Turkish Journal of Engineering
https://dergipark.org.tr/en/pub/tuje
e-ISSN 2587-1366

# Detection of cotton leaf disease with machine learning model

**Unain Hyder** [1], **Mir Rahib Hussain Talpur** [*2]

[1] Sindh Madrassatul Islam University, Department of Computer Science, Karachi, Pakistan, unainansariansari@gmail.com
[2] Sindh Agriculture University, Department of Information Technology Centre, Tandojam, Pakistan, rahibtalpur@gmail.com

Cite this study: Hyder, U., & Talpur, M. R. H. (2024). Detection of cotton leaf disease with machine learning model. Turkish Journal of Engineering, 8 (2), 380-393

https://doi.org/10.31127/tuje.1406755

**Keywords**
Machine learning
Convolutional neural network
Diseases
Artificial neural networks

**Research Article**

Received: 19.12.2023
Revised: 04.01.2024
Accepted: 23.01.2024
Published: 18.04.2024

**Abstract**

This study aims to use a machine learning (ML) model to accurately classify four datasets of cotton crop leaves as either infected or healthy. Bacterial blight, Curly virus, Fussarium Wilt, and healthy leaves were used as the datasets for the study. ML is a useful tool in detecting cotton leaf diseases and can minimize the rate of disease. The problem is that without machine learning technique it is very difficult and time consuming to detect the diseases then to sort out this problem a machine learning model is proposed and to test the accuracy of the proposed model, the confusion matrix concept was used. The researchers have done their research works to diagnose the diseases by using (ML) model but the drawback of their research was that the results which were given by the different (ML) models were not accurate. The target of the study was to identify diseases affecting the cotton plant in the early stages using traditional techniques. However, utilizing various image processing techniques and machine learning algorithms, including a convolutional neural network, proved to be helpful in diagnosing the diseases. This technological approach can simplify the detection of damaged leaves and minimize the efforts of farmers in detecting those diseases. Cotton is a natural fiber produced on a large scale, and it is grown on 2.5% of overall agronomic land. The detection of cotton leaf diseases is crucial to maintain the crop's productivity and provide reliable earnings to farmers. A confusion matrix is N X N matrix used for evaluating the performance of a classification model, where N is the number of target classes. The matrix compares the actual target values with those predicted by machine learning model. This technique has four parameters to test the accuracy of the results which is given in my research work.

## 1. Introduction

### 1.1 Machine learning (ML)

The machine learning (ML) is a field of inquiry developed to understand methods a machine learning is closely related to computational statistics, which focuses on making predictions using computers, but not all machine learning is a statistical learning. machine learning algorithms build a model based on sample data, known as training data, in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as in medicine, email filtering, speech recognition, computer vision and agriculture. Many researchers have used a machine learning concept to detect the cotton leaf diseases but as I have used in my research work to detect the diseases in the cotton leave and show its accurate results which are tested through the confusion matrix method by using its four parameters. Then the accuracy which is given by the proposed model is validated [1].

### 1.2 Cotton crop

Cotton (*Gossypium hirsutum* L.) is a widely cultivated crop that is frequently affected by pests and diseases. The management of cotton pests and diseases usually involves the use of chemical pesticides. Although chemical pesticides are often effective, their long-term use can lead to increase pest resistance to the pesticides, reduce natural enemies, decrease natural control, and environmental degradation. Due to the growing awareness of the importance of ecological sustainability and the need for sustainable cotton production, the use of organic methods such as intercropping is gradually gaining attention as an alternative approach for cotton pest and disease control. Intercropping of cotton with other crops can often improve the overall yield, provide additional income from the intercropped crop, and offer significant economic benefits without compromising cotton quality [2]. The authors used this approach for the detection of cotton leaf diseases.

Agriculture makes a significant contribution to Pakistan's economy, making up around 20% of GDP (Gross Domestic Product) and 65% of the country's exports, 73% of which are cotton-based goods [3]. However, due to rising introduction costs and the harm caused by cotton diseases, cotton production has decreased during the last few years [4] reported that the four datasets of diseases in their research work. The researchers must make decisions in order to maintain the cotton harvest. In this way, our objective proposed a method for identifying bug-based diseases in cotton crops through in-depth learning [5]. The identification of the cotton leaf diseases through the naked eyes creates a blunder that's why machine learning technique is very useful to identify the cotton diseases. Researchers demand a convenient and affordable method to identify diseases in cotton crops due to the remarkable production of cotton [6]. They stated that the four different diseases are differentiated through the machine learning technique a tensor flow framework is used to identify the cotton leaf illnesses. However, the strategies used in previous exams, such as SVM (Support Vector Machine) and SVM-GA (Support Vector machine - Genetic algorithm), suggest that grouping is a low-speed process, necessitates large computational devices, and is constrained for extremely high numbers of diseases, which increases the overall rate of the framework. According to the tools used, the region of cotton diseases is finished with considerably less calculation and execution. The customer may write the supplied method using one's skill. For example, the given method/method may be written in Python, C++, JavaScript (JS), and so forth. The state's economy is being improved by increasing cotton production and making sure that any problems are caught early on and controlled. This requires that the project's progress be monitored at various times as the day's temperature and the sun's position change [7].

### 1.3 Background of cotton crop

Florida in 1556 and Virginia in 1607 are said to have been the first two states to grow cotton crop seeds. Near the middle Nile Basin, where cotton is grown, in eastern Sudan circa 5000 BCE, the cotton crop may have also been domesticated. Since there are so many indications and symptoms of illnesses that may be detected in cotton, it is a fibre crop. And it is by far the most well-known cash crop that has a variety of effects on the Pakistan financial system [8]. These types offer the majority of yield on a worldwide scale. Vicinity wise so if we need to discuss about Pakistan so it's far 1/3 region production smart. So, in recent times many pathogens illnesses arises in a Cotton crop. So, its product should be maintained through the ML model. This crop is very profitable. The Cotton crop is very absorbent and can easily absorb the moisture from the skin of the leaves.

The Figure 1 is taken from the [9] that shows the production rate of the cotton crop of the countries.

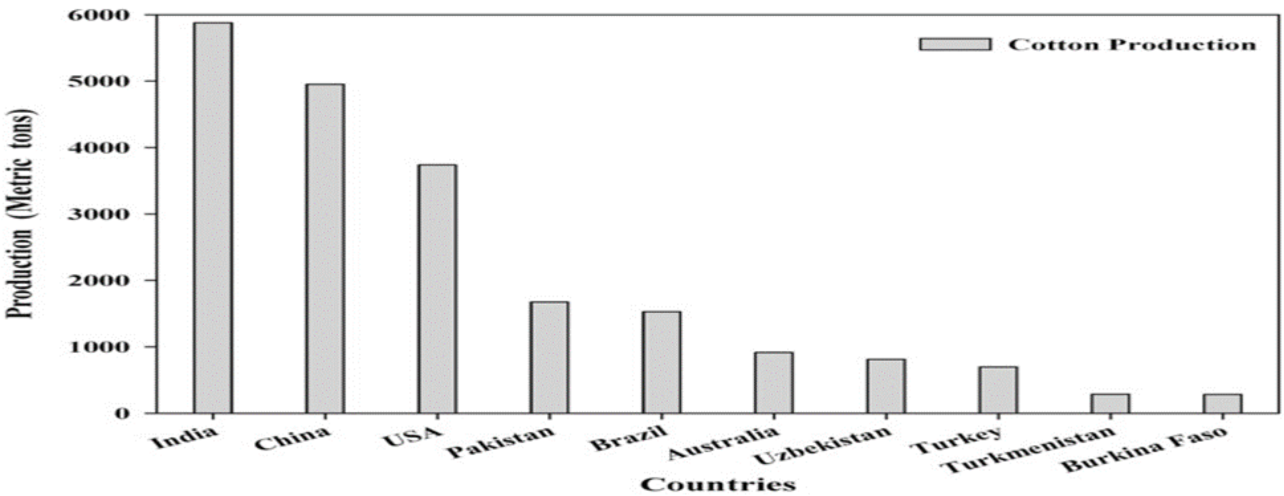



**Figure 1.** Cotton productions by the various countries ("cotton production," 2017).

Cotton is fiber crop in which such plenty of symptoms and symptoms of illnesses arises in it so we are capable of going to detect it [10]. And it is a most famous coins crop that outcomes India's financial device in lots of methods. These varieties yield all over the worldwide. Vicinity smart so if we want to talk about approximately Pakistan so it's far 1/3 location production smart. So, these days many pathogens ailments arise in a Cotton crop [11].

Many studies have been conducted in the past few years to categorize cotton leaf illnesses, and now researchers are attempting to find ways to identify the diseases using ITbased machine learning approaches. Various studies were conducted in 2015 to categorize cotton leaf illnesses utilizing various pattern recognition techniques for classification and identification of cotton leaf diseases. The horticultural area in which the production level of the cotton crop is identified by using different technological strategies [12]. For the detection and categorization of three cotton leaf diseases—Alternaria, Myrothecium, and Bacterial blight—Rothe PR and Kshirsagar RV presented a pattern recognition method. This cotton dataset was gathered from cotton fields in their natural state using two separate Cannon digital cameras, the A460 and EOS550D. Preprocessing was done using a Gaussian filter to reduce noise in the pictures, and image segmentation was done using an active contour model. Image preprocessing and segmentation were followed by the extraction of several characteristics [13].

The deep convolutional neural network (CNN) is being used in this research to distinguish between diseased and healthy leaves as well as to identify sick plants. The CNN model is likewise intended to distinguish between healthy and sick leaves. Photos are also used to train the model, and the input leaf determines the model's output. In order to classify plants and detect plant diseases and illnesses, convolutional neural networks (CNNs) were crucial [14]. Without them, farmers would have to spend a lot of money and effort on the labor intensive and time-consuming task of identifying plant diseases. So, the main benefit of the model is to minimize the time identify and classify the four datasets of the cotton leaf diseases using convolutional neural network in a machine learning model.

The Figure 2 which is taken from the [15] shows the artificial neural network.

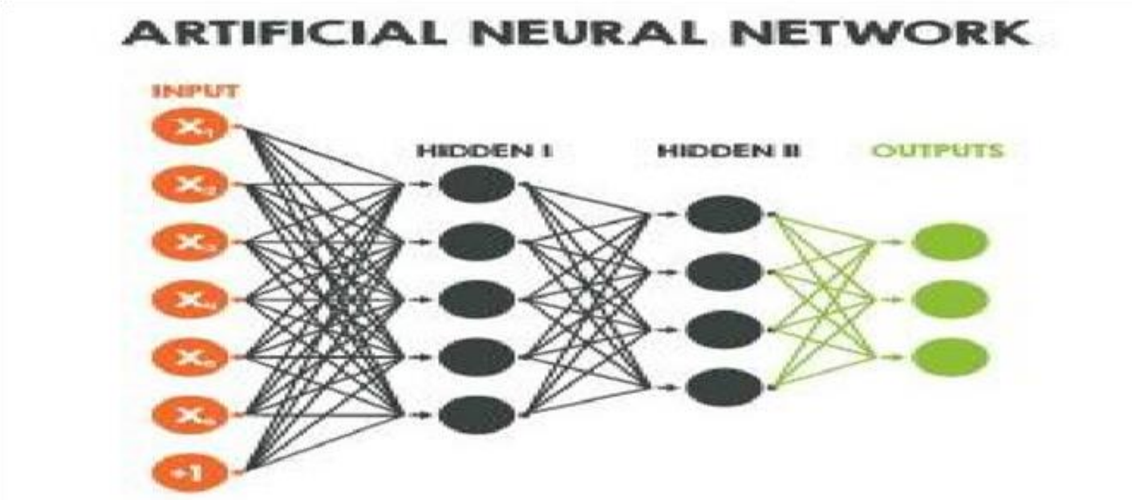



**Figure 2.** Artificial neural networks.

It recognizes through the image in machine learning model and the best classification of the diseases are also done in machine learning model to view the harmful diseases in which we have four diseases which we have already discussed and all are identified by the ML model and it mention its accurate results to the researchers. So, by using the machine learning concept the IT specialist make it easy to detect the diseases in a cotton leaf.

Figure 3 is taken from [16] and some have been taken in a real-world condition.

All these four dataset samples which are given above are the difference between the leaves which are infected or diseased and healthy leaves these all have done through the machine learning model the accuracy rate of the cotton leaves decide that which leaf is how much infected the dark spots shows the infected parts of the leaves.

**Figure 3.** Gallery of four dataset samples which detect from machine learning model (Healthy vs. Infected leaf).

### 1.4. Proposed system to show the viability of the ML model

Pham et al. [17] proposed a system to improve the viability of the Deep Learning model. The dataset is also created using a variety of models, including the Meta Deep Learning model that we have presented as well as bespoke CNN, VGG16, ResNet50, and other models. We discovered several different cotton illnesses in our research, including leaf spot, verticillium shrivel, target spot, fine accumulation, supplement deficiency, and leaf twist. With a precision of 98.53%, the suggested model performed quite well on the cotton dataset. Unmistakable models advance various element extraction techniques in deep learning. The ability to identify leaf diseases is a crucial test that has to be addressed. The suggested methodology will provide a method for identifying diseases in enormously big crops using. The suggested model will provide a mechanism for identifying diseases in tremendously important crops by using a speculating method. The suggested model makes use of a Meta Deep Learning approach [18]. This method takes into account infection identification over different harvests. A compiled technique is wise for effective decision. The suggested model's organization on mobile devices is one of its limitations. Later, the model may also be made better in terms of managing low goal images and by making the model smaller. This will make it easier to submit the model on mobile devices.

### 1.5 Researchers who used framework to detect the cotton leaf diseases

Mhatre and Lanke [19] proposed an online framework which has been effectively carried out for crop illness recognition for cotton leaves utilizing a Convolutional Neural Network. The Convolutional brain network has been created with three secret layers to order the cotton leaf infection pictures. The System effectively processes input from the client and gives yield as illness distinguished. Given adequate information is accessible to preparing, profound learning methods are equipped for perceiving plant leaf illnesses with high precision [20]. The significance of gaining huge datasets with high fluctuation, information expansion, move learning, and perception of CNN enactment maps in further developing grouping precision. India is a country with a rich history of horticulture. Thusly, our work would assist ranchers with forestalling plant illnesses, increment efficiency and benefit. Our future work focuses on a superior informational collection with countless qualities and furthermore execution of yield expectations, preventive activities, and restorative activities, pesticides required and plausible expense for proposed pesticides [21]. This framework can be stretched out to some other yield having the accessibility of enough huge datasets for that harvest. Various different infections can be incorporated for recognition. The System likewise can execute equipment involving IoT for Image catching in fields. The Web connection point can likewise include a gathering for formers to have conversations with respect to the latest things they are looking in changed illnesses.

Ramacharan [22] invented a mechanized sickness discovery framework provides the researcher with a fast and precise finding of the plant illness, permitting the demonstrative cycle to be accelerated, so the researcher can get more harvests out of his fields. Thus, it is vital to make the sickness discovery framework mechanized to accelerate crop analysis. - is paper discusses how to utilize AI and picture handling to sort out whether or not leaves are wiped out. As a beginning stage, this structure can be utilized with an image of a leaf. Most importantly, leaf photographs are tidied up to eliminate any commotion from them. To dispose of clamor, the mean channel is utilized. Division is the demonstration of separating a solitary picture into parts or sections [23]. It can assist you with sorting out how enormous the image is. - e K-implies calculation is utilized to separate the picture into parts. - e head part investigation is utilized to track down highlights. In the subsequent stage, pictures are ordered in light of their substance with assistance from calculations like RBF-SVM, SVM, irregular timberland, and ID3. RBF-SVM performs better in exact leaf sickness location.

### 1.6 Results taken via convolutional neural network by different researchers

Annabel et al. [24] proposed a network which gave us the essential step towards AI by producing a model. Through mirroring neuron communication inside the body, analysts around quite a while back were genuinely ready to overcome something that had never been finished. Before brain nets, there have been a couple, if by any means, models that were really prepared on how our body learned. During this paper we will generally gift a survey of the usage of brain network models inside the field of plant infection discovery. The writing shows that tone, surface and morphological attributes are the most appropriate to the ID and arrangement of sicknesses in plants. Counterfeit brain organizations (ANN) and Convolutional brain organization (CNN) are the most normally utilized brain network models. Programmed discovery of plant sicknesses would settle the question of expensive space proficient. Location of plant illnesses in beginning phase would work with ranchers to help the harvest yield that effectively works on nation's GDP. Future examination can embrace an assessment of the capacity of the calculation rule to analyze the reason for the injury (what nuisance or infection). Besides, the arranged calculation will be upheld with the use of a product which can be used all through real field visits to work with the formation of guides of the degree of pervasion by nuisances and illnesses.

### 1.7 Detection of the diseases by picture strategy done by different researchers

Saha and Nachappa [25] calculated a picture division strategy that may be used for later grouping and programmed identification of plant leaf infections, this research reviews numerous sicknesses organization approaches used for plant leaf sickness discovery. Some of those 10 species on which the suggested computation is tested are bananas, beans, jackfruit, lemons, mangoes, potatoes, tomatoes, and sapota. These ailments were used as differentiating evidence for these plants. The best results were obtained with a remarkably little amount of computing work, demonstrating the effectiveness of the suggested calculation in classifying and recognizing leaf disorders. Utilizing this method has added the benefit of allowing for the early or underlying detection of plant ailments. Artificial neural networks, Bayes classifiers, fuzzy logic, and mixture computations may all be used to further improve the acknowledgement rate in order processes.

Patil et al. [26] stated that the farmers in India rely heavily on agriculture for their income. Depending on where they are, farmers cultivate numerous seasonal local crops. In India, cotton is the most widely grown crop. Cotton is a crop that can be sold, and farmers get a lot of money from it. The farmer will earn more money as a result of this. However, cotton's vulnerability to a wide range of diseases is one of its fundamental flaws. To prevent production loss, these diseases need to be identified as soon as possible. The prediction model is created using the TensorFlow Keras API and the CNN algorithm in this paper.

Singh and Misra [27] stated that cotton is one of the financially critical rural items in Ethiopia; however, it is presented to various requirements in the leaf region. For the most part, these requirements are recognized as illnesses and irritations that are difficult to distinguish with exposed eyes. 'Is a study that uses CNN, a deep learning technique, to create a model that can better detect pests and cotton leaf disease. The researchers have done this by utilizing pests and common cotton leaf diseases like bacterial blight, spider mite, and leaf miner. The K-fold cross validation strategy was used to split the dataset and improve the CNN model's generalizability. For the purpose of training, nearly 2400 specimens (600 images per class) were accessed for this study. Python version 3.7.3 is used to implement the developed model, and Jupyter, TensorFlow-backed, and the deep learning package Keras serve as the development environment. The model was able to identify classes of cotton plant pests and disease with an accuracy of 96.4%. It is shown that its use in real-time applications is possible and that IT-based solutions may be needed to support manual or traditional disease and pest identification.

Tripathy [28] proposed one of the most important cash crops in India is cotton, which is grown in large quantities by many farmers. In the past few decades, cotton diseases ought to have resulted in a significant reduction in productivity. It's important to diagnose cotton diseases early. This study aims to present a method for automatically diagnosing cotton leaf diseases that makes use of convolutional neural networks. Separation based on selecting appropriate Deep Learning-create image features like color and texture. [29] examined to make a framework which will actually want to analyze the sickness present on the leaf of the cotton plant. Since identifying the sickness present on cotton crop with naked eyes could bring a few blunders, so doing likewise with the assistance of machine will decrease the possibilities of misprediction. Wang et al. [30] addressed that plant sickness conclusion framework helps the rancher's general public and agribusiness industry to build the harvest creation limit by shielding their yields from a few illnesses tracked down in their homesteads by recognizing the side effects of infection in its beginning phase and finding a suitable way to control them. The model is intended to recognize the side effects of cotton plant illness. The model effectively pieces the impacted piece of the pictures of leaf tests utilizing thresholding strategies. Zhou et al. [31] summed up and audits various strategies in light of administered, solo, and half breed for multi-crop leaves illness location and characterization. The various leaves sickness identification approaches for multi-crops have been proposed in the space of the horticultural business.

## 2. Material and Method

### 2.1. Dataset

The dataset comprises of four classes 1. Bacterial blight, 2. Curly virus, 3. Fussarium wilt and healthy leaves of cotton, the dataset contains 1710 images those were captured under real world conditions and from internet (Figure 4).

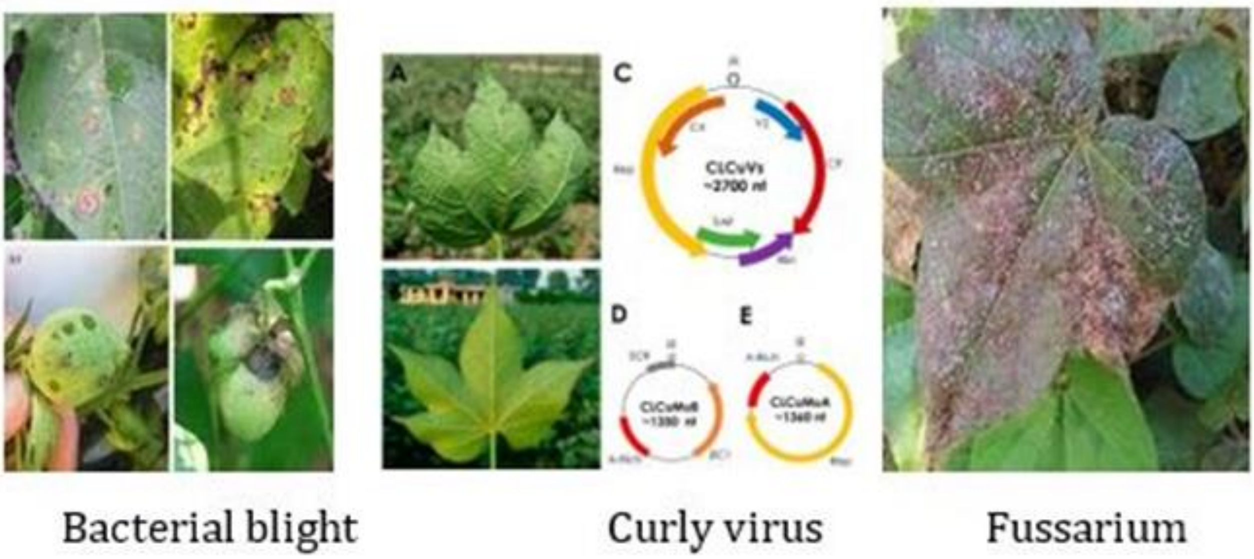



**Figure 4**. Datasets of diseases.

The images were taken from the [8] in which he shows the two datasets of the diseases in a cotton crop leaves.

The pictures in Figure 5 show the two categories of the diseases in cotton crop leaves which are identified through the machine learning model. The results through this model are clearly mentioned the darkspots shows the infected parts of the cotton leaves [8]. They also proposed a machine learning model for cotton leaf disease detection [9].

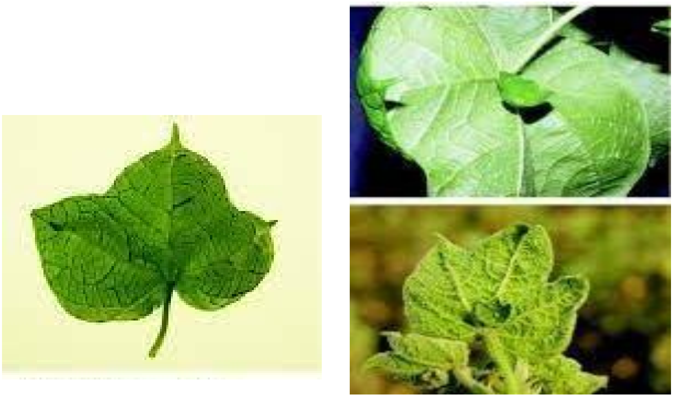
The Figure 6 categories two samples of the leaves which has been taken in the real-world condition.

Some points of the contribution are given below:

1. The numerous databases which have been used to behavior those studies also are mentioned.

2. The comparative analysis of the present strategies is done to evaluate their blessings and drawbacks.
3. The several cotton diseases are labeled and their signs and symptoms also are provided.
4. The numerous performance metrics that can be used to assess the performance of the disorder prediction are also presented.

**Bacterial blight**



**Leaf curly virus**

**Figure 5.** Bacterial blight and curly virus**.**

(a)

(b)

**Figure 6.** Fussarium wilt (a) and healthy leaves (b).

## 2.2. Dataset

A machine learning model has been trained using a cotton disease dataset and has implemented in an android application that has been developed in python. In the android application an image of a cotton leaf will be provided by the end user and it has been predicted by the model whether the leaf is healthy or it contains some diseased based data. The model has been able to detect only three cotton disease that is 1. Bacterial blight 2. Curly virus, 3. Fussarium wilt 4. Healthy leaves of cotton finally the result of the cotton leaf has been displayed with the solution on the android-based application

The Figure 7 which is design to shows the process of the detection of the diseases.

With the aid of Tensor Flow, machine learning is used to develop high-accuracy models for applications that deal with pictures, text, video, or audio. Google has provided a platform for working with deep learning called Tensor Flow. a deep learning subset of machine learning that use CNN for large-scale data processing. Tensor Flow is used by the developers to swiftly launch the apps. Our methodology makes advantage of transfer learning. Transfer learning refers to the process of retraining a model on our data after it has already been trained on a different issue. We don't have to waste time or effort starting over as a result. On the ImageNet dataset, the Deep Neural Network inception-V3 pre-Trained model is employed. In 1000 distinct classrooms, ImageNet is taught alongside Dalmatian and dishwater. Run inference on the pre-trained model to comprehend its input and SoftMax layer before retraining an Inception V3. The Inception of data-file contains this picture of a panda. With a classification score of nearly 89% and the next-highest score being just about 0.8% for an indri, another exotic mammal, the Inception model is pretty certain that this image is a panda. A "SoftMax-function" is what the Inception model produces as its output. Because the SoftMax outputs add to one and range from zero to one, they are sometimes referred to as probabilities rather than possibilities. However, since they are not the results of repeated tests, they are not probabilities in the conventional meaning of the word. Because they show how strongly the neural network thinks the input picture belongs to each potential class, it may be preferable to refer to the output values of a neural network as classification scores or rankings. INCEPTION-V3 INFERENCE: In contrast to training, inference does not reevaluate or modify the neural network's layer structure in light of the findings. Inference takes information from a trained neural network model to infer a conclusion. As a result, when a fresh batch of unknown data is fed into a trained neural network, a forecast is produced depending on the neural network's level of predictive accuracy. Due to the need for a learned neural network model, inference occurs after training. However, the trained model is still on our local host (laptop), thus it must be deployed on a cloud server before it can be used for client inference. This is when platforms like Heroku come into play. RETRAINED INCEPTION-V3 RUNNING INFERENCE: The next phase will be used to do a straightforward inference on a retrained model that is locally accessible (on a laptop).

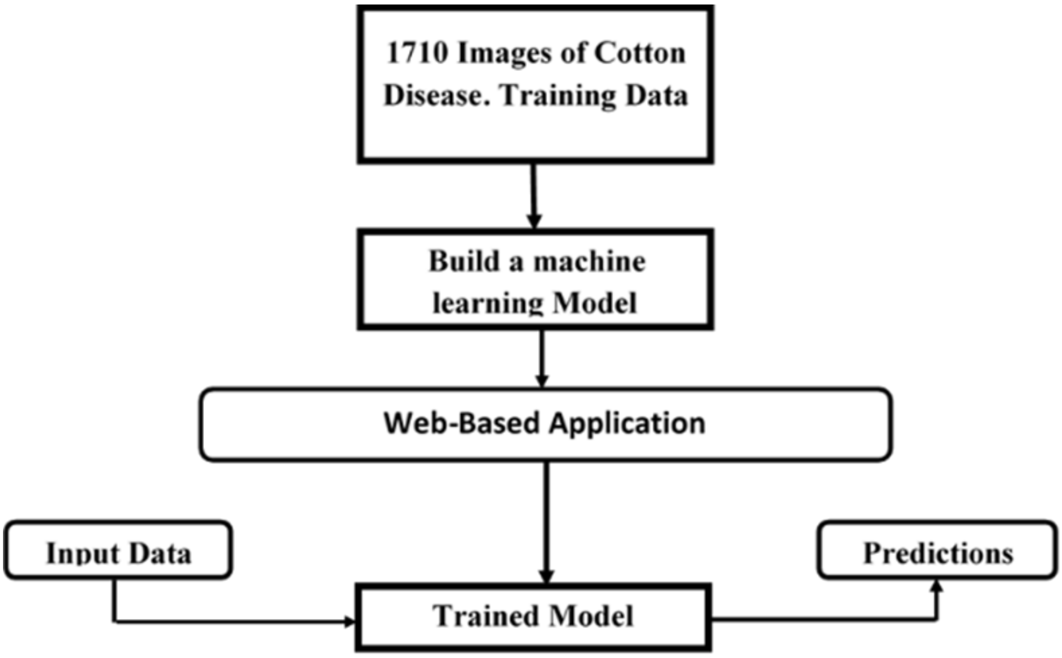



**Figure 7.** The process to detect the diseases effects on cotton leaves.

### 2.3 Machine learning model

The machine learning (ML) which shows the main function the machine learning model is main purpose tool which is used to detect diseases or to predict diseases now we will discuss how it works first of all we capture the pictures in a real world and then put it in a machine learning model to detect the diseases to differentiate the datasets of different diseases as I have already define that the four datasets which we have put in this research work now it's time to detect them with the help of this ML model. This is the main usage of machine learning in disease detection. So, it is minimizing the time of farmers the cotton crop is very healthy and their results through this ML model is accurate. The process of ML model in which the images of cotton crops mean Training data of the images of the cotton crop are defined that some fields shows that the leaves are diseased and some are healthy so the model easily identified this main difference between these two categories. I have two data sets one will be the input data and the other will be the output data of the leaves so it mentioned that the AI played the vital role to prove these two categories. The machine learning model is the tool that enables the system to automatically carry out an action based on the machine learning's experience and does not require explicit programming. The main idea behind machine learning is to let computers learn on their own without human intervention, which simplifies the process. Now, the machine learning model, also known as the ML model, determines what should be done when a particular pattern in a dataset shows up. The creation of computer programs that can access data and a set of algorithms that utilize the data to determine for themselves what action should be taken based on that data, for which we will give this model, is known as machine learning (Figure 8).

### 2.4 Tensor flow framework

The framework (Figure 9) which is used here is Tensor Flow. As we have already discussed above the Tensor Flow is provided by the Google to work with the deep learning. The Tensor Flow is used to flow the process of the detection of diseases now the Tensor Flow framework is very useful in machine learning model. So, we will point out this Framework in our research work. The data of the leaves of the infected and healthy are analyzed by this Framework in a machine learning model. Now the questions arise here shows that how the data is processing in a machine learning model the model uses this framework to process this data by capturing the pictures and put it in a model then by using this framework the model shows the final results. Model tracking, data automation, performance monitoring, and model retraining are all done with or built using the tensor flow platform. Success depends on using production tools to monitor the model training, service, or business process. Developers may now use the framework to build dataflow graphs, which are structures that show how data flows via a network of processing nodes. The connections between the nodes of a network now represent multidimensional data arrays or sensors, and each node itself represents a mathematical action. In the other words tensor flow is the end-to-end encrypted platform which is very secure now data will be hacked out of this platform so it is very beneficial to improve the quality to detect the crop leaves diseases so in this research work it very beneficial to use this work to analyze the best results of the diseased leaves of the cotton crop. The TensorFlow is the open-source software which is used in machine learning and artificial intelligence it is also used across the range of tasks but has a particular neural network. Now the TensorFlow is the helpful to support traditional machine learning. It is also used as an ecosystem of tools, community resources and libraries which help to develop easily build ML applications and the installation of the TensorFlow is directly through pip.python jupyter-notebook.

The link to install the tensor flow is given below and Figure 10 is taken from google.

The work flow of the ML model in Figure 11 is to identified the diseases which are diagnosed so with the help of disease detection the farmers can be easily avoid from the field losses, time consuming or through the detection of these diseases the farmers can get a quantitative crop so it's a main objective of the proposed model is to get the quantitative crop by the diagnosing and removing of the diseases through this ML model.

Figure 12 shows the model accuracy through different ML models custom CNN, VGC16, ResNet50 and proposed model.

The different images of the cotton leaves in Figure 13 which shows the different categories of the diseases which are mentioned in the above pictures these infected leaves are predicted by the technological based concept which is using machine learning algorithms.

The Figure 14 shows the three categories of the dataset which defined that the images of the background soil, lesioned leaf and healthy leaf as well it is a best technique to identify these three categories of the images of the soil which are defined by the machine learning model whose process which we have already defined so by the machine learning farmers get a lot of benefits and they are taking are quantitative cotton crop after the detection of the leaf disease.so the pictures shows the clear differences between these images. In the (b) image category the dark spot shows the infected leaf of the cotton crop.

**Figure 8.** The process of Machine learning model (ML).

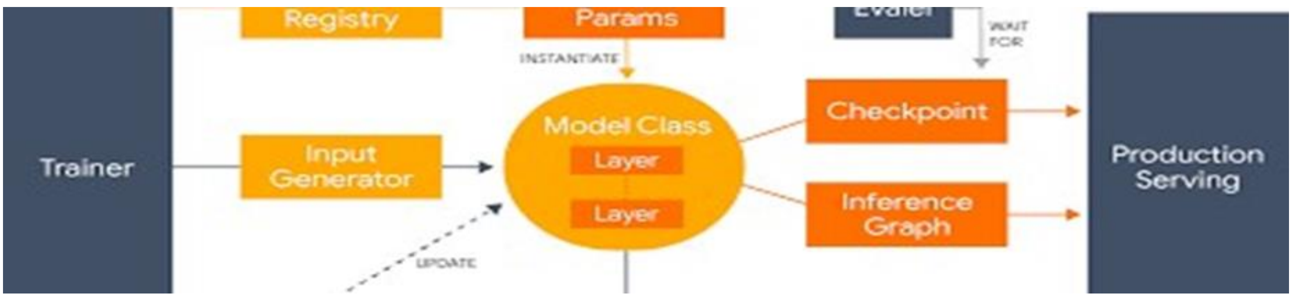



**Figure 9.** The picture shows the Tensor flow framework.

**Figure 10.** The logo of Tensor flow framework.

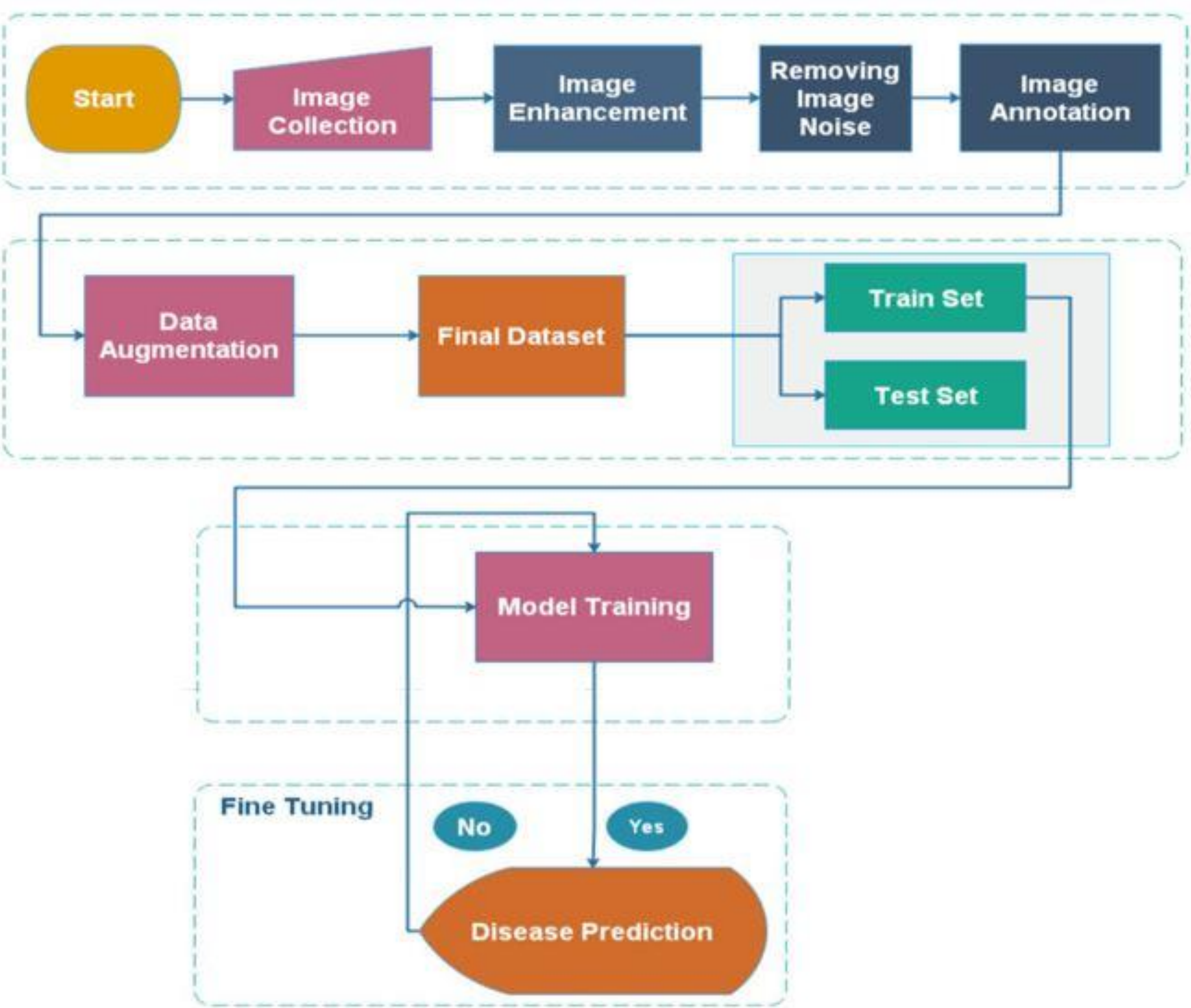



**Figure 11.** The workflow of the model for the leaf disease identification.

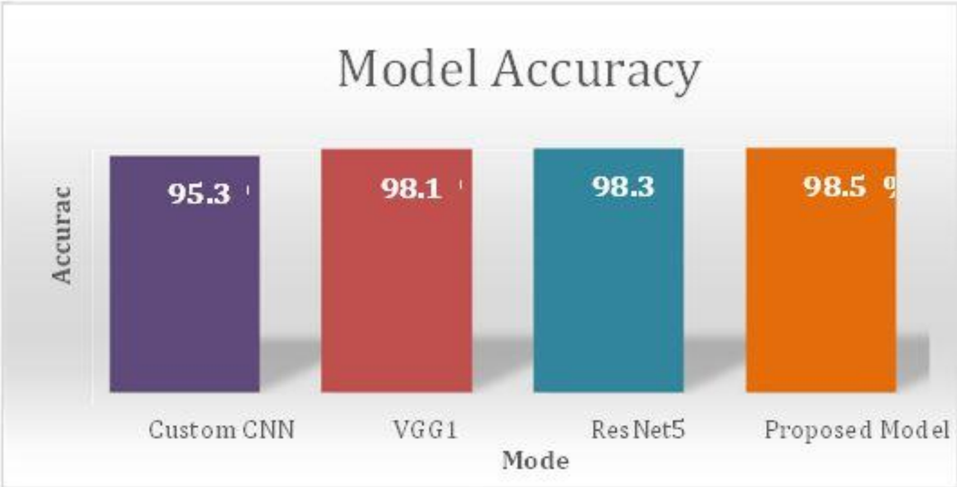



**Figure 12.** The model accuracy through different ML models custom CNN, VGC16, ResNet50 and proposed model.

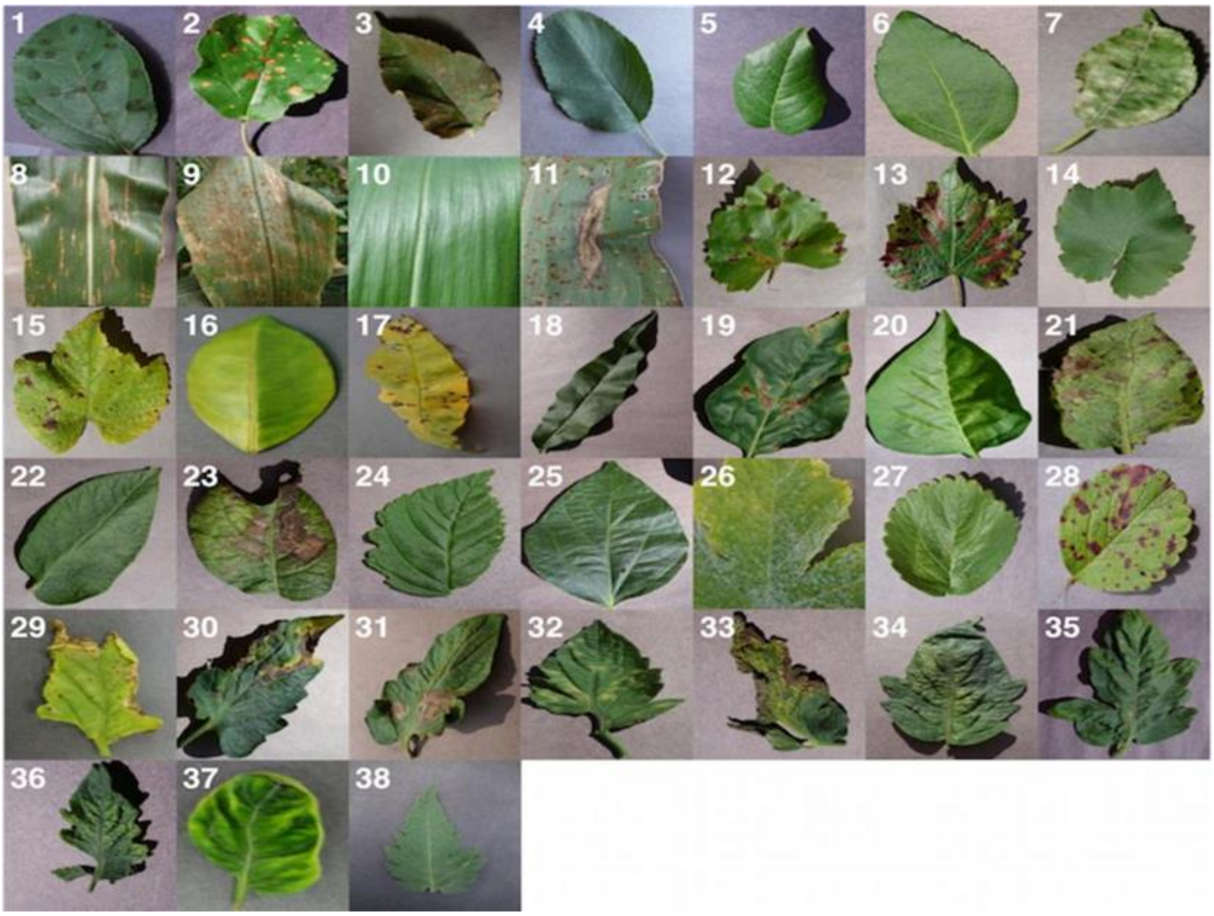


**Figure 13.** The different images of the cotton leave using ML Model.

(a) (b) (c)

**Figure 14.** Images of the background soil (a), lesioned leaf (b) and healthy leaf (c).

## 3. Results

Epochs are single iterations of the data through the model during training. The same data would be periodically cycled across several epochs after each one. When this procedure is repeated, the model is said to be learning. Remember what we mentioned at the conclusion of the network when we indicated that the model is initially initialized with its set of arbitrary weights? To get strong performance on illness classification, the convolutional neural network must be tuned for a few key hyperparameters. By adjusting the specifics of the learning process, outcomes can be improved. STEPS IN TRAINING: The longer a model is trained, the slower the rate of accuracy increase becomes, and eventually it will stop entirely (or possibly decline owing to overfitting). We ran three tests with 500, 1500, and 1000 participants to see which was most effective for our model. HYPER-PARAMETERS: The model may be tweaked with more parameters to see whether the outcomes are improved. The value of the weights that are given to the final layer to be updated during training is controlled by the learning rate hyperparameter. Although it is sense that a lesser learning rate would need more time during training, it may actually improve overall accuracy. But since that isn't always the case, it must be carefully chosen to examine what functions in our situation, where we have found it to be 0.01. As the name implies, the train batch size determines the number of pictures that are weighed during each training phase in order to determine the changes to the final layer. We tried 5 and 10 photographs in our experiments, and it turns out those 10 images per batch work nicely with our model. SETS FOR TESTING, VALIDATION, and AND TRAINING: Three independent data sets—a training, validation, and test set—should be created from the training data. The biggest dataset is often the training set, which includes all of the photo's input into the ML model.

### 3.1 Tested results of different ML models.

Convolutional neural network is trained, and the weights of the model are updated based on the outcomes. A training set is a collection of data that is used to train a model at each iteration (iteration). The model will continue to learn more about the features by being trained again using the same data from the training set. The validation set is different from the training set. In order to prevent model overfitting, it is utilized to verify the model during training. This validation method provides data that might be used to modify the model hyper parameters. The data in the validation set will be validated while the model is being trained on the training set. TESTING Collection: This set of data is used to test the model once it has been successfully retrained. This set is different from the validation and training sets.

The pictures in Figure 15 were taken from the real-world condition that shows the results of the infected and healthy leaves of cotton crop whose accuracy is tested through different (ML) model.

Figure 16 shows the diseased rate detection of the cotton crop leaves using the ML model. The ML model or the machine learning model which is used here is Convolutional neural network (CNN) model which detect the results and shows in the Table 1-4 and in the Figure 15 form which is a particular most commonly used model in machine learning.

**Table 1.** Results of retraining model (Accuracy Graph).

| Parameters | 1st Run | 2nd Run | 3rd Run |
|---|---|---|---|
| Training Steps | 500 | 1500 | 1000 |
| Learning Rate | 0.01 | 0.01 | 0.01 |
| Training Batch Size | 10 | 05 | 10 |
| Misclassified Pictures | 11 | 09 | 05 |
| Testing %age | 5 | 5 | 5 |
| Validation %age | 10 | 10 | 10 |
| Final Test Accuracy (%) | 86.3 | 88.6 | 91 |

**Table 2.** Results summary of different ML Models.

| S.No | Model | Accuracy (%) | Dataset |
|---|---|---|---|
| 1 | Custom CNN | 95.37 | Cotton |
| 2 | VGG16 | 98.10 | Cotton |
| 3 | ResNet50 | 98.32 | Cotton |
| 4 | Proposed Model | 90.53 | Cotton |

**Table 3.** Results to differentiate the diseased and healthy leaves through ML model.

| S.No | Models | Diseased leaf Accuracy (%) | Healthy leaf Accuracy (%) |
|---|---|---|---|
| 1 | Custom CNN | 4.87 | 96.87 |
| 2 | VGG16 | 13.75 | 87.85 |
| 3 | ResNet50 | 4.45 | 96.45 |
| 4 | Proposed Model | 9.64 | 90.56 |

**Table 4.** Cotton crop rate observation per year diseased.

| Disease Names | Diseased Rate per year | Disease Samples |
|---|---|---|
| Bacterial Blight | 56.8% (2020) | 15 |
| Curly Virus | 66.7% (2021) | 25 |

**Figure 15.** Infected and healthy leaves.

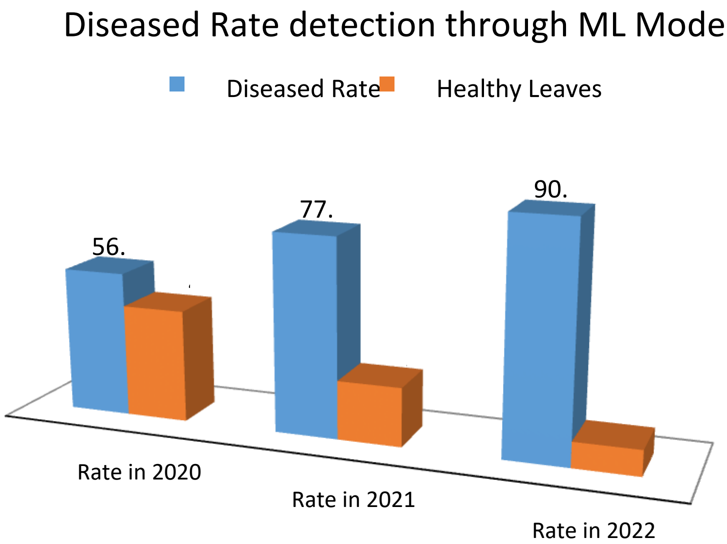



**Figure 16.** Diseased rate detection through ML model per year.

### 3.2. Confusion matrix method

A confusion matrix is a table that is used to define the performance of a classification algorithm. a confusion matrix visualizes and performance of classification algorithm. technique has four parameters to test the accuracy of the results which is given in my research work. The pattern to identify the accurate results through this technique is given in the Figure 17.

The results which are given by the confusion matrix technique are valid I have used its four parameters in my research work to detect proper results so the results through this technique are given in Table 5.

| n=165 | Predicted: NO | Predicted: YES | |
|---|---|---|---|
| Actual: NO | TN = 50 | FP = 10 | 60 |
| Actual: YES | FN = 5 | TP = 100 | 105 |
| | 55 | 110 | |

**Figure 17.** The pattern to identify the accurate results.

**Table 5.** Tested results through confusion matrix method per year.

| Year | Infected leaf (%) | Healthy leaf (%) |
|---|---|---|
| 2019 | 10.45 | 80.55 |
| 2020 | 85.50 | 15.00 |
| 2021 | 80.00 | 20.00 |
| 2022 | 75.56 | 25.00 |

## 4. Conclusion

The goal of this research was to examine and comprehend the illnesses that affect cotton crops and to develop a machine learning model for diagnosing cotton diseases. The work begins with the history and development of the cotton crop, the issues it is currently experiencing, and the goals of our challenge, which may be used to address those issues. After that, a survey of the literature was conducted to have a great idea of the most recent investigations into the identification of illnesses in cotton crops. After reviewing the literature, the thesis focuses on the methods used to create the model; it contains the task description seen via the usage of the extracts and their physical characteristics. The experimental outcome is then presented, and it is concluded that the model was properly built in this thesis work and project designing and can be easily applied to the Go platform. The accuracy ranges for unusual cotton ailments were examined after the model's creation. The readings were obtained at a level of accuracy better than 90%, indicating the success of the project. Because this version is based on machine learning, it will continue to improve accuracy with each use by analyzing each new picture that is taken using the model. On the other hand, a smartphone that is entirely built on technology may provide the option to purchase pesticides online.

## Recommendations

The population of the sector is increasing at a pace of 1.08% per 12 months. With time, the need for meals will increase. To meet the demands, we must thus obtain the majority of our produce from agricultural lands. However, global warming has grown into a significant issue that has to be addressed, and one of the causes of global warming is also agriculture. Therefore, to stop agriculture from contributing to global warming and to increase productivity. In agriculture, monitoring, diagnosing, and ethical usage of pesticides have become crucial. With the aid of system learning, Cotton Care Assignment will help in identifying the illnesses and keeping an eye on the cotton crop. However, it is also necessary to monitor various types of flora and plant life in order to improve production and reduce the usage of chemical compounds.

## Author contributions

**Unain Hyder:** Conceptualization, Literature Review, Methodology, Writing Original.
**Mir Rahib Hussain Talpur:** Machine Learning Model, Data curation, Software, Validation, Visualization, Investigation, Writing-Reviewing and Editing.

## Conflicts of interest

The authors declare no conflicts of interest.

## References


1. Azath, M., Zekiwos, M., & Bruck, A. (2021). Deep learning-based image processing for cotton leaf disease and pest diagnosis. Journal of Electrical and Computer Engineering, 2021, 1-10. https://doi.org/10.1155/2021/9981437
2. Caldeira, R. F., Santiago, W. E., & Teruel, B. (2021). Identification of cotton leaf lesions using deep learning techniques. Sensors, 21(9), 3169. https://doi.org/10.3390/s21093169
3. Chi, B. J., Zhang, D. M., & Dong, H. Z. (2021). Control of cotton pests and diseases by intercropping: a review. Journal of Integrative Agriculture, 20(12), 3089-3100. https://doi.org/10.1016/S2095-3119(20)63318-4
4. Dunne, R., Desai, D., Sadiku, R., & Jayaramudu, J. (2016). A review of natural fibres, their sustainability and automotive applications. Journal of Reinforced Plastics and Composites, 35(13), 1041-1050. https://doi.org/10.1177/0731684416633898
5. Iqbal, Z., Khan, M. A., Sharif, M., Shah, J. H., ur Rehman, M. H., & Javed, K. (2018). An automated detection and classification of citrus plant diseases using image processing techniques: A review. Computers and Electronics in Agriculture, 153, 12-32. https://doi.org/10.1016/j.compag.2018.07.032
6. Amjad, K., & Ghous, H. (2021). Critical review on multi-crops leaves disease detection using artificial intelligence methods. International Journal of Scientific & Engineering Research, 12(2), 879-912.
7. Kumar, A., & Singh, R. (2022). The Different Techniques for Detection of Plant Leaves Diseases. International Journal of Artificial Intelligence, 9(1), 1-7. https://doi.org/10.36079/lamintang.ijai-0901.342
8. Kumbhar, S., Nilawar, A., Patil, S., Mahalakshmi, B., & Nipane, M. (2019). Farmer buddy-web based cotton leaf disease detection using CNN. International Journal of Applied Engineering Research, 14(11), 2662-2666.
9. Liang, X. (2021). Few-shot cotton leaf spots disease classification based on metric learning. Plant Methods, 17, 1-11. https://doi.org/10.1186/s13007-021-00813-7
10. Lambat, R. K. M., Kothari, R., & Mane, M. K. (2022). Plant disease detection using inceptionv3. International Research Journal of Engineering and Technology (IRJET), 9(6), 2295-2300.

11. Memon, M. S., Kumar, P., & Iqbal, R. (2022). Meta deep learn leaf disease identification model for cotton crop. Computers, 11(7), 102. https://doi.org/10.3390/computers11070102
12. Noon, S. K., Amjad, M., Ali Qureshi, M., & Mannan, A. (2021). Computationally light deep learning framework to recognize cotton leaf diseases. Journal of Intelligent & Fuzzy Systems, 40(6), 12383-12398. https://doi.org/10.3233/JIFS-210516
13. Pandhare, N., Panchal, V., Mishra, S. S., & Tambe, D. (2022). Cotton plant disease detection using deep learning. International Research Journal of Modernization in Engineering Technology and Science, 4(4), 1156-1160.
14. Pechuho, N., Khan, Q., & Kalwar, S. (2020). Cotton crop disease detection using machine learning via tensorflow. Pakistan Journal of Engineering and Technology, 3(2), 126-130.
15. Pantazi, X. E., Moshou, D., & Tamouridou, A. A. (2019). Automated leaf disease detection in different crop species through image features analysis and One Class Classifiers. Computers and Electronics in Agriculture, 156, 96-104. https://doi.org/10.1016/j.compag.2018.11.005
16. Tripathi, M. K., & Maktedar, D. D. (2021). Detection of various categories of fruits and vegetables through various descriptors using machine learning techniques. International Journal of Computational Intelligence Studies, 10(1), 36-73. https://doi.org/10.1504/IJCISTUDIES.2021.113819
17. Pham, T. N., Van Tran, L., & Dao, S. V. T. (2020) Early disease classification of mango leaves using feed-forward neural network and hybrid metaheuristic feature selection. IEEE Access, 8, 189960-189973. https://doi.org/10.1109/ACCESS.2020.3031914
18. Thangaraj, R., Anandamurugan, S., & Kaliappan, V. K. (2021). Automated tomato leaf disease classification using transfer learning-based deep convolution neural network. Journal of Plant Diseases and Protection, 128(1), 73-86. https://doi.org/10.1007/s41348-020-00403-0
19. Mhatre, R., & Lanke, V. (2021). Cotton leaves disease detection and cure using deep learning. International Research Journal of Modernization in Engineering Technology and Science, 3(1), 1369-1374.
20. Ranjan, M., Weginwar, M. R., Joshi, N., & Ingole, A. B. (2015). Detection and classification of leaf disease using artificial neural network. International Journal of Technical Research and Applications, 3(3), 331-333.
21. Sarwar, R., Aslam, M., Khurshid, K. S., Ahmed, T., Martinez-Enriquez, A. M., & Waheed, T. (2021). Detection and classification of cotton leaf diseases using faster R-CNN on field condition images. Acta Scientific Agrıculture, 5(10), 29-37.
22. Ramacharan, S. (2021). A 3-stage method for disease detection of cotton plant leaf using deep learning CNN algorithm. International Journal for Research in Applied Science & Engineering Technology., 9(VII), 2503-2510.
23. Kumar, S., Ratan, R., & Desai, J. V. (2022). Cotton disease detection using tensorflow machine learning technique. Advances in Multimedia, 1812025. https://doi.org/10.1155/2022/1812025
24. Annabel, L. S. P., Annapoorani, T., & Deepalakshmi, P. (2019). Machine learning for plant leaf disease detection and classification–a review. In 2019 International Conference on Communication and Signal Processing (ICCSP), 538-542. https://doi.org/10.1109/ICCSP.2019.8698004
25. Saha, P., & Nachappa, M. N. (2022). Cotton Plant Disease Prediction Using Deep Learning. International Journal for Research in Applied Science & Engineering Technology, 10(3), 744-746. https://doi.org/10.22214/ijraset.2022.40731
26. Patil, S. V., Sharma, A. K., Kamble, B. R., & Jadhav, K. B. (2022). Cotton leaf disease detection using deep learning. International Journal of Creative Research Thoughts, 10(5), 9804-9810.
27. Singh, V., & Misra, A. K. (2017). Detection of plant leaf diseases using image segmentation and soft computing techniques. Information Processing in Agriculture, 4(1), 41-49. https://doi.org/10.1016/j.inpa.2016.10.005
28. Tripathy, S. (2021) Detection of cotton leaf disease using image processing Techniques. In Journal of Physics: Conference Series. https://doi.org/10.1088/1742-6596/2062/1/012009
29. Gupta, S., Geetha, A., Sankaran, K. S., Zamani, A. S., Ritonga, M., Raj, R., ... & Mohammed, H. S. (2022). Machine learning-and feature selection-enabled framework for accurate crop yield prediction. Journal of Food Quality, 2022, 1-7. https://doi.org/10.1155/2022/6293985
30. Wang, G., Sun, Y., & Wang, J. (2017). Automatic image-based plant disease severity estimation using deep learning. Computational Intelligence and Neuroscience, 2017, 2917536. https://doi.org/10.1155/2017/2917536
31. Zhou, C., Zhou, S., Xing, J., & Song, J. (2021). Tomato leaf disease identification by restructured deep residual dense network. IEEE Access, 9, 28822-28831. https://doi.org/10.1109/ACCESS.2021.3058947