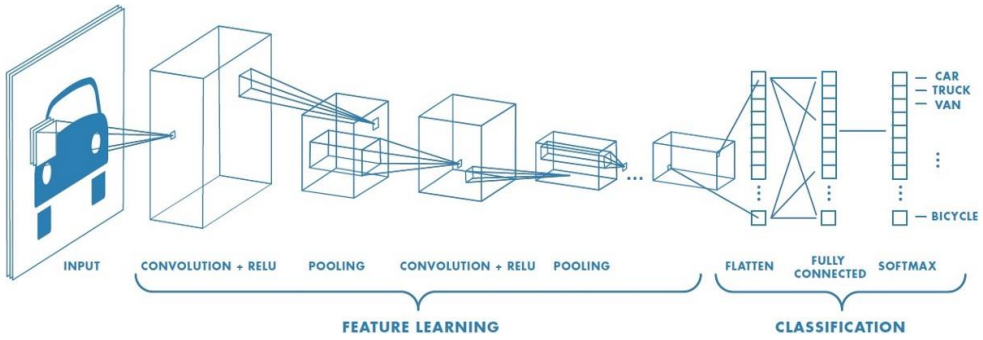


Figure 2-5 CNN architecture

How convolutional layers work? In convolutional neural networks, the convolutional layers are working like building blocks. A convolutional layer has a tensor computation function or an application which creates a feature map that visualize the input. CNN is a short specialized neural network for two-dimensional or three-dimensional image data. This network has a linear operation that can multiple the weights of input.

- Convolutional Layer - This contains with filters and parameters and each kernel has neurons to transfer data. There are 2D or 3D convolutional layers it depends on the project as well.
- Pooling Layer - Pooling layer is capable for decreasing the spatial measure of the Convolved Include. This layer can reduce the computational power which is using to feed data. There are two types of pooling
    1. Max Pooling
    2. Average Pooling
- Activation Functions - Activation functions are numerical equations that decide the output of a network. This function is connected with each neuron. These are some examples for activation function “Relu”, “Softmax”, “Sigmoid”, “tanh”. Relu is the most common activation function which can increase the performance between layers.(unknwon, 2019)
- Fully Connected Layer – Fully connected layer is connected with the output, which produce from last max pooling layer (flatten).

There are advantages and disadvantages of using a CNN. Advantages are

- Most suitable network for image classification

Disadvantages are as follows

- Needs a large dataset to gain a good accuracy
- Needs a processing power compared to other networks

2.5.10 Image Processing

Image processing is a technique that converts the image into digital format to perform operations on an image to extract valuable information from it. In modern days we can see many researchers are observing new technologies related to image processing. Deep learning models like KNN, SVM can be integrated with image processing. There are two methods in image processing namely, **digital and analogue**. When using the digital method image has to undergo with image preprocessing, enhancement and extraction. (unknown, 2014) There are few steps in image processing

- Reading the image via image reading libraries
- Analyzing the image
- The result can be a modified image or report that's based on image examination.

2.5.11 Unsupervised Learning

Unsupervised learning is a method which has machine learning algorithms to analyze and cluster unlabeled datasets. These types of algorithms can work on its own to identify patterns and information still undergoing to detect.

What is clustering? Clustering type of data mining technique which categorizes unlabeled data into groups based on their similarities. K-means, fuzzy C-means are few examples for clustering algorithms.

2.4.11.1 K-means Clustering

This is a common clustering method which has data points assigned to a group. Group represents a value of K and it has cluster values. This is more suitable for image segmentation, document clustering, market segmentation and etc. (Raut and Ingole, 2017) has introduced a system to identify plant diseases using image processing and k means clustering. There are advantages of this algorithm they are

- Easy to implement

- Works perfectly on large datasets
- Clusters can take different shapes and sizes

2.6 Chapter Summary

This chapter consists with the problem domain and has mainly focused to find new solutions to the existing problem. The concept map was used to collect information. This chapter is mainly dived into three categories namely problem domain, algorithmic review and existing work. While reading this content the reader can gather information about *what's machine learning and how it works and the subcategories of machine learning.* The author has given a prominent place to supervise learning and unsupervised learning because these are the main subcategories. In the middle section the algorithms like support vector machine (SVM), Logistic regression (LR), K-nearest neighbor (KNN), unsupervised algorithms like K-means clustering are being well explained with contrasting the advantages and disadvantages as well.

CHAPTER 3: METHODOLOGY

3.1 Chapter Overview

This chapter will examine the choice of the foremost suited research technique, project management methodology, and the selected model for this project. The development methodologies, the designs are discussed in this chapter. While doing the project the risks and mitigation plans are also examined in this chapter.

3.2 Research Methodology

The quality of a research is measured by three concepts feasibility, possibility, and scalability. These concepts need to flow according to an eloquent discipline throughout the project. Detection of banana plant diseases is not an easy task because all the diseases are mostly similar. There were various systems to identify banana plant diseases. The most common diseases like “Sigatoka” and “Banana bunchy top” have identified using different technologies. But few amounts of systems have identified diseases like “Cordona Leafspot” and “banana streak”. According to this research quantitative approach will be used to gather research data.

3.3 Development Methodology

There are different types of development methodologies. Choosing the correct development methodology is soo important for this project. With the expectation of recognizing an adaptable development methodology, the key qualities were in a general sense exacted of several programming development methodologies. Below table 3-1 shows the frequently used methodologies and their advantages and disadvantages.

Methodology	Description	Advantages	Disadvantages
Rapid	This is a type of agile development that focuses on prototype releases and iterations. The process of RAD are gathering the requirements, process of prototyping, receiving	1. Developers have flexibility and adjustability to make changes to the software 2. Quick repetitions can be done	1. Developers with high skills needed for RAD development 2. Compared to other models the

	the feedback, and finalizing the application	3. Time effective	management is complex
Waterfall	The waterfall model was the primary process model and it has a linear-sequential life cycle show.	1. Easy to implement 2. Easy to maintain or to manage. 3. Time timeline can adjust according to the requirement	1. Risk is very high 2. It's difficult to make the changes. 3. Until completing the project cannot do testing.
Agile	Agile is a combination of software development methodologies. Changing the requirements, frequent delivery and customer satisfaction are the key features of agile methodology	1. QA sessions will improve and solve the developer's doubts and concerns 2. Face to face conversations 3. Late changes can be done	1. For large software deliverables, it's difficult to manage. 2. Lack of information and documentation,

Table 3-1 Development methodologies methods

3.3.1 Selected Methodology

While analyzing the above methodologies the author identified that RAD methodology is more complex compared to Agile. And it's not suitable for a project like this because of the technical risk. Waterfall is a good technology but it's rejected because in this project still, the requirements are changing. So, The best solution was to go with the Agile methodology. This suits because while talking with the experts and grabbing their knowledge will help the author to do the changes even if it's late.

3.4 Project Management Methodology

Project management technique is fair a favored way of portraying a system utilized to do something. Before starting a project there are key points to focus on to complete the project

successfully. There are two main project management strategies such as PRINCE2, PMBOK, and Agile. PRINCE2 is a process-based methodology. This project can be managed more efficiently using the PRINCE2 methodology. There are seven concepts in PRINCE2 methodology. This methodology will help to do future implementations without any obstacles.

3.4.1 Project Plan (Gantt Chart)

The Graph can be seen under Appendix A – 1 Gantt Chart

3.4.2 Work Breakdown

The Graph can be seen under Appendix A – 2 Work Breakdown Structure

3.4.3 Deliverables

The above table 3-2 represents the summary of all deliverables that need to be finished during the given time period.

Deliverable	**Date**
Submission of Draft Project Initiation Document	5th November 2020
Submission of Literature Review Document	7th December 2020
SRS Document	4th January 2021
Interim Process Report	26th January 2021
Prototype	5th March 2021
Review Research Paper	18th March 2021
Submission of Draft project Reports	9th April 2021
Thesis	3rd May 2021
Final Research Paper	3rd May 2021

Table 3-2 Time Deliverable

3.4.3 Risks and Mitigations

Risk ID	**Description**	**Probability**	**Impact**	**Mitigation Action**
R01	With the selected model, haven't done any development before	High	Medium	Read more articles related to that model and got advice from the experts

R02	Adding more features to the application and the deadline is about to pass.	High	Very high	The main core functionalities were given a high priority and the rest parts did at last.

Table 3-3Risk mitigation plan

Even though some risks are mentioned as high somehow the author was able to find reliable solutions to keep the project at not a high-risk position.

3.4 Chapter Summary

In this chapter, all the required methodologies to develop this project were examined. The methodologies like research methodology, development methodology, and project management methodology were compared to contrast their importance. The project plan and the work breakdown chart were also summarized under this chapter.

CHAPTER 4: SOFTWARE REQUIRMENT SPECIFICATION

4.1 Chapter Overview

This chapter consists with the requirement engineering which the project needed. A brief explanation about stakeholders and how the stakeholders are engaged under this framework, and their roles have been distinguished. Also, the requirement techniques and the outcomes are discussed in this chapter. Use case diagrams, required descriptions, functional and non-functional requirements of the system are also under this chapter.

4.2 Rich Picture

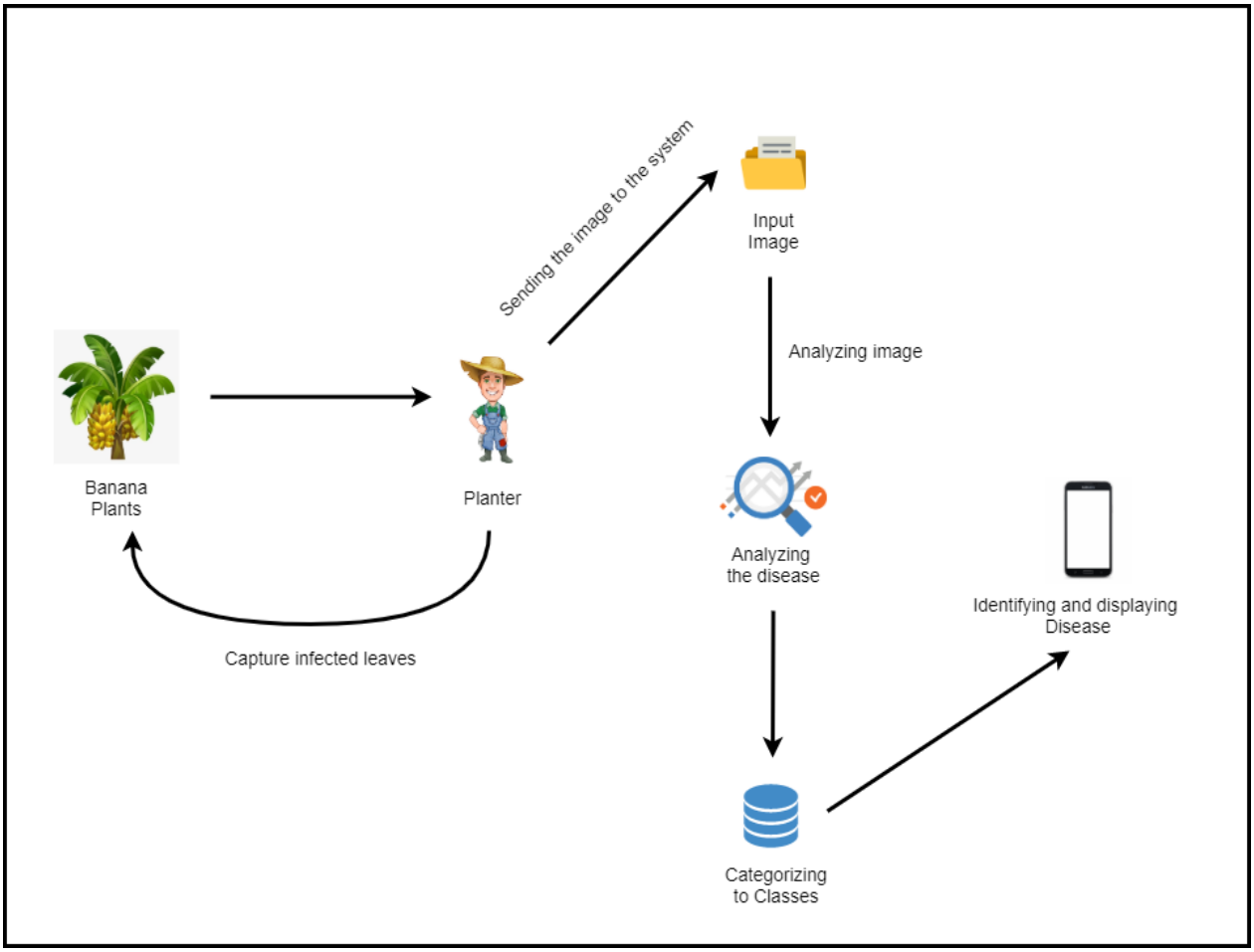


Figure 4-1 Rich picture

4.3 Stakeholder Analysis

The stakeholders represent how they engaged with this system, and to analyze we use onion model to extract their roles of this system.

4.3.1 Onion Model

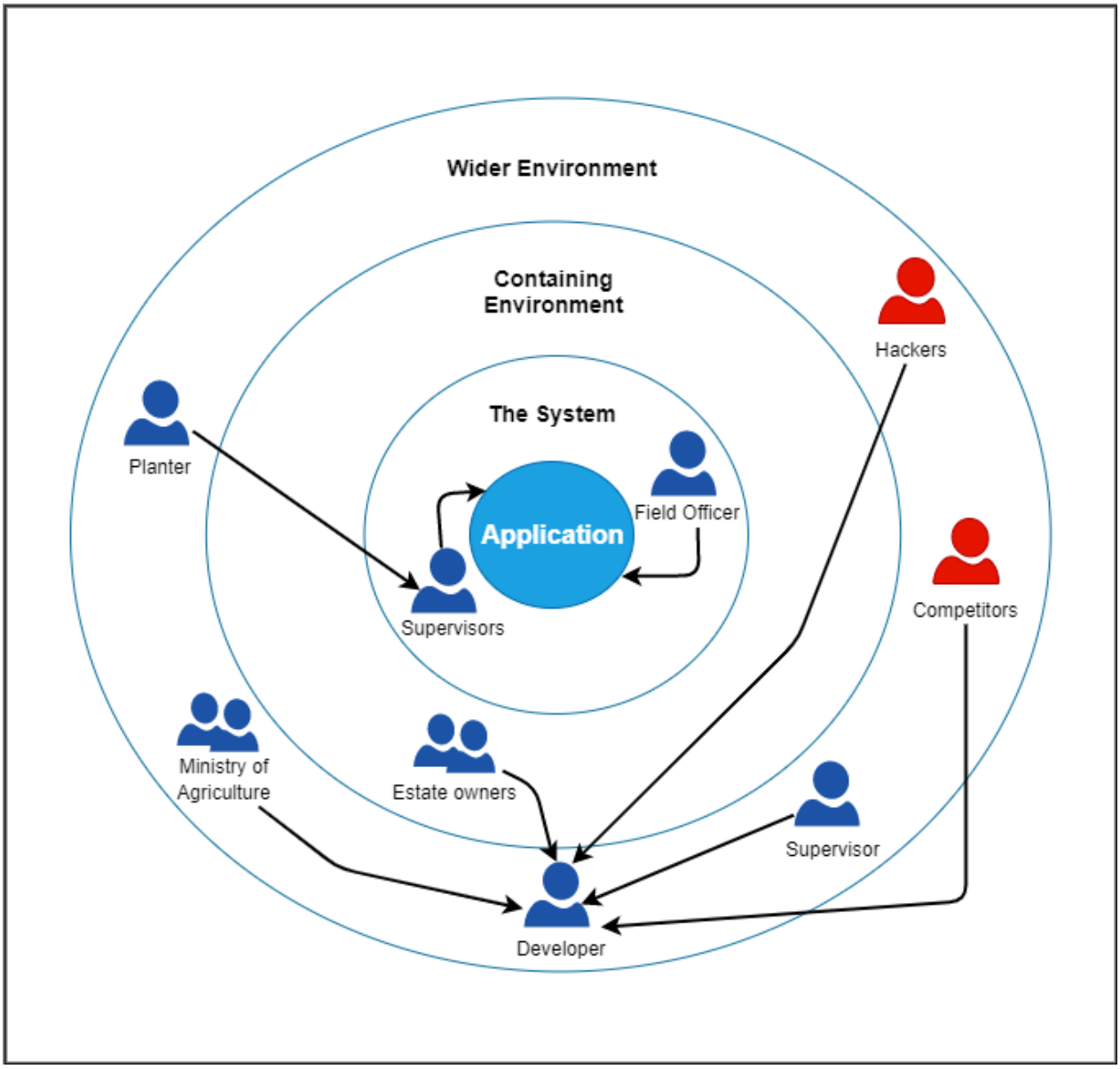


Figure 4-2 Onion model

4.3.2 Stakeholder Viewpoints

The determined stakeholders have different roles when it comes to a processing background. This table shows different viewpoints.

Stakeholder	Role	Viewpoint
Field officer	Operational role	• Can identify banana plant diseases when the planter is reported. • Can take action at the necessary time.
Supervisor		
Estate Owner	Financial beneficiary	Protecting plants from such viruses is a bit difficult task. Increasing plant production will help financially.
Planter	Financial beneficiary	Planter can earn extra salary if the harvest is solid
Ministry of Agriculture	Political beneficiary	Trying to expand the profit by maintaining the standards of the banana plants.
Developer	Develops the system and maintenance	Develops the system, adopting new technologies, training the model, testing, and fixing bugs.
Supervisor		Give direction to the student to complete the project tasks according to the requirements assembly the deadlines.
Hackers	Negative stakeholders	Gain access to the system and interrupt the system and its data.

Competitors	Negative stakeholders	Makes system with coordinate include competitions to the proposed system

Table 4-1 Stakeholder roles and their viewpoints

4.4 Selection of Requirement Elicitation Techniques/Methods

Requirement elicitation approaches are diverse ways of gathering and identifying the requirements. The reason for these strategies is to gather requirements utilizing interviews, surveys, conceptualizing and writing review surveys. L.

4.4.1 Questioning from the audience

Questioning from the planters and agricultural experts was done to identify the requirements of the banana plant diseases identification system. Advantages Disadvantages Identified problems can be discussed easily.

Advantages	**Disadvantages**
Identified problems easily and discussed them.	Difficulty to meet the domain experts.
New issues and existing ones can be analyzed.	
Can gather more information because it's a face-to-face interview.	

Table 4-2 Advantages & disadvantages observing the audience

4.4.2 Analyzing previous systems

While analyzing the existing systems which has already been written in Chapter 2 - Literature review can extract information about the technologies, limitations and etc. This helps to come up with modern prerequisites for the project.

Advantages	**Disadvantages**
For a neural network, the important concepts cant be identifies	Need to spend more time to analyze.
The limitations can be identified.	

Whether the author is satisfied with his work he has done.	

Table 4-3 Advantages & disadvantages of analyzing previous systems

4.4.3 Brainstorming

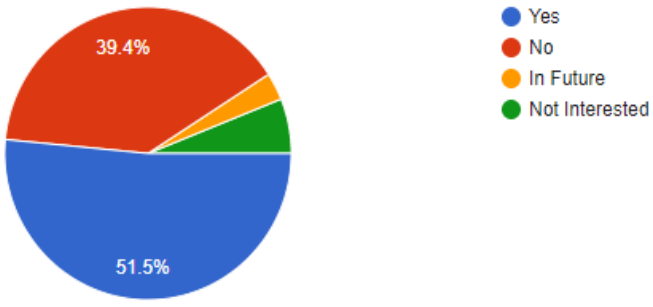
Brainstorming is a strategy plan, groups utilize to produce thoughts to unravel clearly defined plan problems. The author has done several brainstormings from industrial experts and colleagues at different stages of the project.

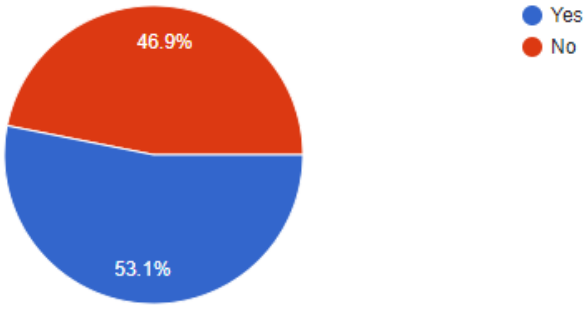
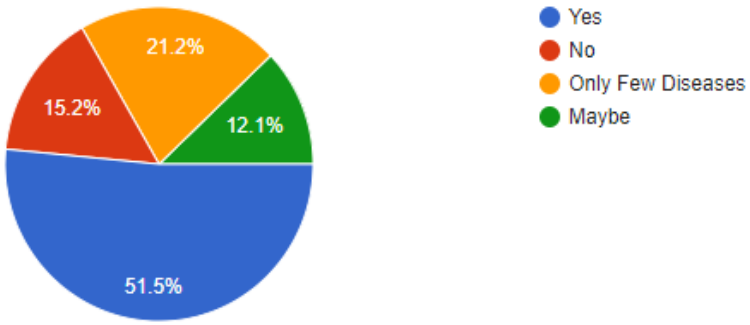
Advantages	**Disadvantages**
Can easily find out the loup holes or a new requirement	It may well be overwhelming for a single individual to figure out.

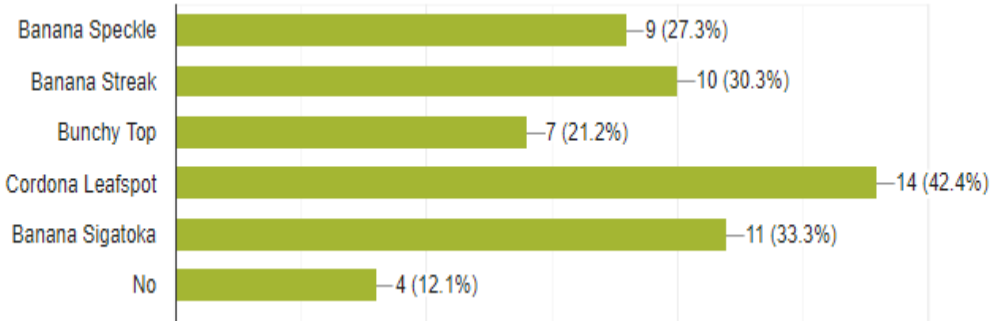
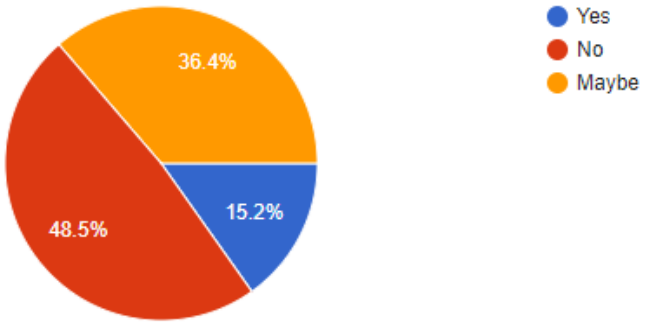
Table 4-4Advantages & disadvantages of brainstorming

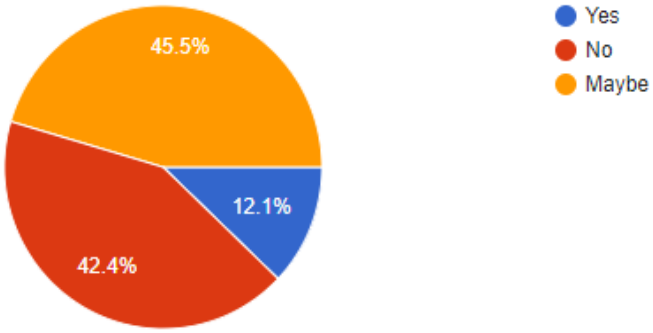
4.5 Discussion of Results

To summarize the findings author has created a Google form survey to gather requirements. The questions were made according to a standard to read and identify easily. The following table includes the aim, observation of the results, and conclusion.

Question	**Do you have experience in plantation?**
Aim	To Analyze user experience in plantation
Observation	While analyzing the user responses it shows that 51.5% were familiar and 39.4% were not familiar with the plantation. Rest 10% maybe planning to plant in future or not interested for plants. 
Conclusion	According to the questionnaire, it shows that most of the people were engaged with plantation.

Question	**Are you or any of your family members/relatives engaged in banana plant cultivation?**
Aim	This is the most awaited questionnaire, and according to this responses the author identify who were interested in banana plant cultivation
Observation	53.1% responded that they are familiar with banana plants and 46.9% responded that they are not capable of banana plant cultivation. 
Conclusion	This shows that most of the people were engaged in banana cultivation. This pie chart explains the progress for a better output.
Question	**Do you know that there are different types of banana plant diseases that destroy the entire cultivation?**
Aim	To identify whether people know about different types of banana plant diseases
Observation	51.5% of respondents said that they are aware of these diseases and 21.2% knew few diseases and the rest 27% are not aware of these diseases. 
Conclusion	According to the responses, most of the users know that the entire cultivation can be destroyed by a virus.

Question	**Have you seen any of these diseases before?**
Aim	To identify the most famous diseases among people
Observation	Cordona leafspot 42.4% and banana sigatoka 33.3% are the most known diseases among users. 30.3% responded that they have seen the banana streak virus. Bunchy top 21.2% and banana speckle 27.3% are the least known viruses. Others haven't seen any of these diseases before. 
Conclusion	Cordona leafspot has the highest popular rating compared to other diseases. But few 12.1% of people don't know about banana plant leaves. But the majority know about banana plant diseases.
Question	**Do you think that you can identify banana plant diseases with your exposed eye?**
Aim	To identify whether people can recognize banana plant diseases from their naked eyes
Observation	48.5% have responded that it's hard or unable to identify banana plant diseases from their naked eyes. 36.4% said sometimes they can identify from their exposed eyes. Rest 15.2% can identify diseases from their exposed eyes. 

Conclusion	The majority can't or hard to identify such diseases from their naked eyes.
Question	**Do you think the farmers have enough knowledge to identify these types of diseases?**
Aim	To find out how the users responded from a farmer's perspective
Observation	45.5% said that farmers might have the experience to identify such diseases and 42.4% responded that they are not capable of identifying these diseases. The remaining 12.1% think that farmers have enough knowledge to control these viruses. 
Conclusion	This shows that farmers or whoever wishes to grow banana plants should know to identify these diseases at the early stages. At this stage, the planters should improve their knowledge through this application.
Question	**Have you used a web/mobile application to identify banana plant diseases?.**
Aim	To check whether the people have used any kind of software to identify banana plant diseases.
Observation	All the reactions show that they haven't used an application to be mindful.

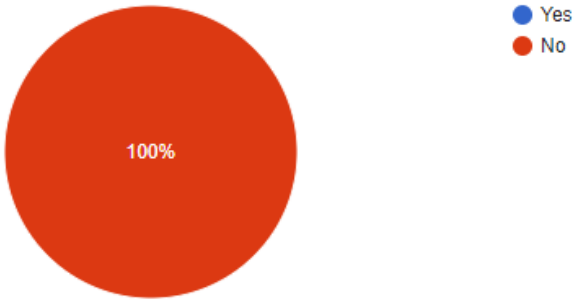
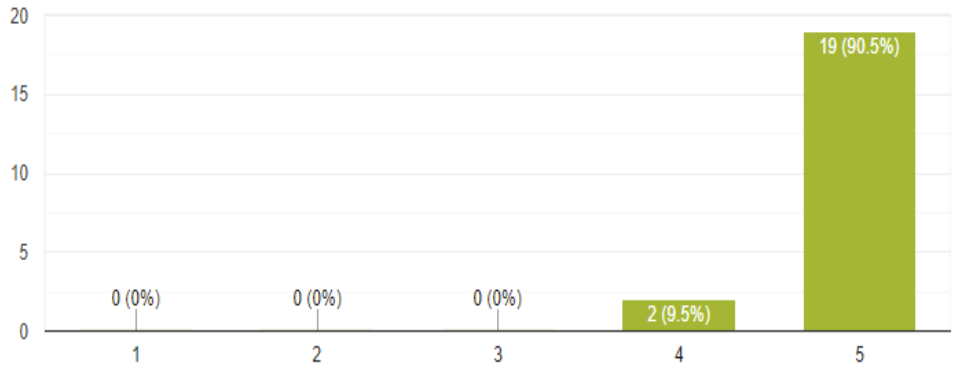
	
Conclusion	There is a market space to introduce this application. Moreover, there should be a place to learn about the solutions to avoid banana plant diseases.
Question	**Do you wish to use a mobile/web application to be aware of these viruses? How satisfied are you with this application?**
Aim	To identify the target audience who is willing to use this application
Observation	90.5% recommended with 5 ratings and 9.5% recommended with 4 ratings. 
Conclusion	According to the feedback, almost everyone is like to use a web/mobile application to aware from these viruses.

Table 4-5 Online survey responses and conclusions

4.6 Summary of Findings

Found Elicitation	Questioning the Audience	Brain Storming	Analyzing previous Systems
Should use more models for prediction.	✓		✓
Should gain a high accuracy for training and testing.		✓	✓
A web application with an interface is required.	✓	✓	
Should display the predicted disease and need an expert to clarify more regarding the disease detected	✓	✓	

Table 4-6 Summary of requirement elicitation

4.7 Context Diagram

This diagram describes the flow of the banana plant disease detection system.

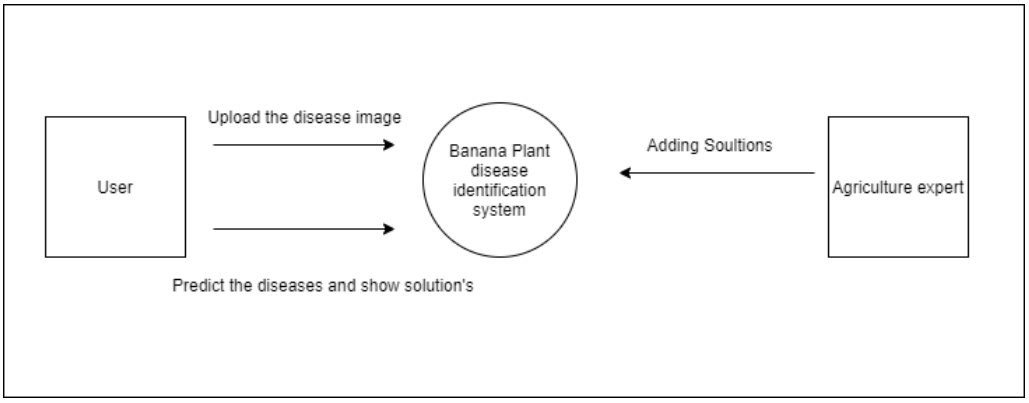


Figure 4-3 Context diagram

4.8 Use Case Diagram

The use case diagram shows the relationship between actors and use cases. Users can register and login into the system. There is an authentication when logging into the system. After a successful login user can upload an image and check whether it's a defective leaf or not. The system will predict the disease and display it to the user. If the user wants more clarification related to pesticides, he/she can contact a specialist.

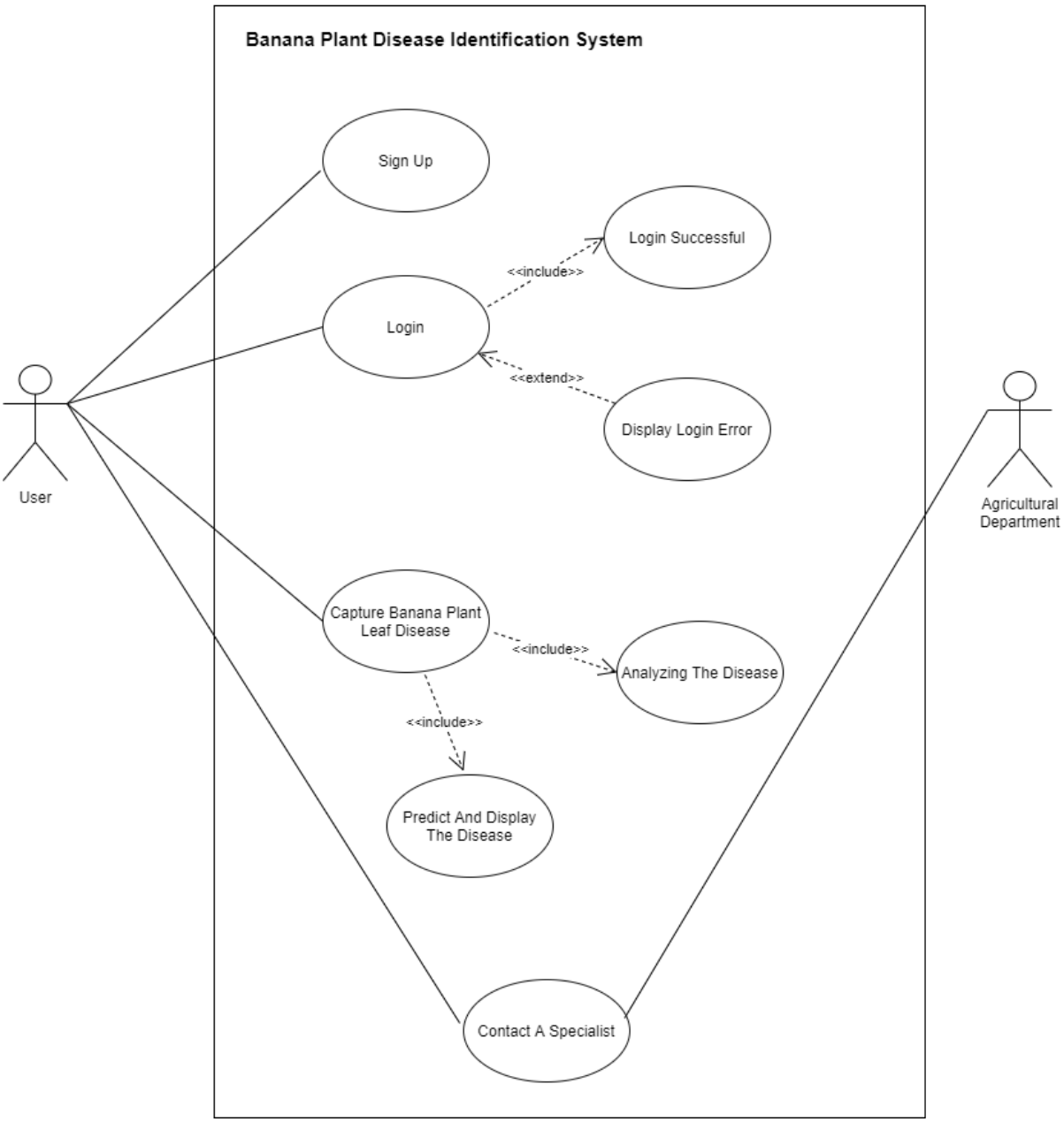


Figure 4-4 Use case diagram

4.9 Use Case Description

According to the use case diagram the main use cases has been shown in the given tables. The rest tables are shown in appendix B1 and B2.

1. Capturing a Banana Plant Leaf Use Case

Use Case	**Capturing banana leaf**
ID	003
Description	The user has to take a photo of a banana leaf
Priority	High
Actors	Users
Summary	This is the most important use case because there are five diseases, which predict according to what user uploads.
Pre Condition	User has to login to the system.
Post Condition	The system will predict the disease class which belongs to.
Triggering Event	Upload the image and click predict button
Main Flow	1. User needs to upload a valid image 2. Needs to click predict button 3. From the system, it will open the upload an image window. 4. System will match the image features with trained image features. 5. System will predict and display the disease

Alternative Flow	**Scenario 1: User uploads an invalid image** 1. System displays a message showing "please upload a valid image" **Scenario 2: User submits predict button directly** 1. The system displays an error message
Exceptional Flow	**1.** Sometimes if the application crashes it will display a message like "connection lost".

Table 4-7 Capturing banana plant leaf use case

2. Contact an Expert Related to Banana Plant Diseases Use Case

Use Case	**Capturing banana leaf**
ID	004
Description	User can contact an expert related to banana plant diseases for more details.
Priority	High
Actors	Users
Pre Condition	User has to log in to the system.
Post Condition	The image has to send to the expert to review.
Triggering Event	Click the contact now button
Main Flow	1. User needs to click the contact now button 2. The image has to send for review.
Exceptional Flow	1. Sometimes if the application crashes it will display a message like "connection lost".

Table 4-8 Contact an expert use case

4.10 Requirements

In this phase, the requirements of the project are listed below according to their priority level.

Priority Level	Description
Critical	A core function of the system
Important	Not essential, but its recommended to be done
Non-important	At this stage, it's out of the scope or it’s future development.

Table 4-9 Requirement priority level

4.10.1 Functional requirements

All the functional requirements related to this project can be seen from the below table.

FR No	Requirement title and description	Priority Level
FR 01	New users need to register with the system to continue with the process. This has a sign-up form.	Important
FR 02	Registered users need their username and password to log in to the system	Important
FR 03	Allow users to upload banana plant leaves with diseased or non-diseased.	Critical
FR 04	Image augmentation	Critical
FR 05	Training the dataset with a combination of algorithms. Each image has been resized to a standard size.	Critical
FR 06	Classify images to different classes using classification algorithms. Each classification has an accuracy.	Critical
FR 07	The final predicted disease or the result will be displayed to the user.	Critical

FR 08	If the user wants to clarify more details related to that particular disease the user can contact an expert related to banana plant diseases.	important

Table 4-10 Functional requirements

4.10.2 Non-Functional requirements

All the non-functional requirements related to this project can be seen from the below table.

NFR No	Requirement title and description	Priority Level
NFR 01	**Performance:** The system should work with a good speed without crashing or any other disturbances.	Critical
NFR 02	**Usability:** The UI should be simple and easy to understand.	Important
NFR 03	**Accuracy:** When the user predicting the disease it should predict with a reasonable level of accuracy.	Critical
NFR 04	**Maintainability:** The developer should be able to maintain the application at any time.	Important
NFR 05	**Scalability:** When the application is hosted with the server it should be able to requests data without breaking.	Desirable

Table 4-11 Non-functional requirements

4.11 Chapter Summary

The System Requirement Specification is a main chapter in the report. This chapter started with the project-rich picture. Before analyzing the prerequisites, the stakeholders of this system were distinguished. The author has identified the real-world stakeholders and drawn a model to explain more briefly. Under this chapter, the use case diagram, use case description the functional requirements and non-functional requirements were defined. The following chapter will be talking about almost the design, flow, and execution of the proposed solution

CHAPTER 5: SOCIAL, LEGEL, ETHICAL AND PROFESSIONAL ISSUES

5.1 Chapter Overview

In this chapter the guidelines or the rules to be followed during designing the system as the system consists of personal data.

5.2 SLEP Issues and Mitigation

1. Social

When developing a system it's necessary to consider the following aspects like public privacy, health, security, environment. Third parties shouldn't allow to grab data from the system. The user responses were added to the thesis and it's not against the user privacy policy.

2. Legal

The dataset was not an open-source database. The office packages and IDE's have licenses and registered under University of Westminster. The publications created as part of the research were legitimately copyrighted and transferred to the distributor.

3. Ethical

From the questioner, no personal details were collected except personal opinions. For the evaluation, the domain experts and evaluators granted permission to display their details on the report.

4. Professional

Computers used to develop the system were password protected and updated with latest OS versions. Haven't used any kind of third-party or pirated software to make this research. Data storing devices like flash drives, external hard drives were password protected. From the beginning to the end all the legalization laws were followed.

5.3 Chapter Summary

This chapter is a proof to provide that the author has followed SLEP laws to develop the system. In this chapter, the four concepts of SLEP are being summarized.

CHAPTER 6: SYSTEM ARCHITECTURE AND DESIGN

6.1 Chapter Overview

Under this chapter, the system design, architecture and the proposed solutions were explained. The class diagram will explain how the relationships were connected with the class instances. The high-level architecture will explain the overall system design. Furthermore, the sequence diagram and wireframes are included in this chapter.

6.2 Design Goals

In this system, the design objectives are as follows.

Design Goal	Description
Performance	Performance is another key aspect to be examined when designing the system. The main purpose of this system is to identify banana plant diseases and if it doesn't identify the disease properly the application is not useful. So the accuracy and the performance should be good.
Usability	All the users may not be technical experts to use this application. So the application should be easy to use with good efficiency.
Scalability	At any time the system should be able to adapt for new diseases or for new datasets. As well as the system should be able to change existing algorithms with new algorithms.
Accuracy	The system efficiency is based on training accuracy. For the selected diseases the accuracy should be high.
Reusability	The trained weights should be reusable for other developments

Table 6-1 Design objectives

6.3 System Architecture Design

What is System Architecture Design? System architecture design is kind of a conceptual representation of the elements and sub-elements that contrasts the behavior of a system. The below graph shows the design of the system. There are benefits of maintaining a system architecture design such as high productivity, quality of the system, ease of code management and many more. This architecture has three tires namely presentation tire, logical tier and data tier.

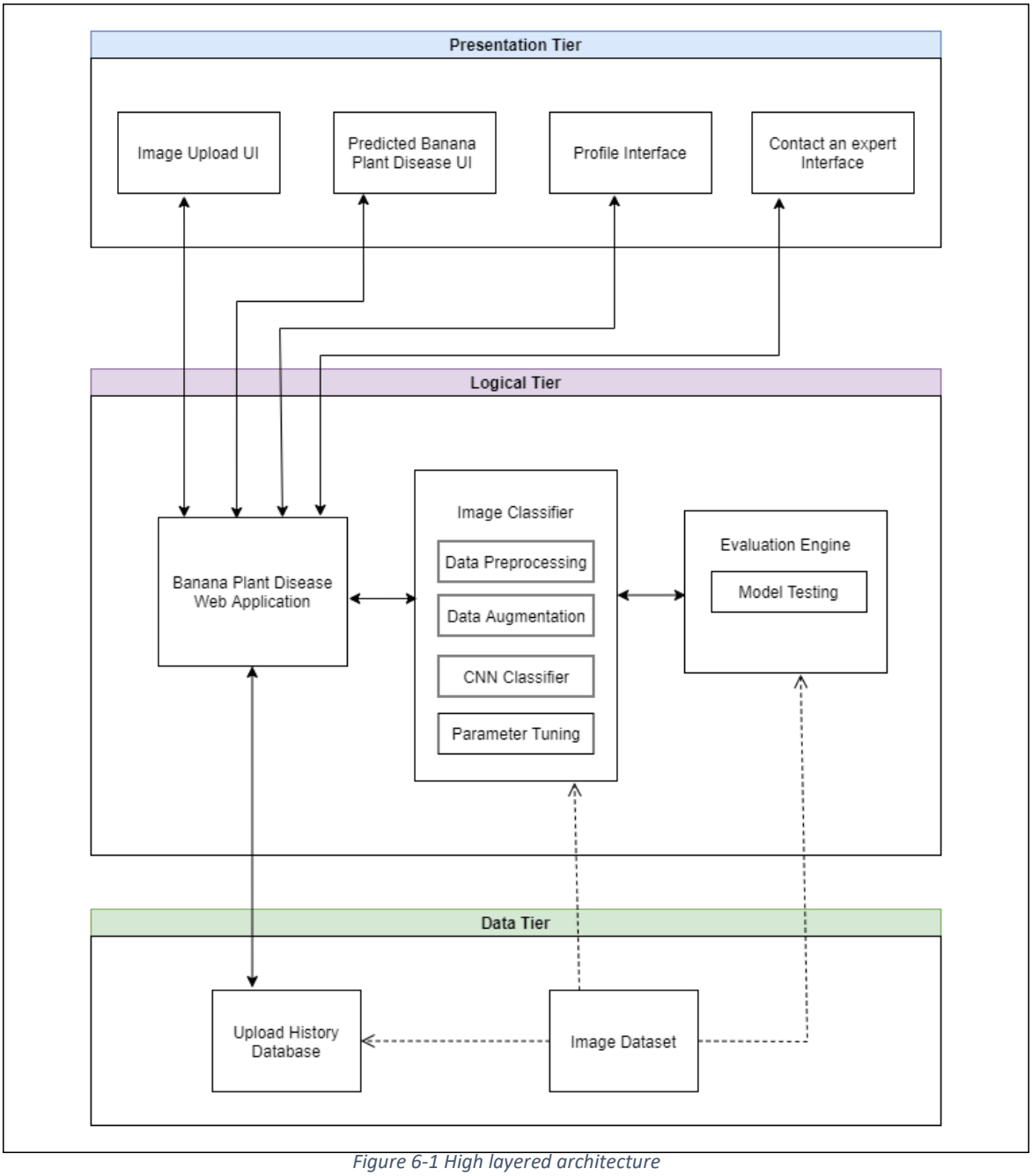


Figure 6-1 High layered architecture

So the image classifier contains the training dataset 80%, testing dataset 20%, image augmentation, training algorithms, parameter tuning and etc. Finally, the web application provides users to identify banana plant diseases with a single click.

1. Data Tier

This is the final layer which provides the facility to store data and to fetch data to the system. In this system, there are two data storage units when executing the program. The profile details and the uploaded history will be saved in the database.

2. Logical Tire

The logical tier is the functionality layer which controls the application by performing a group of processes. Once the user inputs an image it will go through several logics and displays the output. The Functions are

   A. Data Preprocessing - This is an important step before the training process, the dataset needs to preprocess using OpenCV or GNU-Octave.
   B. Data Augmentation - The dataset is not an open-source dataset, so the author needs to augment the dataset because it's very small. To augment data, we can use OpenCV or lightrooms to generate similar images with rotations, flipping, color changes and etc.
   C. Implementing the Model - For better accuracy with 16 layers, a residual neural network has been used. For this Tensorflow and Keras are the main libraries which the author has used.
   D. Training the Model - Residual network can be used to get a good accuracy or to avoid over fittings and under fittings.
   E. Prediction - The Application needs to predict the exact disease according to the trained weights.

3. Presentation Tire

The Presentation tier consists with the system UI’s which users used to interact. This tier has HTML templates to display information to the user. Once the user captures an image and uploads it, the UI’s will change from “upload an image UI” to “Predicted banana plant disease UI”.

6.4 System Design

6.4.1 SSADM Analysis

Structured systems analysis and design methodology is based on the waterfall model. SSADM is the serious user association within the prerequisites analysis stage. The users are given clear, effortlessly reasonable documentation comprising of different diagrammatic representations of the framework. In the SSADM method, there are three important techniques. Such as

1. Logical data modeling
2. Data flow modeling
3. Entity event modeling

The advantages of SSADM

1. For the developers, it's easy and simple to understand
2. The product can be released on time
3. The activities can do according to a sequence.

The disadvantages are

1. This method is not accurate for large systems.
2. This method needs more time to analyze.

6.4.2 Sequence Diagram

This sequence diagram clarifies how a user can be connected with the application to identify banana plant diseases and to aware from them. New users need to register, and the details will store in the database. The next step is to log in to the system using the registered email and password and upload the image to the server. The system will analyze the disease using the trained algorithms and display it to the user.

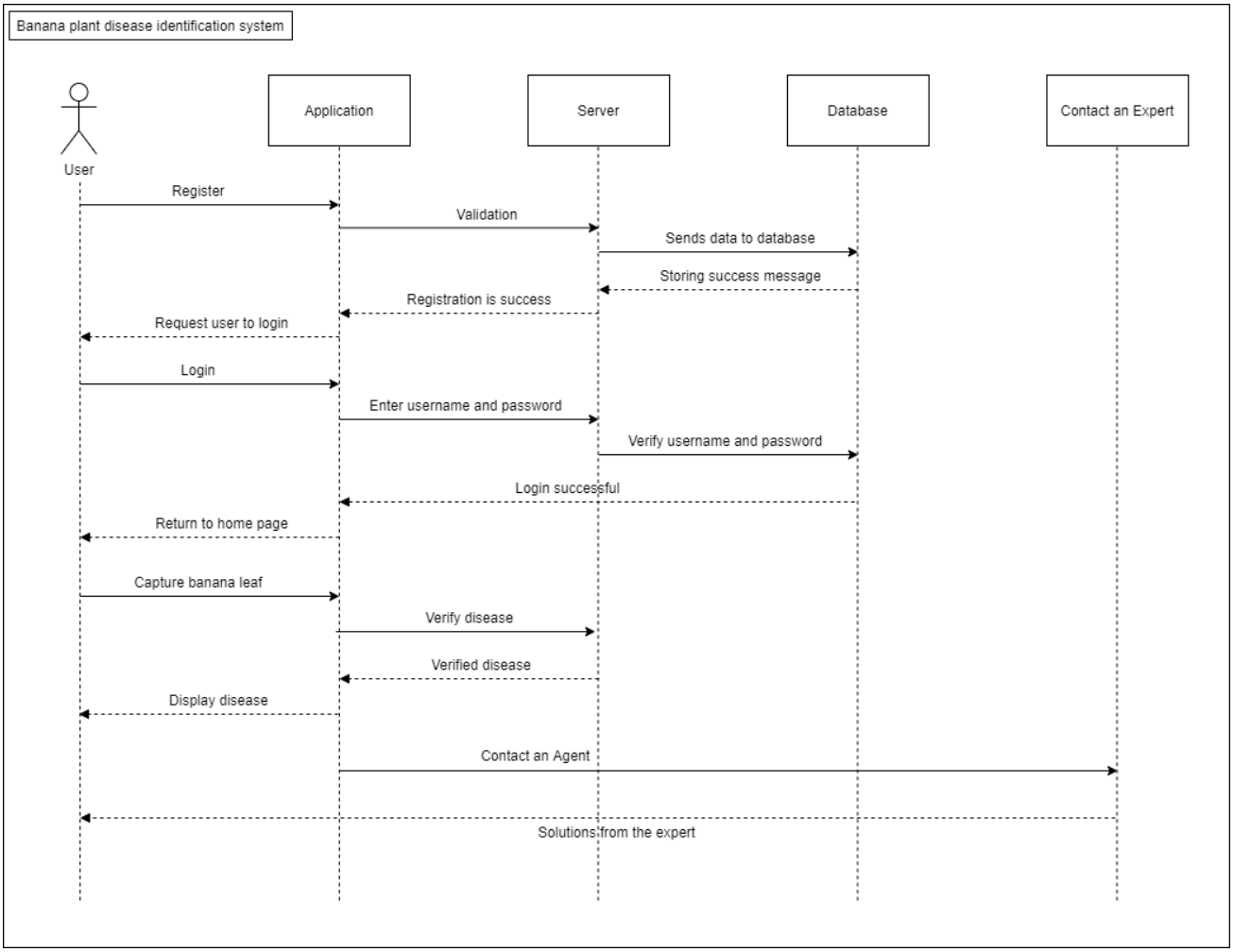


Figure 6-2 Sequence diagram

6.4.3 Component Diagram

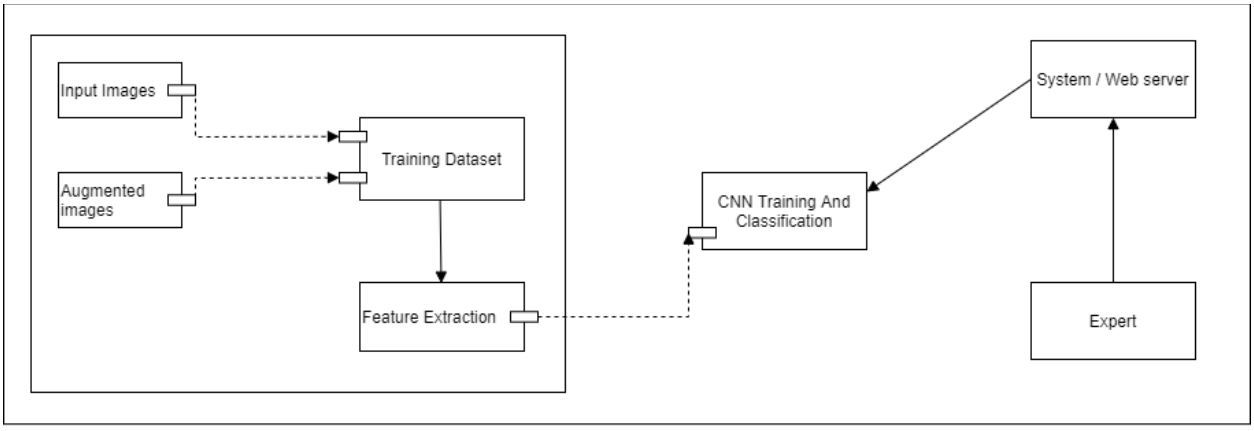


Figure 6-3 Component diagram

6.4.4 UI Wireframes

As the system is developed for the users to upload banana plant leaf diseased images there should be several UI's for users for the inputs and for the output results.

Figure 6-4 Ui wireframes

6.4.5 Process Flow Chart

The following flowchart expresses the flow of the system and its techniques that are used to identify banana plant diseases. According to figure 6.5, the application has used data augmentation techniques like image flips, rotations, background color changes, image blurrings, adding noise, etc.

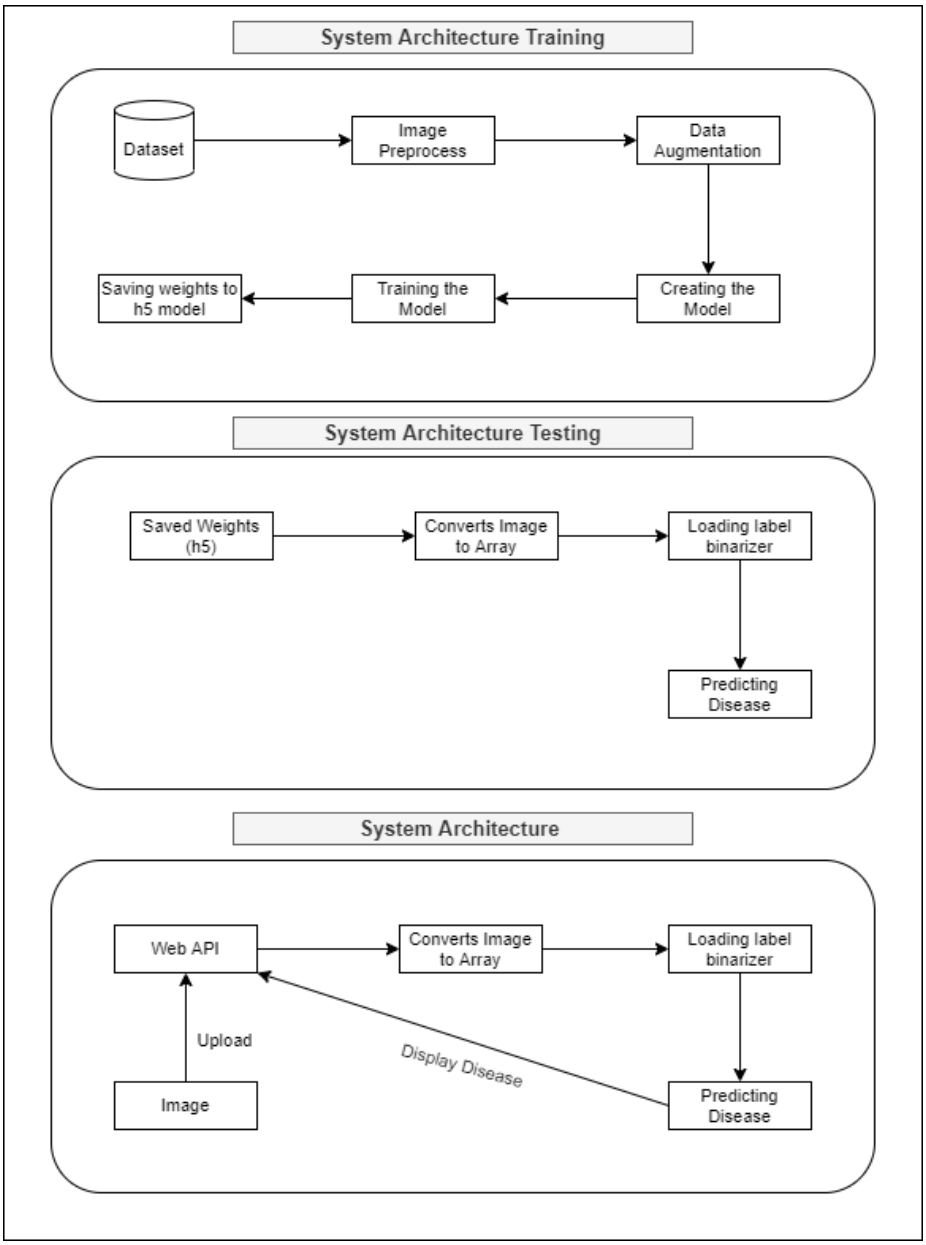


Figure 6-5 Flow chart

6.5 Chapter Summary

This chapter has given a point-by-point analysis of system architecture and design. In the beginning, the design goals were defined. The importance of SSADM analysis, as well as the advantages and disadvantages, are explained in this chapter. The context diagram, data flow diagram, high-level architecture, UI wireframes, and the process flow chart were displayed in this chapter. The next chapter will describe the system implementation.

CHAPTER 7: IMPLEMENTATION

7.1 Chapter Overview

This chapter will show how the proposed solution is going to implement by the author. Mainly this chapter will express how the system design is going to convert to runnable code. Under this chapter, the development tools, the programming language, libraries that are used to utilize the accuracy and the datasets are documented. Finally, the problems faced during development and the problem-solving techniques are presented.

7.2 Technology Selection

7.2.1 Technology Stack

The technology stack chosen to be utilized within the different layers of the framework are as follows.

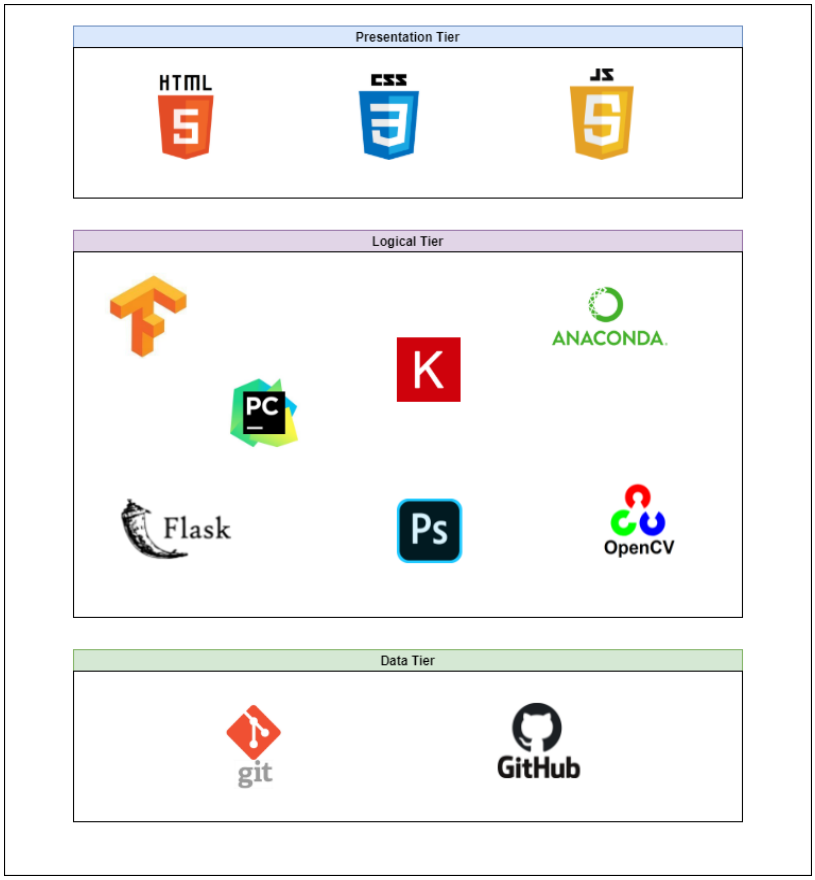


Figure 7-1 Technology selection

7.2.1 Data Selection

For a data science project, we need data. The performance of a project depends on the size of the data. Hence the first doubt that the author had to find out appropriate datasets available for the project to be built. Unfortunately, the author couldn't find a suitable dataset from online resources to carry out the project. The first move was to visit the Kaggle website and find out a valid dataset, and there were similar images but couldn't use them because they are not valid. Then the author had to contact an agent at the Department of Agriculture Central Province to find out a dataset. As the author expected they didn't have a dataset for banana plant diseases. Finally, the author had to take images of the relevant diseases from several areas like Kandy, Colombo, Batticaloa, Mathara. So the author was able to capture images of the banana sigatoka, banana speckle, banana bunchy top, banana streak, cordona leaf spot and the good leaves.

7.2.2 Programming Language

A programming language is a kind of computer language which used by programmers and researchers to develop software applications, scripts or any other application for computers to assassinate. In modern days there are number of programming languages used for several perspectives. For a machine learning project, developers use Python, R, Ruby, C# mainly. For this project, the author has decided to use python because it has many supportive libraries which are useful for ML and AI. And this language has vast community support. So the author has chosen libraries like Scikit-learn, Scikit-image, Tensorflow and Keras to integrate python.

7.2.3 Deep Learning Libraries and Frameworks

Python is a powerful language which has many supportive libraries. Tensorflow, Keras and Pytorch are the most popular libraries. After analyzing several aspects the author has decided to use the following libraries.

Name	Description	Features
Tensorflow	Made by Google one of a powerful library for deep learning based on theano.	Works perfectly for mathematical expressions including multidimensional arrays. Either can be used for CPU or GPU

Keras	This was written in python and it's a deep learning API made for deep learning platforms	This supports almost all the CNN, ANN, RNN models. Works for CPU and GPU as well.
Scikit-learn	Scikit-learn is an open-source machine learning library written in python.	Can be used for regression, clustering, classifications, and SVM machines.
Numpy	Numpy is a useful fundamental package written in python.	Can be used for advanced mathematical functions, multidimensional arrays, and for confusion metrics.
Flask	This was written in python and it's a micro web framework.	Flask doesn't need tools are libraries. Support WSGI 1.0 compliant Unicode based framework
OpenCV	OpenCV is a cross-platform open-source library using for computer vision purposes and Intel has developed this.	Works perfectly for image processing, object detection, and real-time computer vision applications.

Table 7-1 Technologies and tools

7.2.4 IDEs

PyCharm was the main IDE used by the author to develop the project. PyCharm has the ability to switch between project interpreters quickly. Pycharm has used to do the frontend development. It has a powerful debugging console which was very useful when debugging.

Photoshop CC 2019

Photoshop was the IDE used to do data augmentations like image rotations, flipping, color changes, image rotations and etc.

7.2.5 Summary of Technology Selection

Component	Technology/Tool
Programming Language	Python
ML Libraries	Tensorflow, Keras, Scikit-learn, Scikit-image Numpy, Pandas, OpenCV
Virtual Environment	Anaconda
IDEs	PyCharm, Photoshop CC
Version Control	Github
Server	Flask

Table 7-2 Summary of technologies and tools

7.3 Implementation of Core Functionalities

7.3.1 Data Preprocessing

Data preprocessing is an important step in image processing. It's a kind of data mining technique to clear data or to add missing data. In this scenario, most of the images were taken from a mobile phone. Hence the author is trying to gain good accuracy from the training process the first move was to remove the image background and add a white color background instead. For this Photoshop CC was the best software to be used.

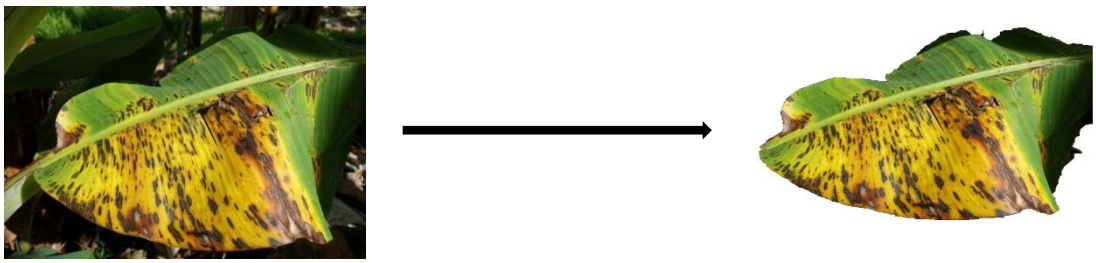


Figure 7-2 Image preprocessing

Tensorflow.keras.preprocessing and img_to_array() libraries were used to convert PIL images to a numpy array. Using the OpenCV library the images were read and resized to 180*180. The size can change at any time but it needs more processing power to process.

```
image = cv2.imread(imagePath)
image = cv2.resize(image, (IMAGE_DIMS[1], IMAGE_DIMS[0]))
image = img_to_array(image)
data.append(image)
```

After loading images to a numpy array, the next step is to load LabelBinarizer to transform class names to binary labels. To call LabelBinarizer we need to import the scikit-learn library.

```
mlb = MultiLabelBinarizer()
labels = mlb.fit_transform(labels)
```

During the training the dataset needs to split 80% for training and 20% for testing. This can be done using sklearn.model_selection method.

```
(trainX, testX, trainY, testY) = train_test_split(data,
                                                  labels, test_size=0.2, random_state=42)
```

7.3.2 Data Augmentation

Data science projects are based on datasets. The system efficiency, performance, reliability are depend on the dataset which is using. Data augmentation is a technique which helps to expand the dataset artificially. To expand the dataset the author has done several techniques. During the training process. From the keras ImageDataGenerator() can augment data while training.

```
aug = ImageDataGenerator(rotation_range=25, width_shift_range=0.1,
                         height_shift_range=0.1, shear_range=0.2, zoom_range=0.2,
                         horizontal_flip=True, fill_mode="nearest")
```

Photoshop CC and OpenCV were used to do image blurrings, adding noise, background color changes, rotations, etc. A sample set of images can be seen below.

Figure 7-3 Image augmentation patterns

7.3.3 Implementing the Residual Network

The different machine learning models were discussed in previous chapters. This project belongs to the supervised learning category. After preprocessing and augmenting data the next step is to implement the network. The residual network belongs to the CNN category which has convolutional blocks and skipping blocks. In this scenario, the author has used 16 convolutional layers combined with 9 convolutional blocks and 3 skipping blocks. The main purpose was to use this residual model is to get a higher accuracy. At the end of the training, there is a dense layer valued with 0.5 to reduce the overfitting problems.

```
# Common Block
X0 = Conv2D(16, (3, 3), padding="same", strides=(1, 1), name="con_layer1")(X_Input)
X0 = BatchNormalization(axis=3, name="B1")(X0)
X0 = Activation("relu")(X0)
X0 = MaxPooling2D(pool_size=(2, 2), strides=(1, 1), name="max1")(X0)

# Block 1
X = Conv2D(32, (3, 3), padding="same", strides=(1, 1), name="con_layer2")(X0)
X = BatchNormalization(axis=3, name="B2")(X)
X = Activation("relu")(X)
X = MaxPooling2D(pool_size=(1, 1), strides=(1, 1))(X)

# Block 2
X = Conv2D(32, (5, 5), padding="same", strides=(1, 1), name="con_layer3")(X)
X = Activation("relu")(X)
X = BatchNormalization(axis=3, name="B3")(X)
X = MaxPooling2D(pool_size=(1, 1), strides=(1, 1))(X)

# Block 3
X = Conv2D(64, (3, 3), padding="same", strides=(2, 2), name="con_layer4")(X)
X = BatchNormalization(axis=3, name="B4")(X)
X = Activation("relu")(X)
X = MaxPooling2D(pool_size=(1, 1), strides=(1, 1))(X)
print(X)
# Skipping Block 1
XS = Conv2D(64, (3, 3), strides=(2, 2), padding='same', kernel_initializer=glorot_uniform(seed=1), name="con_layer5")(X0)
XS = BatchNormalization(axis=3, name="B5")(XS)
XS = Activation("relu")(XS)
print(XS)
#########################################################################################################

# Adding layers
X = tensorflow.keras.layers.Add()([X, XS])
X = Activation("relu")(X)
```

The first block is the common block starting from 16 (filter size) with (3,3) kernel size. The architecture of the residual is documented in **Figure 7-4**.

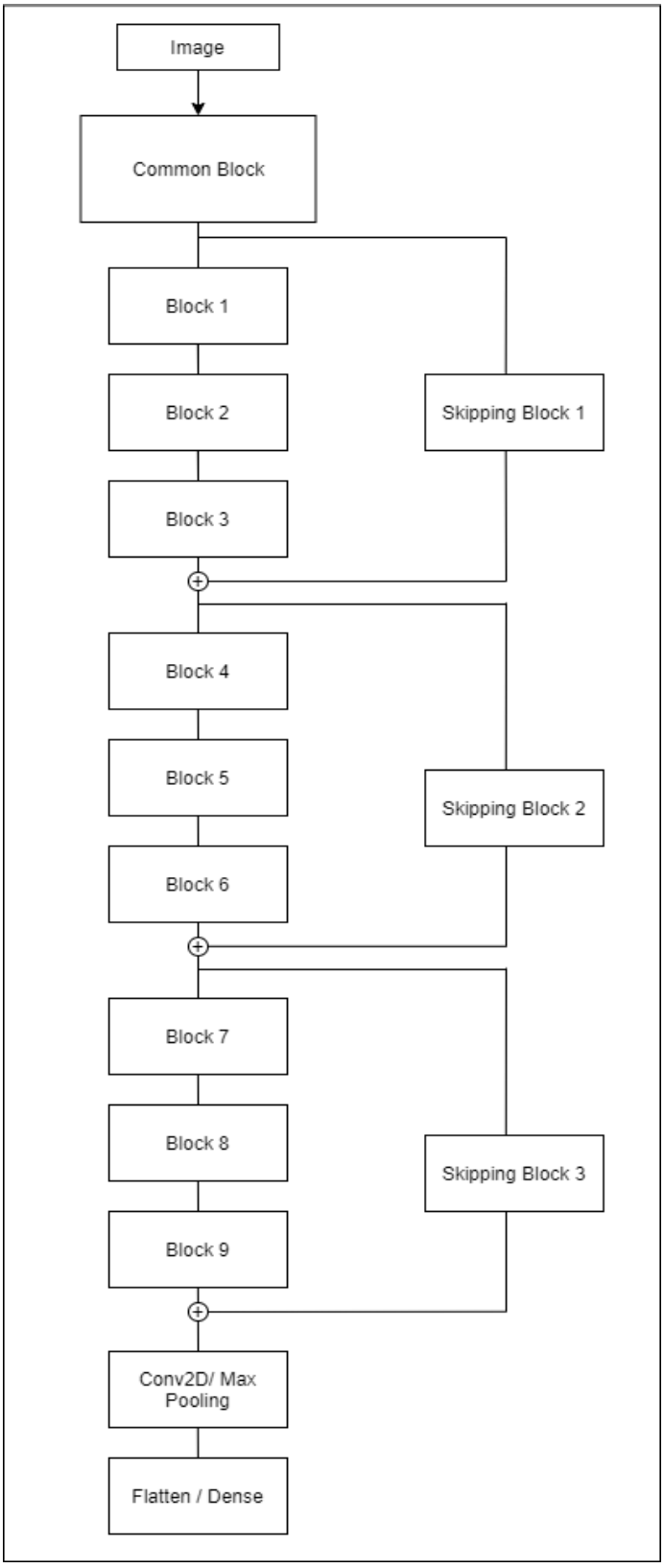


Figure 7-4 Residual model architecture

What is a convolutional layer? A Convolutional layer receives input value and sends tensor outputs. This layer consists with a set of parameters to be learned. Ex: filters, kernels, strides. And moves to the activation function.

What is an Activation function? An activation function is used to identify the neural network mappings. Ex: 0 or 1, 1 or 1 or etc..There are activation functions like Sigmoid, Relu, Tanh, Leakyrelu, softmax. Among them the best activation function was Relu.

What is a Batchnormalization? It's the process of normalizing the input data. Which means it helps to standardize the input data to a layer. This helps to sustain the learning rate significantly decreasing the training epoch size. For this model 13 batch normalization layers were used.

7.3.4 Loss Function / Optimizer

Optimizer is the key feature that parallelly works with the model dataset. Optimizer is an inbuild function on keras. There are optimizers like SGD, Adam, RMSprop, Adadelta. Among them, the author has decided to use the Stochastic Gradient Descent function. There are advantages when using SGD function. The learning rate was set to le-3 and the minimum learning rate was 0.00001

Advantages of SGD

1. It's easy to fit into the memory because it has a single training process
2. It is computationally quick as where one test is processed at a time.

```
opt = SGD(lr=INIT_LR, decay=INIT_LR / EPOCHS)
```

To calculate the loss function first have to consider the crossentrophy. If the model has more than 2 or 3 classes it's better to use categorical_crossentrophy. If the model has less than 3 classes it's better to use binary_crossentrophy with sigmoid function. In this model, there were six classes and had to select categorical_crossentrophy with softmax classifier.

```
model.compile(loss="categorical_crossentropy", optimizer=opt, metrics=["acc"])
```

7.3.5 Early Stoppings / Checkpoints

Early Stopping is a function built-in keras to stop the training when the validation accuracy is not improving.

```
# early sttoping
early = EarlyStopping(monitor='val_loss', patience=8, verbose=1, mode='min')
```

Checkpoints are used to save random stages of the training process. So the weights can be load later and continue the training procedure. In this model, the author has decided to save the best weight point.

```
checkpoint = ModelCheckpoint(weight_name, monitor='val_loss', verbose=1, save_best_only=True, mode='min')
```

After removing the image background the author tried a sample training and analyzed that there is an overfitting problem due to the image background removal. So the best thing was to train both images to overcome the overfitting problems.

7.4 User Interfaces

A new user needs to register with the system. After registering, the user needs to log in to the system. A successful login will redirect to the home. In the home component, the upload function is there to upload a banana leaf shown in figure 7-5. After uploading the image it will classify according to the trained weights and displays the output according to figure 7-7

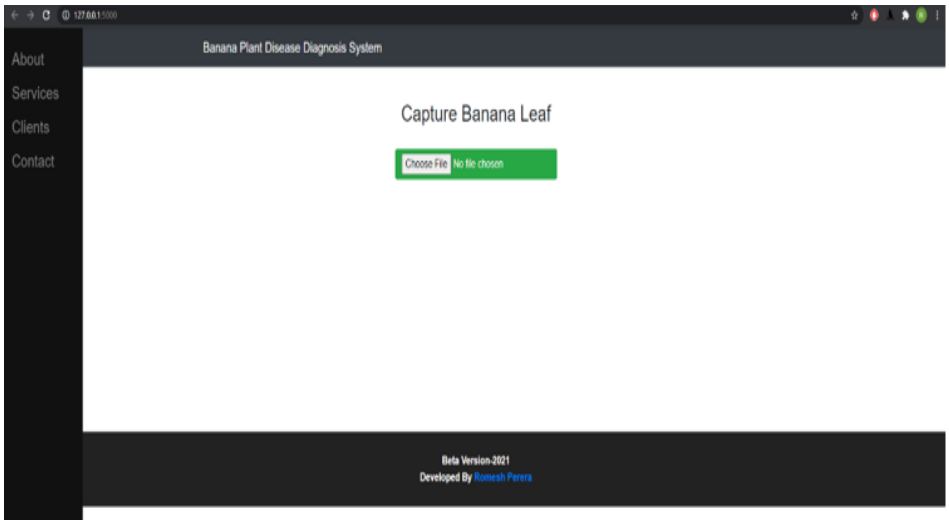


Figure 7-5 Upload an image interface

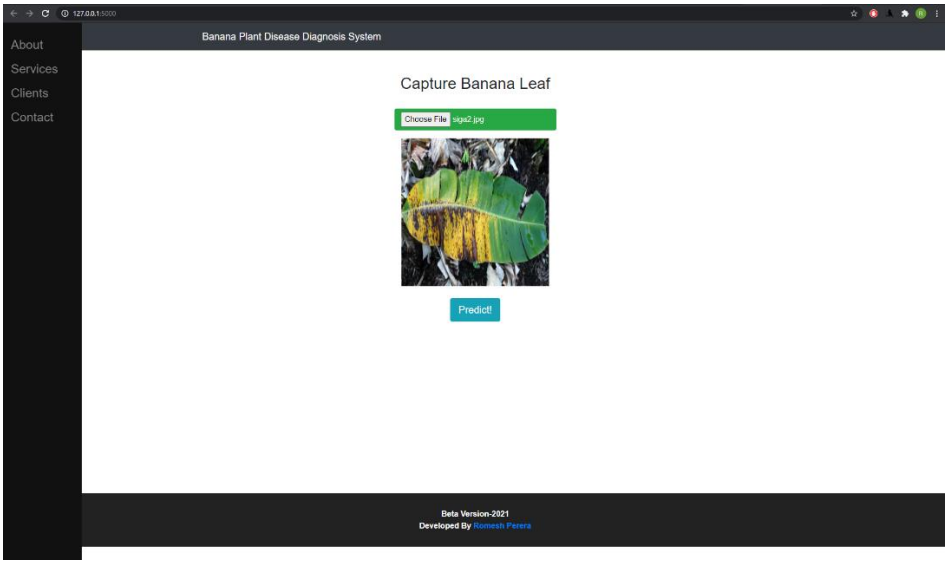


Figure 7-6 Uploaded image interface

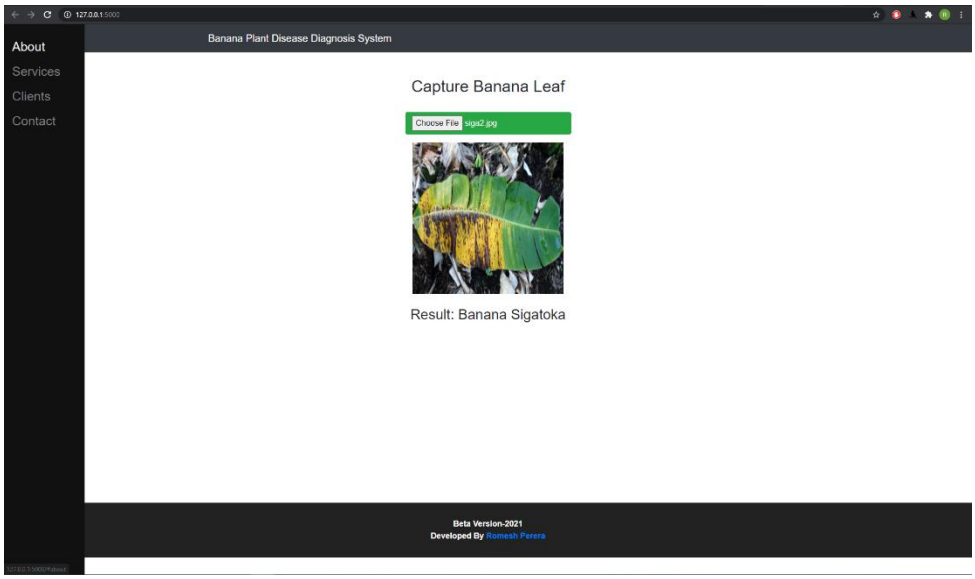


Figure 7- 7 Predicted disease interface

7.5 Chapter Summary

From the beginning of this chapter under technology selection, the technology stack, data selection, programming language used, the technologies and libraries that integrated into the system were discussed. The system core functions like building the CNN model, augmenting the dataset, are also documented in this chapter. Finally, the design of the user interfaces were mentioned in this chapter.

CHAPTER 8: TESTING

8.1 Chapter Overview

This chapter consists with several testing methods to check whether the implemented system functionalities are working properly according to the requirement. In the beginning, the author has discussed testing goals and objectives. The criteria testing is also examined in this chapter. In the end, functional testing, non-functional testing, accuracy testing, and performance testing are defined.

8.2 Objectives and Goals of Testing

Testing is a part of software development to guarantee that the software is met the expected obligation and to maintain the quality of the software. The objectives of testing can be identified as follows

1. To authenticate whether the functional requirements are satisfied with the system.
2. To authenticate whether the non-functional requirements are satisfied with the system.
3. To confirm whether the residual model is working properly.
4. To find out bugs and to resolve them.

8.3 Testing Criteria

As the first step of testing a system, have to concern on two factors to identify the gap between the implemented system and the expected system.

1. Functional Quality - the identified and defined functional requirements which are carried out during the software design phase validate the core functions of the developed solution.
2. Structural Quality – Related to project qualities, the non-functionalities like code quality, numbering, versioning.

8.4 Model-Based Testing

Model testing is a technique in which the test cases are determined from a demonstration that depicts the useful aspects of the framework beneath the test. The advantages of doing model testing are

1. Model changes can be tested easily
2. Can analyze whether the system is performing according to a sequence

8.4.1 Confusion Matrix

Figure 8-1 displays the confusion matrix related to the system classification.

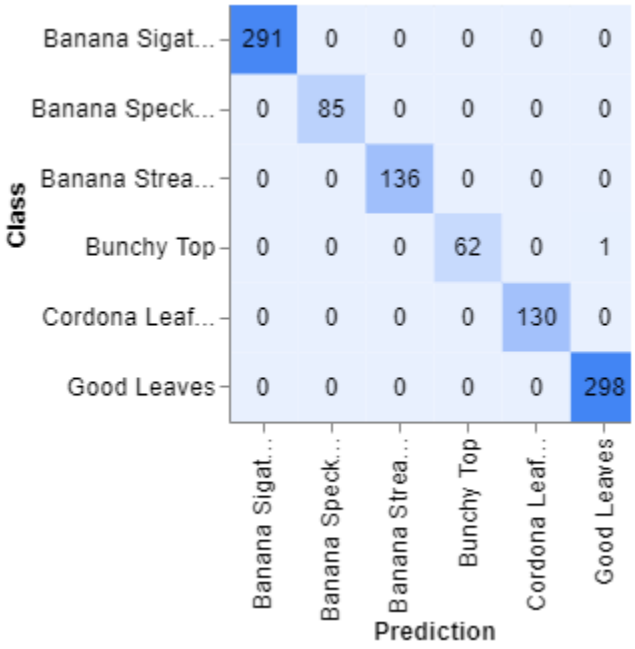


Figure 8-1 Confusion matrix

The class labels represent according to the index. **0, 1, 2, 3, 4, 5** namely **Banana Sigatoka, Banana Speckle, Banana Streak, Bunchy Top, Cordona Leafspot,** and **Good Leaves.** This has plotted using 9 convolutional blocks with 3 skipping blocks. According to the matrix, the highest value indicates the class with high accuracy.

8.4.1.1 Accuracy

Accuray will determine the percentage of correct predictions when executing for test data. Accuracy can be calculated by dividing the value of correct predictions by the number of total predictions.

8.4.1.2 F1 Score

F1 score determines the consonant mean of precision and recall values related to classification problem. The calculation can be done

$$F_1 = \left(\frac{\text{recall}^{-1} + \text{precision}^{-1}}{2} \right)^{-1} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}.$$

8.4.1.3 Precision

Precision can be is defined as the division of significant examples (true positives) among all of the cases which were anticipated to have a particular class.

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$= \frac{True\ Positive}{Total\ Predicted\ Positive}$$

8.4.1.4 Recall

Can be defined as the division of cases which were anticipated to belong to a class with regard to all of the cases that belong in the class.

$$\text{Recall} = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

$$= \frac{True\ Positive}{Total\ Actual\ Positive}$$

8.5 Benchmarking

Benchmarking is a method to evaluate our system with other systems to analyze whether the system attained with qualitative standards. This system was benchmarked with these methods

1. Competitive system benchmarking
2. Internal benchmarking

8.5.1 Competitive system benchmarking

Similar system's to banana plant disease detection systems were identified under chapter-2. For every system they have used different datasets with different architectures. Most of the researchers

were used VGG16, VGG19, LeNet, Resnet50 to train their models. To benchmark, author has used these models with two different datasets.

Dataset	VGG16	VGG19	LeNet	Resnet50	Residual Model
Images without different backgrounds	78.40	81.20	74	84.50	97.14
Images with different backgrounds	76	80.50	65.35	72.65	90.50
Images with both	88.20	90.40	80.60	92.85	98.70

Table 8-1 Competitive benchmark

8.5.2 Internal benchmarking

Banana plant disease detection system needs a high accuracy to detect diseases under different backgrounds and lightings. So the author has changed layers and learning rates to evaluate the accuracy. This benchmarking helps to find out the best architecture.

Optimizer	Learning Rate	4 blocks & 1 skipping block	6 blocks & 2 skipping blocks
Adam	Le-3	90.10	94.65
	Le-4	88.68	93.85
SGD	Le-3	91.45	96.33
	Le-4	90.55	94.49

Table 8-2 Internal benchmark

8.6 Functional Testing

Using black box testing technique, the author has decided to do the functional testing. This includes the expected output and the actual output for a given input. The functional requirements were documented under chapter 4. The summary of black box testing is as follows.

Test Case	Description	Expected Output	Actual Output	Status
01	A new user needs to sign up with the system.	User details will store in the database.	User details will store in the database.	Pass
02	Entering invalid email address and password	User login authentication should fail	User login authentication is failed	Pass
03	Registered user needs to enter valid email address and password	Login authentication should be successful	Login authentication is successful	Pass
03	The user should be able to upload an image of banana plant leaves.	The image should upload successfully	The image uploaded successfully	Pass
04	Image Augmentation	New images should augment	New images are augmented	Pass
05	The model should train and display the summary of the layers.	All the layers are connected with neurons and each layer should work perfectly	All the layers are working without skipping any block or a layer	pass
06	Training and validation accuracy should be acceptable	Training and validation accuracy should get above 90%	Both training and validation accuracy is above 95%	Pass
07	The user Should display the final predicted disease	Predicted disease should display	The predicted disease is displaying to the user.	Pass
08	To get more information user needs to contact an expert	Contact an expert	Contacted an expert	Pass

Table 8-3 Functional testing

8.7 Module and Integration Testing

This banana plant disease detection system has multiple combinations of modules joined together to build the final product. The System Design chapter describes how these modules were combined as a single unit and it's already documented. This testing method will identify whether the system works perfect after combining all modules.

ID	Module	Status
M 01	Preprocessing / Augmentation	Pass
M 02	Residual Network	Pass
M 03	Database / Server	Pass
M 04	User Application	Pass

Table 8-4 Summary of integration testing

8.8 Non-Functional Testing

For a data science project, the system efficiency is based on its accuracy and performance. So the accuracy is an important fact that the author has concerned about. Furthermore, the time consumption is also analyzed under non-functional testing.

8.8.1 Accuracy Testing

The accuracy can be calculated as follows.

$$\text{Accuracy} = \frac{\text{Number of features classified correctly}}{\text{Number of Features}} \text{ X } 100$$

In this project, the author has trained two models using different optimizers and two different datasets.

Model 1

For model 1 to calculate the loss function, the author has used Adam optimizer and categorical_crossentrophy with softmax classifier. The dataset which was used for this model has removed the image background. For model 1, the author has used 30 epochs with a 16 batch size. According to figure 8-3, the training accuracy and the validation accuracy don't show a high amount of over fittings.

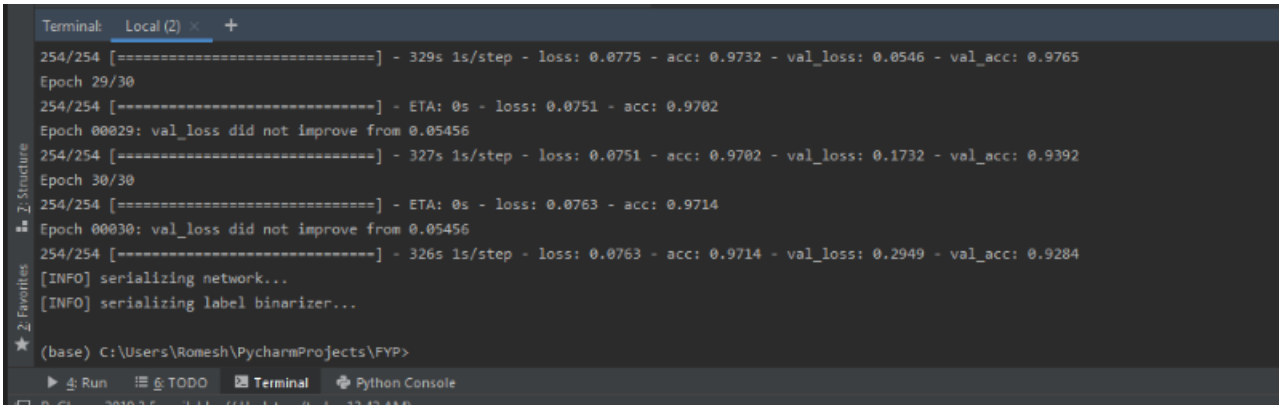


Figure 8-2 Accuracy for model 1

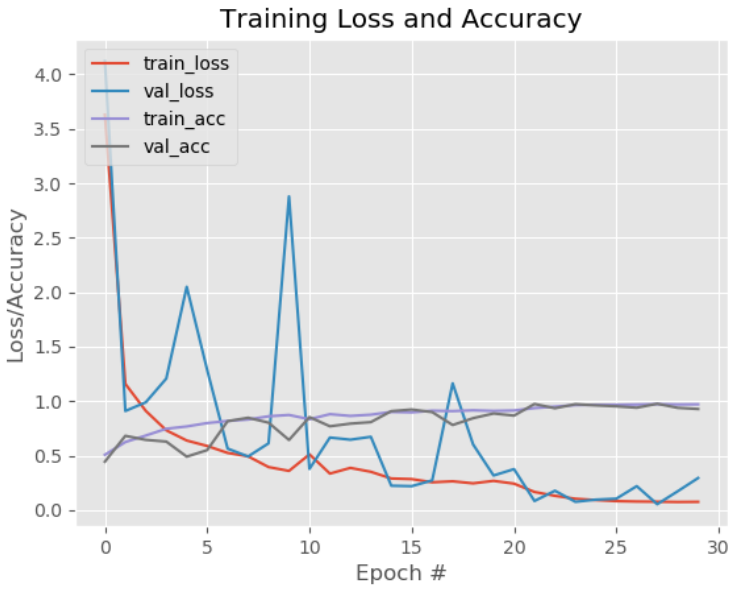


Figure 8-3 training and validation accuracy for model 1

Model 2

For model 2 to calculate the loss function, the author has used SGD optimizer and catrgorical_crossentrophy with softmax classifier. The dataset which was used for this model has removed and not removed background images. For model 2, the author has used 35 epochs with a 16 batch size and gained a training accuracy of 98.7% with 95.4% validation accuracy.

Figure 8-4 Training and validation accuracy for model 2

8.8.2 Performance Testing

The performance was tested for both models listed below.

Model	Number of Images	Epochs	Batch Size	Optimizer	Time Taken
Model 1	6,052	30	16	Adam	**4H : 32M**
Model 2	9,058	35	16	SGD	**5H : 50M**

Table 8-5 Summary of performance testing

The system was tested with the server when calling APIs to predict the disease. This was done using Google Chrome as the web browser.

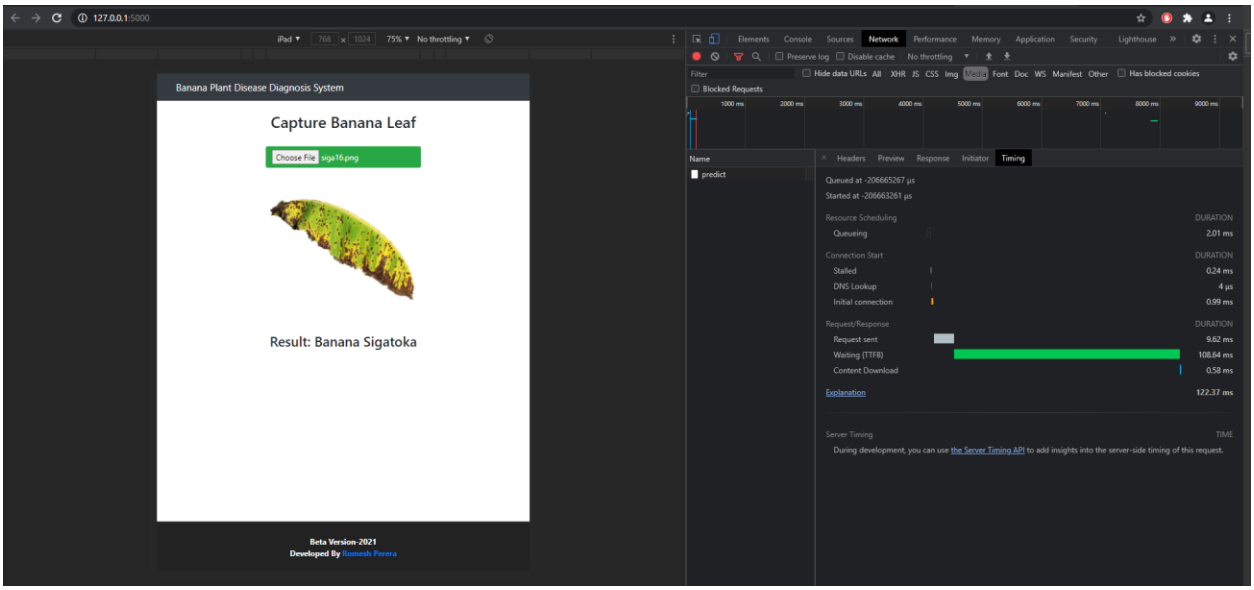


Figure 8- 5 Performance testing with the server and API

8.9 Limitations of the Testing Process

Data limitation is the main limitation that occurred when doing testing. To apply the algorithm to the real environment, need more images with different backgrounds. To test other models like VGG16, VGG19, Xception if the author had a high power GPU the testing would become perfect.

8.10 Chapter Summary

This chapter is fully covered with project testings. The chapter started with the goals and objectives of testing. Then the testing criteria and model testing were identified and divided into four subparts. Using black box testing the functional and non-functional testing has been done. Hence this relates to the data science project benchmarking, accuracy testing, and performance testing were documented.

CHAPTER 9: EVALUATION

9.1 Chapter Overview

The evaluation chapter consists with different evaluations done by the domain experts and the target audience. Most of the evaluators have raised their concerns about the system's importance, usability and conclusions. In addition, this chapter incorporates the evaluations related to the decisions made throughout the project. Finally, the author has provided a self-evaluation.

9.2 Evaluation Methodology and Approach

Since the target audience hasn't used an application like this, the author has decided to create an introductory video explaining the system and to share it among domain experts to gather evaluations. A questioner was sent to technical experts separating different sections to get feedback for each section. The qualitative data was analyzed according to the thematic analyzing method. These evaluators have evaluated whether the system is considering the current situation of banana plant diseases. According to the prototype, a quantitative evaluation has done.

9.3 Evaluation Criteria

To evaluate the developed system the author has decided to go with a quantitative approach. There were evaluators to evaluate the system according to evaluation criteria.

Criteria	Description
The project idea and the concept	To get feedback and insights on the proposed banana plant disease detection system.
Project scope	To analyze the project scope needs to evaluate the system by using domain experts
System architecture and design	To evaluate whether the proposed system architecture and design are completed up to the proper standards
Prototype and solutions	To evaluate whether the developed system is a proof of concept for the proposed solution.
Selection of algorithms	To get feedback related to used algorithms
Accuracy and performance	To evaluate the system non-functional requirements

Future enhancements	To get feedback on whether the proposed system future developments achievable

Table 9-1 Evaluation criteria

9.4 Self Evaluation

Criteria	Self-Evaluation
The project idea and the concept	The concept of developing a disease detection system is to protect our valuable assets from dead full disasters. Many researchers have done related to banana plant disease detection systems. But it's for a limited number of diseases and the performance of the system is not effective. Naked eye observation is the classical and old method way to identify such diseases. This system identifies many diseases and it will be useful for modern society.
Project scope	The scope of the project has been defined by the author in early chapters and according to the timeline needs to finish this project within six months from an undergraduate perspective. But the knowledge has taken from this research is satisfied and needs more to do in future enhancements.
System architecture and design	While analyzing the SRS the best solution was to develop the system under three-tier architecture. The UI's designed well and it's responsive for devices as well.
Prototype and solutions	The author was confused, whether to use keras or pytorch. The implemented system performance can increase in future enhancements.
Accuracy and performance	Old methods and systems like bare eye observation will give slow results. Compared to previous systems the author of this project was able to get a high accuracy and good performance
Future enhancements	Hence the accuracy is high the dataset is limited and there can be small over fittings. In the future to avoid these barriers need to expand the dataset with different backgrounds. The author has

	an idea to integrate this technology with a drone for large-scale analysis.

Table 9-2 Self evaluation

9.5 Selection of the Evaluators

To evaluate the system several evaluators were selected and sent a demonstration video of the implemented system. Due to the prevailing situation in the country, the author of this project couldn't meet domain experts and technical experts face to face. Two types of evaluators were selected to evaluate the system.

1. Domain experts working at the ministry of Agriculture department.
2. Technical experts who have used deep learning.
3. General researchers.

According to these criteria, the evaluators were agreed to display their names in this evaluation form.

Area in experience	**Affiliation**	**Justification**
Domain Experts	**Mr. Chinthaka Silva** Research Assistant Department of Agriculture (Peradeniya)	Domain experts are the ones who directly work and do experiments in plants and different cultivations.
	Mr. Anuja Perera Research Assistant Department of Agriculture University of Peradeniya	
	Mr. Manura Wijethunga Technical Assistant Fruit Research and Development Institute (Horana)	

Technical Experts, Developers, Engineers	**Mr. Varman Nagarasa** Senior Software Engineer Oensys technologies	For the proposed system industrial experts are assembled as the evaluators to get feedback from an industrial perspective.
	Mr. Danushka Dolapihilla Tech-lead iTelasoft	
	Mr. Charith Wijesundara Senior Dev Ops Engineer Edge computing and building solutions	
General Researchers	**Dr. Eranga Jayawickrama** Research Scholar University of Tohoku Japan	To get an idea about the overall concept general researchers were selected as the evaluators.

Table 9-3 Selected evaluators

9.6 Evaluation Results

The following table is the summary of the evaluators based on their experience and knowledge.

9.6.1 Domain Experts

9.6.1.1 The concept

Evaluator	Feedback
Mr. Chinthaka Silva	According to his experience Banana plant diseases is a major threat in the future. As an example, if we take 100 plants 60 of them have a disease. It's hard to identify such diseases from bare eyes. This is a good opportunity for the planters. So he is in favor with the idea.

Mr. Manura Wijethunga	Currently, they don't have a system like this to analyze banana plant diseases. This system can help society to identify such diseases at early stages.
Dr. Eranga Jayawickrama	Nice idea. In Sri Lanka, it's rare to find a system like this. Day by day the dump rate is getting higher due to various diseases.
Mr. Anuja Perera	That's a great idea even inside the university there are trees similar look to these diseases.

Table 9-4 Evaluation feedback: Concept - Domain experts

9.6.1.2 Solution

Evaluator	Feedback
Mr. Chinthaka Silva	It's hard to identify such diseases from bare eyes. This is a good opportunity for the planters and society.
Mr. Manura Wijethunga	To analyze and to confirm from the naked eyes takes a long time. So, developing an application like this will save time and energy.
Dr. Eranga Jayawickrama	Focus to develop the system with a user-friendly interface. Make it possible to do changes after developing.
Mr. Anuja Perera	Naked eye observation is not accurate.

Table 9-5Evaluation feedback: Solution - Domain experts

9.6.2 Technical Experts

9.6.2.1 Scope

Evaluator	Feedback
Mr. Varman Nagarasa	This is a good project to do as a final year student. Image processing is a good area and the hardest part is to find a proper dataset. The project scope should expand more.
Mr. Danushka Dolapihilla	As a tech lead his opinion is to expand the dataset.

Mr. Charith Wijesundara	He has used image processing before. This system will be useful for the farmers. This may bit challenging because of the different image backgrounds
Dr. Eranga Jayawickrama	Haven't done image processing before. The project scope is clear and it's very important.

Table 9-6 Evaluation feedback: Scope - Technical experts

9.6.2.2 Architecture of the Solution

Evaluator	Feedback
Mr. Varman Nagarasa	It's really important to use the correct technologies. Satisfied with the high-level architecture. Especially he appreciated the model which I have chosen (residual network). Give the main priority to the training process and find the loup wholes inside the system
Mr. Danushka Dolapihilla	Data augmentation is really important for an image processing project.
Mr. Charith Wijesundara	Need more details about the system design. Try to send the uploaded image to the agricultural expert through the user account.
Dr. Eranga Jayawickrama	There are system design flows to follow. During the development consider that also.

Table 9-7 Evaluation feedback: Architecture - Technical experts

9.6.2.3 Implementation of the Solution

Evaluator	Feedback
Mr. Varman Nagarasa	Implementation has done according to the design. Want to figure out from where does the scratch development starts from? Find more images without augmenting the same image. Overall, the implementation is good.
Mr. Danushka Dolapihilla	Implementation is good but needs more details about tools and techniques

Mr. Charith Wijesundara	Almost most of the image processing techniques are used. Make the system to predict more diseases in a single process
Dr. Eranga Jayawickrama	If can add more diseases it would be great

Table 9-8 Evaluation feedback: Implementation - Technical experts

9.7 Limitations of Evaluation

Limitations are summarized in the conclusion chapter

9.8 Evaluation of Functional Requirements

FR No	Requirement title and description	Evaluation	Priority Level
FR 01	New users should register with the system to continue with the process. This has a sign-up form.	Not Implemented	Important
FR 02	Users should enter their username and password to log in to the system	Not Implemented	Important
FR 03	Need to upload banana plant leaves with diseased or non-diseased.	Implemented	Critical
FR 04	Image augmentation	Implemented	Critical
FR 05	Training the dataset with a combination of algorithms. Each image has been resized to a standard size.	Implemented	Critical
FR 06	Should classify images to different classes using classification algorithms. Each classification has an accuracy.	Implemented	Critical
FR 07	Should predict the disease or the result.	Implemented	Critical

FR 08	If the user wants to clarify more details related to that particular disease the user can contact an expert related to banana plant diseases.	Implemented	important

Table 9-9 Observation on functional requirements

9.9 Evaluation of Non-Functional Requirements

NFR No	Requirement title and description	Evaluation	Priority Level
NFR 01	**Performance:** The system should work at a good speed without crashing or any other disturbances.	Implemented	Critical
NFR 02	**Usability:** The UI should be simple and easy to understand.	Implemented	Important
NFR 03	**Accuracy:** When the user predicting the disease, it should predict with a reasonable level of accuracy.	Implemented	Critical
NFR 04	**Maintainability:** The developer should be able to maintain the application at any time.	Implemented	Important
NFR 05	**Scalability:** When the application is hosted with the server it should be able to requests data without breaking.	Implemented	Desirable

Table 9-10 Observation on non-functional requirements

9.10 Chapter Summary

This chapter started with evaluation methodology and deeply explains the concepts that the author has followed. The self-evaluation phase explains how the author evaluates his system from an undergraduate aspect. Under evaluation criteria, the used quantitative approaches like project scope, architecture, solutions were defined. Four domain experts, technical experts, and a general researcher selected, and their feedbacks were summarized.

CHAPTER 10: CONCLUSION

10.1 Chapter Overview

This chapter includes the conclusion of the project. The project aim and objectives are mentioned in this chapter. The problems that the author had to face, and the steps followed to solve those problems are summarized in this chapter. The utilization of knowledge and the learning outcome are discussed in this chapter.

10.2 Achievements of Research Aims and Objectives

10.2.1 Aim

This project aims to identify banana plant diseases by analyzing a leaf from that particular plant. The aim has been evaluated by domain experts and technical experts to check the aim is archived successfully or not.

10.2.2 Objectives

The research objectives are as follows.

Objective	Chapter	Status
Identifying the project scope, background, and activity plan.	Chapter 1	Completed
The key literature review is followed by existing systems, research gap and the data set.	Chapter 2	Completed
To gather system and user requirements to validate functional and non-functional requirements.	Chapter 4	Completed
At the implementation to build the residual model to gain high accuracy.	Chapter 7	Completed

To test and find out domain experts and technical experts to evaluate the model.	Chapter 8/9	Completed

Table 10-1Reseach objectives

10.3 Utilizing the Knowledge from the Course

All through the whole degree program, numerous modules were instructed. Even though not many of those modules were straightforwardly connected to this project. But the following modules contributed towards the completion of the venture

Module	Description
Program Principles I and II	The programming foundation was given by these two modules. The knowledge of coding principles from a beginner to advanced was taught from these.
Object Oriented Programming I and II	
Algorithms and Design	This module helped critical thinking to develop this system with several algorithms
SDGP	A second-year module, a group project taught us punctuality. Programming skills and documentation skills were improved from this module.
UI/UX	A mobile application development module which helped to learn about user behavior. For the prototype development, this module helped a lot.
Placement Year	The placement year guided the author to do the final year project with neural networks. The experience grabbed from the placement team was extremely valuable.

Table 10-2 Course modules and their importance

10.4 Use of Existing Skills

Apart from above mention modules, few skills helped the author to complete the project on time.

- Even though web designing was a module the, existing experience in HTML and CSS were helpful when developing the prototype

- Experience in drawing UML diagrams helped to analyze the design and implementation chapter
- Knowledge extract on basic python programming helped to learn neural networks faster.

10.5 Use of New Skills

During the placement year, the author learned some machine learning techniques. To learn more functions he had to invest more time on it. Learning new algorithms and techniques is valuable for future development. While doing this project the author learned many new things and some of them are

- The author learned to analyze a deep research gap and a wide problem domain related to the project.
- Gained knowledge to augment a dataset, preprocess, and the importance of a valid dataset.
- To identify functional and non-functional requirements.
- The author learned CNN models like VGG16, VGG19, Inception, Exception, Resnet50, and their different functioning algorithms.
- Have the knowledge to compare and critically evaluate different techniques and methods and to select the most appropriate.
- After completing the project, the author learned python libraries related to TensorFlow and Keras and their integrations with different API's.
- Learned how to remove image backgrounds via photoshop and OpenCV.

10.6 Achievement of Learning Outcomes

The following points are the learning outcomes gained by the author after completing the project.

Learning Outcomes	**Description**
LO1	Analyzing the most appropriate techniques and methods to implement the project. Selecting and comparing the methodologies. Identifying a deep research gap to complete the LR.
LO2	The needed requirements were gathered and identifying the functional and non-functional requirements.

LO5, LO6	For Deep Learning there are python libraries to be learned before implementing the model. Such libraries like Tensorflow, Keras, Flask, Scikit-learn were learned from online resources. Got an idea about the Anaconda the python interpreter.
LO4	Learned how to conduct an interview and meetings with the domain experts and technical experts and helped the author to develop his communication skills.
LO7	To manage the inquire about project and model development within the given time outline, due dates, and changing requirements.
LO8	Self-learning and problem-solving were one of an imperative results of the project. The author was able to analyze the requirements on his own to develop the system

Table 10-3 Learning outcomes

10.7 Problems and Challenges Faced

Problem	**Description of the methods to overcome the problem**
Lack of experience related to the problem domain	The experience is very low related to the selected project and it was challenging to the author. The solution was to read more research papers related to the banana plant disease identification systems. The author had to visit Agriculture Department to collect more resources related to the project.
Lack of data	For image processing, the network needs a huge amount of images to gain good accuracy. The author couldn't find a dataset from online resources and the solution was to download images from the internet and to create the dataset manually. With the COVID-19 pandemic situation in the country, there was a blockage in creating the dataset.
Overfittings	When using the images without background there was an overfitting problem. The solution was to augment images with different backgrounds and to use them for training.

Hardware and Software troubleshoot	When installing TensorFlow, it requires CUDA and CUDANN to run the system. With the installed TensorFlow version the GPU didn't work due to version conflictions. The solution was to find the correct CUDA version.
Limited Time	Throughout the project, this was the most challenging task, to finish the project before the deadline. Hence the author was able to finish the implementation on time the extra days were given to complete the thesis.
Training Time	To train the dataset with 16 convolutional layers, the recommended hardware requirement is high. Somehow the author was able to train with the existing GPU.

Table 10-4 Problems and challenges faced

10.8 Limitations of the Research

The core functionalities of the system were mentioned in the previous chapter. While designing the system the author has identified some limitations.

1. Dataset

The dataset is not sufficient to train a network with 16 layers. Although the author was able to find an average amount of images. There are images with different backgrounds. Due to lack of data with different backgrounds, there can be minor miss predictions.

2. Contacting an expert through the system

The application allows users to contact an expert via 3rd party application. There's no option to contact from the internal system.

3. A limited number of evaluators

After the implementation, the evaluation is a must. For the evaluation, there were a limited number of domain experts and technical experts to evaluate the system. Author has sent twenty survey emails but only a few were responded.

An overall, there are few more banana plant diseases to be identified. Due to the unavailability of data the author of this project couldn't identify those.

10.9 Future Enhancements

Even though the banana plant disease detection system was able to archive its goals, there are several components to be implemented under future enhancements. When Comparing existing developed systems this system performs equally or better than them. Identifying limitations and working on them is a good move in future development.

Enhancement	Description	Priority
Improving the system to identify more diseases	There are few more diseases to add to the system they are “Panama wilt”, “anthracnose”, “Banana freckle”.	Critical
Improving image data set	To improve the accuracy and to avoid over fittings and under fittings	Critical
Multi-select analysis	Giving the option to the user to compare diseases at the same time with the multi-select upload function.	Important
Contact an expert option	Allowing the users to contact an expert through the application.	Important
Identifying the disease current stage	After analyzing the disease the next step is to identify the stage of that particular disease.	Important
Generate the history	Maintaining a database to store uploaded images where the users can send it directly to the experts.	Important
Making it possible to purchase fertilizers and chemicals.	After predicting the diseases the system will automatically recommend the solution including the treatments. Users can purchase items through the system.	Not important

Table 10-5 Future enhancements

10.10 Achievement of the Contribution to the body of Knowledge

The system is developed to predict banana plant diseases using image processing techniques. When considering existing systems, most of the systems are developed outside of Sri Lanka. So this would be an opportunity for the Sri Lankan community. In Sri Lanka, the author couldn't find a dataset for the project. As a contribution, the author had to create a dataset for banana plant diseases. To create he had to go to several places in Sri Lanka to capture images. At the current development stage, the system will identify the disease and the next step is to display the stage of that particular disease. This system can be use in the Department of Agriculture to guide planters.

10.11 Concluding Remarks

This research is based on banana plant diseases and the author of this project has achieved his requirement finally. During the time of evaluation, the evaluators were given positive feedbacks with some recommendations. This system has developed targeting the banana planters and public society. There are many challenges and obstacles during requirement gathering and the time of implementation. This system can identify five diseases and have the ability to analyze banana plant leaves in different backgrounds. This system can use as a solution for the planters in modern society. Throughout the project, the author was able to learn new concepts, libraries, tools in deep learning. As the final outcome, the author has proven that this system can identify diseases rather than traditional methods.

References

Albawi, S., Mohammed, T.A., Al-Zawi, S., 2017. Understanding of a convolutional neural network, in: 2017 International Conference on Engineering and Technology (ICET). Presented at the 2017 International Conference on Engineering and Technology (ICET), IEEE, Antalya, pp. 1–6. https://doi.org/10.1109/ICEngTechnol.2017.8308186

Amara, J., Bouaziz, B., Algergawy, A., 2016. A Deep Learning-based Approach for Banana Leaf Diseases Classification 10.

Ani Brown Mary N, Robert Singh A., Athisayamani, S., 2020. Banana leaf diseased image classification using novel HEAP auto encoder (HAE) deep learning. Multimed. Tools Appl. 79, 30601–30613. https://doi.org/10.1007/s11042-020-09521-1

Aruraj, A., Alex, A., Subathra, M.S.P., Sairamya, N.J., George, S.T., Ewards, S.E.V., 2019b. Detection and Classification of Diseases of Banana Plant Using Local Binary Pattern and Support Vector Machine, in: 2019 2nd International Conference on Signal Processing and Communication (ICSPC). Presented at the 2019 2nd International Conference on Signal Processing and Communication (ICSPC), IEEE, Coimbatore, India, pp. 231–235. https://doi.org/10.1109/ICSPC46172.2019.8976582

Brownlee, J., 2020. Supervised and Unsupervised Machine Learning Algorithms. URL https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/

Dimaano, F., 2019. What is Machine Learning?

Ferentinos, K.P., 2018. Deep learning models for plant disease detection and diagnosis. Comput. Electron. Agric. 145, 311–318. https://doi.org/10.1016/j.compag.2018.01.009

Gandhi, R., 2018. Support Vector Machine — Introduction to Machine Learning Algorithms.

Gomez Selvaraj, M., Vergara, A., Montenegro, F., Alonso Ruiz, H., Safari, N., Raymaekers, D., Ocimati, W., Ntamwira, J., Tits, L., Omondi, A.B., Blomme, G., 2020. Detection of banana plants and their major diseases through aerial images and machine learning methods: A case study in DR Congo and Republic of Benin. ISPRS J. Photogramm. Remote Sens. 169, 110–124. https://doi.org/10.1016/j.isprsjprs.2020.08.025

Grant, A., 2020. Info On Banana Plant Pests – Learn About Banana Plant Diseases. URL https://www.gardeningknowhow.com/edible/fruits/banana/banana-pests-and-diseases.htm

Nasser, I.M., Abu-Naser, S.S., 2019. Artificial Neural Network for Predicting Animals Category 3, 7.

Nations encyclopedia, 2020. Sri Lanka - Agriculture.

Pant, A., 2019. Introduction to Logistic Regression.

Perming, E., n.d. Paddy and banana cultivation in Sri Lanka 71.

Raut, S., Ingole, K., 2017. Review On Leaf Disease Detection Using Image Processing Techniques 04, 4.

Saranya, N., Pavithra, L., Kanthimathi, N., Ragavi, B., Sandhiyadevi, P., 2020a. Detection of Banana Leaf and Fruit Diseases Using Neural Networks, in: 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA). Presented at the 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), IEEE, Coimbatore, India, pp. 493–499. https://doi.org/10.1109/ICIRCA48905.2020.9183006

Shanika, W., Prasannath, K., 2016. Survey on Diseases of Banana and Their Management in Batticaloa District, Sri Lanka 4.

unknown, unknown, 2014. Digital Image Processing. URL https://sisu.ut.ee/imageprocessing/book/1

unknown, U., 2019. 7 Types of Neural Network Activation Functions: How to Choose? Neural Netw. Concepts.

Weerasooriya, R., 2016. Banana Nation: Your Comprehensive Guide To Sri Lankan Bananas. URL https://roar.media/english/life/food/banana-nation-comprehensive-guide-sri-lankan-bananas

Borah, P., 2018. Wonderful Benefits Of Banana: How To Include The Fruit In Your Daily Diet. Issue https://food.ndtv.com/food-drinks/benefits-of-banana-how-to-include-the-fruit-in-your-daily-diet-1216006.

Mary, A. B., 2020. Banana leaf diseased image classification using novel HEAP auto encoder (HAE) deep learning. Issue https://sci-hub.st/https://link.springer.com/article/10.1007/s11042-020-09521-1.

Ongsulee, P., 2017. Artificial Intelligence, Machine Learning and Deep Learning.

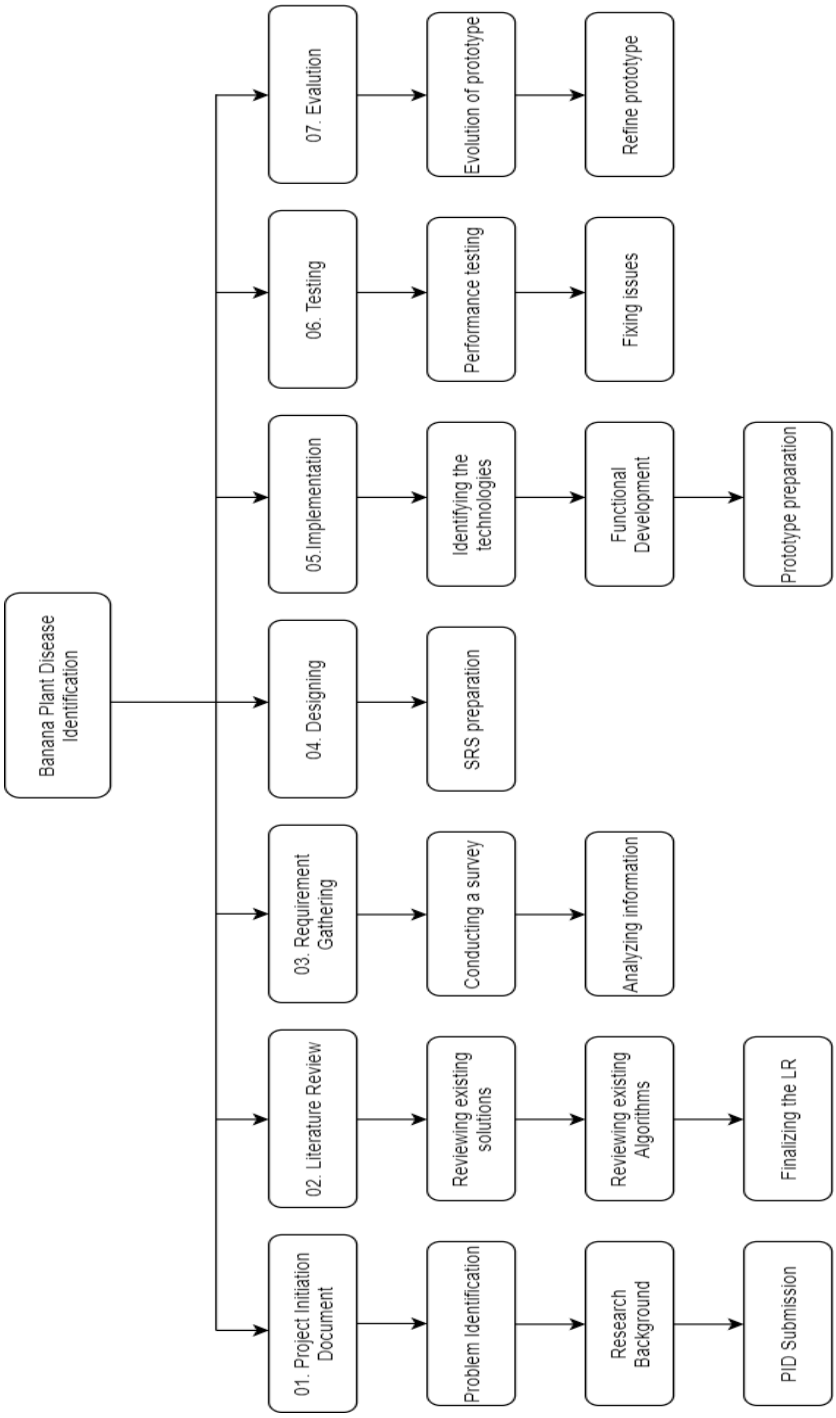
Sandhu, G. K., 2019. “PLANT DISEASES DETECTION TECHNIQUES: A REVIEW. Issue Interenational Conference on Automation, Computational and Technology Management (ICACTA).

Budiyanto, G., Ipnuwati, S., Gifari, S.A.A., Huda, M., Jalal, B., Abdul, A., Maseleno, A., Hananto, A.L., n.d. Web based expert system for diagnosing disease pest on banana plant. Int. J. Eng. 8.

Deltour, P., 2017. Disease suppressiveness to Fusarium wilt of banana in an agroforestry system: Influence of soil characteristics and plant community 9

APPENDIX A: PROJECT MANAGEMENT

Appendix A-1 Work Break Down Structure



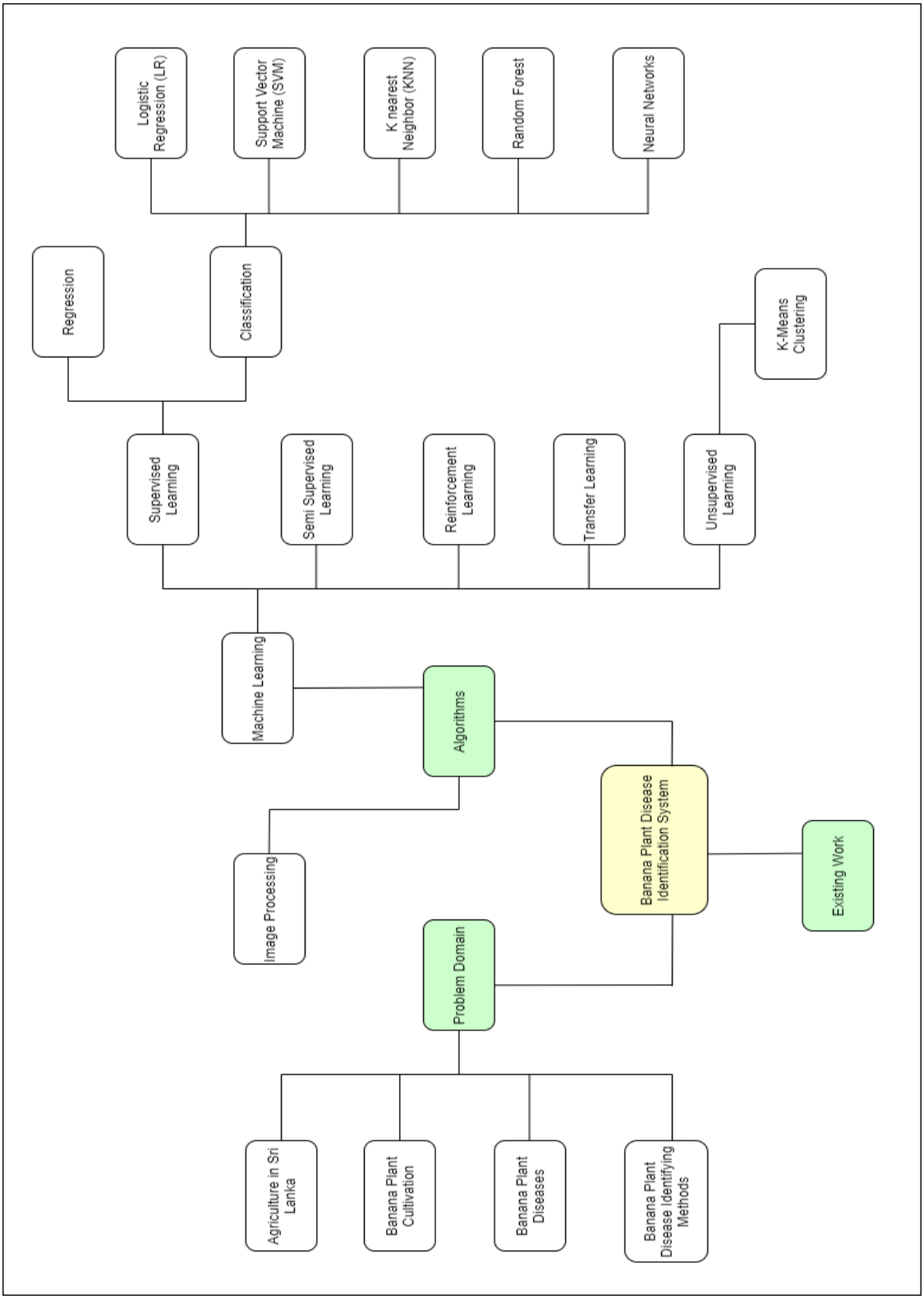
Appendix A-1 Work breakdown structure

Appendix A-2 Gantt Chart

Task	Start	Finish	Timeline	Duration
Total Product Duration	31-Aug-2020	25-May-2021		268
Project Initiation Document	31-Aug-2020	20-Oct-2020		50
Research on the Project Background	31-Aug-2020	10-Sep-2020		10
Reasearch on the Problem Domain	11-Sep-2020	23-Sep-2020		12
Research on the Exsisting Techniques and Algorithms	24-Sep-2020	1-Oct-2020		7
Specify a project scope	2-Oct-2020	7-Oct-2020		5
Identify resource requirement	8-Oct-2020	13-Oct-2020		5
Finalizing the Project Initiation Document	14-Oct-2020	18-Oct-2020		4
Submit Project Initiation Document	19-Oct-2020	20-Oct-2020		1
Liturature Review	20-Oct-2020	14-Nov-2020		25
Deep Research on the Exsisting Techniques and Algorithms	20-Oct-2020	25-Oct-2020		5
Prapare Liturature Review draft	26-Oct-2020	31-Oct-2020		5
Modifications for the Liturature Review	1-Nov-2020	12-Nov-2020		11
Submit Liturature Review	13-Nov-2020	14-Nov-2020		1
Requirement Specifications	14-Nov-2020	14-Jan-2020		61
Requirement gathering	14-Nov-2020	19-Nov-2020		5
Complete SRS document	20-Nov-2020	23-Nov-2020		3
Submit SRS document	27-Nov-2020	1-Dec-2020		4
System Design	2-Dec-2020	13-Dec-2020		11
Defining a system design specification	13-Dec-2020	11-Jan-2020		29
Submit design specification	12-Jan-2020	14-Jan-2020		2
Implementation	14-Dec-2020	8-Apr-2021		116
Build prototype	14-Dec-2020	31-Mar-2021		108
Prototype report	31-Mar-2021	8-Apr-2021		8
Testing and evaluation	9-Apr-2021	30-Apr-2021		21
System Testing	9-Apr-2021	22-Apr-2021		13
Critical evaluation of the system	23-Apr-2021	30-Apr-2021		7
Final report	1-May-2021	25-May-2021		24
Complete document	1-May-2021	23-May-2021		22
Submit document	24-May-2021	25-May-2021		1

Appendix A-2 Gantt chart

Appendix A-3 Concept Map



Appendix A- 3 Concept map

APPENDIX B: REQUIREMENT GATHERING

Appendix B-1 Survey Questioner

Final Year Project Survey: Banana Plant Disease Detection System

Hi, I'm Romesh Perera, a final year student following a B.Eng. (Hons) Software Engineering degree at the Informatics Institute of Technology(IIT), affiliated with the University of Westminster, UK.

This is a system to identify banana plant diseases in early stages using a mobile application.

I really appreciate if you could spare some of your valuable time to fill out the following questionnaire.

Please note that your responses are anonymous and will be handled confidentially. All information collected will be used for academic purposes only.

Thank you so much for your kind support and the valuable time !

Are you or any of your family members / relatives engaged in banana plant cultivation? *

- ○ Yes
- ○ No

Do you have experience in plantation?

- ○ Yes
- ○ No
- ○ In Future

Do you know that there are different types of banana plant diseases that destroy the entire cultivation ?*

- ○ Yes
- ○ No
- ○ Only Few Diseases

Have you seen any of these diseases before ?*

☐ Banana Speckle

☐ Banana Streak

☐ Bunchy Top

☐ Cordona Leafspot

☐ Banana Sigatoka

☐ No

If you know any other diseases except from the above list, please specify.

Your answer

Do you think that you can identify banana plant diseases with your exposed eye? *

◯ Yes

◯ No

◯ Maybe

Do you think that you can identify banana plant diseases with your exposed eye? *

- ◯ Yes
- ◯ No
- ◯ Maybe

Do you think the farmers have enough knowledge to identify these types of diseases ?*

- ◯ Yes
- ◯ No
- ◯ Maybe

Have you used a mobile application to identify banana plant diseases? *

- ◯ Yes
- ◯ No

How satisfied are you with this application

1	2	3	4	5
◯	◯	◯	◯	◯

Submit

Appendix B-2 Sign In Use Case

Use Case	Sign In
ID	001
Description	The user has to sign into the system
Priority	High
Actors	User
Summary	To identify diseases user needs to login to the system to get access
Pre Condition	User has to register to the system.
Post Condition	User needs to upload an image consists banana plant leaf diseased or non-diseased
Triggering Event	Submitting the login button
Main Flow	Log in to the system
Alternative Flow	1. User can enter an invalid username and password 2. The system will display an invalid message 3. Re-enter correct username and password
Exceptional Flow	1. Server response time out 2. Internal failures to login into the system

Appendix B-1 Sign In use case

Appendix B-3 Sign Up Use Case

Use Case	Sign up
ID	002
Description	The user has to sign into the system
Priority	High
Actors	User
Summary	The first step is to create an account for new users.
Pre Condition	User has install the application.
Post Condition	To access the system user needs to login to the system
Triggering Event	Submitting the register button
Main Flow	Signing up to the system
Alternative Flow	1. User can enter an invalid username and password 2. Re-enter correct username and password 3. The system will identify that the user has already registered with the system 4. Sign up success message
Exceptional Flow	1. Server response time out 2. Internal failures to login into the system

Appendix B-2 Sign up use case

APPENDIX C: IMPLEMENTATION

Appendix C-1 Image Blurring

Below figure, C-1 shows the code implementation which the author used to blur images using OpenCV.

```
import cv2
import os
import glob
import shutil

source_dir = "C:\\Users\\Romesh\\Desktop\\source"
target_dir = "C:\\Users\\Romesh\\Desktop\\target"

i = 0
for sub_dir in glob.glob(source_dir):
    for img_path in glob.glob(sub_dir + "\\*.jpg"):
        img = cv2.imread(img_path)
        cv2.imshow(' image', img)
        blurImg = cv2.blur(img, (3, 3))
        cv2.imshow('blurred image', blurImg)
        cv2.waitKey(1)
        cv2.imwrite(os.path.join(target_dir, 'blur{:>05}.jpg'.format(i)), blurImg)
        i += 1
```

Appendix C-1 Image augmentation using OpenCV

Appendix C-2 Image Blurring

Below figure, C-2 shows the code implementation which the author used to add noise to images using OpenCV.

```
import numpy as np
from skimage.util import random_noise
import cv2
import os
import glob

source_dir = "C:\\Users\\Romesh\\Desktop\\final"
target_dir = "C:\\Users\\Romesh\\Desktop\\target"

i = 0
for sub_dir in glob.glob(source_dir):
    for img_path in glob.glob(sub_dir + "\\*.jpg"):
        img = cv2.imread(img_path)
        cv2.imshow(' image', img)
        noise_img = random_noise(img, mode='s&p',amount=0.05)
        noise_img = np.array(255 * noise_img, dtype='uint8')
        cv2.imshow('blurred image', noise_img)
        cv2.waitKey(1)
        cv2.imwrite(os.path.join(target_dir, 'noise{:>05}.jpg'.format(i)), noise_img)
        i += 1
```

Appendix C-2 Adding noise to the image using OpenCV

APPENDIX D: EVALUATION

Appendix D-1 Evaluation Survey

The author has selected some domain experts and technical experts to get the feedback for the developed system. So he has sent a survey to get feedbacks from the experts.

Feedback Forum - Deep Learning Approach to Detect Banana Plant Diseases with Image Processing

Hi, I'm Romesh Perera, a final year student following a B.Eng. (Hons) Software Engineering degree at the Informatics Institute of Technology(IIT), affiliated with the University of Westminster, UK.

This is a system to identify banana plant diseases in early stages using a mobile application

This survey is created to get feedback from domain experts and technical experts to evaluate the research project.

Thank you so much for your kind support and the valuable time !
Link - https://youtu.be/d9RDT9LuZeo

* Required

Name *

Your answer

Designation *

Your answer

Concept of the project

Your answer

Scope of the Project

Your answer

System Design and Architecture

Your answer

Performance and the accuracy

Your answer

Further developments to be added

Your answer

Rate the system *

1	2	3	4	5
○	○	○	○	○

Get link