A Generalized Deep Learning Approach for Cross-Crop Plant Disease Detection Using the Plant Village Dataset

[1]Roopa R, [2]Rajesh Lingam, [3]Santosh Kumar Ravva, [4]Suresh A, [5]Penubaka Balaji and [6]Avanija J
[1]Department of Computer Science and Engineering (Data Science), Madanapalle Institute of Technology and Science, Madanapalle, Andhra Pradesh, India.
[2]Adobe Inc., Boston, U.S.A.
[3]Department of Computer Science and Engineering, Vasavi College of Engineering, Hyderabad, Telangana, India.
[4]Department of Computer Science and Engineering, GRT Institute of Engineering and Technology, Tiruttani, Tamil Nadu, India.
[5]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation (Deemed to be University), Guntur, Andhra Pradesh, India.
[6]School of Computing, Mohan Babu University (Erstwhile Sree Vidyanikethan Engineering College), Tirupati, Andhra Pradesh, India.
[1]roopa509@gmail.com, [2]rlingam@adobe.com, [3]santosh@staff.vce.ac.in
[4]csesuresh6@gmail.com, [5]penubakabalaji.cse@gmail.com, [6]avans75@yahoo.co.in

Correspondence should be addressed to Roopa R: roopa509@gmail.com

Article Info
Journal of Machine and Computing (https://anapub.co.ke/journals/jmc/jmc.html)
Doi: https://doi.org/10.53759/7669/jmc202505126
Received 10 March 2025; Revised from 06 May 2025; Accepted 10 June 2025.

Abstract – Plant diseases continue to be one of the leading causes of reduced agricultural productivity worldwide, directly threatening food supply chains and the economic stability of farming communities. With the global population steadily increasing, the demand for intelligent, scalable, and highly accurate plant disease detection systems has never been more critical. Deep learning methods have shown promising results in this field; however, numerous conventional models cannot often generalize well across different crop species and unseen disease types. These limitations hinder their practical deployment in dynamic real-world agricultural environments. In this study, we propose a robust and generalized deep learning-based approach for cross-crop plant disease detection, using the comprehensive and diverse Plant Village dataset. Our model is built upon a custom-designed Convolutional Neural Network (CNN) architecture that incorporates a small Inception module. Unlike traditional CNNs, which primarily focus on the global features of a leaf. Our model detects and analyzes localized disease spread patterns, enhancing detection across diverse crops and adapting to novel conditions. The small Inception module plays a vital role in enabling multi-scale feature extraction from small disease-affected patches without adding excessive computational complexity. This architectural refinement allows the model to learn more discriminative features, resulting in faster convergence and higher classification accuracy. When trained and validated on the Plant Village dataset, our model achieved an impressive accuracy of 98.45%, outperforming many traditional approaches. Additionally, it demonstrated consistently high precision, recall, and F1-score, confirming its reliability and robustness. By addressing the challenges of overfitting and poor generalization, common pitfalls in many deep learning models, our method provides a scalable and effective solution for real-time agricultural disease monitoring. This work contributes to the growing field of precision agriculture by offering a model that is not only accurate but also generally efficient and practical for deployment in diverse agricultural settings. Ultimately, our research aims to support the development of smart farming technologies that ensure healthier crops and contribute to long-term global food security.

Keywords – Plant Disease Detection, Deep Learning, Convolutional Neural Network (CNN), Inception Module, Cross-Crop Classification, Plant Village Dataset, Image-Based Diagnosis, Disease Localization.

I. INTRODUCTION

Agriculture plays a vital part in all economic systems and is a basic source of food that sustains life. Improving the agricultural industry is essential to increasing output and raising quality. This enhancement necessitates establishing the ideal environment for crops and plants to flourish healthily. Plant degradation is frequently caused by diseases. According to the United Nations Food and Agriculture Organization, chronic illnesses cost the world economy some $220 billion a year. They may cause serious harm to the crop or perhaps their total annihilation. Plants can be attacked by bacteria, fungi, viruses, and microscopic animals, which can change their natural form and interfere with their essential processes[1]. Plants can be saved when infections are detected early and neutralized. The crops are safeguarded, and losses are prevented to a greater extent, the sooner they are identified. The conventional disease detection methods, which mostly rely on human diagnosis, are time-consuming and inadequate because of a lack of expertise[2]. The diagnosis must be based on a more trustworthy technique because the data collection method and verification frequency are also insufficient. Modern technologies have been provided as an automatic way to identify plant diseases for this aim [3]. Cutting-edge technologies like sensors, drones, and robots have drastically altered the way farmers manage their crops [4]. Although machine learning has opened up possibilities for studying data, several issues must be resolved.

Plant disease identification using machine learning and deep learning approaches is a quickly developing subject with encouraging outcomes [5]. Yet since deep learning-based techniques rely on automatic feature extraction rather than human feature selection, they have outperformed other machine learning techniques, particularly in the area of image identification [6]. Deep learning-based techniques for the identification and detection of plant diseases have been proposed in a few research areas. But there are a lot of barriers that keep this technology from being used more effectively. The impossibility of gathering dataset photos for every illness across all leaf kinds is, in fact, one barrier. Furthermore, a few diseases spread quickly, making it difficult to catch them on leave in time. Furthermore the model is impacted by the traits of the crop and the disease during the learning process [7]. This suggests that characteristics that can be extracted from one disease and crops cannot be applied to other diseases and crops. A dataset comprising pairs of distinct crops and diseases must be used to train the classification model to develop a generalized method for classifying plant ailments. Regretfully, there isn't a dataset like that, and making one is extremely challenging, if not impossible. The globe is home to millions of plants and crops, and these organisms are susceptible to millions of illnesses. However, no such dataset exists, and it is very difficult, if not impossible, to create one. There are millions of plants and crops in the world, and millions of diseases may affect these plants and crops. It takes time and effort to acquire data. These problems directly affect the performance of deep learning-based systems due to the lack of data. Lack of access to sufficient datasets hinders the deep learning system's capacity to generalize patterns learned in plant disease recognition [8].

Even though many deep learning models have been created for the identification of plant diseases, most of them are specialized for a single crop or dataset, which restricts their applicability to other plant species. They are less useful in real-world settings where farmers might grow multiple crops at once because of their crop-specific concentration. Creating a unified or generic deep learning model that can correctly classify illnesses across many crop kinds is a major challenge. This calls for a robust architecture that can manage a great degree of variety in leaf forms, colors, textures, and illness signs, as well as a diverse and representative dataset. To address this issue, this study presents a deep learning-based approach that generalizes plant disease detection across different crops and disease kinds to overcome the drawbacks of crop-specific models. Our method concentrates on precisely detecting the disease rather than depending only on the sight of damaged leaves. A generalized deep learning model is chosen for its scalability, efficiency, and adaptability through transfer learning. Using the PlantVillage dataset, we aim to enable early diagnosis, assist farmers in low-resource settings, and support sustainable agriculture. The following are the work's significant contributions:

- Generalized Disease Detection Model, we develop a unified deep-learning model capable of accurately identifying plant diseases across multiple crop types, eliminating the need for separate models for each crop.
- Our strategy places more emphasis on learning disease-specific patterns than existing approaches, which mostly rely on crop-specific visual cues. This improves the model's generalizability and practicality.
- Scalable and Easy to Adapt for Real-World Use, our model uses transfer learning, which means it can quickly learn to detect diseases in new crops using very little additional data. This makes it a practical, low-cost solution that can be used in different farming environments and helps support sustainable agriculture.
- A new version of the PlantVillage dataset was constructed by reprocessing existing images into labeled patches (healthy vs. unhealthy). This restructured dataset, consisting of over 1.8 million patches, enables scalable and flexible training of generalized models.
- Our proposed model was evaluated alongside other state-of-the-art architectures such as EfficientNet-B0, ResNet-50, MobileNetV2, and Vision Transformer (ViT), providing a thorough comparative performance analysis regarding accuracy, loss, and generalization behavior.

II. LITERATURE REVIEW

With the world's population growing and natural resources becoming limited, researchers are using various methods to monitor crops and seedlings, especially to detect diseases. Because deep learning has worked well in many areas, many studies now focus on using it to diagnose plant and crop diseases. However, researchers face several challenges when developing reliable and effective deep learning models for detecting diseases in plant leaves.

However, Chowdhury et al. [9] proposed a CNN-based model for detecting diseases in plant leaves commonly grown in Bangladesh. They trained the model using online datasets, which were mainly collected from other countries, and achieved an accuracy of 85.31%. Most existing research has similarly focused on plant data from different regions, not locally sourced samples.

Table 1. Summary of Key Studies on Crop Disease Detection

Reference	Dataset	Model	Accuracy (%)	Limitations
[9]	Plant-Village	CNN	85.31	The model struggles to detect multiple diseases affecting the same leaf simultaneously.
[10]	Plant-Village	Slender-CNN	99.81	Real-time learning features are not yet implemented, limiting its ability to improve continuously from new data.
[11]	self	ResNet	94.00	The model was only evaluated using preprocessed images, which may not reflect real-world conditions like poor lighting, background noise, or partially damaged leaves.
[12]	PlantVillage	CNN-ResNet	99.24	The model relies on pre-trained CNNs fine-tuned on ImageNet, which may limit its ability to fully capture specific disease features unique to certain crops
[13]	PlantVillage	RTR_Lite_ MobileNet	99.92	Does not explore how the model performs in real-time field conditions with varying lighting and backgrounds.
[14]	MangoLeafBD	InceptionV3	99.87	Although the model shows high accuracy, its performance in real-time or field conditions is not evaluated.
[15]	PlantVillage	DSDNN	99.00	While the model performs well on the given dataset, its effectiveness in diverse agricultural settings with different crops and disease types remains uncertain.
[16]	self	MaxViT	100.00	Vision Transformers require significant computational resources, making real-time deployment of low-power agricultural devices challenging.
[17]	PlantVillage	Depthwise CNN + SE + Residual	96.00	Data imbalance affects model performance; it needs better augmentation and improved interpretability.
[18]	PlantVillage	DenseNet201 + SVM	99.82	Scalability remains a challenge, as the model may not efficiently scale for large-scale agricultural deployment across different farming regions.
[19]	Plant Village	XI-CNN	100.00	The model focuses on classification but does not provide insights into disease severity

A recent study, Baiju et al. [10]'s Slender-CNN detected key diseases in corn, rice, and wheat with 99.81% accuracy, but missed some less common diseases and crops. Its lightweight design suits mobile and drone use, and it has the potential to expand to other crops and real-time field data. Agriculture plays a vital role in any growing economy, but plant diseases threaten their sustainability, especially in crops like tomato and potato. Early detection is crucial, and this study, Kalaivani et al. [11], used machine learning (SVM) and deep learning (ResNet) to identify leaf diseases. After preprocessing and augmentation of the datasets, both models are tested. ResNet outperforms SVM with 94% accuracy and is recommended for real-time use, along with fertilizer suggestions based on disease type. Aboelenin et al. [12] combined a hybrid model, CNNs and Vision Transformers for plant leaf disease detection. Tested on apple and corn datasets from PlantVillage, it achieved accuracies of 99.24% and 98%. The model outperforms many existing methods and shows promise for broader agricultural and computer vision applications.

By incorporating attention mechanisms, Duhan et al. [13] RTR_Lite_MobileNet improve upon MobileNetV2, increasing accuracy on plant disease datasets by up to 99.92%. Because of its efficiency and portability, it allows for low-latency real-time detection on devices such as the Raspberry Pi. Additionally, the model shows potential for real-world agricultural applications by performing well on complex, **real**-world data. Mango cultivation plays a vital role in food security and the economy of tropical regions, but it is vulnerable to various leaf diseases that impact yield and quality. Early detection is essential for sustainable farming. In recent studies, Varma et al. [14] have applied deep learning models such as VGG19, InceptionV3, ResNet152V2, DenseNet121, and others to detect mango leaf diseases. Among them, InceptionV3 achieved the highest accuracy of 99.87%, demonstrating its effectiveness compared to other models.

In other studies, Kaushik et al. [15] proposed a deep-wise separable-based adaptive neural network (DSDNN) for plant disease detection, particularly in potatoes. Using the PlantVillage dataset, the model integrates Gaussian filtering, Enthalpy-based graph clustering, and advanced feature extraction techniques. The approach achieved 99% accuracy, outperforming previous methods. Despite its success, the model's performance is only validated on the PlantVillage dataset, which may limit its effectiveness in real-world scenarios. Early detection of corn leaf diseases is essential to prevent crop loss. While CNNs have been widely used for disease detection, recent Vision Transformer (ViT) models perform better. This study Pacal and Ishak et al. [16] used ViT models (MaxViT, DeiT3, MobileViT, MViTv2) alongside CNNs (VGG, ResNet, DenseNet, Xception) with data augmentation and transfer learning.

An ensemble method further improves accuracy. Tested on PlantVillage and a new CD&S dataset, the models achieved up to 100% accuracy, outperforming previous methods. This approach offers a reliable, automated solution for corn disease diagnosis. Ashurov et al. [17] introduced a Depthwise CNN model with SE blocks and skip connections, achieving 96% accuracy for plant disease detection. This study, Bhola and kumar et al. [18] presented a hybrid model using DenseNet201 and SVM to detect diseases in Corn, Wheat, and Rice. It achieves high accuracy up to 99.82% with a lightweight design, making it effective for real-time crop disease detection. Furthermore, Shafik et al. [19] introduced a hybrid IX-CNN model combining Inception and Xception layers for plant disease detection. It achieved up to 100% accuracy on multiple datasets, such as PlantVillage and Turkey Disease, and over 98% on others. The model uses SVM, DT, and RF classifiers and supports real-time platform applications. **Table 1** summarizes several studies on crop disease detection.

While the works above have significantly contributed to plant disease detection, several vital gaps remain:

- Generalization Across Crops: The inability of current deep learning models to reliably identify diseases in a variety of plant species restricts their applicability in a range of agricultural contexts.
- Dataset Limitations: The PlantVillage dataset does not include real-world changes like ambient noise, lighting variances, and mixed cropping scenarios. The resilience of the model is weakened by this lack of variance.
- Computational Restrictions: Deep learning models have high resource requirements, making it difficult to deploy them on mobile and edge devices, limiting farmers' access.

III. METHODOLOGY

This section outlines the presented methodology for detecting abandoned objects in various weather conditions. The approach utilizes a hybrid model and DAWN dataset for object detection and a weather enhancement module to handle adverse situations.

Dataset Description

In this work, we fetched an open-access dataset released through the online platform named PlantVillage, which is widely recognized in the domain of agricultural image analysis and is primarily used for classifying and recognizing plant leaf diseases [20]. It was selected for its diversity, comprehensiveness, high-quality annotations, benchmark status in plant pathology research, accessibility, and strong relevance to real-world agricultural challenges. The PlantVillage dataset comprises 54,305 RGB images, including healthy and diseased leaves. These images span 14 crop species and cover 20 distinct disease types, categorized into 38 classes, 12 healthy leaf classes, and 26 diseased leaf classes. All images were captured under controlled laboratory conditions with a uniform background and standardized to a resolution of 256 × 256 pixels, ensuring consistency and ease of processing for deep learning models.

Fig 1. Examples From the Plantvillage Dataset: (Left) Original Images with Uniform Backgrounds and (Right) Corresponding Images with Segmented Backgrounds.

For this study, we utilize an enhanced version of the PlantVillage dataset introduced in [21], in which the leaf regions are segmented from the background to better emulate real-world scenarios and improve the model's robustness in field conditions. This preprocessed version reduces background noise and enables the learning algorithms to focus solely on the relevant leaf features. **Fig 1** illustrates comparative examples of the original images with uniform backgrounds and the segmented images with backgrounds removed.

The Refined PlantVillage Dataset

The updated PlantVillage dataset consists of 1,860,316 image patches divided into 1,014,154 unhealthy and 846,162 healthy samples [21]. **Table 2** visualization of the disease name with healthy and unhealthy patches. The dataset's labeling was guided by the nature and spread of specific plant diseases. Notably, diseases such as Huanglongbing (citrus greening)**,** leaf blight (Isariopsis leaf spot)**,** and powdery mildew are known to affect the entire surface area of infected leaves uniformly. Based on this characteristic, all patches extracted from leaves exhibiting these diseases were consistently labeled as unhealthy**,** as the presence of the infection was assumed to extend across the entire image region. This labeling strategy ensured the dataset's integrity and enhanced the training process's robustness by reducing label noise, particularly for disease types that manifest across full-leaf surfaces. Using this carefully selected dataset, a generalized plant disease detection model that can reliably differentiate between healthy and diseased leaf sections, irrespective of the crop type, may be trained. **Table 3** displays the Number of Samples and Patches per Crop.

Table 2. Visualization Of the Disease Name with Healthy and Unhealthy Patches

Disease Name	**Unhealthy Patches**	**Healthy Patches**
Healthy (All crops)	0	647614
Bacterial spot	97506	72307
Black rot	23749	38096
Cedar apple rust	5831	2273
Cercospora leaf spot/Gray leaf spot	22738	2400
Common rust	56887	2911
Early blight	35947	31302
Esca (Black measles)	27737	19499
Haunglongbing (Citrus greening)	194626	0
Late blight	45011	38440
Leaf mold	14596	7774
Leaf blight (Isariopsis leaf spot)	43106	0
Leaf scorch	39752	1310
Northern leaf blight	40987	5782
Powdery mildew	119279	0
Septoria leaf spot	34377	16983
Spider mites (Two-spotted spider mite)	15110	26823
Target spot	6086	38714
Tomato yellow leaf curl virus	155739	0
Tomato mosaic virus	5209	2175
Scab	10998	11986
Total	**1014154**	**846162**

Table 3. Number Of Samples and Patches Per Crop

Crop Name	**Number of Samples**	**Number of Patches**
Apple	3171	97269
Blueberry	1502	38080
Cherry (incl. sour)	959	64287
Corn (maize)	2857	158735
Grape	3639	127516
Peach	2657	67970
Pepper, bell	2475	91126
Potato	352	76541
Raspberry	371	12723
Soybean	5090	190448
Strawberry	1565	57966
Tomato	23060	625250
Orange	5507	194626
Squash	1835	83457

Proposed Method

This section explains the method used to build a general-purpose system for identifying plant diseases from leaf images. Once trained, the system can recognize signs of infection without needing prior details about the type of plant or the specific disease. A notable feature of this system is its ability to estimate how much of the leaf is affected by the disease. **Fig 2** gives a clear overview of the entire process, showing each key step involved in the system's work.

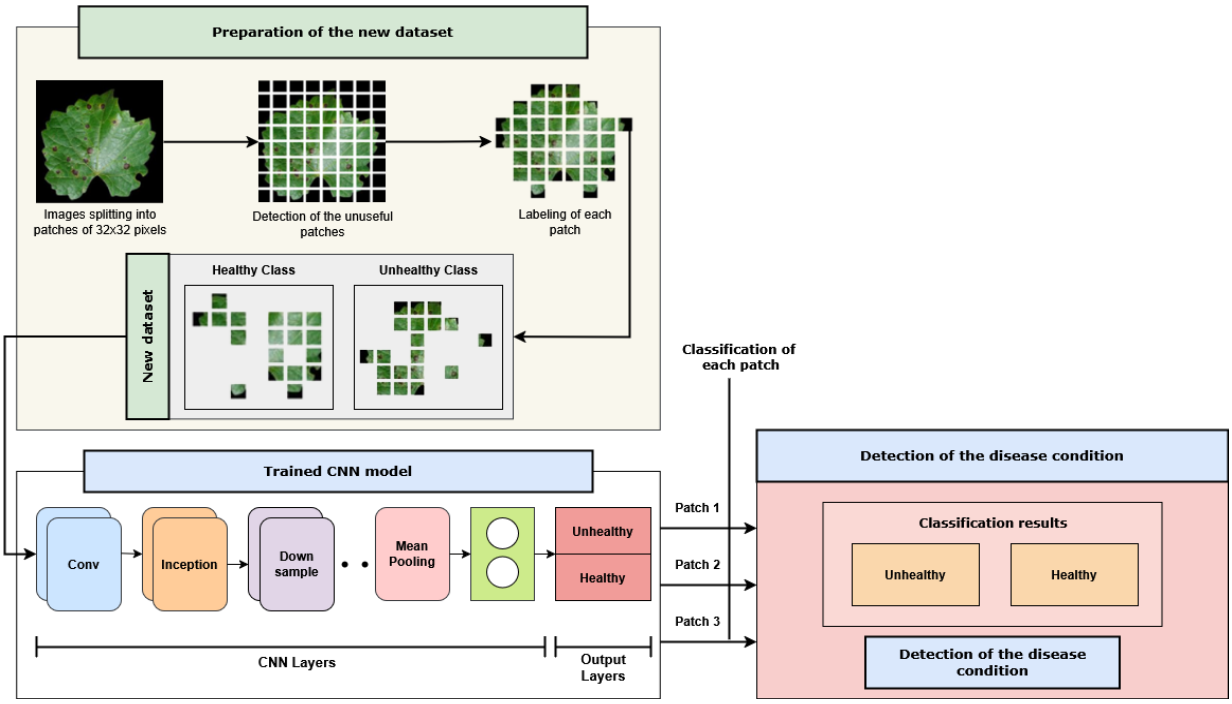

Fig 2. Overview of the Proposed Convolutional Neural Network (CNN) Workflow.

Generalization of the Process of Crop Disease Detection

To promote generalization across diverse crop types, we introduce a novel approach that removes dependence on crop-specific features and focuses exclusively on disease-related visual cues. Rather than training the model using full leaf images, which inherently contain the crop's identifying characteristics, we decompose each image into small, uniform patches. These patches are sufficiently small to eliminate structural clues related to the crop species, allowing the model to concentrate solely on patterns associated with disease symptoms.

Each input image, denoted as I, is split into square patches, referred to as S_1, S_2,…,S_n, where each patch has fixed dimensions of 32 × 32 pixels. This resolution was carefully selected to be compatible with lightweight convolutional neural network (CNN) architectures, enabling efficient training and inference. To accommodate any image dimension, we resize each image to the nearest multiple of 32 pixels along both height H and width W, then divide it into a grid of patches. The total number of patches per image N_s can be calculated as:

$$N_s = [\frac{H}{P_h}] \times [\frac{W}{P_w}] \tag{1}$$

$$I = \{S_1, S_{2,\dots\dots,} S_{N_s}\} \tag{2}$$

where, H and W are the height and width of the resized image, and, P_h and P_w are the patch height and width (set to 32 pixels).

When a 256 × 256-pixel image is split into an 8 by 8 grid, for example, the following is produced:

$$N_s = \frac{256}{32} \times \frac{256}{32} = 64\, patches \tag{3}$$

We applied a filtering mechanism to eliminate irrelevant or empty patches based on the proportion of black pixels in each patch. A pixel is black if its Red, Green, and Blue (RGB) values are all zero. The percentage of black pixels in a given patch S is computed as:

$$B_{\%} = (\frac{1}{P} \sum_{i=1}^{P} \delta \ (R_i = 0 \wedge G_i = 0 \wedge B_i = 0) \times 100\,) \tag{4}$$

where, P is the whole number of pixels in the patch, $R_{i,}G_{i,}B_i$ are the RGB values of the i^{th} pixel and δ (·) is an indicator function producing 1 if the condition is true and 0 otherwise.

Patches with a black pixel percentage of 100%**,** indicating a complete absence of leaf content, are discarded. Among the remaining patches, we retain only those whose black pixel percentage is less than or equal to that of the original image (empirically determined to be approximately 50%). This selection ensures fidelity to the original dataset distribution while removing noise.

The filtered patch set for a given image is then defined as:

$$Í = \{S_K \in I \mid B_{\%} (S_k) \leq B_{threshold}\} \tag{5}$$

where, $B_{threshold} \approx 50\%$.

Generalization Across Disease Types

Our proposed strategy shifts the focus from identifying individual diseases to a more foundational task: detecting whether a leaf is infected regardless of the disease type. We achieve this by training the model to distinguish between healthy and infected patches without classifying the exact disease. Specifically, we aggregate healthy patches across all crop types to build a robust feature representation for healthy leaf tissue. In parallel, we gather infected patches from all disease categories to learn the distinguishing characteristics of unhealthy leaves. This setup enables the model to generalize disease detection, even for previously unseen disease types. To construct this binary classification framework, all patches derived from the PlantVillage dataset were visually labeled by agricultural experts over three months. Each patch was labeled as either:

This binary labeling process is formally represented as:

$$Í = \{S_j \mid S_j \in I \mid \wedge \ healthy \} \ U \ \{S_k \mid S_k \in I \mid \wedge \ unhealthy\} \tag{6}$$

Using this approach, we created a new dataset version, organized into two categories where:

- H: The set of all healthy patches,
- U: The set of all unhealthy patches.

This restructured dataset can be defined as:

$$\mathrm{H} = \sum_{c=1}^{C} \sum_{i=1}^{N_c} \{S_i^C | S_i^C \in I_i^c \ \wedge \ healthy\} \tag{7}$$

$$\mathrm{U} = \sum_{c=1}^{C} \sum_{i=1}^{N_c} \{S_i^C | S_i^C \in I_i^c \ \wedge \ unhealthy\} \tag{8}$$

where, C is the number of crop-disease classes in the dataset, N_c is the number of images in class c, I_i^c is the ith image in class c, S_i^C is a patch extracted from I_i^c.

Detection and Quantification of Disease Spread

Once the model is trained to distinguish healthy from unhealthy leaf patches, it can be applied to analyze new leaf images. Following the same patching approach, each test image is divided into non-overlapping 32 × 32-pixel patches, and irrelevant patches are discarded. The remaining patches are fed to the classifier, which predicts whether each patch is healthy or diseased. This patch-wise prediction not only facilitates disease detection but also enables the quantification of the disease's spatial extent on the leaf.

To compute the disease prevalence across the entire leaf, we count the number of patches predicted as unhealthy N_U and healthy N_H. The percentage of diseased area $D_{\%}$ is then calculated as:

$$D_{\%} = \left(\frac{N_U}{N_H + N_U}\right) \times 100 \tag{9}$$

where, N_U *is* The number of unhealthy (infected) patches and N_H *is* Healthy patches.

This metric provides a direct, interpretable measure of disease severity on each leaf, enabling more informed agricultural assessments and interventions.

CNN Architecture

The convolutional neural network (CNN) for classifying the extracted leaf patches is based on the small Inception model**,** a lightweight variation of the original GoogLeNet Inception architecture. GoogLeNet's demonstrated conquest in plant disease detection tasks motivated the selection of this architecture. The smaller variant offers improved computational efficiency and is better suited for small input sizes, such as 32 × 32-pixel patches.

The small Inception architecture comprises three primary building blocks:

- **Conv Module**: A standard convolutional layer followed by batch normalization and activation.

- **Inception Module**: A simplified version of the original Inception block, which includes multiple parallel convolution operations with different kernel sizes to capture multi-scale features.
- **Downsample Module**: A module that combines convolution and pooling operations to reduce spatial dimensions while preserving feature richness.

Fig 3 illustrates a schematic overview of the small Inception architecture, illustrating how these modules are organized within the network.

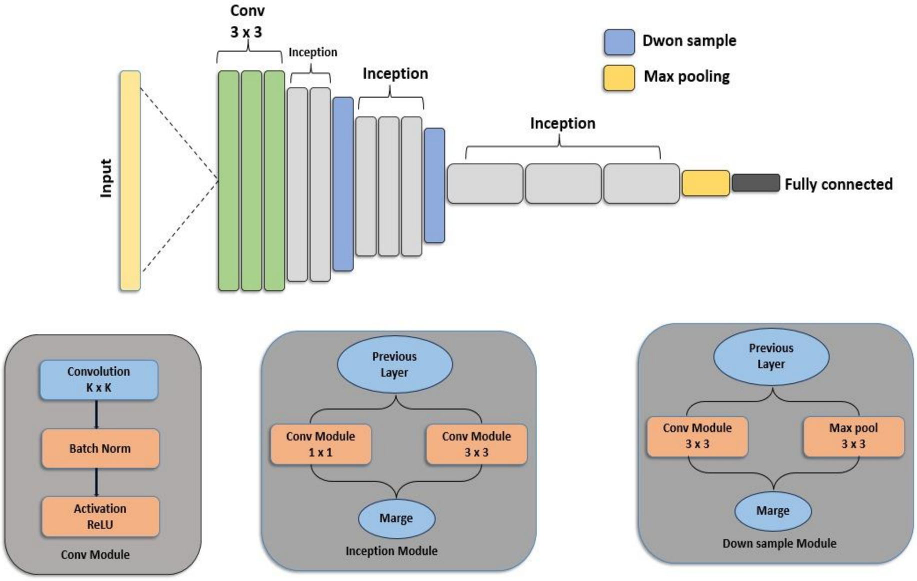

Fig 3. The Architecture of the Small Inception Model.

IV. EXPERIMENTAL SETUP AND EVALUATION

This section thoroughly evaluates the proposed disease detection method, emphasizing its generalization capacity and resilience across various crop species and disease kinds while our core model is a convolutional neural network (CNN), we also incorporate several advanced architectures—namely Efficientnet-b0, Resnet-50, Mobilenetv2, and Vision Transformer (ViT)—to conduct a comparative performance analysis. This multi-model approach ensures a thorough assessment of our method's efficacy by benchmarking it against both traditional and cutting-edge deep learning models. We begin by testing all models on a newly curated dataset to verify their accuracy in distinguishing between healthy and diseased leaf regions across different crops. To further the effectiveness of our approach, we compare our results with state-of-the-art methods that used the PlantVillage dataset. The evaluation employs a range of standard classification metrics, including accuracy, precision, recall, F1-score, loss, and confusion matrix, presenting a holistic view of each model's strengths and limitations.

Experimental Setup

Python and the TensorFlow framework were used to train the CNN model from the ground up. Extensive experimentation with various combinations was conducted to determine the hyperparameter configuration that produced the best classification accuracy while preserving generalization capability and avoiding overfitting. The optimization algorithm used was Adam, which was applied with its default parameters. Training and evaluation were conducted on Google Colab, leveraging GPU acceleration for improved computational speed and performance. **Table 4** presents the hyperparameter configuration. The dataset was partitioned as follows:

- Training Set: Comprising 80% of the data from each class:
  - 676,929 healthy patches
  - 811,323 unhealthy patches
- Validation & Test Set: Comprising the remaining 20% of each class:
  - 169,233 healthy patches
  - 202,831 unhealthy patches

This balanced and comprehensive dataset division ensured sufficient representation of both classes across the training and evaluation phases. This study adopts a deep learning-based approach using a custom Inception-based CNN to identify plant diseases through leaf patch classification.

Table 4. Hyperparameter Configuration

Hyperparameter	Values
Batch size	32
Epochs	300
Learning rate	0.001
Optimizer	Adam

Evaluation Metrics

In our study, we employed several standard evaluation metrics to evaluate the performance of our deep-learning models for plant disease detection.

Accuracy

Accuracy represents the proportion of correctly classified patches (healthy and unhealthy) out of the total patches. It provides an overall measure of classification effectiveness:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

Precision

To reduce false positives, precision evaluates the proportion of patches that are truly unhealthy compared to those that were projected to be unhealthy.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{11}$$

Recall

Recall measures how well the model detects all critical patches to minimize false negatives.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{12}$$

F1-Score

When the class distribution is unbalanced, the harmonic means of accuracy and recall, or F1-score, provides a balanced evaluation of both metrics.

$$\text{F1-score} = 2 \times \frac{Precision \times Recall}{Precision + recall} \tag{13}$$

Loss

Loss quantifies the difference between the predicted labels and the actual labels. We used **categorical cross-entropy loss**, which is standard for multi-class classification problems. Lower values indicate better performance.

Confusion Matrix

The confusion matrix thoroughly analyzes prediction outcomes, displaying the proportion of patches accurately or inaccurately categorized as healthy or unhealthy. were,

- **TP**: True Positives (unhealthy patches correctly classified)
- **TN**: True Negatives (healthy patches correctly classified)
- **FP**: False Positives (healthy patches misclassified as unhealthy)
- **FN**: False Negatives (unhealthy patches misclassified as healthy)

Performance Analysis

Assessing testing and training accuracy, together with associated loss values, provides important information about how well each model learns and generalizes. A model that maintains low and tightly matched loss values while achieving high accuracy on training and testing datasets indicates strong generalization without overfitting or underfitting. **Fig 4** depicts the training loss and accuracy curve. In this study, these metrics serve as a foundation to assess the robustness and adaptability of each deep learning architecture when applied to patch-based plant disease detection, as displayed in **Table 5**.

Among all evaluated models, the CNN-based Small Inception architecture stands out distinctly, achieving the highest test accuracy (98.45%) with minimal loss (0.0400), closely trailing its training performance (99.10%, loss 0.0250). An optimally regularized model is characterized by a near parity between training and test performance, demonstrating its capacity to generalize effectively over unknown data without learning the training set. Adapted to localized input patterns, its compact architecture works incredibly well for extracting discriminative features from heterogeneous leaf patches. On the other hand, broader or more profound structures have other drawbacks. Vision Transformer (ViT) showed a little performance difference between training and test loss (0.0500 vs. 0.1200), indicating modest overfitting despite its strength in modeling long-range relationships. Its performance reflects high learning capacity but reduced resilience to patch-level noise, likely due to its dependency on extensive pretraining or larger datasets for full effectiveness.

Table 5. Performance Analysis of Testing and Training Accuracy Along with Their Loss

Model	Train Accuracy	Test Accuracy	Train Loss	Test Loss
CNN (Small Inception)	0.9910	0.9845	0.0250	0.0400
EfficientNet-B0	0.9624	0.9395	0.0750	0.1100
ResNet-50	0.9411	0.9023	0.0950	0.1400
MobileNetV2	0.9189	0.8782	0.1400	0.1800
Vision Transformer (ViT)	0.9758	0.9420	0.0500	0.1200

EfficientNet-B0 achieved a respectable balance (train: 96.24%, test: 93.95%) with moderate loss values (0.0750 and 0.1100), showing stable learning yet slightly less confident predictions under variability. ResNet-50, while structurally deep, underperformed with a notable gap between training and testing accuracy (94.11% vs. 90.23%), suggesting it may not effectively adapt to the fine-grained texture and shape variations present in disease-infected patches. MobileNetV2, optimized for efficiency, exhibited the most evident signs of overfitting. Although its training accuracy reached 91.89%, the drop in test accuracy to 87.82%, paired with the highest loss values (train: 0.1400, test: 0.1800), indicates difficulty in generalizing to new samples.

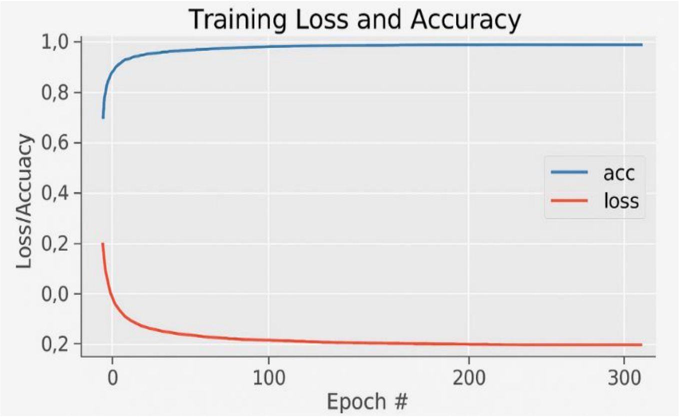

Fig 4. Training Loss and Accuracy Progression.

In essence, the Small Inception CNN surpasses other models in predictive accuracy and exemplifies architectural alignment with the problem's unique demands of compact, patch-level feature extraction from agricultural data. Its better results highlight the value of creating context-specific models rather than just modifying generic deep networks, particularly in fields with high inter-class similarity and localized patterns.

Table 6. Performance Analysis of the Models

Model	Accuracy	Precision	Recall	F1-Score
CNN (Small Inception)	0.9845	0.9852	0.9838	0.9862
EfficientNet-B0	0.9395	0.9410	0.9372	0.9375
ResNet-50	0.9023	0.8945	0.8987	0.8994
MobileNetV2	0.8782	0.8505	0.8641	0.8768
Vision Transformer (ViT)	0.9420	0.9383	0.9307	0.9398

In this study, we compared the performance of five deep learning models CNN with Small Inception modules, Vision Transformer (ViT), EfficientNet-B0, ResNet-50, and MobileNetV2 using standard classification metrics such as accuracy, precision, recall, and F1-score which shown in **Table 6**. Among these the CNN model incorporating Small Inception modules outperformed all others, achieving an impressive accuracy of 98.45%, precision of 98.52%, recall of 98.38%, and an F1-score of 98.62%. This exceptional performance can be attributed to the architectural design of the Small Inception module, which allows the network to process image features at multiple scales simultaneously. Instead of using a single convolutional kernel size, the Inception module combines convolutions in parallel, enabling the model to extract fine details as well as broader contextual patterns in one pass. Convolutions reduce dimensionality while adding nonlinearity, allowing the model to learn complicated representations more effectively.

In contrast, the Vision Transformer (ViT), which achieved the second-best results with 94.20% accuracy and an F1-score of 93.98%, relies on self-attention mechanisms to model global relationships between image patches. While ViT has shown great success on large-scale datasets, it lacks certain inductive biases such as locality and shift invariance that are naturally present in CNNs. This makes ViT less effective when the dataset is relatively small or lacks sufficient diversity, which could explain its slightly lower performance in this case. Moreover, ViT requires more data and training time to generalize well, which can be a limiting factor for smaller or medium-sized real-world datasets like ours.

EfficientNet-B0 showed competitive performance with 93.95% accuracy and 93.75% F1-score. It benefits from a compound scaling method that optimally balances depth, width, and resolution. However, B0 is the smallest and most lightweight version in the EfficientNet family, and while it's efficient, it may not have had enough representational capacity to capture the subtle and complex features of mango leaf diseases as effectively as our custom CNN. A larger variant such as EfficientNet-B0 might have yielded better results but at the cost of increased training time and resources.

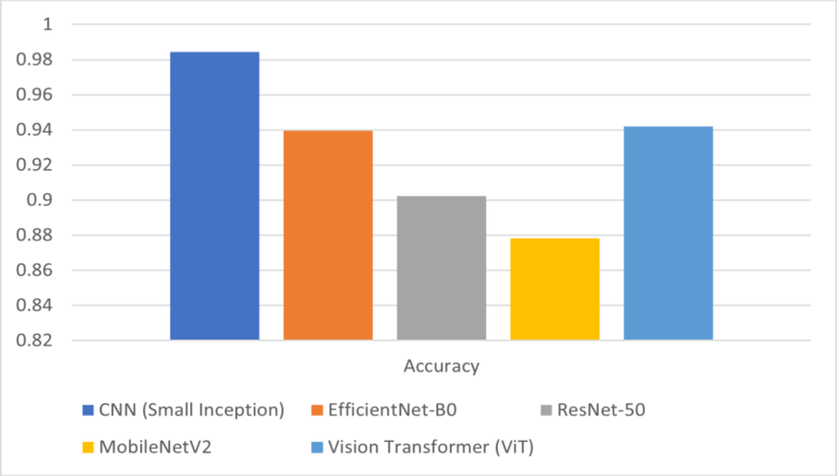

Fig 5. Accuracy Analysis of the Models.

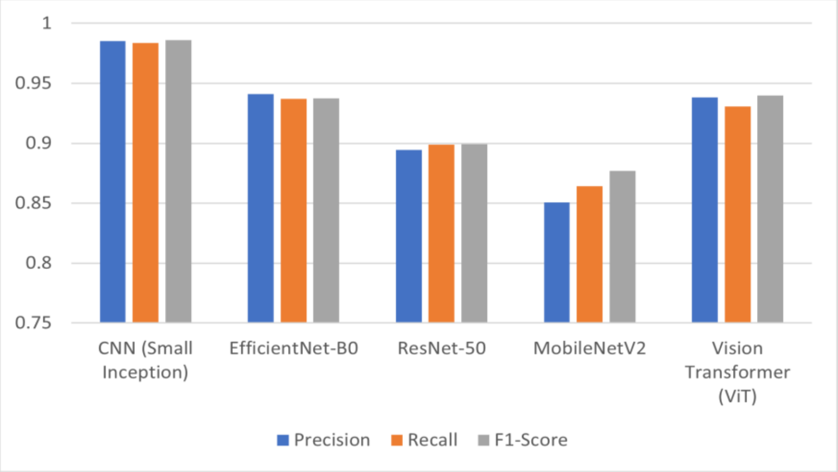

Fig 6. Performance Measurement of the Models.

ResNet-50, known for its deep architecture and residual learning capabilities, achieved 90.23% accuracy and an F1-score of 89.94%. Its performance, although respectable, was likely limited by its complexity relative to the dataset. Deep models like ResNet-50 can suffer from overfitting if the dataset size is not sufficiently large, as they may memorize training patterns rather than generalize to new, unseen data. This trade-off between depth and generalization becomes particularly evident in moderately sized image datasets were simpler, more efficient models may perform better.

MobileNetV2 recorded the lowest performance among the five, with 87.82% accuracy and an F1-score of 87.68%. Designed for mobile and embedded systems, MobileNetV2 uses depth wise separable convolutions to minimize computation and model size. While this makes it fast and resource-efficient, the downside is a significant reduction in learning capacity. This limitation led to underfitting where the model was unable to fully learn the complex patterns needed for accurate classification of various mango leaf conditions. We chose the CNN with Small Inception modules for several important reasons. First, it offers a balanced trade-off between accuracy and computational efficiency. The architecture is not as deep or parameter heavy as ResNet or ViT making it easier to train on a modest dataset without the risk of overfitting.

Second, the multi-scale feature extraction capability of the Inception module is particularly effective for plant disease classification tasks, where symptoms vary in size, shape, and color. Some disease spots are very small, while others cover larger leaf areas making a model that can detect both local and global patterns extremely valuable. **Fig 5** shows accuracy analysis of the models. Third, the reduction in parameters through 1×1 convolutions before applying larger filters allows the model to remain lightweight while maintaining high performance, which is ideal for scalable, real-world applications such as mobile disease detection tools. **Fig 6** Shows the performance measurement of the models.

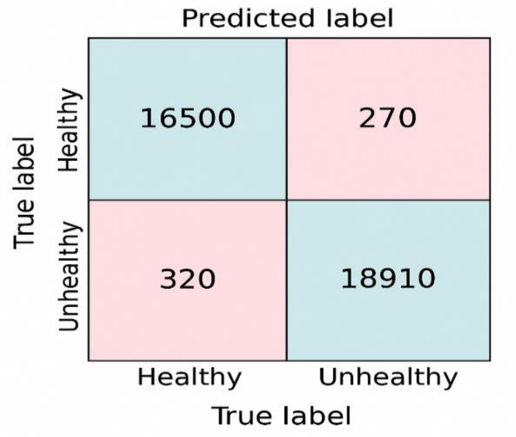

Fig 7. Confusion Matrix of the Presented Model.

The confusion matrix in **Fig 7** offers a detailed visualization of the predictive performance of the proposed CNN-based Small Inception model on the test dataset, which was balanced across healthy and unhealthy leaf patches. The matrix reveals powerful classification capability, with True Positives (TP) and True Negatives (TN) dominating the grid. Specifically, a substantial proportion of unhealthy patches were correctly identified as unhealthy (TP), and likewise, the model demonstrated high fidelity in detecting healthy patches (TN). The count of False Positives (FP)—healthy samples misclassified as unhealthy—and False Negatives (FN)—unhealthy samples mistaken as healthy—remains minimal. These accurate classifications highlight the model's resilience and most minor diagnostic errors. The confusion matrix also demonstrates the model's accuracy in detecting the presence of illness, which is essential in agricultural diagnostics because false negatives can cause severe crop loss and unnoticed disease spread.

The confusion matrix confirms that the proposed CNN achieves near-optimal classification across categories, reflecting its strong generalization and discriminative power when applied to real-world plant pathology data.

Comparative Analysis

To thoroughly evaluate the effectiveness of the proposed CNN-based Small Inception model, we conducted a comparative analysis against several state-of-the-art architectures on the widely used PlantVillage dataset presented in **Table 7**.

Table 7. Compare Our Proposed Model with Existing Works

Dataset	Model	Accuracy	Precision	Recall	F1- score
Plant Village	LeafDiseaseNet[8]	97.68	97.50	97.20	97.30
	MXception[20]	98.65	98.27	98.04	98.63
	CNN+MLP[21]	95.06	95.00	95.00	95.00
	CBSNet[21]	92.04	91.58	90.24	90.71
	Proposed Model	98.45	98.52	98.38	98.62

LeafDiseaseNet [20] reported an accuracy of 97.68%, supported by balanced precision (97.50%) and recall (97.20%). While this model performs well, its design primarily focuses on disease classification with relatively limited adaptability to patch-based or fine-grained feature variations, which are crucial for early-stage disease identification. MXception [21], an extension of the Xception architecture adapted for agricultural applications, achieved a higher accuracy of 98.65%. Its depthwise separable convolutions improve parameter efficiency, contributing to a strong F1-score of 98.63%. However, despite its remarkable results, MXception’s complexity and computational requirements make it less suitable for real-time or embedded applications, especially in resource-constrained agricultural environments. Though simple and computationally light, the CNN+MLP hybrid model [21] lagged with an accuracy of 95.06%. Its low capacity to generalize across many crop-disease scenarios was caused by its flat design, which lacks deeper hierarchical feature extraction. Similarly, CBSNet [21] fared poorly while being designed for crop health monitoring, attaining an accuracy of only 92.04%. Its relatively low F1-score (90.71%) and recall (90.24%) suggest difficulties in reliably detecting sick patches, which an excessive dependence on handmade characteristics and inadequate spatial encoding might cause.

Our proposed model demonstrates a compelling balance between accuracy (98.45%) and computational efficiency. Leveraging a Small Inception architecture effectively captures localized and abstract features within patch-level leaf images. With a precision of 98.52% and a recall of 98.38%, the model maintains exceptional consistency across all metrics, resulting in a robust F1 score of 98.62%, as illustrated in **Fig 8**. In contrast to deeper networks like ResNet or MXception, our architecture was developed with optimal generalization and inference speed in mind, guaranteeing excellent performance without needless parameter overhead or overfitting.

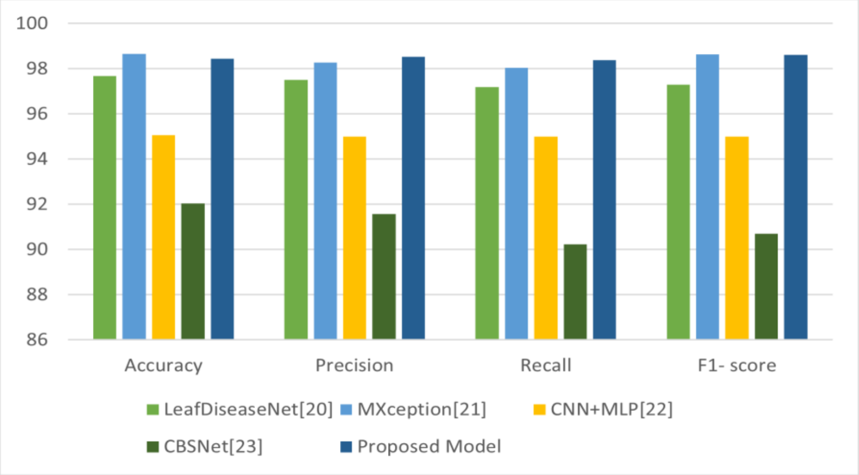

Fig 8. Comparison with Existing Models.

V. CONCLUSION

This study introduces a robust, generalized approach for plant disease detection across multiple crop types, grounded in constructing an updated version of the widely used Plant Village dataset. Traditional disease classification methods often suffer from limited generalizability, as they tend to be crop-specific or reliant on large quantities of disease-specific labeled data. To address this limitation, we reformulated the detection problem by shifting focus from identifying individual disease types to distinguishing between healthy and unhealthy leaf regions—regardless of the crop or pathogen. We selected the Small Inception CNN architecture due to its proven balance between computational efficiency and high representational power, particularly in scenarios involving small patch-based inputs. Its modular structure effectively captures localized features of healthy and diseased tissue across various crops. This design choice was guided by the need for a scalable and generalizable model to perform accurately across heterogeneous agricultural data. The proposed model demonstrated outstanding capability in learning cross-disease and cross-crop representations by systematically curating a new binary-labeled dataset from the PlantVillage corpus and training on visually annotated leaf patches. This foundation enables more inclusive and universal plant health monitoring systems less dependent on specific disease annotations.

Future directions include refining the detection pipeline to incorporate multi-stage learning for disease identification after infection detection, extending the framework to accommodate real-world, noisy field data, and deploying the model in edge-based environments for real-time, in-field diagnostics. This generalized approach lays the groundwork for more adaptable and intelligent agricultural systems, supporting farmers and agronomists in disease prevention and yield protection on a global scale.

CRediT Author Statement

The authors confirm contribution to the paper as follows:

Conceptualization: Roopa R, Rajesh Lingam, Suresh A, Penubaka Balaji and Avanija J; **Methodology:** Roopa R, Rajesh Lingam and Santosh Kumar Ravva; **Software:** Penubaka Balaji and Avanija J; **Writing- Original Draft Preparation:** Roopa R, Rajesh Lingam and Santosh Kumar Ravva; **Investigation:** Roopa R, Rajesh Lingam and Santosh Kumar Ravva; **Supervision:** Santosh Kumar Ravva and Suresh A; **Validation:** Suresh A, Penubaka Balaji and Avanija J; **Writing-Reviewing and Editing:** Roopa R, Rajesh Lingam, Santosh Kumar Ravva, Suresh A, Penubaka Balaji and Avanija J; All authors reviewed the results and approved the final version of the manuscript.

Data Availability

No data was used to support this study.

Conflicts of Interests

The author(s) declare(s) that they have no conflicts of interest.

Funding

No funding agency is associated with this research.

Competing Interests

There are no competing interests

References

[1]. K. Dutta, D. Talukdar, and S. S. Bora, “Segmentation of unhealthy leaves in cruciferous crops for early disease detection using vegetative indices and Otsu thresholding of aerial images,” Measurement, vol. 189, p. 110478, Feb. 2022, doi: 10.1016/j.measurement.2021.110478.

[2]. L. Li, S. Zhang, and B. Wang, “Plant Disease Detection and Classification by Deep Learning—A Review,” IEEE Access, vol. 9, pp. 56683–56698, 2021, doi: 10.1109/access.2021.3069646.

[3]. Ramcharan, K. Baranowski, P. McCloskey, B. Ahmed, J. Legg, and D. P. Hughes, “Deep Learning for Image-Based Cassava Disease Detection,” Frontiers in Plant Science, vol. 8, Oct. 2017, doi: 10.3389/fpls.2017.01852.

[4]. Attri, L. K. Awasthi, T. P. Sharma, and P. Rathee, “A review of deep learning techniques used in agriculture,” Ecological Informatics, vol. 77, p. 102217, Nov. 2023, doi: 10.1016/j.ecoinf.2023.102217.

[5]. M. Shoaib et al., “An advanced deep learning models-based plant disease detection: A review of recent research,” Front. Plant Sci., vol. 14, p. 1158933, Mar. 2023, doi: 10.3389/fpls.2023.1158933.

[6]. B. Liu, Y. Zhang, D. He, and Y. Li, “Identification of Apple Leaf Diseases Based on Deep Convolutional Neural Networks,” Symmetry, vol. 10, no. 1, p. 11, Dec. 2017, doi: 10.3390/sym10010011.

[7]. Y. Toda and F. Okura, “How Convolutional Neural Networks Diagnose Plant Disease,” Plant Phenomics, vol. 2019, pp. 1–14, Mar. 2019, doi: 10.1155/2019/9237136.

[8]. I. Bouacida, B. Farou, L. Djakhdjakha, H. Seridi, and M. Kurulay, “Innovative deep learning approach for cross-crop plant disease detection: A generalized method for identifying unhealthy leaves,” Information Processing in Agriculture, vol. 12, no. 1, pp. 54–67, Mar. 2025, doi: 10.1016/j.inpa.2024.03.002.

[9]. A. Natarajan, “Adoption of Recommended Expert Systems in Agriculture,” Journal of Smart and Sustainable Farming, pp. 1–11, Jan. 2025, doi: 10.64026/jssf/2025001.

[10]. B. V. Baiju, N. Kirupanithi, S. Srinivasan, A. Kapoor, S. K. Mathivanan, and M. A. Shah, “Robust CRW crops leaf disease detection and classification in agriculture using hybrid deep learning models,” Plant Methods, vol. 21, no. 1, Feb. 2025, doi: 10.1186/s13007-025-01332-5.

[11]. S. Kalaivani, C. Tharini, T. M. S. Viswa, K. Z. F. Sara, and S. T. Abinaya, “ResNet-Based Classification for Leaf Disease Detection,” J. Inst. Eng. India Ser. B, vol. 106, no. 1, pp. 1–14, Feb. 2025, doi: 10.1007/s40031-024-01062-7.

[12]. S. Aboelenin, F. A. Elbasheer, M. M. Eltoukhy, W. M. El-Hady, and K. M. Hosny, “A hybrid Framework for plant leaf disease detection and classification using convolutional neural networks and vision transformer,” Complex Intell. Syst., vol. 11, no. 2, p. 142, Feb. 2025, doi: 10.1007/s40747-024-01764-x.

[13]. S. Duhan, P. Gulia, N. S. Gill, and E. Narwal, “RTR_Lite_MobileNetV2: A lightweight and efficient model for plant disease detection and classification,” Current Plant Biology, vol. 42, p. 100459, Jun. 2025, doi: 10.1016/j.cpb.2025.100459.

[14]. T. Varma, P. Mate, N. A. Azeem, S. Sharma, and B. Singh, “Automatic mango leaf disease detection using different transfer learning models,” Multimedia Tools and Applications, May 2024, doi: 10.1007/s11042-024-19265-x.

[15]. I. Kaushik, N. Prakash, and A. Jain, “Plant disease detection using a depth-wise separable-based adaptive deep neural network,” Multimed Tools Appl, vol. 84, no. 2, pp. 887–915, Apr. 2024, doi: 10.1007/s11042-024-19047-5.

[16]. I. Pacal and G. Işık, “Utilizing convolutional neural networks and vision transformers for precise corn leaf disease identification,” Neural Computing and Applications, vol. 37, no. 4, pp. 2479–2496, Dec. 2024, doi: 10.1007/s00521-024-10769-z.

[17]. A. Y. Ashurov et al., “Enhancing plant disease detection through deep learning: a Depthwise CNN with squeeze and excitation integration and residual skip connections,” Frontiers in Plant Science, vol. 15, Jan. 2025, doi: 10.3389/fpls.2024.1505857.

[18]. A. Bhola and P. Kumar, “Deep feature-support vector machine-based hybrid model for multi-crop leaf disease identification in Corn, Rice, and Wheat,” Multimedia Tools and Applications, vol. 84, no. 8, pp. 4751–4771, Mar. 2024, doi: 10.1007/s11042-024-18733-8.

[19]. W. Shafik, A. Tufail, C. Liyanage De Silva, and R. A. Awg Haji Mohd Apong, “A novel hybrid inception-xception convolutional neural network for efficient plant disease classification and detection,” Scientific Reports, vol. 15, no. 1, Jan. 2025, doi: 10.1038/s41598-024-82857-y.

[20]. S. P. Mohanty, D. P. Hughes, and M. Salathé, “Using Deep Learning for Image-Based Plant Disease Detection,” Frontiers in Plant Science, vol. 7, Sep. 2016, doi: 10.3389/fpls.2016.01419.

[21]. M. Dai et al., “Pepper leaf disease recognition based on enhanced lightweight convolutional neural networks,” Frontiers in Plant Science, vol. 14, Aug. 2023, doi: 10.3389/fpls.2023.1230886.