

Disease Identification using Deep Learning in Agriculture: A Case Study of Cotton Plant

Jawad Hassan¹, Kaleem Razzaq Malik^{*1}, Ghulam Irtaza², Ali Ghulam³, Ashfaq Ahmad⁴

¹ Department of Computer Science and Engineering, Air University Multan Campus, Multan, Punjab, Pakistan

² Department of Information Sciences, University of Education Multan Campus, Multan, Punjab, Pakistan

³ Information Technology Centre, Sindh Agriculture University, Tandojam, Sindh, Pakistan

⁴ Department of Computer Science, Abdul Wali Khan University, Mardan, Khyber Pakhtunkhwa, Pakistan

*Corresponding Email: krmalik@aumc.edu.pk

ABSTRACT

Among all the agrarian products, cotton is known as “Ready Cash Crop” and it plays the significant role in the stability of the economy of a country. Therefore, it is extremely important to monitor the cotton crop from the numerous diseases. Unfortunately, sometimes human eyes not be able to analyze these diseases at earlier stage and that will affect not only the quality and also the quantity of the cotton crops. To address this early monitoring issue we proposed an interactive framework based on target feature extraction and deep learning model for cotton leaf screening to deal with these well-known dangerous diseases; Grey Mildew, Cercospora, Bacterial Blight and Alternaria. In this study we chosen our own collected dataset that contains 522 images of cotton leaves that were collected from the field (Cotton agricultural areas near the Multan city). The performance evaluation matrix indicates the algorithm secure; 85.42% overall accuracy, 0.8542 precision, 0.8542 recall, 0.854 F1 score and 0.817 kappa coefficient indicates the generalization and acceptability of the model. The proposed framework not only assists the agronomist but also the farmer because of early identification of diseases from cotton crop and to avoid from the massive loss. It make better decisions for cotton crop management and contributes in the sustainability of the economy

KEYWORDS

Cotton leaf, leaf diseases, feature extraction, CNN

JOURNAL INFO

HISTORY: Received: November 16, 2022

Accepted: December 27, 2022

Published: December 30, 2022

INTRODUCTION

Pakistan's economy is mostly dependent on agriculture. Cotton is an essential component of it. Pakistan is a major producer of cotton, a cash commodity valued across the world [9]. It provides enough natural fiber to meet the needs probably it is one third of the world's population. It is the backbone of the country's agrarian industrial economy [1]. Both Punjab and Sindh, with respective shares of 65% and 35%, are major centers of cotton cultivation in Pakistan. The country of Pakistan ranks sixth in global cotton production. Cotton is grown by around 1.5 million farmers on a total area of 2,527 hectares, and the annual output is about 1.37 million tons [9]. It was Pakistan's very first commercially successful harvest. Cotton has an essential function in our medical institutions. Cotton is used to make a wide variety of items, such as thread, fishnets, breathable fabrics, animal feed, bookbinding, edible oil, coffee filters, first aid kits, high quality paper, automobile-tops, canvas, curtains, and carpets [6].

The cotton crop faces a number of different types of diseases and pest attacks during its growth [3]. These

types of attacks play a hazardous role to reduce the quality and quantity of the cotton crop. Therefore, it is extremely important to detect these diseases at earlier stage to increase the cotton crop's production and contribute handsomely in the economy of country. Some of most dangerous diseases as reported are Grey Mildew, Cercospora, Bacterial Blight and Alternaria.

A. Grey Mildew

Grey Mildew is a type of fungal disease and it can cause significant harm to a wide variety of plants, including vegetables, flowers, and fruits. It is not difficult to determine which leaf is impacted. It is made up of a powdered material that has a bluish-gray color in appearance. It manifests itself most frequently on the upper surface of the upper leaf, while it can sometimes be seen anywhere on the plant [3]. This situation is present in the environment when the temperature is between 20 and 30 degrees Celsius, and there is a high level of humidity (at least 80%) [3].





Figure 1: Grey Mildew Disease.

B. Cercospora

Cercospora is a type of fungal disease. It can be found on both outdoor and indoor plants. It first appears as small, round brown and grey-brown to tan patches that range in size from 1 to 10 mm, and grow large enough to touch each other [2]. The areas around developed spots have dry, lifeless and grey centers that fall out and break on a regular basis. It happens when the environmental conditions are warm and wet [4].



Figure 2: Cercospora Disease

C. Bacterial Blight

Bacterial Blight is a common bacterial disease. In Europe during the 1840s, one million people died from eating potatoes that were affected by blight. However, blight affects other plants, including cotton, in addition to potatoes [1]. It is propagated by spores that are carried by the wind. These spores can easily cover a large number of areas, speedily and rapidly spreading the infection in plants. It can only spread under warm temperature and high humid conditions, especially occur when the temperature of the environment is 33-40 °c in daytime and 16-20°C in nighttime for two consecutive days and the humidity is higher than 90% for eleven hours [3].



Figure 3: Bacterial Blight Disease

D. Alternaria

Alternaria is a fungal disease that manifests itself after the seedlings have been wounded. At the base of the stem, spindle-shaped streaks of a reddish-brown color form, and these streaks are split lengthwise. On the margins of the cotyledons, brown lesions with a semicircular shape frequently develop. The majority of the time, it is connected to development later in the season [3]. Small circular patches ranging in size from one to ten millimeters in diameter form on it during periods of damp weather when the temperature is between twenty and thirty degrees Celsius [4].



Figure 4: Alternaria Blight Disease

Cotton has a relatively long growth period, about 180 days, and has the habit of unlimited growth. During the long growth process of cotton, it is very necessary to monitor the crop regularly with detail observation. For this cause in this study, we design and develop a feature extraction-based AI system that provides long-term monitoring and detect the disease from the cotton's leaf. Deep Learning [4] and Computer Vision are two examples of modern technologies that aid in the autonomous, human-free diagnosis of cotton diseases. A automated method for recognizing plant diseases [5] would be more precise, less expensive, and more efficient. Cotton leaf diseases cost the global agriculture business a

great deal of money [6]. Management of the disease is critical for guaranteeing both quality and quantity. Early diagnosis of cotton leaf diseases [7] is necessary to limit disease transmission and increase the efficacy of treatment methods. Old methods permit a subjective approach to diagnosing the condition. In contrast, emerging technologies give a means for objectively diagnosing cotton leaf diseases. Automated technologies, such as Deep Learning [8], facilitate the precise and reliable identification of cotton leaf diseases, therefore saving time and money. Because of the disease and the environment's dynamic nature, early and precise identification of infected cotton leaves is a crucial issue in Computer Vision. The intricacy is enhanced by the leaf image's rich background, angle, and tiny symptoms. The identification of the disease is aided by visual signs [9]. In contrast, the photograph with a complex backdrop is a tough task.

PROBLEM STATEMENT

There are many causes that reduce the quantity and quality of cotton crop's production while the most common cause is the attack of diseases. It decreases the cotton crop's production by more than 9.8 Million tons each year world wide. [5]. But the early detection of disease from cotton crop might be save from the massive loss. Therefore, need to design an effective and automate deep learning based system for properly screening of cotton leaf. To boost the generalization capability of the model needs to trained the model on local dataset.

LITERATURE REVIEW

Cotton is a major natural fiber crop. It's cultivated in more than 100 countries, covering 2.5% of arable land [1]. In 2019/2020, 17 million hectares produced 28.2 million tons of cotton [2]. Despite Brazil's enormous size, its agriculture is focused on Mato Grosso and Bahia, hot and humid regions that require intense cultivation and major technical investment to battle fungal disease outbreaks, especially ramularia leaf spot.

Despite hurdles, plant disease detection is a popular topic. Many approaches have been presented throughout time. SVM algorithms may be used to identify and distinguish plant diseases in a systems manner. Another technique employs leaf images and an ANN-based algorithm to detect plant diseases. Baquero et al.[3] used color structure descriptors and closest neighbors. This approach classifies chlorosis, sooty molds, and early blight.

Shah et al. created an artificial neural network to identify cotton leaf diseases [4]. The basic strategy of the project is to identify cotton illnesses using an artificial neural network tool that employs picture pre-processing to highlight the damaged leaf based on color changes and determines the disease based on data. 18 leaf photographs were used to diagnose six diseases. Diseased leaves were coded as 1 and healthy leaves as 0.

[4] employed image processing to recognize cotton leaves. K-means segments the datasets. Herbivorous insects threaten cotton crops. Cotton herbivores devour leaves during the vegetative phase [5] and flower buds, blooms, and bolls during the reproductive phase [6]. They affect cotton's amino acid, water, and oxidative conditions [7]. Spodoptera frugiperda, the fall armyworm [8], causes irreversible plant damage. Early detection is key. Farmers use precision agriculture and remote sensing to maintain or enhance crop field productivity.

IMAGE PROCESSING

Revathi and Hemalatha [24] also developed image processing. They first digitized photos with a high-tech handheld camera. After that, image processing removes RGB highlights for further study. Then, explanatory approaches were employed to categorize the photos by subject. Edge recognition algorithms transform the RGB image to a dark scale image to identify leaf spot photo damage. Using shading channel checks and RGB highlighting helps differentiate unhealthy portions of photos. Kirthi Pilli et al. [25] discussed eAGRO for early crop disease detection. eAgro advice bridges farmers with agricultural specialists [26].

Tripathy proposed an image processing methodology for identifying cotton plant disease. The model fragments leaf sample images using thresholding. The model pinpoints the diseased leaves. Finally, sick area GLCM features were employed to categorize the condition. The result helps farmers create fruitful resolve and safeguard their crops [27].

Automatic disease detection and classification utilizing leaf image processing to identify plant disease indicators is a basic requirement. This can help farmers detect a disease early on before it spreads. First, a color transformation of the RGB leaf was produced, then color space transformation was applied. Image segmentation follows. The second phase eliminates green leaf regions. In the third phase, the sick sections' textural properties are examined. Karthika et al. recommended image processing for cotton disease detection [28-29]. Classification requires training and testing. The classifier is given a random picture from the data set of infected and normal cotton leaves. Bacteria blight, black arm spot, and leaf spot are classified using multi-class SVM. Cotton disease is identified by multi SVM and is called the Black spot arm. Inconsistent behavior sends a bad message to the farmer [30].

Azath et al. used CNN to enhance detection of cotton leaf disease and pests. They utilized bacterial blight, spider mite, and leaf miner. K-fold cross validation increased the CNN model's generalization. This study used 2400 specimens (600 photographs each class) for training. The produced model is written in Python 3.7.3 and based on Keras, TensorFlow, and Jupyter. The model is 96.4% accurate in recognizing cotton leaf diseases and pests. This showed its potential in real-time applications and the

necessity for IT-based solutions to supplement manual disease and pest detection [31].

Anton L.P. de Ocampo and Elmer P. Dadios [32] designed a computationally light mobile platform. This article uses MobileNet to classify plant diseases. CNN lightweight is used server-side. Model accuracy is 87.75%.

John C. Valdoria, Arlene R. Caballeo, Brian I. D. Fernandez, and John M. M. Condino [33] produced an android app that detects plant illness using image processing and a deep learning neural network. This Android app diagnoses illness using photos. The paper uses tensor flow server-side machine learning library.

Kumara et al. designed a machine learning-based cotton disease app. The software uses a predictive iOS ML model. CreateML for mac builds the ML model. All labeled data sets are imported into CreateML for picture categorization. Then, mac program trains and evaluates the model. 91% accuracy. App development uses Apple's core ml framework. The app is designed on VIPER iOS app architecture, which has numerous levels [34].

Using server-level integration, all machine learning logic may be done on the server for improved accuracy and performance. This approach allows the software to operate online and offline and use machine learning algorithms. More cotton diseases can be added to the same project and data combined.

Recent study has used machine learning to model plant remote sensing data [35]. Machine learning approaches identify plant illnesses and pests [36]. Tageldin et al. [37] used machine learning to forecast caterpillar (*Spodoptera littoralis*) attacks on cotton plants with 84% accuracy. Another recent addition showed how remote sensing can quantify fall armyworm damage [38].

Operational Land Imager (OLI) and Multispectral Instrument (MSI) are two contemporary orbital sensors used for agriculture mapping. Previous study [39] has proven these sensors' agronomic-related efficacy. Validated simulation methods started on a terrestrial scale and converged on aerial or orbital scales. This notion is untapped in agriculture.

Ramos et al. modeled the spectrum response of autumn armyworm-infested cotton plants. It found the best learner in a state-of-the-art comparison, the ideal day for analysis, and the most helpful wavelengths and spectral areas. It also developed a theoretical model for evaluating damage utilizing multispectral contemporary orbital sensors, which are used to map fields, crops, and farmlands. Random Forest outperformed other machine learning algorithms in discriminating healthy from unhealthy cotton plants, with Precision, Recall, and F1-measure values exceeding 0.912 on the last day of research (8th day). The framework can accurately identify cotton plants under insect herbivory attack, and trials are planned to evaluate its feasibility for other forms of insect damage in related cultivars. Future research will evaluate deeper learning approaches like deep neural networks with bigger in-field data [40].

Shah and Jain [41] reviewed ANN's applicability to cotton leaf diseases. The layman's technique to differentiating a cotton plant infection was to inspect it physically, which resulted in many errors owing to visual observations and lighting. However, the artificial nervous system made it easier to determine the nature of the cotton plant's leaves. Six disorders were diagnosed using 18 paper pictures. Scores were "1" and "0" for disease-free leaves.

Pujari et al. [42] proposed researching early pathogen contamination indicators. This initiative supports picture capture, preprocessing, feature selection, and infectious illness testing. The relationship between image processing technologies that characterize parasite disease eradication and agricultural testing has grown.

Umair Ayyub delivered a study on agricultural disturbances and data mining in 2018 [43]. Their study paper focused on Pakistani farmers' losses due to insect-caused agricultural diseases. Neural networks, SVMs, decision trees, and K-nearest neighbors were employed to analyze the perturbation. James examined GM crops and world crop demand in 2002 [44].

In this study, Saha et al. discuss neural network models for plant disease detection. According to the research, chromatic, texture, and morphological features are excellent for recognizing plant diseases. Most popular neural network models are ANNs and CNNs (CNNs). Plant sickness detection would solve a costly profession. Early diagnosis of plant diseases will help farmers raise food production, boosting the country's GDP [45].

Yang et al. [46] developed an algorithm to detect carnivorous *Phymatotrichopsis*. Convolutional neural networks evolved from conventional artificial neural networks because data is represented hierarchically. They learn faster than regular ANNs. When taught correctly, these techniques classify patterns accurately. CNN architecture affects visual resource extraction and categorization. Abade et al. [47] found that CNNs enhance plant disease identification. Even with real-world images, the accuracy is good.

Barbedo [48] identified plant ailments based on pests and leaf spots with above 70% accuracy for many crops. Deebe and Amuthe [49] employed several convolutional neural network designs to obtain 62 to 96% accuracy, however discovered the networks function better for the present controlled data set due to lesser accuracy of real-world data. CNNs perform well in applications related to crop disease classification and detection, as well as sophisticated imaging challenges [50].

Caldeira et al. show that deep learning may help diagnose agricultural pests and diseases by identifying blemishes on cotton leaves. Traditional image processing has been replaced by two deep convolutional network models. Deep learning improves accuracy. Data balance boosts performance, too. The ResNet50 convolutional network was better at finding lesions in the overall assessment and across classes. However, the average difference between its results

and those of GoogleNet is insignificant. We believe the current findings may be used outside of academia as operational tools [51].

Rajasekar et al. developed a deep learning approach for cotton identification, opening the way for a fast, automated, less costly, and trustworthy solution. CNN-based deep learning labels most cotton disease issues well. ResNet on imagenet is combined with the Xception module for image

categorization by building a new fully integrated Softmax layer. Focal Loss, not Cross Entropy Loss, is used to increase the learning ability of smaller features. In this method, the merged ResNet boosts feature extraction while lowering model calculation time. The recommended approach performed effectively on the big repository and cotton disease image dataset [54].

Ref	Segmentation Technique	Feature Extraction	Learning	Disease Detection	Dataset		Classification	Acc
					Image	Resolution		
[1]	Active Contour Method	Color and Texture	Supervised	Bacterial Blight, Alternaria and Grey-mildew	300	256*256	Random Forest	96.40%
[12]	Region based method Clustering	Color, Texture and Feature Location	Supervised	Bacterial Blight, Alternaria and Myrothecium	600	256 * 256	Nearest Neighbor	490.11292 sharpness, 502.62412 Myrothecium
[2]	Feature Extraction Network	Pixels with similar spectral and texture	Unsupervised	White Mildew	1951	256 * 256	ResNet 50 ResNet 150V2	82.55%
[3]	Feature Extraction Network	Pixels with similar spectral and texture	Unsupervised	White Mildew	1951	128 * 128	CNN (500 epoch)	90%
[4]	Histogram Method	Frequency and Gaps in data	Unsupervised	White Mildew	1951	256 * 256	ResNet 50	98.40 %
[31]	Feature Extraction Network	Color and Texture	Unsupervised	Leaf disease and pest diagnosis	2400	N/A	CNN (100 epoch)	89.0 %
[52]	Graph Cut Method.	Extract Color Layout	Supervised	Bacterial disease and Myrothecium	N/A	N/A	Neuro Fuzzy Inference	N/A
[28]	Clustering	color and texture	Unsupervised	Grey mildew	180	300 * 300	KNN technique	85%
[26]	Clustering	Contrast, Correlation, color and texture	Unsupervised	Bacterium blight and deficiency of magnesium		640x480	Neural Network	98.46%
[13]	K Means Clustering	Color and Texture	Supervised	Bacterial blight, Deficiency of magnesium	130	N/A	SVM	98%
[24]	Edge Detection	Boundary, Shape, Color and Texture	Unsupervised	leaf virus, grey mildew	N/A	N/A	K-Means nearest neighbor	94%
[53]	N/A	N/A	Supervised	Grey mildew,	N/A	N/A	Decision Tree	490.11292 sharpness, 502.62412

				Anthracnose, wilt				Myrothecium
[54]	C-means Clustering	N/A	Supervised	Cotton aphids	20	N/A	Fuzzy Techniques	85%
[41]	Colour Image Segmentation Method	Color and Contrast	Unsupervised	Rust and Scarab Diseases	N/A	N/A	Artificial Neural Networks	86.83 %

From the existing literature review it is conclude that lake of work has been done to detect the diseases (Grey Mildew, Cercospora, Bacterial Blight and Alternaria) from the cotton leaf using feature extraction and deep learning algorithm. All the above mentioned work done on public available dataset but for local generalization and acceptability of proposed model, collected the data set from cotton agricultural areas near the Multan city

METHODOLOGY

Agriculture plays an essential role in the economic growth of any nation. Cash crops include important agricultural products like cotton. Cotton is susceptible to the

vast majority of diseases that result in substantial crop losses. There are several diseases that impact yield through the leaves. If diseases are identified and treated in a timely manner, further crop loss can be avoided. Cotton is prone to a wide range of illnesses, some of which include Alternaria, Cercospora, Bacterial Blight, and Grey Mildew diseases. The correct diagnosis of the condition is essential for taking appropriate action. The diagnosis of plant diseases is significantly aided using deep learning. This work proposes a deep learning-based interactive solution for cotton leaf monitoring, as shown in Figure. 5. This technique can monitor cotton crop health and make better decisions for cotton crop management. This study was done to assist in solving this issue.

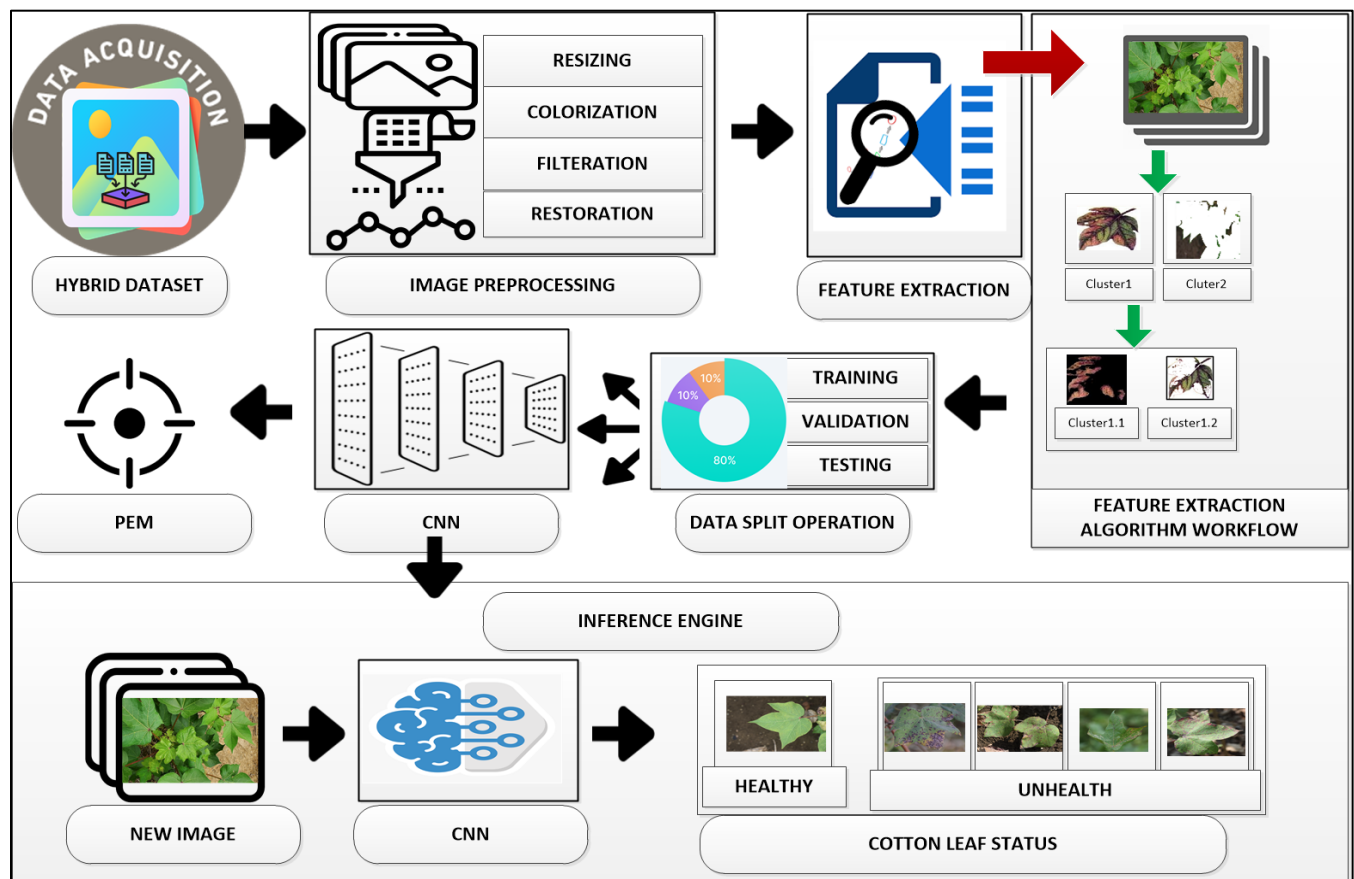


Figure 5: Proposed methodology for cotton leaf disease recognition

In this study we choose our own collected dataset that contains images of cotton leaves that were collected from

the field (Cotton agricultural areas nearby the Multan city) as shown in the Figure 6. The collection method is simple and

there are no high conditions. It contains a total of 522 images of healthy and diseased leaves. Images containing different diseases are taken from multiple angles at different time

periods. Therefore, to enhance the reliability and generalization of proposed model, need to preprocess all the data.



Figure 6: Dataset collection

The images need to be preprocessed first to improve the predictability and generalization of the algorithm and therefore following operations are performed on the given dataset to eliminate irrelevant information from the images.

A. Image Resizing

Images were collected from different sources; therefore, it is necessary to adjust the all the images of data set to standard size $256 * 256$. For the image resizing applied OpenCV library function, that improved the effectiveness of data.

B. Color Transformation

Before the image is formally analyzed color transformation of the images are often carried out, in order to analyze the image in best manners. In this step of preprocessing, all the images of the dataset are converted into a 3-axis color approach named LAB, here L is to represent the lightness of an image and A and B both indicate the chromaticity with no specific numeric limits. It is useful because in LAB color approach, Lightness is kept separate from color, so both can be easily adjusted without affecting each other.

C. Image Filtration

During the acquisition of images, noise may come from the environmental conditions therefore image is the most commonly used method in image processing that suppress noise from the images of dataset. Noise of an image refers to unnecessary or redundant information existing in image data that affects the quality of dataset, so it must be corrected. To apply filter2D function to enhance the quality of images.

D. Image Restoration

Image restoration is an important step in the field of image preprocessing that deals with this deal with blur images. For this utilized OpenCV to restore the original appearance of the scene of the degraded image.

Feature Extraction

The process of feature extraction addresses a feature, also known as a piece of information, which is connected to the accomplishment of a computational task linked with an application. Specific structures inside a picture, such as points, edges, or objects, might be referred to as features. While feature extraction is a crucial step in the process of many computer vision applications, it is also extremely useful in a variety of other areas, such as motion structure, picture retrieval, object identification, and many more.

Customized CNN

In recent years, convolutional neural networks, also known as ConvNets, have emerged as one of the finest learning algorithms for comprehending the content of images. These networks have also demonstrated exceptional performance in tasks linked to detection and retrieval. Active research groups have been created by a wide variety of organizations, including Google, Microsoft, AT&T, NEC, and Facebook, in order to investigate novel CNN architectural designs.

In this study, used CNN model as shown in this figure 7, there are primarily three different kinds of layers: convolutional layer, pooling layer and fully connected layer

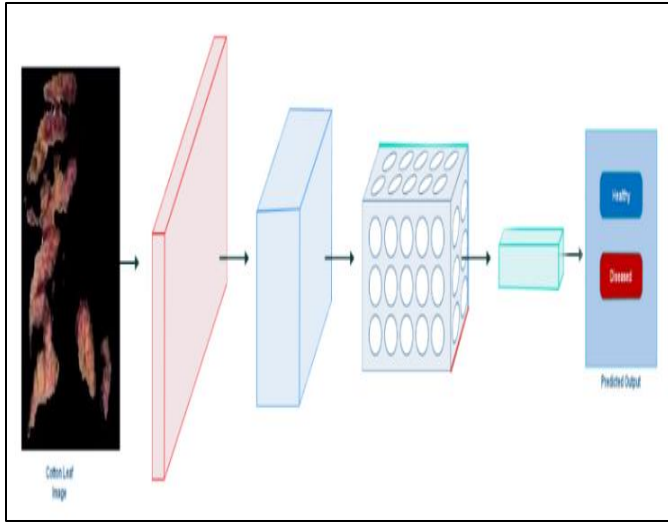


Figure 7: CNN Model

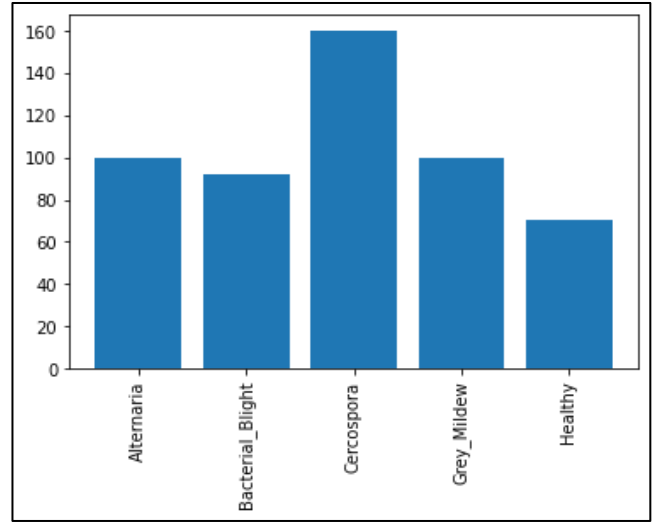


Figure 8: Dataset class distribution

A. Convolutional Layer

The input data is passed through a series of filters (sometimes called kernels) in a convolutional layer. To better extract features from the input data, each kernel has a width of 3, a height of 3, and a weight of 3x3. The kernel's weights are set to random values at the start of training, but they will gradually become more accurate as they process the data in the training set.

B. Pooling layer

By using pooling layers (or downsampling layers), the most important aspects of feature maps may be preserved while maintaining a manageable number of dimensions. The max pooling process is performed by filters as they glide over the input data in a pooling layer.

C. Fully connected layer

There are three distinct layers that make up a multilayer perceptron: the input layer, the hidden layer, and the output layer. Features produced by the CNN are sent into the input layer. Each neuron in the "hidden layer" will have a weight that is determined by the previous layers' weights and is learnt throughout the training process. There are several hidden levels in an MLP. Additionally, there is a layer of neurons responsible for outputting data. However, it has a relu function in hidden layers while the output layer used the softmax activation function to generate the probability of each class of healthy and diseased cotton leaf.

Inference Engine

For this study's purpose of inferring whether a particular cotton leaf is healthy or dead, the researchers employed an inference engine, a component of AI-based systems that interprets and executes the information obtained in accordance with a method chosen by the researchers.

RESULTS AND DISCUSSIONS

In this section discuss the result to ensure the reliability, acceptability and trustworthiness of CNN learning-based proposed method that monitors cotton crop health and make better decisions for cotton crop management.

The dataset consists of 522 images as shown in Figure 8 and contains five classes; 100 images of Alternaria, 80 images of Bacterial Blight, 160 images of Cercospora, 100 images of Grey Mildewand 80 images of Healthy. LAB Color is a global model of coloring it is a more accurate color space and based on 3 axes and each has its own numeric value range; for all dataset images the level of lightness is 0; while A is second axis that deals with the level of greenness to redness and set the value 1 from the range; and last parameter of LAB color is B that indicates the blueness to yellowness effects on image with the value 2 as shown in the Figure 9.

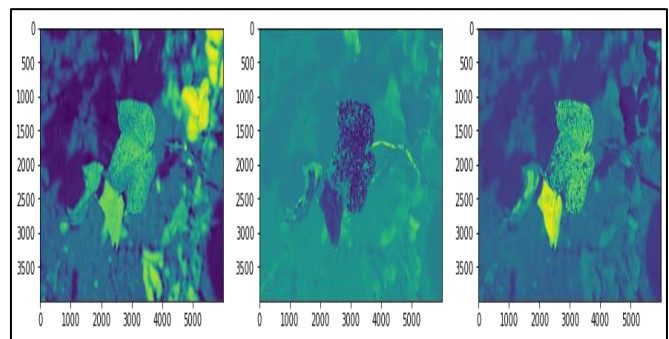


Figure 9: LAB color Scheme

In this work, built four number of clusters (Cluster 1, 2, 1.1 and 1.2) for the extraction of features from the dataset to train the model for better prediction and these clusters are based on k means cluster algorithm. Cluster 1 based on foreground (Leaf Parts) while the cluster 2 contains background area of the image, remaining part from the leaf. After that, further built two more clusters (cluster 1.1 and

cluster 1.2) that separates the disease part of leaf from the remaining leaf part as shown in the Figure 10.

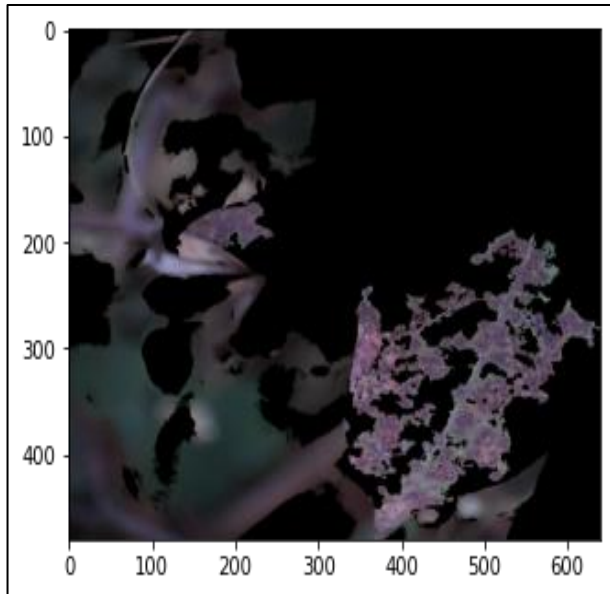


Figure 10: Diseased Leaf after Clustering

The training and validation accuracy are both hovering near to each other which doesn't indicate overfitting of any kind that enhance the generalized acceptability of the model. Training loss is measured during each epoch while validation loss is measured after each epoch and it is metric used to assess the performance of a deep learning model on the validation set. Both losses curves are decreasing 90% nearly to each other that indicates the consistency of the model as shown in Figure 11.

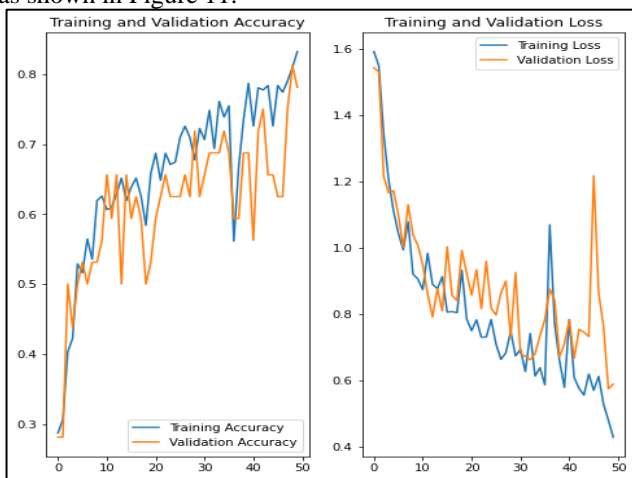


Figure 11: Accuracy and loss Result

To evaluate the predictive ability of the model from confusion matrix as shown in Figure 12 used Model Performance Evaluation criteria that defines the trustworthiness of results. In this calculate the Accuracy, Error Rate, Precision Rate, Recall Rate, and F1 Measurement.

Alternaria	72	2	3	4	5
Grey Mildew	2	74	3	6	4
Bacterial Blight	2	2	73	3	5
Cercospora	2	1	4	75	4
Healthy	2	2	4	4	79
	Alternaria	Grey Mildew	Bacterial Blight	Cercospora	Healthy

Figure 12: Confusion Matrix

Accuracy:

Accuracy refers to the proportion of the number of correctly classified samples to the total number of samples. It is a commonly used method to evaluate the generalization ability of a learning model. It can be measured by Eq. no. 1

$$Accuracy = \frac{TP + FN}{N} \quad (1)$$

Accuracy = 85.42%

Error Rate:

Error rate refers to the proportion of misclassified samples to the total number of samples for selective algorithms (14). It can be calculated using Eq. no 2

$$Error Rate = \frac{TN + FP}{N} \quad (2)$$

Error Rate = 14.58.

Precision:

Precision is the ratio of TP to the sum of TP and FP. It refers to a classifier's ability to not to label a negative instance as positive. It is the accuracy of positive predictions (45). Formula of precision is shown in Eq no. 3.

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

Table 1 indicates the values of Precision for each class,

Alternaria	0.84
Grey Mildew	0.83
Bacterial Blight	0.86
Cercospora	0.87
Healthy	0.87

For single value of precision of the whole model we need to calculate the average precision using the Eq. no. 4

$$Average\ Precision = \frac{TTP}{TTP + TFP} \quad (4)$$

Here TTP is Total number of True Positive and TFP are the Total number of False Positive. So,

$$TTP = 72+74+73+75+79 = 373$$

$$TFP = (2+3+4+5) + (2+3+6+4) + (2+2+3+5) + (2+1+4+4) + (2+2+4+4) = 64$$

$$Average\ Precision = 0.8542$$

Recall:

Recall is the ratio of TP to the sum of TP and FN. It refers to a classifier's ability to find all positive instances. It is the fraction of correct predictions (45). Formula of recall is shown in Eq no. 5.

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

Table 2 indicates the values of Recall for each class, Table 2 Class-wise Recall

Alternaria	0.90
Grey Mildew	0.91
Bacterial Blight	0.84
Cercospora	0.82
Healthy	0.81

For single value of Recall of the whole model we need to calculate the average Recall using the Eq. no. 6

$$Average\ Recall = \frac{TTP}{TTP + TFN} \quad (6)$$

Here TTP is Total number of True Positive and TFN are the Total number of False Positive. So,

$$TTP = 72+74+73+75+79 = 373$$

$$TFN = (2+2+2+2) + (2+2+1+2) + (3+3+4+4) + (4+6+3+4) + (5+4+5+4) = 64$$

$$Average\ Recall = 0.8542$$

F1 Score:

It is the evaluation standard, which is the harmonic average of the precision and recall. Best F1 score is 1.0 and the worst is 0.0 (45). Formula of F1 score is shown in Eq no. 5.7.

$$F1\ Score = 2 \times \frac{(Recall \times Precision)}{(Recall + Precision)} \quad (7)$$

Table 3 indicates the values of F1 Score for each class,

Alternaria	0.87
Grey Mildew	0.87
Bacterial Blight	0.85
Cercospora	0.84
Healthy	0.84

For single value of F1-score of the whole model we need to calculate the average F1-Score also known as Macro F1 and can be calculate by using the Eq. no. 8

$$Average\ F1 = \frac{Class - wise\ F1\ Score\ Value}{Number\ of\ Classes} \quad (8)$$

Here we have 05 number of classes and class-1 F1 Score values are 0.87, 0.87, 0.85, 0.84 and 0.84 So,

$$Average\ F1\ score\ (Macro\ F1) = (0.87+0.87+0.85+0.84+0.84)/5$$

$$Average\ F1\ score = 0.854$$

Kappa Coefficient:

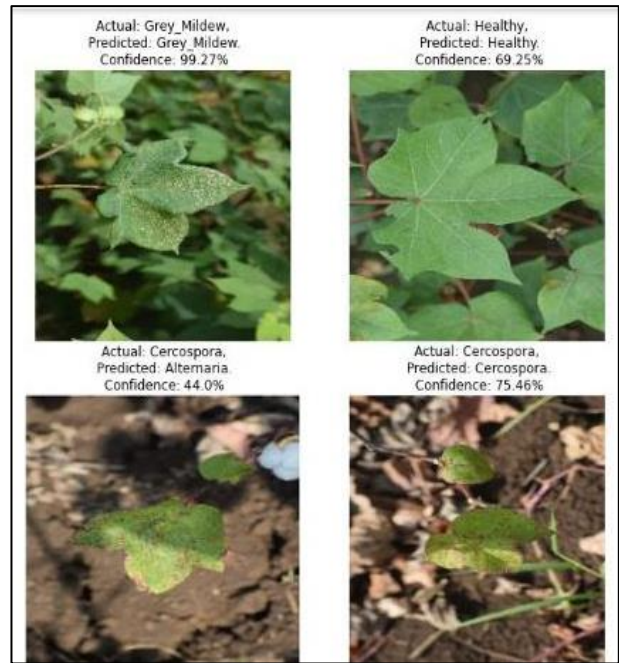


Figure 13. Results

It is statistical function which is calculated by the Eq. no. 9. It is used to measure the inter-reliability of qualitative items. Generally, a kappa value less than 0.20 is considered as a poor agreement. Kappa value from 0.20 to 0.39 is considered as a fair agreement. Kappa value from 0.40 to 0.59 is considered as moderate agreement. Kappa value from 0.60 to 0.78 is considered as good agreement and values of kappa more than 0.80 represents excellent agreement.

$$Kappa\ Coefficient = \frac{OA - AC}{1 - AC} \quad (9)$$

$$OA = (72 + 74 + 73 + 75 + 79) / 437 = .8542$$

$$AC = (80 / 437) * (86 / 437) + (81 / 437) * (89 / 437) + (87 / 437) * (85 / 437) + (92 / 437) * (86 / 437) + (97 / 437) * (91 / 437) = 0.20$$

$$Kappa\ Coefficient = .817$$

Test model performance on random images and a real-time environment. The model predicts perfectly but, on some images, or expressions shows some false results just

because of small dataset. Some samples are shown in Figure 13

CONCLUSION:

The cotton crop is vital to the development of any country and it is mostly affected by diseases that cause significant crop losses. Many diseases affect yield through leaves. Early and accurate disease identification is key to effective action. In this study, we proposed an interactive framework based on features extraction and deep learning algorithm that an important role in cotton disease identification. It identified the leaf according to its class (Healthy, Grey Mildew, Cercospora, Bacterial Blight, and Alternaria). The proposed deep learning-based cotton leaf disease screening model was trained on 522 cotton leaves' which own collected dataset from the field. As results show that the model secure; 85.42% overall accuracy, 0.8542 precision, 0.8542 recall, 0.854 F1 score, and 0.817 kappa coefficient that indicates the good generalization capabilities of the model. This work is considered a useful tool, especially for the naive farmer.

In future work, the accuracy of identification and detection may be increased even further by integrating more images in the dataset and fine-tuning the CNN model's parameters. For a mobile user, it can be further improved by handling low-resolution images and reducing model size.

CREDIT AUTHOR STATEMENT

Jawad Hassan: Methodology, Data curation, Software, Writing- Original draft. **Kaleem Razzaq Malik:** Conceptualization, Supervision. **Ghulam Irtaza:** Writing- Reviewing and Editing. **Ali Ghulam and Ashfaq Ahmad:** New experiments were conducted as suggested by reviewers.

REFERENCES

- [1] R. F. Caldeira, W. E. Santiago, and B. Teruel, "Identification of cotton leaf lesions using deep learning techniques," *Sensors (Basel)*, vol. 21, no. 9, p. 3169, 2021.
- [2] D. Baquero, J. Molina, R. Gil, C. Bojacá, H. Franco, and F. Gómez, "An image retrieval system for tomato disease assessment," in *2014 XIX Symposium on Image, Signal Processing and Artificial Vision, IIEEE*, 2014, pp. 1–5.
- [3] N. Shah and S. Jain, "Detection of disease in cotton leaf using artificial neural network," in *2019 Amity International Conference on Artificial Intelligence (AICAI), IIEEE*, 2019, pp. 473–476.
- [4] P. P. Warne and S. R. Ganorkar, "Detection of diseases on cotton leaves using K-mean clustering method," *International Research Journal of Engineering and Technology (IRJET)*, vol. 2, no. 4, pp. 425–431, 2015.
- [5] A. A. Sarangdhar and V. R. Pawar, "Machine learning regression technique for cotton leaf disease detection and controlling using IoT," in *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)*, vol. 2, IIEEE, 2017, pp. 449–454.
- [6] L. T. Pascua and M. E. Pascua, "The distribution and movement of cotton bollworm, *Helicoverpa armigera* Hubner (Lepidoptera: Noctuidae) Larvae on Cotton," *Philippine Journal of Science*, vol. 131, no. 2, pp. 91–98, 2002.
- [7] E. S. Gomes, V. Santos, and C. J. Ávila, "Biology and fertility life table of *Helicoverpa armigera* (Lepidoptera: Noctuidae) in different hosts," *Entomological science*, vol. 20, no. 1, pp. 419–426, 2017.
- [8] M. Eisenring, S. E. Naranjo, S. Bacher, A. Abbott, M. Meissle, and J. Romeis, "Reduced caterpillar damage can benefit plant bugs in Bt cotton," *Scientific Reports*, vol. 9, no. 1, pp. 1–9, 2019.
- [9] Y. C. Zhu, C. A. Blanco, M. Portilla, J. Adamczyk, R. Luttrell, and F. Huang, "Evidence of multiple/cross resistance to Bt and organophosphate insecticides in Puerto Rico population of the fall armyworm, *Spodoptera frugiperda*," *Pesticide Biochemistry and Physiology*, vol. 122, pp. 15–21, 2015.
- [10] A. Parikh, M. S. Raval, C. Parmar, and S. Chaudhary, "Disease detection and severity estimation in cotton plant from unconstrained images," in *2016 IEEE, international conference on data science and advanced analytics (DSAA), IIEEE*, 2016, pp. 594–601.
- [11] P. R. Rothe and R. V. Kshirsagar, "Adaptive neuro-fuzzy inference system for recognition of cotton leaf diseases," in *2014 Innovative Applications of Computational Intelligence on Power, Energy and Controls with their impact on Humanity (CIPECH), IIEEE*, 2014, pp. 12–17.
- [12] P. R. Rothe and R. V. Kshirsagar, "Cotton leaf disease identification using pattern recognition techniques," in *2015 International Conference on Pervasive Computing (ICPC)*, 2015.
- [13] N. R. Bhimte and V. R. Thool, "Diseases detection of cotton leaf spot using image processing and SVM classifier," in *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), IIEEE*, 2018, pp. 340–344.
- [14] C. U. Kumari, S. J. Prasad, and G. Mounika, "Leaf disease detection: feature extraction with K-means clustering and classification with ANN," in *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), IIEEE*, 2019, pp. 1095–1098.
- [15] S. Kumar et al., "A Comparative Analysis of Machine Learning Algorithms for Detection of Organic and Nonorganic Cotton Diseases," *Mathematical Problems in Engineering*, 2021.
- [16] G. Karthik and N. Praburam, "Detection and prevention of Banana leaf diseases from banana plant using Embedded Linux board," in *Online International Conference on Green Engineering and Technologies (IC-GET), IIEEE*, 2016, pp. 1–5.
- [17] S. Jaisakthi, P. Mirunalini, and D. Thenmozhi, "Grape Leaf Disease Identification using Machine Learning Techniques," in *Second International Conference on Computational Intelligence in Data Science, IIEEE/ICCIDS 2019*, .
- [18] M. Punn and N. Bhalla, "Classification of Wheat Grains Using Machine Algorithms," *International Journal of Science and Research*, pp. 363–366, 2013.
- [19] P. R. Rothe and R. V. Kshirsagar, "Adaptive neuro-fuzzy inference system for recognition of cotton leaf diseases," in *Proceedings of the International Conference on Innovative Applications of Computational Intelligence on Power, Energy, and Controls with their Impact on Humanity (CIPECH-14), Ghaziabad, India*, 2014.
- [20] X. Cheng, Y. Zhang, Y. Chen, Y. Wu, and Y. Yue, "Pest identification via deep residual learning in complex background," *Computers and Electronics in Agriculture*, vol. 141, pp. 351–356, 2017.
- [21] Z. Iqbal, M. A. Khan, M. Sharif, J. H. Shah, M. H. Ur Rehman, and K. Javed, "An automated detection and classification of citrus plant diseases using image processing techniques: A review," *Computers and electronics in agriculture*, vol. 153, pp. 12–32, 2018.

- [22] M. J. Brewer and J. P. Glover, "Boll injury caused by leafhopper bug in late-season cotton," *Crop Protection*, vol. 119, pp. 214–218, 2019.
- [23] A. Abdelraheem, N. Esmaili, M. O'connell, and J. Zhang, "Progress and perspective on drought and salt stress tolerance in cotton," *Industrial Crops and Products*, vol. 130, pp. 118–129, 2019.
- [24] P. Revathi and M. Hemalatha, "Classification of cotton leaf spot diseases using image processing edge detection techniques," in *2012 International Conference on Emerging Trends in Science, Engineering and Technology (INCOSET)*, IEEE, 2012, pp. 169–173.
- [25] M. Lezoche, J. E. Hernandez, M. del M. E. Alemany Díaz, H. Panetto, and J. Kacprzyk, "Agri-food 4.0: A survey of the supply chains and technologies for the future agriculture," *Comput. Ind.*, vol. 117, no. 103187, p. 103187, 2020.
- [26] S. K. Pilli, B. Nallathambi, S. J. George, and V. Diwanji, "eAGROBOT-A robot for early crop disease detection using image processing," in *2015 2nd International Conference on Electronics and Communication Systems (ICECS)*, IEEE, 2015, pp. 1684–1689.
- [27] S. Tripathy, "Detection of cotton leaf disease using image processing techniques," *J. Phys. Conf. Ser.*, vol. 2062, no. 1, p. 012009, 2021.
- [28] N. Krithika and A. G. Selvarani, "An individual grape leaf disease identification using leaf skeletons and KNN classification," in *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, 2017.
- [29] S. Tripathy and T. Swarnkar, "A comparative analysis on filtering techniques used in preprocessing of mammogram image," in *Advances in Intelligent Systems and Computing*, Singapore: Springer Singapore, 2020, pp. 455–464.
- [30] J. Karthika, K. Mathan kumar, M. Santhosha, T. Sharan, and S. Sri hariharan, "Disease detection in cotton leaf spot using image processing," *J. Phys. Conf. Ser.*, vol. 1916, no. 1, p. 012224, 2021.
- [31] M. Zekiwo and A. Bruck, "Deep learning-based image processing for cotton leaf disease and pest diagnosis," *Journal of Electrical and Computer Engineering*, 2021.
- [32] D. Ocampo and A. L. P. Dadios, "Mobile platform implementation of lightweight neural network model for plant disease detection and recognition," in *2018 IEEE 10th International Conference on Humanoid, IEEE*, 2018, pp. 1–4.
- [33] J. C. Valdivia, A. R. Caballeo, B. I. D. Fernandez, and J. M. M. Condino, "October. iDahon: An android based terrestrial plant disease detection mobile application through digital image processing using deep learning neural network algorithm," in *2019 4th International Conference on Information Technology (InCIT)*, IEEE, 2019, pp. 94–98.
- [34] S. Kumar and J. Desai, "Development And Simulation Of Ios App To Detect Cotton Plant Diseases Using Machine Learning Techniques."
- [35] A. P. M. Ramos et al., "A random forest ranking approach to predict yield in maize with uav-based vegetation spectral indices," *Computers and Electronics in Agriculture*, vol. 178, 2020.
- [36] V. Srikanth et al., "Modelling and optimization of developed cocoa beans extractor parameters using box behnken design and artificial neural network," *Computers and Electronics in Agriculture*, vol. 177, 2020.
- [37] A. Tageldin, H. Mostafa, and H. S. Mohammed, "Applying Machine Learning Technology in the Prediction of Crop Infestation with Cotton Leafworm in Greenhouse," *bioRxiv*, 2020.
- [38] M. L. Buchaillet et al., "Multi-scale remote sensing for fall armyworm monitoring and early warning systems," in *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, 2020, pp. 4886–4889.
- [39] L. P. Osco et al., "Leaf nitrogen concentration and plant height prediction for maize using UAV-based multispectral imagery and machine learning techniques," *Remote Sensing*, vol. 12, no. 19, 2020.
- [40] A. P. M. Ramos et al., "Detecting the attack of the fall armyworm (*Spodoptera frugiperda*) in cotton plants with machine learning and spectral measurements," *Precision Agriculture*, pp. 1–22, 2021.
- [41] V. A. Gulhane and A. A. Gurjar, "Detection of diseases on cotton leaves and its possible diagnosis," *International Journal of Image Processing (IJIP)*, vol. 5, no. 5, pp. 590–598, 2011.
- [42] J. Pujari, R. Yakkundimath, and A. Byadgi, "Identification and classification of fungus disease affected on agriculture/horticulture crops using image processing techniques," in *IEEE International Conference on the Computational Intelligence and Computing Research*, 2014.
- [43] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, "Deep neural networks based recognition of plant diseases by leaf image classification. Computational intelligence and neuroscience. 2016."
- [44] L. J. Thomson, S. Macfadyen, and A. A. Hoffmann, "Predicting the effects of climate change on natural enemies of agricultural pests," *Biological control*, vol. 52, no. 3, pp. 296–306, 2010.
- [45] P. Saha and M. N. Nachappa, "Cotton Plant Disease Prediction Using Deep Learning. ."
- [46] C. Yang, G. N. Odvody, C. J. Fernandez, J. A. Landivar, R. R. Minzenmayer, and R. L. Nichols, "Evaluating unsupervised and supervised image classification methods for mapping cotton root rot," *Precis. Agric.*, vol. 16, no. 2, pp. 201–215, 2015.
- [47] A. Abade, P. Ferreira, and F. Vidal, "Plant Diseases recognition on images using Convolutional Neural Networks: A Systematic Review. arXiv 2020."
- [48] J. G. Arnal Barbedo, "Plant disease identification from individual lesions and spots using deep learning," *Biosyst. Eng.*, vol. 180, pp. 96–107, 2019.
- [49] K. Deeba and B. Amutha, "WITHDRAWN: ResNet - deep neural network architecture for leaf disease classification," *Microprocessor. Microsyst.*, no. 103364, p. 103364, 2020.
- [50] J. Boulent, S. Foucher, J. Théau, and P. L. St-Charles, "Convolutional Neural Networks for the Automatic Identification of Plant Diseases," *Front. Plant Sci*, vol. 10, 2019.
- [51] V. Rajasekar, K. Venu, S. R. Jena, R. J. Varthini, and S. Ishwarya, "Detection of Cotton Plant Diseases Using Deep Transfer Learning," *Journal of Mobile Multimedia*, vol. 18, no. 2, pp. 307–324, 2021.
- [52] P. Revathi and M. Hemalatha, "Advance computing enrichment evaluation of cotton leaf spot disease detection using image edge detection," in *Proceedings of the 2012 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT 12)*, Coimbatore, India, 2012.
- [53] J. Chopda, H. Raveshiya, S. Nakum, and V. Nakrani, "Cotton crop disease detection using decision tree classifier," in *2018 International Conference on Smart City and Emerging Technology (ICSCET)*, IEEE, 2018, pp. 1–5.
- [54] J. Zhao, M. Liu, and M. Yao, "Study on image recognition of insect pest of sugarcane cotton aphid based on rough set and fuzzy C-means clustering," in *2009 Third International Symposium on Intelligent Information Technology Application*, vol. 2, IEEE, 2009, pp. 553–555.