CSE 3200: System Development Project

Crop Prediction & Disease Detection System Using Machine Learning Algorithms

by

Name: Md Mossadek Touhid - Roll: 1707068
Name: Rakib Hasan - Roll: 1707014

Supervisor:
Abdul Aziz
Assistant Professor

Department of Computer Science and Engineering
Khulna University of Engineering & Technology

June 24, 2021

Approval

This Project Report has been submitted for examination with the approval of our supervisor.

Abdul Aziz
Assistant Professor
Department of Computer Science and Engineering
Khulna University of Engineering & Technology

Acknowledgement

After praising the Almighty Allah, We would like to express our sincere gratitude to our Supervisor, Assistant Professor Abdul Aziz for the continuous support, patience, motivation, enthusiasm and immense knowledge in our project. His guidance helped us in all the time during this project.
Finally, We wish to thank all the people who gave us support during this project.

Md Mossadek Touhid
&
Rakib Hasan

Abstract

Farmers usually plan the cultivation process based on their previous experiences. Due to the lack of precise knowledge about cultivation, they end up cultivating undesirable crops. To help the farmers take decisions that can make their farming more efficient and profitable, the research tries to establish an intelligent information prediction analysis on farming in Bangladesh. However, this way of farming here is still at the initial stage. The research suggests weather based beneficial crop rank before the cultivation process. To achieve these results, we are considering four major crops which are Rice, Potato, Jute, Tobacco and Maize. The prediction is based on analyzing a static set of data using Machine Learning techniques. This static dataset contains previous year's data according to the weather and Climate data. The project has an intent to use Support Vector Machine classification algorithm. In this report we will also show the detection of diseases of plants by getting their images of leaves. For this, We have an intent to use Convolutional Neural Network algorithm.And Finally we will use a Graphical User Interface (GUI) using tkinter for detecting plant diseases.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

In this era of technology, applying scientific methods and automated learning in problem solving has come a long way from being a trend to become a necessity. This realization has made the first world countries more advanced in fields like agriculture, medicine, education etc. The use of technology resulted in a booming economy and made them richer each day, whereas developing countries like Bangladesh, lags behind for turning a blind eye to it by giving the excuse of the process being too costly. Though agriculture is a crucial part of our economy, due to the lack of use of technology and scientific methods to farming, farmers in most cases do not get the preferred output. Success in agricultural growth, it is believed, is based on an ecologically adapted and economically viable agricultural technology which involves a continuous adaptation to available resources. The objective of the research is to provide a learning agent that can aid in taking decisions to make the farming more efficient and profitable through technology.

1.2 Problem Statement

In crop selection method makes use of technique where it recommends different set of crops for specific area. There are various options are available to select for farmers. They can choose one of the options and observe the results. The combination which will give high yield for a specific area is generated as output for that area according to temperature, rainfall, humidity and pH.
Some of the common symptoms of disease in plants disease are Early Blight, Late Blight, Bacterial Spot etc. These diseases can be identified by physi-

cal condition of plants leaves. The experts can identify either the plant is defected or not by looking at leaves stems or fruit. This approach requires having lots human resources for this particular job. In this era of technology and automation it is not very efficient approach, it would be much better if we have an automated system which detects disease in plants automatically. There are many researches already done to fill this purpose most of them utilize traditional machine learning approaches. The purpose of this study is to create such automated system for detecting diseases in plants by using deep learning technique.

1.3 Objective

The goal of this project is to help the farmers maximize their profit margin by providing predictions on crops that will give the maximum output. The objective of the project is to provide a learning agent that can aid in taking decisions to make the farming more efficient and profitable through technology.

- To develop a system that selects a right crop according to the given climate and soil properties data. A yield forecast system based on fewer data requirements compared to existing yield forecast solutions which demand large amounts of data.
- To design the system for predicting crop and detecting crop disease.
- To implement the designed system with a tkinter based GUI.
- To carry out the testing process of the implemented system.

1.4 Significance of the study/system

Crop prediction may be a complicated development that's influenced by soil and environmental condition input parameters. Agriculture input parameters vary from field to field and farmer to farmer. Agriculture should adapt to those climate changes, and it will do therefore by developing models which will in theory optimize management practices, maximize the rotations of the new crop to manage the changes of soil, novel breeding programs. By maximizing the worth of foretelling, the seasonal climate changes may be ascertained and recorded in an exceedingly timely manner. Later on, by victimization computer code supported machine learning, one will timely assess

the temperature change impact and check attainable situations that incorporate ascertained changes in climatic conditions and water distribution. data processing is that the process of analyzing the experimental knowledge collected over a amount and varied locations from completely different views, extract trends or patterns of data of knowledge of info and switch them into helpful information for users. Users will then additionally reason and summarize the relationships ascertained from the collected knowledge and typically predict what knowledge to expect. Machine learning techniques are a part of data processing and knowledge exploration and focus exclusively on characteristic correlations or patterns among massive datasets or massive relative databases. The patterns, associations, or relationships among all this knowledge will additional be reborn into information that's offered to the user as historical patterns and future trends. This information provided by machine learning will facilitate farmers with crop cultivation by predicting probabilities of crop losses or stop losses altogether.

In this era of technology, applying scientific methods and automated learning in problem solving has come a long way from being a trend to become a necessity. This realization has made the first world countries more advanced in fields like agriculture, medicine, education etc. The purpose of this study is to create such automated system for detecting diseases in plants by using deep learning technique. Deep learning is subset of machine learning. The advantage of deep learning over machine learning is that one does not need to worry about domain expertise as no feature engineering is required in this, unlike traditional machine learning approaches. Our system just like other previous researches utilizes images of plants leaves to detect disease in plants. Plant disease detector is computer vision based automated plant disease diagnostic system which utilizes machine learning techniques to correctly identify disease and healthy plants also the type of disease.

Chapter 2

Literature Review

An introduction to plan diseases.[1]Current and prospective methods for plant disease detection.[2]Deep learning models for plant disease detection and diagnosis. [3]How plants respond to climate change: migration rates, individualism and the consequences for plant communities.[4]Development of automated hybrid intelligent system for herbs plant classification and early herbs plant disease detection.[6]A Scalable Machine Learning System for Pre-Season Agriculture Yield Forecast.[7]Machine learning in detecting and classifying diseases of a plant leaf.[8]Agricultural Production Output Prediction Using Supervised Machine Learning Techniques.[9]A Review on Machine Learning Classification Techniques for Plant Disease Detection.[10]Machine Learning Approach for Forecasting Crop Yield Based on Climatic Parameters.[11]

Chapter 3

Methodology

3.1 Support Vector Machine or SVM

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine.

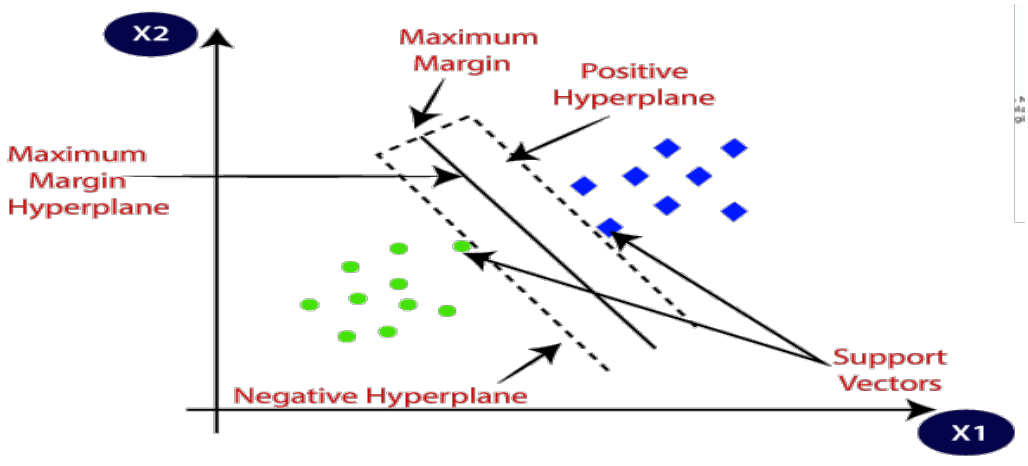

Figure 3.1: Applied Methodology of SVM

3.2 Convolution Neural Network (CNN)

Deep learning is powerful machine learning approach which have mitigated the traditional machine learning headache of feature engineering. It doesn't need any domain expertise now and all credit goes to deep learning. The core of deep learning is artificial neural network (ANN). Artificial neural networks are mathematical models that replicate with their neurons and synapses interconnecting them the general principles of brain function. To implement neural network one of the most standard library is TensorflowIt provides all libraries related to artificial neural network. With the help of Tensorflow one can perform classification tasks on text as well as images.
Convolution Neural Networks (CNNs)[5] are used to detect the disease in plant's leaves. CNN is an evolution of simple ANN that gives better result on images. Because images contains repeating patterns of particular thing (any image). Four important functions of CNN are convolution, pooling, flattening and full connection. Convolution is used to detect edges of patterns in an image and pooling is used to reduce the size of an image. Moreover training of these models are done using spyder and Keras API of Tensorflow. Keras is tensor flow's high level API for building and training deep learning models.

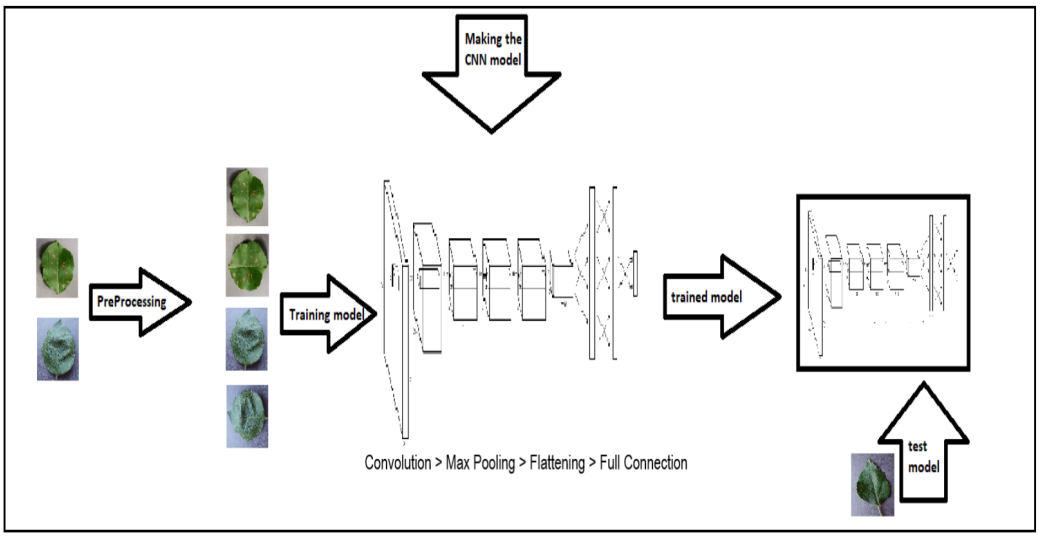

Figure 3.2: Applied Methodology of CNN

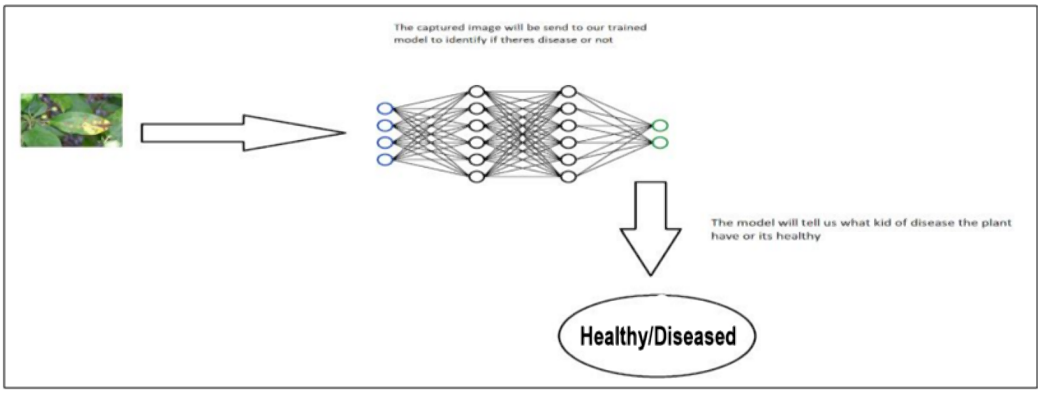

Figure 3.3: Testing of an image

Chapter 4

System Architecture

4.1 Overview of the System

Architecture Structure depicts how the system interacts and the control flows from one point in the cycle to another. There we will discuss the hardware control flow of the system right from capturing from the image to detection of the disease and displaying it. The figure below shows the hardware and software control flow of the system as a whole.

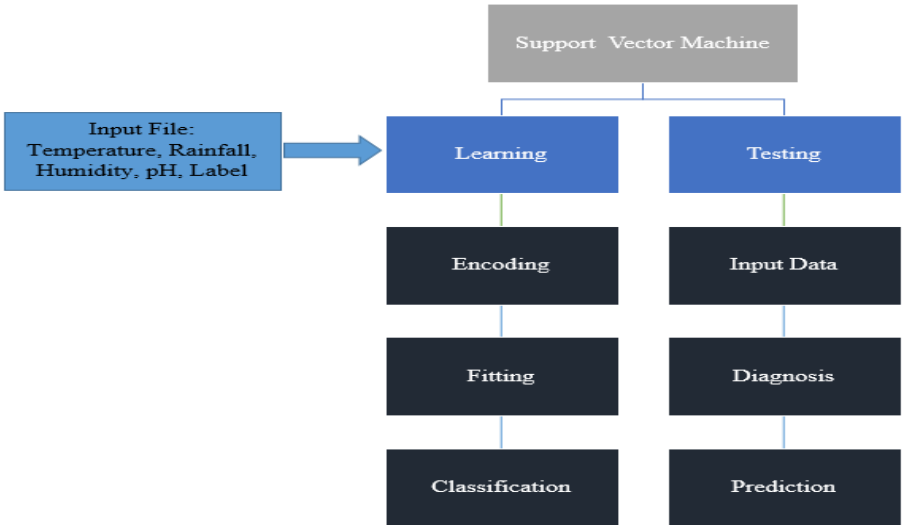

Figure 4.1: Crop Prediction System Architecture

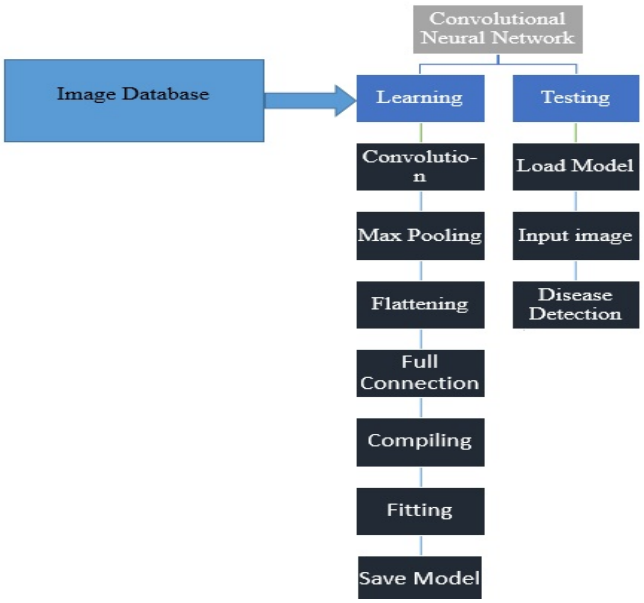

Figure 4.2: Crop Disease Detection System Architecture

4.2 Design and Development

For crop prediction,we have computer based program that takes simply humidity and rainfall and predict which crop will be good for that weather. Secondly,for disease detection,Our proposed system is an tkinter based software. We have tkinter based software which simply takes the image. Then this image is sent through a Convolutional Neural network which encodes this image into a numerical array and classifies it with the other numerical arrays in the model. The model is a tensorflow model which is made into a tensorflow lite model because of the large size of the normal tensorflow model. This model helps classify the uploaded image numerical value to the dataset values. When a numerical array matches it calculates the confidence and displays the value which has the highest confidence. In this way, we can ensure that we always have the highest confidence value showing in the results. The proposed system analysis is as follows: Pre-processing of the weather dataset: Pre-processing is a very important step in SVM as the data in the dataset may have some inconsistency which may affect the accuracy of the system. In this approach, for each category of a feature, we create a new column (sometimes called a dummy variable) with binary encoding (0 or 1) to denote whether a particular row belongs to this category. After encoding and fitting we have the dataset of weather condition . Now to testify the predicted crop we classify the data where it belongs using svm classifier.

4.3 Dataset for crop prediction

One weather dataset including temperature,rainfall,humidity and pH is used to detect what crop will be good in particular weather condition.

Temperature	**Rainfall**	**Humidity**	**pH**	**Label**
20.87974371	202.9355	82.00274	6.502985	Rice
21.77046169	226.6555	80.31964	7.038096	Rice
23.00445915	263.9642	82.32076	7.840207	Rice
24.21744605	166.1366	82.85284	7.479248	Jute
24.06475727	153.639	71.31343	6.509175	Jute
24.85441411	175.573	74.4407	6.572561	Jute
18.51816776	55.53128	90.98805	6.641906	Maize
22.53510514	67.99257	64.40866	6.48904	Maize
23.02038334	61.89472	63.03843	5.680361	Maize
23.78218473	83.63618	65.00372	5.807673	Tobacco
28.22598038	75.81132	69.43291	5.863311	Tobacco
27.77164653	78.83942	65.41395	6.05028	Tobacco

Table 4.1: Dataset Description for Crop Prediction

4.4 Pre-processing of the Images

Pre-processing is a very important step in CNN as the images in the dataset may have some inconsistency which may affect the accuracy of the system.

4.5 Dataset for disease prediction

One dataset is used to perform plant disease detection.Dataset consists of 21 classes of images of each plant. Both training and test datasets have total 46937 images. Description of these classes and dataset is given in following Table- 4.2 .

Plant Name	**Healthy or Diseased**	**Disease Name**	**Images (Number)**
Rice	Diseased	Bacterial Blight	2112
Rice	Diseased	Blast	1760
Rice	Diseased	Brown spot	2000
Rice	Diseased	Tungro	1862
Potato	Healthy	-	2280
Potato	Diseased	Early Blight	2424
Potato	Diseased	Late Blight	2424
Maize	Healthy	-	2324
Maize	Diseased	Common Rust	2384
Maize	Diseased	Cercospora /Gray Leaf spot	2052
Maize	Diseased	Northern Leaf Blight	2385
Tomato	Healthy	-	2407
Tomato	Diseased	Early Blight	2400
Tomato	Diseased	Late Blight	2314
Tomato	Diseased	Bacterial Spot	2127
Tomato	Diseased	Leaf Mold	2352
Tomato	Diseased	Septoria Leaf Spot	2181
Tomato	Diseased	Spider mites	2176
Tomato	Diseased	Target Spot	2284
Tomato	Diseased	Mosaic Virus	2238
Tomato	Diseased	Yellow Leaf Curl Virus	2451
Total			46937

Table 4.2: Dataset Description for Disease Detection

4.6 Disease Detection

Detection of disease is performed in two steps i.e. detection of the type of crop and detection of type of disease. This takes place with the help of Convolutional Neural Network .We will be using Transfer Learning for building the Model. It is a technique where the pertained models are used to create the current models. Classification also acts as fully connected classifiers which are formed using various learnings done by the model. We do the following by flattening of images which convert the pooled images to single dimension vectors. Once the images are converted to the vectors it gets quite easy to classify the images. Through the trained model we get certain numerical values with respect to various classes.

Parameter	**Value**
Epochs	25
Batch Size	32
Learning Rate	$1e^{-1}$
Activation in initial layers	Relu
Activation in final layers	Softmax

Table 4.3: CNN Training Parameters

When the leaf is healthy and there is no classification the results are shown as healthy and when there is a disease which when grey scaled shows black spots, it classifies them so they are shown as which disease they are and the confidence of the classification. Classification takes place between two numerical arrays. If the numerical arrays match, then it is a healthy or a diseased leaf, depending upon the dataset provided. Classification is a simple but relevant procedure which gives a proper result and is used in plant disease detection.

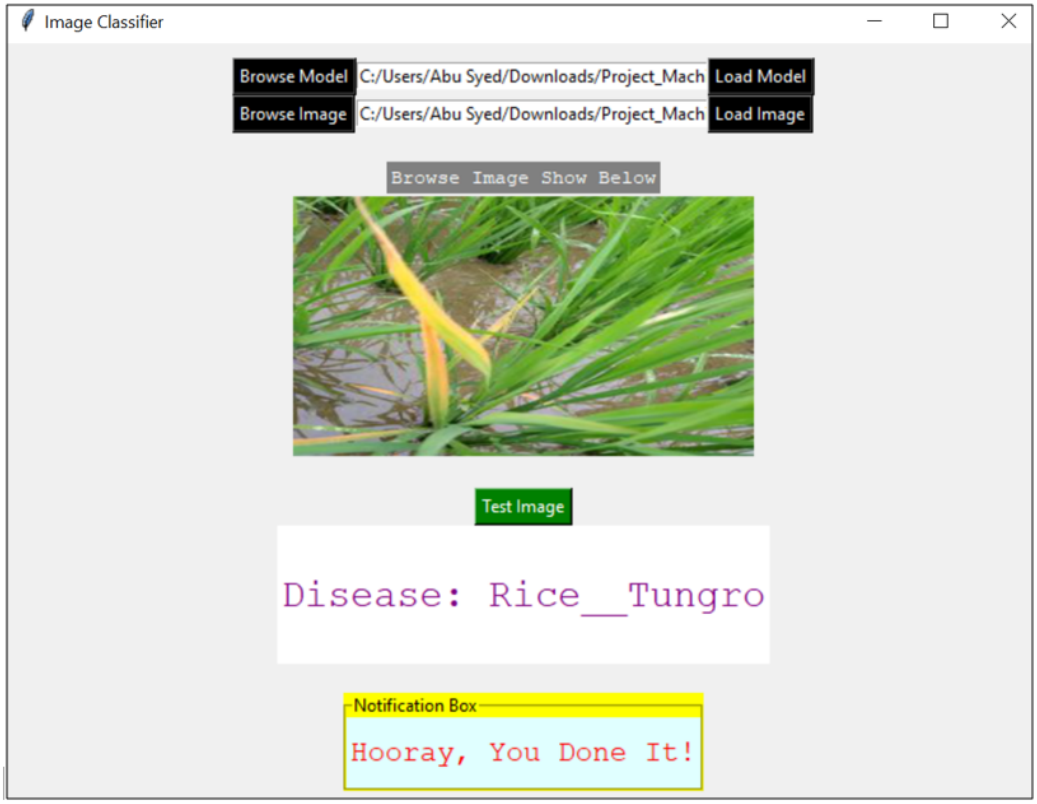

Figure 4.3: Disease of detection using GUI

Chapter 5

Result Analysis

This model were developed using Deep Learning in python. These images are from 21 different classes. Some real time images were also used. Those images were captured from local environment. They do not belong to any class which are present in dataset. But model give us more than 80% accuracy on those images as well by telling either leaf is healthy or diseased. Testing dataset gives accuracy more than 80%. Below is the Train-Test accuracy and Train-Test loss graph generated respectively by our model on testing dataset.

This study has utilized deep learning capabilities to achieve crop prediction and crop disease detection system. This system is based on a simple classification mechanism which exploits the functionalities of SVM & CNN. For detecting image, finally, the model utilizes the fully connected layers. The research was carried out using the publically accessible collection of 46937 images. The system has achieved an overall 84.18% testing accuracy on publically accessible dataset. It is concluded from accuracy that CNN is highly suitable for detection and diagnosis of plants.

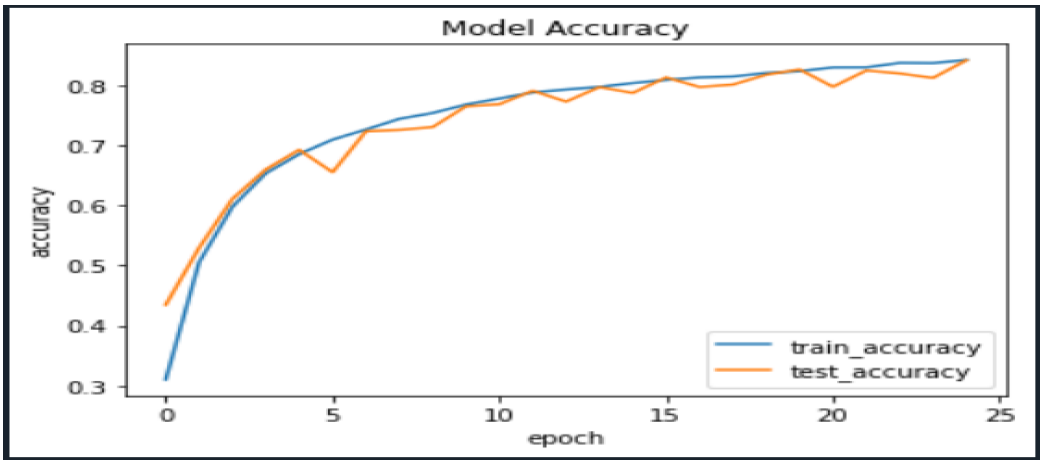

Figure 5.1: Training and Testing Accuracy on test dataset

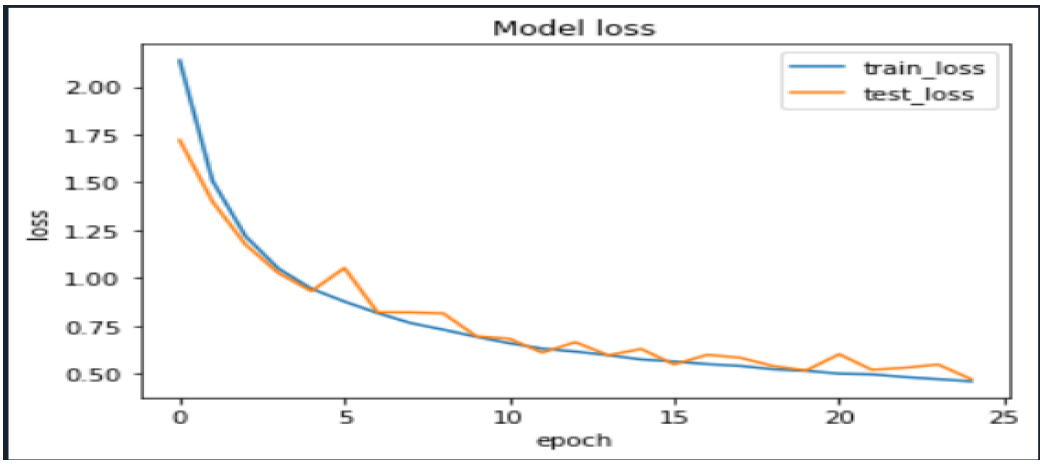

Figure 5.2: Training and Testing Loss on test dataset

Chapter 6

Limitation, Future Work & Conclusion

6.1 Limitation and Future Work

Since currently the system is trained using small dataset, the model is trained to detect only 21 types of plant diseases. We propose to train the system with much more data of various other plants and diseases to further increase the scope of the system. By adding images of many other plants, it will help in extracting many more features of the plants which certainly help in improving the accuracy of the system. The users using the system may also contribute to the system by capturing different types of plant images which can be added to the dataset. This dataset can be further used to build better models Also the may be improved in terms of accuracy by implementation of better algorithms in the coming future. We also propose to provide certain remedies for the crop diseases to the user by analyzing the diseases. This will certainly help the users to avoid such diseases in the future. Also the remedies will help the user to get rid of the diseases hence, improving their yield. This system can be integrated into mini-drones to live detection of diseases from plants in cultivated areas. Though this system is trained with only 21 classes it could tell if the plant has a disease or not as somehow symptoms are same in all kinds of plants. In addition, more actual environment images can be added to the dataset to improve the accuracy on real-condition images of leaves and classify more plant types as well as disease types. In the future, this system can also adopt 3 layer approach where the first layer detects if there's any plant in an image or not, second layer tells the plant type and the third layer tells if there is any disease or not and what type of disease is there if any.

6.2 Conclusion

The proposed methodology in the crop prediction and crop's leaf disease detection system focus on generating an advance and efficient system which makes the process of creating high yield of crop much more easier for the farmers. The project aims to predict crop and detect the most common diseases occurring on a crop's leaf using image processing technique under upbringing technology i.e., machine learning. In easier terms, the farmer will be able to accurately detect the type of disease a particular plant is having using the image of the plant.

In this study, we describe the comparison of our system with preexisting systems with proper methodology and implementation. It aims to make the life of farmers easier. The system can be a boon to the agricultural sector as it advances the crop production and management process, as agriculture is of the major reason to facilitate growth of per capita income of our country.

References

[1] Murk Chohan, Adil Khan, Rozina Chohan, Saif Hassan Katpar, and Muhammad Saleem Mahar. Plant disease detection using deep learning.

[2] Yi Fang and Ramaraja P Ramasamy. Current and prospective methods for plant disease detection. *Biosensors*, 5(3):537–561, 2015.

[3] Konstantinos P Ferentinos. Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture*, 145:311–318, 2018.

[4] Brian Huntley. How plants respond to climate change: migration rates, individualism and the consequences for plant communities. *Annals of Botany*, pages 15–22, 1991.

[5] Yann LeCun, Yoshua Bengio, et al. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10):1995, 1995.

[6] MS Mustafa, Zulkifli Husin, WK Tan, MF Mavi, and RSM Farook. Development of automated hybrid intelligent system for herbs plant classification and early herbs plant disease detection. *Neural Computing and Applications*, 32(15):11419–11441, 2020.

[7] I. Oliveira, R. L. F. Cunha, B. Silva, and M. Netto. A scalable machine learning system for pre-season agriculture yield forecast. *2018 IEEE 14th International Conference on e-Science (e-Science)*, pages 423–430, 2018.

[8] L. Sherly Puspha Annabel, T. Annapoorani, and P. Deepalakshmi. Machine learning for plant leaf disease detection and classification – a review. In *2019 International Conference on Communication and Signal Processing (ICCSP)*, pages 0538–0542, 2019.

[9] Md. Tahmid Shakoor, Karishma Rahman, Sumaiya Nasrin Rayta, and Amitabha Chakrabarty. Agricultural production output prediction using supervised machine learning techniques. In *2017 1st International Conference on Next Generation Computing Applications (NextComp)*, pages 182–187, 2017.

[10] U Shruthi, V Nagaveni, and BK Raghavendra. A review on machine learning classification techniques for plant disease detection. In *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*, pages 281–284. IEEE, 2019.

[11] S. Veenadhari, Bharat Misra, and CD Singh. Machine learning approach for forecasting crop yield based on climatic parameters. In *2014 International Conference on Computer Communication and Informatics*, pages 1–5, 2014.