e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:05/May-2023 Impact Factor- 7.868 www.irjmets.com

COTTON CROP DISEASE PREDICTION USING CNN

Ms. Rutika Khedkar[*1], Ms. Gayatri Kurhekar[*2], Mr. Dhananjay Khodaskar[*3], Ms. Anushka Sirsat[*4], Ms. Aishwarya Deshmukh[*5], Prof. A. R. Bhuyar[*6]

[*1,2,3,4,5]Students, Dept of I.T, Sipna COET, Amravati, India

[*6]Assistant Professor, Dept of I.T, Sipna COET, Amravati, India

DOI : https://www.doi.org/10.56726/IRJMETS38922

ABSTRACT

India is the largest country in terms of production of cotton crop. This crop is the widely occupied all over the India. But this cotton crops get affected by most of the diseases like Bacterial blight, Curl Virus, Fussarium wilt and many more. General observation by farmers may be time-consuming, expensive and sometimes inaccurate. The Cotton leaf Disease Detection and identifying the disease at an early stage is a very difficult task for the farmers. To beat this problem, a machine learning approach is proposed which can assess the image of the leaf of plant and detect the disease over the cotton crop using machine learning approach. For availing this user got to upload the image then with the assistance of image processing ,we can proceed with applying CNN to predict cotton leaf disease. CNN extract each features of an image and deeply processed at each layer. Every disease on a crop has different features which are extracted at each layer of the convolution network. The goal of this application is to develop a system that recognizes crop diseases and also provide the required solution to overcome this disease by suggesting the particular fertilizer for particular disease.

Keywords: Cotton disease, CNN, Analysis, Designing, Algorithm, Results

I. INTRODUCTION

Agriculture is a field that significantly affects both life and economy. Each year, a number of illnesses that damage cotton plants create substantial financial losses for farmers that grow cotton. The most prevalent diseases that affect cotton plants include Bacterial Blight, Fusarium Wilt, and Curl Virus. If farmers can identify these diseases early on and treat them appropriately, they can reduce waste and save financial loss.

The goal of this study is to use CNN's deep learning architecture to detect or identify several diseases that affect cotton leaves. because Convolutional Neural Network can readily classify them but our unaided vision cannot. By offering them treatments for the disease, this fully automated technology helps farmers save time and effort when manually identifying cotton plant diseases. You won't believe it when I tell you that the top 5% error of human eyesight is even lower than the error of some pre-trained Neural Network Architectures, which has an error of about 3%. Farmers have a wide variety of options from which to choose crops that are right for their farms. However, those crops must be grown in a technical manner to produce an optimal yield of high-quality produce. Human vision is limited in its ability to diagnose diseases because the majority of the basic symptoms are tiny. This procedure is time-consuming and tedious. Identification of the disease's symptoms via image processing is a current research field that is of high priority. The farmers are battling the cotton leaf disease as a result of their way of life. A disease-diagnosis system that will aid farmers is necessary. This method focuses on identifying diseases by processing digital photographs of plant leaves that have been acquired. The key benefit of our initiative is that we offer disease solutions and information on which pesticides or insecticides are best for a given condition. By treating crops properly, farmers may prevent the spread of disease and improve the quality of their harvest. In order to predict the development and spread of illnesses in cotton crops, a complicated field known as cotton crop disease prediction is used. Since cotton is a significant crop in the world and its diseases can have serious negative effects on farmers, consumers, and the industry as a whole, this is an important area of research. We will give a thorough introduction to cotton crop disease prediction in this article, including the numerous methods that have been employed, their advantages and disadvantages, as well as the difficulties and potential in this area.

II. AIMS & OBJECTIVES

The motivation behind the cotton crop disease prediction project is to help farmers and the cotton industry as a whole to improve crop yields and reduce economic losses caused by disease outbreaks. Cotton is a major crop worldwide, and its diseases can have significant economic and social impacts on farmers, consumers, and the industry. By predicting the likelihood and severity of diseases, farmers can take timely and appropriate measures to prevent or control disease outbreaks, such as using fungicides, modifying irrigation practices, or adjusting planting schedules. Furthermore, disease prediction models can help farmers optimize their crop management practices and minimize losses. This can have significant economic benefits, such as increasing crop yields and improving the quality of cotton fiber. It can also contribute to the sustainability of the cotton industry by reducing the use of chemical inputs and promoting more efficient use of resources.

III. LITERATURE REVIEW

1. Pranita P.Gulve,Sharayu S.Tambe, Madhu A.Pandey,Mrs S.S. Kanse This paper uses GLCM, thresholding, segmentation, feature extraction for detecting the diseases. They create colour transformation structure for the rgb leaf image , apply colour space transformation structure then image is segmented.Unnecessary parts (green area) within leaf area is removed.Calculate the texture features for the segmented infected object.Extracted features are passed through a pre-trained neural network. Image is taken with the help of digital camera and all images are stored in a JPEG format. Image resize and image filtering is done. Here filtering is important to remove any noise content. For this they used gaussian low pass filter is applied and positive standard deviation sigma. In this the classifier used is euclidian distance classifier.[1]

2. A. Jenifa, R. Ramalakshmi, V. Ramchandran In this paper uses convolution neural networks, deep neural networks, leaf disease, classification, image processing. This training model contains 500 leaf images and testing model contains 100 leaf images. Cotton leaf diseases to be classified in this are - cercospora, bacterial blight, ascochyta blight, target spot. The tool which they used is MATLAB tool. Diseased cotton leaf taken as input image then input image converted into grey converted image then the noise co-efficient image, then it converted into filtered image, next morphological image then clustered image the segmentation image and finally the output image takes out the diseased part. This system shows the 96% accuracy for the classification of diseased cotton leaves.[2]

3. Rakesh Chaware, Rohit Karpe, Prithvi Pakhale, Prof. Smita Desai This system identify three diseases-aleternaria atternata, anthracnose and bacterial blight. This system works parallel for healthy and defected leaf image. Two images has Been taken one for the healthy leaf other for the defected leaf. Disease detection starts from training process. In training process resizing of the healthy and defected image has been done. Then convert rgb to grey scale image then apply stem, stairs, canny edge detection, surf, entropy, warp images. This technique is applied on both the samples healthy as well as defected. Once the training process of first phase samples is finished, comparison has been done on the basis of values obtained for all the parameters used.[3]

IV. PROPOSED SYSTEM

In a system for a specific purpose, we employ a real-time dataset that includes various images of cotton disease, such as bacterial blight, fusarium wilt, and curl virus disease, along with a prediction of the healthy plant. Some of the photographs are for testing, while others are for training. We first feed the model the photos from the real-time dataset for recognising the cotton illness. As soon as we input the images into our system, it displays the results in the form of the disease's classification as well as the specific fertiliser names we suggest using to diagnose the condition. It also specifies the fertiliser doses that should be applied to the cotton crop to achieve the best results.

The methodology of cotton disease detection using image processing has the following steps:

1) Cotton Images' Sample Digitization:

In order to provide accurate, unbiased, and simplified digital images of leaves in the cotton plant sample database for additional analysis and processing, the data acquisition was utilised. The goal was to deliver balanced illumination or consistent lighting to the digitising system. The photos taken with a smartphone camera and a digital camera are then uploaded to a computer, shown on a screen, and saved to the hard drive as digital colour photographs in the PNG, JPG, and JPEG formats.

2) Image pre-processing:

In this stage, higher-quality, higher-resolution photos are needed. All of these photographs have been cropped and resized in a particular way. Using a data augmentation approach, we rotate these photos and remove any noisy material.

Image segmentation is the process of breaking up a digital image into several portions. It can be used to readily identify which section of the model is contaminated by removing the affected pixel region from the leaf.

3) Feature extraction:

In this procedure, some of the crucial elements of the flawed leaf should be extracted. It is capable of generating coloured structures and converting the colour value from the RGB components of the cotton leaf image's defective areas. We can train our neural network using this feature. Once all procedures have been completed, the CNN algorithm is applied to the model and the train and test data are provided. The model's operation is depicted in the following flowchart.

On the colab server, the data is accessible. The photos on the route will be read one at a time by the path variable. The opencv library is used to read each image. The photos are then downsized using dyadic image processing. The paths of the photos also include information about each image's Class, which is extracted and kept in a label variable.

The data and label lists are converted into numpy arrays for the purpose of training the model. The data train: test split ratio is 75:25. Runtime data augmentation during training is also made available to the optimizer. For building the model, the classifier is notified that the image dimensions are 150*150*3 along with 4 classes. We have used sparse categorical cross entropy for measuring the losses during training process which are monitored continuously. We have used Adam optimizer which is latest optimizer that SGD. The training process gives a trained model which can further be used for testing purpose. For testing, we have uploaded the leaf images from validation dataset and check the obtained label with the actual one.

4.1. Deep Learning Architecture : CNN(Convolution Neural Network)

In the last few years of the IT industry, there has been a huge demand for once particular skill set known as Deep Learning. Deep Learning a subset of Machine Learning which consists of algorithms that are inspired by the functioning of the human brain or the neural networks.

These structures are called as Neural Networks. It teaches the computer to do what naturally comes to humans. Deep learning, there are several types of models such as the Artificial Neural Networks (ANN), Autoencoders, Recurrent Neural Networks (RNN) and Reinforcement Learning. But there has been one particular model that has contributed a lot in the field of computer vision and image analysis which is the Convolutional Neural Networks (CNN) or the ConvNets.

CNN has high accuracy, and because of the same, it is useful in image recognition. Image recognition has a wide range of uses in various industries such as medical image analysis, phone, security, recommendation systems, etc.

CNN has high accuracy, and because of the same, it is useful in image recognition. Image recognition has a wide range of uses in various industries such as medical image analysis, phone, security, recommendation systems, etc.

The term 'Convolution" in CNN denotes the mathematical function of convolution which is a special kind of linear operation wherein two functions are multiplied to produce a third function which expresses how the shape of one function is modified by the other. In simple terms, two images which can be represented as matrices are multiplied to give an output that is used to extract features from the image.

4.2 Basic Architecture

There are two main parts to a CNN architecture

- A convolution tool that separates and identifies the various features of the image for analysis in a process called as Feature Extraction.
- The network of feature extraction consists of many pairs of convolutional or pooling layers.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:05/May-2023 Impact Factor- 7.868 www.irjmets.com

- A fully connected layer that utilizes the output from the convolution process and predicts the class of the image based on the features extracted in previous stages.
- This CNN model of feature extraction aims to reduce the number of features present in a dataset. It creates new features which summarises the existing features contained in an original set of features. There are many CNN layers as shown in the CNN architecture diagram.

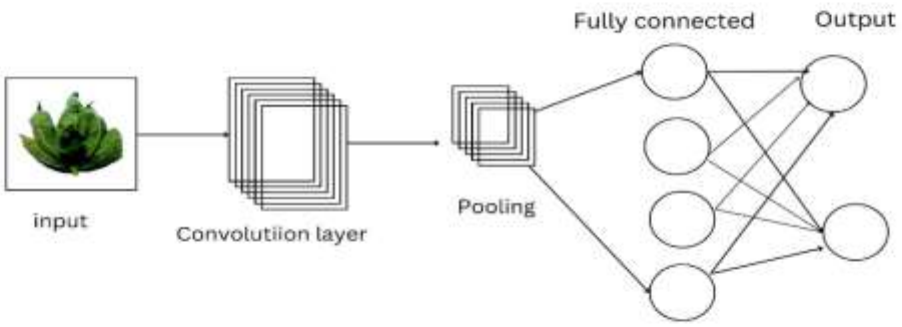

Fig.4.1.1 CNN Architecture

The CNN is made up of three different kinds of layers: fully-connected (FC), pooling, and convolutional layers. A CNN architecture is created when these layers are stacked.

4.2.1. Convolutional layer

The first layer utilised to extract the different features from the input photos is this one. Convolution is a mathematical process that is carried out at this layer between the input image and a filter of a specific size, MxM. The dot product between the filter and the portions of the input image with regard to the filter size (MxM) is taken by sliding the filter across the input image.

The result is known as the Feature map, and it provides details about the image, including its corners and edges. This feature map is later supplied to further layers to teach them additional features from the input image.

4.2.2. Pooling

After a convolutional layer, a pooling layer is frequently applied. The primary objective of this layer is to reduce the size of the convolved feature map in order to reduce computing costs. Using fewer links between layers and independently modifying each feature map, this is accomplished. Different pooling operations exist, depending on the technique used. It is essentially a summary of the features that a convolution layer produced.

The feature map provides the largest contribution to Max Pooling. Through average pooling, the average of the components in a region of an image with a predetermined size is calculated. Sum Pooling calculates the components' cumulative sums inside the given section.

4.2.3. Fully Layered Connection

The Fully Connected (FC) layer, which also has weights and biases, is used to connect the neurons between two layers. These layers are often positioned before the output layer and make up the last few layers of a CNN design.

This flattens the input image from the layers beneath and provides it to the FC layer. The typical operations on mathematical functions are then performed on the flattened vector through a few more FC levels. At this moment, the classification process begins to take place. Two layers are connected because two fully connected layers perform better than one connected layer. The degree of human oversight is reduced by these CNN levels.

4.4.4 Algorithm Used

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. 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.

1. Accepts a volume of size$W1 \times H1 \times D1$
2. Enquires four hyper parameters: Number of filters K Their spatial extent F The stride S The amount of zero padding P
3. Produces a volume of size $W2 \times H2 \times D2$ where:
4. W2= (W1-F+2P)/S+1
5. H2= (H1−F+2P)/S+1(i.e. width and height are Computed equally by symmetry)
6. D2=K
7. 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.
8. A common setting of the hyper parameters is F=3, S=1, P=1 However, there are common conventions and rules of thumb that motivate these hyper parameters.

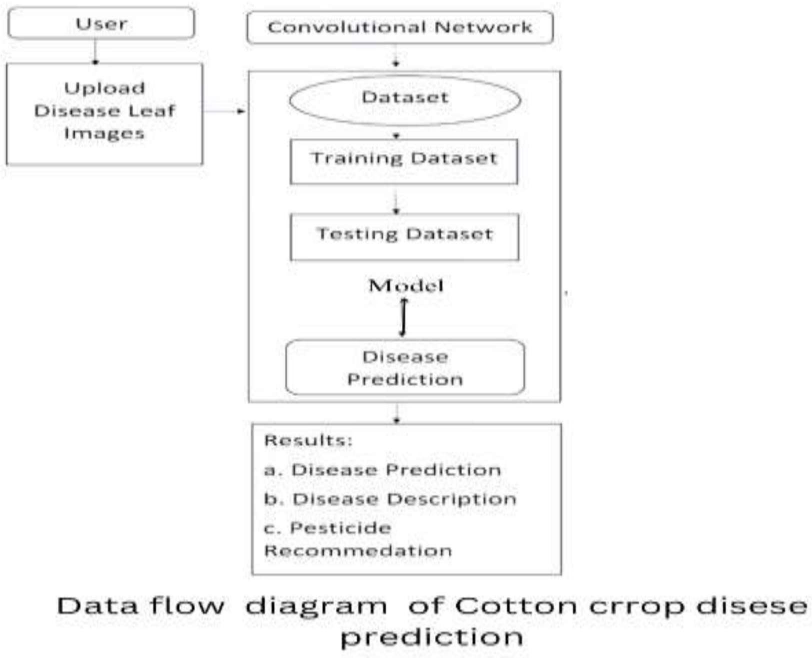

Fig 4.1.2 Data flow diagram

V. SYSTEM ANALYSIS AND DESIGN

The design goals of a cotton crop disease prediction system could include the following:

Accuracy: The system should be designed to accurately predict the occurrence of diseases in cotton crops. This can be achieved by using high-quality data and reliable algorithms to analyse the data.

Speed: The system should be able to provide predictions quickly, as early detection of diseases can be critical for preventing their spread and minimizing damage to the crop.

Scalability: The system should be able to handle large amounts of data, as well as a large number of users accessing the system simultaneously.

User-friendly interface: The system should be designed to be easy to use for both farmers and agricultural experts, with a user-friendly interface that allows users to input data and view results in a clear and understandable way.

Integration: The system should be designed to integrate with other agricultural systems, such as weather monitoring systems and irrigation systems, to provide a more comprehensive picture of the crop's health.

Accessibility: The system should be accessible to farmers in remote and rural areas, who may not have access to high-speed internet or advanced technology.

Cost-effectiveness: The system should be cost-effective to implement and maintain, so that it can be adopted by farmers with different budgets and resources.

By achieving these design goals, a cotton crop disease prediction system can help farmers to make informed decisions about disease prevention and management, ultimately leading to increased crop yield and profitability.

Here are the things that this system will perform:

1. Input the diseased plant image.
2. Output prediction
3. Information of crop cultivation

5.1 Input disease plant image: We have to enter the diseased plant leaf through the file choosing option in our system and click on predict.

5.2 Output Prediction: In output prediction the output is based on user input leaf and it predicts the class of disease based on the model prediction.

5.3 Information of crop cultivation: Here the information is given to the user by providing the youtube videos through which they can gain the required knowledge.

5.4 User Interface

The user interface design of system consist of following modules which are listed as below.

1. Introduction about System
2. Input field and predict button
3. Output display page

The screenshots of all the section along with information about its contents are mentioned below:

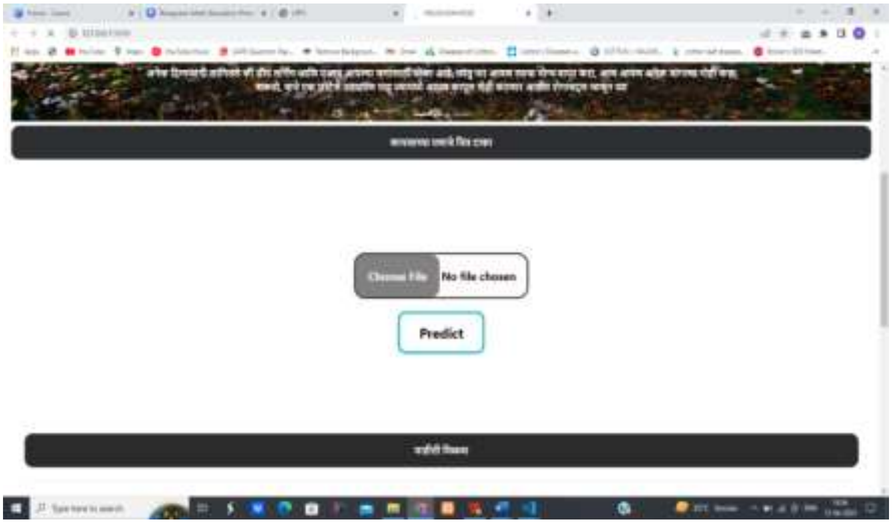

Fig 5.4.1 Prediction part

Fig 5.4.2 Output Predicted

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:05/May-2023 Impact Factor- 7.868 www.irjmets.com

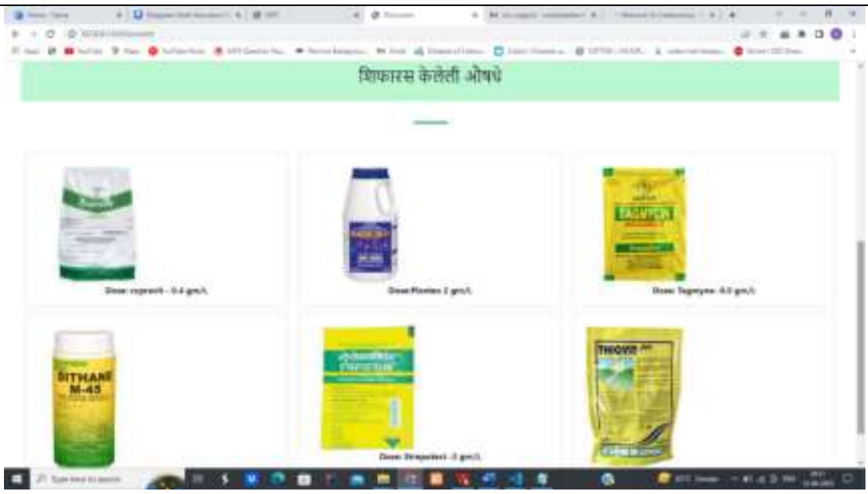

Fig 5.4.3 Fertilizer Suggested

VI. RESULTS AND DISCUSSION

1. Image processing result

Image augmentation parameters that are generally used to increase the data sample count are zoom, shear, rotation, preprocessing function and so on. Usage of these parameters results in generation of images having these attributes during training of Deep Learning model. Image samples generated using image augmentation, in general results, in increase of existing data sample set by nearly 3x to 4x times. In our project we use data generator library of keras python. This is a result of image augmentation.

Fig 6.1 Image Augmentation

2. CNN model building result

For this project, we have created deep learning model convolutional neural network (CNN) using keras library for our project cotton disease prediction. First of all, why we were 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.

3. Epochs

Every sample in the training dataset has had a chance to update the internal model parameters once during an epoch. One or more batches make up an epoch. The batch gradient descent learning algorithm, for instance, is used to describe an Epoch that only contains one batch.

Learning algorithms take hundreds or thousands of epochs to minimize the error in the model to the greatest extent possible. The number of epochs may be as low as ten or high as 1000 and more. A learning curve can be plotted with the data on the number of times and the number of epochs. This is plotted with epochs along the x-axis as time and skill of the model on the y-axis. The plotted curve can provide insights into whether the given model is under-learned, over-learned, or a correct fit to the training dataset.

An epoch consists of passing a dataset through the algorithm completely. Each Epoch consists of many weight update steps. To optimize the learning process, gradient descent is used, which is an iterative process. It improves the internal model parameters over many steps and not at once. Hence the dataset is passed through the algorithm multiple times so that it can update the weights over the different steps to optimize learning.

4. Test Cases

Test Case 1: Train model with 30 epoch In test case 1 we perform operation on 30 times of single image the training accuracy of 1st epoch is 0.3253 and valid accuracy of 1 st epoch is 0.4068 so our model is not that much good but when we process done in till 30th epoch the training accuracy was 0.8951 and valid accuracy is 0.9224 the result is model was perform good and its give the accuracy of 86% accurate .

Moreover, the accuracy of a model is not solely dependent on the number of epochs it has been trained for. While increasing the number of epochs can sometimes lead to improved accuracy, it can also cause overfitting, which results in a model that performs well on the training set but poorly on the test set ,but the different tests carried out to check or increase the accuracy of the model

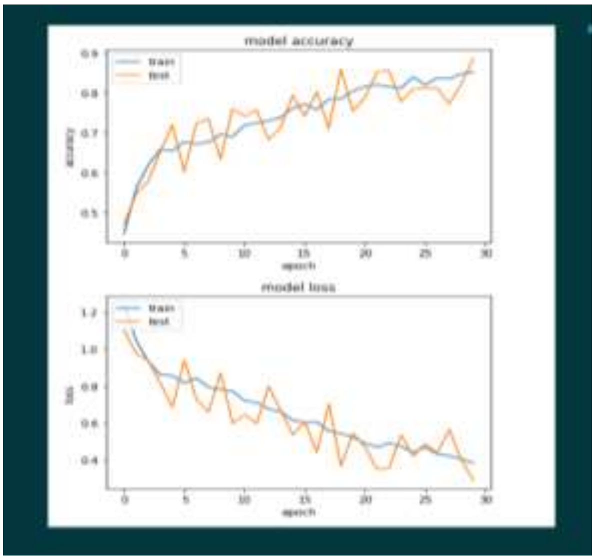

Fig: 6.1.1 model with 30 epoch

Test Case 2: Train model with 50 epoch In test case 2 we perform operation on 50 times of single image the training accuracy of 1st epoch is 0.4239 and valid accuracy of 1 st epoch is 0.4566 so our model is finding more details about the features so the accuracies of similarity decreases as the epoch values increases .Here is the graph

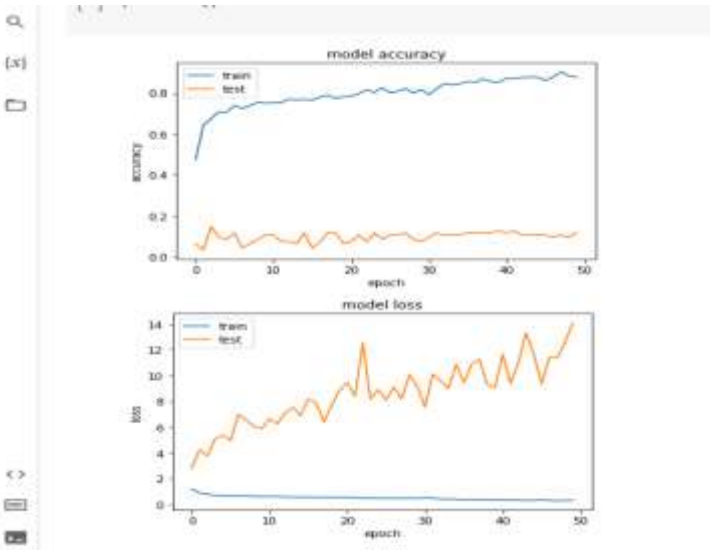

Fig: 6.1.2 with 50 epoch

VII. CONCLUSION

In conclusion, cotton crop disease prediction is a critical area of research that can provide numerous benefits to farmers, the cotton industry, and society as a whole. Disease prediction models can help farmers optimize their crop management practices, reduce economic losses, improve resource efficiency, promote better pest management, improve food security, and reduce the use of chemical inputs. By using disease prediction models, farmers can save time and reduce the efforts required to monitor their crops, while also improving the accuracy of disease diagnosis and taking proactive measures to prevent or control disease outbreaks.

The future scope of cotton crop disease prediction is promising, with several potential avenues for further research and development. By integrating multiple data sources, using machine learning techniques, developing disease forecasting models, creating mobile applications, and promoting collaboration between researchers and farmers, disease prediction models can become even more accurate and useful for farmers.

VIII. REFERENCES

[1] Pranitha P. Gulve, Sharayu S. Tambe, Madhu A. Pandey, Mrs S.S.Kanse, “Leaf Disease Detection of Cotton Plant Using Image Processing Technique”, 2015

[2] A. jenifa, R. Ramalakshmi, V. Ramachandran,” Classification of cotton leaf Disease using Multi-Support Vector Machine”, 2019

[3] Rakesh Chaware, Rohit Karpe, Prithvi Pakhale, Prof. SmitaDesai, "Detection and Recognition of Leaf Disease Using Image Processing", (2017)