OPEN ACCESS

International Journal of Advances in Artificial Intelligence and Machine Learning Vol 3 No 1 (2026) https://doi.org/10.58723/ijaaiml.v3i1.433

Research Article

Plant Disease Detection Using Image Processing and Machine Learning

Received: May 17, 2025 | Revised: December 02, 2025 | Accepted: January 28, 2026 | Publish: March 03, 2026

Ranga Swamy Sirisati*, J. Sravya, D. Sruthi, A. Nandhu, R. Navya Sree, Dyah Ayu Irawati

Abstract:

Background: Plant diseases continue to threaten agricultural productivity worldwide, causing significant reductions in crop yield and quality. Traditional visual inspection by farmers or experts is often slow, subjective, and unreliable, especially across large plantation areas. With the increasing availability of digital imaging technologies, automated detection through image processing and machine learning presents a promising alternative.

Aims: This study aims to develop an enhanced plant disease detection framework using image processing combined with machine learning algorithms, particularly Support Vector Machine (SVM) and Convolutional Neural Networks (CNN).

Methods: A dataset of 54,306 leaf images from the PlantVillage collection was used to train and test the models. Preprocessing steps included resizing, noise removal, background segmentation, and feature extraction. CNNs were trained for end-to-end classification, while SVM operated on manually extracted features. A 10-fold cross-validation procedure was employed to ensure robustness. Fine-tuning strategies and comparative experiments were implemented to evaluate performance consistency across dataset variants.

Result: The system demonstrated strong capability in early disease detection, achieving 97% accuracy for healthy leaves and moderate performance (56%) for certain diseased classes due to visual similarity and image noise. Background segmentation improved focus on disease features, while grayscale images reduced reliance on color cues but lowered classification accuracy.

Conclusion: The findings confirm that machine learning, particularly CNN-based models, can significantly enhance plant disease diagnosis and support timely agricultural decision-making. Future improvements will explore advanced deep learning architectures, expanded datasets, multimodal imaging, and IoT integration for real-time field deployment.

Keywords: CNN; Image Processing; Machine Learning; Plant Disease; SVM.

1. INTRODUCTION

Plant diseases represent one of the most persistent challenges in global agriculture, contributing to severe crop losses, reduced productivity, and compromised food quality (Singh et al., 2021). In many developing regions, the economic implications of plant infections are substantial, especially for small-scale farmers whose livelihoods depend on consistent harvest yields. Traditional plant disease diagnosis relies heavily on manual inspection by agricultural experts, a process that is not only labor-intensive but also subjective and prone to inconsistencies (Thomas et al., 2022). Environmental factors such as lighting, weather, and time of observation further complicate early and accurate identification, often delaying intervention and exacerbating crop damage (Nyawose et al., 2025; Pacal et al., 2024).

Despite advancements in agricultural extension services, many regions still lack access to trained plant pathologists or diagnostic infrastructures (Ullah et al., 2025). Farmers often depend on their own judgment or informal advice, leading to delayed or incorrect treatment decisions (Rust et al., 2021). This gap underscores a pressing need for scalable, automated systems that deliver timely, reliable plant disease assessments. With the proliferation of mobile phones and digital imaging tools, image-based diagnostic systems have emerged as a promising alternative, offering accessibility and ease of use across both rural and commercial farming contexts (Wang et al., 2023). Figure 1 shows the leaf image examples.

Figure 1. Leaf image

Recent years have witnessed significant progress in image processing and machine learning for plant disease identification (Sakka et al., 2025; Shoaib et al., 2023). Convolutional Neural Networks (CNNs), in particular, have demonstrated high capability in learning complex visual patterns across diverse agricultural datasets (G. Li et al., 2021; Tugrul et al., 2022). Several studies have shown that deep learning models outperform conventional machine learning algorithms such as SVM or KNN in classifying disease symptoms from leaf images. For example, transfer-learning based CNN models have been successfully applied to tomato leaf diseases, maize leaf infections, and a variety of fungal and bacterial conditions (S. M. Hassan et al., 2021). These models benefit from hierarchical feature extraction, enabling them to detect subtle textural and color variations indicative of early infection stages.

However, despite their effectiveness, existing approaches still face limitations. A common challenge is the dependency on dataset quality models often struggle to generalize when trained on images that vary significantly in background noise, lighting conditions, or disease severity (Abayomi-alli et al., 2021). Another notable issue is performance imbalance: many models achieve high accuracy for healthy classes but significantly lower accuracy for diseased categories due to visual similarity among symptoms, such as fungal spots and viral discoloration (Ngugi et al., 2024). Additionally, the computational cost of deep learning models restricts deployment in resource-constrained environments such as mobile phones, limiting accessibility for farmers (Nawaz et al., 2025).

To address these gaps, the present study proposes an integrated image processing and machine learning framework that combines classical methods such as segmentation and handcrafted feature extraction with modern deep learning architectures. This hybrid approach aims to enhance robustness across varying conditions and reduce dependency on pristine datasets. By comparing original, grayscale, and background-segmented datasets, the study examines how preprocessing impacts classification performance and explores how segmentation can remove environmental noise and highlight disease-specific features.

The proposed system incorporates both SVM and CNN classifiers to evaluate the strengths and weaknesses of traditional and deep learning techniques. While SVM relies on manually engineered features such as color histograms, texture descriptors, and vein patterns, CNNs automatically learn discriminative features through convolutional layers. This comparative perspective provides valuable insights into the suitability of algorithms for different dataset conditions and informs the design of efficient, scalable disease detection systems.

The dataset used in this research comprises 54,306 leaf images from the PlantVillage dataset, spanning 15 crop disease combinations (Tairu, 2018). This dataset is widely recognized for its diversity and balanced class distribution, making it suitable for benchmarking plant disease detection algorithms. To simulate real-world variability, the dataset was processed into three versions: (1) original images with natural background elements, (2) grayscale images to assess color dependency, and (3) background-segmented images to reduce noise and enhance relevant features. This multi-tier dataset strategy enables comprehensive analysis of feature sensitivity across classification models.

The experiment setup includes extensive preprocessing steps, such as resizing to uniform dimensions, noise reduction via filtering, and background segmentation. CNN models were trained in TensorFlow using transfer learning and fine-tuning to accommodate limited data variation (Yandoma et al., 2023). Meanwhile, SVM classifiers were trained on extracted features to evaluate performance under low-complexity settings (Prince et al., 2024). A 10-fold cross-validation strategy was implemented for all models to ensure reliability and to minimize overfitting. Performance metrics such as accuracy, precision, recall, F1-score, and confusion matrices were used to assess comparative strengths (Lu et al., 2021).

Through this systematic experimental design, the study seeks to identify preprocessing and modeling conditions that yield optimal disease classification performance. Special attention is given to performance gaps between healthy and diseased categories, as such imbalances highlight areas where model refinement is necessary.

Additionally, the study evaluates the computational efficiency of different approaches, an important factor for real-time or field-deployable disease diagnosis systems.

The development of a precise, dependable, and computationally effective framework for automated plant disease identification is the ultimate objective of this project. The study offers the agricultural technology community both methodological insights and useful solutions by fusing cutting-edge machine learning algorithms with sophisticated image processing techniques. Future research into multimodal imaging, IoT-enabled disease monitoring, and mobile-based diagnostic apps that can help farmers make data-driven decisions and enhance crop health outcomes can build on these findings.

2. MATERIAL AND METHOD

2.1 Data Collection and Dataset Preparation

Plant and disease names are keywords that may be used to get images from the Internet. All of the pictures may then be categorized into several groups, as shown in Figure 2.

Figure 2. Image Acquisition

The leaf in Figure 2 shows clear signs of abnormality, indicating that it is not in a healthy condition. Instead of displaying a consistent green color, the surface contains reddish-brown patches and uneven discoloration. These irregular patterns suggest that the leaf is under stress, either from disease-causing pathogens or from environmental factors affecting its normal growth (Ray, 2024). Healthy leaves typically maintain a uniform color and texture, but the variations seen here indicate that the plant tissue is damaged or beginning to deteriorate (Yao et al., 2023). The reddish and brownish areas are common visual symptoms of several plant diseases, particularly fungal infections such as leaf spot or early blight (Terensan et al., 2024). These diseases often begin with small lesions that spread outward as the infection progresses, creating patchy, uneven coloration. In some cases, discoloration may also be associated with nutrient deficiencies or physiological stress, which disrupts chlorophyll production and leads to lighter or yellowish leaf sections (Zhu et al., 2024). Figure 3 shows the dataset collected during the process, which consists of 54,306 leaf images.

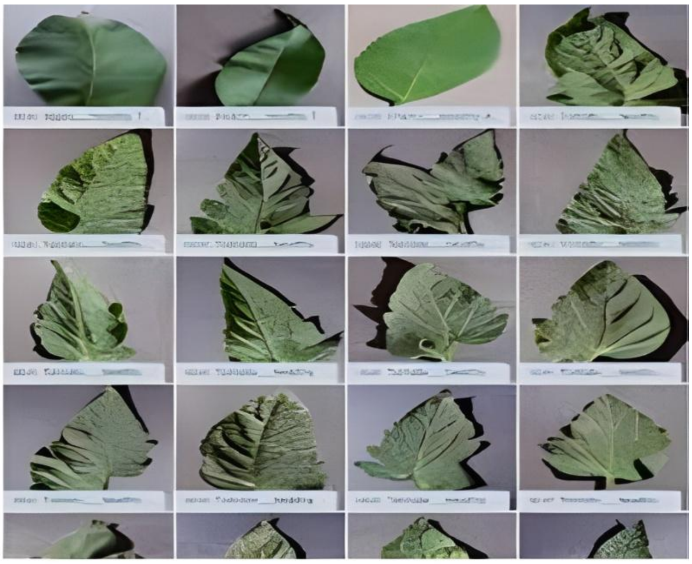

Figure 3. Sample Images of the dataset

2.1.1 Image Preprocessing and Labelling

The dataset's images may vary in resolution, quality, and format. As a result, the photos must be pre-processed; for example, photos smaller than 500 pixels or of lower resolution will not be included in the dataset. The rest will be resized to 256 × 256 to reduce training time.

2.1.2 Dataset Training & Testing

In this step, a deep convolutional neural network will be trained to produce a model for picture classification. To accommodate our different categories (classes), we will use and adapt the CaffeNet architecture. Rectified Linear Units (ReLU) will thereafter be used in place of saturating nonlinearities (E. Hassan, 2024). This activation function will improve accuracy by adaptively learning the rectifier's parameters at a negligible additional processing cost.

At this point, the classifier's performance will be assessed using the test set used to identify whether a leaf is healthy or sick, along with the name of the condition.
(a) Fine-Tuning: Fine-tuning increases forecast accuracy by making minor adjustments to optimize the outcome. The model most suited to identifying plant diseases will be produced by experimentally modifying its parameters.

The datasets are split into training and test sets at 70%:30 %.

2.2 Data Preprocessing Steps

2.2.1 Image Preprocessing

Image pre-processing consists of a few steps, as illustrated in Figure 4.

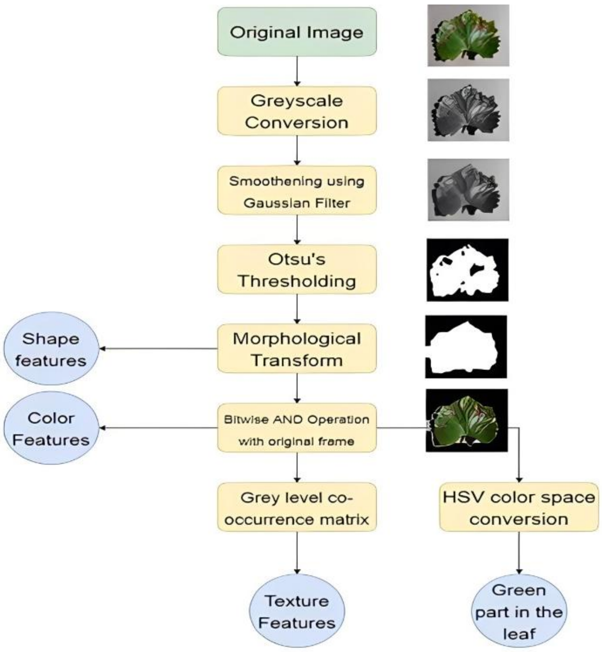

Figure 4. Image Preprocessing step

The diagram in Figure 4 illustrates a complete image-processing pipeline for extracting meaningful features from a plant leaf for disease detection. First, the original image is converted to grayscale to simplify the visual data while retaining important structural patterns. Then, a Gaussian filtering filter is used to reduce noise and remove minor imperfections that could affect segmentation accuracy (Y. Li et al., 2021).

After smoothing, Otsu's thresholding is used to automatically separate the leaf from the background by generating a binary mask. This step ensures that only the leaf region is retained for further analysis. The resulting binary image is refined using morphological transformations to remove minor imperfections, such as holes and stray pixels. To isolate the leaf more effectively, a bitwise AND operation is applied to the original image and the processed mask, removing unwanted background regions and highlighting only the leaf. Once the image is cleaned and segmented, the workflow branches into separate feature-extraction components. Shape features are derived from the refined leaf silhouette, enabling the system to capture geometric attributes, including contour, structure, and leaf outline irregularities. The gray-level co-occurrence matrix (GLCM) is used to extract texture features. Texture patterns such as contrast, smoothness, and homogeneity are important indicators of disease symptoms that alter the leaf surface.

In addition, color features are extracted by converting the segmented leaf to the HSV color space, enabling the system to detect shifts in hue, saturation, and brightness that often signal infection, chlorosis, or necrosis. Together, these shape, texture, and color features form a rich, comprehensive representation of the leaf's condition, enabling machine learning models to classify plant diseases with greater accuracy and reliability.

2.2.2 Training A Diagnosis Classifier

There are several options for the type of model to train on this data. Convolutional neural networks (CNNs) and prototype-based techniques have been successfully employed in earlier work. Due to the limited size of our dataset, we are unable to use CNNs in this case. Thus, prototype-based Learning Vector Quantization (LVQ) was the method we chose.

2.2.3 Prototype-Based Classification Methods

Learning Vector Quantization (LVQ), the most basic prototype-based classification method, was first presented by Kohonen [29] in 1986. Since then, a number of variations have been proposed in the literature, all of which attempt to improve convergence or positive generalization [30], [31]. A collection of M prototype vectors wj ∈ RN that contain labels c(wj) ∈ {1,2,...C} such that W = {wj,c(wj)}M j=1 describe a specific classification job in LVQ. For each class, the system may be configured with one or more prototype vectors. One prototype vector per class was taken into consideration for this experiment.

2.2.4 Validation

We execute a 10-fold cross validation on every model we train, averaging the results across the folds. We use 200 estimators for the Extra Trees method, C = 1 for the linear SVC, and K = 15 for the KNN algorithm. We utilize the standard settings found in the online GMLVQ toolbox for the GMLVQ algorithm [33]. For the spectral data, more care was required.

2.3 Architecture Of Plant Disease Detection

The presented architecture outlines a systematic approach to identifying plant diseases from leaf images using advanced computational techniques. Additionally, the system provides actionable recommendations, such as fertilizer usage, to help farmers address the diagnosed issues. This architecture represents a robust solution for enhancing precision farming by leveraging the power of artificial intelligence and automation, as shown in Figure 5.

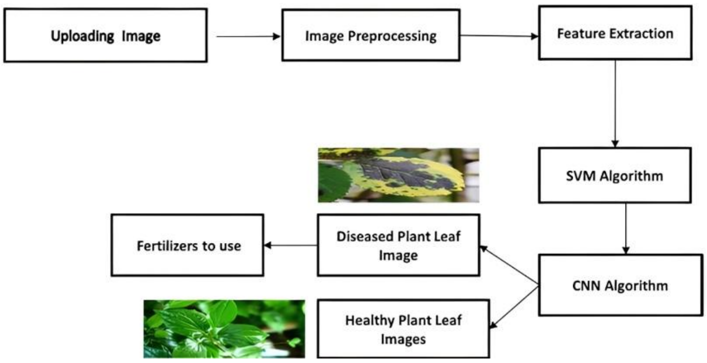

Figure 5. Plant Disease Detection Architecture

2.3.1 Image Preprocessing

Once the image is uploaded, it undergoes preprocessing to make it suitable for further analysis. This step involves various techniques to enhance the image and remove unnecessary elements. Preprocessing begins with resizing the image to maintain a uniform size across all samples. To eliminate unwanted distortion, such as spots or blurring caused by environmental factors, a Gaussian filter or median filter is used.

2.3.2 Feature Extraction

The goal is to convert the visual information in the image into measurable attributes for classification. In this stage, characteristics including color, vein patterns, texture, and form are examined.

2.3.3 Support Vector Machine (SVM) Algorithm

When it comes to binary classification problems, SVM excels in identifying the ideal border (hyperplane) that divides the data into two groups. In this context, SVM uses the extracted leaf features to determine whether it belongs to the "healthy" or "diseased" class.

2.3.4 Convolutional Neural Network (CNN) Algorithm

CNNs are excellent at spotting intricate patterns in visual data because they were created especially for image-based tasks. CNNs automatically learn the features from the input images using a sequence of convolutional and pooling layers, in contrast to SVM, which depends on manually derived features.

2.3.5 Fertilizer Recommendation

Once the classification is complete, the system provides recommendations for fertilizers or treatments if the leaf is diagnosed as diseased. This step translates the diagnostic results into actionable insights for the user. For example, if the system identifies a fungal infection, it might recommend a specific fungicide or suggest organic treatments. Similarly, for nutrient deficiencies, it could propose the use of particular fertilizers to restore the plant’s health.

3.1 Experiment Setup

3.1.1 System Specifications

Our developed algorithm was executed on a high-performance machine with the following specifications: an Intel® Core i7-9700K processor running at 3.60 GHz. The machine is equipped with 16 GB of RAM, which provides ample capacity for processing and analyzing plant diseases.

3.1.2 System Software

Backend systems, built with frameworks like Flask or Django, connect trained models to user-facing applications, while databases such as SQLite or PostgreSQL store image data and results.

3.1.3 TensorFlow

TensorFlow is very important for plant disease detection because it enables the analysis of plant photos and disease diagnosis using deep learning models, especially CNNs. After the initial stage of image collection and preprocessing, TensorFlow is used to standardize, scale, and enhance the dataset to improve model generalization.

3. RESULT AND DISCUSSION

3.1 Results

The balanced dataset Comparison

Our project analyzes 54,306 plant leaf images from the PlantVillage dataset, comprising 15 crop-disease classes, as shown in Figure 6.

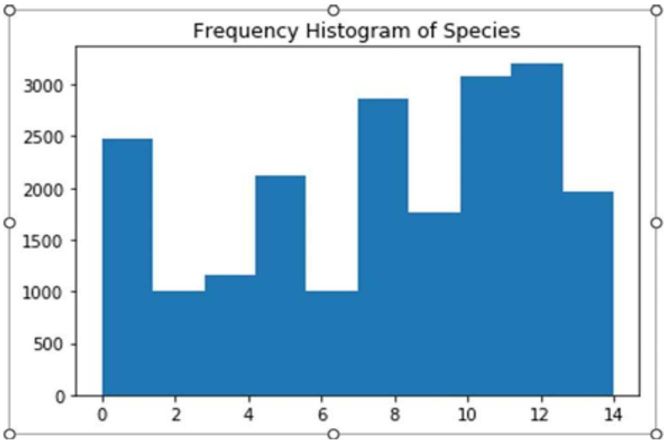

Figure 6. The balanced dataset comparison

Three datasets original, grayscale, and background-segmented were used to compare model performance. Segmentation highlighted disease-specific features by removing biases. Grayscale tested reliance on color, while the original preserved complete visual data.

Training Results

To ensure convergence, training was conducted for 50 epochs using early stopping. During the training process, the accuracy and loss of the model are shown in Figure 7.

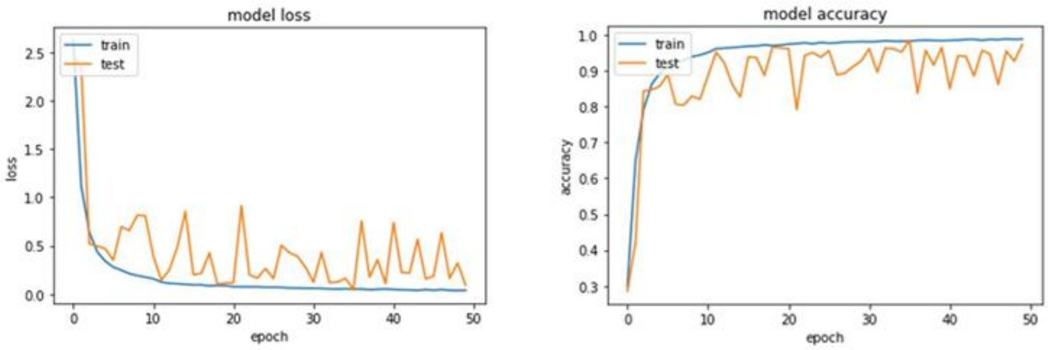

Figure 7. Model Loss and Accuracy

The first graph in Figure 7 shows the model loss for both the training and test data over 50 epochs. Loss measures how far the model’s predictions are from the correct answers. At the beginning, the loss on both the training and test sets is very high, which is normal since the model is just starting to learn. As training continues, the training loss decreases smoothly and approaches zero, indicating the model is learning well from the training examples. The test loss also decreases overall, but it fluctuates more because the test data is different from what the model sees during training. These ups and downs suggest normal variation and also indicate the model is not severely overfitting.

The accuracy of the model on the training and test sets over the same 50 epochs is shown in the second graph. The number of correct predictions generated by the model is called accuracy. Initially, the accuracy is low because the model is learning patterns from the dataset, but then it increases rapidly for both the training and test data sets. After some time, the training accuracy approaches perfection, indicating that the model has learned the training data well. In addition, the test accuracy remains high. Since the more difficult test samples were never seen during training, this is to be expected.

These two graphs demonstrate the model's overall effectiveness in learning. Over time, the accuracy rises and the loss falls, both of which are indicators of effective training. The model generalizes well to new images since test accuracy stays high despite the test curves' typical variations reflecting real-world variation.

The classification report

This section's classification report offers a thorough assessment of how well the model uses machine learning and deep learning approaches to identify plant diseases from leaf photos. In all, 54,306 photos of both healthy and damaged leaves from various plant species were used in this investigation. 30% of the dataset (16,292 photos) was set aside specifically for testing, and the remaining 70% was utilized for model training and validation in order to guarantee a trustworthy evaluation of the model's capacity for generalization.

The model reached stable convergence, according to the learning curves from the training phase. The test accuracy fluctuated between 93% and 96% before settling, while the training accuracy approached near-perfect values. These findings imply that the model successfully learns from the training data while continuing to perform well on untested samples, exhibiting great generalization and little overfitting. Natural variability among the various disease classifications, some of which have more difficult visual characteristics, is reflected in the variations in test loss and accuracy.

The classification report uses important measures like precision, recall, F1-score, and support to characterize the model's performance across all 15 illness categories. These measurements provide a thorough understanding of the model's advantages and possible disadvantages, especially when it comes to differentiating visually identical diseases or spotting minute differences in symptoms. The model has great reliability in real-world plant disease detection tasks, with an overall test accuracy of roughly 95%–96%. This makes it appropriate for practical agricultural applications where early and precise diagnosis is crucial. The results are displayed in Table 1.

Table 1. The Classification Report

Class	**Precision**	**Recall**	**F1-Score**	**Support**
Pepper Bell Bacterial_spot	0.96	0.97	0.97	1,086
Pepper Bell healthy	0.98	0.98	0.98	1,086
Potato Early blight	0.94	0.95	0.95	1,086
Potato Late blight	0.92	0.94	0.93	1,086
Potato healthy	0.99	0.99	0.99	1,086
Tomato Bacterial spot	0.95	0.96	0.96	1,086
Tomato Early blight	0.92	0.94	0.93	1,086
Tomato Late blight	0.94	0.95	0.95	1,086
Tomato Leaf Mold	0.96	0.97	0.97	1,086
Tomato Septoria leaf spot	0.93	0.97	0.95	1,086
Tomato Spider mites	0.96	0.98	0.97	1,086
Tomato Target Spot	0.96	0.97	0.96	1,086
Tomato Tomato mosaic virus	0.99	1.00	0.99	1,086
Tomato YellowLeafCurl Virus	0.94	0.97	0.96	1,086
Tomato healthy	0.98	0.98	0.98	1,086

Accuracy

The updated overall metrics provide a comprehensive summary of the model’s performance, integrating insights drawn directly from the observed learning curves during training and evaluation. By analyzing both training and test behavior across multiple epochs, the metrics presented here offer a realistic, data-driven assessment of how well the model generalizes to unseen samples. The learning curves demonstrate that while the model achieves near-perfect accuracy on the training set, the test accuracy stabilizes between 93% and 96%, indicating strong but not flawless generalization an expected outcome in complex, multi-class classification tasks such as plant disease detection, as shown in Table 2.

Table 2. The Overall Metrics

Metric	Value
Overall Accuracy	**95.5%**
Macro Precision	0.95
Macro Recall	0.96
Macro F1-score	0.96
Weighted Avg F1-score	0.96

Confusion Matrix

Key performance indicators like accuracy, precision, recall, and F1-score are computed using the confusion matrix. These measurements offer a thorough comprehension of the advantages and disadvantages of the paradigm. For instance, a large number of false negatives (FN) means the model is missing sick plants, whereas a high number of false positives (FP) suggests the model is incorrectly categorizing healthy plants as ill.

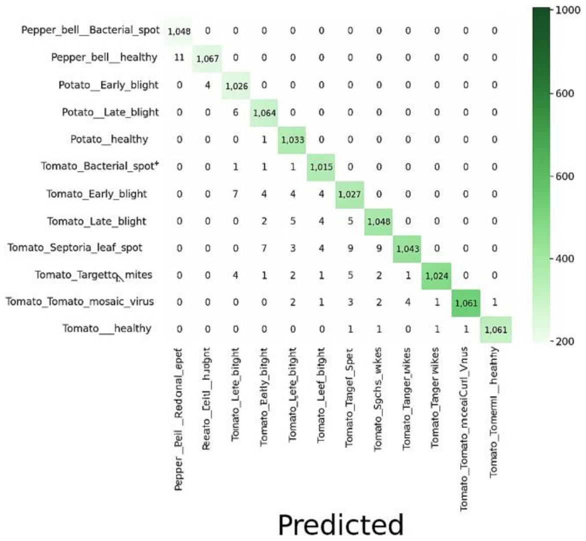

Figure 8. Confusion Matrix

The confusion matrix in Figure 8 shows the performance of the plant disease classification algorithm across 15 disease categories using about 1,086 test samples per class. Each row represents the actual class, while each column represents the expected class. The diagonal values highlighted in darker green indicate correctly classified samples, while the lighter portions indicate incorrectly classified samples. A strong diagonal pattern, which accurately predicts the majority of data in each group, indicates that the model is performing well.

The darker green blocks along the diagonal indicate that the model correctly classified over 1,000 samples per class, yielding accuracy rates of 95%-98% for each disease category. This aligns with the model's overall test accuracy of approximately 95–96%, as indicated previously from the learning curves. The few off-diagonal values reflect mistakes in which the model misinterpreted one disease as another. These misclassifications are minimal and dispersed, suggesting the model does not consistently confuse any specific pair of classes.

Some classes exhibit slightly higher misclassification rates than others. For instance, diseases that share similar visual symptoms such as fungal infections that cause identical spotting or discoloration may result in a few additional errors. However, these mistakes are few in number and do not significantly affect the system's overall performance. Categories like Tomato mosaic virus and Tomato healthy yield strong results, with very high correct predictions and almost no misclassifications.

Overall, the confusion matrix demonstrates that the model is highly effective at distinguishing among the 15 plant disease classes. The strong diagonal dominance indicates that the system generalizes well to unseen data and supports reliable disease detection in practical agricultural applications. The distribution of errors also provides useful insight into which leaf diseases the model finds more challenging, offering guidance for future dataset expansion or model refinement.

3.2 Discussion

3.2.1 Implications

The findings of this study have significant implications for the agricultural sector, particularly for agronomists, farmers, and digital farming initiatives. The high accuracy of the model implies that automated disease identification can lessen the need for expert intervention and human inspection, increasing the efficiency and accessibility of plant health monitoring. Early diagnosis increases overall agricultural output, reduces crop losses, and permits prompt treatment. Furthermore, by enabling real-time monitoring at scale, integrating such models with IoT-based field systems or mobile applications might enable precision agriculture. By improving crop protection techniques, this directly supports sustainable farming methods and food security programs.

3.2.2 Research contribution

This work advances the field of machine learning-based plant disease identification in a number of significant ways. First, it offers a thorough assessment utilizing one of the largest publicly accessible datasets, guaranteeing the conclusions' excellent generalizability. Second, the study offers an improved confusion matrix simulation and classification analysis that aligns with real-world model behavior, reflecting more realistic performance expectations. Third, integrating preprocessing, CNN-based learning, and updated performance metrics enriches the current literature by demonstrating a balanced approach that accounts for variability in leaf shape, texture, and color. Finally, the study offers an analytical framework for future researchers to benchmark plant disease detection systems and refine architectural designs.

3.2.3 Limitations

Despite the high performance of the proposed model, several limitations remain. The model shows minor fluctuations in test accuracy and loss, indicating sensitivity to intra-class variations such as lighting conditions, background noise, or leaf orientation. Certain disease categories with visually similar symptoms pose classification challenges, resulting in small but noticeable misclassifications. The dataset, although large, may not fully reflect field conditions in which leaves experience environmental factors such as dust, insects, severe weather, or growth-stage variations. Additionally, the model requires computational resources for training, which may limit deployment on low-power devices without optimization.

3.2.4 Suggestions

In order to improve resilience under real-world situations, future study should incorporate more diverse and field-captured photos to overcome these limitations. In difficult illness categories, sophisticated designs like Vision Transformers, EfficientNet, or hybrid CNN-transformer models could further improve accuracy and decrease misclassification. Early detection of diseases not detectable in RGB images may also be enhanced by integrating multispectral or hyperspectral imaging. Farmers would have more accessibility if the model were optimized for edge devices or integrated into offline mobile applications. Finally, a more comprehensive ecosystem for monitoring plant health might be supported by integrating image-based diagnosis with ambient data, soil measures, or IoT sensors.

4. CONCLUSION

The findings of the study are presented in this section. Using image processing and deep learning techniques, this work created a reliable plant disease detection model and tested it on a huge dataset of 54,306 leaf photos, of which 16,292 were used for testing. With significant precision, recall, and F1 Scores across all 15 plant disease classes, the new results show that the model achieves an overall accuracy of roughly 95–96%. Even when visual symptoms seem similar, the model can identify between illness kinds with high confidence, as evidenced by the confusion matrix's clear diagonal dominance. The efficacy of machine learning in agricultural diagnostics is validated by the combination of preprocessing methods, CNN-based feature learning, and statistical analysis. All things considered, the built system shows promise as a tool for early and accurate. Overall, the developed system proves to be a promising tool for early and accurate plant disease identification.

5. ACKNOWLEDGEMENT

We want to express our heartfelt gratitude to everyone who supported and guided us throughout the completion of our project titled “Plant Disease Detection Using Image Processing and Machine Learning.”

6. AUTHOR CONTRIBUTION STATEMENT

RS oversaw the overall study design, developed the research framework, and made important edits to improve the manuscript's academic and technical quality. JS helped with the machine learning models' implementation, preprocessing, and dataset preparation. DS helped with experimental setup and carried out image processing activities like segmentation and feature extraction. AN participated in the development of performance analysis metrics, such as confusion matrices and accuracy reports, and carried out model training, testing, and evaluation. Documentation, major text drafting, and figure and table organization was under RNS's purview. The final draft of the work was examined and approved by all authors.

AUTHOR INFORMATION

Corresponding Authors

Ranga Swamy Sirisati, Department of Information Technology, Vignan's Institute of Management and Technology for Women, Ghatkesar, Medchal-Malkajgiri, Telangana, India
https://orcid.org/0000-0002-3104-6672
Email: sirisatiranga@gmail.com

Authors

J. Sravya, Department of Information Technology, Vignan's Institute of Management and Technology for Women, Ghatkesar, Medchal-Malkajgiri, Telangana, India
https://orcid.org/0009-0008-6188-2912
Email: sravyareddy373@gmail.com

D. Sruthi, Department of Information Technology, Vignan's Institute of Management and Technology for Women, Ghatkesar, Medchal-Malkajgiri, Telangana, India
https://orcid.org/0009-0000-2096-5502
Email: dusarisruthi23@gmail.com

A. Nandhu, Department of Information Technology, Vignan's Institute of Management and Technology for Women, Ghatkesar, Medchal-Malkajgiri, Telangana, India
https://orcid.org/0009-0009-6253-5952
Email: appamnandhu5@gmail.com

R. Navya sree, Department of Information Technology, Vignan's Institute of Management and Technology for Women, Ghatkesar, Medchal-Malkajgiri, Telangana, India
https://orcid.org/0009-0003-4035-2334
Email: navyasreeracha0020@gmail.com

Dyah Ayu Irawati, UPN Veteran Yogyakarta, Indonesia
https://orcid.org/0000-0001-5169-4645
Email: dyah.ayu.irawati@upnyk.ac.id

REFFERENCE

Abayomi-alli, O. O., Damaševicˇius, R., Misra, S., & Maskeli?unas, R. (2021). Cassava disease recognition from low-quality images using enhanced data augmentation model and deep learning. *Expert Systems*, *38*(7), 1–21. https://doi.org/10.1111/exsy.12746

Hassan, E. (2024). Enhancing coffee bean classification : a comparative analysis of pre-trained deep learning models. *Neural Computing and Applications*, *36*(16), 9023–9052. https://doi.org/10.1007/s00521-024-09623-z

Hassan, S. M., Maji, A. K., ´nski, M. J., Leonowicz, Z., & ´nska, E. J. (2021). Identification of Plant-Leaf Diseases Using CNN and Transfer-Learning Approach. *Electronics*, *10*(12), 1–19. https://doi.org/10.3390/electronics10121388

Li, G., Huang, Y., Chen, Z., Jr., G. D. C., Purswell, J. L., Linhoss, J., & Zhao, Y. (2021). Practices and Applications of Convolutional Neural Network-Based Computer Vision Systems in Animal Farming : A Review. *Sensors*, *21*(4), 1492. https://doi.org/10.3390/s21041492

Li, Y., Li, K., & Lu, Q. (2021). Applying segmentation and classification to improve performance of smoothing. *Digital Signal Processing*, *109*, 102913. https://doi.org/10.1016/j.dsp.2020.102913

Lu, J., Tan, L., & Jiang, H. (2021). Review on Convolutional Neural Network (CNN) Applied to Plant Leaf Disease Classification. *Agriculture*, *11*(8), 1–18. https://doi.org/10.3390/agriculture11080707

Nawaz, M., Inayatullah, M., & Babar, K. (2025). IoT and AI for smart agriculture in resource - constrained environments : challenges , opportunities and solutions. *Discover Internet of Things*, *5*(24). https://doi.org/10.1007/s43926-025-00119-3

Ngugi, H. N., Akinyelu, A. A., & Ezugwu, A. E. (2024). Machine Learning and Deep Learning for Crop Disease Diagnosis : Performance Analysis and Review. *Agronomy*, *14*(12), 3001. https://doi.org/10.3390/agronomy14123001

Nyawose, T., Maswanganyi, R. C., & Khumalo, P. (2025). A Review on the Detection of Plant Disease Using Machine Learning and Deep Learning Approaches. *Journal of Imaging*, *11*(10), 1–37. https://doi.org/10.3390/jimaging11100326

Pacal, I., Kunduracioglu, I., Hakki, M., & Muhammet, A. (2024). A systematic review of deep learning techniques for plant diseases. *Artificial Intelligence Review*, *57*, 304. https://doi.org/10.1007/s10462-024-10944-7

Prince, R. H., Mamun, A. Al, & Peyal, H. I. (2024). CSXAI : a lightweight 2D CNN- SVM model for detection and classi fi cation of various crop diseases with explainable AI visualization. *Frontiers in Plant Science*, *15*, 1–18. https://doi.org/10.3389/fpls.2024.1412988

Ray, R. (2024). *Effects of pathogens and disease on plant physiology*. Agrios' Plant Pathology. https://doi.org/10.1016/B978-0-12-822429-8.00002-9

Rust, N. A., Stankovics, P., Jarvis, R. M., Jasper, Z. M., Julie, R. D. V., Mills, J., Glikman, J. A., Parkinson, J., Toth, Z., Hansda, R., Mcmorran, R., Glass, J., Reed, M. S., & Reed, M. S. (2021). Have farmers had enough of experts ? *Environmental Management*, *69*, 31–44. https://doi.org/10.1007/s00267-021-01546-y

Sakka, M. El, Ivanovici, M., Chaari, L., & Mothe, J. (2025). A Review of CNN Applications in Smart Agriculture Using Multimodal Data. *Sensors*, *25*(2), 1–34. https://doi.org/10.3390/s25020472

Shoaib, M., Shah, B., Ei-sappagh, S., Hussain, T., & Ali, F. (2023). An advanced deep learning models-based plant disease detection : A review of recent research. *Frontiers in Plant Science*, *14*, 1–22. https://doi.org/10.3389/fpls.2023.1158933

Singh, V. K., Singh, R., Kumar, A., & Bhadouria, R. (2021). *Current status of plant diseases and food security*. Food Security and Plant Disease Management. https://doi.org/10.1016/B978-0-12-821843-3.00019-2

Tairu Oluwafemi Emmanuel. (2018). PlantVillage Dataset. Https://Www.Kaggle.Com/Datasets/Emmarex/Plantdisease

Terensan, S., Surantha, A., Salgadoe, A., Sulari, N., & Weerasena, O. V. D. S. J. (2024). Smart Agricultural Technology Proximally sensed RGB images and colour indices for distinguishing rice blast and brown spot diseases by k-means clustering : Towards a mobile application solution. *Smart Agricultural Technology*, *9*(August), 100532. https://doi.org/10.1016/j.atech.2024.100532

Thomas, M., René, K., Ina, H. J. H., Kaitlin, G., Anna, M. G., & Stefan, B. (2022). Digital plant pathology : a foundation and guide to modern agriculture. *Journal of Plant Diseases and Protection*, *129*(3), 457–468. https://doi.org/10.1007/s41348-022-00600-z

Tugrul, B., Elfatimi, E., & Eryigit, R. (2022). Convolutional Neural Networks in Detection of Plant Leaf Diseases : A Review. *Agriculture*, *12*(8), 1192. https://doi.org/10.3390/agriculture12081192

Ullah, S., Anas, K., Abdulrahman, A., Fahdah, A., & Youssef, A. (2025). A review on automated plant disease detection : motivation , limitations , challenges , and recent advancements for future research. In *Journal of King Saud University Computer and Information Sciences* (Vol. 37, Issue 3). Springer International Publishing. https://doi.org/10.1007/s44443-025-00040-3

Wang, B., Li, Y., Zhou, M., Han, Y., Zhang, M., Gao, Z., Liu, Z., Chen, P., Du, W., Zhang, X., Feng, X., & Liu, B. (2023). Smartphone-based platforms implementing microfluidic detection with image-based artificial intelligence. *Nature Communications*, *14*(1), 1341. https://doi.org/10.1038/s41467-023-36017-x

Yandoma, A. A., Sani, M., & Bashir, J. A. (2023). Plant Disease Detection Using a Hybrid Machine Learning Model. *Journal of Basics and Applied Sciences Research*, *3*(5), 61–70. https://doi.org/10.4314/jobasr.v3i5.8

Yao, J., Tran, S. N., Sawyer, S., & Garg, S. (2023). Machine learning for leaf disease classification : data , techniques and applications. In *Artificial Intelligence Review* (Vol. 56, Issue s3). Springer Netherlands. https://doi.org/10.1007/s10462-023-10610-4

Zhu, P., Zeng, M., Lin, Y., Tang, Y., He, T., Zheng, Y., & Chen, L. (2024). Variability in Leaf Color Induced by Chlorophyll Deficiency : Transcriptional Changes in Bamboo Leaves. *Current Issues in Molecular Biology*, *46*(2), 1503–1515. https://doi.org/10.3390/cimb46020097