A Project Report

on

COTTON DISEASE PREDICTION USING DEEP LEARNING

Submitted in partial fulfillment of the

requirement for the award of the degree of

Bachelor of Technology in Computer Science and Engineering

GALGOTIAS UNIVERSITY

(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

Under The Supervision of
Dr. T. Poongodi
Associate Professor

Submitted By

18SCSE1130006 – TUSHAR GUPTA

18SCSE1180051 – MOHIT PANKAJ

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

GALGOTIAS UNIVERSITY, GREATER NOIDA, INDIA

DECEMBER - 2021

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
GALGOTIAS UNIVERSITY, GREATER NOIDA

CANDIDATE'S DECLARATION

I/We hereby certify that the work which is being presented in the project, entitled **" Cotton Disease Prediction Using Deep Learning "** in partial fulfillment of the requirements for the award of the **BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE AND ENGINEERING** submitted in the **School of Computing Science and Engineering** of Galgotias University, Greater Noida, is an original work carried out during the period of **JULY-2021 to DECEMBER-2021**, under the supervision of **Dr. T. Poongodi, Associate Professor**, **Department of Computer Science and Engineering** of School of Computing Science and Engineering , Galgotias University, Greater Noida

The matter presented in the project has not been submitted by me/us for the award of any other degree of this or any other places.

18SCSE1130006 – TUSHAR GUPTA
18SCSE1180051 – MOHIT PANKAJ

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Supervisor

Dr. T. Poongodi

(Associate Professor)

CERTIFICATE

The Final Thesis/Project/ Dissertation Viva-Voce examination of Tushar Gupta and Mohit Pankaj Admis No.: - 18SCSE1130006, 18SCSE1180051 has been held on 20 Dec 2021 and his/her work is recommended for the award of B. Tech Computer Science and Engineering.

Signature of Examiner(s) **Signature of Supervisor(s)**

Signature of Project Coordinator **Signature of Dean**

Date: December, 2021
Place: Greater Noida

ABSTRACT

Agriculture is one of the important professions in many countries including India. As most part of the Indian financial system is dependent on agriculture production, the keen attention to the concern of food production is necessary. The taxonomy and identification of crop infection got much importance in technical as well as economic in the Agricultural Industry. While keeping track of diseases in plants with the help of specialists can be very costly in agriculture region.

There is a need for a system which can automatically detect the diseases as it can bring revolution in monitoring large fields of crop and then plant leaves can be taken cure as soon as possible after detection of disease.

The aim of the proposed system is to develop an application which recognizes cotton leaf diseases. For availing this user need to upload the image and then with the help of image processing we can get a digitized color image of a diseased leaf and then we can proceed with applying CNN to predict cotton leaf disease. For this model, nearly 2400 specimens (600 images in each class) were accessed for training purposes.

This developed model is implemented using python version 3.7.3 and the model is equipped on the deep learning package called Keras, TensorFlow backed, and Jupyter which are used as the developmental environment. This model achieved an accuracy of 96.4% for identifying classes of leaf disease and pests in cotton plants. This revealed the feasibility of its usage in real-time applications and the potential need for IT-based solutions to support traditional or manual disease and pest’s identification.

Table of Contents

List of Figures

List of Tables

Acronyms

SVM	Support Vector Maching
ML	Machine Learning
DL	Deep Learning
CNN	Convolutional Neural Networks

CHAPTER-1
Introduction

India is an agricultural country as per the Observation in India Most of the people depend upon agriculture. Farmers have a good range of multiplicity to pick suitable crops for his or her farm. However, the cultivation of those crops for optimum yield and quality produce is usually technical. The disease diagnosis is restricted by human visual capabilities because most of the primary symptoms are microscopic. This process is tedious, time-consuming. Nowadays within the area of research, a major concern is an identification of the symptoms of the disease by means of image processing. The farmers are struggling during their lifestyle for a way to affect the disease of the cotton leaf. There a requirement for a disease diagnosis system which will support farmers. This technique focuses on disease identification by processing acquired digital images of leaves of the plant. The many papers we refer they use different algorithms for disease detection of leaf many of use support vector machine (SVM), artificial neural network (ANN) so these all-different approaches use to detect disease. The main part or advantage of our project is we provide solution for disease and give the information which pesticide or insecticides are suitable for that disease. That can help farmers to stop disease from spreading and crops give better results when we give proper treatment of these crops. The following sample images from our dataset.

The goal of this application is to develop a system which recognizes crop diseases. In this the user has to upload an image on the system, Image processing starts with the digitized color image of the diseased leaf. Finally, by applying the CNN plant disease can be predicted.

A. Purpose of Proposed System:

1. Developing a user-friendly web-based system for farmers
2. Recognizing Cotton leaf diseases accurately from input images
3. Providing corrective and preventive measures for the detected diseases

For this project, I have created deep learning model convolutional neural network (CNN) using keras library for our project cotton disease prediction. First of all, why I am using CNN because CNNs are used for image classification and recognition because of its high accuracy. The CNN follows a hierarchical model which works on building a network, like a funnel, and finally gives out a fully-connected layer where all the neurons are connected to each other and the output is processed.

Even though agriculture is the backbone of Ethiopia, so far no advanced technologies have been explored in the development of automation in agricultural science and also there are high problems in production and quality due to different diseases and pests. In recent times, the sophisticated emerging technology has attracted many researchers in the field of detection and classification of cotton leaf diseases and pests. In Ethiopia, there are several constraints which reduce the yield and quality of the product. Particularly, identification of potential diseases or pests on Ethiopian cotton is based on traditional ways. There is a wide area of farm suitable for cotton plantation, but only limited research attention is given to cotton crop production. Traditionally, experts detect and identify such plant diseases and pests on bared eyes. Bared eye determination is considered as a loss of low-level accuracy in order to detect any diseases. On high demand, different advanced technologies were aided for structuring the systems to assist nonautomatic recognition of the plant diseases and pests to increase the accuracy for any corrective measures. With the help of advanced technologies, the plant diseases were reduced, thus increasing the productivity which helped to raise the economy via boosting the production. For that reason, the implementation of information technology-based solutions in the sector of agriculture had high level of significance for Ethiopia's development in monetary, community, and eco-friendly developments by increased cotton crops' productions.

Day by day, all over the world, agriculture land is going to be reduced because the population is increasing rapidly and lack of water resources. Disease in the plant is one of those hazards that have to be examined at this stage. In contrast, the isolation of plants from their natural environment is being happened, and they are grown in unusual conditions. Many valuable crops and plants are very vulnerable to disease. They would have a great struggle to survive in nature without human involvement. Yield loss in harvests is regularly connected with plant illness or factors, such as climate, water availability, and supplement accessibility. To improve the productivity of the crop, environmental factors or product resources such as temperature and humidity are important. An important role is played by the root exudates of the plant, which helps in improving the nutrients of the soil. Compared to their wild relations, cultivated plants are always more flexible to diseases. This is the large numbers of the same species or different kinds, having a similar gene grown together, sometimes over many kilometers of distance.

The many papers we refer they use different algorithms for disease detection of leaf many of use support vector machine (SVM), artificial neural network (ANN) so these all-different approaches use to detect disease. The main part or advantage of our project is we provide solution for disease and give the information which pesticide or insecticides are suitable for that disease. That can help farmers to stop disease from spreading and crops give better results when we give proper treatment of these crops. The following sample images from our dataset.

The goal of this application is to develop a system which

recognizes crop diseases. In this the user has to upload an image on the system, Image processing starts with the digitized color image of the diseased leaf. Finally, by applying the CNN plant disease can be predicted.

A. Purpose of Proposed System:

1. Developing a user-friendly web-based system for farmers
2. Recognizing Cotton leaf diseases accurately from input images
3. Providing corrective and preventive measures for

the detected diseases

B. Cotton leaf diseases focused:

1. Alternaria Macrospora
2. Bacterial Blight

This developed model is implemented using python version 3.7.3 and the model is equipped on the deep learning package called Keras, TensorFlow backed, and Jupyter which are used as the developmental environment. This model achieved an accuracy of 96.4% for identifying classes of leaf disease and pests in cotton plants. This revealed the feasibility of its usage in real-time applications and the potential need for IT-based solutions to support traditional or manual disease and pest’s identification.

1.2. Problem Formulation:

The cotton plant is susceptible to several disorder (biotic and abiotic constraints) attacks due to temperature fluctuation, diseases, and pests. Indeed, the whole world produced nearly 576 kg per hectare of cotton crops, where only 10% of production loss occurred due to different cotton leaf diseases. The United States of America (USA) is a major exporter of cotton in the world and it obtained 5.1 billion US dollars in 2016, but there are well-known native pests which were the reason for the distraction of cotton farms. And, India has 24 percent of cotton land of the world and got 4.6 billion of dollars in 2016, from which generally 18% of cotton crops' production was lost every year due to different diseases that attacked the cotton plants which had its impacts on losing almost nine hundred thousand of Indian rupees. Presently, in Ethiopia, nearly 12–15% of cotton crop plants are infected due to different diseases. In Ethiopia, performance evaluation of GTP-I showed that these diseases and pests are the main constraints of the world standards in cotton quality and quantity of production. This results in the downfall of the economy of both the farmer and the country.

The farmers are struggling during their lifestyle for a way to affect the disease of the cotton leaf. There a requirement for a disease diagnosis system which will support farmers. This technique focuses on disease identification by processing acquired digital images of leaves of the plant. The many papers we refer they use different algorithms for disease detection of leaf many of use support vector machine (SVM), artificial neural network (ANN) so these all-different approaches use to detect disease.

India is an agricultural country as per the Observation in India Most of the people depend upon agriculture. Farmers have a good range of multiplicity to pick suitable crops for his or her farm. However, the cultivation of those crops for optimum yield and quality produce is usually technical. The disease diagnosis is restricted by human visual capabilities because most of the primary symptoms are microscopic. This process is tedious, time-consuming. Nowadays within the area of research, a major concern is an identification of the symptoms of the disease by means of image processing. The farmers are struggling during their lifestyle for a way to affect the disease of the cotton leaf. There a requirement for a disease diagnosis system which will support farmers. This technique focuses on disease identification by processing acquired digital images of leaves of the plant. The many papers we refer they use different algorithms for disease detection of leaf many of use support vector machine (SVM), artificial neural network (ANN) so these all-different approaches use to detect disease. The main part or advantage of our project is we provide solution for disease and give the information which pesticide or insecticides are suitable for that disease. That can help farmers to stop disease from spreading and crops give better results when we give proper treatment of these crops.

Detecting these diseases with bare eyes increased the complexity of cotton crops productivity which decreased the accuracy in identification precision. Even an expert would fail to assess and diagnose the diseases with their bare eyes, and this inadequate technique leads to more wastage of cotton crops. Due to these mistaken conclusions, most of the time, certain unnecessary pesticides which badly affect healthy cotton are applied. Leaving the farm for even a short time interval without production will affect overall nation GDP.

The researchers forwarded the following research questions with consideration of the issues cited in the statement of problems:

(1) What is the suitable technique used for diagnosing cotton disease and pests?

(2) How to develop an automatic cotton disease and pests' diagnosis system?

(3) How to determine the acquisition of the model?

Deep learning incorporates image processing and data analysis as a path for more possible findings. As it has been a successful application, it has now entered the domain of agriculture. Today, several deep learning-based computer vision applications such as CNN (convolutional neural network), RNN (recurrent neural network), DBN (deep belief network), and DBM (deep Boltzmann Machine) are performing tasks with high accuracy. However, the most prominent application for this research work is CNN .

Nowadays, CNN techniques are used to detect different objects and to perform automatic drawings of instructions for analysis purposes. K-fold cross-validation strategy recently recommended dataset splitting and boosted generalization of the CNN model. Generally, the model developed at the end was from scratch rather than any transferred learning model or pertained model.

Deep learning draws an attention in order to maximize the performances to classify different tasks which help to promise the human intervention data. In this real world, the usage of deep learning shows the major interest for decoding human brain activities. The problem is faced between intertrial and inter subject variability in electro encephalography signals, an indigenous access for attention-based bidirectional long-short-term memory. Convolutional neural network was analyzed among different factors that are classified into four classes of electro encephalography motor imaginary functions. Here, the usages of bidirectional long-short-term memory with the attention model accomplished the extraction of different features from the raw electro encephalography signals. Advancement of the clinical translation of the electro encephalography motor imaginary-based brain computer interface technology is applicable for varied request, where this system supports the paralyzed patients. The unusual achievements include the maximum accuracy and time-resolved predictions.

To make an efficient and effective interface system, the human plays an important role. Graph convolutional neural networks, a novel deep learning framework, addressed the issues in order to differentiate the four-class motor imaginary intentions by mutually agreeing through the similarity of electro encephalography electrodes. To find the motor imaginary, four tasks are preferred with the prediction of highest accuracy.

1.3. Merits of Proposed System

This research study focused on developing an identification model for cotton leaf diseases and pests using deep learning technique called convolutional neural networking. Three common types of disease and pests such as bacterial blight, leaf miner, and spider mite have been affecting cotton productivity and quality. Also, the model applied made a supervised learning technique on datasets with four prime feature extraction process and 2400 datasets. The datasets are limited to four different feature descriptors. Taking into consideration the time constraints and reach of the regions that grow cotton, the research focused in the southern part of Ethiopia such as Arba Minch, Shele, and Woyto. MelkaWorer agricultural research center was also proposed as a focus area because it is responsible for cotton farms in SNNPR. Deep learning techniques were used to perform the automatic feature eradication from the different input datasets.

The aim of the proposed system is to develop an application which recognizes cotton leaf diseases. For availing this user need to upload the image and then with the help of image processing we can get a digitized color image of a diseased leaf and then we can proceed with applying CNN to predict cotton leaf disease. For this model, nearly 2400 specimens (600 images in each class) were accessed for training purposes.

However, the cultivation of those crops for optimum yield and quality produce is usually technical. The disease diagnosis is restricted by human visual capabilities because most of the primary symptoms are microscopic. This process is tedious, time-consuming. Nowadays within the area of research, a major concern is an identification of the symptoms of the disease by means of image processing. The farmers are struggling during their lifestyle for a way to affect the disease of the cotton leaf. There a requirement for a disease diagnosis system which will support farmers. This technique focuses on disease identification by processing acquired digital images of leaves of the plant. The many papers we refer they use different algorithms for disease detection of leaf many of use support vector machine (SVM), artificial neural network (ANN) so these all-different approaches use to detect disease. The main part or advantage of our project is we provide solution for disease and give the information which pesticide or insecticides are suitable for that disease. That can help farmers to stop disease from spreading and crops give better results when we give proper treatment of these crops.

In Cotton plants, diseases like yellow, brown spots early and late burnt are the few examples leaf diseases. These diseases may be contagious or non-contagious viral, non-viral, bacterial and non-bacterial diseases. To identify the disease part in the cotton leaf and group into the type of diseases. Extracting the affected zone of the leaf through the image visualization. As the naked eye detection is not efficient and an expert advice is required which is more costly. Through this system the detection of the diseases by analyzing the different symptoms of the disease is more cost efficient.

CHAPTER-2
Literature Survey

They consist of 4 main phases within the primary phase, we create a color transformation structure for the RGB leaf image, and then, we apply color space transformation for the color transformation structure. Then the image is segmented. Within the second phase, the unnecessary part (green area) within the leaf area is removed. Within the third phase, we calculate the feel features for the segmented infected object. Finally, within the fourth phase, the extracted features are skilled a pre-trained neural network. The feature vector of input images is compared with the database. The category with which minimum distance is obtained is identified, class. The classifier utilized in this is often the Euclidean distance classifier. The system is developed to detect cotton disease spots. The system effectively segments the diseased portion of the image of the leaf sample using thresholding-based region extraction (diseased spots).

In this paper they're using Crop Image, Agriculture image Processing, Image segmentation, Histogram Equalization.

Accurate recognition and disease classification help us improve growth, leading to the development of several advanced techniques to be used in agriculture. Due to rapid growth in various diseases and adequate knowledge, identification and detection have become a major challenge. By monitoring the stages continuously, one can notice the disease with the naked eye, but the only drawback would be the time-consuming and increasing cost. It shows that the image processing had substantiated itself as a viable instrument for recognizable proof and arrangement of plant diseases, while the computerized camera capacities are viewed as a superior alternative for the human eye subbed by a learning calculation. To conquer the troubles of the procedure physically, a few strategies dependent on computer technology vision are created lately to perceive the disease and recognizable proof of agribusiness and agriculture crops. The quality analysis becomes difficult in captured images as they contain noise. Another restriction is that the images are to be captured under the environment in the conditions of lightning controlled. The methods of image processing execution in genuine use might be difficult because of a portion of the issues. To reduce the impact of light, point of the camera, catching gadget, and separation between objects, powerful adjustment is required. Different issues may emerge because of the changeability of shading under characteristic conditions. They divide the diseases within the following three categories: Bacterial disease, fungal diseases, viral disease. The image processing technique is used for detecting diseases on cotton leaf early and accurately. The processing scheme consists of image acquisition through camera or web, image pre-processing includes image enhancement and image segmentation where the affected and useful areas are segmented, feature extraction and classification. Finally, the presence of diseases on the plant leaf are going to be identified.

For feature extraction, they are using the K-mean clustering algorithm method for classification and Neural-network as recognizer.

The step-by-step procedure is shown as below.

1) RGB image acquisition

2) Pre-processing of image using histogram equalization

3) Resize the image

4) K-mean Algorithm for image segmentation

5) Computing features extraction

6) Classification & Recognition using neural networks

7) Statistical analysis.

Study of diseases on the cotton leaf studied by using the image processing toolbox and also the diagnosis Using MATLAB helps to suggest necessary remedies for that disease arising on the leaf of a cotton plant. The accurate recognition for using K- Mean Clustering method the Euclidean distance is 89.56% and the execution time for K-Mean Clustering method using Euclidean distance is 436.95 second and also, thresholding is completed by a dynamical range [0,1] counting on the color intensity from leaves image.

Hence that disease detects using K-mean Clustering method using Euclidean distance and it's excellent methods of disease detection on cotton leaves.

This system consists of two parts:

1) digital image assessment and feature extraction of sample cotton leaf

2) to implement the back propagation artificial neural network in machine learning.

This process has the following five steps- Image acquisition, Image Pre-processing, Image enhancement, Image segmentation and Feature extraction. To classify the quality of cotton leaf diseases, the Artificial Neural Network tool of MATLAB is used. This quality identification is done based on the RGB and HSV components of the image. This ANN tool works on neurons, neurons are further connected to hidden layers of neurons. The ANN tool also has a back-propagation process. The prediction of the outcome is taken randomly by the neural network process. The advantage of this method is it can predict the data correctly with minimized error. This method was able to detect cotton leaf with or without defects from the image.

This system classifies the leaf image using image classification algorithm CNN. It can automatically detect and recognize diseases supported extracted features at each convolution layer. The system used an image processing technique for disease detection. The user must upload the cotton leaf image. The system can pre-process the uploaded image then apply the CNN technique. By using CNN technique system can test the image with the trained dataset and extract the features. This technique is based on the infected images of various plants. Images of the infected plants are captured by a camera and process using segmentation techniques to easily identify the infected area.

1. Image Acquisition: during this phase, the raw image is taken as input from the user and Converted into an equivalent grayscale image. Also, the image is resized into the size 128*128

2. Convolutional Layers: After the alteration of the captured image, the processed image Further passes through three different hidden layers during which feature extraction, pooling and Flattening layer is additionally performed.

3. Disease Prediction: Applying Convolutional Neural Network using Softmax layer the leaf image is predicted with a disease which is gaining absolutely the best probability of occurrence. The goal is to develop a better system which recognizes crop diseases and displays user the results as

detected disease, pesticides recommended and price of pesticides recommended, and for that user need to upload a picture then, Image processing starts with the digitized colour image of the diseased leaf.

Finally by applying the CNN disease are often predicted.

According to Shuyue, they outlined the different formats of graph convolutional neural network. It was prepared to process the uniform electro encephalography data for the purpose of predicting the four classes of motor imaginaries to relate with electro encephalography electrode. They addressed their data with the transformation of 2D to 3D perspectives. The structure was processed through these dimensional units.

A study stated that, in order to utilize the dynamic route of deep learning, they proposed short-term voltage stability. They managed the clustering algorithm to obtain short-term voltage stability to increase the reliability.

In, it is stated that deep learning technique was applied to identify the leaf diseases in different mango trees. The researchers used five different leaf diseases from various specimens of mango leafs, where they addressed nearly 1200 datasets. The CNN structure was trained with more than 600 images, where 80% are used for training and 20% are used for testing. Remaining 600 images were used to find the accuracy and to identify the mango leaf diseases which showed the feasibility of its usage in real-time applications. The classification accuracy can be further increased if more images in the dataset are provided by tuning the parameters of the CNN model.

The research study states that the mechanism for the identification and classification of rice plant datasets are used to process the CNN model. For training, nearly 500 different images with diseases were collected for processing from the rice experimental field.

In, detection of cotton leafs was addressed with image processing. Here, K-means algorithms are used to segment the datasets.

The research showed the identification of diseases in banana plants which infect their leaf. In this research study, 3700 images were used for training, but there is no balanced dataset in each class. Researchers performed different experiments, for example, the training mode by using colored and grayscale image datasets and also by using different dataset splitting techniques. They obtained the best accuracy of 98.6% in colored image and 80% and 20% training to the validation dataset.

A study stated that, in order to utilize the dynamic route of deep learning, they proposed short-term voltage stability. They managed the clustering algorithm to obtain short-term voltage stability to increase the reliability.

In, it is stated that deep learning technique was applied to identify the leaf diseases in different mango trees. The researchers used five different leaf diseases from various specimens of mango leaf's where they addressed nearly 1200 datasets. The CNN structure was trained with more than 600 images, where 80% are used for training and 20% are used for testing. Remaining 600 images were used to find the accuracy and to identify the mango leaf diseases which showed the feasibility of its usage in real-time applications. The classification accuracy can be further increased if more images in the dataset are provided by tuning the parameters of the CNN model.

In this model they're using Crop Image, Agriculture image Processing, Image segmentation, Histogram Equalization. They divide the diseases within the following three categories: Bacterial disease, fungal diseases, viral disease. The image processing technique is used for detecting diseases on cotton leaf early and accurately. The processing scheme consists of image acquisition through camera or web, image pre-processing includes image enhancement and image segmentation where the affected and useful areas are segmented, feature extraction and classification. Finally, the presence of diseases on the plant leaf are going to be identified. For feature extraction, they are using the K-mean clustering algorithm method for classification and Neural-network as recognizer.

Data Collection and Sampling Technique

The sample leaf images which the researchers have used in this research are both primary as well as secondary types of datasets. Primary data is a type of data collected fresh for the first time. In this study, the primary types were collected from July to August 2019 from Arba Minch, Shele, and Woyto cotton farms where cotton plants are widely planted and there is high infection in SNNPR, whereas secondary data collected in each class were obtained from Melaka-Worker agricultural research center founded in the Afar region and SNNPR.

For this study, the researcher has used purposive or judgmental sampling techniques, selecting three infected and a healthy sample from the population, which is non probabilistic. During data collection, 2400 images of data are captured and distributed into four equal classes such as bacterial blight, healthy, leaf miner, and spider mite used to train with balanced dataset.

a. Cotton Images Sample Digitization

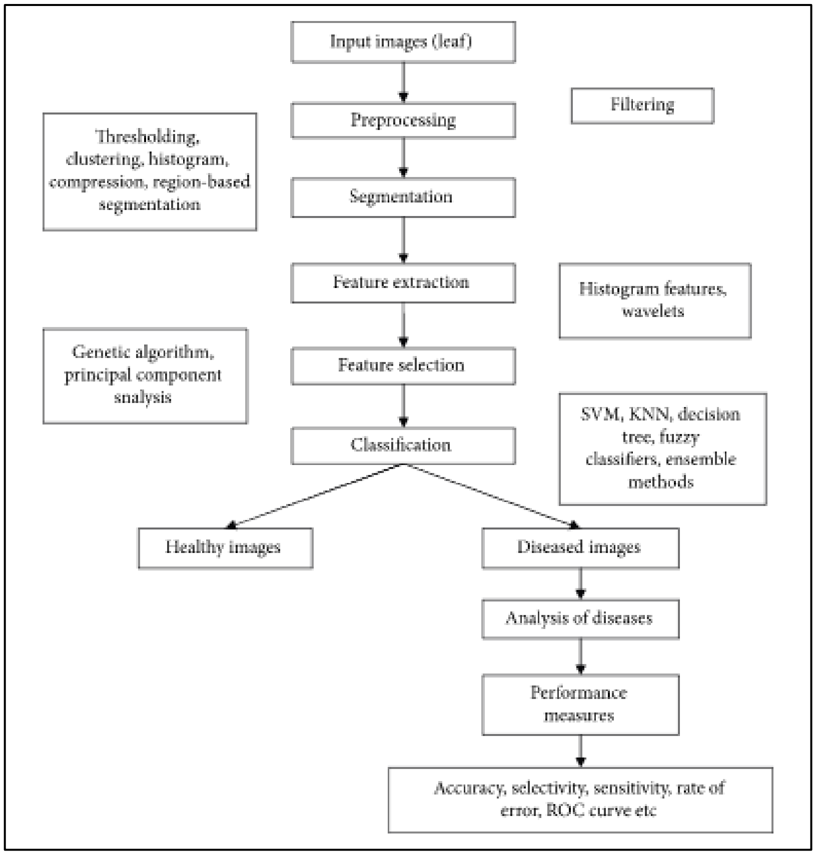
The data acquisition system in this research was used with regard to generate clear, unbiased, and simplified digital images of leaf in the cotton plant sample database for further analysis and processing. The aim was to provide the digitizing system with uniform lightning or balanced illumination. The images captured using a smartphone camera and digital camera are then transferred to a computer, displayed on a screen, and stored on the hard disk in the PNG format as digital color images.

b. Image Data Preprocessing

Inserting preprocessed images into a network is the first and basic task in all image processing projects. Common image preprocessing tasks in any image processing project are vectorization, normalization, image resizing, and image augmentation. In this research, these image preprocessing tasks are carried out before going to further deep learning processing using OpenCV library in python. Data augmentation is also used to generate more training datasets from the real sets for data samplings.

c. Feature Extraction

Deep learning solves different short comes of machine learning feature extraction such as extracting features manually by using the best and robust technique called a CNN. The layers are used to learn the knowledge. With the use of filtering mechanism, the data are used to match and extract their values.



System Architecture of Cotton leaf detection

Dataset Partitioning and Model Selection Methodology

The used dataset partitioning technique is K-fold cross-validation which is partitioned as K values, where K + 1 have to be obtained for the upcoming divisions. For this research, the study researcher has assigned the K value as 10 because it is recommended for deep learning [8, 20]. Therefore, K = 10 means 10-fold cross-validation, so dividing the total dataset into 10. D = 2400/10 = 240 data for each fold are used. From this routine activity, 80% (2160 leaf images) yield the most appropriate performance which are trained and rest 20% (240 leaf image) are used for testing; thus, the system was validated.

The dataset contains two types of diseased leaf images that are Alternaria Macrospora and Bacterial Blight and also Healthy Leaf images. The training dataset consists of total 513 images while the testing dataset consists of 207 images. The accuracy of training is 80% whereas the accuracy of testing is 89%.

Tool Selection

To collect cotton leaf images for this research, two image capturing devices were used such as a smartphone and digital camera. The proposed model was implemented using python version 3.7.3 for its usages. Also, the model is trained on the deep learning package called Keras, Version: 2.2.4-tf TensorFlow backed. TensorFlow, Version: 1.14.0 was recommended to adopt the proposed system. To evaluate the performance, many experimental setups were conducted with the help of a graphical user interface using Tkinter. From hardware, training and test was carried out on CPU instead of GPU.

Evaluation Techniques

To evaluate the routine of the structure, the researchers used various techniques in different periods, such as the developmental stage and at the end. First, the researchers evaluate the acquirements of the prototype using the confusion matrix and four evaluation metrics for confusion matrix reports such as F1-score, Precision, Recall, and Accuracy on the test dataset. Secondly, in this study for subjective evaluation, the researcher has used a questionnaire to measure the performance of a prototype by domain experts. The dataset contains two types of diseased leaf images that are Alternaria Macrospora and Bacterial Blight and also Healthy Leaf images. The training dataset consists of total 513 images while the testing dataset consists of 207 images. The accuracy of training is 80% whereas the accuracy of testing is 89%. An objective evaluation has been made using the experimental analysis to test an artifact. Finally, the result of the evaluation depicts the practical applicability of the model.

A study stated that, in order to utilize the dynamic route of deep learning, they proposed short-term voltage stability. They managed the clustering algorithm to obtain short-term voltage stability to increase the reliability.

In, it is stated that deep learning technique was applied to identify the leaf diseases in different mango trees. The researchers used five different leaf diseases from various specimens of mango leaf’s where they addressed nearly 1200 datasets. The CNN structure was trained with more than 600 images, where 80% are used for training and 20% are used for testing. Remaining 600 images were used to find the accuracy and to identify the mango leaf diseases which showed the feasibility of its usage in real-time applications. The classification accuracy can be further increased if more images in the dataset are provided by tuning the parameters of the CNN model.

In this model they're using Crop Image, Agriculture image Processing, Image segmentation, Histogram Equalization. They divide the diseases within the following three categories: Bacterial disease, fungal diseases, viral disease. The image processing technique is used for detecting diseases on cotton leaf early and accurately. The processing scheme consists of image acquisition through camera or web, image pre-processing includes image enhancement and image segmentation where the affected and useful areas are segmented, feature extraction and classification. Finally, the presence of diseases on the plant leaf are going to be identified. For feature extraction, they are using the K-mean clustering algorithm method for classification and Neural-network as recognizer.

During experimentation, different experiments were undergone to get an efficient model by customizing various parameters that provided different results. Those parameters are dataset color, number of epochs, augmentation, optimizer, and dropout. According to Serawork Wallelign [19], augmented RGB colored images provided about 15% improvement on accurate than that of not augmented.

For this new model, the researcher has trained three different numbers of epochs such as 50, 100, and 150. However, the model achieved the best performance on 100 epochs. Nitish Srivastava added a dropout in the CNN given additional performance (2.7%). Therefore, during the experiment, the researcher used 0.25 and 0.5 dropout percent in each layer and achieved the best performance in 0.5 dropout percent. Finally, a very important experiment was carried out on the regularization method that optimization algorithms' usage could minimize the loss through iterations by updating means according to a gradient. From Figure 8, it is observed that the effects of numbers on epochs and regularization methods are identified. For this research, two most recent and used optimization algorithms are used such as RMSProp and Adam, but the Adam optimization algorithm reduces loss by 2.5%

Disease of Cotton and their symptom

The natural and most frequently occurring cotton disorders are due to the loss of nutrition, environmental stress, and chemical factors that cause imbalances. These factors affect the production of cotton during the growth of the crop. There is a contrasting difference between the modification of the plant and the harmful environments, including the disorders. Verticillium wilt and cotton leaf curl diseases are major limiting factors for cotton production. Verticillium wilt is a highly hazardous disease that is strongly epidemic. The abundance of wilt pathogens is one of the many factors which results in disease occurrence.

A study stated that, in order to utilize the dynamic route of deep learning, they proposed short-term voltage stability. They managed the clustering algorithm to obtain short-term voltage stability to increase the reliability. CNN architecture consists of two broad sections such as feature learning and classification section. In general, the cotton images feed into an input layer and end with an output layer. The hidden layer consists of different layers. Here, a cotton leaf and the output will be the class name of such an image also called the label of cotton leaf diseases or pests. In general, for this proposed architecture, each cotton leaf images with addition of neurons are augmented with considerable weights. Output of the augmentation process to the upcoming layers are processed and duplicated to next layer. Output layers show the prediction tasks for calculating neurons for this research.

In, it is stated that deep learning technique was applied to identify the leaf diseases in different mango trees. The researchers used five different leaf diseases from various specimens of mango leaf's where they addressed nearly 1200 datasets. The CNN structure was trained with more than 600 images, where 80% are used for training and 20% are used for testing. Remaining 600 images were used to find the accuracy and to identify the mango leaf diseases which showed the feasibility of its usage in real-time applications. The classification accuracy can be further increased if more images in the dataset are provided by tuning the parameters of the CNN model.

In this model they're using Crop Image, Agriculture image Processing, Image segmentation, Histogram Equalization. They divide the diseases within the following three categories: Bacterial disease, fungal diseases, viral disease. The image processing technique is used for detecting diseases on cotton leaf early and accurately. The processing scheme consists of image acquisition through camera or web, image pre-processing includes image enhancement and image segmentation where the affected and useful areas are segmented, feature extraction and classification. Finally, the presence of diseases on the plant leaf are going to be identified. For feature extraction, they are using the K-mean clustering algorithm method for classification and Neural-network as recognizer. To evaluate the routine of the structure, the researchers used various techniques in different periods, such as the developmental stage and at the end. First, the researchers evaluate the acquirements of the prototype using the confusion matrix and four evaluation metrics for confusion matrix reports such as F1-score, Precision, Recall, and Accuracy on the test dataset. Secondly, in this study for subjective evaluation, the researcher has used a questionnaire to measure the performance of a prototype by domain experts, as shown in Figures 3and 4. An objective evaluation has been made using the experimental analysis to test an artifact. Finally, the result of the evaluation depicts the practical applicability of the model.

A study stated that, in order to utilize the dynamic route of deep learning, they proposed short-term voltage stability. They managed the clustering algorithm to obtain short-term voltage stability to increase the reliability.

The cotton boll is also a chief component of fiber production. Boll injury and symptoms of cotton boll with leaf-footed bugs were similar to those of plant bugs. Early detection of bollworm and pink bollworm with the help of field evolved resistance was also discussed.

(1) Angular Leaf Spot or Black Arm Disease.

Symptoms. Small spots appear under the cotyledons, which may dry and fall off. Such similar spots may also appear on the leaves. They become angular bound by leaflets and turn brown. Several small spots may combine. The infected petiole may get damaged. Elongated dark brown lesions appear on the stem as well as petioles and branches. The young stems may be enclosed and killed in the dark arm phase. Sunken black lesions are seen on the bolls. Young bolls may wither. The disease attacked stem may become weak. Discoloration of lint may appear.

(2) Vascular Wilt Disease.

Symptoms. The soils with pH values ranging from 6 to 8.00 will affect this disease. Almost all the parts of the plant get affected. The yellowing of cotyledons, sautéing of petioles, and filling of the dried leaves occur at the seedling stage. In contrast, in the young and grown-up plants, deficiency of bloat, leaf hanging, delicate shoots, caramelizing, and, lastly, the demise of the plants takes place.

(3) Grey Mildew or Dahiya Disease.

Symptoms. The organism typically assaults the more seasoned leaves making nonuniform rakish, pale, and lackluster spots. They are generally bound by the veinlets and are generally below or above the piece of the leaf. Usually, some spots which might be up to hundreds are recognized on the leaf surface. In severe infections, the leaves fall off earlier, changing their color to yellowish-brown.

(4) Anthracnose Disease.

Symptoms. In the seedling stage, minor, reddish round spots appear on the cotyledons and primary leaves. When the lesions are on the upper region, the stem may be enclosed, causing seedlings to wilt and die. In adult plants, the fungus infects the stem, causing it to separate and shred-off the bark. The symptoms are prominent on bolls as water-soaked, circular, slightly sunken reddish-brown spots which turn black. As a result of boll infection, they open prematurely. As a result, the lint becomes stained, hard, and compact.

(5) Root Rot Disease.

Symptoms. The symptoms might be visible in the seedling phase, where we can find brown spots in color on the cotyledons. At the collar area, there is dark shading, which might be extended towards the lower parts. The sinewy roots go through rotting. Rotting and shredding of the roots appear by the bark of the roots. Influenced plants can be pulled out effortlessly. The patches of infection are shown up in the field.

(6) Boll Rot Disease.

Symptoms. In the beginning, the disease’s small spots, which are brown or black in color, appear that expand to enclose the complete bolls. The contamination goes to tissues present inside, resulting in the infected stem or lint. The bolls never crack open, and they do not immediately fall off in the early stages. The rotting of the stem may be external in some cases, which causes infection to the pericarp that leaves the internal tissues freely. A large number of fungi may be observed on the infected bolls.

(7) Leaf Spot or Blight Disease.

Symptoms. The disease can occur in many stages. However, the most severe one occurs in the leaves after 45 to 60 days of planting. Tiny, pale brown, irregular round spots, measuring 0.5 to 6 mm diameter, may be present on the leaves. Each spot contains a central lesion surrounded by concentric rings. Several spots merge to form blighted areas. The affected leaves become weak and wither. Stem lesions are observed rarely. But in extreme cases, these spots may appear on bracts and bolls. During experimentation, different experiments were undergone to get an efficient model by customizing various parameters that provided different results. Those parameters are dataset color, number of epochs, augmentation, optimizer, and dropout. According to Serawork Wallelign, augmented RGB colored images provided about 15% improvement on accurate than that of not augmented.

For this new model, the researcher has trained three different numbers of epochs such as 50, 100, and 150. However, the model achieved the best performance on 100 epochs. Nitish Srivastava added a dropout in the CNN given additional performance (2.7%). Therefore, during the experiment, the researcher used 0.25 and 0.5 dropout percent in each layer and achieved the best performance in 0.5 dropout percent. Finally, a very important experiment was carried out on the regularization method that optimization algorithms' usage could minimize the loss through iterations by updating means according to a gradient. From Figure 8, it is observed that the effects of numbers on epochs and regularization methods are identified. For this research, two most recent and used optimization algorithms are used such as RMSProp and Adam, but the Adam optimization algorithm reduces loss by 2.5%

The natural and most frequently occurring cotton disorders are due to the loss of nutrition, environmental stress, and chemical factors that cause imbalances. These factors affect the production of cotton during the growth of the crop. CNN architecture consists of two broad sections such as feature learning and classification section. In general, the cotton images feed into an input layer and end with an output layer. The hidden layer consists of different layers, as shown in Figure 6. Here, a cotton leaf and the output will be the class name of such an image also called the label of cotton leaf diseases or pests. In general, for this proposed architecture, each cotton leaf images with addition of neurons are augmented with considerable weights. Output of the augmentation process to the upcoming layers are processed and duplicated to next layer. Output layers show the prediction tasks for calculating neurons for this research. There is a contrasting difference between the modification of the plant and the harmful environments, including the disorders. Verticillium wilt and cotton leaf curl diseases are major limiting factors for cotton production. Verticillium wilt is a highly hazardous disease that is strongly epidemic. The abundance of wilt pathogens is one of the many factors which results in disease occurrence.

A study stated that, in order to utilize the dynamic route of deep learning, they proposed short-term voltage stability. They managed the clustering algorithm to obtain short-term voltage stability to increase the reliability. CNN architecture consists of two broad sections such as feature learning and classification section. In general, the cotton images feed into an input layer and end with an output layer. The hidden layer consists of different layers. Here, a cotton leaf and the output will be the class name of such an image also called the label of cotton leaf diseases or pests. In general, for this proposed architecture, each cotton leaf images with addition of neurons are augmented with considerable weights. Output of the augmentation process to the upcoming layers are processed and duplicated to next layer. Output layers show the prediction tasks for calculating neurons for this research.

In, it is stated that deep learning technique was applied to identify the leaf diseases in different mango trees. The researchers used five different leaf diseases from various specimens of mango leaf's where they addressed nearly 1200 datasets. The CNN structure was trained with more than 600 images, where 80% are used for training and 20% are used for testing. Remaining 600 images were used to find the accuracy and to identify the mango leaf diseases which showed the feasibility of its usage in real-time applications. The classification accuracy can be further increased if more images in the dataset are provided by tuning the parameters of the CNN model.

In this model they're using Crop Image, Agriculture image Processing, Image segmentation, Histogram Equalization. They divide the diseases within the following three categories: Bacterial disease, fungal diseases, viral disease. The image processing technique is used for detecting diseases on cotton leaf early and accurately. The processing scheme consists of image acquisition through camera or web, image pre-processing includes image enhancement and image segmentation where the affected and useful areas are segmented, feature extraction and classification. Finally, the presence of diseases on the plant leaf are going to be identified. For feature extraction, they are using the K-mean clustering algorithm method for classification and Neural-network as recognizer. To evaluate the routine of the structure, the researchers used various techniques in different periods, such as the developmental stage and at the end. First, the researchers evaluate the acquirements of the prototype using the confusion matrix and four evaluation metrics for confusion matrix reports such as F1-score, Precision, Recall, and Accuracy on the test dataset. Secondly, in this study for subjective evaluation, the researcher has used a questionnaire to measure the performance of a prototype by domain experts, as shown in Figures 3and 4. An objective evaluation has been made using the experimental analysis to test an artifact. Finally, the result of the evaluation depicts the practical applicability of the model.

Eclipse 3.3 Indigo IDE is used for overall development as a standard platform. JSP is used to design GUI of the Web Application. The database is basically used for storing user details like usernames, disease details, pesticide information and pesticide costs. We used the MYSQL database as the data to be stored was mostly structured. The tool used for DB functionalities was MYSQL GUI Browser.

By comparing different classification algorithm we come to know that CNN is a feed-forward neural network that is used to analyze visual images by processing data. CNN can give a higher accuracy in the prediction on plant disease due to different advantages. We detect whether the plant is having some disease or not and what pesticides are needed for recovery from that disease with the help of CNN. The datasets are divided into Training and testing datasets in the ratio 70:30 and cross-validation is applied to achieve better accuracy. CNN architecture consists of two broad sections such as feature learning and classification section. In general, the cotton images feed into an input layer and end with an output layer. The hidden layer consists of different layers, as shown in Figure 6. Here, a cotton leaf and the output will be the class name of such an image also called the label of cotton leaf diseases or pests. In general, for this proposed architecture, each cotton leaf images with addition of neurons are augmented with considerable weights. Output of the augmentation process to the upcoming layers are processed and duplicated to next layer. Output layers show the prediction tasks for calculating neurons for this research.

For building CNN Python 3.6 is used along with libraries such as OpenCV for image processing and TensorFlow for Machine learning. Using libraries allows us to extend the functionalities for using it on larger database. CNN consists of three Hidden layers i.e. Convolution layers; each convolution layer consists of Pooling and Relu layers. Relu is the activation function used to remove the negative weights. MaxPooling is used in the pooling layers. 64 different filters are applied in each layer for extracting features. Convolving window size is 3*3 and stride size is 4. Softmax layer at the end linearize the probabilities of the diseases in a single dimensional matrix to choose maximum probability. Xampp server is used to deploy the Web Application on local computer. Server handles the request and response to system user.

In, it is stated that deep learning technique was applied to identify the leaf diseases in different mango trees. The researchers used five different leaf diseases from various specimens of mango leaf's where they addressed nearly 1200 datasets. The CNN structure was trained with more than 600 images, where 80% are used for training and 20% are used for testing. Remaining 600 images were used to find the accuracy and to identify the mango leaf diseases which showed the feasibility of its usage in real-time applications. The classification accuracy can be further increased if more images in the dataset are provided by tuning the parameters of the CNN model.

A study stated that, in order to utilize the dynamic route of deep learning, they proposed short-term voltage stability. They managed the clustering algorithm to obtain short-term voltage stability to increase the reliability.

Prototype Development and Evaluation

For the prototype, the researchers focused on the convention of the digital forensic investigation process, which is ISO and IEC to evaluate the prototype in terms of efficiency, effectiveness, fault tolerance, helpfulness, learn ability, and the control to assess the quality of the prototype. For the time being, the system prototypical test is carried out as a desktop application which is conducted with the help of Tkinter, a graphical user interface in Python programming language.

For the prototype, researchers focused on the convention of the digital forensic investigation process, which is ISO and IEC, to evaluate the prototype in terms of efficiency, effectiveness, fault tolerance, helpfulness, learn ability, and to control the assess quality of the prototype. For questioners, evaluators were allowed to rate the options as extremely satisfied, very satisfied, somewhat satisfied, not so satisfied, and not at all satisfied for five closed-ended questions and one open-ended question.

The CNN structure was trained with more than 600 images, where 80% are used for training and 20% are used for testing. Remaining 600 images were used to find the accuracy and to identify the mango leaf diseases which showed the feasibility of its usage in real-time applications. The classification accuracy can be further increased if more images in the dataset are provided by tuning the parameters of the CNN model.

The questionnaires are distributed to Ethiopian cotton farm experts, as shown in Figure 11. The data obtained from the farmers are recorded in Table 1 to evaluate the model. For questionnaire's, evaluators were allowed to rate the options as extremely satisfied, very satisfied, somewhat satisfied, not so satisfied, and not at all satisfied for five closed-ended questions.

CHAPTER-3

FUNCTIONALITY/ WORKING

Tools and Technology Used:

I. MATLAB Tools

There are 4 classes of diseases for which the images were collected. The images were less in number. As deep learning-based approaches need more images, we have performed data augmentation using MATLAB. This is one time process which gives 10 rotated versions of original image. Therefore, lot of data is generated which is helpful to training model.

II. Google Collab

We have used various data science related libraries like keras, tensorflow, sklearn, opencv, matplotlib, numpy etc. For the purpose of building keras model we have used sequential modelling technique. The architecture of model consists of two conv2D layers. Each conv2D layer is followed by activation layer named 'relu' and maxpooling layer. MA polling is that the highest value they can catch and declare. Once the data is available at final maxpooling layer, it is subjected to set of fully connected neuron as they are in ANN. For this purpose, flattening is done and dense layers are added. Flatten is convert output to in one dimensional array that work is done by flatter. This creates the architecture of deep learning model which will be trained using the data which was Uploaded earlier the data is available on the colab server. Path variable will read the images from the path one by one. Each image is read using opencv library. Subsequently, the images are resized with dyadic image processing. The paths of the images also talk about the Class of each image which is extracted and stored in a variable called label. The data and label lists are converted into numpy arrays for the purpose of training the model.

Convolutional Neural Networks

Introduced in the area of machine learning in 1990 by Lecun et al., CNNs represent the continuing development of traditional artificial neural networks; they have a greater capacity for learning due to the representation of the data in hierarchic form in various levels of abstraction.

Basically, a CNN functions by performing various convolutions in different layers of the network. This creates different representations of the set of learning data, starting with the most general in the first layers and becoming more specific in the deeper layers. The convolutional layers act as a kind of extractor of characteristics, since the reduction of the dimensionality of the entrance data groups them in layers.

The convolutional layers codify various basic resources into more discriminative ones, creating a series of intermediate layers separated by rectified linear units (ReLU) and layers of enveloping (pooling). In a more generic explanation, the convolutional layers act as filters, which transform one entrance image into another, highlighting specific patterns. In the next to last layer of the network, the final characteristics of the image are emphasized; thus, the final stage (layer) of the network acts as a linear classifier. A more detailed description of this kind of model can be found in Lecun et al.

With standard image processing algorithms, the impact on the result occurs as a function of image quality, type and resolution; it also depends on the type of descriptor of the characteristics and the classifier selected. In the case of a CNN, it is the structure, which has the greatest influence on performance in the extraction and classification of the resources/aspects of the images. Ososkov and Goncharov mention that shallow networks limit the learning capacity because of the limited number of layers.

For the recognition of lesions on cotton leaves based on image processing, the basic architecture for a CNN is found with Google Net and ResNet50 .

The processing structure of images for deep learning considers a stage for the weighting of the data and another for the artificial increase of data. Once the data is separated into training and testing sets, the number of images in the class with the smallest number of images is defined as the maximum value for the counting of images in each class for each of the two sets of data (training or test).

The final step involves the artificial increase of the data, with images redimensioned to a standard size and defined for the first layer of the CNNs, i.e., the images were redimensioned to 224 × 24 × 3 pixels.

Both the standard image processing algorithms and CNNs, as well as the training and testing processes, were implemented using the Toolbox of Image Processing Computer Vision and the Toolbox of Machine Learning of the MATLAB 2018a software (Mathworks), installed in an HP Z800 computer with two 2.66 GHz Xeon X5650 Intel processors with a 12 M cache, 128 Gb Ram, 6.40 GT/s Intel QPI, and a Nvidea FX 3800 video board based on the Windows 10 operational system.

For the analysis of the deep learning algorithm, the confusion matrix was used as a metric for defining applicability which could be compared with traditional image processing algorithms. The main diagonal of the confusion matrix provides a summary of the performance of the models, since the correct classifications are found there.

One single image contains number of pixels. These pixels group together it makes entire image. CNN uses feature detector. Feature detector used to detect significant features of image data in order to provide detection. It is the smallest matrix of weights. To reduce the bigger images into smaller images strides are used. Stride is the number of pixels by which we slide our filter matrix over the input matrix. This process is called convolution. By this the shape of the input image is modified feature detection there by detecting the particular feature from the input image and to get the information about that feature. This is called the feature map. Large images take lot of time. It is easier to process small images in faster manner.

According to Shuyue, they outlined the different formats of graph convolutional neural network. It was prepared to process the uniform electro encephalography data for the purpose of predicting the four classes of motor imaginaries to relate with electro encephalography electrode. They addressed their data with the transformation of 2D to 3D perspectives. The structure was processed through these dimensional units.

A study stated that, in order to utilize the dynamic route of deep learning, they proposed short-term voltage stability. They managed the clustering algorithm to obtain short-term voltage stability to increase the reliability.

In, it is stated that deep learning technique was applied to identify the leaf diseases in different mango trees. The researchers used five different leaf diseases from various specimens of mango leaf's where they addressed nearly 1200 datasets. The CNN structure was trained with more than 600 images, where 80% are used for training and 20% are used for testing. Remaining 600 images were used to find the accuracy and to identify the mango leaf diseases which showed the feasibility of its usage in real-time applications. The classification accuracy can be further increased if more images in the dataset are provided by tuning the parameters of the CNN model.

Eclipse 3.3 Indigo IDE is used for overall development as a standard platform. JSP is used to design GUI of the Web Application. The database is basically used for storing user details like usernames, disease details, pesticide information and pesticide costs. We used the MYSQL database as the data to be stored was mostly structured. The tool used for DB functionalities was MYSQL GUI Browser.

By comparing different classification algorithm we come to know that CNN is a feed-forward neural network that is used to analyze visual images by processing data. CNN can give a higher accuracy in the prediction on plant disease due to different advantages. We detect whether the plant is having some disease or not and what pesticides are needed for recovery from that disease with the help of CNN. The datasets are divided into Training and testing datasets in the ratio 70:30 and cross-validation is applied to achieve better accuracy.

For building CNN Python 3.6 is used along with libraries such as OpenCV for image processing and TensorFlow for Machine learning. Using libraries allows us to extend the functionalities for using it on larger database. CNN consists of three Hidden layers i.e. Convolution layers; each convolution layer consists of Pooling and Relu layers. Relu is the activation function used to remove the negative weights. MaxPooling is used in the pooling layers. 64 different filters are applied in each layer for extracting features. Convolving window size is 3*3 and stride size is 4. Softmax layer at the end linearize the probabilities of the diseases in a single dimensional matrix to choose maximum probability.

In this model they're using Crop Image, Agriculture image Processing, Image segmentation, Histogram Equalization. They divide the diseases within the following three categories: Bacterial disease, fungal diseases, viral disease. The image processing technique is used for detecting diseases on cotton leaf early and accurately. The processing scheme consists of image acquisition through camera or web, image pre-processing includes image enhancement and image segmentation where the affected and useful areas are segmented, feature extraction and classification. Finally, the presence of diseases on the plant leaf are going to be identified. Eclipse 3.3 Indigo IDE is used for overall development as a standard platform. JSP is used to design GUI of the Web Application. The database is basically used for storing user details like usernames, disease details, pesticide information and pesticide costs. We used the MYSQL database as the data to be stored was mostly structured. The tool used for DB functionalities was MYSQL GUI Browser.

Introduced in the area of machine learning in 1990 by Lecun et al., CNNs represent the continuing development of traditional artificial neural networks; they have a greater capacity for learning due to the representation of the data in hierarchic form in various levels of abstraction.

Basically, a CNN functions by performing various convolutions in different layers of the network.

This creates different representations of the set of learning data, starting with the most general in the first layers and becoming more specific in the deeper layers. The convolutional layers act as a kind of extractor of characteristics, since the reduction of the dimensionality of the entrance data groups them in layers.

The convolutional layers codify various basic resources into more discriminative ones, creating a series of intermediate layers separated by rectified linear units (ReLU) and layers of enveloping (pooling). In a more generic explanation, the convolutional layers act as filters, which transform one entrance image into another, highlighting specific patterns. In the next to last layer of the network, the final characteristics of the image are emphasized; thus, the final stage (layer) of the network acts as a linear classifier. A more detailed description of this kind of model can be found in Lecun et al.

With standard image processing algorithms, the impact on the result occurs as a function of image quality, type and resolution; it also depends on the type of descriptor of the characteristics and the classifier selected. In the case of a CNN, it is the structure, which has the greatest influence on performance in the extraction and classification of the resources/aspects of the images. Ososkov and Goncharov mention that shallow networks limit the learning capacity because of the limited number of layers.

By comparing different classification algorithm we come to know that CNN is a feed-forward neural network that is used to analyze visual images by processing data. CNN can give a higher accuracy in the prediction on plant disease due to different advantages. We detect whether the plant is having some disease or not and what pesticides are needed for recovery from that disease with the help of CNN. The datasets are divided into Training and testing datasets in the ratio 70:30 and cross-validation is applied to achieve better accuracy.

For building CNN Python 3.6 is used along with libraries such as OpenCV for image processing and TensorFlow for Machine learning. Using libraries allows us to extend the functionalities for using it on larger database. CNN consists of three Hidden layers i.e. Convolution layers; each convolution layer consists of Pooling and Relu layers. Relu is the activation function used to remove the negative weights. MaxPooling is used in the pooling layers. 64 different filters are applied in each layer for extracting features. Convolving window size is 3*3 and stride size is 4. Softmax layer at the end linearize the probabilities of the diseases in a single dimensional matrix to choose maximum probability.

For feature extraction, they are using the K-mean clustering algorithm method for classification and Neural-network as recognizer. The step-by-step procedure is shown as below.

1) RGB image acquisition

2) Pre-processing of image using histogram equalization

3) Resize the image

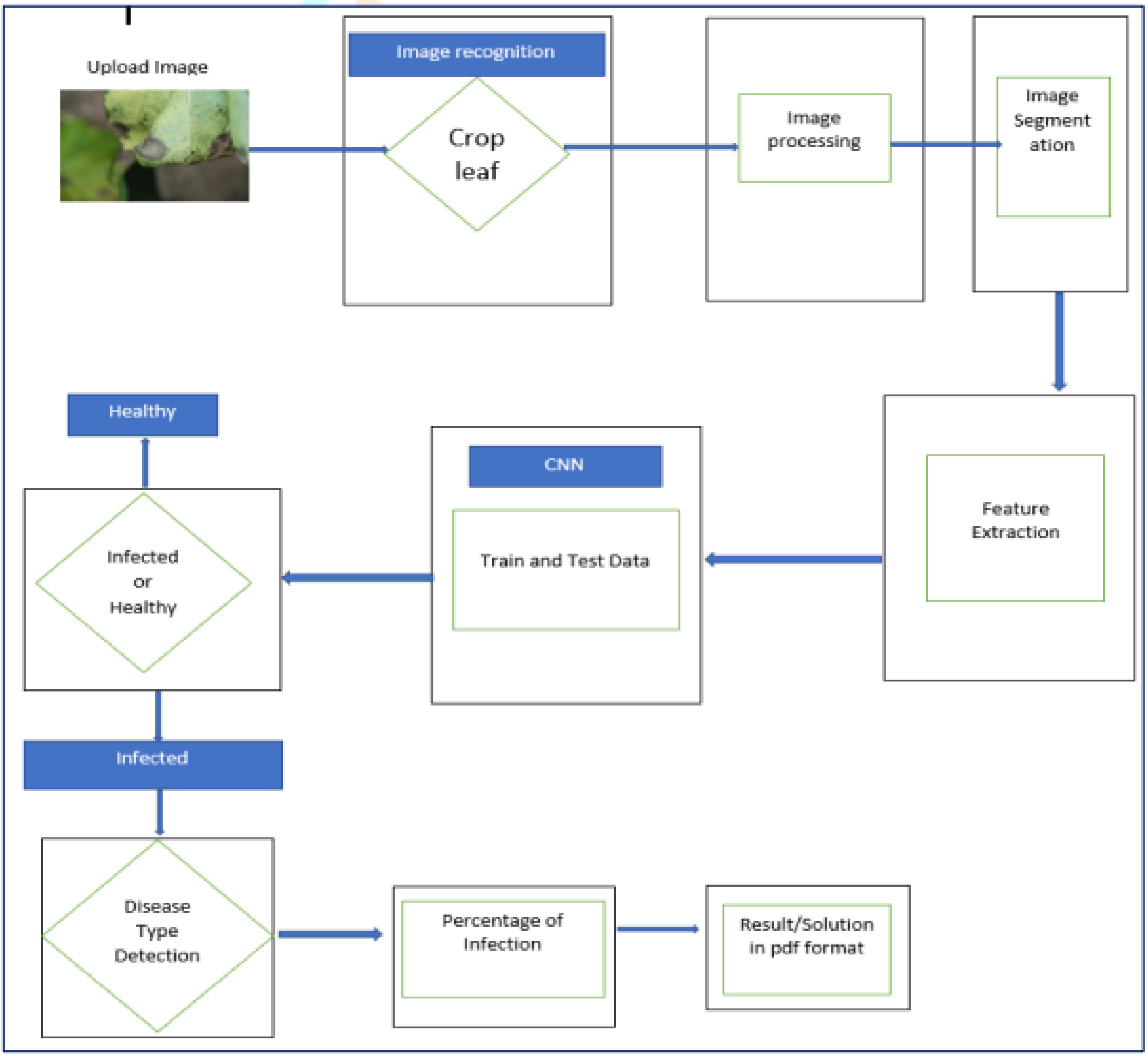
4) K-mean Algorithm for image segmentation

5) Computing features extraction

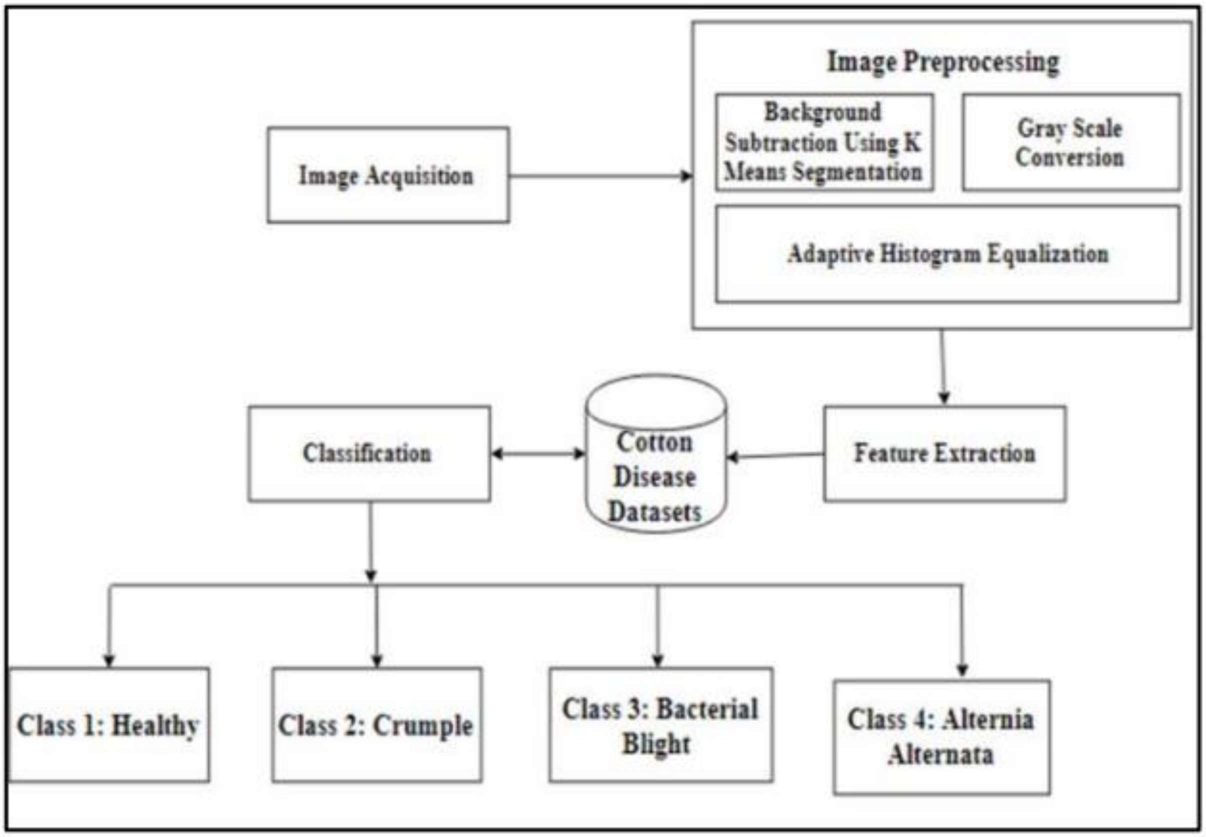
6) Classification & Recognition using neural networks

7) Statistical analysis.

Study of diseases on the cotton leaf studied by using the image processing toolbox and also the diagnosis Using MATLAB helps to suggest necessary remedies for that disease arising on the leaf of a cotton plant. The accurate recognition for using K- Mean Clustering method the Euclidean distance is 89.56% and the execution time for K-Mean Clustering method using Euclidean distance is 436.95 second and also, thresholding is completed by a dynamical range [0,1] counting on the color intensity from leaves image. Hence that disease detects using K-mean Clustering method using Euclidean distance and its excellent methods of disease detection on cotton leaves.



Block Diagram



System Architecture of Cotton Disease Prediction

Algorithm

Traditional feature learning methods rely on semantic labels of images as supervision. They usually assume that the tags are evenly exclusive and thus do not point out towards the complication of labels. The learned features endow explicit semantic relations with words.

We also develop a novel cross-model feature that can both represent visual and textual contents. CNN itself is a technique of classifying images as a part of deep learning. In which we apply single neural network to the full image.

i. Accepts a volume of size$W1\times H1\times D1$

ii. Requires four hyper parameters:

- Number of filters K
- Their spatial extent F
- The stride S
- The amount of zero padding P

iii. Produces a volume of size $W2\times H2\times D2$ where:

a. $W2=(W1-F+2P)/S+1$

b. $H2=(H1-F+2P)/S+1$(i.e. width and height are computed equally by symmetry)

c. $D2=K$

iv. With parameter sharing, it introduces F*F*D1 weights per filter, for a total of (F*F*D1)*K weights and K biases. In the output volume, the dth depth slice (of size W2*H2) is the result of performing a valid convolution of the dth filter over the input volume with a stride of S, and then offset by dth bias.

v. A common setting of the hyper parameters is F=3, S=1 P=1

Implementation

Eclipse 3.3 Indigo IDE is used for overall development as a standard platform. JSP is used to design GUI of the Web Application. The database is basically used for storing user details like usernames, disease details, pesticide information and pesticide costs. We used the MYSQL database as the data to be stored was mostly structured. The tool used for DB functionalities was MYSQL GUI Browser.

By comparing different classification algorithm we come to know that CNN is a feed-forward neural network that is used to analyze visual images by processing data. CNN can give a higher accuracy in the prediction on plant disease due to different advantages. We detect whether the plant is having some disease or not and what pesticides are needed for recovery from that disease with the help of CNN. The datasets are divided into Training and testing datasets in the ratio 70:30 and cross-validation is applied to achieve better accuracy.

For building CNN Python 3.6 is used along with libraries such as OpenCV for image processing and TensorFlow for Machine learning. Using libraries allows us to extend the functionalities for using it on larger database. CNN consists of three Hidden layers i.e. Convolution layers; each convolution layer consists of Pooling and Relu layers. Relu is the activation function used to remove the negative weights. MaxPooling is used in the pooling layers. 64 different filters are applied in each layer for extracting features. Convolving window size is 3*3 and stride size is 4. Softmax layer at the end linearize the probabilities of the diseases in a single dimensional matrix to choose maximum probability.

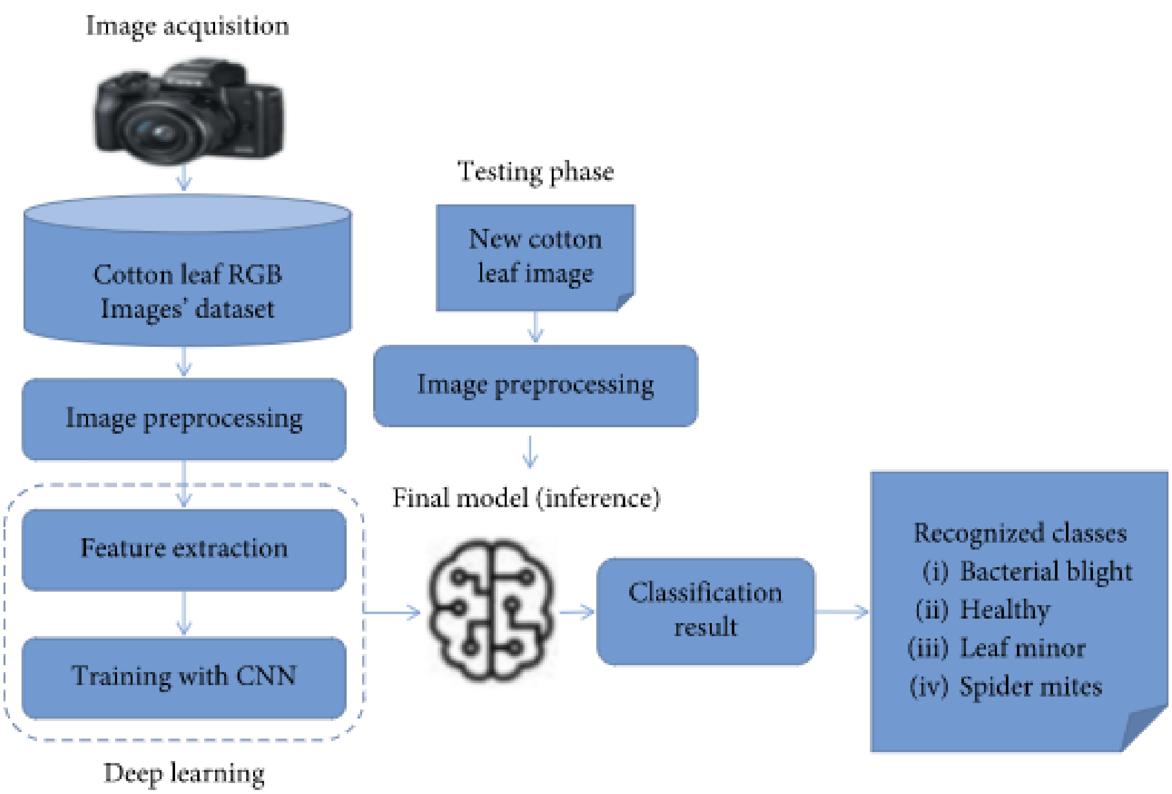
Xampp server is used to deploy the Web Application on local computer. Server handles the request and response to system user.

In, it is stated that deep learning technique was applied to identify the leaf diseases in different mango trees. The researchers used five different leaf diseases from various specimens of mango leaf's where they addressed nearly 1200 datasets. The CNN structure was trained with more than 600 images, where 80% are used for training and 20% are used for testing. Remaining 600 images were used to find the accuracy and to identify the mango leaf diseases which showed the feasibility of its usage in real-time applications. The classification accuracy can be further increased if more images in the dataset are provided by tuning the parameters of the CNN model.

In this model they're using Crop Image, Agriculture image Processing, Image segmentation, Histogram Equalization. They divide the diseases within the following three categories: Bacterial disease, fungal diseases, viral disease. The image processing technique is used for detecting diseases on cotton leaf early and accurately. The processing scheme consists of image acquisition through camera or web, image pre-processing includes image enhancement and image segmentation where the affected and useful areas are segmented, feature extraction and classification. Finally, the presence of diseases on the plant leaf are going to be identified. For feature extraction, they are using the K-mean clustering algorithm method for classification and Neural-network as recognizer.

Designing of Cotton Plant Disease Model

The first task in this model designing is image acquisition from the field with digital camera and smartphone. Then, image preprocessing techniques were applied to prepare acquired images for further analysis. After this, preprocessed images were inserted into the CNN algorithm to feature extraction with neural network. Then, best-suited extractions to represent the image are extracted from the image using an image analysis technique. Based on the extracted features, the training and testing data that are used to identify are extracted. Finally, a trained knowledge base classifies a new image into its class of syndromes.



Cotton leaf diseases recognition model process

Now a day's financial impact of agriculture is increase day by day with economic growth of our country, still agriculture is one wide sectors and that is play very important role for our county. While keeping track of diseases in plants with the help of specialists can be very costly in agriculture region. There is a need for a system which can automatically detect the diseases as it can bring revolution in monitoring large fields of crop and then plant leaves can be taken cure as soon as possible after detection of disease. There are so many diseases that affect for cotton and many more crop that effect many filed of agriculture. So those identify this disease and how to recover from it. This objective will satisfy though our application which is helps for do prediction of cotton disease as well how to overcome from it.

For this project I have download dataset form Kaggel . In this dataset there are three folders like train , test and validation folder . In this dataset there are 4 category like diseased cotton leaf, diseased cotton plant, fresh cotton leaf , fresh cotton plant .

For this project, I have created deep learning model convolutional neural network (CNN) using keras library for our project cotton disease prediction. First of all, why I am using CNN because CNNs are used for image classification and recognition because of its high accuracy. The CNN follows a hierarchical model which works on building a network, like a funnel, and finally gives out a fully-connected layer where all the neurons are connected to each other and the output is processed.

After getting the dataset next step is to pass our training data for our CNN model to learn to identify or classify different classes of images . The model architecture used was:

```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #   
=================================================================
conv2d_1 (Conv2D)            (None, 148, 148, 32)      896       
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 74, 74, 32)        0         
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 72, 72, 64)        18496     
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 36, 36, 64)        0         
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 34, 34, 128)       73856     
_________________________________________________________________
max_pooling2d_3 (MaxPooling2 (None, 17, 17, 128)       0         
_________________________________________________________________
conv2d_4 (Conv2D)            (None, 15, 15, 256)       295168    
_________________________________________________________________
max_pooling2d_4 (MaxPooling2 (None, 7, 7, 256)         0         
_________________________________________________________________
dropout_1 (Dropout)          (None, 7, 7, 256)         0         
_________________________________________________________________
flatten_1 (Flatten)          (None, 12544)             0         
_________________________________________________________________
dense_1 (Dense)              (None, 128)               1605760   
_________________________________________________________________
dropout_2 (Dropout)          (None, 128)               0         
_________________________________________________________________
dense_2 (Dense)              (None, 256)               33024     
_________________________________________________________________
dropout_3 (Dropout)          (None, 256)               0         
_________________________________________________________________
dense_3 (Dense)              (None, 4)                 1028      
=================================================================
Total params: 2,028,228
Trainable params: 2,028,228
Non-trainable params: 0
_________________________________________________________________
```

The loss function used was "sparse_categorical_crossentropy" and optimizer used was "Adam" . For training the model we used Keras API with tensorflow at backend.

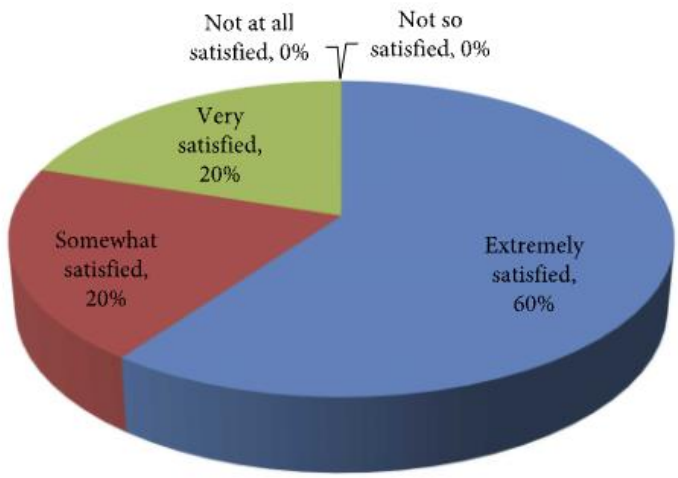
After the creating CNN model , as per the data science life cycle we have to deploy our model . For deploy our model into web application we use python framework Flask to connect CNN model with HTML , CSS & JavaScript file . For using model into flask we have to save our model .

After saving a model create web application using HTML , CSS & JAVASCRIPT and then connect with Flask .

So , all the things have finally done , I have attach some of the output the give from my web application .

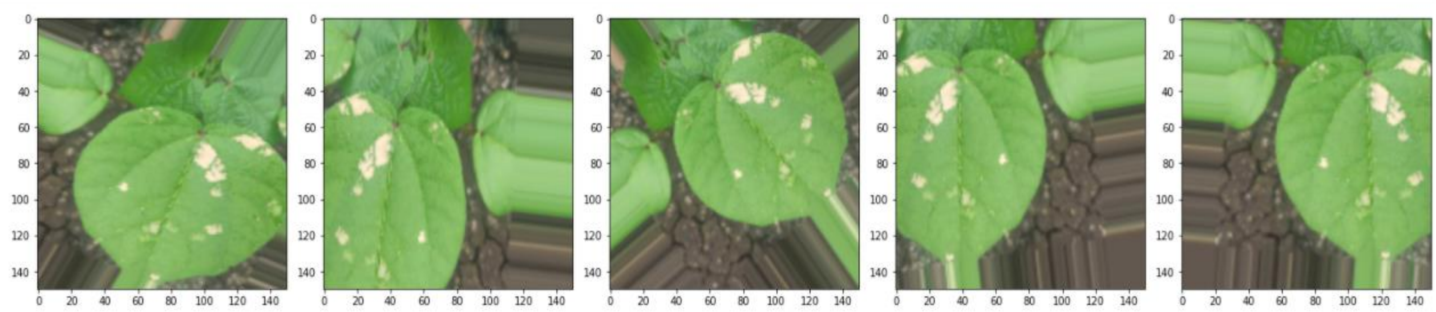
Survey

In the survey of users, the overall performance of the cotton disease and pest identification prototype evaluation selected by the evaluator was 60% extremely satisfied option for all questions and 20% of very satisfied and somewhat satisfied option. Also, for the open-ended question, almost all experts reflect constrictive thoughts on the overall performance of the system and prototype. So, this result shows that the prototype of cotton diseases and pests was performed well in problem-solving ability and making a correct prediction.



CODE:-

```
[ ] # showing augmented images
    images = [training_data[0][0][0] for i in range(5)]
    plotImages(images)
```

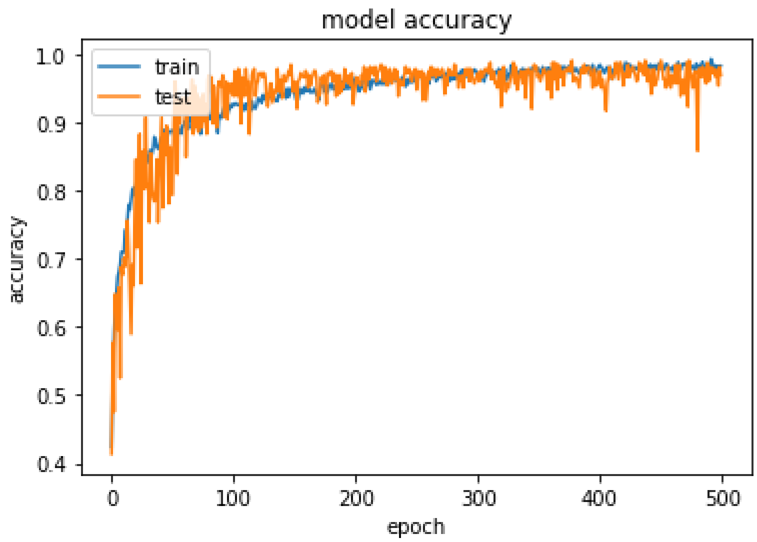


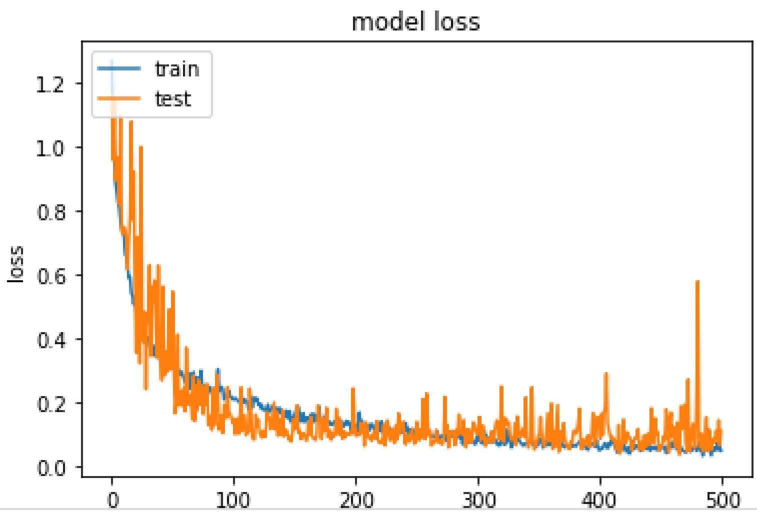
```
[ ] # save best model using vall accuracy
```

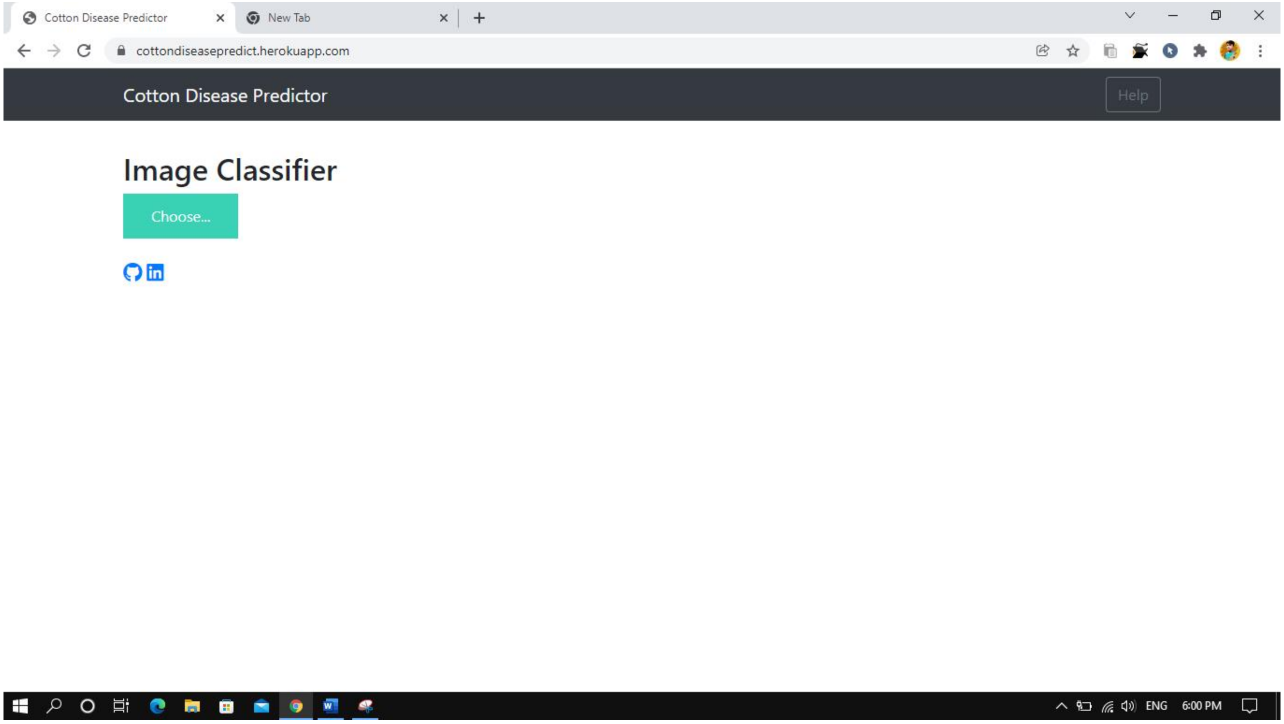
```
[ ] # train cnn model
    history = cnn_model.fit(training_data,
                            epochs=500,
                            verbose=1,
                            validation_data= valid_data,
                            callbacks=callbacks_list) # time start 16.06
```

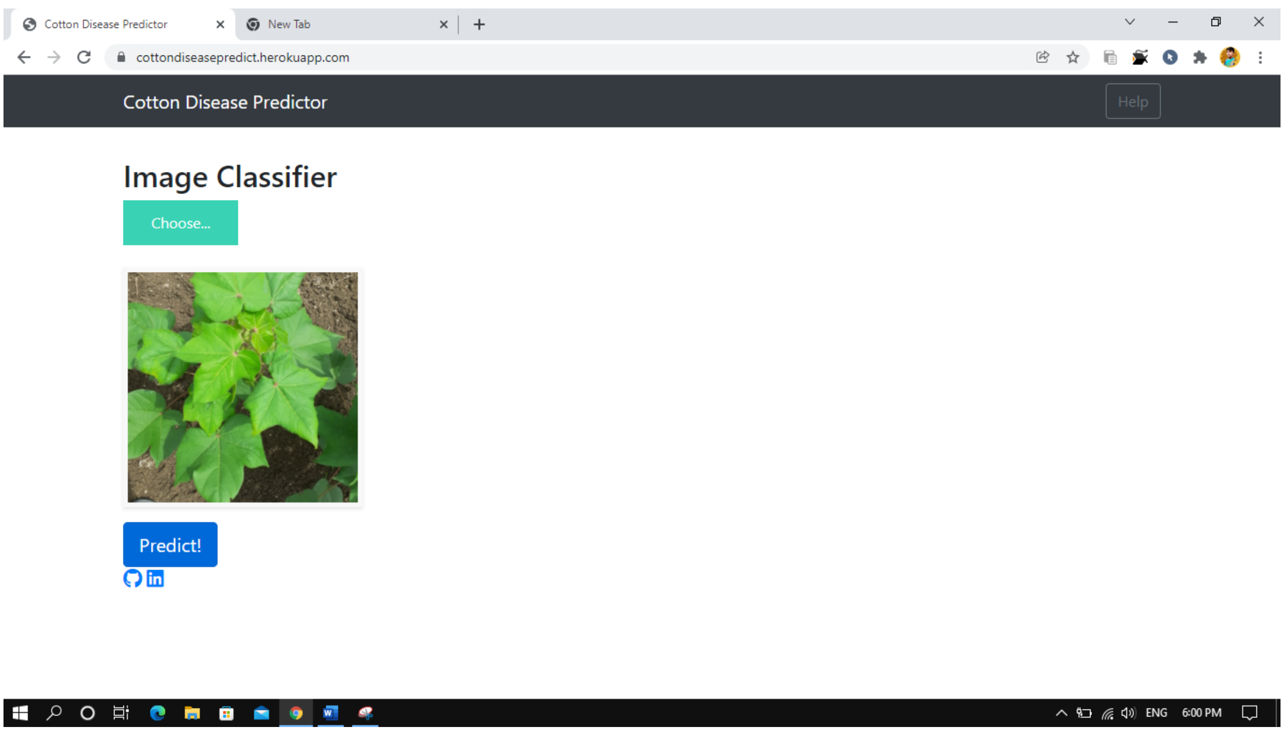
```
Epoch 1/500
61/61 [==============================] - ETA: 0s - loss: 1.2668 - accuracy: 0.4239
Epoch 00001: val_accuracy improved from -inf to 0.41358, saving model to /content/drive/My Drive/My ML Project /DL Project/CNN/cotton plant disease prediction/v4_pred_cott_dis.h5
61/61 [==============================] - 30s 499ms/step - loss: 1.2668 - accuracy: 0.4239 - val_loss: 1.1550 - val_accuracy: 0.4136
Epoch 2/500
61/61 [==============================] - ETA: 0s - loss: 1.0705 - accuracy: 0.5295
Epoch 00002: val_accuracy improved from 0.41358 to 0.57716, saving model to /content/drive/My Drive/My ML Project /DL Project/CNN/cotton plant disease prediction/v4_pred_cott_dis.
61/61 [==============================] - 31s 516ms/step - loss: 1.0705 - accuracy: 0.5295 - val_loss: 0.9566 - val_accuracy: 0.5772
Epoch 3/500
61/61 [==============================] - ETA: 0s - loss: 0.9644 - accuracy: 0.5920
Epoch 00003: val_accuracy did not improve from 0.57716
61/61 [==============================] - 29s 478ms/step - loss: 0.9644 - accuracy: 0.5920 - val_loss: 1.1414 - val_accuracy: 0.4753
Epoch 4/500
61/61 [==============================] - ETA: 0s - loss: 0.8841 - accuracy: 0.6284
Epoch 00004: val_accuracy improved from 0.57716 to 0.64815, saving model to /content/drive/My Drive/My ML Project /DL Project/CNN/cotton plant disease prediction/v4_pred_cott_dis.
61/61 [==============================] - 33s 545ms/step - loss: 0.8841 - accuracy: 0.6284 - val_loss: 0.8928 - val_accuracy: 0.6481
Epoch 5/500
61/61 [==============================] - ETA: 0s - loss: 0.8573 - accuracy: 0.6509
```

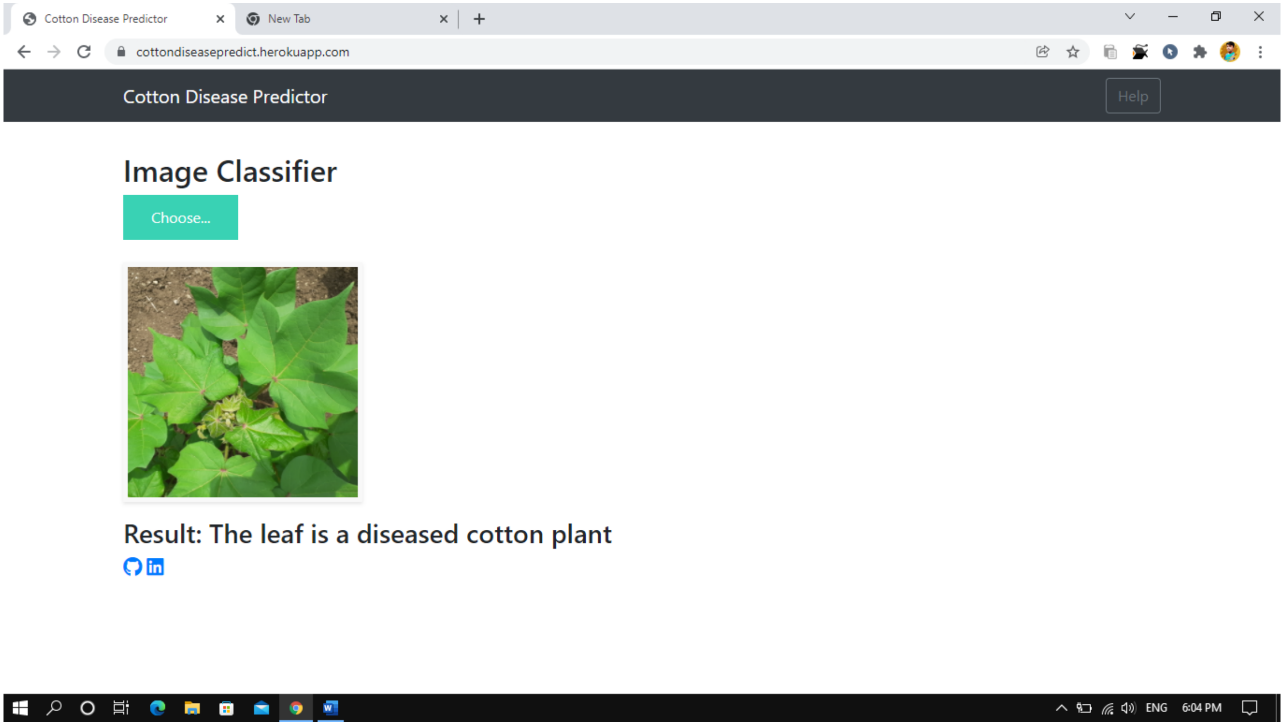
```
# summarize history for accuracy
plt.plot(history.history['accuracy'])
plt.plot(history.history['val_accuracy'])
plt.title('model accuracy')
plt.ylabel('accuracy')
plt.xlabel('epoch')
plt.legend(['train', 'test'], loc='upper left')
plt.show()
# summarize history for loss
plt.plot(history.history['loss'])
plt.plot(history.history['val_loss'])
plt.title('model loss')
plt.ylabel('loss')
plt.xlabel('epoch')
plt.legend(['train', 'test'], loc='upper left')
plt.show()
```

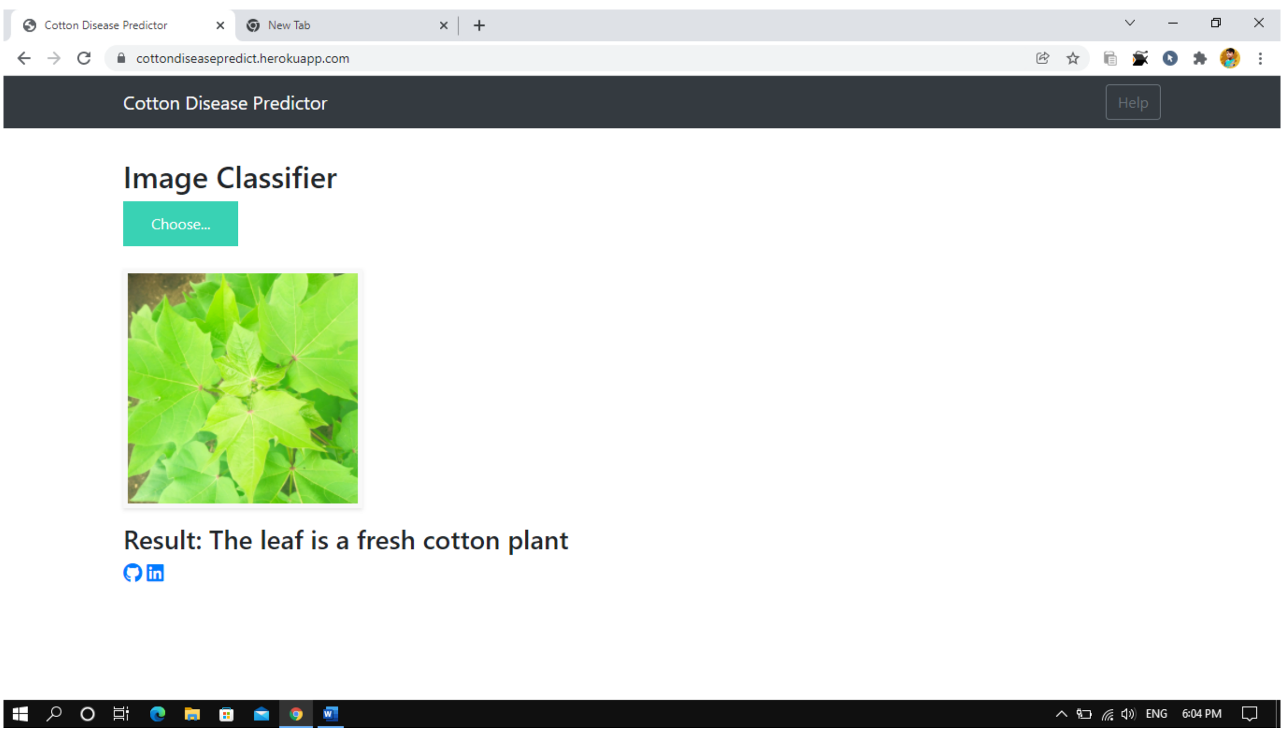

model accuracy
train
test
accuracy
epoch
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0
100
200
300
400
500


model loss
train
test
loss
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
100
200
300
400
500


Cotton Disease Predictor
New Tab
cottondiseasepredict.herokuapp.com
Cotton Disease Predictor
Help
Image Classifier
Choose...
ENG
6:00 PM


Cotton Disease Predictor
New Tab
cottondiseasepredict.herokuapp.com
Cotton Disease Predictor
Help
Image Classifier
Choose...
Predict!
ENG
6:00 PM


Cotton Disease Predictor
New Tab
cottondiseasepredict.herokuapp.com
Cotton Disease Predictor
Help
Image Classifier
Choose...
Result: The leaf is a diseased cotton plant
ENG
6:04 PM


Cotton Disease Predictor
New Tab
cottondiseasepredict.herokuapp.com
Cotton Disease Predictor
Help
Image Classifier
Choose...
Result: The leaf is a fresh cotton plant
ENG
6:04 PM

CHAPTER-4
Results and Discussion

The goal of this application is to develop a system which recognizes crop diseases and displays user the results as detected disease, pesticides recommended and cost of pesticides recommended, and for that user have to upload an image then, Image processing starts with the digitized color image of the diseased leaf. Finally, by applying the CNN plant disease can be predicted.

The dataset contains two types of diseased leaf images that are Alternaria Macrospora and Bacterial Blight and also Healthy Leaf images. The training dataset consists of total 513 images while the testing dataset consists of 207 images. The accuracy of training is 80% whereas the accuracy of testing is 89%.

The results are obtained by running the Cotton leaf disease detection system (GUI) module as follows:

The dataset is created by background subtraction and feature extraction and stored in a feature vector and that done for around 500+ images and stored in a .MAT file (A Mat Data file) consists of a 2 variables Dataset and Disease type vector array.

- The input image is loaded from the system into the Cotton leaf disease detection system.
- Then the read image is stored in a matrix for image processing.
- The system Checks the input image is cotton leaf or not.
- Background Subtraction through K-means is performed on the cotton image and displayed in the GUI.
- Adaptive Histogram is applied to the Image
- Using the Multiclass Support Vector Machine for 100 images 70% accuracy is achieved.
- Using the K-Nearest Neighbor Classifier 86% accuracy is achieved.
- Using the Artificial Neural Network 86% accuracy is achieved.

CHAPTER-5
Conclusion and Future Scope

5.1 Conclusion

A web-based system has been successfully implemented for crop disease detection for cotton leaves using Convolutional Neural Network. Two diseases viz. Alternaria Macrospora and Bacterial Blight are successfully being detected by system with training accuracy 80% using training dataset of images for each disease. Healthy leaf image set of images is also used for detecting healthy images. Convolutional neural network has been developed with three hidden layers to classify the cotton leaf disease images. System successfully takes image input from user and provides input in the form of disease detected, preventive measures, corrective measures, pesticides required and probable cost for suggested pesticides.

System can be extended to any other crops having availability of enough large dataset for that crop. Number of other diseases can be included for detection. System also can implement hardware using IOT for Image capturing in fields. The Web interface can also involve a forum for formers to have discussions regarding the current trends they are facing in different diseases.

5.2 Future Scope

Currently the Cotton Disease Detection System is capable of detecting diseases such as Alternaria alternate, Bacterial blight of cotton and Crumple leaf.

- Further the system can be trained to recognize different types of diseases.
- The current system can be simplified by developing an android application, which could be convenient for users.
- The current system uses 500+ images for the training and classification. The image dataset can be increased and thus that can provide more efficient results, and other efficient methods for classification can be experimented for the accurate results.
- Experimenting with noisier and the distorted images can be tried.
- Applying this project for real time image capturing and classification can be done, by interfacing with camera module into the system. Which captures images and upload to the system for the classification results can be experimented.

Reference

- Kumar, S., Jain, A., Shukla, A.P., Singh, S., Raja, R., Rani, S., Harshitha, G., AlZain, M.A. and Masud, M., 2021. A Comparative Analysis of Machine Learning Algorithms for Detection of Organic and Nonorganic Cotton Diseases. Mathematical Problems in Engineering, 2021.
- P. R. Rothe R. V. Kshirsagar "Cotton Leaf Disease Identification using Pattern Recognition Techniques" International Conference on Pervasive Computing (ICPC) IEEE pp. 1-6 2015.
- Ratih Kartika Dewi, R. V. Hari Ginardi, "Feature Extraction for Identification of Sugarcane Rust Disease" 2014 International Conference on Information, Communication Technology and System.
- R. W. Briddon and P. G. Markham, "Cotton leaf curl virus disease," Virus Research, vol. 71, no. 1-2, pp. 151–159, 2000.
- W. Zhang and Y. Du, "Rhizosphere fungal community structure succession of Xinjiang continuously cropped cotton," Fungal Biology, vol. 123, pp. 42–50, 2018.

.