Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 01 , January - 2026

A Survey on Deep Learning Techniques for Plant Leaf Disease Detection

Pramod Mishra
Student (M. Tech),
Eshan college of engineering, Farah, Mathura, Uttar Pradesh

Pawan Yadav
Assistant Professor, CSE Department.
Eshan college of engineering, Farah, Mathura, Uttar Pradesh

Abstract: Animal diseases of the plant leaf have a serious impact on the global food security and farm output. These diseases require early and proper detection in order to manage the crops. Deep learning (DL) and computer vision methods have become effective methods of automated plant disease detection on leaf images in recent years. The paper is an in-depth literature review of the deep learning-based methods used in the detection of plant leaf disease, especially Vision Transformer (ViT)-based models and their combination with Convolutional Neural Networks (CNNs). Different state-of-the-art models, such as CNN, DCNN, YOLO, RSNSR-LDD, ensemble CNN-ViT architecture and attention-based transformers, are discussed systematically. The publicly available datasets, metrics of evaluation, trends of performance, research gaps, limitations, and future research directions are addressed. The review points out that the transformer based and hybrid deep learning models are always more effective in comparison to the traditional CNNs particularly when dealing with high dimensional and complex image data and therefore potential solutions to smart and precision agriculture.

Keywords: Convolutional Neural Network, Deep learning, Plant leaf disease detection, Smart agriculture Vision Transformer,

1. INTRODUCTION

Agriculture is the main source of most developing economies and is an essential source of food security the world over. Nonetheless, the disease of plants, especially of leaves, a disease caused by fungi, bacteria, viruses, and pests, is a great menace to crop productivity and quality. The world has been estimated to lose about 10-16 percent of the crop yearly due to plant pests and diseases causing a huge economic loss. Having a previous experience and visual inspection of the disease are very subjective, time consuming, and prone to errors, which are traditionally used to carry out disease diagnosis.

The high rate of development of artificial intelligence (AI), especially deep learning, has transformed the process of automated plant diseases diagnostics. The use of convolutional neural networks (CNNs) on image-based disease detection has been popular because they are very effective in feature extraction. Nevertheless, CNN-based methods are usually limited by factors like a lack of generalization, sensitivity to background noise, data imbalance and long-range dependencies modeling in images.

ViTs are image classification and object detector models inspired by the introduction of transformers in natural language processing systems. ViTs have self-attention to get global contextual information, which allows it to be better represented by features than traditional CNNs. ViTs have demonstrated good outcomes in detecting diseased areas on plant leaves with high precision in agricultural applications. However, studies of pathological models of transformers in agronomy are not extensive.

The proposed review will fill this gap by conducting a systematic analysis of the existing deep learning-based methods of detecting plant leaf diseases, particularly Vision Transformer-based models. The research paper links the trends in research conducted in the past, currently, and in the future, determines the weaknesses of the current methods, and presents future research in smart agriculture (Gupta and Kumar Pal, 2025).

2. METHODOLOGY OF LITERATURE REVIEW

The systematic literature review was carried out to find the appropriate research articles on the topic of plant leaf disease detection with the help of deep learning techniques. Peer-reviewed journal articles, conference papers, and review studies that were published in the period between 2019 and 2024 were taken into consideration. The keywords were also used to search databases like IEEE Xplore, ScienceDirect, SpringerLink, MDPI, and Google Scholar, which included plant leaf disease detection, deep learning, CNN, Vision Transformer, and smart agriculture.

3. DEEP LEARNING TECHNIQUES FOR PLANT LEAF DISEASE DETECTION

3.1 Convolutional Neural Networks (CNNs)

The most popular deep learning model to use in the detection of plant leaf diseases is the Convolutional Neural Networks (CNN) as a result of high ability to automatically learn about hierarchical and discriminative features of raw image data. The CNNs are

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 01 , January - 2026

composed of the convolutional layers, pooling layers, and fully connected layers, which combine to form spatial features and classification. Initial research in plant disease identification was largely based on manual characteristics like color, texture, and shape but CNNs do not require manual features engineering; they learn features themselves.

A number of established CNN structures have been used with success with regard to plant leaf disease classification. VGG networks are characterized by simplicity, depth, which makes them effective in extracting the features but at the expense of high computation complexity. The ResNet architectures added the residual connections as a solution to the vanishing gradient problem to enable deeper networks to be effectively trained. MobileNet and EfficientNet are lightweight CNN models with a smaller size of parameters and suitable to mobile and edge devices, which makes them the best choice in smart agriculture applications. DenseNet goes further to improve feature reuse by interconnecting layers with all its other layers, which results in an optimal flow of gradients and learning (Bektaş, 2024).

The CNN-based models have provided high accuracy in benchmark datasets like PlantVillage, Tomato Leaf Disease Dataset and Cassava Leaf Disease Dataset. Nevertheless, even being successful, CNNs have a number of limitations. They usually find it hard to cope with complicated and messy backgrounds, differences in lighting, occlusion, and swiveling of leaves. Moreover, CNNs require the presence of massive annotated data to be effective, which is hard to acquire in the actual farming conditions. A second significant weakness is that they have a localized receptive field, which inhibits the choice of long-range dependencies and global contextual relations among leaf images (Kamilaris and Prenafeta-Boldú, 2018).

3.2 Vision Transformer (ViT)

Vision Transformers (ViTs) are a new generation in detection of plant diseases based on images since the convolutional operations are substituted with self-attention mechanics. ViTs are based on models of transformers in natural language processing but apply the transformation to an input image consisting of fixed-size patches which are flattened and then projected to a sequence of embeddings. These embeddings are then fed through transformer encoder layers which employ multi-head self-attention to obtain relationships between image patches.

The main benefit of ViTs is that they can capture global contextual data and long-range correlations that tend to be essential in determining patterns of diseases that are actually spread out in leaf surfaces. In contrast to CNNs, which pay attention to the local characteristics of the image, ViTs perceive the entire image as a whole, which makes it possible to better distinguish between normal and diseased areas (Giakoumoglou, Pechlivani and Tzovaras, 2023).

Recent research has shown that ViT-based models are 1-3 percent better than traditional CNNs in their accuracy to classify different crops including cassava, rice, tomato, apple, and grape leaves. In addition, ViTs give higher interpretability in terms of attention maps that enable researchers and farmers to see the areas on the leaf that give the highest contributions to disease prediction. This is a very useful attribute in precision agriculture and decision support systems.

Nonetheless, ViTs also have a problem. They generally need huge datasets to be trained successfully and are computationally more burdensome than CNNs. ViTs can perform worse as they do not learn to focus on patterns of attention when trained on small datasets. In order to solve these problems, scientists frequently use transfer learning, data augmentation, and hybrid architectures (Murugavalli and Gopi, 2025).

3.3 Hybrid CNN–ViT Models

Hybrid CNN-ViT models are proposed to address the weaknesses of each of the two approaches by integrating the advantages of both convolutional neural networks and vision transformers. In such architectures, CNNs are usually applied to low-level and mid-level feature extraction, to detect local spatial patterns (edges, textures, lesion boundaries, etc.). These features are removed and added to ViT modules that encode global contextual relationship based on self-attention.

PlantXViT, SLViT and ensemble CNN ViT models have demonstrated better performance than single CNN or ViT models. These hybrid designs are more accurate and yet they are computationally efficient and thus is suitable to be deployed in Internet of Things (IoT)-based smart agriculture systems and edge devices. Hybrid models have much lower training time and memory needs by minimizing the number of transformer parameters, and by using CNN-based feature extraction.

The other strength of the hybrid models is that they are effective in real field scenarios where the background clutter, contrasting light, and partial blockage are prevalent. The CNN component has proven to be effective in noise and irrelevant information filtering

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 01 , January - 2026

where the transformer component is effective in disease discrimination using global attention. Consequently, the use of hybrid CNNViT as a viable and scalable solution to real-time plant disease detection is becoming more popular (Aboelenin *et al.*, 2025).

3.4 Object Detection and Super-Resolution Models

Other than classification of images, object detection and super-resolution can also be important in improving the process of detecting plant leaf diseases in a real-life situation. Real-time localization and identification of the diseased parts of the leaf images in leaf images can be achieved with the help of object detection models, especially the YOLO (You Only Look Once) family. YOLO based architectures are very efficient and they can be used in real time applications like drone based crop monitoring and mobile disease diagnosis systems (Dalal and Mittal, 2025).

Super-resolution models are used to solve problems that arise because of low-quality and low-resolution images that are usually taken in the field. Solutions like Residual Skip Network-based Super-Resolution for Leaf Disease Detection (RSNSR-LDD) has been used to improve image resolution prior to classification enabling the deep learning models to extract higher meaningful features. These models are able to enhance the accuracy of disease detection when the input images are corrupted by noise, blur, or compression artifacts, and, as a result, provide a significant improvement in accuracy in detection.

Objects detection and super-resolution systems have been successfully merged with CNN and ViT based classifier. All these strategies can not only enhance the level of classification but also allow exact localization of diseases, measurement of severity, and detecting the disease at early stages. This is a prerequisite of precision agriculture, where timely action can save the massive crop losses (Kamalesh Kanna *et al.*, 2024).

4. DATASETS AND EVALUATION METRICS

4.1 Publicly Available Datasets

The datasets that are publicly available are important in training, developing and evaluating deep learning models to detect plant leaf diseases. The majority of the available researches are based on benchmark datasets like PlantVillage, PlantDoc, Cassava Leaf Disease Dataset, Grape Leaf Disease Dataset and Tomato Leaf Disease Dataset. Such datasets offer marked images of both healthy and sick leaves of various crop types and disease kinds so as to make a reasonable comparison and replicate the research findings.

The most popular of these is the PlantVillage dataset which has over 50,000 images, 38 species of plant leaf diseases of multiple crops and is the biggest. It is so popular due to the quality of images, the uniformity of labeling, and the moderate acquisition conditions. In a similar fashion, the Cassava Leaf Disease Dataset has been eminent because of its relevance in the developing nations, with a variety of disease categories, including cassava mosaic disease and bacterial blight. The datasets related to plantDoc and grape localization are often employed in object location and disease detection because these datasets contain data that are obtained in the natural setting and have diverse backgrounds.

Although they have numerous benefits, publicly available datasets have a number of limitations. Images are often gathered in a controlled laboratory or greenhouse environment, with the background of the image being homogenous, the lighting being constant, and environmental sources of noise minimized. As a result, the models that are trained on these datasets do not tend to generalize to real-field conditions, whereby the leaves can be partially covered, impacted by shadows, integrate with other leaves, or have different illumination and weather conditions. Also, the majority of datasets have imbalanced classes whereby there are a large number of samples representing specific diseases in comparison with others that might have biased results in model execution.

The other key problem is that there are no large scale, professionally labeled field datasets. The creation of a dataset is time-consuming and expensive because the domain knowledge of plant pathologists or agricultural specialists is needed to annotate the plant diseases. Consequently, scholars often re-use available data, restricting the richness and representativeness of training data. These problems can be solved by the use of real-field data collection, domain adaptation and data augmentation, which is one of the pressing research directions.

4.2 Evaluation Metrics

The metrics of evaluation are necessary to determine the effectiveness, strength, and stability of deep learning models to detect the disease in plant leaves. Classification accuracy is the most frequently described measure in the literature that is gauged by the

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 01 , January - 2026

measure of the proportion of correctly classified samples. Although accuracy gives a general measure of the performance of a model, it is not accurate in skewed datasets, where prevailing classes can distort the accuracy scores.

To eliminate this shortcomings, researchers tend to provide other measures like precision, recall, and F1-score. Precision is a measure of the rate at which the positive samples were accurately predicted and it indicates how the model does not falsely conclude that a sample is positive. As mentioned above, also referred to as sensitivity, is the percentage of the true positive samples that are classified correctly, which shows that the model detects diseased leaves. The harmonic mean of the precision and recall is the F1-score which gives a balanced measure in case of unequal distribution of classes.

Specificity and Matthews Correlation Coefficient (MCC) are other significant measures that can be used to measure the model to identify healthy leaves correctly and to solve imbalanced classification problems respectively. MCC evaluates the true positives, true negatives, false positives and false negatives together giving a more inclusive and objective evaluation of the model performance.

Mean Average Precision (mAP) and Intersection over Union (IoU) are also additional measures that are typically used in object detection and localization problems. Such metrics are not only evaluated in regards to whether a disease is appropriately detected, but it gauges how well is the location of the diseased area precisely detected in the image.

On the whole, the applicability of plant disease detection models in the real world cannot be evaluated by using a single evaluation metric. A blend of complementary measurements is needed to provide strength, the ability to generalize and reliability especially when used in smart agriculture and precision farming systems.

5. Comparative Analysis and Performance Trends

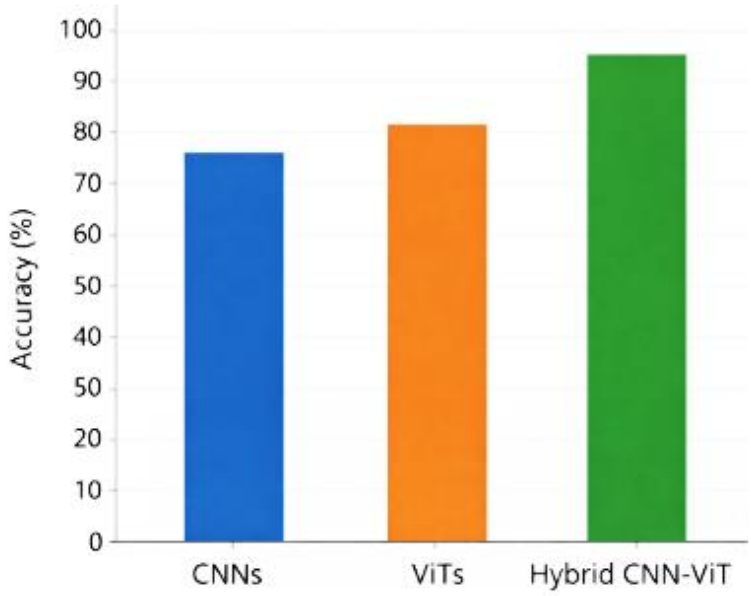

Figure 1. Accuracy Comparison of Deep Learning Models

The comparative analysis of classification accuracy of three big groups of deep learning models applied to detect plant leaf diseases are provided in Figure 1 and include Convolutional Neural Networks (CNNs), Vision Transformer (ViTs) and Hybrid CNN -ViT models. The bar chart has given an overview of the representative values of accuracy that have been reported in various studies that were involved in this review.

Based on the comparison, traditional CNN-based models have good baseline performance that is normally at 9295 percent accuracy on the popular benchmark datasets, including PlantVillage and Tomato Leaf Disease datasets. Their performance is largely due to their effective local feature extraction, but the performance is likely to be poor in complex backgrounds as well as in different illuminating conditions.

Models based on the Vision Transformer (ViT) demonstrate a significant progress and the accuracy values in this approach tend to be 95-97% and higher. This has been improved by the fact that the self-attention mechanism allows ViTs to focus on long and global

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 01 , January - 2026

contextual dependencies on leaf images. Consequently, ViTs work better in differentiating faint patterns of disease spread across the leaf surface.

The use of hybrid CNNViT networks is the most accurate as they usually reach over 9799 percent, drawing the advantages of both of these models. CNN layers give strong local features and transformer layers give global relationships resulting in enhanced generalization and robustness. These models are more effective than standalone CNN and ViT models, especially in real-field and multi-class disease recognition setting.

On the whole, Figure 2 introduces the most evident trend of performance: Hybrid CNN-ViT models > Vision Transformers > Conventional CNNs, which means that integrated architectures are the most promising to achieve accurate and scalable systems of plant leaf disease detection in smart agriculture.

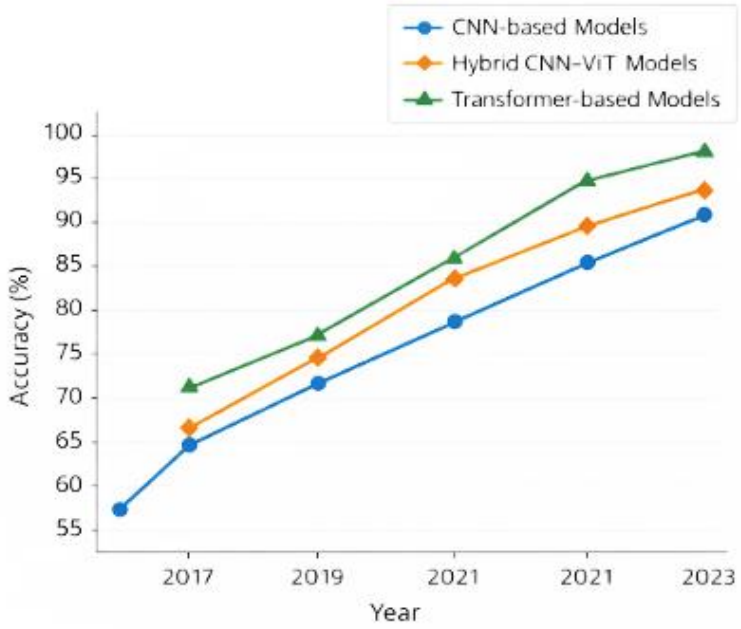

Figure 2. Accuracy Trends Over Time

The classification accuracy attained by various deep learning methods of detecting plant leaf diseases over the last few years in time is shown in Figure 2. Three significant categories of models are compared in the line graph CNN-based models, hybrid CNN-ViT models, and transformer-based models, in which the improvement of deep learning architecture has gradually enhanced the performance of detection.

The previous years used CNN-based models to prevail in plant disease detection studies with an accuracy of between 55-65. The essential cause of these improvements was in the form of deeper architectures, better optimization methods and benchmark datasets like PlantVillage. Nonetheless, the increase in accuracy on CNNs also slowed down progressively because these models could not capture long-range dependencies and global contextual information on leaf images.

The innovation of hybrid CNN ViT models was the phase of a transition. Figure 3 demonstrates that hybrid models are always superior and attain greater accuracy of between 65-93 with time when compared with traditional CNNs. This has been improved by the CNN-based local feature extraction and transformer-based global attention mechanisms which have demonstrated better generalization and have been shown to be robust against complex background and illumination conditions.

Models based on transformers, especially Vision Transformers (ViTs) and their versions, show the strongest upward trend. Starting with accuracies of greater than 70, transformer-based methods quickly develop to have an accuracy approaching that of 97-99% in recent years. Such models are able to model relationships throughout the whole leaf image, as they make use of the self-attention mechanism, which make them very useful in identifying small and spatially localized disease patterns.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 01 , January - 2026

On the whole, Figure 2 shows a clear increasing trend of research interests in moving the focus on conventional CNN-based architectures to transformer-based and hybrid models. The observed increasing trend attests that the attention-based deep learning models are the most promising direction of future plant leaf disease detection frameworks especially when it comes to real-field, high-dimensional, and multi-class agricultural data.

The evaluation of the available literature shows that transformer-based and hybrid models are always superior to traditional CNNs, especially with high-dimensional and multi-class data. Mechanisms of attention and pruning are also efficient and accurate.

6. LIMITATIONS OF CURRENT RESEARCH

Despite promising results, several limitations persist:

- Heavy reliance on publicly available datasets with limited diversity
- Requirement of large, accurately annotated datasets
- High computational cost of transformer-based models
- Limited deployment in real-field agricultural environments

These challenges hinder the widespread adoption of deep learning models in practical agricultural settings.

7. FUTURE RESEARCH DIRECTIONS

Future research should focus on:

- Developing large-scale, real-field annotated datasets
- Integrating reinforcement learning for adaptive disease diagnosis
- Exploring lightweight and energy-efficient transformer models
- Leveraging federated learning for data privacy
- Deploying models on edge and IoT devices for real-time applications

8. CONCLUSION

The review is a systematic review of the article discussing the field of deep learning in the detection of plant leaf diseases, focusing particularly on models based on Vision Transformer. The results prove that ViT and hybrid CNN ViT architectures have better performance than traditional CNNs. Although the existing set of issues concerning the data availability and the complexity of a calculation is still the case, the continuous improvement of deep learning structures and intelligent agriculture systems is likely to make the early disease detection and crop control much more efficient in the nearest future.

REFERENCES

[1] Aboelenin, S. *et al.* (2025) 'A hybrid Framework for plant leaf disease detection and classification using convolutional neural networks and vision transformer', *Complex and Intelligent Systems* [Preprint]. Available at: https://doi.org/10.1007/s40747-024-01764-x.

[2] Bektaş, J. (2024) 'Evaluation of YOLOv8 Model Series with HOP for Object Detection in Complex Agriculture Domains', *International Journal of Pure and Applied Sciences* [Preprint]. Available at: https://doi.org/10.29132/ijpas.1448068.

[3] Dalal, M. and Mittal, P. (2025) 'A Systematic Review of Deep Learning-Based Object Detection in Agriculture: Methods, Challenges, and Future Directions', *Computers, Materials and Continua* [Preprint]. Available at: https://doi.org/10.32604/cmc.2025.066056.

[4] Giakoumoglou, N., Pechlivani, E.M. and Tzovaras, D. (2023) 'Generate-Paste-Blend-Detect: Synthetic dataset for object detection in the agriculture domain', *Smart Agricultural Technology* [Preprint]. Available at: https://doi.org/10.1016/j.atech.2023.100258.

[5] Gupta, G. and Kumar Pal, S. (2025) 'Applications of AI in precision agriculture', *Discover Agriculture* [Preprint]. Available at: https://doi.org/10.1007/s44279-025-00220-9.

[6] Kamalesh Kanna, S. *et al.* (2024) 'YOLO deep learning algorithm for object detection in agriculture: a review', *Journal of Agricultural Engineering* [Preprint]. Available at: https://doi.org/10.4081/jae.2024.1641.

[7] Kamilaris, A. and Prenafeta-Boldú, F.X. (2018) 'A review of the use of convolutional neural networks in agriculture', *Journal of Agricultural Science* [Preprint]. Available at: https://doi.org/10.1017/S0021859618000436.

[8] Murugavalli, S. and Gopi, R. (2025) 'Plant leaf disease detection using vision transformers for precision agriculture', *Scientific Reports* [Preprint]. Available at: https://doi.org/10.1038/s41598-025-05102-0.