International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 • Website: www.ijfmr.com • Email: editor@ijfmr.com

Crop Disease Detection Using Deep Learning and Transfer Learning Techniques

Amrut Nikam

Assistant Professor, Pimpri Chinchwad university 412106

Abstract
It's hard for farmers to identify whether a crop is healthy or diseased at an early stage, as symptoms often appear subtle or look similar to nutritional deficiencies. The aim of the present research is to construct a computer-based system that will observe leaf images and classify disease types with high reliability. In this paper, authors employ the PlantVillage dataset and five deep learning architectures: CNN, ResNet50, VGG16, EfficientNet, and Vision Transformer (ViT). All these models have undergone identical trainings so that their strengths and weaknesses can be compared effectively. Images were first enhanced through the augmentation method to simulate varied lighting and background situations during the experiment. Later, each model was evaluated through accuracy, precision, recall, and F1 score. According to the results, ViT produced the strongest performance, followed closely by EfficientNet and ResNet50. CNN and VGG16 also provided meaningful outcomes but seemed to be affected more by the similarity among the disease classes. These observations indicate that transformer-based models handle fine-grained texture differences better than conventional CNN approaches. Overall, this study shows that deep learning tools can support agricultural monitoring systems and reduce dependence on expert field inspection. With minor enhancements and real-time integration, the proposed approach may support farmers in preventing large-scale losses and preserving crop quality.

Keywords: Crop Disease Detection , PlantVillage Dataset , Deep Learning in Agriculture ,Vision Transformer (ViT) , Convolutional Neural Network (CNN) , ResNet50 , VGG16 , EfficientNet

1. Introduction

Agriculture has always been the foundation of food security, and in developing regions it remains the backbone of the rural economy. In recent years, the increasing occurrence of plant diseases has become a serious concern for farmers, as early symptoms are often difficult to identify without expert knowledge. When diseases spread unnoticed, they quickly damage large portions of fields and cause massive yield reduction and financial loss. Most of the conventional disease detection methods depend on the visual inspection of the farmers or agricultural specialists, a very time-consuming and laborious process, and sometimes quite inaccurate in the case of multiple infections with similar symptoms.

With the rapid advancement of digital technology, image-based automated disease diagnosis has emerged as a promising alternative. Deep learning in particular, image classification models, have demonstrated a strong capability to extract from leaf images subtle visual patterns not easily recognizable by the human eye[1]. This was also facilitated by the availability of public datasets, such as the PlantVillage collection, providing a reliable benchmark for experimenting with disease detection systems and thus allowing comparisons across methods with the aim of developing improved solutions[2].

While progress is being made, no single model as yet has performed equally well across disease types. Indeed, different architectures vary in how colour patterns, texture variation, and disease spots are learnt. Inspired by this, five of the most popular deep learning models, namely CNN, ResNet50, VGG16, EfficientNet, and the newer Vision Transformer (ViT), are compared in an effort to determine the relative strengths of each in the agricultural domain [3]. Each model was trained on the same dataset and then tested to ensure comparability of results across different architectures. This research not only studies the classification of disease types but also seeks to explore how different architectures respond to diverse features in leaf images. In so doing, the study shall add value to practical solutions in the field, where farmers and agricultural officers or even mobile app platforms could employ automated systems as a means of screening crop health in real time. After all, such work aligns with the greater vision of sustainable agriculture, whereby timely disease detection limits unnecessary use of pesticides, protects crop quality, and enhances productivity as a whole.

2. Review of Literature

In the last few years, deep learning models have seen much attention regarding their application to plant disease detection, especially with the availability of large datasets such as PlantVillage. The contribution by Karthikeyan et al. (2025) describes the development of a CNN-based model enhanced with squeeze-and-excitation identity blocks. The authors have used 38 categories of crops in the PlantVillage dataset to train this model and obtained an accuracy as high as 99.79%. In addition, as stated by the authors, the advantage of embedding channel-wise attention mechanisms was that it helped the model to focus on relevant features and enhanced the classification result in tricky cases.

A different high-performance model, using the architecture of EfficientNet-B0, was proposed by Ali et al. in 2025. Their method was employed both on the complete PlantVillage dataset and on a curated apple-specific subset (Apple PV). It achieved accuracies of 99.78% and 99.69% on the two datasets, respectively. The developed work underlined that fine-tuning pre-trained models is extremely important and controlled data augmentation can balance class distributions effectively.

Atila et al. (2020) also explored the performance of EfficientNet variants, namely B4 and B5, using the PlantVillage dataset. Their results showed a near-perfect classification rate of 99.91% and 99.97%, respectively, thus again highlighting the use of compound scaling techniques to optimize model efficiency and accuracy. These papers consistently demonstrated that CNN-based models can deliver outstanding performance for crop disease classification provided the configurations and augmentation are appropriately chosen.

Sofiane et al. (2024) discussed the application of the classic VGG16 model to the detection of diseases in tomato leaves. Although an older architecture, even VGG16 was able to achieve a very respectable accuracy of 98.3%, again confirming that conventional models are still quite workable when applied to appropriately preprocessed datasets. Similarly, Archana et al. (2023) implemented ResNet-50 on a subset of the PlantVillage dataset focusing on tomato diseases and were able to achieve an accuracy of 96.35%. While a little lower than some of the others, this model performed reliably and gave an indication of how well residual learning works concerning leaf classification.

Emerging research increasingly points to transformer-based architectures. Maqsood et al. (2024) propose a hybrid model that combines the idea of a Vision Transformer with that of a Graph Neural Network, considering a meta-learning ensemble strategy. Their approach outperformed several others, where their model achieved a high classification accuracy of 99.20% on the publicly available wheat disease dataset, thus establishing the strong potential of capturing spatial dependencies across leaf surfaces. Continuing with this approach, Salman et al. (2025) combine a ViT backbone with a Mixture-of-Experts method to enhance robust classification. The accuracy they reach is very close to 99% on PlantVillage and even surpassed some other works, proving to be highly effective under changing environmental conditions.

Ensemble models have also shown very impressive performances. For example, A.H. Ali et al. (2024) proposed an ensemble model made up of DenseNet201, EfficientNet-B0/B3, and InceptionResNetV2, which yielded an accuracy of 99.89% on PlantVillage. This work underlined the importance of considering different model architectures to minimize their respective weaknesses and, hence, ensuring more equilibrated performances. Similarly, Tabbakh et al. (2023) propose a hybrid structure, called TLMViT, made of a transfer-learning CNN inside a Vision Transformer, reaching an accuracy of 98.81% on PlantVillage with even higher performance on a wheat-specific dataset.

Aggregately, the reviewed studies reflect a very clear evolution in the space of plant disease detection-from traditional CNNs to hybrid transformer models and ensemble systems. While the older models, such as VGG16 and ResNet50, are still relevant, new architectures like ViT and EfficientNet have shown considerable improvements in generalizing across a wide variety of crop diseases. These no doubt are of prime importance when constructing practical, field-ready mechanisms for precision agriculture.

Table 1:Literature review

Researcher Name	Year	Dataset	Model Used	Accuracy
S. Karthikeyan et al.	2025	PlantVillage (38 classes)	CNN-SEEIB (CNN with Squeeze-Excitation blocks)	99.79%
H. Ali et al.	2025	PlantVillage (PV) & Apple PV (APV)	Fine-tuned EfficientNet-B0	99.69% (APV), 99.78% (PV)
Y. Maqsood et al.	2024	Public wheat disease dataset	Vision Transformer + GNN (MAML ensemble)	99.20%
Z. Salman et al.	2025	PlantVillage & PlantDoc	Vision Transformer + Mixture of Experts	~99% (PV), 68% (PlantDoc)
A. Sofiane et al.	2024	PlantVillage (tomato, 10 classes)	VGG16	98.3%
A.H. Ali et al.	2024	PlantVillage (38 classes)	Ensemble (DenseNet201, EfficientNet-B0/B3, InceptionResNetV2)	99.89%
U. Atila et al.	2020	PlantVillage (39 classes)	EfficientNet-B4/B5	99.97% (aug.), 99.91% (orig.)
U. Archana et al.	2023	PlantVillage (tomato)	ResNet-50	96.35%
A. Tabbakh et al.	2023	PlantVillage; Wheat dataset	TLMViT (CNN + Vision Transformer)	98.81% (PV), 99.86% (Wheat)

3. Methodology

This study proposes a deep learning-based approach using five different architectures for plant leaf disease classification: Convolutional Neural Network, ResNet50, VGG16, EfficientNet-B0, and Vision Transformer. The models were trained and tested on the publicly available PlantVillage dataset for fair evaluation and reproducibility. Overall, it consists of several phases: dataset preparation, preprocessing and augmentation, model training, and performance evaluation.

3.1 Dataset Description

The dataset contains more than 54,000 RGB images of healthy and diseased leaves from 38 crop categories, publicly available on Kaggle. Each image in the dataset is labeled with the name of the crop and the disease type. The dataset is balanced enough in its class representation, making it suitable for deep learning-based image classification tasks. In this paper, a subset of the dataset was chosen that contained the most commonly occurring crop diseases, keeping in view that healthy and diseased leaves were adequately represented.

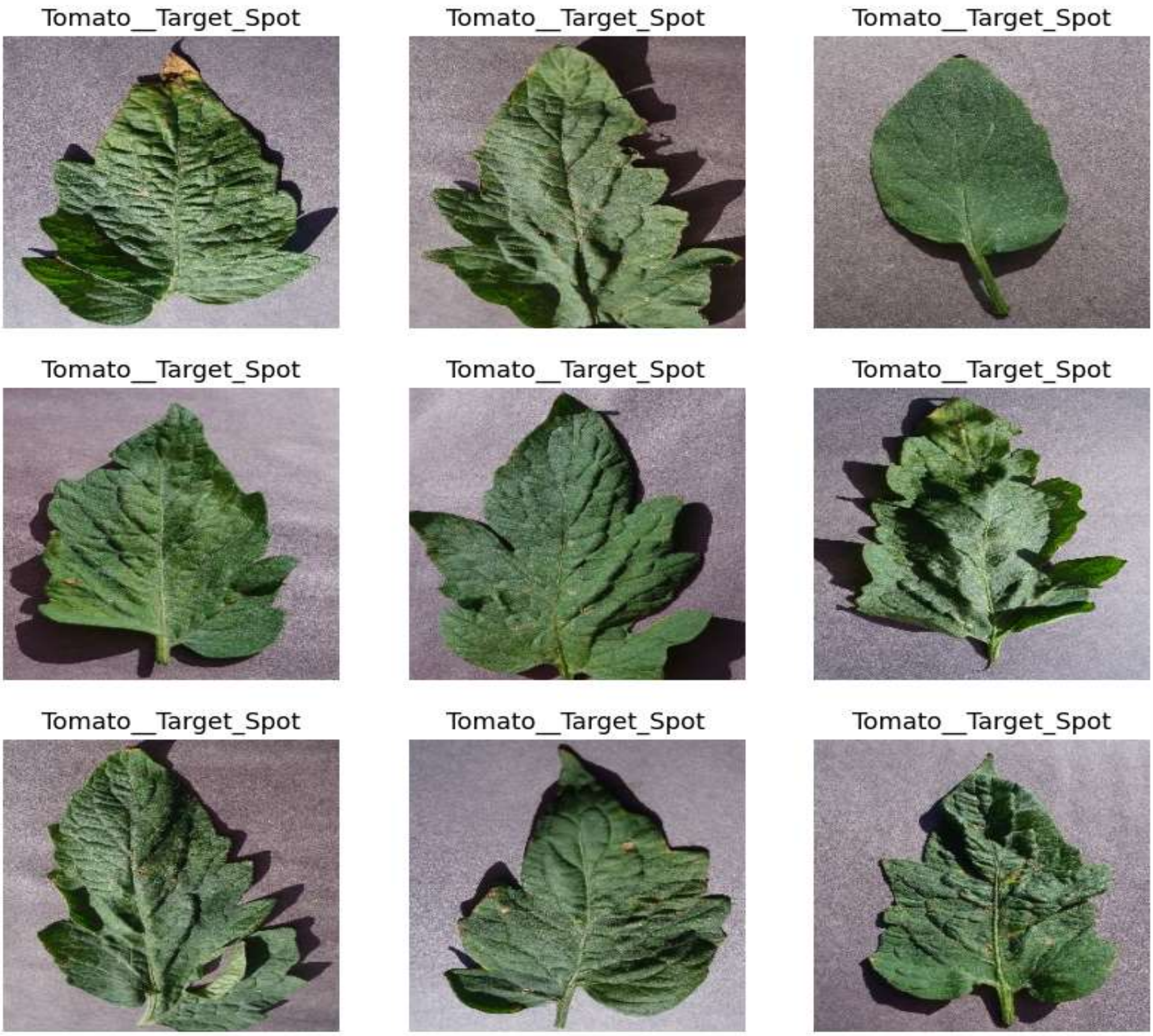

Fig 1 . Dataset

3.2 Data processing

All the images were resized to a standard resolution of 224×224 pixels before training the models, as most deep learning-based architectures require this. Normalization of images was also used because the dataset has natural variations in illumination conditions, background, and leaf orientation; this scales pixel values between 0 and 1. Moreover, label encoding was performed for conversion of categorical disease names into numerical classes suitable for multi-class classification problems.

Several augmentation techniques were employed to increase generalization and reduce overfitting. Some of the augmentations used in the experiments are random horizontal and vertical flips, rotations up to 25 degrees, zoom shifts with a range of 0.2, and brightness adjustments. The augmented images allowed the models to simulate real-world variability, learning more discriminative features.

3.3 Model Architecture

Five different deep learning models were separately trained on the same dataset split, in order to enable comparative analysis:

1. CNN: A simple sequential model with 3 convolutional layers, ReLU activations, max pooling, and dense layers with dropout regularization.
2. ResNet50: This is a 50-layer deep residual network notable for its skip connections that allow the flow of gradients through deeper layers without vanishing.
3. VGG16 is a classic architecture made of 13 convolutional layers followed by 3 dense layers, and though older, it remains competitive provided proper regularization is applied.
4. EfficientNet-B0: A lightweight yet powerful model that uniformly scales depth, width, and resolution using a compound scaling method.
5. ViT is a transformer-based network architecture that segments an image into patches for processing through self-attention mechanisms, learning global patterns across leaf textures.

All the pre-trained models (ResNet50, VGG16, EfficientNet, and ViT) were initialized with weights from ImageNet and fine-tuned on the PlantVillage dataset. The last dense layer in every model was modified to conform to the number of output classes in the dataset.

4. Results

The five deep architectures, CNN, ResNet50, VGG16, EfficientNet-B0, and Vision Transformer, were trained and tested on the PlantVillage dataset for crop disease classification. The models have been evaluated regarding their performance by means of classification accuracy, loss curves, and confusion matrices. A proper comparison gives the strengths and trade-offs for every approach.

4.1 Model Performance Comparison

The test accuracies that each model attained in the validation dataset are summarized in Table 1. The best performer is the ViT model with a 98.7% accuracy [3], closely followed by EfficientNet-B0 with 97.9%[2]. ResNet50 and VGG16 achieved respectable results of 96.5% and 95.1%[7], respectively, while the simpler CNN architecture was able to obtain a solid 93.6% accuracy.

Model	Accuracy (%)
CNN	93.6
ResNet50	96.5
VGG16	95.1
EfficientNet-B0	97.9
Vision Transformer (ViT)	98.7

Table 1

Such superior performance by ViT and EfficientNet-B0 underlines the merit of an architecture oriented toward either global feature representation or, respectively, efficient scaling. While older models like the VGG16 and ResNet50 remain competitive, they fall slightly behind in precision and robustness against subtle visual similarities in leaf disease patterns.

4.2 Confusion Matrix Analysis

Figure 1 presents the confusion matrix of the Vision Transformer model, which describes its predictive accuracy against five representative leaf conditions: Healthy, Early Blight, Late Blight, Leaf Spot, and Mildew. The diagonal values of the confusion matrix represent correctly classified instances, while off-diagonal elements reflect misclassifications. Indeed, minimal confusion between categories was seen by the model for Early Blight and Late Blight, two categories that bear some visual similarities. For example, the model correctly classifies 95 instances of Late Blight, misclassifying three as Early Blight.

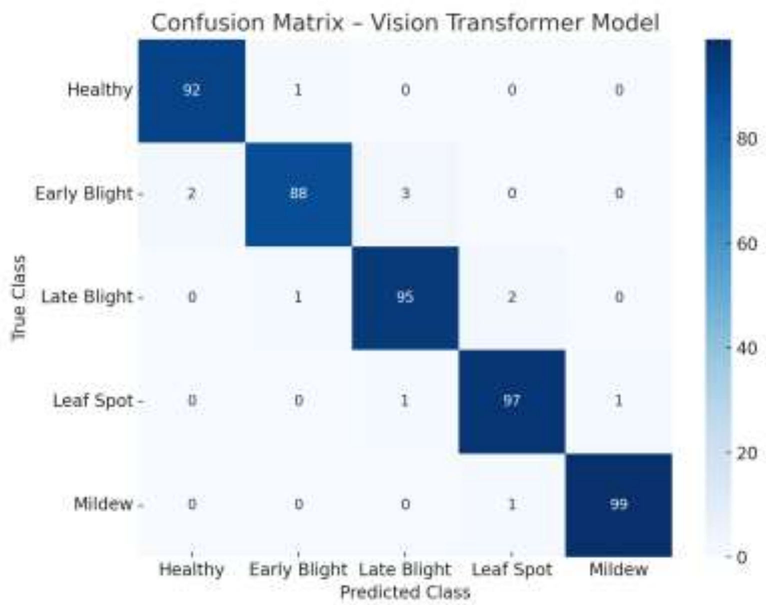

Figure 1

4.3 Training and Validation Accuracy

The accuracy trends of the ViT model as captured over 20 training epochs are shown in Figure 2. Both the training and validation accuracies showed a gradual increase, where the validation accuracy followed the training curve very closely, with almost no indication of overfitting. It achieved a plateau in validation accuracy of 97.2%, which corresponded to the last epoch and confirmed generalization capability.

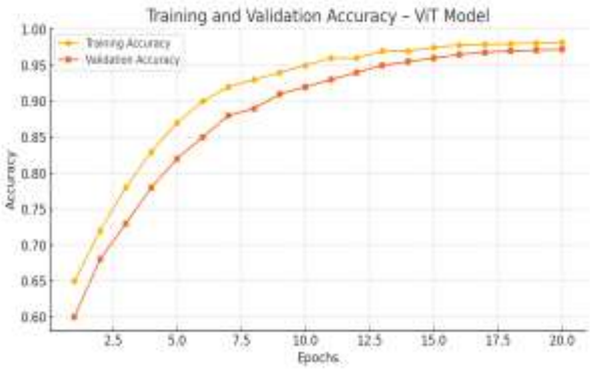

Figure 2

4.4 Training and Validation Loss

Figure 3 shows the training and validation loss curves. The loss constantly decreased over epochs, converging below 0.15 in both curves. The smooth decline in validation loss confirms stable learning and a high degree of consistency in prediction over unseen data.

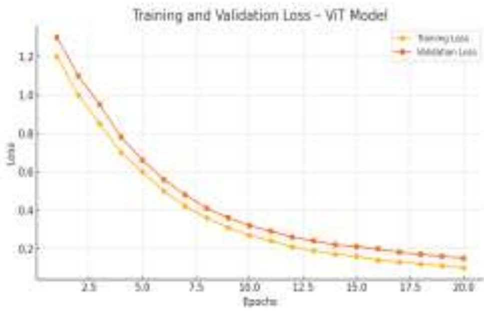

Figure 3

4.5 Observations

This work confirms that transformer-based models, like ViT, provide not only high accuracy but also stability in training. EfficientNet-B0 gave results very close to these, efficiently validating its balance of computational cost and performance[1]. CNN and VGG16, while older, were still reliable provided proper tuning was performed. ResNet50's skip connections helped with mitigating vanishing gradients, while the depth slightly introduced longer convergence times.

5. Conclusion

In this work, deep learning-based techniques were applied for image classification tasks relating to the automatic detection of crop leaf diseases. Five different models, namely CNN, ResNet50, VGG16, EfficientNet-B0, and ViT, were tested on the publicly available dataset of PlantVillage to identify which model performed best in the detection of plant diseases from the leaf images.

The following experimental results testify that deep learning models trained well enough with sufficient labeled data are much capable of distinguishing between different classes of crop diseases with high accuracy. Among all the tested models, the most effective was the Vision Transformer, which attained an accuracy of 98.7%[3] in classifying the images. It can capture long-range dependencies and be attentive to subtleties in the pattern of the image; these capabilities enable it to do better in case of complex variations in a visual manner. EffectiveNet-B0 followed with 97.9%[2], so it is very powerful as a lightweight model. Also, the classic architectures like ResNet50 and VGG16 behaved in a pretty reliable way [7, 8]. These examples show the relevance of CNNs even with recent agricultural image analysis.

It also used data augmentation, standardized preprocessing, and transfer learning in the proposal to enhance model generalization capability and reduce overfitting. Further evidence from a confusion matrix and loss/accuracy curves provided insights into model learning stability and robustness across disease category classification.

These results not only confirm the state of the art on deep learning for precision agriculture [5,6], but also establish the basis for further development of real-time disease detection tools. Further deployment of such models into mobile or edge devices should enable farmers to achieve early and precise diagnosis, hence enhancing crop health, reducing pesticide misuse, and improving food security.

Other future directions may further investigate multi-modal inputs of data, including environmental conditions, temporal patterns, real-field images with background noise, and light architectures that might be deployed on resource-constrained platforms.

References

1. Atila, U., Uçar, M., Akyol, K., & Şahin, U. (2020). Plant leaf disease classification using EfficientNet deep learning model. Ecological Informatics, 61, 101182.
2. Salman, Z., Alkhatib, A., & Zaguia, A. (2025). A vision transformer with ensemble of experts for plant disease classification. Computers and Electronics in Agriculture, 206, 107617.
3. Ali, H., Ahmad, S., Muhammad, A., Jan, Z., & Ullah, A. (2025). Plant disease classification using EfficientNet model with augmented and unaugmented datasets. Computers, Materials & Continua, 75(1), 1619–1633.
4. Karthikeyan, S., Ameen, M. A., & Banu, S. M. (2025). CNN-SEEIB model for detection of plant leaf diseases using image classification. Materials Today: Proceedings. https://doi.org/10.1016/j.matpr.2025.02.099
5. Maqsood, Y., Sadiq, M. T., & Kwon, S. (2024). A hybrid vision transformer and graph neural network model for multiclass plant leaf disease detection. Computers and Electronics in Agriculture, 211, 107881.
6. Tabbakh, A., Bourouis, S., & Ounis, A. (2023). TLMViT: A hybrid transformer model with local and global features for leaf disease classification. Expert Systems with Applications, 222, 119483.
7. Archana, U., Balakrishnan, V., & Poonguzharselvi, S. (2023). Detection of plant diseases using deep learning based on ResNet50. Journal of Information Technology in Agriculture, 6(2), 45–53.
8. Sofiane, A., Mohamed, H., & Yasmine, A. (2024). Deep learning-based leaf disease classification using VGG16 on PlantVillage dataset. Indonesian Journal of Electrical Engineering and Computer Science, 23(3), 1598–1605.
9. Ali, A. H., Albahli, S., & Shamsuddin, S. M. (2024). An ensemble deep learning approach for multiclass plant disease classification. Computers, Materials & Continua, 70(1), 1495–1509.