A Project Report

on

Crop Disease Detection and Diagnosis

by

1. Aayush Singh

2. Deepanshu Sonparote

3. Nitish Talekar

under the guidance of

Prof. Dilip S. Kale

MANJARA CHARITABLE TRUST

RAJIV GANDHI INSTITUTE OF TECHNOLOGY, MUMBAI

Department of Computer Engineering

University of Mumbai

October - 2019

MANJARA CHARITABLE TRUST

RAJIV GANDHI INSTITUTE OF TECHNOLOGY, MUMBAI

Juhu-Versova Link Road Versova, Andheri(W), Mumbai-53.

Certificate

Department of Computer Engineering

This is to certify that

1. Aayush Singh
2. Deepanshu Sonparote
3. Nitish Talekar

Have satisfactory completed this synopsis entitled

Crop Disease Detection and Diagnosis

Towards the partial fulfillment of the

BACHELOR OF ENGINEERING
IN
(COMPUTER ENGINEERING)

as laid by University of Mumbai.

Guide | **Head of Department**

Prof. Dilip S. Kale | **Dr.Satish Y. Ket**

Principal

Dr.Sanjay Bokade

Internal Examiner | **External Examiner**

Declaration

We wish to state that the work embodied in this synopsis titled "Crop Disease Detetction and Diagnosis" forms our own contribution to the work carried out under the guidance of "Prof Dilip S. Kale" at the Rajiv Gandhi Institute of Technology.

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. we also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. we understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

(Students Signatures)
Aayush Singh (758)
Deepanshu Sonparote (759)
Nitish Talekar (762)

Acknowledgement

We wish to express our sincere gratitude to Dr. Sanjay U. Bokade, Principal and Dr. Satish. Y. Ket , H.O.D of Department Computer Engineering of Rajiv Gandhi Institute of Technology for providing us an opportunity to do our project work on “Crop Disease Detection and Diagnosis”.

This project bears on imprint of many peoples. We sincerely thank our project guide Prof. Dilip S. Kale for his guidance and encouragement in carrying out this synopsis work.

Finally, we would like to thank our colleagues and friends who helped us in completing project work successfully

1. Aayush Singh
2. Deepanshu Sonparote
3. Nitish Talekar

Abstract

The crop agriculture industry faces the economic losses due to the pest infections, bacterial or viral contagions, the farmers lose nearly 10-20 percent of the total profit on an average annually in India. This project proposes a solution to the agricultural problem, which involves crop disease recognition by using machine learning and deep learning techniques. In this project, the system sets out to classify cotton crop images into classes, whether the crop is infected by a disease or not. Also, we endeavour applications that give the farmer readily available means to identify the diseases on their crop and take appropriate damage control actions.

The dataset used to train the model was user created (mobile capture images with high-resolution camera) from various crop farms. Cotton crops, of different varieties, containing four classes of diseases, namely Rust, Mosaic Virus, Woolyaphids and Healthy plants are taken as classification ideals.

The trained models have provided a performance reaching a 91.05% success rate in identifying the corresponding cotton plant disease. The model used in the study delivers significant accuracies of classification on the dataset used by employing Residual Network techniques. The model is very useful advisory or early warning tool for the farmers for identification of diseases in the early stage so that immediate action can be taken.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction Description

Agriculture is the primary occupation in India. Crops are part and parcel of the agricultural industry. Nowadays, a tremendous loss in the quality and quantity of crop yield is observed, subject to various diseases in the plant. Crop Plant disease classification is a critical step, which can be useful in early detection of pest, insects, disease control, increase in productivity, among other examples. Farmers recognize disease manually with foregoing symptoms of plants, and with experts, whereas the actual diseases are hard to distinguish with naked eye, and it is time-consuming to predict whether the crop is healthy or not. Cotton is one of the major agricultural crops in India and it has a dominant impact on the overall Indian agriculture sector. Cotton plant leaf disease diagnosis is very difficult through observation to find the symptoms on plant leaves, incorporates it's a part of a high degree of complexity. Due to complexity even experienced agronomists and plant pathologists often fail to successfully diagnose specific diseases and are consequently led to mistaken conclusions and treatments.

This project proposes to help predict crop diseases in plants by processing the images of the crop. For this, Image Processing techniques are used for the very fast, accurate and appropriate classification of diseases. Symptoms of diseases in cotton predominantly come out on leaves of plants. The existence of an automated system for the detection and diagnosis of plant diseases would offer a support system to the agronomist who is performing such diagnosis through observation of the leaves of infected plants.

The existing techniques for disease detection have utilized various image processing methods followed by various classification techniques. Crop Yield Forecasting has been an area of interest for producers, agricultural-related organizations. Timely and accurate crop yield forecasts are essential for crop production. The proposed system uses an artificial neural network to classify the health of a cotton leaf plant.

1.2 Organization of Report

- **Ch.1 Introduction:** An introduction to motivation for the project, proposed plan and system and a brief description of existing system.

- **Ch.2 Review of Literature:** The survey of existing systems and current methods of dealing with the problem of crop diseases.

- **Ch.3 Proposed System:** The detailed description of proposed system using Classification techniques with help of neural networks to determine disease in crops.
- **Ch.4 Implementation Plan:** The proposed time allocation and plan of action for the development of project.
- **Ch.5 Simulation, Results & Analysis:** The resulting output system and analysis of results obtained during development.
- **Ch.6 Conclusion:** The final results and conclusion of results obtained during the development and testing of project.

Chapter 2

Literature Review

Crop Yield Forecasting has been an area of interest for producers, agricultural-related organizations. Timely and accurate crop yield forecasts are essential for crop production. The existing techniques for disease detection have utilized various image processing methods followed by various classification techniques. However, some unconventional approaches have led to classification of diseases using unconventional factors. Chopda et al. [1] propose a system which can predict the cotton crop diseases using decision tree with the help of the parameters like temperature, soil moisture, etc. based on the previous year data and through sensors. However, these data might not be fully dependable to predict or classify diseases.

Image classification and regression techniques play a very important role because it allows identifying, group, and properly of organisms from a standardized system. We apply an algorithm for image segmentation technique on data for automatic detection and classification of plant leaf diseases. In [2], Kamble defines the application of texture analysis for detecting plant diseases with the help of different image processing technique. Further with the use of Decision-Making Module, the disease is classified. Singh and Misra [8] suggest different diseases classification techniques that can be used for plant leaf disease detection and an algorithm for image segmentation, the advantage of using this method is that the plant diseases can be identified at an early stage or the initial stage.

Deep learning is a set of learning methods attempting to model data with complex architectures combining multiple non-linear transformations. The element of deep learning is the neural networks that are combined to form the deep neural networks. These techniques have enabled significant progress in the fields of image processing and image classification. Kulkarni [3] formulates an application of Deep Convolutional Neural Network to identify and classify crop disease on images, testing it on five classes of crops and three types of diseases for each class. Mique Jr, Eusebio L [4] proposed an application that will help farmers in detecting cotton insect pests and diseases using Convolutional Neural Network (CNN) and image processing. The searching and comparison of captured images to a stack of cotton pest images was implemented using a model based on CNN. Collected images were pre-processed and were used in training the model.

There exist several types of architectures for neural networks:

1. The multilayer perceptions, that are the oldest and simplest ones,

2. The Convolutional Neural Networks (CNN), particularly adapted for image processing

3. The recurrent neural networks and dense neural network used for sequential data such as text or times series.

A different approach is taken by Petrellis [6] where mobile phone application for plant disease diagnosis is presented which is based on the detection of the disease signature that is expressed as a number of rules that concern the color, the shape of the spots, historical weather data among other factors..

2.1 Survey Existing system

2.1.1 Plant Disease Detection using CNN

CNN architectures vary with the type of the problem at hand. The proposed model consisted of three convolutional layers each followed by a max-pooling layer. The final layer was fully connected MLP. ReLu activation function was applied to the output of every convolutional layer and fully connected layer. The first convolutional layer filters the input image with 32 kernels of size 3x3. After max-pooling is applied, the output is given as an input for the second convolutional layer with 64 kernels of size 4x4. The last convolutional layer has 128 kernels of size 1x1 followed by a fully connected layer of 512 neurons. The output of this layer is given to softmax function which produces a probability distribution of the four output classes. The model is trained using adaptive moment estimation (Adam) with batch size of 100 for 1000 epochs.

2.1.2 Cotton Blast Disease Recognition Using a CNN and SVM

The SVM is a powerful classifier that works well on a wide range of complex classification problems. SVM with different kernel functions can transform a nonlinear separable problem into a linear separable problem by projecting data into a higher dimensional space to maximize the classification distance and achieve the desired classification. In this study, the radial basis function (RBF), a popular kernel function of SVM, is chosen as the kernel function. The LIBSVM, as an efficient open source tool, is chosen to build SVMs in experiments. CNN works as a trainable feature extractor and SVM performs as a recognizer. This hybrid model automatically extracts features from the raw images and generates the predictions.

2.2 Limitation Existing system or research gap

- No public platform to help farmers
- Does not use actual real life plant dataset

The occurrence of cotton disease is regular, and the type and the probability of the cotton disease vary with the stages of cotton growth. Therefore, different cotton disease identification systems should and can be established using the method presented by this study, and then the automated cotton disease diagnosis can be realized by combining identification models and domain knowledge of cotton disease. Although our method of automatic identification of cotton blast has achieved satisfactory results, substantial further work is needed to improve its accuracy and reliability in cotton disease diagnosis systems. In particular, we plan to address the following two issues in future studies:

1. Expand the dataset of cotton disease, and establish a comprehensive tool for cotton disease diagnosis system. Te data augmentation method will be employed for building a good classifier when the number of samples is insufficient.

2. Study other deep neural network architectures and take full advantage of the deep learning algorithms to improve the classification accuracy, and enhance the reliability and robustness of the cotton disease diagnosis systems.

2.3 Problem Statement and objectives

A tremendous loss in crop yield has been observed globally every year, subject to various diseases which are infested upon them. Detection of plant diseases is an essential topic as it may prove beneficial in monitoring large fields of crops, and thus automatically detect the diseases and provide appropriate diagnosis as soon as they appear on plant leaves. The proposed system is a web-based solution for automatic detection and classification of plant leaf diseases. Once analyzed, the system provides methods and techniques to control and limit the spread of disease.

2.3.1 Objectives

- To compare different models of classification on basis on accuracy and efficiency and select the best fitting model for disease detection in crops.

- To create a web-platform for detection of disease in crops by use of classification.

- To provide treatment and precautionary information for the related disease.

2.4 Scope

The future work of this project is to develop a complete system consisting of server-side components containing a trained model and an application for smart mobile devices with features such as displaying recognized diseases in plants, based on leaf images captured by the mobile phone camera. This application will serve as an aid to farmers, enabling fast and efficient recognition of plant diseases and facilitating the decision-making process when it comes to the use of chemical pesticides.

Chapter 3

Proposed System

The proposed system is a web-based solution for automatic detection and classification of plant leaf diseases. Once analyzed, the system provides methods and techniques to control and limit the spread of disease.

Frontend: The system will include a front-end that will display an option to upload an image of diseased crop leaf. It will display the image as well as the detected disease of the plant. Along with the detection, it will also show symptoms, triggers and precautionary measures for the detected disease.

Backend: On the backend of the system, models of appropriate algorithms will be used. These models are trained models with sufficient accuracy of the detection of disease. The uploaded image will be uploaded to the model and model will return prediction for the same.

Classifier: Different classifiers will be used for the proposed system and most accurate will be chosen. Future scope of proposed system includes the option to select model of detection too.

3.1 Framework

The proposed system looks forward to use one of the following frameworks: CNN, DNN, SVM, Random Forest, or ResNet

3.1.1 Convolutional Neural Network

A convolutional neural network (CNN) is a specific type of artificial neural network that uses perceptrons, a machine learning unit algorithm, for supervised learning, to analyze data. CNNs apply to image processing, natural language processing and other kinds of cognitive tasks. A convolutional neural network is also known as a ConvNet. Like other kinds of artificial neural networks, a convolutional neural network has an input layer, an output layer and various hidden layers. Some of these layers are convolutional, using a mathematical model to pass on results to successive layers. This simulates some of the actions in the human visual cortex.CNNs are a fundamental example of deep learning, where a more sophisticated model pushes the evolution of artificial intelligence by offering systems that simulate different types of biological human brain activity.

3.1.2 Deep Neural Network

Each neuron in a layer receives an input from all the neurons present in the previous layer—thus, they're densely connected. In other words, the dense layer is a fully connected layer, meaning all the neurons in a layer are connected to those in the next layer. A densely connected layer provides learning features from all the combinations of the features of the previous layer, whereas a convolutional layer relies on consistent features with a small repetitive field.

3.1.3 Support Vector Machine

A support vector machine (SVM) is a machine learning algorithm that analyzes data for classification and regression analysis. SVM is a supervised learning method that looks at data and sorts it into one of two categories. An SVM outputs a map of the sorted data with the margins between the two as far apart as possible. SVMs are used in text categorization, image classification, handwriting recognition and in the sciences. A support vector machine is also known as a support vector network (SVN).A support vector machine is a supervised learning algorithm that sorts data into two categories. It is trained with a series of data already classified into two categories, building the model as it is initially trained. The task of an SVM algorithm is to determine which category a new data point belongs in. This makes SVM a kind of non-binary linear classifier.

3.1.4 Random Forest Classifier

A random forest is a data construct applied to machine learning that develops large numbers of random decision trees analyzing sets of variables. This type of algorithm helps to enhance the ways that technologies analyze complex data.n general, decision trees are popular for machine learning tasks. In a random forest, engineers construct sets of random decision trees to more carefully isolate knowledge from data mining, with different applied variable arrays. One way to describe the philosophy behind the random forest is that since the random trees have some overlap, engineers can build systems to study data redundantly with the various trees and look for trends and patterns that support a given data outcome.

3.1.5 Residual Network

A deep residual network (deep ResNet) is a type of specialized neural network that helps to handle more sophisticated deep learning tasks and models. It has received quite a bit of attention at recent IT conventions, and is being considered for helping with the training of deep networks. In deep learning networks, a residual learning framework helps to preserve good results through a network with many layers. One problem commonly cited by professionals is that with deep networks composed of many dozens of layers, accuracy can become saturated, and some degradation can occur. Some talk about a different problem called "vanishing gradient" in which the gradient fluctuations become too small to be immediately useful.

3.2 Details of Hardware and Software

3.2.1 Software Requirements

- Windows, MacOS or Linux
- Python - Tensorflow, Scipy, Pillow, Flask, Scikit-Learn

3.2.2 Hardware Requirements

- Processor - Intel Core i5+
- RAM - 8GB+
- ROM - 1GB+
- GPU - 2GB+

3.3 Design Details

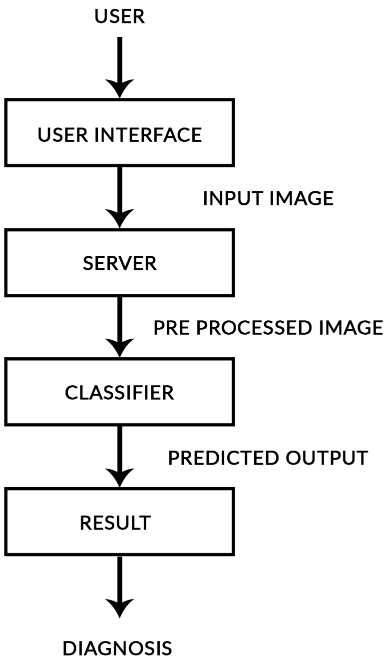


Figure 3.1: The Proposed System

- When the user inputs an image of a diseased crop, the system considers that as input data and sends it along the server
- The server preprocesses the image in a required format 200x200 and calls for a pre-trained classification model to pass the image into it. The model does the following:

1. Runs the model on the image and classify the picture into 1 of 4 categories. (Rust, Mosaic Virus, Woolly Aphids, Healthy)
2. The models predicts the most accurate result as diagnosis

- Upon successfully enrolling running the model, the diagnosis of the disease is displayed on the user interface for the user

the system uses various classification models and is mode to help farmers and crop owners to help prevent diseases in their crop.

3.3.1 Detailed Design

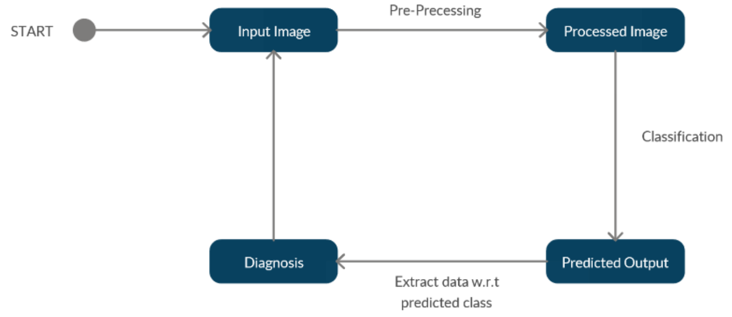


Figure 3.2: State Diagram

State diagram of Crop Disease Detection Model.

1. State 1: Initial state includes input state for system; The user selects an input image for the system and passes it on to undergo pre-processing
2. State 2: Second state receives processed data as input as submits this data to the classifier model present in the system as required.
3. State 3: The output state receives an output class from the classifier model after it is done with prediction. Output is then submitted to database to retrieve diagnostic information which is sent to diagnosis state.
4. State 4: Diagnosis state displays the diagnosis provided from the model and the database. After display, the state return to initial state to restart the process with new input.

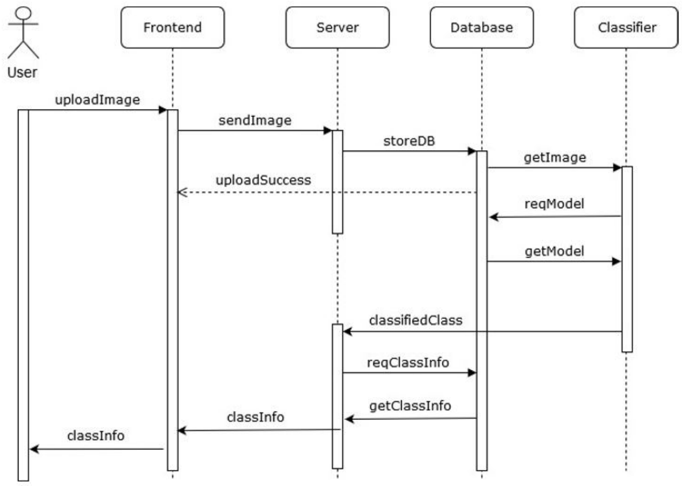


Figure 3.3: Sequence Diagram

3.4 Methodology

Input Images: For the initial process, images of high resolution (4160 x 3120) are taken from datasets as input by setting IMGW, IMGH with 3 channels (RGB), for better visibility, preferred with dimensions > 180. The features that are used for classification of the images. The images are foremost pre-processed into a 4160x 3120, RGB format with pixel values ranging from 0 to 255. The feature normalization used in the study is the min-max normalization. It is the ratio of the difference between the instance's feature value and the minimum value of a feature in the instance to the difference between the maximum and minimum values of features in the instance.

Dataset and their Labels: The dataset contains a total of 1676 images. This is constituted by 411 healthy plants images, 360 having the Mosaic virus, 431 plants being infected by Rust and 474 samples of plants having Woolly aphids. Table 3.1 shows the composition of dataset of leaf images with class name and the number of plant leaf samples.

Table 3.1: Dataset Counts

Image Class	Images Count
Healthy	411
Mosaic Virus	360
Rust	431
Wooly Aphids	474
Totals	1676

3.4.1 Procedures

1. Input images are read using OpenCV or Keras imaging library

2. Input image names are then processed with OS library and added to respective lists

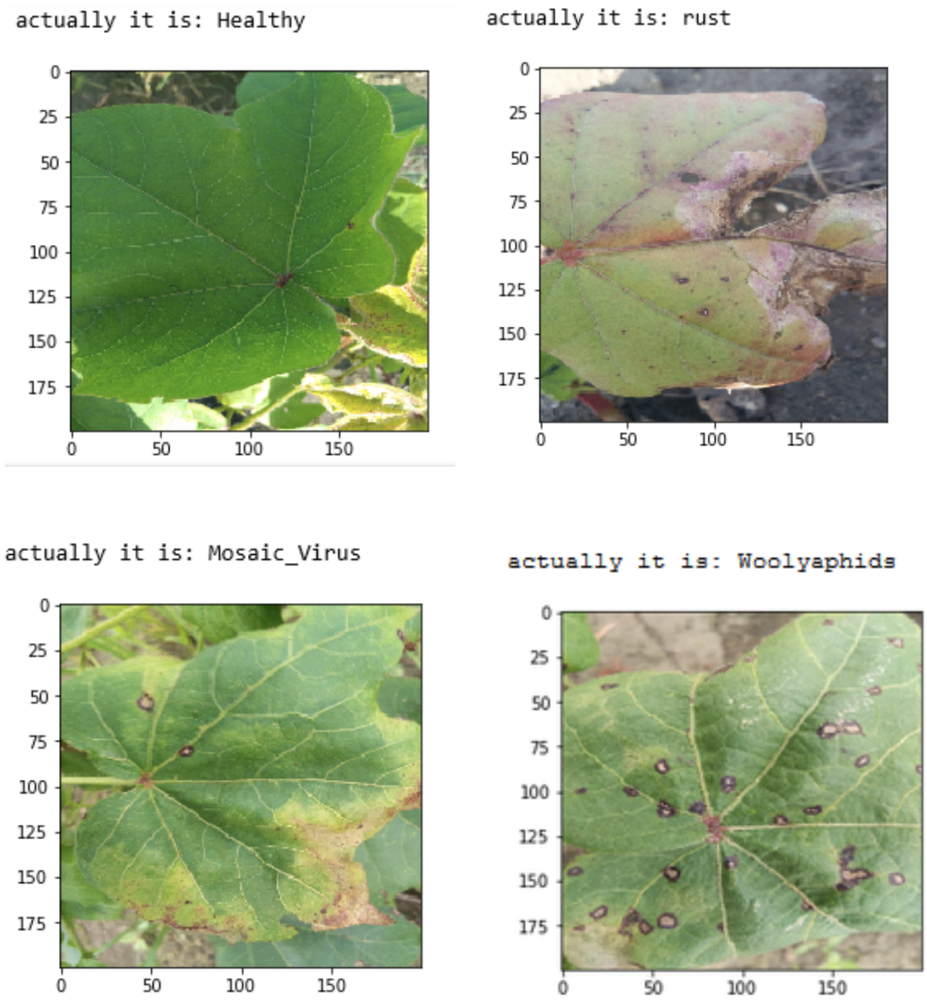


Figure 3.4: Dataset and their Labels

3. Image is split into 3 RGB channel (red, blue and green) and then merged together and then are resized for uniformity

4. Image is reshaped and divided into Training and Testing Data (70:30) ratio

5. The images are passed selected model out of a range of different classification models to a

   (a) Deep Neural Network with Leaky ReLu activation
   (b) Convolutional Neural Network with ReLu activation
   (c) Support Vector Machine Algorithm with 4 classes
   (d) Random Forest Algorithm
   (e) Residual Network framework

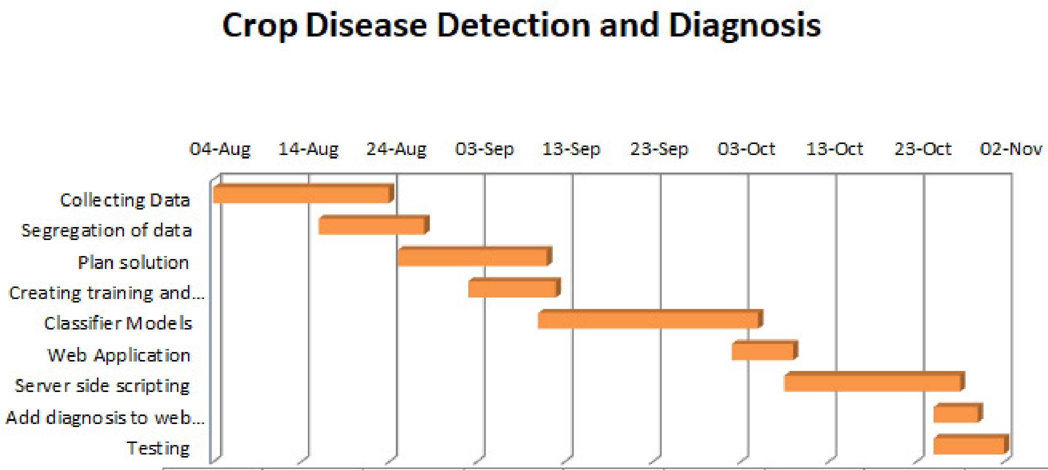
6. The Trained model is then used to test the Testing Data to get the Accuracy Percentage

Chapter 4

Implementation Plan

4.1 About Plan

1. Collection of dataset samples from various locations as well as internet and other notable and trusted sources.
2. Segregation and ordering of data by labels of diseases known via inspection by the naked eye and experience.
3. Selection of algorithms and image classification techniques to use in detection of crop diseases
4. Dividing the dataset into training and testing set and training a model based on different classification algorithms.
5. Saving model that is trained and accurate up to highest accuracy and efficiency that is possible.
6. Loading saved models onto the web platform as backend structure
7. Taking user input image of disease and passing input through model to analyse image and detect disease in crop image and return detected disease.
8. Displaying disease as well as an efficient way to deal with disease and precautionary means to take for the same
9. Deploy model along with web app to spread awareness and also to help farmers and other crop owners to care for their crops.



	Testing	Add diagnosis to web app	Server side scripting	Web Application	Classifier Models	Creating training and testing data	Plan solution	Segregation of data	Collecting Data
Start Date	25-Oct	25-Oct	08-Oct	02-Oct	10-Sep	02-Sep	25-Aug	16-Aug	04-Aug
Days To Complete	10	5	20	7	25	10	17	12	20

Figure 4.1: Gantt Chart:Implementation Plan

Chapter 5

Simulation, Results & Analysis

5.1 Simulation

The project runs on web application on localhost that will be deployed.

Web application: Takes input image from user of the diseased crop via browse and upload button on the screen. And then predicts the disease using model and displays the symptoms, triggers and precautionary measures of the predicted disease.

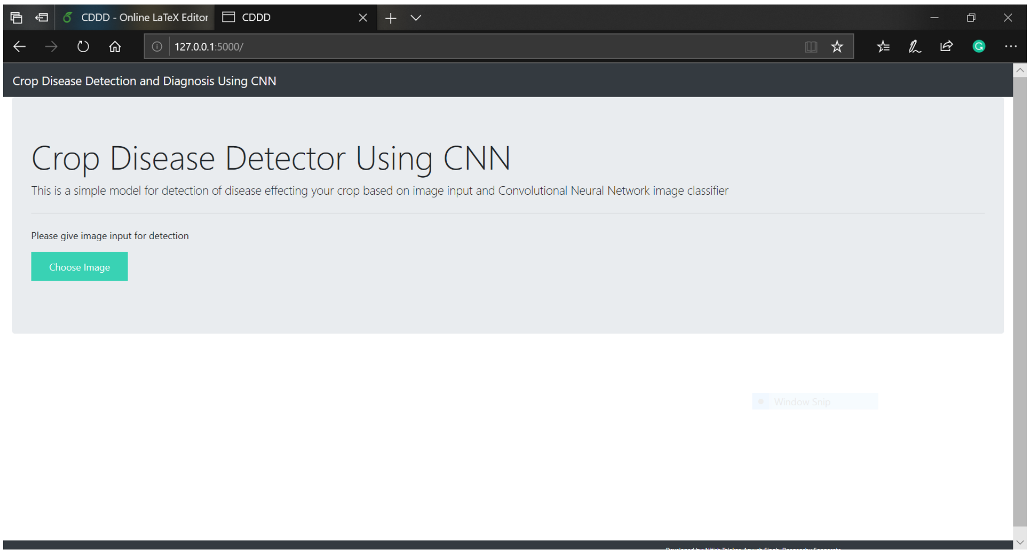


Figure 5.1: Browse button on screen

Once uploaded successfully on the server the model helps in prediction of disease with maximum prediction accuracy. The user interfaces displays the predicted disease as well as the symptoms that accompany the crop. It also displays triggering factors of the disease to help understand cause. A list of precautionary measures is also shown so as to spread awareness among the users After one image, you can detect for a new crop plant as well via the choose image link provided.

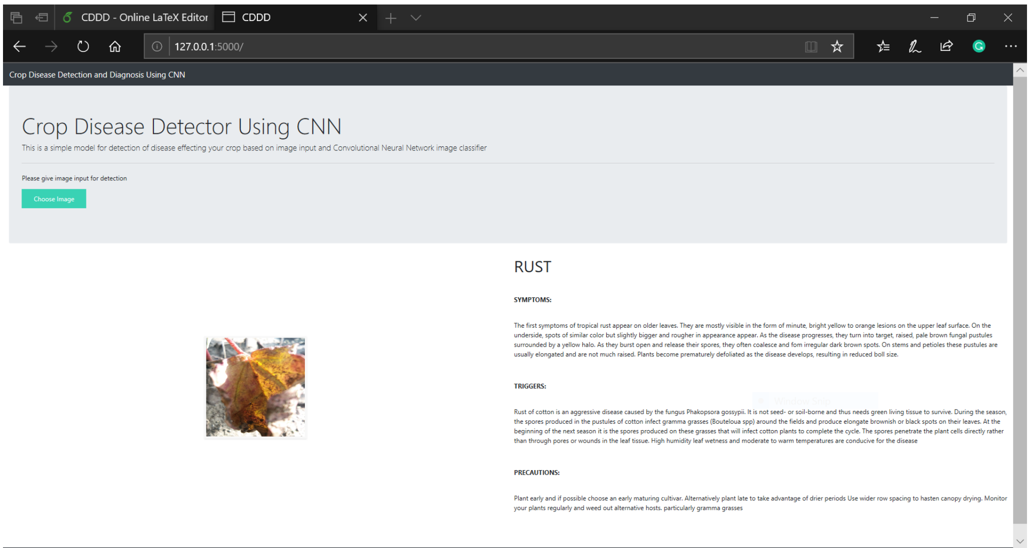


Figure 5.2: Predicted value and diagnosis

5.2 Analysis

5.2.1 Comparison

Five classification architectures and systems have been developed and tested during the development of the system. Following is comparison of maximum accuracy obtained during development and testing of each model and architecture .

Table 5.1: Accuracy Comparison

Classification Architecture	Accuracy
Random Forest Classifier	79.23%
Support Vector Machine	80.77%
Convolutional Neural Network	86.52%
Deep Neural Network	89.12%
Residual Network	91.05%

Chapter 6

Conclusions

This project is using a system dealing with plant recognition based on the leaves images and a new approach of deep learning method which automatically classifies and detects plant diseases from leaf images. With a dataset of over 1600 images, the project has created a working prediction model based on labeled data. In conclusion the project is develped using Dense Neural Network (DNN), Convolution Neural Network (CNN), Residual Network, Support Vector Machine (SVM) and Random Forest Algorithm for better classification accuracy. A web application is deployed with a complete system consisting of server-side components containing a trained model and an application for smart mobile devices with features such as displaying recognized diseases in plants. The maximum accuracy of trained model is 91.05%

Bibliography

[1] Chopda, J., Raveshiya, H., Nakum, S. and Nakrani, V., Cotton Crop Disease Detection using Decision Tree Classifier, *International Conference on Smart City and Emerging Technology (ICSCET)*, IEEE, pp. 1-5, January 2018.

[2] Kamble, J.K., Plant Disease Detector, *International Conference on Advances in Communication and Computing Technology (ICACCT)*, pp. 97-101, February 2012.

[3] Kulkarni, O., Crop Disease Detection Using Deep Learning, *Proceedings of Journal of Global Research in Computer Science* , Volume 4, No.4, April 2013.

[4] Mique Jr, E.L. and Palaoag, T.D., Rice Pest and Disease Detection Using Convolutional Neural Network, *International Conference on Information Science and System*, ACM, pp. 147-151, April 2018.

[5] Padol, P.B. and Yadav, A.A., SVM classifier based grape leaf disease detection, *Conference on advances in signal processing (CASP)*, IEEE, pp. 175-179, June 2016.

[6] Petrellis, N., Mobile Application for Plant Disease Classification Based on Symptom Signatures, *21st Pan-Hellenic Conference on Informatics*, ACM, p. 1, September 2017.

[7] Zeng, W., Li, M., Zhang, J., Chen, L., Fang, S. and Wang, J., High-Order Residual Convolutional Neural Network for Robust Crop Disease Recognition,*2nd International Conference on Computer Science and Application Engineering*, ACM, p. 101, October 2018.

[8] Singh, V. and Misra, A.K., 2015, March. Detection of unhealthy region of plant leaves using image processing and genetic algorithm, *International Conference on Advances in Computer Engineering and Applications*, IEEE, pp. 1028-1032, March 2015.

[9] Sardogan, M., Tuncer, A. and Ozen, Y., 2018, September. Plant leaf disease detection and classification based on CNN with LVQ algorithm, *3rd International Conference on Computer Science and Engineering (UBMK)*, IEEE, pp. 382-385, September 2018.