



Automated Detection and Classification of Vegetable Leaf Disease using Machine Learning Techniques

Shilpa T S ^{1,*}, Kumar Siddamallappa U ¹ and Anusha Jajur J ²

¹ Department of Studies in Computer Applications (MCA), Davangere University, Davangere, Karnataka, India.

² Department of Studies in Computer Science, Davangere University, Davangere, Karnataka, India.

Kumar Siddamallappa U; ORCID: 0000-0002-1975-3868

Anusha Jajur J; ORCID: 0009-0003-9835-624X

World Journal of Advanced Engineering Technology and Sciences, 2026, 20(01), 251–260

Publication history: Received on 14 June 2026; revised on 20 July 2026; accepted on 22 July 2026

Article DOI: <https://doi.org/10.30574/wjaets.2026.20.1.0373>

Abstract

Plant leaf diseases are a major concern in agriculture because they reduce crop quality, lower productivity, and cause economic losses to farmers. Early detection of these diseases is important for protecting plants and improving overall crop management. However, manual identification through visual inspection is often slow and may produce inaccurate results, especially in the early stages of infection. To overcome this problem, this work proposes an automated leaf disease detection system using image processing and deep learning techniques. OpenCV is used for preprocessing the leaf images through resizing, noise removal, color normalization, and enhancement, while a Convolutional Neural Network (CNN) is trained to extract important features and classify the images into different disease categories. The dataset consists of healthy and diseased leaf images collected from publicly available sources and is divided into training, validation, and testing sets for effective model development. For practical implementation, the trained model is integrated into a Flask-based web application that allows users to upload leaf images and obtain predictions easily. The performance of the system is evaluated using accuracy, precision, recall, and F1-score, and the model achieved an accuracy of 95%. The results show that the proposed system can detect leaf diseases effectively and efficiently, making it a useful and cost-effective solution for supporting farmers and agricultural experts in early disease diagnosis.

Keywords: Leaf Disease Detection; OpenCV; Convolutional Neural Network; Image Processing; Deep Learning; Flask.

1. Introduction

Agriculture is a vital activity that ensures food security and sustainable economic growth, and vegetable crops are among the most demanded agricultural products in terms of nutritional and economic value. Nevertheless, plants are subject to a variety of fungal, bacterial, and viral infections that affect their yield and quality. Some of the most harmful diseases reported include Downy Mildew, Powdery Mildew, Fusarium Wilt, Mosaic Virus, and Bacterial Leaf Spot [6]. Therefore, farmers and agricultural experts need to ensure that crop leaves are frequently inspected for pathogens and receive proper treatment once detected. However, the conventional manual inspection approach is costly, tedious, laborious, and error-prone. The rise of ML-driven computer vision technology has led to the development of automated image recognition and identification solutions for the early detection of plant diseases.

Specifically, Convolutional Neural Networks (CNN) have proven effective in image processing tasks such as pattern recognition, identification, sorting, and classification [7]. Preprocessing techniques also play a critical role in computer vision applications by enhancing image quality, reducing noise, and improving model performance during training and prediction. This study proposes an automated leaf disease detection system based on ML algorithms for vegetable crops,

* Corresponding author: Shilpa. T.S

including Bitter Gourd, Cucumber, and Pumpkin. Notably, the system will be trained on leaf images with healthy and infected samples of the aforementioned crops' categories.

The proposed methodology will be divided into several stages, namely, preprocessing, partitioning of the dataset into train/validation/test sets, extracting features using CNN, and classifying leaf images into their corresponding categories. Moreover, the model's performance will be evaluated using classification performance metrics. The proposed solution is expected to achieve high accuracy in detecting and classifying diseases affecting Bitter Gourd, Cucumber, and Pumpkin leaves. The developed system will offer a reliable, affordable, and accurate solution for plant disease detection and classification. the proposed methodology can be modified to detect crop diseases in other vegetables and utilized in mobile or IoT-enabled applications to offer automated real-time disease monitoring and control.

1.1. Problem Statement

Plant diseases are a serious problem in agriculture because they reduce crop quality, lower yield, and affect food security. Early detection is important to stop diseases from spreading and damaging healthy plants. Manual identification through visual inspection can be slow, inconsistent, and dependent on expert knowledge. Many diseases also show similar symptoms, which makes accurate diagnosis difficult. To solve this issue, an automated system is needed for fast and reliable detection. The proposed approach uses image preprocessing to improve leaf image quality before analysis. A Convolutional Neural Network is then used to learn important disease patterns from the images. This helps in accurately identifying and classifying plant leaf diseases. The model is deployed in a Flask-based web application for easy user access and real-time prediction. Overall, the system provides an efficient, practical, and accurate solution for early leaf disease detection.

1.2. Research Objectives

The main objective of this research is to develop an intelligent and user-friendly system for automatic leaf disease detection that can support farmers and agricultural professionals in early diagnosis and timely decision-making.

- To develop an automated leaf disease detection system using OpenCV and CNN that can identify plant diseases from digital leaf images.
- To train, deploy, and evaluate the model through a Flask web application using performance metrics such as Accuracy, Precision, Recall, and F1-Score.

2. Literature Review

Sujatha et al. [1] proposed an innovative approach based on the concepts of machine learning and deep learning to classify plant leaf diseases. The article highlights several research works and concludes that the algorithm of deep learning is more efficient than other algorithms because it detects features automatically extracted from leaf images. The authors highlight the importance of image pre-processing, data augmentation, and transfer learning to enhance classification efficiency.

Liu et al. [2] designed an intelligent deep learning system that uses advanced algorithms to detect and classify diseases in vegetable leaves. The researchers highlight the role of their model's exceptional ability to extract prominent features and utilize attention to enhance classification accuracy. The proposed model showed high accuracy and processing speed during the experiment.

Hai et al. [3] proposed a few-shot learning framework with attention prototypes to identify and segment leafy vegetable diseases. The segmentation model was compared to the widely applied convolutional neural networks. The experiment revealed that Few- Shot Learning could effectively recognize disease categories and segment leaf diseases.

Naqvi et al. [4] proposed a hybrid disease recognition model composed of a custom convolutional neural network and a shallow classifier. The work illustrates how CNN can capture abstract feature presentations from leaf images and employs a shallow classification algorithm to predict disease category. The model showed efficient recognition accuracy and reduced complexity in portable agricultural devices.

Girmaw et al. [5] proposed a deep learning approach to identify and classify diseases affecting cabbage leaves. The article highlights the role of preprocessing steps, convolutional neural networks, and data augmentation techniques in

classifying cabbage leaf diseases. The proposed model revealed exceptional performance in detecting and classifying healthy and diseased cabbage leaves.

The reviewed articles [1]-[6] discussed advanced machine learning and deep learning approaches to identify and classify diseases in vegetable leaves as shown in table 1. Most of the discussed research works highlight the significance of CNNs, Attention Mechanisms, Few-shot learning, and multisource information fusion to enhance classification performance. However, most of the discussed research works focused on a single crop or addressed a limited range of diseases. This study aims to propose a machine learning framework to detect diseases in Bitter Gourd, Cucumber, and Pumpkin leaves. This will contribute to enhanced classification accuracy in detecting diseases affecting multiple crops.

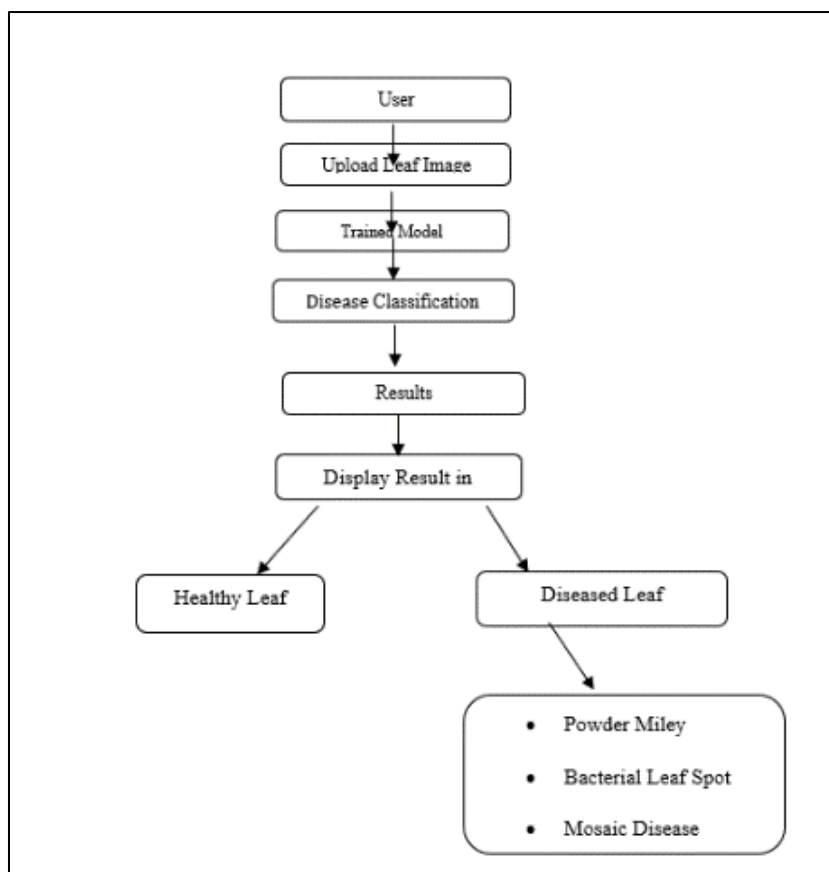
Table 1 Comparison Table

Ref. No	Author(s) & Year	Method	Key Findings	Limitations
[1]	Sujatha et al. (2025)	Machine Learning + Deep Learning	Improved disease detection accuracy using deep learning, preprocessing, and transfer learning.	Limited evaluation on multiple vegetable crops.
[2]	Liu et al. (2025)	Intelligent Deep Learning with Attention Mechanism	High classification accuracy and real-time disease detection.	Requires high computational resources and large datasets.
[3]	Hai et al. (2025)	Few-Shot Learning with Prototype Attention	Performs well with limited training samples and supports segmentation.	Complex implementation and mainly applicable to few-shot learning scenarios.
[4]	Naqvi et al. (2024)	Custom CNN + Shallow Classifier	High accuracy with reduced computational complexity.	Generalization across different crops is limited.
[5]	Girmaw et al. (2024)	CNN-Based Deep Learning Model	Accurate classification of cabbage leaf diseases.	Focused only on cabbage leaf diseases.
Proposed	This Work (2026)	Convolutional Neural Network (CNN) with Image Preprocessing and Data Augmentation	Provides automated multi-crop vegetable leaf disease detection with high accuracy, improved feature extraction, and robust classification across multiple disease classes.	Can be extended in future to include additional vegetable crops and real-time mobile/IoT-based disease detection.

3. Proposed Methodology

The proposed research develops an intelligent Leaf Disease Detection System that uses image processing and deep learning to automatically identify diseases from plant leaf images. It is designed to support farmers and agricultural professionals by offering an efficient and easy to use solution for early disease diagnosis. The system combines image preprocessing, disease classification, and real-time prediction in a single framework.

The overall workflow includes image acquisition, preprocessing, feature extraction, classification, evaluation, and web deployment. Leaf images are collected from a public dataset, organized into healthy and diseased categories, and divided into training, validation, and testing sets. OpenCV is used to improve image quality through resizing, color normalization, noise reduction, and enhancement, while a CNN learns important disease patterns such as spots, discoloration, and texture variations for accurate classification. The trained model is then evaluated using Accuracy, Precision, Recall, and F1-Score and deployed in a Flask web application for instant prediction through a simple browser interface. In this, we use Convolutional Neural Network (CNN) Machine Learning Model and below is the Architecture Diagram of our Project.

**Figure 1** System Architecture

3.1. Dataset Description

The study used a custom dataset of Cucumber, Pumpkin, and Bitter Gourd leaf images—both healthy and diseased sourced from reliable agricultural repositories and public datasets and organized into class-specific folders. Images were split into training, validation, and testing sets to enable effective learning, performance monitoring, and unbiased evaluation. All images were resized to a uniform dimension, and preprocessing (normalization and data augmentation) was applied to reduce computational cost and boost dataset diversity. The dataset captures variations in leaf color, texture, shape, and disease symptoms, supporting accurate disease classification. Including multiple crop species enhances the model's adaptability and eases future extensions to additional crops.

Table 2 Dataset Split

Dataset Category	Training data	Validation data	Testing data
Percentage	70%	15%	15%

Table 3 Dataset Characteristics

Parameter	Description
Image Type	RGB Color Images
Image Format	JPG / PNG
Input Size	224 × 224 Pixels
Dataset Categories	Healthy and Diseased Leaves

Crop Types	Cucumber, Pumpkin, Bitter Gourd
Deep Learning Model	Convolutional Neural Network (CNN)
Preprocessing Tool	OpenCV
Programming Language	Python

Table 4 Leaf Disease Description

Type	Sample Image	Healthy/Unhealthy	Number of Images
Bitter Gourd Healthy		Healthy	551
Bitter Gourd Downey Mildew		Unhealthy	570
Bitter Gourd Fusarium Wilt		Unhealthy	502
Bitter Gourd Mosaic Virus		Unhealthy	600
Cucumber Healthy		Healthy	1229
Cucumber Downy Mildew		Unhealthy	1284
Cucumber Powdery Mildew		Unhealthy	1190
Pumpkin Healthy		Healthy	400
Pumpkin Bacterial Leaf Spot		Unhealthy	400
Pumpkin Mosaic Disease		Unhealthy	400

3.2. Image Pre-processing

Image pre-processing is an important stage in the proposed Leaf Disease Detection System because image quality directly affects model performance. Since leaf images may contain variations in lighting, background, resolution, and noise, preprocessing is used to standardize them before classification. The uploaded image is first checked for a valid format, then resized to 224×224 pixels to match the CNN input size. After that, color normalization and noise reduction are applied to improve clarity and remove unwanted disturbances.

Table 5 Image Pre-processing Techniques

Pre-processing Step	Purpose
Image Validation	Ensures only valid image files are processed
Image Resizing	Converts all images to a fixed size of 224×224 pixels
Color Normalization	Reduces variations caused by lighting conditions
Noise Reduction	Removes unwanted image noise using Gaussian Blur
Image Enhancement	Improves contrast and highlights disease symptoms
Data Augmentation	Increases dataset diversity and reduces overfitting

3.3. Feature Extraction using Convolutional Neural Network (CNN)

Feature extraction in the proposed Leaf Disease Detection System is handled automatically by the Convolutional Neural Network (CNN), eliminating the need for manual feature engineering. Preprocessed images pass through multiple convolutional layers that detect simple patterns (edges, curves, color variations) in early layers and capture complex traits (lesions, discoloration, irregular textures, infected regions) in deeper layers. ReLU activation introduces non-linearity while Max Pooling reduces feature map size, preserves key information, and lowers computational cost. The resulting feature maps are flattened and fed to fully connected layers, with Softmax producing the final disease-class prediction. This automatic feature-learning pipeline enhances the accuracy and reliability of disease classification.

3.4. Disease Detection and Classification

After extracting meaningful features from the input images, the trained CNN model performs disease classification. The classifier analyzes the learned feature representation of each leaf image and compares it with the disease patterns learned during training. The proposed system classifies the uploaded image into one of the predefined categories, including healthy leaves and various disease classes associated with Cucumber, Pumpkin, and Bitter Gourd plants. During prediction, the Softmax layer generates probability scores for all available classes. The disease category with the highest confidence score is considered the final prediction. The trained model is integrated into a Flask web application, enabling users to upload leaf images through a web browser.

3.5. Flask Web Application

To make the proposed model easily accessible, a web-based application was developed using the Flask framework. The web application serves as the interface between the user and the trained deep learning model. The application begins with a professional landing page that introduces the objectives and features of the system. Users can create an account using the registration page and securely access the application through the login module. Once authenticated, users are redirected to a personalized dashboard where they can upload plant leaf images for disease analysis.

3.6. Experimental Setup

The proposed system was implemented using Python as the primary programming language. OpenCV was employed for image preprocessing, while TensorFlow and Keras were used to develop and train the Convolutional Neural Network. Flask was utilized to deploy the trained model as a web application. Training was performed on a computer equipped with a modern processor and sufficient memory to handle deep learning operations. The dataset was divided into training, validation, and testing subsets to evaluate the performance of the proposed model under unseen conditions.

Table 6 Hardware Configuration

Component	Specification
Processor	Intel Core i3
RAM	8 GB or Higher
Storage	256 GB SSD
Operating System	Windows 11

Table 7 Software Configuration

Software	Version
Python	3.10
OpenCV	2.13
TensorFlow	2.7
Keras	3.7
Flask	5.6
VS Code	13.33
SQLite	8.1

3.7. Performance Evaluation

The effectiveness of the proposed system was evaluated using standard classification performance metrics. These metrics provide a comprehensive understanding of the model's ability to correctly identify plant diseases.

Accuracy measures the percentage of correctly classified images among all test samples.

Formula:
$$Accuracy = \frac{TP+FP}{TP+FP+TN+FN}$$

Where:

TP – True Positive

TN – True Negative

FP – False Positive

FN – False Negative

Precision indicates how many images predicted as a particular disease actually belong to that disease category.

Formula:
$$Precision = \frac{TP}{TP + FP}$$

Recall measures the ability of the model to correctly identify all positive disease samples.

Formul:
$$Recall = \frac{TP}{TP+FN}$$

F1-Score provides a balanced evaluation by combining Precision and Recall into a single performance metric.

Formula:

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

A confusion matrix was also generated to visualize the classification performance across different disease categories and identify any misclassifications.

4. Results and Discussion

The trained Convolutional Neural Network demonstrated effective performance in identifying diseases from leaf images belonging to Cucumber, Pumpkin, and Bitter Gourd plants. During testing, the model successfully recognized disease symptoms such as discoloration, leaf spots, and infected regions with high consistency. Image preprocessing using OpenCV significantly improved the quality of input images and contributed to more stable feature extraction. The combination of preprocessing and deep learning enabled the proposed system to achieve reliable classification performance across different plant species. The Flask web application successfully integrated the trained model into an interactive platform where users could upload leaf images and receive prediction results in real time. The response time remained satisfactory, making the application suitable for practical agricultural use. The evaluation metrics demonstrated that the proposed system can accurately distinguish healthy and diseased leaves while maintaining good generalization capability on unseen test images.

Table 7 Performance Comparison

Metric	Accuracy	Precision	Recall	F1-Score
Value	95.67%	94.71%	77.44%	94.06%

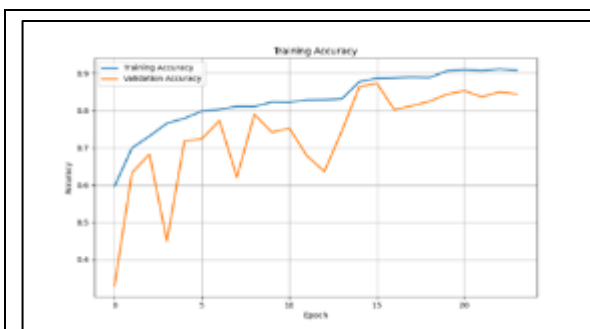


Figure 2 Training Accuracy

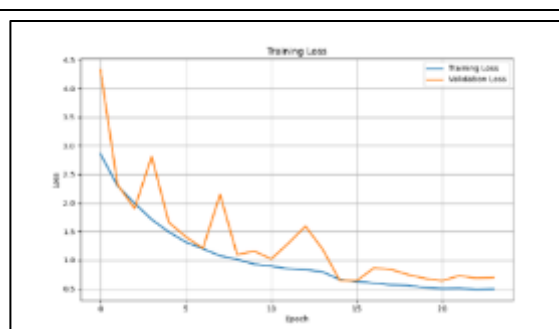
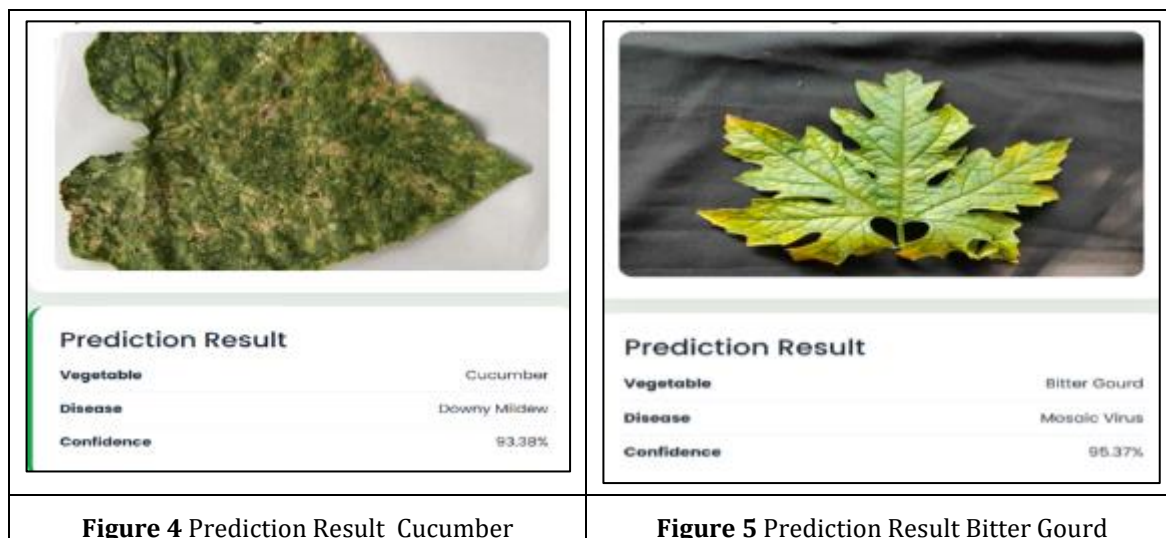


Figure 3 Training Loss



5. Conclusion

This research presents an intelligent Leaf Disease Detection System that combines OpenCV-based image preprocessing with a Convolutional Neural Network to automatically identify plant diseases. The developed framework analyses leaf images, extracts important visual features, and accurately classifies diseases affecting Cucumber, Pumpkin, and Bitter Gourd plants. By integrating image processing, deep learning, and a Flask-based web application, the system offers an efficient and user-friendly solution for early disease diagnosis. It reduces the need for manual inspection and provides quick prediction results that can help farmers and agricultural professionals make timely crop management decisions. The experimental results show that the proposed framework delivers reliable disease classification while remaining practical for real-world use. Overall, this system shows strong potential for improving agricultural productivity through intelligent plant health monitoring.

Future Scope

Although the proposed system achieves effective disease detection, several improvements can be incorporated in future research to enhance its capabilities.

- Extend the system to support additional crop species and a larger variety of plant diseases.
- Support multilingual interfaces to improve accessibility for farmers from different regions.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

- [1] R. Sujatha, S. Krishnan, J. M. Chatterjee, et al., "Advancing plant leaf disease detection integrating machine learning and deep learning," *Scientific Reports*, vol. 15, Art. no. 11552, Apr. 2025.
- [2] J. Liu, X. Wang, Q. Chen, P. Yan, and D. Guo, "Intelligent deep learning architecture for precision vegetable disease detection advancing agricultural new quality productive forces," *Frontiers in Plant Science*, vol. 16, Aug. 2025.
- [3] T. Hai, Y. Shao, X. Zhang, et al., "An efficient model for leafy vegetable disease detection and segmentation based on few-shot learning framework and prototype attention mechanism," *Plants*, vol. 14, no. 5, Art. 760, Mar. 2025.
- [4] S. A. F. Naqvi, M. A. Khan, A. Hamza, et. al, "Fruit and vegetable leaf disease recognition based on a novel custom convolutional neural network and shallow classier," *Frontiers in Plant Science*, vol. 15, Sept. 2024.

- [5] D. W. Girmaw, A. O. Salau, B. S. Mamo, et. al, “A novel deep learning model for cabbage leaf disease detection and classification,” Discover Applied Sciences, vol. 6, 6Art. no. 521, Oct. 2024.
- [6] Anusha Jajur, J., Kumar Siddamallappa, U. (2026). A Descriptive Study of Fungal Diseases in Onion Classification. In: Bhattacharyya, S., Samanta, D. (eds) Recent Trends in Intelligence Enabled Research. ISSIP 2025. Lecture Notes in Electrical Engineering, vol 1626. Springer, Cham. https://doi.org/10.1007/978-3-032-25038-4_29
- [7] J. Anusha Jajur and U. Kumar Siddamallappa (2026), “Onion Image Classification Using CNN Models”, ICTACT Journal on Image and Video Processing, Volume: 16 , Issue: 3, Page No. : 3839 – 3847, ISSN No. : 0976-9102, 10.21917/ijivp.2026.0541