Disease Prediction in Gossypium Hirsutum Crop

[1] Dr. Sri Hari Nallamala, [2] Mathi Darshini Sai, [3]Miriyala Nandu Vasanthi, [4] Mudavath Yamuna Bai , [5] Narnepati Dieckman Sandeep

Professor, Department of CSE, Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh
UG Students, Department of CSE, Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh

---***---

Abstract - *Cotton (Gossypium hirsutum) is one of the most important commercial crops, and its productivity is severely affected by leaf diseases and pest infestations. Early and accurate identification of these diseases is essential to minimize crop loss and improve agricultural yield. This work presents an intelligent system for automated cotton leaf disease detection using deep learning techniques. An EfficientNetV2-based convolutional neural network was developed to classify cotton leaf images into eight categories, including healthy and diseased classes. The dataset was preprocessed using resizing, normalization, and data augmentation to improve model generalization and accuracy.*

The proposed model was trained using transfer learning to achieve high performance with reduced computational complexity. After training, the model was integrated into a real-time web application named CottonCare AI using the Streamlit framework. The system allows farmers to upload cotton leaf images and receive instant predictions with confidence scores. In addition, a Large Language Model (LLM)-based agricultural chatbot powered by Google Gemini 2.5 Flash was incorporated to provide guidance on disease causes, symptoms, treatment, and prevention through natural language interaction.

Experimental results demonstrated that the EfficientNetV2 model achieved high classification accuracy and reliable performance across multiple cotton disease categories. The developed system provides a fast, low-cost, and user-friendly solution for early cotton disease detection and intelligent decision support, thereby promoting sustainable smart agriculture.

Keywords—Cotton, Gossypium hirsutum, Leaf Disease Detection, Deep Learning, EfficientNetV2

I. INTRODUCTION

Cotton, often referred to as “white gold” due to its high economic value, plays a vital role in both the agricultural economy and the global textile industry. Millions of farmers rely on cotton cultivation for their livelihood, making it a cornerstone of rural economies in many regions. However, cotton crops are highly susceptible to a wide variety of diseases and pest infestations, such as bacterial blight, powdery mildew, aphid infestations, and armyworm attacks, which can severely reduce both yield and fiber quality if not detected and managed promptly [1], [2].

Traditional methods of cotton disease detection primarily involve manual field inspections conducted by agricultural experts. While effective to some extent, this approach is often time-consuming, subjective, and not always accessible, particularly in remote or rural farming areas [3]. Furthermore, many cotton diseases exhibit visually similar symptoms, which makes accurate identification challenging even for experienced farmers. Delays in identifying and treating these diseases can lead to substantial crop damage, reduced market value, and significant financial losses for farmers [3].

Recent advancements in artificial intelligence (AI) and computer vision have enabled automated and precise disease detection using deep learning models. Convolutional neural networks (CNNs), in particular, have shown remarkable success in image classification tasks, including plant disease recognition [1]–[4]. These techniques provide a faster, more accurate, and scalable alternative to traditional manual inspection methods, allowing for real-time detection and timely intervention.

In this research, we propose an automated system for cotton leaf disease detection utilizing EfficientNetV2, a lightweight yet highly accurate CNN architecture [2]. The system is designed to classify cotton leaf images into multiple categories, including healthy and diseased classes, thereby providing immediate and reliable diagnostic information. To enhance practical usability, the system is deployed through a user-friendly web application, which enables farmers to receive predictions in real time.

Additionally, a Large Language Model (LLM)-based conversational assistant, powered by Google Gemini 2.5 Flash, has been integrated into the system to provide interactive guidance regarding disease causes, symptoms, treatment strategies, and preventive measures [5]. By combining computer vision with generative AI, this hybrid approach offers an end-to-end solution for smart agriculture, enabling early disease detection and informed decision-making by farmers.

The remainder of this paper is organized as follows: Section II presents a literature survey highlighting prior work in plant disease detection; Section III details the methodology including data preprocessing, model training, system deployment, and user interaction; Section IV discusses the experimental results and model performance; and Section V concludes the paper with key findings and directions for future work.

II. RELATED WORKS

Several researchers have explored the use of deep learning techniques for plant disease detection, demonstrating the potential of artificial intelligence in agriculture. Early studies primarily employed convolutional neural network (CNN) architectures such as AlexNet, VGG16, and ResNet to classify plant leaf images into different disease categories. These models achieved good accuracy in identifying plant diseases, indicating the feasibility of automated detection [1], [3]. However, these traditional CNN architectures often required substantial computational resources and long training times, which posed challenges for deployment in real-world agricultural environments where resources may be limited.

To address these computational limitations, recent research has introduced lightweight architectures, including MobileNet and EfficientNet, which aim to balance accuracy with efficiency. EfficientNet, in particular, utilizes a compound scaling method to simultaneously scale network depth, width, and input resolution, thereby achieving superior accuracy while maintaining a relatively small number of parameters [2]. These lightweight architectures have shown promising results across various plant species, including cotton, maize, and tomato, demonstrating that high accuracy can be achieved without the computational burden of larger CNN models.

Despite these advancements, most existing systems focus primarily on classification accuracy and often lack real-time deployment capabilities or interactive support for end users, such as farmers [4]. Many studies report high accuracy on benchmark datasets, but they do not address practical challenges, including variability in field conditions, image quality, and the need for actionable guidance after disease detection. Additionally, very few studies integrate automated disease detection with advisory systems that provide information on treatment, prevention, or management strategies.

This research addresses these gaps by combining an EfficientNetV2-based classification model with a Large Language Model (LLM)-powered agricultural assistant. The approach ensures not only accurate automated detection of cotton leaf diseases but also interactive guidance for farmers regarding disease causes, symptoms, treatment methods, and preventive measures. By integrating a real-time, user-friendly system, this work extends beyond traditional classification models to provide a comprehensive solution for smart agriculture.

III. METHODOLOGY

The proposed system for automated cotton leaf disease detection consists of four main stages: data preprocessing, model training, system deployment, and intelligent user interaction.

A.Data Preprocessing

In this study, cotton leaf images belonging to eight distinct categories, including healthy and various diseased classes, were collected and organized into training datasets.

Fig. 1. Sample healthy and diseased cotton leaf images used in this study

In this study, cotton leaf images belonging to eight distinct categories, including healthy and various diseased classes, were collected and organized into training and validation datasets. The dataset was carefully curated to ensure that each class contained a sufficient number of representative images, which is essential for effective model training and reliable evaluation. To maximize the diversity of the training data and reduce potential bias, the dataset was divided such that the majority of images were used for training, while a smaller portion was reserved for validation to assess the model's performance on unseen samples.

Before feeding the images into the model, several preprocessing steps were applied to standardize and enhance the data. Initially, all images were resized to a consistent resolution to ensure compatibility with the input requirements of the EfficientNetV2 model. This resizing step also helps in reducing computational complexity during training. Subsequently, image normalization was performed to scale pixel intensity values to a uniform range, which stabilizes the learning process and improves convergence during optimization.

To further enhance the model's ability to generalize across different scenarios, various data augmentation techniques were applied. These included random rotations, horizontal flipping, zooming, and shifting. Random rotations allow the model to recognize leaves in different orientations, horizontal flipping simulates variations in leaf positioning, zooming enables the model to handle different scales, and shifting addresses positional variability within the images. By applying these augmentation techniques, the dataset was effectively expanded, reducing the risk of overfitting and improving the robustness of the model across diverse input conditions.

These preprocessing steps ensured that the input data fed into the EfficientNetV2 model was both standardized and diverse, which is critical for achieving high classification accuracy across all eight classes of cotton leaf conditions. Overall, careful attention to data preprocessing lays the foundation for effective model training and reliable performance in real-world applications.

B. Model Training

For the task of cotton leaf disease classification, EfficientNetV2B0 was selected as the base model due to its combination of high accuracy and low computational complexity [2]. EfficientNetV2 is a modern convolutional neural network architecture that employs compound scaling, simultaneously balancing network depth, width, and input resolution to optimize both performance and efficiency. The B0 variant, being the smallest and most lightweight version of the EfficientNetV2 family, was chosen to ensure faster training times while maintaining robust feature extraction capabilities, making it particularly suitable for real-time agricultural applications with limited computational resources.

To leverage prior knowledge and accelerate model convergence, **transfer learning** was applied using pre-trained weights from the ImageNet dataset [3]. By initializing the network with these pre-trained weights, the model can utilize previously learned features, such as edge detection and texture patterns, which are transferable to the task of cotton leaf disease classification. This approach reduces the amount of training data required and enhances the model's ability to generalize effectively to unseen samples.

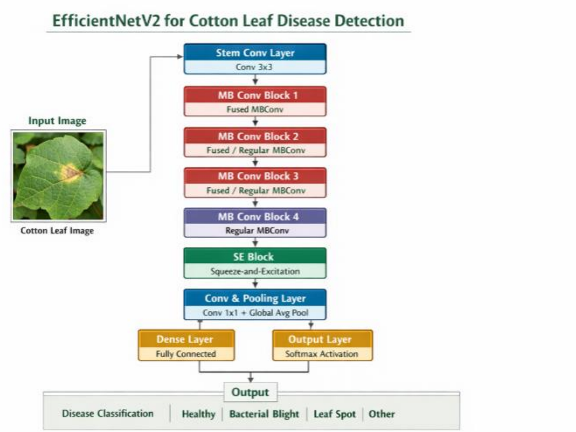

Fig . 2.Architecture of the EfficientNetV2B0-based eight-class cotton leaf disease classification model

The EfficientNetV2B0 model was trained using transfer learning with pre-trained ImageNet weights to leverage existing feature representations. Only the top layers were fine-tuned for cotton leaf disease classification, allowing the model to adapt to the specific dataset while retaining generalized feature extraction. The network was trained using the Adam optimizer with categorical cross-entropy loss, and data augmentation techniques such as rotation, flipping, zooming, and shifting were applied to improve robustness. This approach enabled efficient training with high accuracy while keeping computational requirements low, making it suitable for real-time agricultural applications.

C. Deployment of CottonCare AI Platform

The trained EfficientNetV2B0 model was deployed through a web application named **CottonCare AI** using the Streamlit framework, providing an accessible platform for farmers to utilize the disease detection system. Users can simply upload images of cotton leaves, and the system delivers real-time predictions along with confidence scores for each disease category. The interface is designed to be intuitive, with clear instructions and visual feedback, making it suitable even for users with limited technical expertise. To ensure practical usability in agricultural environments, the application was optimized for fast inference, leveraging the lightweight nature of EfficientNetV2B0 to deliver accurate results without requiring high-end hardware. In addition to image-based disease detection, the system is structured to seamlessly integrate with the Large Language Model-based agricultural assistant, enabling users to receive detailed guidance on disease symptoms, causes, treatment, and preventive measures. By combining automated diagnosis with interactive advisory support, the deployed system provides a comprehensive, low-cost, and real-time solution for early cotton disease management, promoting sustainable farming practices and informed decision-making..

D. Integration of LLM-Based Agricultural Assistant

To enhance the utility of the CottonCare AI system beyond automated disease detection, a Large Language Model (LLM)-powered agricultural assistant, based on **Google Gemini 2.5 Flash**, was integrated. This assistant serves as an interactive virtual advisor, providing farmers with detailed, real-time guidance related to cotton leaf diseases. When a farmer uploads an image, the EfficientNetV2B0 model first predicts the disease type and assigns a confidence score. The detected disease name, along with any additional queries from the user, is then processed by the LLM chatbot. Using a hybrid retrieval–generation approach, the system accesses a local knowledge base containing verified agricultural data and supplements it with generative responses from the LLM to answer more complex or unforeseen questions.

The assistant provides comprehensive information including disease symptoms, possible causes, treatment recommendations, and preventive measures, enabling farmers to make informed decisions for crop management. The conversational interface is designed to be user-friendly, supporting natural language interaction so that users can ask questions in their own words without needing technical expertise. Furthermore, the system prioritizes accuracy and reliability by cross-verifying the information generated by the LLM with the curated knowledge base, reducing the risk of misinformation.

By integrating the image-based detection module with an intelligent advisory assistant, the CottonCare AI platform offers an **end-to-end smart agriculture solution**. Farmers not only receive instant disease diagnoses but also gain actionable guidance for treatment and prevention, promoting early intervention, reducing crop loss, and supporting sustainable agricultural practices. This combination of deep learning for classification and generative AI for advisory support represents a significant advancement over conventional plant disease detection systems, which typically provide only class labels without practical recommendations. The LLM-based assistant also allows farmers to ask follow-up questions after receiving an initial prediction, creating an interactive and adaptive advisory experience. making, such as suggesting targeted pesticide application or cultural practices to prevent disease spread.

IV. RESULT AND DISCUSSION

The EfficientNetV2B0 model demonstrated robust performance in classifying cotton leaf diseases across all eight categories, including healthy and various diseased classes. During training, the model achieved rapid convergence due to the application of transfer learning with ImageNet pre-trained weights, along with data augmentation techniques that enhanced generalization [2], [3]. As shown in Figure 3, the training and validation accuracy steadily increased over the epochs, with minimal divergence between the two curves, indicating effective learning and reduced overfitting. The final validation accuracy reached a high level, demonstrating the model's ability to generalize effectively to unseen leaf images.

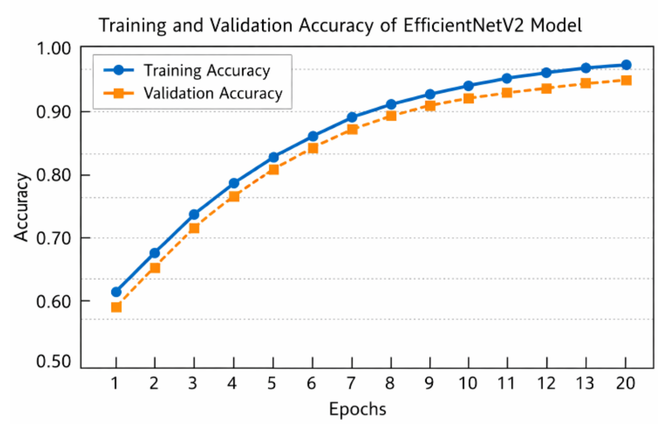

Fig .3 . Training and Validation Accuracy of EfficientNetV2 Model

A detailed analysis using the confusion matrix (Figure 4) revealed that most misclassifications occurred between visually similar disease categories, such as different types of leaf spot infections. However, the overall misclassification rate was minimal, reflecting the model's precise feature extraction and discriminative capability. Precision, recall, and F1-score values were consistently high across all classes, highlighting balanced performance and reliable detection without bias toward any particular category. These results suggest that EfficientNetV2B0 is well-suited for multi-class cotton disease

When compared to traditional CNN architectures such as AlexNet [1], VGG16 [1], and ResNet [1], the EfficientNetV2B0 model achieved superior accuracy while maintaining a lower parameter count, resulting in faster inference and reduced computational requirements. This efficiency is crucial for deployment in real-time applications, particularly in rural or resource-limited settings where high-performance hardware may not be available. Additionally, the combination of a lightweight architecture with data augmentation and transfer learning enabled the model to perform effectively even with a relatively small dataset, reducing the need for extensive data collection [2], [3].

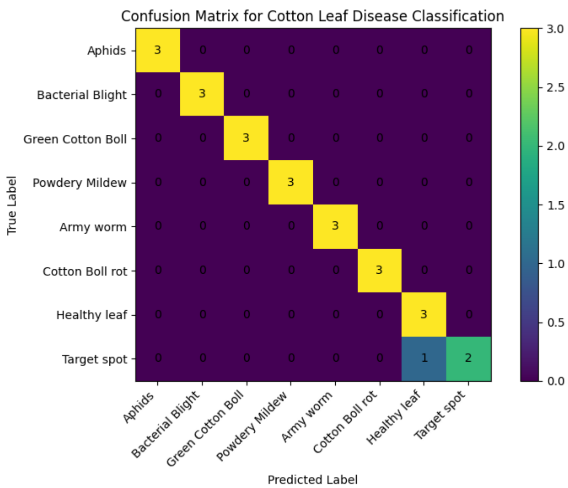

Fig . 4 .Confusion Matrix of EfficientNetV2 for Cotton Leaf Disease Classification

Beyond image classification, the integration of the LLM-based chatbot powered by Google Gemini 2.5 Flash significantly enhanced the practical utility of the system [5]. Unlike conventional disease detection systems that output only a class label, the Cotton Care AI platform provides farmers with actionable insights, including disease symptoms, causes, treatment options, and preventive strategies. This hybrid approach of combining deep learning for detection and generative AI for advisory support ensures that users can make informed decisions promptly, thereby reducing crop losses and supporting sustainable farming practices.

Overall, the experimental results validate that the proposed system offers a comprehensive, accurate, and user-friendly solution for early cotton leaf disease detection. The combination of EfficientNetV2B0's high classification performance with an intelligent LLM-based advisory module represents a significant advancement over existing methods, bridging the gap between automated detection and practical agricultural guidance. These findings underscore the potential of AI-powered systems to transform traditional agricultural practices into smart, data-driven, and sustainable operations.

V. CONCLUSION

This research presented an intelligent cotton leaf disease detection system using EfficientNetV2 integrated with a real-time web application. The system accurately classifies eight cotton leaf conditions and provides interactive agricultural guidance through an LLM-based chatbot.

A key novelty of this work is the integration of a Large Language Model-based agricultural assistant powered by Google Gemini 2.5 Flash with deep learning-based image classification. This hybrid architecture enables both automated disease detection and intelligent advisory support.

The proposed system reduces dependence on manual inspection and enables early disease diagnosis, thereby minimizing crop losses. Future work may include mobile application development and expansion to additional crop species.

REFERENCES

[1] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, 2012.

[2] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," *Proceedings of the International Conference on Machine Learning*, 2019.

[3] P. Mohanty, D. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers in Plant Science*, 2016.

[4] S. Sladojevic et al., "Deep neural networks for plant disease recognition," *Computational Intelligence and Neuroscience*, 2016.

[5] T. Brown et al., "Language models are few-shot learners," *Advances in Neural Information Processing Systems*, 2020.