iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: III Month of publication: March 2025

DOI: https://doi.org/10.22214/ijraset.2025.67475

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue III Mar 2025- Available at www.ijraset.com

Identification and Classification of Plant Disease

Yash More[1], Nithin Suvarna[2], Om Patil[3], Rudresh Thakur[4], Sandesh Patil[5]
Department of Data Engineering, Mumbai University, India

***Abstract:** Agricultural sciences have long struggled with challenges such as plant diseases, declining crop yields, and the need for sustainable farming solutions. Accurate and rapid plant identification plays a crucial role in biodiversity monitoring, crop management, and early disease detection. However, traditional plant identification methods are often time consuming, prone to human error, and require expert knowledge. Deep learning, a branch of artificial intelligence, provides an effective solution to these challenges. By leveraging neural networks, deep learning can recognize complex patterns in plant images, enabling precise identification and classification based on features such as leaf shape, colour, and texture. Among deep learning models, Convolutional Neural Networks (CNNs) have proven particularly effective in automating feature extraction and classification processes, significantly enhancing plant recognition capabilities. This project explores the application of deep learning techniques to accurately classify plant species and detect diseases. By training a deep learning model on a dataset of plant images, the goal is to develop a tool that can assist farmers, botanists, and researchers in identifying plants and diagnosing diseases more efficiently. This could lead to improved agricultural productivity, better disease management, and enhanced biodiversity conservation.*
***Keywords:** Deep Learning, Plant Identification, Convolutional Neural Networks (CNNs), Agriculture, Disease Detection*

I. INTRODUCTION

Global food security and economic stability, especially in developing nations, are largely dependent on agriculture. However, crop yields and quality are seriously threatened by the rising incidence of plant illnesses brought on by pathogens like fungi, bacteria, and viruses, which can result in large financial losses and food shortages. In order to ensure sustainable agricultural practices, early detection and management of these illnesses are essential. Conventional disease detection techniques, which depend on expert knowledge and manual inspection, are frequently labour-intensive, time-consuming, and prone to human mistake [1][6]. There is an increasing chance to create automated, effective, and precise methods for plant disease diagnosis because to developments in deep learning, especially Convolutional Neural Networks (CNNs) [2][7]. The goal of this project is to develop an automated system for recognizing and categorizing plant diseases by utilizing deep learning techniques, particularly CNNs. utilizing a CNN model, the suggested approach achieves an astonishing 99% accuracy utilizing a dataset of over 64,000 photos spanning 20 categories of healthy and ill plants. The study's objectives are to, create a strong deep learning model for precise disease identification assess how well it performs in comparison to current techniques and offer an affordable, scalable solution that can be included into farming operations. A user-friendly web application will make the technology available, allowing farmers to input leaf photos for immediate disease detection and treatment.

II. LITERATURE REVIEW

The desire to increase agricultural output and guarantee global food security has spurred notable breakthroughs in the field of deep learning-based plant disease identification in recent years. In order to create automated systems for identifying and categorizing plant diseases from leaf photos, researchers have investigated a variety of deep learning approaches, most notably Convolutional Neural Networks (CNNs). In this part, important works are reviewed, their contributions are highlighted, and research gaps are identified.

1) Abraham Hirani [1] investigated the use of CNNs in the detection of plant diseases, highlighting how well they perform computer vision tasks such segmentation, feature extraction, and image classification. Transformer networks, which are excellent at identifying long-range dependencies in data, were compared to conventional CNN techniques in this study. Transformer networks demonstrated promise for better performance in particular situations, but CNNs are still a reliable option.
2) Ansari In order to identify diseases in maize plants, Fatima Anass [2] created a leaf disease detection system utilizing the VGG16 and VGG19 architectures. The study showed that VGG19 outperformed traditional feature-based techniques with an accuracy of 96.5% using a dataset of 3,000 coloured maize leaf images. This demonstrates how deep CNN models may be used to detect diseases in agricultural systems in real time.

3) With an emphasis on automating the identification of sick leaves, Sandeep Kumar [3] presented a machine learning-based system for leaf disease detection and classification. The research highlighted the financial consequences of plant illnesses and showed that the suggested algorithm outperformed current methods in terms of accuracy, demonstrating machine learning's potential to simplify farming methods.
4) To identify and categorize tomato leaf diseases, Melike Sardogan [4] used CNNs in conjunction with the Learning Vector Quantization (LVQ) technique. The study demonstrated the efficacy of this hybrid approach for precise and automated disease identification using a dataset of 500 photos encompassing four diseases, highlighting the significance of early detection for crop health.
5) Using a dataset of 160 photos, Shima Ramesh [5] used the Random Forest method to detect plant diseases, with a 70% classification accuracy. The study highlighted how machine learning techniques can be used to detect diseases on a broad scale, improving agricultural output and food security.

Although there are still a number of gaps, the field of deep learning-based plant disease diagnosis has advanced significantly. Lack of interpretability, high computing requirements of models, limited diversity and number of datasets, and inadequate interaction with agricultural operations are some of the difficulties. Furthermore, evaluations of the socioeconomic and environmental effects of these technologies are required, as are comparative studies of various architectures and real-time monitoring systems. Future studies should concentrate on building bigger and more varied datasets, making effective and understandable models, making user-friendly interfaces, and investigating real-time solutions to improve the viability and sustainability of deep learning-based agricultural systems.

III. METHODOLOGY

The Data gathering, preprocessing, model creation, training, validation, and deployment are some of the crucial steps in the Plant Disease Detection System technique. This all-encompassing strategy guarantees that the model accurately detects and categorizes plant diseases while giving end users reliable data.

1) *Data Collection:* To start the study, data was gathered using the Plant Village dataset as well as self-collected plant photos, which include a variety of pictures of both healthy and sick plants from different species [1][6]. To make the training process easier, this information was categorized into distinct directories according to classifications (healthy and various disease categories).
2) *Data Preprocessing:* To improve their quality and usability, the gathered photos were pre-processed. Among them was resizing. To guarantee uniformity, images were scaled to 256x256 pixels.

- Normalization In order to enhance model convergence during training, pixel values were normalized to fall within [0, 1].
- Augmenting Data The training data was made intentionally more diverse by using techniques including flips, random rotations, and brightness modifications [3][9]. By exposing the model to different image variations, this helped avoid overfitting.

3) *Model Development:* TensorFlow and Keras were used to create the deep learning model, which implemented a Convolutional Neural Network (CNN) architecture. To extract pertinent characteristics from the input images, the model consists of many convolutional layers followed by pooling layers. Important elements consist of: o Convolutional Layers To record spatial hierarchies, several convolutional layers with ReLU activation functions are used.

- Layers of Pooling To lower dimensionality while keeping significant characteristics, use max pooling layers.
- The Flattening Layer creates a 1D vector from 2D feature maps.
- Completely Interconnected Layers SoftMax activation is used for classification in order to generate probability for every class.

4) *Training of the Model and Dataset Split:* Ten percent of the dataset was used for testing, ten percent for validation, and eighty percent for training.

- The Optimizer The loss function (Sparse Categorical Cross Entropy) was optimized using the Adam optimizer.
- Measures of Performance Accuracy and loss measures were used to assess the model's performance.
- The Process of Training To make sure the model performed effectively when applied to new data, the training procedure entailed tracking both training and validation accuracy. When validation accuracy plateaued, early halting was used to avoid overfitting.

5) *Model Evaluation:* The model's classification performance was assessed using the test dataset following training. Among the important metrics computed are:

- Accuracy Measures the model's overall accuracy.
- Precision shows the percentage of actual favourable results among those that were anticipated.
- Recall shows the percentage of real positives that are true positives.
- The F1 Score offers a fair assessment of memory and precision.

The Confusion Matrix shows how well the model performs in various classifications.

6) *Model Deployment:* The finished model was put into use in a Web service that let users upload plant photos for in-the-moment disease identification. By integrating the trained model, the application gives users instant feedback on the plants' health. It has intuitive user interfaces that make interaction and result display simple.

IV. SYSTEM ARCHITECTURE

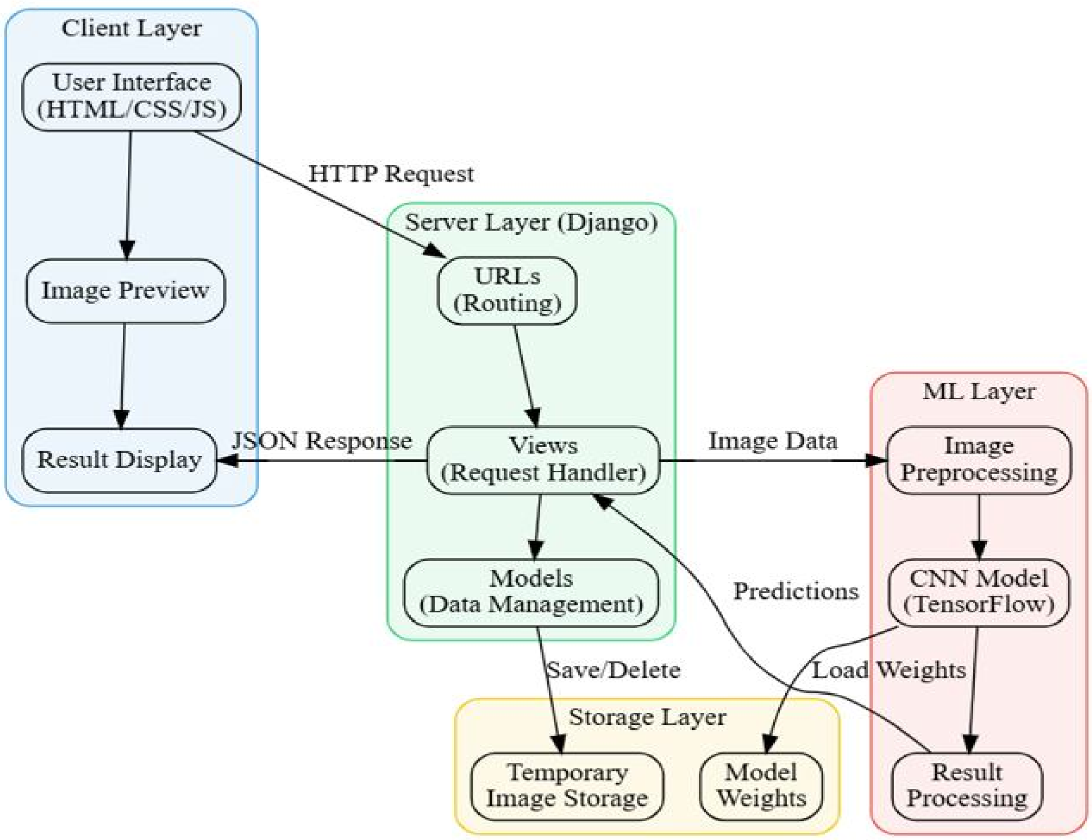


Fig. 4.1 System Architecture

The Frontend Layer, Backend Layer, and ML Component make up the three-tier architecture of the Plant Disease Detection System. From user contact to disease prediction and result dissemination, every layer is essential to maintaining smooth operation.

A. *Layer Frontend*

In addition to handling user interaction, the frontend layer offers a user-friendly interface for processing and uploading plant photos. One of the main features is the HTML5 Drag and Drop Interface, which makes it simple for users to upload photographs by simply dragging and dropping files into the program. Instantaneous Picture Preview ensures user trust in the input by showing a preview of the uploaded image as soon as it is selected. Loading animations and progress indicators to improve user experience, it offers visual feedback while processing images and making predictions about the model. Components of Responsive Design Ensures the application is accessible and functional across various devices, including desktops, tablets, and mobile phones.

B. Layer of the Backend

Core processing and communication between the frontend and the ML model are managed by the backend layer. Django Server is a crucial component that controls user authentication, HTTP requests, and data routing between the frontend and backend. Module for Image Preprocessing resizes, normalizes, and applies augmentation algorithms to uploaded photos in order to prepare them for analysis. In order to detect and classify diseases, ML Model Integration interfaces with the trained CNN model. The submitted photos and prediction results are stored and arranged by the file management system for further use and examination.

C. The ML Component

The system's central machine learning component is in charge of image analysis and plant disease prediction. Important components consist of CNN Model a deep learning model for feature extraction and classification that consists of several convolutional layers, pooling layers, and fully connected layers. Pipeline for Image Preprocessing applies augmentation techniques, normalizes pixel values, and resizes photos to guarantee consistency in input data. Class Prediction Using Scores for Confidence provides the anticipated illness class as well as a confidence score that shows how certain the model is of its prediction. With features to update or retrain the model as necessary, Model State Management makes sure the model is loaded and prepared for inference.

D. Data Movement

To guarantee effective processing and result delivery, the system adheres to a streamlined data flow.

1) Image Upload: Using the frontend interface, the user uploads a picture of a plant.
2) Preprocessing: To get the image ready for analysis, the backend performs preprocessing operations such scaling and normalization.
3) Model Prediction: The CNN model receives the preprocessed image and uses it to forecast the disease class and produce a confidence score.
4) Result Display: After being returned to the frontend, the prediction results are presented to the user in an understandable and intuitive manner.

V. EXPERIMENTATION & RESULTS

Fig 5.1 Plant Disease Detection System Application Interface

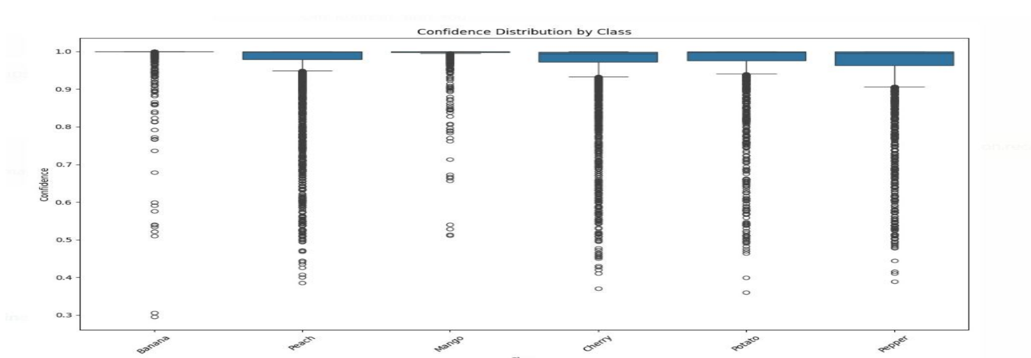


Fig 5.2 Confidence Score Distribution Across Different Plant Classes in Disease Classification

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue III Mar 2025- Available at www.ijraset.com

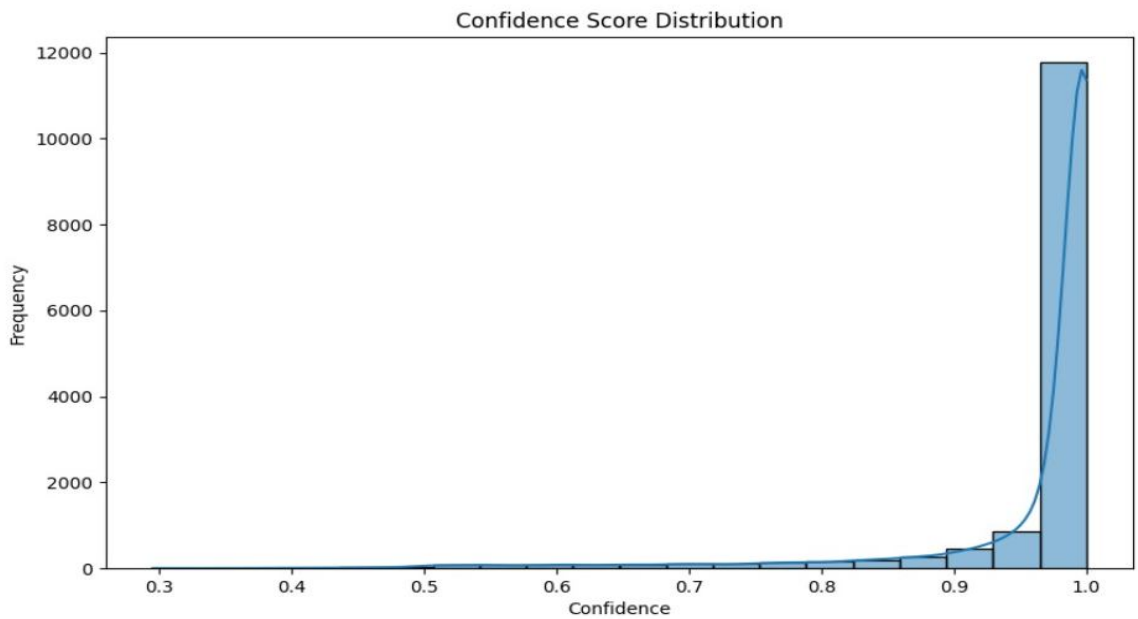


Fig. 5.3 Distribution of Confidence Scores in Disease Classification Predictions

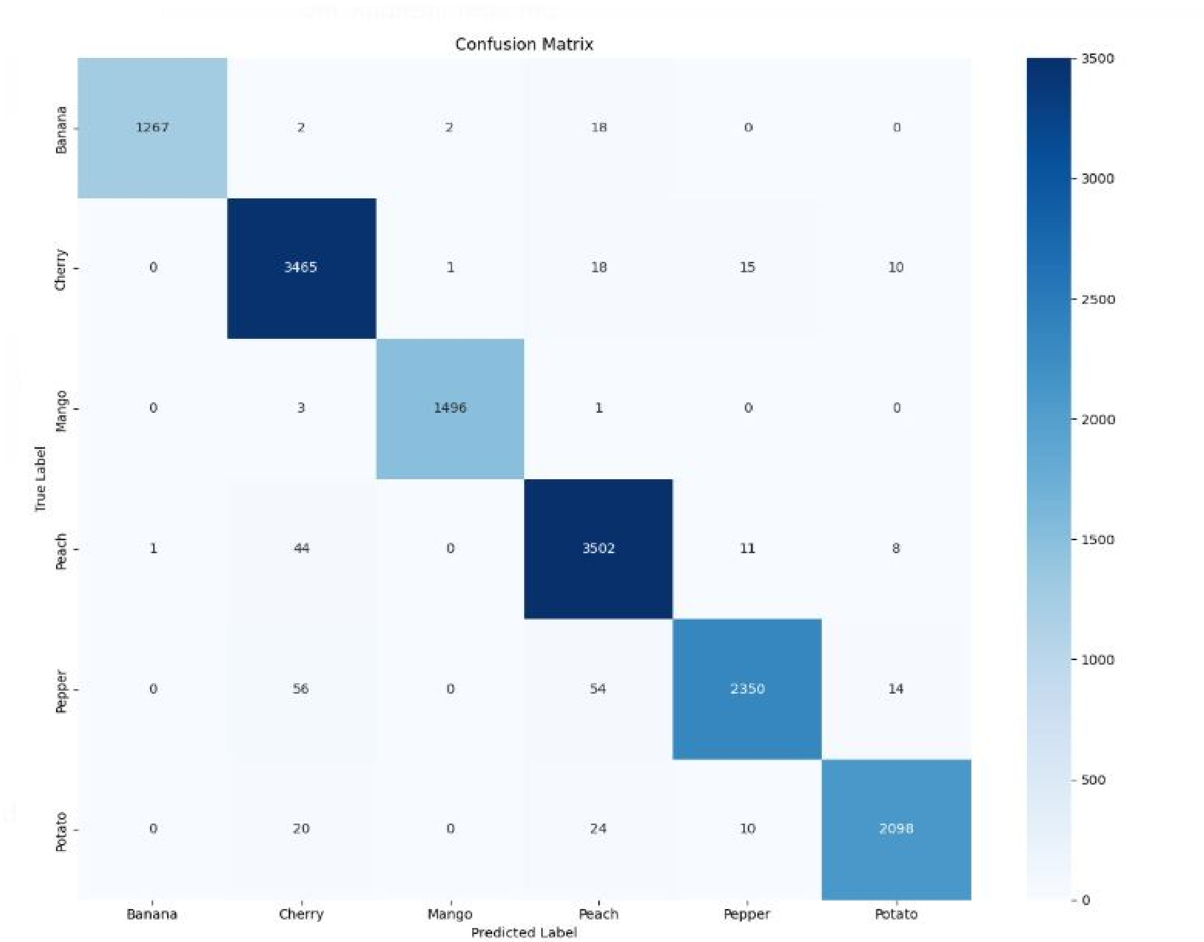


Fig 5.4 Confusion Matrix for Plant Disease Classification Model

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue III Mar 2025- Available at www.ijraset.com

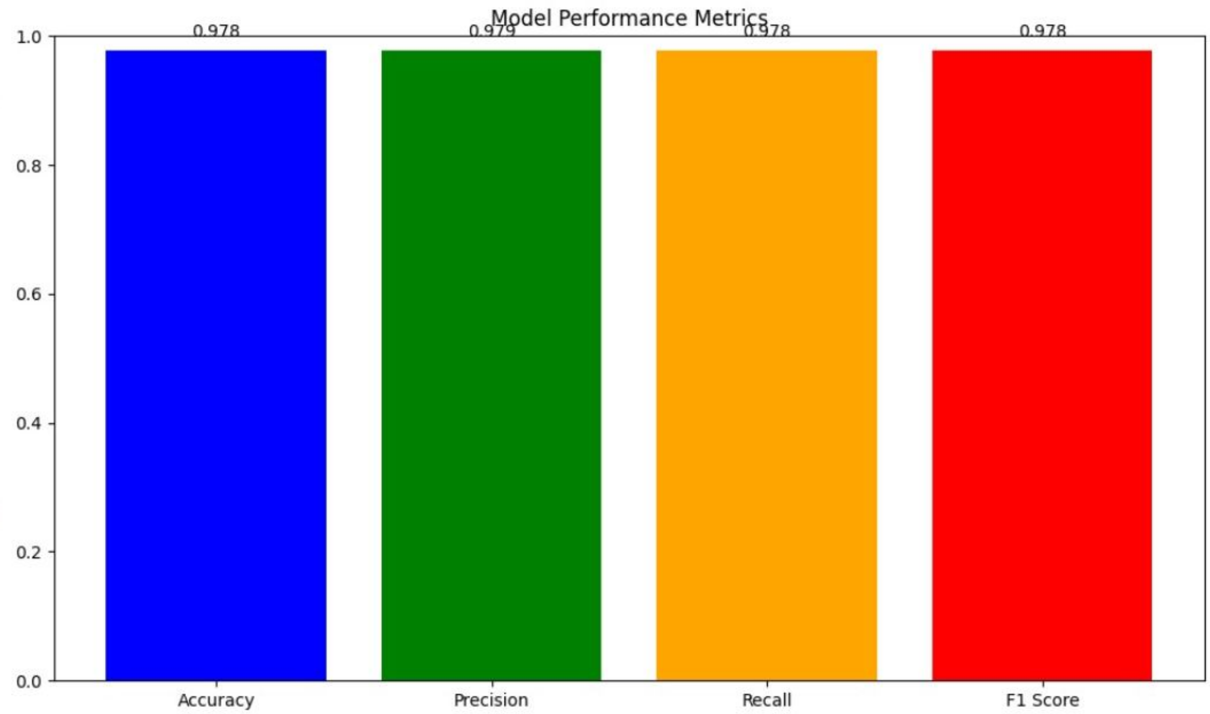


Fig 5.5 Performance Metrics of the Plant Disease Classification Mode

1) Fig 5.1: Plant Disease Detection System Application Interface

The user interface of a plant disease detection system is displayed in this picture. Users can upload and analyse photos of plant leaves for disease identification with ease thanks to the interface's intuitive and user-friendly design.

A Selected Picture:

- A chosen picture of a plant leaf is shown on the interface. In this instance, the picture displays a solitary, seemingly healthy mango leaf.
- By using the "Change Image" button, users can submit their own photos to replace the one that is now shown.
- Buttons for Analysis
- Below the picture are two primary buttons:

Analyse Plant: Pressing this button starts the analysis. After clicking, the system analyses the uploaded image using machine learning techniques to look for any possible illnesses.

- Change Image: If viewers like to examine a different leaf or plant, they can use this button to choose a different image.

Analysis Results: The interface shows the analysis results beneath the buttons:

- Plant: This field shows the kind of plant that the image's system has detected. The leaf in this instance has been accurately recognized by the system as being from a mango plant.
- Status: Based on the analysis, this field offers a rating of the plant's health. According to the system, this plant is "Healthy."
- Confidence: This field displays, as a percentage, the system's analysis's degree of confidence. The system has 99.97% confidence in determining that the mango leaf is healthy in this case. The model achieved an accuracy of 99%, which is higher than the 96.5% accuracy reported by Ansari Fatima Aness [2] using the VGG19 architecture.

2) Fig 5.2: Confidence Score Distribution Across Different Plant Classes in Disease Classification

The distribution of confidence scores for distinct classes—banana, peach, mango, cherry, potato, and pepper—is shown in this box plot solely for experimental reasons. With the majority of results grouped around the upper quartile, the plot demonstrates that confidence scores are often high across all classes. Interestingly, a number of outliers show that the model's confidence was lower than usual, especially in the Banana and Peach classes.

3) Fig 5.3: Distribution of Confidence Scores in Disease Classification Predictions

The overall distribution of confidence scores for the model's predictions is shown by this histogram. The distribution has a notable concentration of scores around 1.0 and is strongly skewed towards higher confidence levels. This implies that the model typically makes predictions with a high degree of confidence, which is further supported by the presence of a long tail of lower confidence scores.

4) Fig 5.4: Confusion Matrix for Plant Disease Classification Model

A thorough analysis of the model's performance in each of the six classes is given by the confusion matrix. While off-diagonal elements show misclassifications, diagonal elements show accurate predictions. The model does remarkably well for peaches and potatoes, according to the matrix, with most forecasts coming true. Nonetheless, there are observable misclassifications between related classes, including Cherry and Potato and Banana and Mango, indicating possible difficulties in differentiating between these groups.

5) Fig 5.5: Performance Metrics of the Plant Disease Classification Mode

The model's performance metrics, such as Accuracy, Precision, Recall, and F1 Score, are compiled in this bar chart. The values of all metrics, which range from 0.978 to 0.979, are quite high. This suggests that the model's predictions for every class are incredibly accurate and trustworthy. These metrics' consistency indicates that the model is correctly calibrated and consistently performs well across a range of prediction types.

VI. CONCLUSION

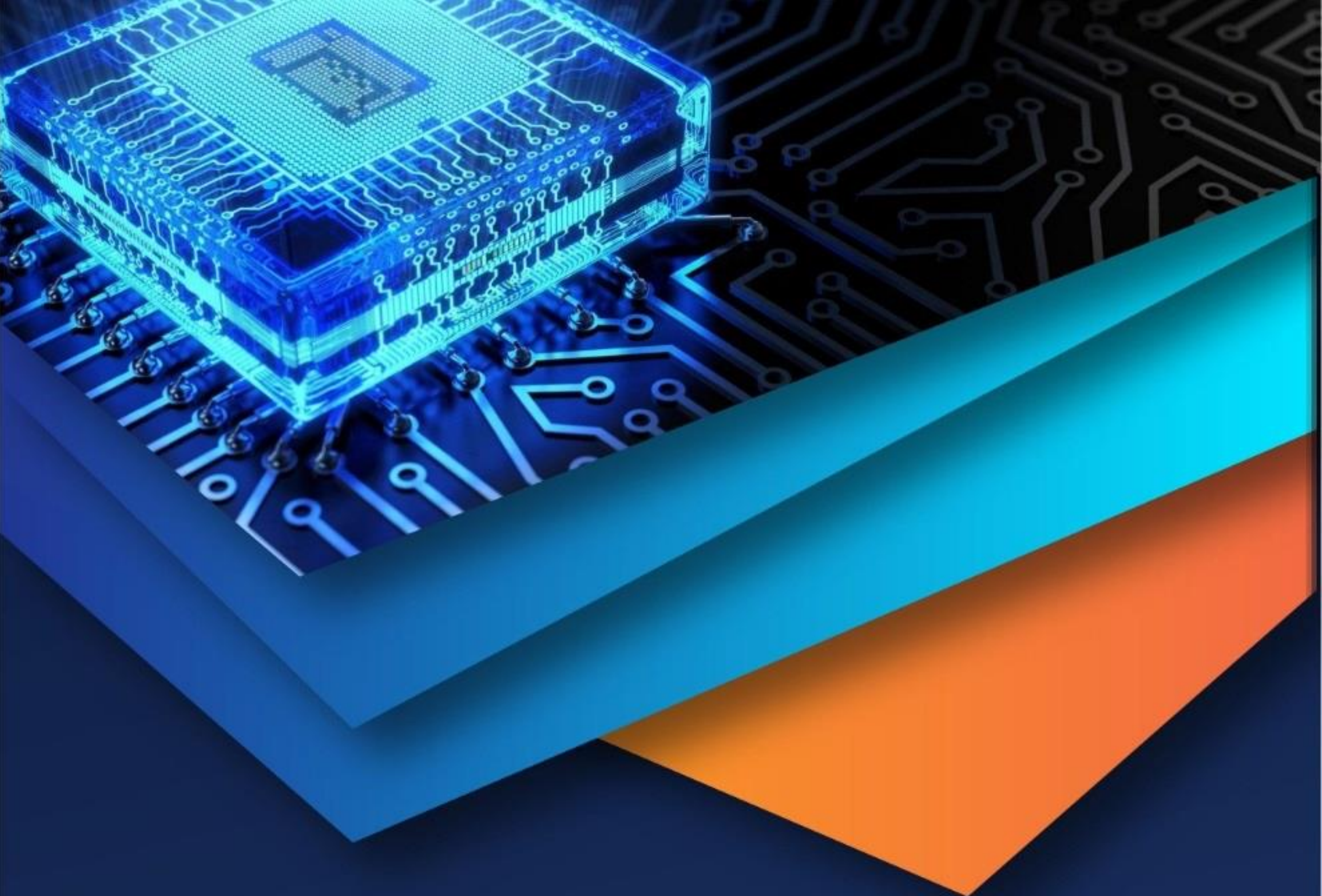
High accuracy, precision, recall, and F1 scores of 0.978 were attained by the study's deep learning-based model for automated plant disease detection and categorization. Compared to conventional machine learning techniques, the model provides a more advanced and precise approach by utilizing convolutional neural networks. Its wider ramifications include improving crop management techniques, offering an affordable substitute for manual inspections, and acting as a teaching tool. The model can adjust to bigger datasets thanks to its scalable architecture. Future plans call for growing the dataset, adding more data modalities, testing in the real world, enhancing the user interface, investigating transfer learning and model optimization strategies, and working with practitioners and researchers.

VII. ACKNOWLEDGEMENT

We would want to sincerely thank you to all those who helped this study to be successfully completed. First and most importantly, we sincerely thank you to Universal College of Engineering for giving the tools, equipment, required to complete this effort. Particularly thanks to Mr. Sandesh Patil, our project supervisor, for his great direction, perceptive criticism, and ongoing support over the study process. The trajectory of this endeavour was much shaped by his knowledge and mentoring. Our thanks also go especially to the Plant Village dataset developers and collaborators, whose work formed the basis for our tests. Their efforts in gathering and distributing such a thorough dataset have been absolutely vital in improving studies on plant disease identification. We also value the technical support, constructive comments, and technical assistance of the faculty members and colleagues in the Department of Data Engineering at Universal College of Engineering. We would also want to thank our friends and relatives for their support and understanding, which kept us going during demanding stages of our endeavour. At last, we appreciate the time and effort editors and reviewers have spent offering comments meant to raise the calibre of this work. Without the combined support of all these people and organizations, this study would not have been feasible.

REFERENCES

[1] E. Hirani, V. Magotra, J. Jain, “Plant Disease Detection Using Deep learning”, 2021 6th International Conference for Convergence in Technology.
[2] A. F. Aness; R. Fernandes; Bilal Shaikh; Kamran Shaikh; Nida Shaikh, “Leaf Disease Detection”, August 2022
[3] S. Kumar; KMVV Prasad; A. Shrilekha; T. Suman; B. Pranav Rao, “Leaf Disease Detection and Classification based on Machine Learning”, October 2020.
[4] M. Sardogan, A. Tuncer, Y. Ozen, “Plant Leaf Disease Detection and Classification based on CNN with LVQ Algorithm”, IEEE Xplore (2020 September).
[5] S. Ramesh, R. Hebbar, "Plant Disease Detection Using Machine Learning",2018 International Conference.
[6] J. A. Pandian, et al., “Plant Disease Detection Using Deep Convolutional Neural Network,” July 2022.
[7] A. Ahmad, D. Saraswat, and A. E. Gamal, “A survey on using deep learning techniques for plant disease diagnosis and recommendations for development of appropriate tools,” June 2022.
[8] S. Iniyan, et al., “Plant Disease Identification and Detection Using Support Vector Machines and Artificial Neural Networks,” 2020.
[9] M. S. Arya, K. Anjali, and D. Unni, "Detection of unhealthy plant leaves using image processing and genetic algorithm with Arduino," 2018.
[10] B. Liu, Y. Zhang, D. He, and Y. Li, “Identification of Apple Leaf Diseases Based on Deep Convolutional Neural Networks,” Dec. 2017.
[11] P. Sharma, Y. P. S. Berwal, and W. Ghai, "KrishiMitr (Farmer's Friend): Using Machine Learning to Identify Diseases in Plants," 2018.
[12] T. Tawde, et al., “Rice Plant Disease Detection and Classification Techniques: A Survey,” July 2021.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY