



Machine Learning–Based Prediction of Cotton Crop Diseases: A Systematic Study

Sagar Padiya¹, Jayashri Salampuria²

¹Assistant Professor, Shri Sant Gajanan Maharaj College of Engineering, Shegaon, India

²Student, Shri Sant Gajanan Maharaj College of Engineering, Shegaon, India

Abstract: Agriculture is one of the basic entities in Indian economy and to maintain the progressive production of food it is considered to be the most important aspects. This project uses machine learning to create a smart system for detecting and predicting cotton crop diseases. By employing image recognition and analysing datasets of disease symptoms, the system will enable farmers to identify diseases early, leading to healthier crops and increased yields. The system integrates data preprocessing, model training, and real-time disease detection using deep learning. This technology will improve farmers' decision-making, minimize crop losses, and support sustainable agriculture. Accuracy, scalability, and user-friendliness for cotton farmers are key focuses of this project.

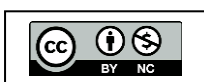
Keywords: Machine Learning, Recognition, Disease, Datasets, Accuracy.

I. INTRODUCTION

Cotton is a critical global commodity, supporting millions of farmers and a substantial portion of the textile industry, particularly in regions like India, China, and the United States. However, cotton crops are vulnerable to diseases such as bacterial blight, root rot, and leaf curl virus, which can drastically reduce yields and fiber quality. Effective disease management is therefore essential for profitable and sustainable cotton farming. Improved disease prediction and intervention are crucial for maintaining a stable cotton supply and protecting the livelihoods of those involved in its production.

Traditional disease management relies on periodic inspections by experts or farmers, who visually assess symptoms and apply treatments. These methods are labor-intensive, dependent on the observer's expertise, and often fail to detect diseases in their early stages. Small-scale farmers, often lacking access to expert advice and diagnostic tools, face significant challenges in effective disease management. Furthermore, delayed disease identification can result in excessive pesticide use, increasing costs and harming the environment. This necessitates more efficient, accessible, and accurate disease management solutions.

Machine learning (ML), a branch of artificial intelligence, trains algorithms to identify patterns and make data-driven predictions. ML has demonstrated promising results in various agricultural applications, including crop yield prediction, weed detection, pest control, and soil health monitoring. In disease prediction, ML can analyze extensive image datasets (e.g., infected leaf images) alongside environmental data like soil moisture, temperature, and humidity. By recognizing complex patterns, ML models can accurately detect diseases early. For cotton crops, ML-based disease prediction offers the potential for real-time diagnosis and tailored recommendations, leading to more efficient and precise disease management practices.



II. LITERATURE REVIEW

The Cotton Disease Prediction System represents a groundbreaking approach in agricultural management, addressing the persistent threats posed by diseases to this pivotal global cash crop. With the imperative need for timely disease detection to safeguard yield and quality, this research introduces a novel methodology leveraging transfer learning with Convolutional Neural Networks (CNNs). By analyzing high-resolution images of cotton leaves, the system utilizes a userfriendly web interface developed using HTML and CSS in Visual Studio Code. The backend functionality, orchestrated through Jupyter Notebook, enables real-time processing and predictions. Through rigorous testing, the system demonstrates remarkable robustness and accuracy, offering farmers a proactive tool for disease management. By seamlessly integrating frontend and backend components, this system not only enhances efficiency but also empowers.

Cotton farming, an essential component of global agriculture, confronts challenges from diseases jeopardizing crop health and productivity. The timely identification of these diseases is critical for mitigating their impact. In response, this research introduces the Cotton Disease Prediction System, a pioneering solution leveraging transfer learning with Convolutional Neural Networks (CNNs). Developed through a user-friendly web interface, HTML, and CSS in Visual Studio Code, the system ensures accessibility for farmers. The backend, powered by Jupyter Notebook, enables real-time processing, making swift disease predictions. This paper delves into the methodology, results, and future directions, highlighting the transformative potential of this system in precision agriculture. (Jadhav, et. al. 2024)

This paper delves into the innovative utilization of computer technology as a means to assist farmers in the critical task of identifying and managing diseased cotton plants. The study encompasses a comprehensive data collection effort, comprising a diverse array of images portraying various states of cotton plant health, ranging from robustly healthy specimens to those afflicted with various diseases. Leveraging sophisticated machine learning methodologies, specifically Convolutional Neural Networks (CNNs), a meticulously crafted computational model was trained to discern subtle visual cues indicative of plant health or disease within the collected dataset. Special emphasis was placed on ensuring the model's adaptability to a wide spectrum of environmental conditions, encompassing variations in lighting, angles, and background clutter commonly encountered in real-world agricultural settings. Rigorous evaluation and validation procedures were employed to assess the model's performance, revealing commendable levels of accuracy and reliability in its ability to classify cotton plants with precision.

The agricultural sector plays a pivotal role in ensuring food security and economic stability worldwide. Among the various crops cultivated globally, cotton holds significant economic importance, with its cultivation spanning vast regions across the globe. However, cotton production faces numerous challenges, chief among them being the rampant spread of diseases that can devastate entire crops if left unchecked. The early detection and management of these diseases are paramount to ensuring optimal crop yield and safeguarding farmers' livelihoods. Traditional methods of disease identification often rely on visual inspection by experienced agronomists, which can be time-consuming, subjective,



and prone to errors. In recent years, the advent of advanced technologies, particularly in the field of machine learning and computer vision, has offered promising solutions to the challenges faced by the agricultural sector. (Kshirsagar, et. al. 2024)

The creation and use of a comprehensive cotton leaf disease dataset offer significant benefits in agricultural research, precision farming, and disease management. This dataset enables the development of accurate machine learning models for early disease detection, reducing manual inspections and facilitating timely interventions. It serves as a benchmark for testing algorithms and training deep learning models, aiding in automated monitoring and decision support tools in precision agriculture. This leads to targeted interventions, reduced chemical use, and improved crop management. Global collaboration is fostered, contributing to the development of disease-resistant cotton varieties and effective management strategies, ultimately reducing economic losses and promoting sustainable farming.

Field surveys conducted from October 2023 to January 2024 ensured meticulous image capture under diverse conditions. The images are categorized into eight classes, representing specific disease manifestations, pests, or environmental stress in cotton plants. The dataset comprises 2137 original images and 7000 augmented images, enhancing deep learning model training. The Inception V3 model demonstrated high performance, with an overall accuracy of 96.03%. This underscores the dataset's potential in advancing automated disease detection in cotton agriculture. The presence of diseases such as Cotton Leaf Curl Disease and leaf hopper in cotton plants poses significant challenges to farmers worldwide, leading to substantial yield losses, reduced crop quality, and economic hardships. Timely detection and effective management of these diseases are critical for sustaining cotton production and ensuring food security. Therefore, identifying plant diseases at the early stage will benefit in diagnosing and preventing unnecessary crop loss.

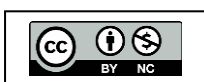
Among different parts of the plant, the leaf is the part that affects the crop yield if it gets affected. Visible symptoms can help in the detection of disease, and plant pathologists can suggest a suitable pesticide. Traditional methods of disease detection often rely on manual inspection by agronomists, which can be time-consuming, subjective, and prone to human error. By contrast, utilizing computer vision and machine learning techniques with the aid of a comprehensive cotton leaf disease dataset can revolutionize disease detection and management practices. (Bishshash, et. al. 2024).

III. PROPOSED METHODOLOGY

The methodology employed to develop and implement the Cotton Crop Disease Prediction System using Machine Learning. It outlines the process from data collection to model evaluation, explaining how each step contributes to a robust prediction system. The methodology encompasses data preprocessing, model selection, training, evaluation, and deployment.

1. Data Collection and Dataset Preparation:

The initial step involves creating a comprehensive image dataset of healthy and diseased cotton plants under various environmental conditions.





- **Data Sources:** Images will be sourced from agricultural research institutions, public repositories, and local farms, capturing various cotton diseases (e.g., bacterial blight, leaf spot, fungal infections) and their symptoms (e.g., leaf discoloration, lesions, spots).
- **Data Labeling:** Agricultural experts will manually label the images, classifying plants as healthy or affected by specific diseases. Labels may also specify the disease type for multi-class classification.
- **Data Augmentation:** Techniques like rotation, flipping, cropping, and scaling will artificially expand the dataset to improve model robustness and handle variations in orientation, scale, and lighting.
- **Data Preprocessing:** Images will be preprocessed by:
 - Resizing to a consistent dimension (e.g., 224x224 pixels).
 - Normalizing pixel values (0-1).
 - Removing irrelevant background information.
 - Splitting the dataset into training, validation, and testing sets

2. Feature Extraction:

This focuses on identifying key image features that differentiate healthy and diseased plants. While CNNs learn features automatically, initial feature extraction can enhance performance.

- **Color Features:** Leaf color variations (e.g., yellowing, browning, dark spots) will be analyzed.
- **Texture Features:** Leaf texture (roughness, smoothness, patterns) will be analyzed using methods like GLCM.
- **Shape Features:** Leaf shape deformations and irregularities will be analyzed using techniques like contour analysis and edge detection.

3. Model Selection and Architecture:

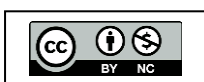
A Convolutional Neural Network (CNN) will be used for image classification due to its proven effectiveness.

- **CNN Architecture:** The CNN will consist of convolutional, pooling, and fully connected layers. Convolutional layers learn local patterns, pooling layers reduce dimensionality, and fully connected layers combine features for prediction.
- **Transfer Learning:** A pre-trained CNN model (e.g., VGG16, ResNet, MobileNet) will be fine-tuned on the cotton disease dataset, leveraging previously learned features and accelerating training.
- **Fine-Tuning:** The top classification layer of the pre-trained model will be adapted for the specific classification task (binary or multi-class) and retrained on the cotton disease dataset.

4. Model Training and Hyperparameter Tuning:

Preprocessed images will be fed into the CNN for training.

- **Training Process:** Backpropagation and gradient descent optimization will be used to minimize the difference between predicted and actual labels, using cross-entropy loss.



- **Hyperparameter Tuning:** Techniques like grid search or random search will optimize hyperparameters (learning rate, batch size, epochs, dropout rate).
- **Cross-Validation:** k-fold cross-validation will prevent overfitting and provide a more reliable performance estimate.

5. Model Evaluation and Performance Metrics:

The trained model will be evaluated on the unseen testing dataset.

- **Metrics:** Accuracy, precision, recall, and F1-score will be calculated.
- **Confusion Matrix:** A confusion matrix will visualize model performance and identify potential biases.

6. Model Deployment and User Interface:

The trained and evaluated model will be deployed in a user-friendly application.

- **Web Application:** A web-based app will allow farmers to easily upload images and receive instant disease predictions.

IV. SYSTEM ARCHITECTURE

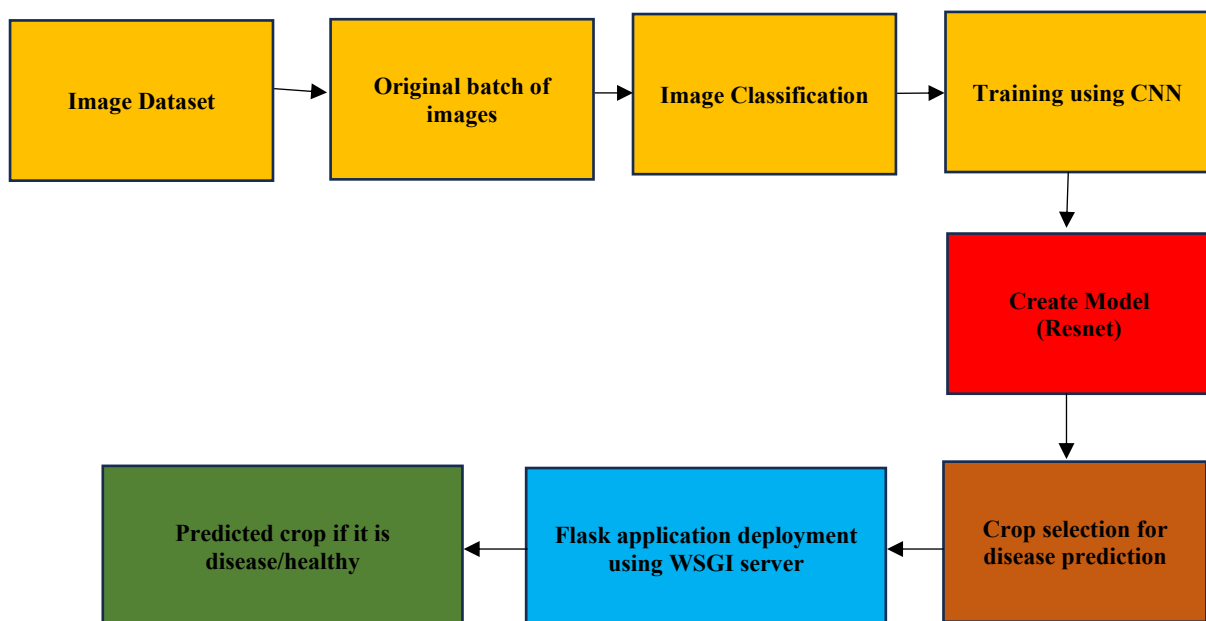


Figure 1: System Architecture

1. Data collection is the very first step. We use a self-generated dataset to detect whether the leaf is of cotton or not, another dataset is used to detect disease type.
2. Preprocessing of collected dataset, Model created by keras library through which we were able to do image processing tasks.

3. Building CNN (Convolutional Neural Network) Model (ResNet50 Architecture) for classification of various cotton diseases.
4. The Developed Model will be deployed on WSGI Server.
5. Developing a flask web application that uses API deployed on the local host through WSGI server.

V. SYSTEM MODULES

Modules Used in Project

1. Tensorflow
2. Numpy
3. Pandas
4. Matplotlib
5. Scikit – learn
6. Nltk

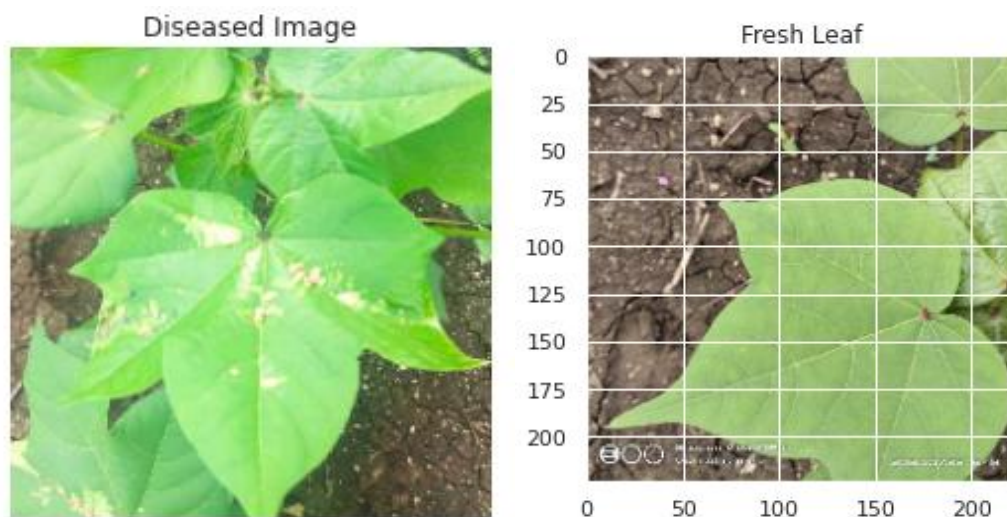


Figure 2: Plant Visualization

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 222, 222, 32)	896
max_pooling2d (MaxPooling2D)	(None, 111, 111, 32)	0
conv2d_1 (Conv2D)	(None, 109, 109, 64)	18496
max_pooling2d_1 (MaxPooling2)	(None, 54, 54, 64)	0
conv2d_2 (Conv2D)	(None, 52, 52, 64)	36928
flatten (Flatten)	(None, 173056)	0
dense (Dense)	(None, 64)	11075648
dense_1 (Dense)	(None, 4)	260
Total params: 11,132,228		
Trainable params: 11,132,228		
Non-trainable params: 0		

VI. RESULTS

The Cotton Crop Disease Prediction System using Machine Learning is designed to enhance the efficiency and accuracy of disease diagnosis in cotton crops. This chapter outlines the anticipated results and benefits of implementing this system, focusing on the improvements it will bring to agricultural practices, the advantages of using machine learning for disease prediction, and the potential impact on cotton farming at large.

RUN COTTON CROP DISEASE PREDICTION

Step 1: Choose image | Step 2: Once image upload | Step 3: Click predict for disease predict

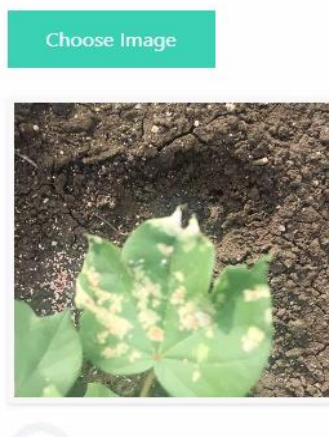
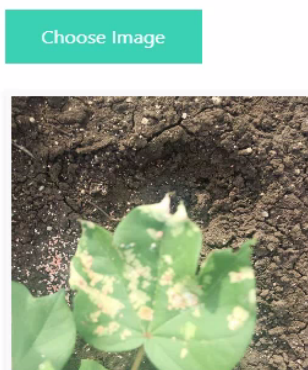


Figure 3: Image Selection

RUN COTTON CROP DISEASE PREDICTION

Step 1: Choose image | Step 2: Once image upload | Step 3: Click predict for disease predict



The leaf is diseased cotton leaf

Figure 4: Prediction of Leaf as Diseased

RUN COTTON CROP DISEASE PREDICTION

Step 1: Choose image | Step 2: Once image upload | Step 3: Click predict for disease predict



The leaf is fresh cotton leaf

Figure 5: Prediction Result as Fresh Leaf with No Disease

VII. CONCLUSION

In this project, we have proposed a deep learning-based approach using CNNs for cotton disease prediction. We have collected a self-generated dataset consisting of images of cotton plants infected with bacterial blight, leaf curl, leaf spot, powdery mildew, rust, and wilt. We have used transfer learning to fine-tune a pre-trained CNN model for our specific classification task and evaluated its performance on a test dataset. The proposed system has the potential to be a reliable tool for the early detection and accurate diagnosis of cotton diseases, which can significantly improve crop management practices and reduce economic losses.

The results of our experiments show that the proposed approach achieves high accuracy in classifying the six types of cotton diseases, demonstrating the effectiveness of CNNs in the field of agricultural technology. The self-generated dataset can also be used for further research and development in the field of computer vision and agricultural technology. Overall, this project contributes to the development of technology that can enhance the efficiency and sustainability of cotton cultivation practices.

The proposed system can provide a more accurate and efficient alternative to traditional methods of cotton disease diagnosis and management, which can lead to better crop yields and quality, as well as reduced economic losses. Further research can be done to improve the accuracy and efficiency of the proposed approach, and to expand its application to other crops and agricultural practices.

This project introduces a deep learning-based solution for the early detection and accurate classification of cotton diseases using Convolutional Neural Networks (CNNs). We have leveraged a self-generated dataset, encompassing various images of cotton plants affected by bacterial blight, leaf curl, leaf spot, powdery mildew, rust, and wilt, to train a CNN model. The dataset, curated specifically for this research, provides a robust foundation for the model's training and evaluation.



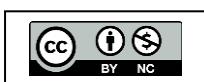
By implementing transfer learning techniques, we fine-tuned a pre-trained CNN architecture to cater to our specific classification task, ensuring the model efficiently distinguishes between the different disease types. Our experiments demonstrate that the proposed system achieves high accuracy in classifying the six cotton diseases, confirming the viability of CNNs in solving complex problems in agricultural technology. The results highlight the potential of deep learning methods to significantly improve the accuracy and speed of disease diagnosis, which, in turn, can help mitigate the impact of these diseases on cotton crop yields.

The self-generated dataset, along with the trained CNN model, contributes to the advancement of research in both agricultural technology and computer vision, offering a valuable resource for further exploration. The application of CNNs in this domain is a step toward automating the monitoring and management of crops, ultimately leading to smarter farming practices. By providing an efficient, reliable, and scalable method for disease detection, this project presents a more effective alternative to traditional manual methods of diagnosis, which can often be slow, error-prone, and costly.

Moreover, the system's implementation can lead to better crop management decisions, optimized resource use, and a reduction in the use of pesticides, thereby promoting more sustainable agricultural practices. Beyond cotton, this approach has the potential to be adapted to other crops, enabling broader applications across different agricultural domains. Future research can focus on further enhancing the model's accuracy, incorporating additional disease types, and exploring real-time applications in field settings. In essence, this project marks a significant contribution to the evolving intersection of technology and agriculture, offering a path toward more sustainable and productive farming practices.

REFERENCES

- [1] Ranjana Jadhav, Karanjawane Vaishnavi, Mahajan Isha, Karpe Vedant, Gaikwad Bhagyesh, Ingale Mohit, "Cotton Disease Prediction System", Educational Administration: Theory and Practice 2024, 30(6), 661-666 ISSN: 2148-2403
- [2] Ketki Kshirsagar, Anup Ingle, Ruqaiyya Ansari, Vaishnavi Sambare, Achal Niswade, "Cotton Leaf Disease Detection Using Machine Learning", Afr.J.Bio.Sc. 6(16) (2024) ISSN: 2663-2187
- [3] Prayma Bishshash, Asraful Sharker Nirob, Habibur Shikder, Afjal Hossan Sarower, Touhid Bhuiyan, Sheak Rashed Haider Noori, "A comprehensive cotton leaf disease dataset for enhanced detection and classification", Data in Brief 57 (2024) 110913
- [4] Amit Jadhav, Shubham Chaudhari, Ankush Gupta, Dr. S. A. Mithavkar, "Cotton Disease Detection and Cure Using CNN", International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
- [5] Vishal Pallagani; Vedant Khandelwal; Bharath Chandra; Venkanna Udutalapally; Debanjan Das; Saraju P. Mohanty V. Pallagani, V. Khandelwal, B. Chandra, V. Udutalapally, D. Das and S. P. Mohanty, "dCrop: A Deep-Learning Based Framework for Accurate Prediction of Diseases of Crops in Smart Agriculture, " 2019 IEEE International Symposium on Smart Electronic Systems (iSES) (Formerly iNIS), Rourkela, India, 2019, pp. 29-33, doi: 10.1109/iSES47678.2019.000 20.
- [6] Al-Amin; Tasfia Anika Bushra; Md Nazmul Hoq M. Al-Amin, T. A. Bushra and M. Nazmul Hoq, "Prediction of Potato Disease from Leaves using Deep Convolution Neural Network towards a Digital Agricultural System, " 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), Dhaka, Bangladesh, 2019, pp. 1-5, doi: 10.1109/ICASERT.2019.89349 33.
- [7] P. Revathi and M. Hemalatha, "Classification of cotton leaf spot diseases using image processing edge detection techniques," 2012 International Conference on Emerging Trends in Science, Engineering and Technology (INCOSSET), Tiruchirappalli, India, 2012, pp. 169-173, doi: 10.1109/INCOSSET.2012.6513900.





- [8] Islam, Manowarul & Talukder, Md. Alamin & Uddin, Md Ashraf & Akhter, Arnisha & Sharmin, Selina & Mamun, Md & Debnath, Sumon. (2023). A Deep Learning Model For Cotton Disease Prediction Using Fine- Tuning With Smart Web Application In Agriculture. Intelligent Systems with Applications. 20. 200278. 10.1016/j.iswa.2023.200278.
- [9] Rani, S., Mishra, A.K., Kataria, A. et al. Machine learning-based optimal crop selection system in smart agriculture. Sci Rep 13, 15997 (2023). <https://doi.org/10.1038/s41598-023-42356-y>.
- [10] Kumar, Kadem & Ramesh Chandra, Gollapudi & Sukheja, Deepak. (2020). Cotton Disease Detection using Deep Learning. International Journal of Innovative Technology and Exploring Engineering. 9. 152-156. 10.35940/ijitee.D1391.029420.
- [11] Sandeep Kumar, Rajeev Ratan, J. V. Desai, "Cotton Disease Detection Using TensorFlow Machine Learning Technique", Advances in Multimedia, vol. 2022, Article ID 1812025, 10 pages, 2022. <https://doi.org/10.1155/2022/1812025>
- [12] Azath M. , 1 Melese Zekiwo,2 and Abey Bruck1. Journal of Electrical and Computer Engineering Volume 2021, Article ID 9981437, 10 pages <https://doi.org/10.1155/2021/9981437> . Deep Learning Based Image Processing for Cotton Leaf Disease and Pest Diagnosis.
- [13] Sharvari V. Patil, Anjali K. Sharma, Bhagyashree R. Kamble, Kajal B. Jadhav International Journal of Creative Research Thoughts (IJCRT) Volume 10, Issue 5 May 2022 | ISSN: 2320-2882 COTTON LEAF DISEASE DETECTION USING DEEP LEARNING.
- [14] Bhoomika K1, L N Sowmya Shree2, Monisha A3, Vinaya Hegde4, Mr. Dheeraj.D5 International Research Journal of Engineering and Technology (IRJET) Volume: 10 Issue: 03 | Mar 2023 SURVEY ON COTTON PLANT DISEASE DETECTION.
- [15] Pratiti Saha, Dr. Nachappa MN International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538 Volume 10 Issue III Mar 2022- Available at www.ijraset.com.
- [16] Rahul Mhatre, Vishal Lanke e-ISSN: 2582-5208 International Research Journal of Modernization in Engineering Technology and Science Volume:03/Issue:01/January-2021 COTTON LEAVES DISEASE DETECTION AND CURE USING DEEP LEARNING.
- [17] Rafael Faria Caldeira, Wesley Esdras Santiago, and Barbara Teruel Caldeira RF, Santiago WE, Teruel B. Identification of Cotton Leaf Lesions Using Deep Learning Techniques. Sensors (Basel). 2021 May 3;21(9):3169. doi: 10.3390/s21093169. PMID: 34063578; PMCID: PMC8124293.
- [18] Rubini PE , Dr.Kavitha P P. Rubini and P. Kavitha, "Deep Learning model for early prediction of plant disease, " 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), Tirunelveli, India, 2021, pp. 1104-1107, doi: 10.1109/ICICV50876.2021.9388538.
- [19] Shaikh Farhad Hossain; Sony Hartono Wijaya; Ming Huang; Irmanida Batubara; Shigehiko Kanaya; Md. Altaf-UIAmin Farhad S. F. Hossain, S. H. Wijaya, M. Huang, I. Batubara, S. Kanaya and M. A. -U. -A. Farhad.

