DOI: 10.15849/IJASCA.250330.05

Int. J. Advance Soft Compu. Appl, Vol. 17, No. 1, March 2025
Print ISSN: 2710-1274, Online ISSN: 2074-8523

Automated System for Detecting, Identifying, and Preventing Cotton Leaf and Boll Diseases Using Deep Learning

Saeed Azfar [1], Adnan Nadeem[2, *], Kamran Ahsan[3], Amir Mehmood[4], Abdul Ali[5], Saad S. Al Qahtany[2] and Hani Almoamari[2]

Department of Computer Science, Institute of Business Management, Karachi 75300, Pakistan
saeed.azfar@iobm.edu.pk
Faculty of Computer and Information Systems, Islamic University of Madinah, Medina 42351, Saudi Arabia
adnan.nadeem@iu.edu.sa, s.alqahtany@iu.edu.sa, h.moamari@iu.edu.sa
Dept. of Computer Science, Federal Urdu University of Arts, Science and Technology, Karachi,
Kamran.ahsan@fuuast.edu.pk
Department of Computer Science and IT, Sir Syed University of Engineering and Technology, Karachi 75300, Pakistan
Amir.mehmood@ssuet.edu.pk
University of Loralai, Loralai, Baluchistan, Pakistan
Abdul.ali@uoli.edu.pk

Abstract

The Internet of Things (IoT) technology facilitates real-time data collection through sensors, enhancing disease detection and management accuracy and crop efficiency. Pest infestations in cotton crops adversely impact its production and ultimately affect the nation's economy which depends on agriculture. Previous research has primarily focused on detecting cotton leaf diseases through imagery, while this study addresses both cotton leaf and its boll diseases. An existing dataset of cotton leaf images was customized by incorporating classes of cotton boll images, resulting in a comprehensive dataset of 7289 images categorized into eight distinct classes, including healthy and diseased leaves and bolls. Deep learning models, including VGG16, InceptionV3, and a tailored model, were applied to evaluate the accuracy of disease identification and the detection of infection levels. The tailored model demonstrated performance comparable to pre-trained models, achieving high accuracy in classification tasks. The IoT-integrated, perception-based decision system proposed in this study enables early detection of diseases and supports the implementation of preventive measures, contributing to improved crop yields and mitigating the economic losses caused by pest infestations and diseases.

Keywords: *Convolutional neural network (CNN), Cotton Dataset, Cotton disease, Crop Disease identification, Deep learning, Internet of Things (IoT), Perception-based decision Making.*

Received 6 October 2024; Accepted 30 January 2025

1 Introduction

With the growing global population, the demand for precision agriculture, both in terms of product quality and quantity, is also increasing. Technology plays a crucial role in addressing current agricultural demands, including crop monitoring, pest management, remote field monitoring, and food productivity. The cotton crop holds importance for producing thread and fiber, meeting clothing needs. Cotton supports textile industry growth, contributing to national revenue. Plant protection is critical in cotton production, alongside other factors [1]. Similarly, crop diseases and harmful pests pose risks to reducing product quality and quantity [2]. Specifically, identifying diseases and detecting crop pests are vital tasks for improving yield and production quality, with early disease detection enhancing crop quality [3].

The traditional method of detecting and identifying cotton diseases is challenging, human-dependent, and requires significant experience and expertise. As a result, it does not always provide farmers with an accurate assessment of their crops. Technologies such as remote sensing and artificial intelligence improve the reliability of disease diagnosis. Machine learning, deep learning algorithms, and computer vision [4] techniques efficiently identify crop pests and their diseases. A computerized system for identifying crop pests or diseases can achieve high accuracy [5, 6]. Incorporating IoT devices allows these systems to collect continuous data, improving the accuracy and timeliness of disease detection, preventing further spread, and enabling automated drone-based spraying.

The integration of IoT (Internet of Things) in agriculture, particularly in cotton farming, has transformed crop management and monitoring. IoT devices such as soil sensors, weather stations, and drone-based imaging systems provide real-time data on crop health, soil moisture levels, and environmental conditions, enabling precise and timely interventions to enhance crop yield and quality. Early diagnosis of diseases is critical to preventing their spread, as timely identification of plants' local seasonal viruses ensures effective crop management and a good harvest [7]. Research has shown that automatic disease identification systems, such as Deep Learning [8], save time and money while offering more reliable diagnoses.IoT-enabled cameras and sensors can supply live data to CNN models, making the identification process more dynamic and responsive. Additionally, an IoT-based alert system, combined with drone-based monitoring and spraying, allows farmers to monitor their crops closely and prevent damage from pests and diseases.

Convolutional Neural Networks (CNNs) [9, 10] are widely used for identifying plant diseases and agricultural pests. In CNNs, input images pass through convolutional layers for feature extraction, followed by pooling layers to reduce spatial dimensions, and fully connected layers to process feature vectors. The model outputs the probability of leaf health or condition, learning through training on labeled images of healthy and diseased plants. This study focuses on testing CNNs for efficient identification of Cotton Plant Leaf and Bolls diseases. Fig. 1. Illustrates the severity levels of Boll Rot and various Leaves diseases. The reason for incorporating these levels is that when a farmer identifies disease symptoms at an initial stage, they can take prompt action to minimize crop loss and mark and isolate it for treatment. Proposed system incorporates IoT sensors that can immediately alert farmers when symptoms are detected, enabling early intervention. By integrating IoT technology, it enhance the system's ability to deliver real-time monitoring and prompt alerts, significantly improving disease management and crop protection in precision agriculture.

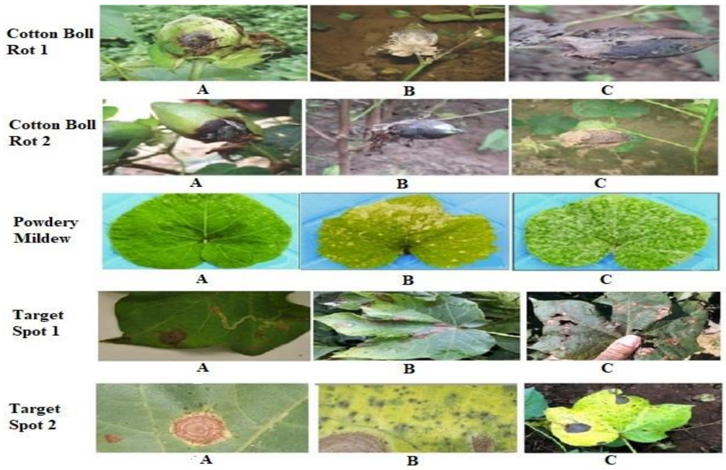


Fig. 1. Cotton bolls and Leaves exhibiting symptoms at various stages of decay are depicted as a) Initial stage b) Intermediate stage and c) Final stage of Disease.

In conclusion, following the introduction, the paper is structured as follows: the second section provides an overview of the background and related work on applying machine learning to plant diseases. The third section details the methodology, including raw data collection, results, and comparisons with other studies, as well as pre-trained models. The final section presents the research findings and discusses future directions.

2 Background with related work

Examining the background of this work reveals that many researchers have already explored this area. In this segment of the research, a comparative review of the work of selected researchers has been conducted, presented in table 1. A thorough analysis of these works has been done to understand the technologies utilized in this field and determine how to advance further.

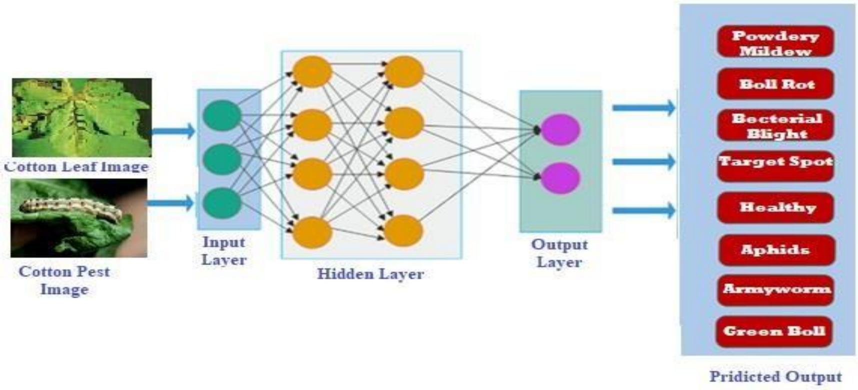
2.1. Cotton Disease Monitoring Using Deep Learning

Deep learning is a rapidly emerging technology in the industry and a highly specialized branch of machine learning that has gained significant popularity in recent years in various fields such as crowd management [12, 13] and healthcare [14, 15]. Among the many features of this technology, its noteworthy ability is to handle a vast amount of data, whether that data is in text format or in the form of audio or videos. As a result, it has found wide applications in various domains, particularly in agriculture and plant sciences, where the diagnosis of diseases and pests relies on image analysis [2, 7 and 8]. Among the many features of deep learning, a particularly unique aspect is its ability to extract desired features from any image based on specific requirements. It employs an automatic method to extract features from images, utilizing a sophisticated and intricate model. This intricate model ensures the rapid and efficient acquisition of desired properties. Unlike conventional methods, deep learning enables us to obtain highly accurate data without errors. In the present research, the aim is to build a convolutional neural network that will provide information about cotton pests and their diseases to farmers, as illustrated in Fig.2, to achieve accurate disease identification and pest detection results.

Fig. 2. Proposed Conceptual CNN Architecture for Disease Detection.

2.2 IoT based Pests / Disease Prediction and Control system

A review of the existing literature, as summarized in Table 1, shows that significant



progress has been made in pest monitoring and disease detection using AI and IoT. While several systems have been proposed for various crops, most lack comprehensive solutions to prevent pest and disease damage. Only a few offer full solutions, but even those have limitations. In [75], IoT and machine learning are used to identify insects, with an ultrasonic generator to repel pests, but this system focuses solely on pests. Another AI-based system in [76] provides alerts but lacks preventive solutions. Recent research [77] proposed an IoT-based crop disease prediction system, but it is limited to weather data and offers no preventive measures. Additionally, it is generalized and not specific to any pest or disease, relying on monitoring traps for predictions.

2.3. GAP Analysis

Table 1 focuses on the detection of Pests or Diseases by the utilization of AI tools and methods. This table presents relevant and contemporary research studies closely related to our current study. Upon reviewing these investigations, it is observed that some of them [24-33] address multiple crop pests and diseases simultaneously. Noteworthy studies include research on pests and diseases in crops such as Tomato [10], Palm and Coconut [9], Tea Plants [34], Millet and Rice [35-38], and Sugarcane [39]. However, these studies are limited in that they do not offer preventive measures. Review papers [36, 40] provide a detailed analysis of machine learning techniques used for detecting plant and foliar diseases across various crops, but these too are restricted to disease detection. Additionally, studies on cotton crop pests and related diseases, such as [29, 41, and 43], align with our research but are similarly limited to detection and identification only.

Among these investigations, certain studies have put forth comprehensive frameworks, whereas others have solely employed artificial intelligence tools to detect the presence of pests or diseases using provided datasets. Upon reviewing the existing research on cotton, it has been observed that [41] focus on ontology-based methods, [28] develop a disease database and knowledge base, [42] utilize weather variations for pest detection establish a database based on cotton disease symptoms, leading to the design of a web-based expert system. In these three conference papers [43-45], detection systems for various cotton diseases have been proposed. One paper utilizes climatic parameters, while the other two employ image processing techniques.

In contrast, our work markedly distinguishes itself from all other studies conducted within this field. Except for two studies [46-47] none of the other reviewed studies have delved into cotton Pests or Cotton bollworm disease datasets. The present study acquired a dataset encompassing six distinct classes of cotton diseases from the widely acknowledged dataset platform, namely Kaggle. Additionally, two classes were developed in-house and seamlessly integrated into the Kaggle-procured dataset.

Furthermore, a prominent facet of our research resides in the utilization of a dataset that encompasses three distinct image categories. Within the first image category, select classes representing cotton pests have been encompassed. The second category comprises images pertinent to foliar diseases, while the third image category encompasses unique images of affected cotton bolls, a facet hitherto unexplored by previous researchers. Our dataset stands out due to its incorporation of these three distinct image categories, each underscoring a unique facet of cotton-related issues. Another distinctive aspect of our research lies in our endeavor to detect diseases affecting cotton bolls, a novel approach that has not been undertaken previously. The rationale behind this lies in the fact that cotton boll diseases wield the potential to decimate the entire cotton yield. A unique aspect of our research is its comprehensive approach to providing a complete solution. First, pests and diseases in the field are detected using machine learning algorithms. Then, based on the identified pests or diseases, preventive measures are offered, which may include pesticide spraying, farmer intervention, or the removal of infected plants and cotton bolls. The severity level of the disease determines the appropriate preventive action, as outlined in Figure 6 of the methodology.

A comprehensive comparison was conducted between the proposed perception-based system and existing trained models, yielding valid results. Generating and training the model with the dataset represented a distinct task, demonstrating higher accuracy compared to other tasks. The identification of plant or foliar diseases in a dynamic environment poses considerable challenges, primarily due to complex backgrounds. However, utilizing Deep Learning technology for disease identification proves to be a reliable approach. In the present era, machine-learning algorithms are extensively employed for disease identification, as demonstrated in this section and referenced in [46-56]. A concise overview of these studies is presented in Table 1, providing a summary of the key findings and methodologies utilized in disease identification using machine-learning algorithms as well as their proposed solutions. Some of these studies have employed cross-validation techniques. Surprisingly, these researchers assert their research to be comprehensive. However, it is observed that unless results undergo cross-validation, research is not deemed comprehensive.

In the literature survey, crop disease diagnosis with the help of Artificial Intelligence and Deep Learning tools was also reviewed. Apart from cotton, many diseases are being diagnosed successfully, but there is no mention of the severity of the diseases anywhere, and regarding the diseases of cotton, there is no mention of cotton boll diseases. Hence, both of these aspects were included to increase the uniqueness of this work. What sets this work apart is its specific focus on leaf diseases in cotton crops. A unified technique was successfully employed to identify both cotton boll diseases and cotton pests, which has not been accomplished simultaneously in previous studies. Similarly, a complete and comprehensive model for protecting crops from diseases and pests was proposed, which is rarely found in previous studies and current literature. Additionally, an algorithm for drone movement was developed based on various scenarios.

To address these gaps, the objective is to develop a complete, integrated system. As depicted in Fig. 3, the system will offer a full solution based on an 8-class dataset, providing comprehensive recommendations for each class. If the number of classes is expanded or applied to other crops, the system will adapt to offer relevant solutions.

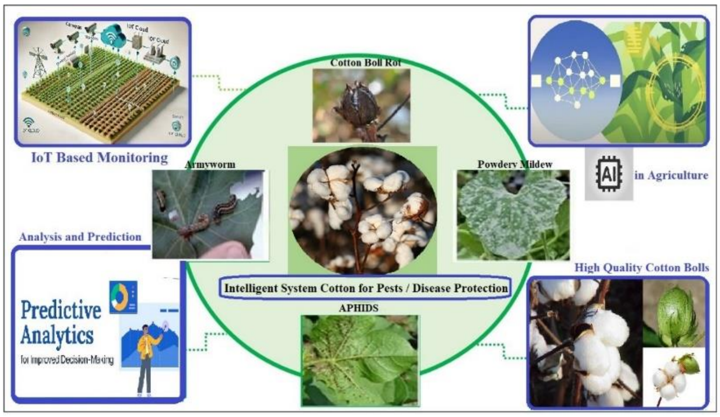


Fig. 3. Overview of the IoT-based automated system for prevention and control for cotton crop diseases and insect pests.

Comparison section with similar studies tabulated in Table 1 and Table 4, researchers have commonly employed CNNs to detect various crop diseases. However, in this study is has used for cotton pests, leaf diseases, and Cotton boll diseases. After reviewing the literature, it is concluded that while researchers in this field, particularly through the use of machine learning, have raised awareness about the presence of diseases, they have not proposed a comprehensive system. There is still a need for an integrated solution that covers all stages, from pest monitoring and disease detection to the eradication of infected plants, as well as continuous monitoring through IoT-based devices. Such an automated system should also include automated preventive measures and provide recommendations to guide farmers.

In this research, the focus is on creating a customized dataset encompassing all three, enabling multifaceted analysis from a single source. The custom dataset expands the scope of the study, as the prior dataset contained fewer classes. By incorporating additional classes, the breadth of the work is enhanced. Augmenting the dataset through techniques like augmentation aims to bolster accuracy and improve the overall quality of the research. Notably, prior research has not addressed cotton boll diseases, and CNNs have not been utilized for this purpose. Cotton crop leaves and cotton bolls exhibit limited symptoms, and the dynamic environment makes disease identification challenging; seasonal variation is considered alongside visible symptoms [11]. A partial dataset was also generated, contributing to this work. The study categorizes identified diseases based on their severity levels, dividing leaf and cotton boll diseases into three distinct stages: initial, intermediate, and final.

The disease detection process is completed by proposing a perception-based intelligent decision system that not only identifies diseases but also includes an alert mechanism to notify farmers of suspicious pest activity and disease identification, enabling timely preventive measures. These measures may involve spraying pesticides or removing infected plants and cotton bolls. Additionally, the system employs drones to spray treatments on affected areas, ensuring the timely eradication of diseased plants.

The diseases of focus in this study include Bacterial Blight, Powdery Mildew, Target Spot, and Cotton Boll Rot. Additionally, the study aims to detect the presence of Armyworms and Aphids on cotton leaves and plants. Fig. 4 visually depicts both Aphids and Armyworms, which the intelligent system is designed to identify, as well as leaves affected by the aforementioned diseases. One key aspect of this study is its ability to provide highly accurate predictions despite working with a relatively small dataset.

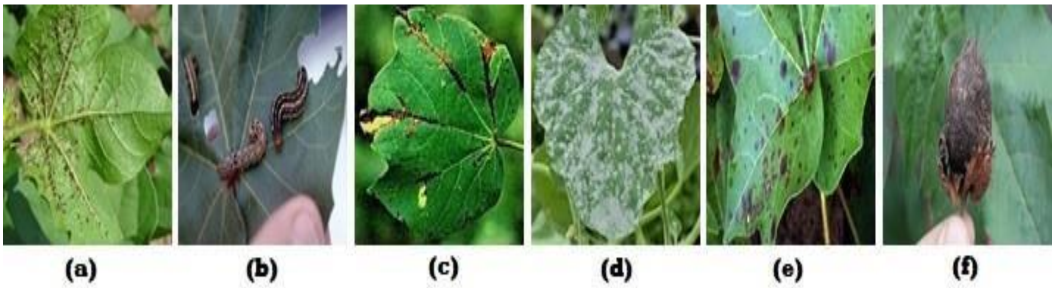


Fig. 4. Cotton disease images a) Aphid b) Armyworm c) Bacterial Blight d) Powdery Mildew f) Target Spot g) Boll Rot

Table 1. Pest-Disease Detection Techniques through AI Methods.

References: Year Tests (F: Field, L: Labs)	Crop	Pests / Disease	Proposed Systems	AI / ML Technologies	Cross Validation	System Scope	Accuracy
Wang, B et al. L 2021	Palm / Coconut	Red Palm Weevil (RPW)	Distributed Acoustic System (DAS) with Fiber Optic Cabling	Convolutional Neural Network (CNN), AI-based Neural Network (ANN)	Not available	Detection	99.9% ANN 99.7% CNN
Bhanu, K.N. L, F 2021	Multiple	Animals	Animal Intrusion Detection System	Support Vector Machine (SVM) Convolutional Neural Network (CNN)	Not available	Detection	Irrigation 90% using SVM IDS 93% CNN
Agarwal M. L 2015	Tomato	Bacterial Pith *Nacrosis*, Early Bligh, Target Spot, White Trail	Weed and Pest Detection System	Machine Vision, Image Processing, Euclidean Distance	Not available	Detection	92% Overall.
Faria F.A et al. F 2014	Multiple	Fruit Flies *Anastrepha*	Automatic Identification System	Machine Learning, Multimodal Fusion	5-Fold Validation	Identification	98%
Silva, D.F L, F 2015	Flying Insect Species	Species of Mosquitoes. *Aedes Aeygipti, Anopheles gambiae, Culex quinquefasciatus*	Novel Sensor and Tool to Control Disease Vector	kNN, SVM, Random Forest (RF)	2-Fold Validation	Pests and Disease detection	90%
Liu, Z. F 2022	Tea Plant *(Camellia sinensis)*	Blister Blight (*Exobasidium vexans)* for Tea Plant	Prediction Model for Disease	Multiple Linear Regression (MLR) Model with Python	Not Available	Disease detection	91%
Titya M.D L, F 2017	Cotton	Multiple Pests and Diseases	Ontology-based Prediction Model	Ontology, Semantic Web, Expert System	Not Available	Pests and Disease detection	NA

Sharma, R.P F 2023	Millet Rice	Pink and Millet Stem Borers, Gundhi, Chinch, and Green Bugs	Complete Pest Monitoring System	Fuzzy Inference System (FIS)	Not Available	Proactive measurement system	98.21%
H. Li et al. L 2011	Cotton	Multiple	Web-Based Intelligent Diagnostic System	Neural Network, Web Technologies	Not Available	Diagnostic System	89.5%
Xiao et al. L 2018	Cotton	Multiple Pests and Diseases	Diagnostic System Proposed for Disease and Pests	Recurrent Neural Network (RNN) Long Short-Term Memory (LSTM) Architecture,	Not Available	Detection / Diagnostic System	96% 91% 86%
Saleem R.M F 2021	Cotton Wheat	White Flies (*Bemisia tabaci)*	Insect Pest Prediction System (IPPS)	Neural Network Radial Basis Function Network (RBFN)	Not available	Pests Prediction System	82.88%
Nadeem R. M et al. F 2021	Sugarcane	Stem Borer	Stem Borer Attack Prediction System	Machine Learning / Naïve Bays	Not Available.	Prediction System	91%
Sharma, R.P et al. F 2021	Rice and Sugarcane	Pink and Millet Stem Borers, Gall Midge, Brown Plant Hoper Gundhi, Chinch	Inference System	Fuzzy Logic, Genetic Algorithms	Not Available.	Prediction System	NA.
Shital B et al. L 2014	Multiple	Multiple	Identifying Method of Plant Disease	Image Processing, Color Histogram, CANNY Edge Detection	Not Available.	Identification System	NA
Dong, Y. et al. L, F 2020	Cotton	Multiple Disease	Cotton Disease Detection System (CDDS)	Fuzzy Logic, Case-Based Reasoning (CBR)	Not Available.	Detection System	Above 90%
Pooja, V L, F 2021	Multiple	*Phytophthora* and Other Insects	Review of Systems	Multiple Technologies	Not Available.	Early Stage Detection	NA
Ramalingam et al. F 2020	Multiple	*Flying and Crawling Insects*	Deep Learning-based IoT Framework	IoT, Faster Region-based Neural Network (RCNN), Deep Learning	10-Fold Cross Validation	Trap Monitoring System	94%
Gawande, A. R. L, F 2020	Multiple	Multiple	IoT-based Monitoring System	Random Forest (RF) along with Support Vector Machine (SVM)	Not available.	Detection System	NA
Chen W. L L, F 2020	Rice	Rice Blast Disease	Rice-Talk-An IoT-based Framework	AI, Machine Learning, Image Processing	Not available.	Detection and Identification system	89.4%
Khattab et al. L, F 2019	Multiple	Epidemic Disease	IoT-based Monitoring System	Expert System, Prediction Algorithm,	Not available.	Detection System	NA

A. Thorat et al. L 2017	Multiple	Powdery Mildew, Black Spot, Leaf Spot, Botrytis Disease	Web Application for Disease Detection	K-Means Clustering, Image Processing,	Not Available.	Detection System	99%
Parikh et al. L 2016	Cotton	Grey Mildew	Detection System with Algorithm	Image Processing Technique	Not Available.	Detection System	82%
Chopda J. et al. L 2018	Cotton	Multiple Disease	Temperature, Soil moisture-based Android Application for Diseases.	Decision Tree Classifier-based Predictions.	Not Available.	Disease Detection System	NA
S. Phadikar et al. 2012	India	Rice / Millet	Fermi Energy Based region extraction method	Rule base Genetic Algorithm applied	10-Fold Cross Validation	Detection and Identification System	90%
Memon, M.S. et al. F 2022	Cotton	Target Spot, Leaf Spot, Nutrient Deficiency, Powdery Mildew, and Verticillium Wilt	Meta Deep learning strategy.	CNN Detection Model Proposed	Not Available.	Disease Detection System	98%
Varun S. et al. F, L 2021	Cotton	Bacterial Blight, Target Spot, and Cercospora	Image Processing by using CNN.	CNN-based detection Model Proposed	Not Available.	Detection System	96%
Sharvari V. Patil et al. F, L 2022	Cotton	Sucking and Chewing pest Disease, Bacterial Blight, and Curl Virus	Image Processing-based CNN	CNN Model	Not Available.	Disease Detection System	90%
G. Harshitha et al. F 2021	Cotton	Gray Mildew, Bacterial Blight, Alternaria, Leaf spot, and deficiency of Magnesium	CNN (ResNet, LSTM)	LSTM Model	Not Available.	Disease Detection System	97%
Rajasekar V. et al. 2021	Cotton	Cotton Plant Diseases	CNN	(ResNet-50)	Not Available.	Detection System	95%

3. Materials and Methods

According to a recent literature survey, IoT and Artificial Intelligence experts are designing tools and applications for agricultural fields so that our Entomologists and Agronomists can benefit from their usefulness. Below are some recent researches in which computer experts have done very useful work with the help of technology in agricultural fields as well. Nowadays, all around the globe, efforts are underway to enhance agriculture through the utilization of IoT, Artificial Intelligence or Machine Learning technologies that are helping to meet the growing food needs.

1. Convolutional Neural Network

In recent days, several machine learning algorithms have been employed by researchers to learn about crop pests and related diseases and to develop preventive measures. Examples include Support Vector Machine [16] while k-nearest Neighbor (k-NN) [17], Random Forest [18] and Convolutional Neural Network (CNN) [19]. While these algorithms faced challenges in recognizing the presence of a face in an image, CNN has exhibited exceptional performance, setting it apart from other algorithms. CNN excels in processing all the pertinent features for disease detection within each layer, as depicted in Fig. 5. The architecture of CNN is specifically designed for complex image recognition tasks, rendering it highly efficient and reliable. Given the diverse symptoms associated with different crop diseases, CNN enables the extraction of relevant symptoms through layered processing. It comprises multiple convolutional layers, each consisting of image filters that convolve with images or feature maps, alongside other accompanying layers [20, 21, and 22]

.

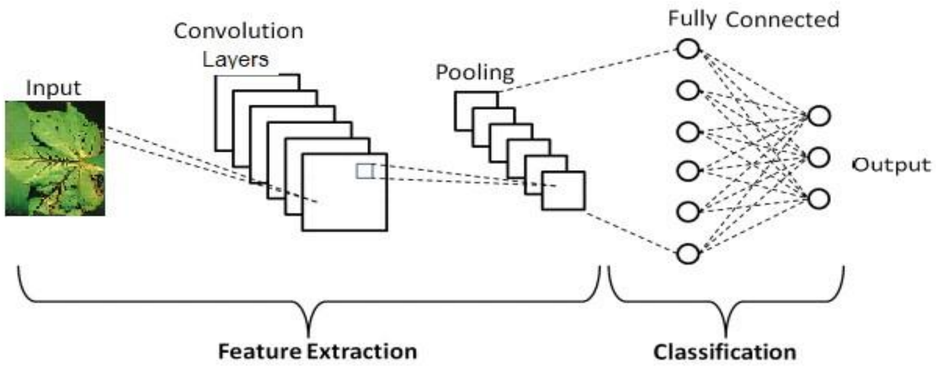


Fig. 5. General CNN Scheme for disease detection

3.2. Transfer Learning in Deep Learning

Transfer learning [23] is also a unique way of working that is used in a few fields other than machine learning. In this, if a model is created for a particular task, it can be used to initialize another task. This concept is widely employed in deep learning, where pre-trained models are leveraged to tackle new issues and tasks, such as natural language processing, image analysis, and computer vision. By adopting transfer learning, the need to build models from scratch, which requires significant computational resources and time, is avoided. Fig. 6 illustrates the specific transfer learning approach employed in the model.

In this research, the objective is to enhance the optimization of the proposed model by integrating pre-trained CNN models. A thorough analysis and comparison of the outcomes achieved by these pre-trained models with the results of the proposed model are conducted. This comparative analysis allows an assessment of how transfer learning enhances the model's capabilities in terms of effectiveness and performance. Table 1 provides detailed information on the features and working methods of these technologies. While some models focus on identifying specific diseases or insects using particular images, this approach employs a dataset that facilitates the identification of multiple diseases. To expand disease identification further, two additional classes have been introduced in the dataset, with plans to incorporate more classes in future iterations.

One limitation of dataset acquisition is the inability to gather data from numerous locations simultaneously; as a result, data must be collected from specific locations. Despite this, the dataset remains non-specific to any single area. It was obtained from a dataset repository representing a part of the subcontinent and supplemented with local data, including two additional classes. Consequently, the dataset is less generalized, encompassing data from various regions. Furthermore, future work will focus on making the dataset more generalized

.

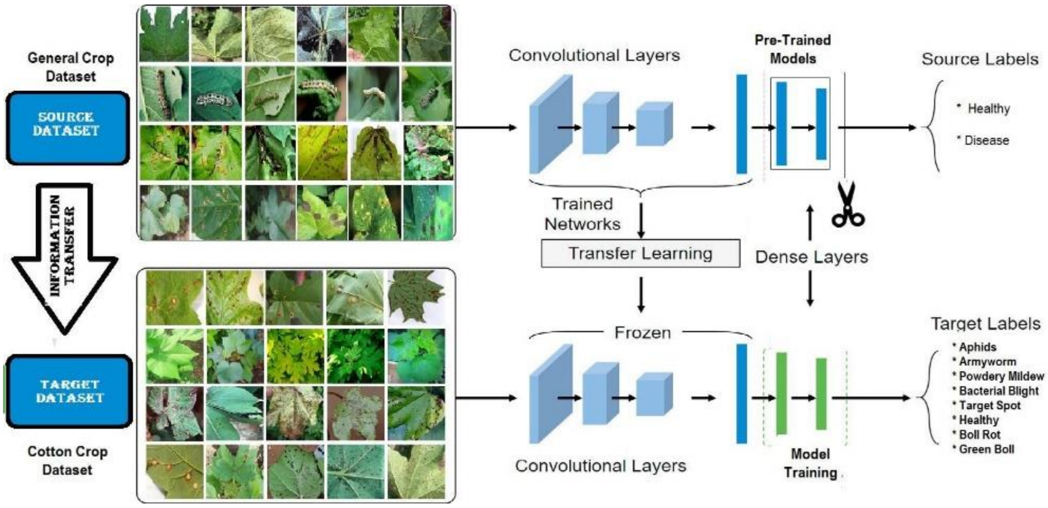


Fig. 6. The Architecture of our Transfer Learning Model

4. Proposed Cotton Disease Identification Methodology

Among all deep learning models, the Convolutional Neural Network (CNN) has a unique advantage in that it is capable of performing tasks that other models cannot. For example, it can extract the features of any image using convolutional layers, which other models cannot extract. This model consists of three primary and fundamental layers. Among these three layers, the Convolutional Layer, which is considered the first layer, works to take images as input. While the second layer is called pooling and the last layer is known as the fully connected layer. The convolutional layer also includes five additional layers. Operating within a grid-like topology, CNN represents a specific class of Artificial Neural Network (ANN), with the Convolutional Layer assuming a crucial role in performing vital computations.

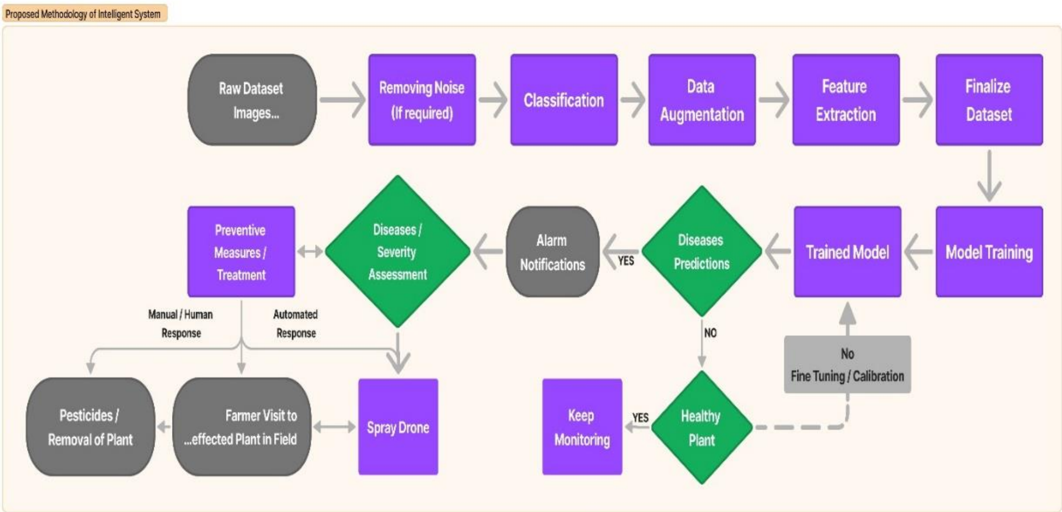


Fig. 7. Workflow of Proposed Model for Cotton Pest/Disease Prediction.

In the scope of this study, cotton diseases are identified by training the raw dataset images using the convolutional process. The proposed methodology of the research is illustrated in Fig. 7, covering all stages from dataset acquisition to disease prediction, identification, alarm notifications, and the provision of preventive measures. By utilizing this comprehensive approach, accurate disease identification within the dataset is aimed for. Based on the severity of the disease and the level of pest infestation, the intelligent system provides farmers with recommendations on the appropriate actions to take.

4.1. Dataset

A different type of dataset is used in which some work has already been done, and some datasets have been created independently. A cotton disease dataset, with a total of six classes, was obtained from Kaggle [57]. The existing dataset was modified, adding two classes named Cotton Ball Rot and Green Cotton Balls. Pictures were obtained from various sources, including fields in Sindh province with the help of IoT-based sensors and cameras, as well as from pesticide companies. In the second step, some images were enhanced and named accordingly. These two classes were added to the Kaggle [57] dataset to allow for more disease classification (see Fig. 8). The custom dataset comprised a total of 7289 images, with 90% designated for training and the remaining 10% for testing and validation. Notably, when forming the dataset, images representing various disease severity levels within the classes were included. This addition aims to enhance the system's ability to predict the severity of diseases. While certain diseases are easily detectable due to their pronounced characteristics, others may pose challenges for the system due to their milder nature. The dataset is currently limited in its generalizability, and further refinement is needed to enhance its broader applicability. This refinement is a part of future plans and intentions.

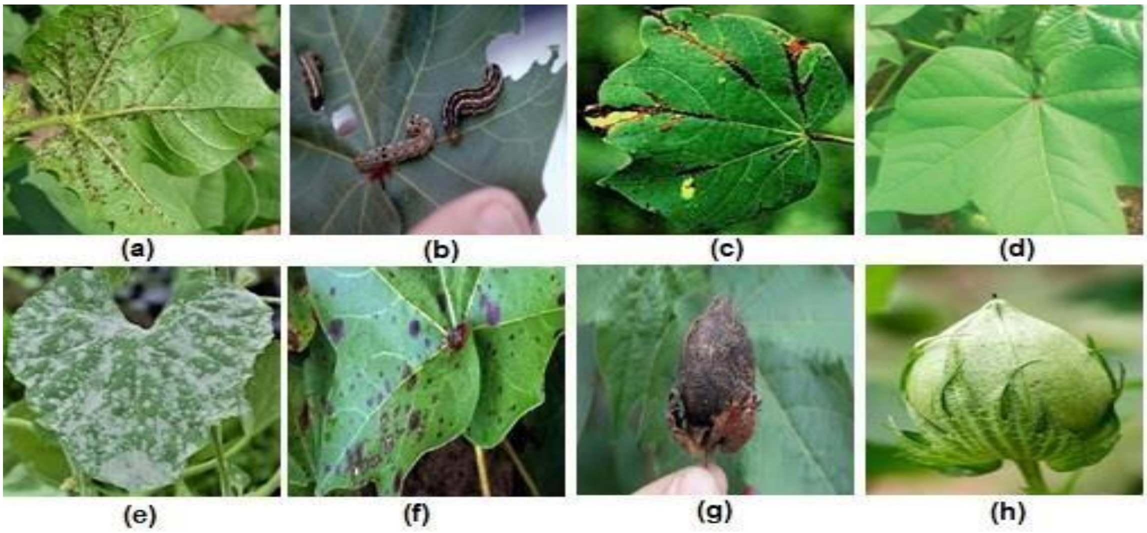


Fig. 8. Dataset Classes a) Aphids b) Armyworm c) Bacterial Bligh d) Healthy e) Powdery Mildew f) Target Spot g) Boll Rot h) Green Cotton Boll

4.2 Data Collection Methodology

Various methods can be employed to collect data, including the use of IoT devices such as sensors and cameras in the field, as illustrated in Fig. 9. These devices can capture images of cotton crop leaves and flowers at different times of the day. This method allows for gathering a vast number of images from various harvest periods within a few days. Another approach involves manually taking pictures during different harvest periods with cameras, although this method is labor-intensive and time-consuming. Alternatively, drones can be deployed to capture images of the crops at different times, though this method might not achieve the same level of accuracy. By utilizing these methods, an extensive dataset can be created to fulfill data requirements. Additionally, this system can be adapted to various types of fields to collect data efficiently.

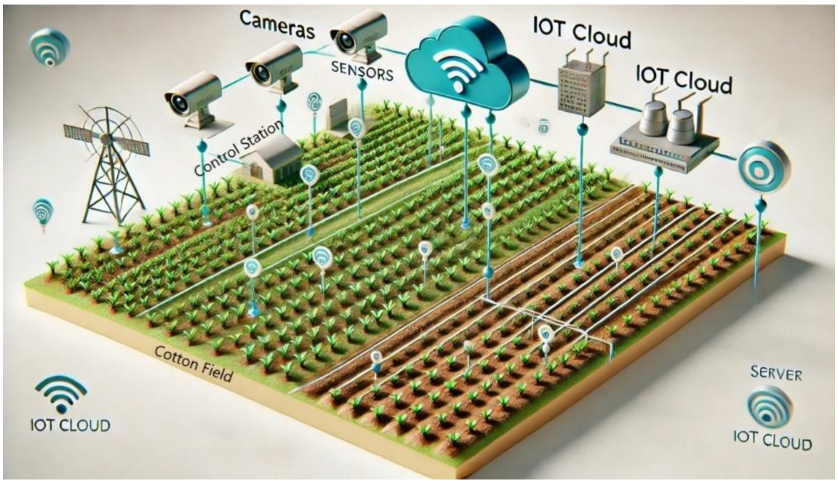


Fig 9. Data Collection Methodology for Cotton dataset.

4.3. Data Processing

Image pre-processing is a technique aimed at improving an image by suppressing distortion, enhancing its features for further critical tasks. An array of these filtered images is created to apply deep learning techniques, such as CNN, k-NN, etc. The specific technique used on this data depends on the results provided by each method. The most appropriate technique, delivering the best results, will be employed.

4.4. Augmentation

In the third step, augmentation techniques were implemented on the dataset, as the number of images was considered small. The number of images for the newly added classes was kept in line with the existing classes. It is known that a small number of images for model training can lead to overfitting. To address this, the dropout function was also used to further optimize the data for prediction by removing some nodes using the dropout filter. The dataset, initially containing about 2508 images across eight classes, was expanded to 7289 images after applying the augmentation technique.

4.5. Feature Extraction

The symptoms and characteristics of crop diseases are crucial factors in disease diagnosis. Unique features emerge in the leaves and fruits of plants when a disease is present, allowing us to identify the specific disease. Similarly, the presence of agricultural insects in crops can be determined by observing the distinct features of these insects. Machine vision and convolutional neural network techniques utilize these distinctive features to diagnose diseases. During the creation and training of our system's dataset, it is ensured that all images were of the same size (K x K blocks) to extract various features. Multiple features such as leaf and cotton ball color, size, texture, presence of black or white spots, and more were employed for machine learning. These features were also utilized to extract background information, ensuring the accuracy of the proposed system's results. While our model does not convert images to grayscale, it is possible to convert certain color images to grayscale to extract specific features. Additionally, the grey level co-occurrence matrix, a common method in pattern recognition for grayscale images, is calculated. These specific features enable the machine to perform segmentation, a critical process that divides images into different parts to obtain desired objects or results.

4.6. Model Training

The subsequent step involved training our model using the augmented dataset to improve its performance and predictive capabilities. For optimized model optimization, The Adam optimizer, known for its effectiveness in deep learning tasks, was employed. The Rectified Linear Unit (ReLU) was chosen as the activation function due to its widespread adoption and proven effectiveness in various applications. The criticality of meticulous model training in achieving accurate disease identification was recognized. Thus, the training process was approached diligently, incorporating both training and validation stages to ensure the model's precision and efficacy. Table 2 provides an overview of the distribution of images used for training, testing, and validation within our model.

Understanding the usefulness of a dataset is crucial, and utilized the k-fold cross validation technique to assess its efficacy. The dataset is divided into k subsets, with k-1 subsets allocated for training and the remaining subset for testing. This procedure evaluates dataset accuracy and other parameters to ensure model quality and guard against bias in image representation within the dataset.

Table 2. Training, Testing and Validation Image.

S. No.	Disease Classes	Training	Validation	Testing
1	Apids	820	40	44
2	Armyworm	840	42	37
3	Bacterial Bligh	850	40	40
4	Cotton Boll Rot	810	58	62
5	Green Cotton Boll	810	45	44
6	Healthy Leaf	827	40	40
7	Powdery Mildew	810	48	48
8	Target Spot	810	42	42
	Total Images	**6577**	**355**	**357**

4.7. Proposed Convolutional Neural Network Methodology

In this proposed method, the process begins with the preprocessing and dataset preparation phase. The images are then fed into the convolutional neural network (CNN), which acts as the network for training the model to recognize different cotton diseases and pests. The first convolutional layer acts as a filter, which is applied to an image to capture its unique features, locations, and strengths. After this work, batch normalization is applied between the two convolutional layers to enhance both the network's speed and accuracy, as well as normalize the input layer. Just after this, the pooling layer starts working and it reduces the size of the filtered image. It is capable of using max, min, and average pooling and these three pools have their advantages and properties.

To mitigate the risk of overfitting in the network, a dropout layer is incorporated. This layer randomly ignores neurons during the training process. A fully connected layer establishes connections between the neurons of one layer and those of the next layer, with each neuron assigned a distinct value that contributes to the final decision. In our model, this layer contains 256 neurons. Finally, the output layer processes the input based on the weighted values of the previous layer's neurons. See Fig.10.

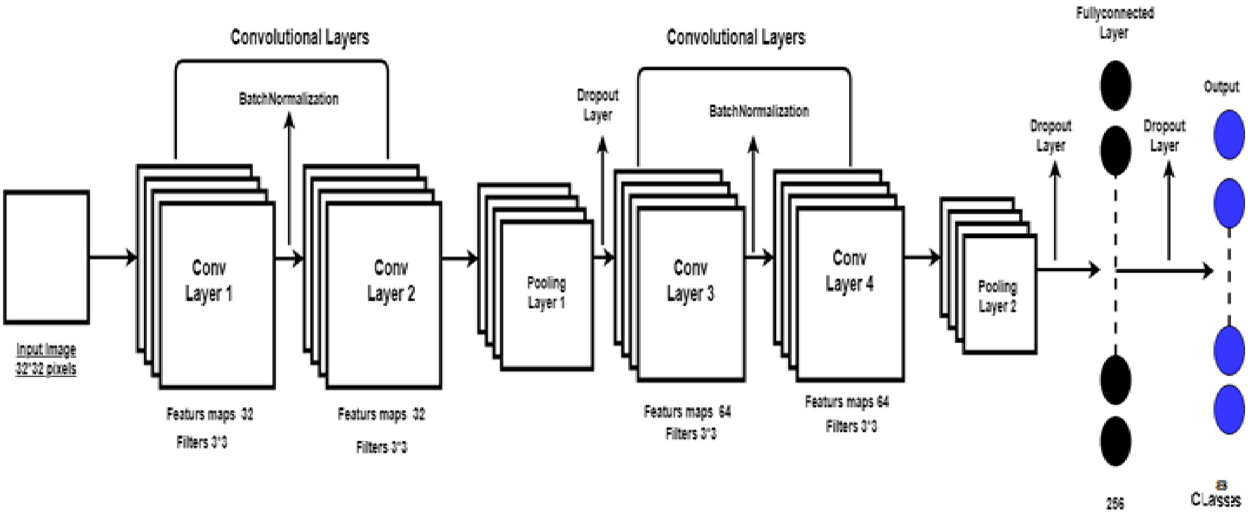


Fig. 10. Proposed CNN internal block diagram.

Now the next few sections cover the steps that our created model is performing. The entire process and the details of what the customized model is doing in the backend will be explained.

4.7.1. Input image and formatting

In the initial stage, the initial layer is presented with an image of dimensions WxHxD, where W represents width, H represents length, and D signifies the depth or the number of channels within the image. In this case, the image has 32 x 32 x 1 pixels. A pickling technique was used that converts a pixel-based image into a character-based stream. This technique makes it easy to save the data on the disk while retaining almost all the information of the data. With the help of this pickling technique, the pixel data or model can be trained on Google GPU. Google's TensorFlow Processing Unit (TPU) is specially designed to easily train large amounts of data, which is very difficult on a local machine.

For this work, Jupyter Notebook was used. The images were transformed into X-pickle and Y-pickle formats. X-pickle contains the image dataset as a byte stream, while Y-pickle holds the class labels. The model was trained using Google Colab, where X-pickle and Y-pickle were loaded and displayed in image format along with the corresponding class labels.

4.7.2. Convolutional Layer

The initial layer in our model is the convolutional layer. It works by taking images as input. Here, the images are taken as MB by M input and the N-by-N filter is applied. The result of the convolutional layer will be (MxN+1) x (MxN+1). This layer divides the image into overlapping matrices through a Kernel; it is generating the filter map and extracts features from the image such as edges, corners, and endpoints.

4.7.3. Normalization and Pooling Layer

As mentioned earlier, batch normalization is used to speed up the processing speed of the network and for training accuracy, here it is also used for the same purpose. It is also used here to standardize the input to a layer for every mini-batch.

A pooling layer has been used to reduce the dimensions of any image. There are various types of polling available such as Min, Average, and Max polling. In this research, max pooling is implemented, which entails choosing the highest value within each sub-matrix and generating a matrix of maximum values. The output size from the pooling layer is given by following equation (1).

$$\mathbf{(M - N + 1)/2} \tag{1}$$

4.7.4. Dropout Layer

To address the issue of overfitting in proposed model, a dropout layer included. Dropout randomly ignores neurons during the training process, preventing the ConvNet from overfitting.

4.7.5. Fully connected layer

After multiple convolutional layers and pooling layers, a fully connected layer is introduced, that takes the data and features of all the previous layers as input and produces the desired output. The flattened matrix from this layer is also passed through the SoftMax classifier, which computes the probability of each class and selects the class with the maximum probability, representing a disease or pest in our case. Parameter-tuning techniques were extensively employed to optimize the performance of the model. Multiple iterations of fine-tuning were conducted, resulting in an impressive accuracy of 92%. Parameters such as augmentation, color, image contrast, dropout, and epoch were systematically adjusted to meet specific requirements and enhance overall results. The training process for the dataset took approximately 40 minutes, with a total of 154 layers in the base model.

4.7.6. Diseases and Pest Predictions

In the context of disease and pest predictions, our experimental approach involves leveraging the CNN network to achieve accurate predictions. The trained model has demonstrated remarkable proficiency in this task. By exploring different numbers of Epochs, substantial variations in the results has been observed. Furthermore, the introduction of previously unseen images into the model's training further enhanced its disease and pest prediction capabilities. Notably, our model achieved exceptional prediction rates of up to 99% in certain classes. The accuracy of both Training and Validation can be observed in the graphs presented in Fig.11. These graphs depict a high level of accuracy in the model training, reaching approximately 92% with minimal discrepancies or errors.

The disease patterns predicted by our model are visually illustrated in Fig.12. The model's lower accuracy in certain classes can be attributed to insufficient images available for training. To address this limitation, a more comprehensive dataset specifically targeting these classes will be acquired during the upcoming cotton crop season. The confusion matrix of the proposed model is displayed in Fig.13. Out of 357 images, our proposed model correctly predicted 336.

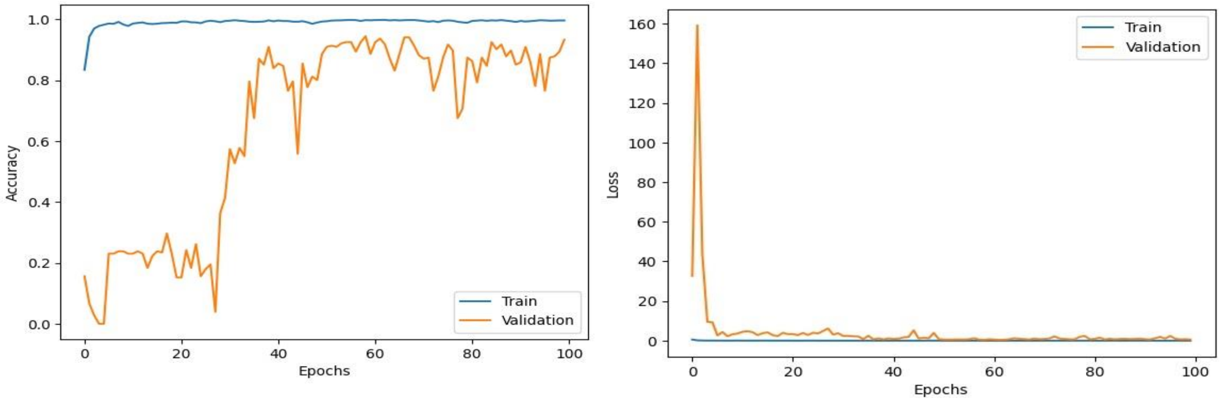


Fig. 11. Training and Validation Accuracy and Losses of Proposed Model.

Fig. 12. Correctly predicted images of different diseases.

A comprehensive response model is being developed for this intelligent system, which will include a complete and integrated response system triggered by the output of the model. This system will involve notifying the farmer of the disease presence at the field coordinates, prescribing the appropriate medicine to the inspector for timely application on the crop, and dispatching assistance to the affected plant in the field. A contingency plan is also incorporated in the proposal

4.7.7. Performance Analysis of Tested Algorithms

It is essential to note that our research is conducted within the context of our specific conditions, while other researchers have approached the problem based on their circumstances. Numerous pre-trained models are available for comparison with our proposed model. To evaluate their accuracy and results using the same dataset, VGG [58] and InceptionV3 [59, 60] have been chosen as two alternative models in the transfer learning approach. In addition to these models, there are many pre-trained models available such as ResNet [61] and LeNet [62] etc.

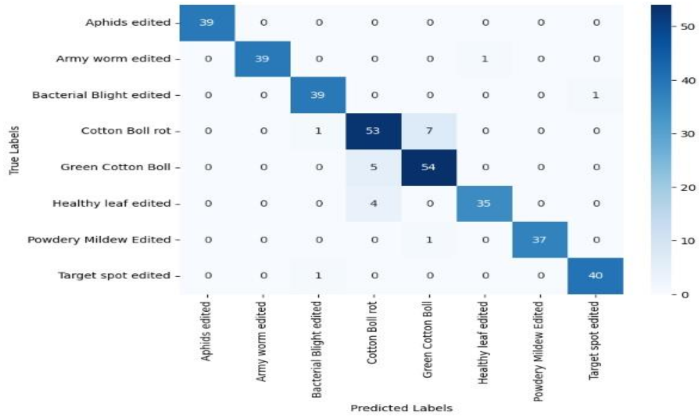


Fig.13. Heat map of Proposed Model.

Using an equal number of images for testing, our customized proposed model achieved an impressive 92% accuracy in predicting the images. Fig.12 visually displays the correctly predicted images and their corresponding classes. Notably, our proposed model surpasses InceptionV3 and Compete with VGG16, as their accuracies and validities are very similar to our model's performance.

Fig.14 illustrates the accuracy and validity graphs of InceptionV3, while Fig.15 presents the accuracy and validity values of VGG16. Furthermore, Fig.18 provides a comparative analysis of the accuracies achieved by all evaluated models along with a summary of results in Table 3. With a training accuracy of approximately 93% and a validation accuracy of 93%, our proposed model demonstrates good performance compared to both InceptionV3 and VGG16, despite the relatively smaller size of our dataset.

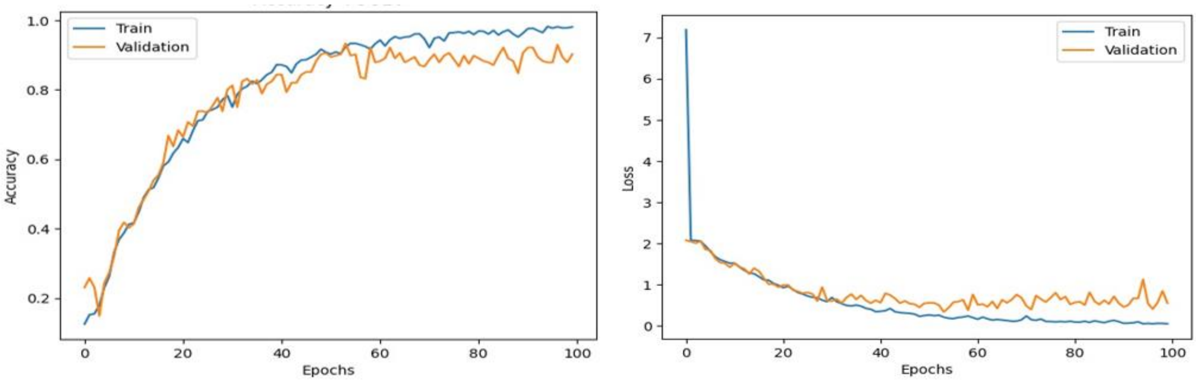


Fig. 14. Training and Validation accuracy and losses of InceptionV3

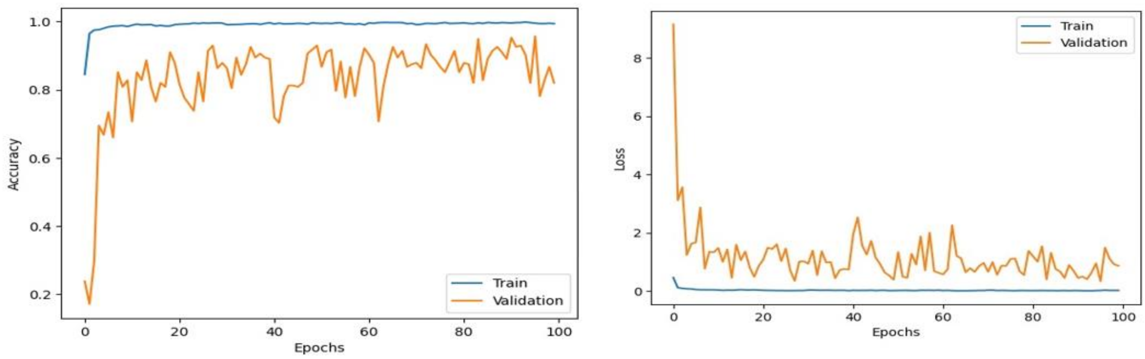


Fig. 15. Training and Validation accuracy and losses of VGG16.

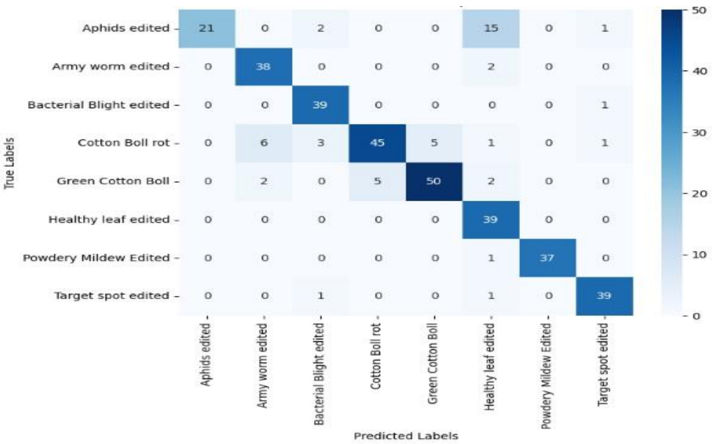


Fig. 16. Heat map of InceptionV3 Mode.

It is worth mentioning that InceptionV3 and VGG16, while regarded as established models, fail to outperform our proposed model in terms of accuracy, thereby validating the effectiveness of our approach. In Fig. no. 16 and 17, the confusion matrix of both models InceptionV3 and VGG16 is presented.

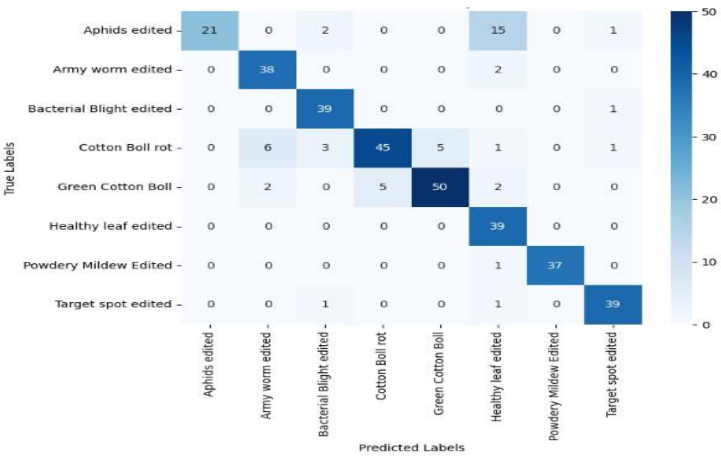


Fig. 17. Heat map of VGG-16 Model.

After training and evaluating all the models, as shown in Table 3, it is apparent that the outcomes of one of the pre-trained models and our customized model are comparable. The accuracy of their results is nearly identical. This indicates that pre-fitted models do not necessarily produce the same results for all datasets. These models were created for a specific dataset and may produce similar results on the same dataset or another specific dataset. However, their usefulness may vary across different datasets. This is why people resort to using customized CNN models and achieve favorable outcomes by training their models with their specific data sets. Compared to the Inception model, our model demonstrates high accuracy and produces excellent results.

A limitation of pre-trained models is that their default layers cannot be modified, resulting in limited control over their results. In contrast, significant control is available over certain parameters of customized models, they are still capable of producing satisfactory results on our data sets. Apart from the default layers, many other factors affect the training, accuracy, and results of any model.

Fig. 18. Training and Validation accuracy graph of our proposed Model, InceptionV3 andVGG-16.

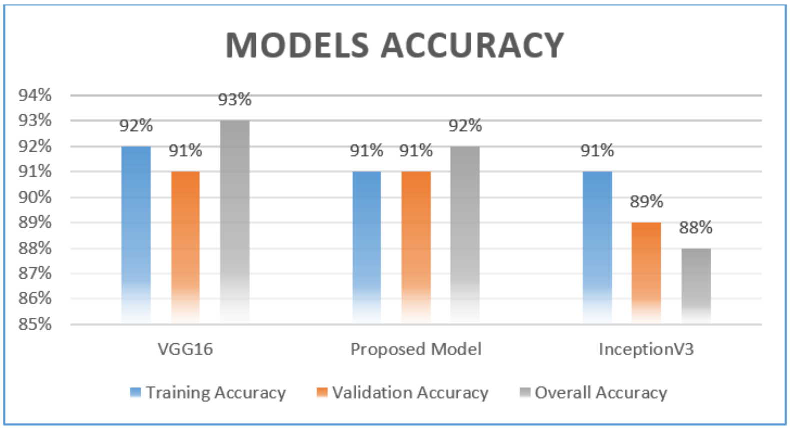


Table 3. Result summary of different Models and their Accuracy.

Model	Architecture	Number of Layers	Parameter Size (approx.)	Receptive Field Size	Accuracy
VGG-16	Deep (16 Layers)	16	138 Million	224x224	93%
Proposed Model	Very Deep (50 Layers)	50	25.6 Million	224x224	92%
InceptionV3	Deep (48 Layers)	48	23.8 Million	299x299	88%

Tuning of any model also plays a crucial role in obtaining good results. Some pre-trained models are well-tuned, while others require a specific time for tuning. On the other hand, customized models offer more flexibility for tuning. For this reason, these customized models are capable of producing accurate results on certain data sets. Something similar happened in this case where our models started to produce good results at lower epochs on our dataset. However, running the pre-trained model for more epochs could have also yielded good results. The underlying structure of any model also plays a significant role in producing accurate results. Now, it depends on the type of dataset for which the model was primarily designed. Some models designed to recognize faces may not perform as well on vegetation datasets as they do on face datasets. These models may need further tuning or an increase in their epochs to achieve good results.

K-fold cross validation technique employed to validate model results. To provide a proof of concept, A sample was taken to perform this cross validation. From the sample, select k=10 subsets and implement k fold cross validation with 9 training and 1 testing subset which is re order in each fold. It is observed that in k-fold cross-validation, the accuracy typically falls between 80 and 90 percent. Specifically, some cases reach 83 percent, and others reaches 85 percent, and some even 87 percent. Given the small sample size used in this experiment, the level of accuracy is deemed acceptable. In the future, a comprehensive cross-validation across the entire dataset will be conducted.

Firstly, proposed model demonstrates superior performance compared to a pre-trained model in both training and testing, showcasing a significant achievement in itself. Secondly, although the accuracy of suggested model increases with additional epochs, training was capped at 100 epochs to maintain consistency with other models for comparison. Thirdly, images were incorporated into the dataset based on disease severity levels see Figure 2, allowing for testing across three distinct stages. This unique aspect of our research enables easier identification of diseases in advanced stages, contrasting with the challenges posed by milder symptoms at earlier stages. Lastly, our model requires considerably less time for training and testing compared to others. While some researchers omit time data, our model consistently delivers superior results in less time.

Table 4. Comparing our proposed approach in contrast to other Machine Learning based cotton disease detection techniques.

Author and Year	Classes of Dataset	No. of Images	No of Diseases and Pests	ML or DP Technique	Computational Time	Results Accuracy and Loss
Memon, M.S. et al. [46] 2021	Own Dataset 7 Classes	2384 (Approx. 340 images in each Class)	**6** 1. Leaf Spot 2. Target Spot 3. Nutrient Deficiency 4. Powdery Mildew 5. Verticillium Wilt 6. Leaf Curl	CNN	Average	98 % Accuracy 0.1 Loss
Varun S. et al. [39] 2021	Own Dataset 4 Classes	Not Mentioned	**4** 1. Bacterial Blight 2. Bronze Wilt 3. Curly Lift 4. Fouler Fungal Disease	CNN	Not Mentioned	96% Accuracy 0.24 Loss
Sharvari V. Patil et al. [63] 2022	Own Dataset 4 Classes	1752 (Approx. 440 images in each Class)	**3** 1. Sucking and Chewing pest Disease 2. Bacterial Blight 3. Curl Virus	CNN	Average	90% Accuracy 0.2 Loss
G. Harshitha et al. [64] 2021	Kaggle [1] Dataset for Cotton Diseases	Not mentioned	**6** 1. Bacterial Blight 2. Alternaria Leaf Spot 3. Gray Mildew 4. Magnesium Deficiency 5. Cercospora 6. Fusarium Wilt	CNN (ResNet, LSTM)	Not Mentioned	97.13 % Accuracy Not mentioned.
Rajasekar V. et al. [65] 2021	CIFAR-10 Dataset (Generalized Dataset)	60000 Images (Different types of Images not limited to Cotton)	Not mentioned Cotton Plant Diseases (They did not mentioned disease name)	CNN (ResNet-50)	Not mentioned	Training accuracy 95% Validation Accuracy 98% Training Loss 0.33 Validation Loss 0.5
Latif, M. R et al. [66] 2021	Own Dataset 3 Classes	Original Images are 95 (After augmentation 1000 images for each class were used)	**3** 1. Mildew 2. Leaf spot 3. Soreshine	ResNet101 (Cubic SVM used for Validation)	Best	98 % Accuracy Loss – Not mentioned
Azath M. et al. [45] 2021	Own Dataset 4 Classes	2400 (600 for each Class)	**4** 1. Bacterial Blight 2. Leaf miner 3. Spider mite 4. Healthy	CNN	Average	96.4% Accuracy Training Loss 0.2 Validation Loss 0.3
Gülmez, B.[67] 2023	Own Dataset 4 Classes	2057 total images with different no. of images in each Class	Not mentioned Cotton Plant Diseases (They did not mentioned disease name)	CNN Grey Wolf Optimization	Average	96 to 100% Accuracy without Training Loss
Proposed Perception-based Decision Making System Methodology	**Kaggle Dataset + Own Dataset (Two Classes)**	**2508 Original Images, 350 for each Class.** **(7289 Augmented Images)**	**8** **1. Aphids** **2. Armyworm** **3. Bacterial Blight** **4. Powdery Mildew** **5. Target Spot** **6. Cotton Boll Rot** **7. Green Cotton Boll** **8. Healthy Leaf**	**CNN InceptionV3 VGG-16**	**Best**	**93% Accuracy Training Loss 0.1 Validation Loss 0.2**

4.7.8. Pests / Disease Detection: Comparison with Similar Studies

A comparison of recent research using CNNs on cotton diseases and pests is provided in Table 4. If researchers compare our research with the current researches, it is very different from the rest of the studies [47, 63-68]. In this study, datasets of cotton pests, foliar diseases, and cotton bolls are used. In contrast, they have only worked on leaf disease datasets. Some results are better than our results, the accuracy of their work is up to 97%, while the accuracy of our work is coming to 93%. This may be because cotton leaf diseases are comparatively easier to identify through their images as the leaf disease is very noticeable and observable. On the other hand, the identification of diseases of cotton balls and cotton pests is much more difficult than the identification of leaf diseases. Their training losses were also low, while our models made more losses during training. Computational time is also very important and our proposed work time is better than others.

It is also mentioned in our gap analysis above that existing models do not necessarily give the same results on every dataset as they did on the dataset for which they were created. Proposed model has identified pests, foliar diseases, and diseases of cotton bolls using the same dataset while their dataset consisted of only leaf images. This is also a reason why our training losses are high and the accuracy of the results is somewhat less compared to them.

Table 5. Performance of non-Deep Learning Approaches based cotton disease detection techniques.

Researcher / Year	Disease or Pest	Proposed System	Technology	Results Available
[40] L, F 2017	Multiple Pests and Diseases	Ontology-based Prediction Model	Ontology, Semantic Web, Expert System	NA
[27] L, 2011	Multiple	Web-Based Intelligent Diagnostic System	Web Technologies	89.5 %
[28] F, 2021	White Flies (*Bemisia tabaci)*	Insect Pest Prediction System (IPPS)	Radial Basis Function Network (RBFN)	82.88%
[30] L, F 2020	Multiple Disease	Cotton Disease Detection System (CDDS)	Case-Based Reasoning (CBR)	Above 90%
[43] L, 2016	Grey Mildew	Detection System with Algorithm	Image Processing Technique	82%
[68] L, 2018	Multiple Disease	Temperature, Soil moisture-based Android Application for Disease Detection.	Decision Tree Classifier-based Predictions.	NA
[45] L , 2014	Bacterial Blight, Alternaria	Graph Cut-based Approach System for Selected Cotton Diseases.	Gaussian Filter Image Processing.	NA.

Accuracy is also influenced by the image quality. Our dataset's image quality may not match that of other researchers' studies. However, our dataset had more images and increased their number by using the augmentation technique. Training also plays a very important role in machine learning algorithms. Maybe there is some deficiency during training that the accuracy of our results is less than the accuracy of other researchers. It is aimed to remove all these shortcomings in our next step and further refine this model with better images, training, and better accuracy.

While this work primarily emphasizes technical aspects, it's crucial to consider its commercial implications. If farmers embrace our system, they stand to benefit significantly: both crop quantity and quality are likely to improve, leading to increased income. Furthermore, by adopting this system, farmers can reduce expenditure on medications, thereby enhancing their profitability.

The field of cotton disease identification has indeed seen relatively limited research, with a predominant focus on deep-learning technologies. However, our study seeks to broaden this perspective by comparing various technologies, not solely confined to deep learning. Through Table 5, the performance of conventional non deep learning methods in identifying cotton disease has been presented. The comparison of these methods with our proposed results in table 4 reveals that our approach yields better output. This analysis underscores the effectiveness and novelty of our model within the context of cotton disease identification.

4.8 Automated Response Mechanism

As outlined in the methodology, this system integrate with image analysis tools and spraying drones to deliver targeted treatments (e.g., pesticides, fungicides) based on the detected severity level and specific location. Additionally, proposed a recommendation engine that suggests specific treatments or preventive measures depending on the disease type, severity, and presence of pests. If the system detects a high severity level, it can recommend a farmer's visit or the removal of infected plants to prevent further spread. Using incoming signals from field sensors, A case or rule-based algorithm [54] has been designed for automated spraying and additional recommendations.

Rule Based Algorithm [54]

1. Initialize serial communication with sensor in the field.
2. Set sensor (S) to Cluster (C) and Control station (B).
3. Initially, Drone (D) in hanger (H).
4. Packets (P) are messages between Sensors and Base, Base and Drone.
5. Hanger (H) is the Drone Station.
6. Farmers (F) or Human (H) assists in monitoring.
7. Camera Drone (D-Cam) starts beacons.
8. While loop starts
   If B receive P from S (n) at time T1:
   Send D to S = =C at time T2 to spray pesticide
   For each sensor from i=n to s(max)-n:
   If B receives P from S(i) = =C(i):
   Send D to S(i) = = C(i) to spray pesticides
   Else if B doesn't get P from S (n) (False Alarm):
   Move D to hanger
   Else if B receives P from S(2), S(3), S(1), or S(n) (Multiple insects):
   Send message (M) to D via GSM.
   Send D to S(2), S(3), S(1), or S(n) to spray pesticides.
   Else if B receives continuous P from multiple S(n) manually.
   If B receive continuous P from S(n)
   Send message (M) to Farmer (F) to visit the Field.
   Perform manual monitoring through the human (H) assistant
9. End while loop.
10. Deactivate serial and wireless communication with sensor.

5. Summary and Future Directions

This research focuses on the development of a perception-based prediction model for the identification of diseases affecting cotton plants and their fruits (Cotton Bolls) and automated Pests/ Diseases prevention system. The research incorporates deep learning, neural network and IoT technologies to achieve accurate disease detection. Initially, a dataset was obtained from a website, but since it only consisted of six classes, we expanded it by adding two additional classes to enable the prediction of a wider range of diseases in this project.

Its high accuracy, versatility, and thorough training characterize our model. Despite being trained on a relatively small dataset, our model outperforms one of the existing pre-trained models, delivering remarkably accurate predictions and yielding favorable results. This is particularly significant considering the challenging environmental conditions and limited data availability. Local farmers will greatly benefit from the convenience provided by our model. Efforts have been made during this work to predict the degree of various diseases, enabling farmers to apply medicines based on these predictions. Additionally, automated pesticide spraying via drones can be employed, and affected plants can be isolated from the rest to prevent significant crop damage.

Nevertheless, this study also presents some limitations. One suggestion for future development is the creation of a real-time Android application that enables farmers to receive expert advice while monitoring their fields, thereby mitigating potential losses. However, the unavailability of datasets for other crops at the local level poses a challenge. It is worth noting a limitation of our work: our focus has been exclusively on the detection, identification, and prevention of pests and diseases in cotton crops. The work constrained by the specificity of the dataset, which is compiled from limited sources. Enhancing the quality of our data collection process could significantly improve the overall quality of our work. Utilizing high-quality cameras for image capture and advanced machinery for data processing would not only enhance picture resolution but also increase the accuracy of disease prediction.

Our future goal is to generalize this model to encompass various crops, making it widely accessible. Additionally, the goal is to enhance the model's performance on low-quality photos by expanding and refining the dataset. An effort will be made to create distinct datasets concerning disease levels so that the model can also tell at what level any disease is present, allowing the crop or plant to be managed at that specific level. The extent to which the disease is present will be accurately assessed using IoT-enabled sensors, which provide real-time data and updates on crop health.

In addition to these efforts, our future direction aims to go beyond disease identification by developing more sophisticated and comprehensive prevention strategies. These strategies will be activated based on our model's predictions, providing a complete action plan for effective disease management. This perception-based system will involve dispatching drones equipped with IoT technology to apply medicine in the affected area of the field or uprooting and discarding the infected plants, along with other preventive measures to prevent disease recurrence. The drones will use real-time data from IoT sensors to precisely target the affected areas, ensuring efficient and effective disease management. For future projects, the method will continue to be used the method of deploying sensors and cameras in the field for easy and efficient data collection. Additionally, the images can be analyzed in real time in a nearby control room, where the IoT-based cameras will transmit the images.

By addressing these limitations and further improving the model, a powerful tool is envisioned to empower farmers across different crops and enables precise disease identification, ultimately enhancing agricultural practices and minimizing crop losses. Finally, comprehensive cross validation will be performed in our future research work which will enable everyone to evaluate its statistical significant.

Acknowledgement:

This research is funded by the Deanship of Scientific Research, Islamic University of Madinah, Madinah, Saudi Arabia

Dataset Availability:

In this study, we utilized a tailor-made dataset that is accessible on this website/link, https://www.kaggle.com/datasets/saeedazfar/customized-cotton-disease-dataset serving as a valuable resource for fellow researchers to enhance their research endeavors.

Implementation Code Availability:

The complete code of our trained model will be available on this public repository. http://github.com/kolarkhan/Cotton-Disease-Classification

Author Contributions:

Every author played a role in formulating the approach, revising the manuscript, reading, and endorsing the submitted version.

Conflict of Interest:

The authors assert that the research was carried out without any commercial or financial associations that could be interpreted as a potential conflict of interest.

Author Contributions: For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used "Conceptualization, SA and AN.; methodology, SA, AM and AN.; software, AM and AA.; validation, KA., AA. and HA.; formal analysis, AN.; investigation, AM.; resources, SAQ and HA.; data curation, AA.; writing—original draft preparation, SA.; writing—review and editing, AN, HA and SAQ.; visualization, AM.; supervision, AA.; project administration, AN and KA.; funding acquisition, AN, SAQ and HA.

References

[1] Farooq, Muhammad Shoaib, Shamyla Riaz, Adnan Abid, Tariq Umer, and Yousaf Bin Zikria. (2020). "Role of IoT Technology in Agriculture: A Systematic Literature Review." Electronics 9 (2): 319.

[2] Toda, Yosuke, and Fumio Okura. (2019). "How Convolutional Neural Networks Diagnose Plant Disease." Plant Phenomics.

[3] Azfar, S., Adnan N., and Abdul B. (2015). "Pest Detection and Control Techniques Using Wireless Sensor Network: A Review." Journal of Entomology and Zoology Studies 3 (2): 92–99.

[4] Zhou, Changjian, Sihan Zhou, Jinge Xing, and Jia Song. (2021). "Tomato Leaf Disease Identification by Restructured Deep Residual Dense Network." IEEE Access 9: 28822–31.

[5] Azfar, S, Adnan N, Kamran A, Amir M, Muhammad S. S, Muhammad S, and Mohammad A. (2023). "An IoT-Based System for Efficient Detection of Cotton Pest." Applied Sciences 13 (5): 2921.

[6] Iqbal, Zahid, Muhammad Attique Khan, Muhammad Sharif, Jamal Hussain Shah, Muhammad Habib ur Rehman, and Kashif Javed. (2018). "An Automated Detection and Classification of Citrus Plant Diseases Using Image Processing Techniques: A Review." Computers and Electronics in Agriculture 153: 12–32.

[7] Saleem, M H, Johan Potgieter, and Khalid M A. (2020). "Plant Disease Classification: A Comparative Evaluation of Convolutional Neural Networks and Deep Learning Optimizers." Plants 9 (10): 1319.

[8] Yan, Qian, Baohua Yang, Wenyan Wang, Bing Wang, Peng Chen, and Jun Zhang. (2020). "Apple Leaf Diseases Recognition Based on an Improved Convolutional Neural Network." Sensors 20 (12): 3535.

[9] Wang, Biwei, Yuan Mao, Islam Ashry, Yousef Al-Fehaid, Abdulmoneim Al-Shawaf, Tien Khee Ng, Changyuan Yu, and Boon S Ooi. (2021). "Towards Detecting Red Palm

Weevil Using Machine Learning and Fiber Optic Distributed Acoustic Sensing." Sensors 21 (5): 1592.

[10] Agarwal, M, Abhishek S., Siddhartha A., Amit S, and Suneet G. (2019). "Tomato Leaf Disease Detection Using Convolution Neural Network." In International Conference on Computational Intelligence and Data Science (ICCIDS).

[11] Khan, Muhammad Attique, Tallha Akram, Muhammad Sharif, Muhammad Awais, Kashif Javed, Hashim Ali, and Tanzila Saba. (2018). "CCDF: Automatic System for Segmentation and Recognition of Fruit Crops Diseases Based on Correlation Coefficient and Deep CNN Features." Computers and Electronics in Agriculture 155: 220–36.

[12] Nadeem, A, Muhammad A, Nauman Q, Kashif R, Amir M, Ali AlZahrani, Fazal N, and Qammer H Abbasi. (2022). "Tracking Missing Person in Large Crowd Gathering Using Intelligent Video Surveillance." Sensors 22 (14): 5270.

[13] Nadeem, A., Muhammad A, Kashif R, Nauman Qadeer, Ali AlZahrani, Amir Mehmood, and Qammer H Abbasi. (2022). "A Novel Integration of Face-Recognition Algorithms with a Soft Voting Scheme for Efficiently Tracking Missing Person in Challenging Large-Gathering Scenarios." Sensors 22 (3): 1153.

[14] Mehmood, Amir, Adnan Nadeem, Muhammad Ashraf, Turki Alghamdi, and Muhammad Shoaib Siddiqui. (2019). "A Novel Fall Detection Algorithm for Elderly Using SHIMMER Wearable Sensors." Health and Technology 9: 631–46.

[15] Siddiqui, Abdul Hameed, Syed Akhtar Raza, Arifa Ghazipura, Muhammad Azhar Hussain, Sarwat Iqbal, Kamran Ahsan, and Adnan Nadeem. (2016). "Analysis of Association between Type of Amblyopia and Gender at a Tertiary Care Hospital in Karachi." JPMA J Pak Med Assoc 66 (5): 545–48.

[16] Zhang, Yongli. (2012). "Support Vector Machine Classification Algorithm and Its Application." In Information Computing and Applications: Third International Conference, ICICA 2012, Chengde, China, and September 14-16, 2012. Proceedings, Part II 3, 179–86. Springer.

[17] K. Taunk, S. De, S. Verma and A. Swetapadma, "A Brief Review of Nearest Neighbor Algorithm for Learning and Classification," *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, Madurai, India, 2019, pp. 1255-1260, doi: https://doi.org/10.1109/ICCS45141.2019.9065747.

[18] Niklas D. (2021). "Random Forest: A Complete Guide for Machine Learning," 176.

[19] Alzubaidi, Laith, Jinglan Z, Amjad J H, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, José S, Mohammed A F, Muthana Al-A., and Laith F. (2021). "Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions." Journal of Big Data 8: 1–74.

[20] Liakos, Konstantinos G, Patrizia Busato, Dimitrios Moshou, Simon Pearson, and Dionysis Bochtis. (2018). "Machine Learning in Agriculture: A Review." Sensors 18 (8): 2674.

[21] S. Ashraf, S. Saleem, T. Ahmed, Z. Aslam and M. Shuaeeb, "Iris and Foot based Sustainable Biometric Identification Approach," 2020 International Conference on

Software, Telecommunications and Computer Networks (SoftCOM), Split, Croatia, 2020, pp. 1-6, doi: https://doi.org/10.23919/SoftCOM50211.2020.9238333

[22] Kumbhar, Shantanu, A Nilawar, S Patil, Bodireddy Mahalakshmi, and Manasi Nipane. (2019). "Farmer Buddy-Web Based Cotton Leaf Disease Detection Using CNN." Int. J. Appl. Eng. Res 14 (11): 2662–66.

[23] Brownlee, J. (2019). "A Gentle Introduction to Transfer Learning for Deep Learning." Machine Learning Mastery, A Web Resource, no. 11: 26–29.

[24] Sahana, K and others. (2021). "Farm Vigilance: Smart IoT System for Farmland Monitoring and Animal Intrusion Detection Using Neural Network." In 2021 Asian Conference on Innovation in Technology (ASIANCON), 1–6. IEEE.

[25] Faria, Fábio Augusto, Paula Perre, Roberto A Zucchi, Leonardo Ré Jorge, TM Lewinsohn, Anderson Rocha, and R da S Torres. (2014). "Automatic Identification of Fruit Flies (Diptera: Tephritidae)." Journal of Visual Communication and Image Representation 25 (7): 1516–27. In review

[26] Silva, Diego F, Vinícius MA Souza, Daniel PW Ellis, Eamonn J Keogh, and Gustavo EAPA Batista. (2015). "Exploring Low Cost Laser Sensors to Identify Flying Insect Species: Evaluation of Machine Learning and Signal Processing Methods." Journal of Intelligent & Robotic Systems 80: 313–30.

[27] Li, Hui, Ronghua Ji, Jianhua Zhang, Xue Yuan, Kaiqun Hu, and Lijun Qi. (2011). "WEB-Based Intelligent Diagnosis System for Cotton Diseases Control." In Computer and Computing Technologies in Agriculture IV: 4th IFIP TC 12 Conference, CCTA 2010, Nanchang, China, October 22-25, 2010, Selected Papers, Part III 4, 483–90. Springer.

[28] Saleem, R M, R Kazmi, Imran Sarwar Bajwa, Amna Ashraf, Shabana Ramzan, and Waheed Anwar. (2021). "IOT-Based Cotton Whitefly Prediction Using Deep Learning." Scientific Programming 2021: 1–17.

[29] Sharma, R, Dharavath Ramesh, Pankaj Pal, Sachin Tripathi, and Chiranjeev Kumar. (2021). "IoT-Enabled IEEE 802.15. 4 WSN Monitoring Infrastructure-Driven Fuzzy-Logic-Based Crop Pest Prediction." IEEE Internet of Things Journal 9 (4): 3037–45.

[30] Bankar, S., Ajita Dube, Pranali Kadam, and Sunil Deokule. (2014). "Plant Disease Detection Techniques Using Canny Edge Detection & Color Histogram in Image Processing." Int. J. Comput. Sci. Inf. Technol 5 (2): 1165–68.

[31] Dong, Yuhong, Zetian Fu, Stevan Stankovski, Yaoqi Peng, and Xinxing Li. (2021). "A Cotton Disease Diagnosis Method Using a Combined Algorithm of Case-Based Reasoning and Fuzzy Logic." The Computer Journal 64 (2): 155– 68.

[32] Gawande, Apeksha R, and Swati S Sherekar. (2023). "Analysis of Crop Diseases Using IoT and Machine Learning Approaches." In International Conference on Applications of Machine Intelligence and Data Analytics (ICAMIDA 2022), 78–85. Atlantis Press.

[33] Vajpayee, Pooja, and Kuldeep Kr Yogi. (2021). "Recognition and Early-Stage Detection of Phytophthora in a Crop Farm Using IoT." Agro-Economic Risks of Phytophthora and an Effective Biocontrol Approach, 95.

[34] Liu, Zhiyan, Rab Nawaz Bashir, Salman Iqbal, Malik Muhammad Ali Shahid, Muhammad Tausif, and Qasim Umer. (2022). "Internet of Things (IoT) and Machine

Learning Model of Plant Disease Prediction–Blister Blight for Tea Plant.” Ieee Access 10: 44934–44.
[35] Sharma, R P, Ramesh Dharavath, and Damodar R Edla. (2023). “IoFT-FIS: Internet of Farm Things Based Prediction for Crop Pest Infestation Using Optimized Fuzzy Inference System.” Internet of Things 21: 100658.
[36] Goel, Lavika, and Jyoti Nagpal. (2023). “A Systematic Review of Recent Machine Learning Techniques for Plant Disease Identification and Classification.” IETE Technical Review 40 (3): 423–39.
[37] Ramalingam, B, Rajesh Elara Mohan, Sathian Pookkuttath, Braulio Félix Gómez, Charan Satya Chandra SairamBorusu, Tey Wee Teng, and Yokhesh Krishnasamy Tamilselvam. (2020). “Remote Insects Trap Monitoring System Using Deep Learning Framework and IoT.” Sensors 20 (18): 5280.
[38] Chen, W, Yi-Bing Lin, Fung-Ling Ng, Chun-You Liu, and Yun-Wei Lin. (2019). “RiceTalk: Rice Blast Detection Using Internet of Things and Artificial Intelligence Technologies.” IEEE Internet of Things Journal 7 (2): 1001–10.
[39] Nadeem, R M, Arfan J, and Rana M S. (2022). “IoT and Machine Learning Based Stem Borer Pest Prediction.” INTELLIGENT AUTOMATION AND SOFT COMPUTING 31 (3): 1377–92.
[40] Praba, V, and K Krishnaveni. (2023). “A Review on Crop Disease Detection Techniques.” In International Conference on IoT, Intelligent Computing and Security: Select Proceedings of IICS 2021, 435–48. Springer.
[41] Titiya, Mahesh D, and Vipul A Shah. (2018). “Ontology Based Expert System for Pests and Disease Management of Cotton Crop in India.” International Journal of Web Portals (IJWP) 10 (2): 32–49.
[42] Xiao, Qingxin, Weilu Li, Yuanzhong Kai, Peng Chen, Jun Zhang, and Bing Wang. (2019). “Occurrence Prediction of Pests and Diseases in Cotton on the Basis of Weather Factors by Long Short-Term Memory Network.” BMC Bioinformatics 20: 1–15.
[43] Thorat, Apeksha, Sangeeta Kumari, and Nandakishor D Valakunde. (2017). “An IoT Based Smart Solution for Leaf Disease Detection.” In 2017 International Conference on Big Data, IoT and Data Science (BID), 193–98. IEEE.
[44] Parikh, A, Mehul S Raval, Chandrasinh Parmar, and Sanjay Chaudhary. (2016). “Disease Detection and Severity Estimation in Cotton Plant from Unconstrained Images.” In 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA), 594–601. IEEE.
[45] Devi, T Gayathri, A Srinivasan, S Sudha, and D Narasimhan. (2019). “Web Enabled Paddy Disease Detection Using Compressed Sensing.” Mathematical Biosciences and Engineering 16 (6): 7719–33.
[46] Rothe, PR, and RV Kshirsagar. (2014). “Automated Extraction of Digital Images Features of Three Kinds of Cotton Leaf Diseases.” In 2014 International Conference on Electronics, Communication and Computational Engineering (ICECCE), 67–71. IEEE.
[47] Memon, Muhammad Suleman, Pardeep Kumar, and Rizwan Iqbal. (2022). “Meta Deep Learn Leaf Disease Identification Model for Cotton Crop.” Computers 11 (7): 102.

[48] Pantazi, X Eirini, Dimitrios Moshou, and Alexandra A Tamouridou. (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.

[49] Zekiwos, Melese, and Abey Bruck. (2021). "Deep Learning-Based Image Processing for Cotton Leaf Disease and Pest Diagnosis." Journal of Electrical and Computer Engineering 2021: 1–10.

[50] Kumar, Satyam, Dayima Musharaf, and Anil Kumar Sagar. (2022). "Comparative Study of Pre-Trained Models on Cotton Plant Disease Detection Using Transfer Learning." In Advanced Computing and Intelligent Technologies: Proceedings of ICACIT 2022, 147–55. Springer.

[51] Sivakumar, P, N Sri Ram Mohan, P Kavya, and P Vinay Sai Teja. (2021). "Leaf Disease Identification: Enhanced Cotton Leaf Disease Identification Using Deep CNN Models." In 2021 IEEE International Conference on Intelligent Systems, Smart and Green Technologies (ICISSGT), 22–26. IEEE.

[52] Patil, B V, and Pravin S Patil. (2021). "Computational Method for Cotton Plant Disease Detection of Crop Management Using Deep Learning and Internet of Things Platforms." In Evolutionary Computing and Mobile Sustainable Networks: Proceedings of ICECMSN 2020, 875–85. Springer.

[53] Zambare, Rajani, Rashmi Deshmukh, Chetan Awati, Suresh Shirgave, Sandeep Thorat, and Sheetal Zalte. (2022). "Deep Learning Model for Disease Identification of Cotton Plants." Special usis Ugdymas 1 (43): 6684–95.

[54] Azfar, S, Adnan N., Kamran A, Amir M., Hani Almoamari, and Saad Said Alqahtany. (2023). "IoT-Based Cotton Plant Pest Detection and Smart-Response System." Applied Sciences 13 (3): 1851.

[55] Atila, Ümit, Murat U, Kemal Akyol, and Emine Uçar. (2021). "Plant Leaf Disease Classification Using EfficientNet Deep Learning Model." Ecological Informatics 61: 101182.

[56] Pandhare, N., Vrunali P, Shivam S M, and D Tambe. (2022). "Cotton Plant Disease Detection Using Deep Learning." Int. Res. J. Modern. Eng. Technol. Sci 4 (04).

[57] Dhamodharan. (2023). "Cotton Plant Disease." http://www.Kaggle.com.

[58] Simonyan, Karen, and Andrew Zisserman. (2014). "Very Deep Convolutional Networks for Large-Scale Image Recognition." arXiv Preprint arXiv:1409. 1556.

[59] Vihar Kurama. (2020). "A Review of Popular Deep Learning Architectures: ResNet, InceptionV3, and SqueezeNet." PaperSpace.Com, A Web Resource, no. 11: 10–17.

[60] Lin, Chunmian, Lin Li, Wenting Luo, Kelvin CP Wang, and Jiangang Guo. "Transfer learning based traffic sign recognition using inception-v3 model." Periodica Polytechnica Transportation Engineering 47, no. 3 (2019): 242-250.

[61] Feng, Z. (2022). "An Overview of ResNet: Architecture and Its Variant," https://builtin.com/artificial-intelligence/resnet-architecture (Web Resource.)

[62] Banger, S. (2022). "The Architecture of LeNet-5". https://medium.com/@siddheshb008/lenet-5-architecture-explained-3b559cb2d52b Web Resource.

[63] Suryavanshi V., Bhamare Y., Rahul B., Komal C., and Bhushan M. (2021). "Cotton Leaf Disease Detection Using CNN." IJCRT 9 (6): 76.

[64] Sharvari V.P., Kajal B.J, Anjali K.S., Bhagyashree R.K. (2022). "Cotton Leaf Disease Detection Using Deep Learning." IJCRT 9 (6): 47.

[65] Harshitha, Gnyana, Sandeep Kumar, Shilpa Rani, and Arpit Jain. (2021). "Cotton Disease Detection Based on Deep Learning Techniques." In 4th Smart Cities Symposium (SCS 2021), 2021:496–501. IET.

[66] Rajasekar, V, K Venu, SoumyaRanjan Jena, R Janani Varthini, and S Ishwarya. (2021). "Detection of Cotton Plant Diseases Using Deep Transfer Learning." Journal of Mobile Multimedia 18 (2): 307–24.

[67] Latif, Muhammad Rizwan, Muhammad Attique Khan, Muhammad Younus Javed, Haris Masood, Usman Tariq, Yunyoung Nam, and Seifedine Kadry. (2021). "Cotton Leaf Diseases Recognition Using Deep Learning and Genetic Algorithm." Computers, Materials & Continua 69 (3).

[68] Gülmez, Burak. (2023). "A Novel Deep Learning Model with the Grey Wolf Optimization Algorithm for Cotton Disease Detection." Journal of Universal Computer Science 29 (6): 595.

[69] Chopda, Jj, Hiral R, Sagar Nakum, and Vivek Nakrani. (2018). "Cotton Crop Disease Detection Using Decision Tree Classifier." In 2018 International Conference on Smart City and Emerging Technology (ICSCET), 1–5. IEEE.

[70] Khattab, Serag ED, A. H., Haythem Ismail, Sahar Zayan, Yasmine Fahmy, and Mohamed M Khairy. (2019). "An IoTBased Cognitive Monitoring System for Early Plant Disease Forecast." Computers and Electronics in Agriculture 166: 105028.

[71] Phadikar, Santanu, et al. "Rice diseases classification using feature selection and rule generation techniques." Computers and Electronics in Agriculture 90 (2013): 76-85.

[72] Guo, Yan, Jin Zhang, Chengxin Yin, Xiaonan Hu, Yu Zou, Zhipeng Xue, and Wei Wang. (2020).

[73] Arsa, D. Made Sri, and Anak Agung Ngurah Hary S. (2019). "VGG16 in Batik Classification Based on Random Forest." In 2019 International Conference on Information Management and Technology (ICIMTech), 1:295–99. IEEE.

[74] Bhannu. K. N and S. K, "Farm Vigilance: Smart IoT System for Farmland Monitoring and Animal Intrusion Detection using Neural Network," 2021 Asian Conference on Innovation in Technology (ASIANCON), PUNE, India, 2021, pp. 1-6,
https://doi.org/10.1109/ASIANCON51346.2021.9544926

[75] K. Saranya, P. Uva Dharini, P. Uva Darshni and S. Monisha, "IoT Based Pest Controlling System for Smart Agriculture," *2019 International Conference on Communication and Electronics Systems (ICCES)*, Coimbatore, India, 2019, pp. 1548-1552,
https://doi.org/10.1109/ICCES45898.2019.9002046

[76] J. C. Sargunar, S. Manikan darajan, and T. Kamalak kannan Subha, , "AI based pest detection and alert system for farmers using IoT", ICSERET-2023, Web of Conferences 387, 05003 (2023) https://doi.org/10.1051/e3sconf/202338705003

[77] Wang Z, Qiao X, Wang Y, Yu H and Mu C (2024) IoT-based system of prevention and control for crop diseases and insect pests. Front. Plant Sci. 15:1323074. https://doi.org/10.3389/fpls.2024.1323074