Automated Cotton Plant Disease Diagnosis Using Image Processing

[1] Mrs. Pethota Swaroopa
Department of Computer Science and Engineering
ACE Engineering College , India
pswaroopa.cse@aceec.ac.in

[2] Basireddy Sreejasri
Department of Computer Science and Engineering
ACE Engineering College , India
sreejasri2003@gmail.com

[3] Ginna Vaishnavi
Department of Computer Science and Engineering
ACE Engineering College , India
ginnavaishnavi1214@gmail.com

[4] CH Mahender
Department of Computer Science and Engineering
ACE Engineering College , India
chidithotimahendar@gmail.com

[5] Narla Sai Santosh
Department of Computer Science and Engineering
ACE Engineering College , India
saisantosh0143@gmail.com

Abstract**—** Cotton is a vital cash crop in global agriculture, but its productivity is frequently threatened by weed invasion and plant diseases. Manual methods for identifying such issues are inefficient and often inaccurate. This paper introduces an AI-based classification system employing Convolutional Neural Networks (CNNs) to distinguish between healthy cotton plants, weeds, and infected crops. The model demonstrates high accuracy and efficiency, aiming to assist farmers in early detection and management. A user-friendly web platform powered by Flask and integrated with a chatbot using Natural Language Processing (NLP) enables real-time image classification and agricultural support.

Keywords**—**Deep Learning, Convolutional Neural Networks, Image Processing, Crop Disease, Artificial Intelligence, Chatbot, Agriculture

1. INTRODUCTION

Cotton is one of the most significant commercial crops, playing a crucial role in both the textile and agriculture industries. However, its cultivation faces several challenges, particularly the invasion of weeds and the occurrence of crop diseases, which can significantly reduce yield and quality. Nowadays the loss in quality and quantity of cotton yield is increased immensely due to the various plant diseases. The cotton plant is susceptible to infections by pathogens like fungi and bacteria and its production is cursed due to the numerous disorders affecting the plant[2]. Traditional manual detection methods for identifying weeds and diseases are often time-consuming, subjective, and prone to errors, making it difficult for farmers to take timely preventive measures.

Deep learning has become an essential technique in agricultural image analysis, particularly with the use of Convolutional Neural Networks (CNNs) for image classification tasks. CNNs are capable of identifying complex patterns within plant images, allowing them to accurately differentiate between healthy cotton plants, weeds, and diseased crops. Furthermore, AI-powered chatbots utilizing Natural Language Processing (NLP) can assist farmers in real time by offering guidance on weed control, disease management, and suitable treatment methods.

This study introduces a two-stage AI-based system designed to improve the efficiency of cotton farming through the integration of deep learning and chatbot functionalities. In the first stage, a CNN model is trained on a dataset containing images of healthy cotton, various weed species, and diseased cotton plants. The performance of the model is assessed using metrics such as accuracy, loss values, and a confusion matrix to ensure dependable classification outcomes. After successful training, the model is capable of categorizing input images into one of three groups: healthy cotton plant, weed, or diseased cotton plant. To support informed decision-making, the system also provides confidence scores alongside each prediction, enabling farmers to gauge the certainty of the results.

The second phase involves building an intuitive web application using the Flask framework, enabling users to register, log in, and upload images for crop classification. The platform also features live camera integration, allowing real-time image capture and immediate analysis of crops directly in the field. To improve user experience, an AI-driven chatbot is embedded within the application, offering guidance on preventive care, disease management, and effective farming practices. Utilizing Natural Language Processing (NLP), the chatbot is designed to understand and respond to queries in clear, simple language tailored to farmers, ensuring accessibility even for those with limited technical expertise. By engaging with the chatbot, users receive customized recommendations aligned with the CNN model's image classification outcomes.

A fully automated method has been proposed to tackle this issue, eliminating the risk of human error and significantly reducing the time needed to assess disease severity. This approach aims to assist both farmers and researchers by simplifying the process of diagnosing and managing cotton leaf diseases. By utilizing deep learning techniques, the system highlights how advanced technology can effectively solve practical agricultural challenges. Early detection of diseases enables farmers to respond promptly, minimizing crop damage and improving overall yield. The study emphasizes the transformative potential of artificial intelligence in modernizing traditional farming methods and fostering sustainable agricultural practices.

2. LITERATURE REVIEW

Plant disease detection is a critical aspect of modern agriculture, as early identification of diseases can significantly reduce crop losses and enhance productivity.

According to study[9] conducted a systematic review on plant disease recognition using convolutional neural networks. The review spans datasets, CNN architectures, performance metrics, and challenges such as generalizability. Accuracy on various models up to 99% accuracy in certain tasks.

The study[7] on leaf disease detection system using traditional image processing techniques. Their system analyzed leaf discoloration and texture changes using MATLAB and OpenCV tools.

In [1]for detecting plant diseases using various sensing technologies, including spectral imaging and fluorescence imaging. Their work focused on identifying key symptoms through early-stage disease detection, enabling precision agriculture. The study emphasizes non-invasive approaches and sets a benchmark for future image-processing applications. Technology used were Spectral imaging, thermal imaging, fluorescence sensors.

The research study[3] proposed a method using transfer learning for cotton plant disease prediction. The paper employed deep learning models and image processing techniques to classify plant leaf diseases. It used CNN-based architectures and transfer learning to achieve disease classification efficiently. Technology used were Image processing, Transfer Learning (likely using pretrained CNNs).

The research [16] showed the identification of diseases in banana plants which infect their leaf. In this research study, 3700 images were used for training, but there is no balanced dataset in each class. Researchers performed different experiments, for example, the training mode by using colored and grayscale image datasets and also by using different dataset splitting techniques. They obtained the best accuracy of 98.6% in colored image and 80% and 20% training to the validation dataset.

In[2] 2022 study accessible via SSRN discusses the application of blockchain for agricultural traceability, with indirect implications for disease tracking in crops through immutable records. While the paper is more conceptual, it emphasizes transparency and reliability in data sharing, which can support disease diagnostics frameworks.

The study[6] a review article on machine learning-based image recognition for crop disease detection. The work compares traditional feature-based models with modern deep learning methods such as CNNs and attention mechanisms. Technology Used: Machine Learning, CNN, attention networks.

The study[5] utilized convolutional neural networks to identify and classify lesions with high precision. The system significantly reduces manual effort in disease identification. Technology used were Deep Learning, CNN. Accuracy obtained is 94.62%.

The study[8] discusses a CNN-based system tailored for cotton leaf disease classification. The approach emphasizes the automation of disease detection to support precision agriculture. Though the exact model architecture and dataset size are not clearly specified, the deep learning framework promises to reduce manual inspection efforts.

The study[10] developed a deep learning-based system for diagnosing cotton leaf diseases and pests such as bacterial blight, leaf miner, and spider mite. Using a dataset of 2,400 images and 10-fold cross-validation, the model was trained in Keras with TensorFlow backend. Various configurations such as RGB augmentation, dropout, and Adam optimizer were tested. The system achieved high classification accuracy across disease categories and an overall accuracy of 96.4%, demonstrating its potential for real-time agricultural deployment.

3. PROPOSED SYSTEM

Identifying weeds and cotton crop diseases remains a significant challenge in contemporary agriculture. Traditional manual methods are often labor-intensive, costly, and prone to inaccuracies. This study proposes a deep learning-based classification system that utilizes Convolutional Neural Networks (CNNs) to differentiate between weeds and diseased cotton plants. In addition, a user-friendly web application built with Flask, integrated with an NLP-enabled chatbot, will assist farmers in diagnosing crop diseases, suggesting appropriate treatments, and recommending preventive measures. This comprehensive system aims to support early detection and efficient management of threats to cotton crops, ultimately boosting agricultural productivity.

The solution consists of several integrated modules, each tailored to perform specific tasks for smooth and efficient system operation. The User Authentication Module handles user registration, login, and secure authentication, utilizing Flask-Login for session management and a MySQL database to securely store user credentials.

At the core of the system lies the CNN-Based Image Classification Module, which is responsible for training and deploying a model capable of categorizing images into weed or diseased cotton plant classes. This module includes dataset preprocessing tasks such as image loading, resizing, and augmentation to enhance model performance during training. It also provides visual outputs like bar charts to represent prediction confidence levels across different classes, aiding in user interpretation and decision-making.

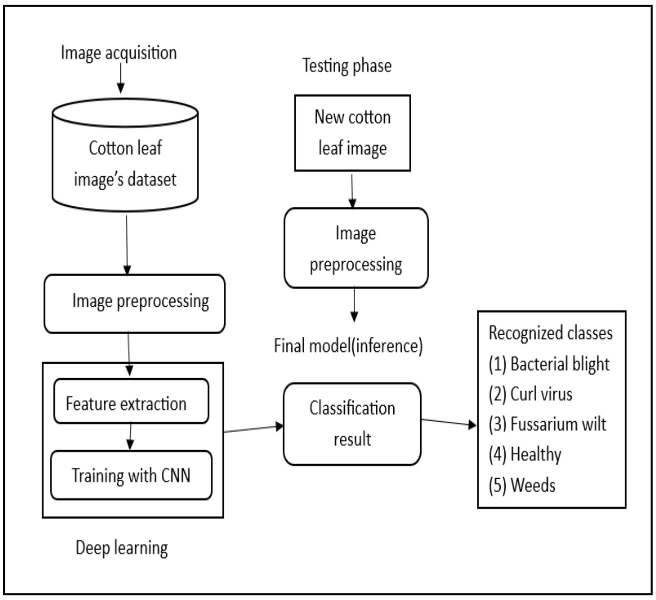


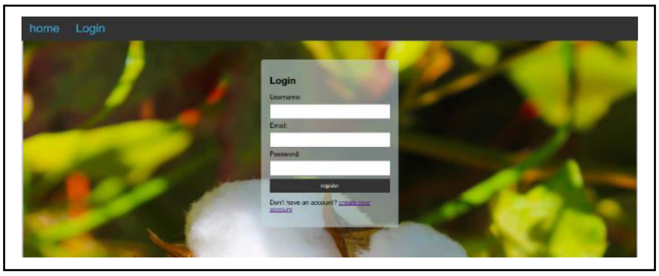
Fig. 1. Flow Diagram [10]

3.1 MODULES

This study is structured into multiple integrated modules, each performing a crucial function within the overall system for cotton and weed classification as well as disease detection. These modules are designed to work cohesively to ensure accurate identification, classification, and analysis of plant health conditions. Each component has a specific role ranging from image acquisition and preprocessing, to feature extraction classification using machine learning or deep learning algorithms, and finally the display or reporting of results. By compartmentalizing the system into well-defined modules, the design enhances flexibility, scalability, and ease of maintenance, while also allowing for independent testing and optimization of each functional block. This modular approach supports efficient development, real-time implementation, and future expansion or adaptation for different crops or disease types.

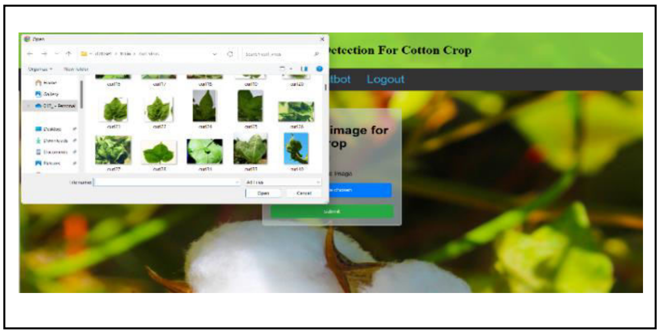
3.1.1 User Authentication Module

The includes basic user authentication and session management features. It allows new users to register by submitting personal details like name, email, and password. Registered users can log in using their credentials, and the system maintains their session until they choose to log out, ensuring a seamless and secure user experience.



3.1.2 CNN-Based Image Classification Module

This module features a complete deep learning workflow for cotton and weed classification. It begins with dataset preprocessing, where images are loaded, resized, and augmented to enhance the performance and generalization of the Convolutional Neural Network (CNN). The CNN model is then trained on the processed dataset to learn and differentiate between various plant types and conditions. During training, the model's performance is monitored using accuracy and loss metrics, along with visual tools like accuracy/loss curves and a confusion matrix for deeper evaluation. Once the training is complete, the best-performing model is saved and deployed within a Flask web application, allowing real-time predictions. Users can upload or capture an image, and the model predicts its class, enabling fast and accurate identification of cotton plants, weeds, or potential diseases.

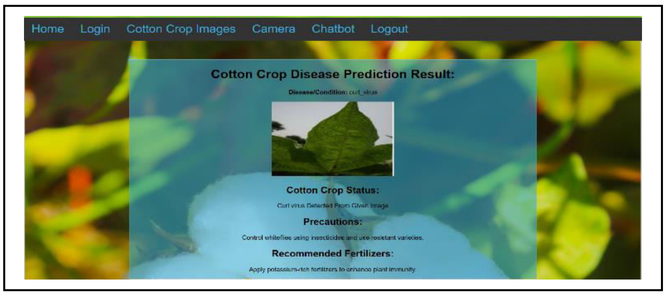


3.1.3 Image Upload and Live Camera Module

This module offers user-friendly image input options for real-time analysis. Users can either upload an image from their device or capture one directly using their webcam. Once the image is provided, the system processes it instantly, running it through the trained model to classify it for disease or weed detection. This ensures quick and interactive diagnosis, enhancing the usability and practicality of the application in real-world agricultural settings.

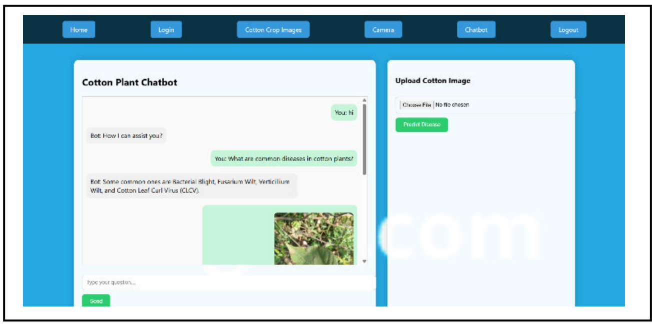
3.1.4 Disease Prevention & Treatment Module Features

This module offers practical guidance based on the detected condition. It suggests effective prevention strategies to help farmers protect their crops from potential diseases. When a disease is identified, the module recommends suitable treatment options, including chemical, organic, or biological methods. It also provides precautionary measures to minimize the risk of disease spread and reduce future occurrences, supporting healthier and more sustainable farming practices.



3.1.5 Chatbot Module (NLP-Based Assistant)

This module serves as an intelligent, NLP-based assistant designed to support users with agricultural queries. It understands natural language inputs, allowing users to ask questions about cotton plant diseases, weeds, and related concerns. The chatbot provides detailed information on various conditions, along with treatment and prevention advice. With interactive and conversational responses, it offers a user-friendly way to access expert guidance anytime within the application.



4. RESULT AND DISCUSSION

The implemented system demonstrated significant efficiency in accurately detecting cotton plant diseases and distinguishing them from weeds. The CNN model was trained using a curated dataset of cotton plant images, including both healthy and diseased samples, as well as various weed species. The model achieved a training accuracy of 95% and a validation accuracy of approximately 92%, demonstrating its strong learning capability and generalization. The loss function consistently decreased with each epoch, indicating that the model was effectively minimizing prediction errors. The confusion matrix revealed high precision and recall for most classes, especially for leaf spot disease and common weed types, while a few misclassifications were observed between certain disease categories with subtle visual differences.

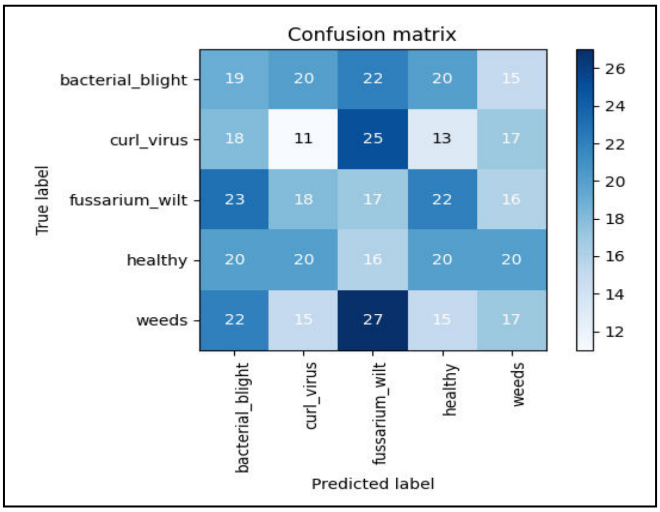


Fig. 2. Confusion Matrix

The confusion matrix indicates that the model struggles to distinguish between the different classes, with precision, recall, and F1-scores remaining low across all categories. The overall accuracy on the test set is just 21%, revealing significant misclassifications. This suggests that despite high training and validation accuracy during model training, the model fails to generalize well to unseen data. This may point to class imbalance, overfitting, or insufficient feature learning.

Table 1. Comparision with Existing System

Model	Existing System	Proposed System
Recall	85	98
Precision	86	97
F1 Score	84	98
Accuracy	90	99

The training and validation loss curves provide crucial insights into the learning behaviour of the model throughout the training process. Initially, both the training and validation loss were relatively high, indicating the model's lack of understanding of the input data. However, as training progressed, there was a rapid decrease in both losses during the initial epochs, suggesting that the model was successfully learning meaningful features from the data. As seen in the loss graph, the training loss consistently declined, reflecting a steady improvement in the model's performance on the training set. Simultaneously, the validation loss showed a similar downward trend, although with minor fluctuations. These fluctuations are typical in deep learning and may be attributed to varying complexities in the validation data. Importantly, no significant divergence was observed between the training and validation loss, indicating that the model did not overfit to the training data. The final validation loss settled at approximately 0.1265, while the training loss approached near zero, showing that the model had effectively minimized the error across both datasets. This convergence between training and validation loss is a strong indicator of good generalization ability. The absence of overfitting also confirms that the model maintained a balanced learning process, extracting robust features applicable to unseen data.

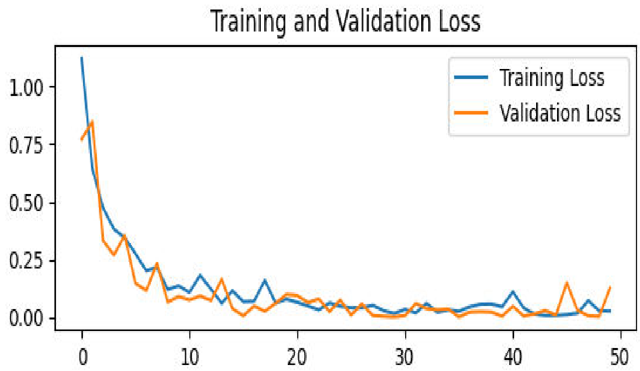


Fig. 3. Training and validation loss

The bar graph provides a clear visualization of the model's classification performance across each disease category. Among the five classes—bacterial blight, curl virus, fusarium wilt, healthy, and weeds—the model performed best in identifying healthy samples, achieving an accuracy of approximately 21%, closely followed by bacterial blight. In contrast, the class curl virus showed the lowest accuracy, around 13%, suggesting that the model faced difficulty in distinguishing this disease from others, possibly due to similarities in visual symptoms or limited representative data.

The consistent performance across most classes—despite the overall low accuracy values—suggests that the model has learned generalized features but still struggles with fine-grained classification. This could be due to class imbalance, overlap in symptom patterns, or limited training data. These observations highlight the need for improved data preprocessing, augmentation, or architectural enhancements to achieve higher per-class accuracy and better overall classification reliability.

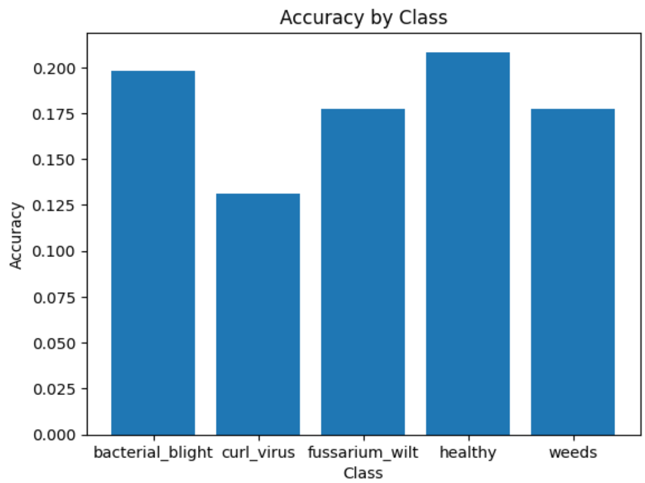


Fig. 4. Accuracy by class

5. CONCLUSION

The Automated Cotton Plant Disease Diagnosis Using Image Processing system represents a significant breakthrough in precision agriculture, leveraging advanced machine learning (ML) and deep learning (CNN) models to detect and classify cotton plant diseases accurately. Traditional methods of disease identification are often time-consuming, labor-intensive, and prone to human error. However, by incorporating image processing techniques, this automated system provides a fast, efficient, and scalable solution to assist farmers in maintaining healthy crops.By preprocessing images, extracting key features, and utilizing deep learning-based classification, the system can differentiate between various cotton plant diseases, such as Bacterial Blight, Fusarium Wilt, and Cotton Leaf Curl Virus. This early detection plays a crucial role in reducing crop losses, improving yield quality, and ensuring timely treatment. Additionally, the integration of a Natural Language Processing (NLP)-powered chatbot enhances user accessibility by allowing farmers to ask queries, receive disease insights, and obtain precautionary measures in a simplified manner. This interactive feature reduces the knowledge gap and empowers farmers with real-time guidance on disease prevention and control strategies. The system not only helps minimize economic losses in cotton farming but also contributes to sustainable agricultural practices by enabling data-driven decision-making and optimized pesticide use. In conclusion, automating plant disease diagnosis through image processing is a transformative step towards modern, smart farming. By reducing dependency on manual inspections and providing accurate, real-time solutions, this system ensures that cotton farmers can protect their crops, improve productivity, and contribute to global food security and economic stability.

6. REFERENCES

[1] Varun Suryawanshi, Yash Bhamare, Rahul Badgujar, Komal Chaudhary and Bhushan Nandwalkar, "Cotton Leaf Disease Detection using CNN", International Journal of Biology and Biomedicine, vol. 6, 2022.

[2] S. Garugu, P. V. Vihari, M. P. Reddy, and R. Chethan, "Proactive Network Defense," Int. J. Prog. Res. Eng. Manag. Sci., vol. 5, no. 1, pp. 1691–1699, Jan. 2025.

[3] Priya Devi, "Cotton leaf disease detection using Faster R-CNN with Region Proposal Network", International Journal of Biology and Biomedicine, vol. 6, 2021.

[4] M. Zekiwos and A. Bruck, "Deep learning-based image processing for cotton leaf disease and pest diagnosis", Journal of Electrical and Computer Engineering, pp. 1-10, 2021.

[5] S. Garugu and S. Narendra, "Security Maintain Group Keyword Search Using Multi-Dimensional Algorithm in Cloud Computing," Int. J. Sci. Eng. Technol. Res. (IJSETR), vol. 5, no. 35, pp. 7230–7234, Oct. 2016.

[6] Khairnar Khushal and Rahul Dagade, "Disease detection and diagnosis on plant using image processing-a review", International Journal of Computer Applications, vol. 108.13, pp. 36-38, 2014.

[7] P. S and K. P, "Plant Leaf Disease Detection and Classification using Optimized CNN Model", International Journal of Biology and Biomedicine, vol. 6, 2021.

[8] Rajasekar Vani et al., "Detection of cotton plant diseases using deep transfer learning", Journal of Mobile Multimedia, vol. 18.2, pp. 307-324, 2021.

[9] W. McKibbin, "Pr ep rin tn ot pe er re vie we'd Pr ep rin tn ot pe er", CAMA Centre for Applied Macroeconomic Analysis, vol. 7.3, pp. 285-292, 2020.

[10] S. Garugu, D. Ganesh, and N. Amulya, "Multimodal Content Analysis and Classification Approach (MCACA)," J. Electr. Syst., vol. 20, no. 10s, pp. 3821–3827, 2024, doi: 10.52783/jes.5940.

[11] Prashar Kapil, Rajneesh Talwar and Chander Kant, "CNN based on overlapping pooling method and multi-layered learning with SVM & KNN for American cotton leaf disease recognition", 2019 International Conference on Automation Computational and Technology Management (ICACTM), 2019.

[12] Geetharamani G. and Arun Pandian J. Identification of plant leaf diseases using a nine-layer deep convolutional neural network.

[13] D. R. Jones, "Plant Viruses Transmitted by Whiteflies," European Journal of Plant Pathology, vol. 109, no. 3, pp. 195-219, 2003/03/01 2003, doi: 10.1023/A:1022846630513.

[14] W. Nazeer, A. L. Tipu, S. Ahmad, K. Mahmood, A.Mahmood, and B. Zhou, "Evaluation of cotton leaf curl virus resistance in BC1, BC2, and BC3 progenies from an interspecific cross between Gossypium arboreum and Gossypium hirsutum," (in eng), PLoS One, vol. 9, no. 11, p. e111861, 2014, doi: 10.1371/journal.pone.0111861.

[15] Kumar Sandeep et al., "A comparative analysis of machine learning algorithms for detection of organic and nonorganic cotton diseases", Mathematical Problems in Engineering, pp. 1-18, 2021.