International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:3 Volume No.10 May - June – 2026
DOI:10.46647/ijetms.2026.v10i03.032 ISSN: 2581-4621

VISION TRANSFORMER-BASED PLANT LEAF DISEASE DETECTION

Pradeepkumar R[1], Vigneshwaran P[2], Yogesh Kumar S[3], Mrs.V.Vinothini AP[4]
[1]UG Scholar, Dept. of CSE, Sri Ranganathar Institute of Engineering and Technology (SRIET), Coimbatore, Tamil Nadu, India
[2]UG Scholar, Dept. of CSE, Sri Ranganathar Institute of Engineering and Technology (SRIET), Coimbatore, Tamil Nadu, India
[3]UG Scholar, Dept. of CSE, Sri Ranganathar Institute of Engineering and Technology (SRIET), Coimbatore, Tamil Nadu, India
[4]Associate Professor, Dept. of CSE, Sri Ranganathar Institute of Engineering and Technology (SRIET), Coimbatore, Tamil Nadu, India

Abstract
Plant leaf diseases significantly affect agricultural productivity and crop yield. Manual identification methods are time-consuming and prone to human error. This study proposes an automated plant leaf disease detection system using the Vision Transformer architecture. The model classifies 33 categories of healthy and diseased leaves using self-attention mechanisms. Experimental results show 87% training accuracy and 84% validation accuracy. The proposed system provides an efficient and scalable solution for precision agriculture.
Keywords: Deep Learning; Plant Disease Detection; Precision Agriculture; Self-Attention; Vision Transformer

1. INTRODUCTION

Agriculture plays a critical role in ensuring global food security and economic sustainability. Plant leaf diseases significantly reduce crop yield, degrade quality, and lead to substantial financial losses for farmers. Early and accurate detection of plant diseases is essential to prevent large-scale agricultural damage and to optimize the use of pesticides. However, traditional disease identification methods rely heavily on manual visual inspection by experts, which is time-consuming, subjective, and not scalable for large agricultural fields.

With the rapid advancement of artificial intelligence, deep learning techniques have been widely adopted for automated image-based classification tasks. Convolutional Neural Networks (CNNs) have traditionally been used for plant disease detection due to their strong local feature extraction capabilities. However, CNN-based models primarily focus on localized receptive fields and may not effectively capture long-range dependencies across the entire image.

Recently, transformer-based architectures have demonstrated remarkable performance in computer vision tasks. The Vision Transformer (ViT) model utilizes a self-attention mechanism to model global contextual relationships between image patches. By dividing an image into fixed-size patches and processing them as sequences, the model captures both local and global features efficiently. This capability makes Vision Transformers highly suitable for distinguishing subtle variations in plant leaf disease patterns.

In this research, a Vision Transformer-based framework is proposed for multi-class plant leaf disease detection. The system is designed to classify 33 categories of healthy and diseased leaves. The objective is to develop an automated, scalable, and efficient solution that supports precision agriculture and enhances early disease diagnosis.

1.1 METHOD

The proposed plant leaf disease detection system is developed using a Vision Transformer (ViT) architecture for multi-class image classification. The overall workflow of the system consists of image acquisition, preprocessing, patch generation, transformer encoding, and classification.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:3 Volume No.10 May - June – 2026
DOI:10.46647/ijetms.2026.v10i03.032 ISSN: 2581-4621

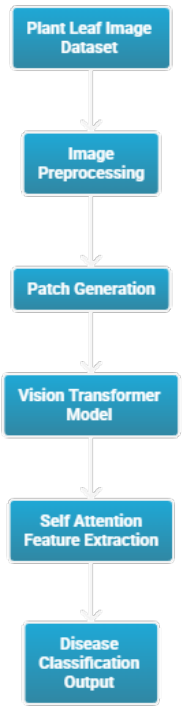

Figure 1: Workflow of Vision Transformer-based Plant Leaf Disease Detection

The overall workflow of the proposed system is illustrated in Figure 1.

Image Acquisition

The dataset used in this research is derived from the PlantVillage dataset and consists of labeled images representing 33 categories of plant leaves. The dataset was divided into training and validation sets to ensure proper evaluation of model performance. The data split helps in assessing generalization capability and avoiding overfitting.

Data Preprocessing

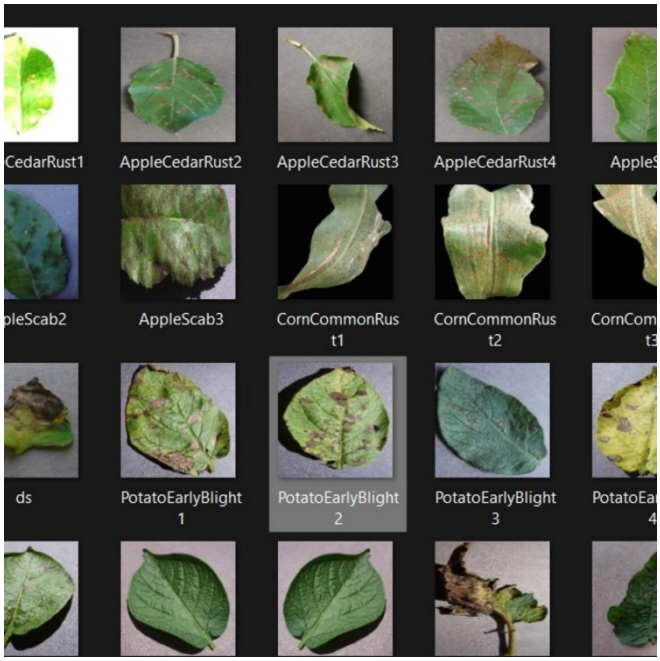

Figure 2: Sample Plant Leaf Disease Images from the Dataset.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:3 Volume No.10 May - June – 2026
DOI:10.46647/ijetms.2026.v10i03.032 ISSN: 2581-4621

Preprocessing plays a vital role in improving model performance. The following steps were applied:

- Image resizing to a fixed input dimension suitable for the Vision Transformer
- Pixel normalization to standardize input values
- Data augmentation techniques such as rotation, horizontal flipping, zooming, and brightness adjustment

These preprocessing steps enhance data diversity and improve model robustness.

Patch Generation and Embedding

In the Vision Transformer architecture, each input image is divided into fixed-size patches. These patches are flattened and converted into vector representations. A linear embedding layer maps each patch into a feature vector of equal dimension. Positional encoding is then added to retain spatial information, since transformer models do not inherently encode positional relationships.

Vision Transformer Architecture

The Vision Transformer model consists of the following core components:

- Patch Embedding Layer
- Positional Encoding
- Multi-Head Self-Attention Mechanism
- Feed-Forward Neural Network
- Residual Connections
- Layer Normalization
- Final Classification Head

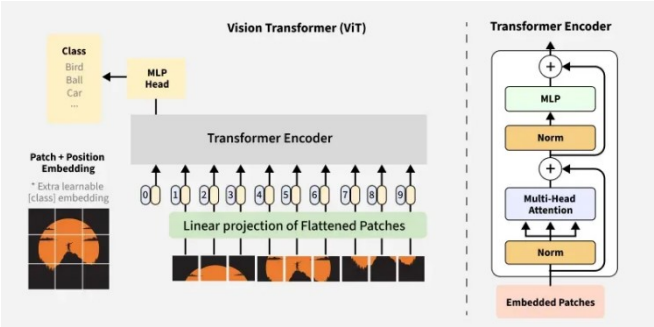

Figure 3: Architecture of the Vision Transformer Model

The architecture of the Vision Transformer used in this study is illustrated in Figure 3.

The self-attention mechanism enables the model to learn relationships between different image regions, allowing it to focus on infected portions of the leaf. This global attention capability improves classification accuracy for visually similar disease patterns.

Model Training

The model was trained using a multi-class classification approach. Cross-entropy loss was used as the objective function, and the Adam optimizer was employed for parameter updates. The training process was monitored using validation accuracy and loss metrics to ensure stable convergence and to prevent overfitting.

2. RESULTS AND DISCUSSION

2.1. Results

The proposed Vision Transformer-based model was trained and evaluated on 33 categories of plant leaf images. The dataset was divided into training and validation subsets to assess generalization performance.

The model achieved the following performance metrics:

- Training Accuracy: 87%
- Validation Accuracy: 84%

DOI:10.46647/ijetms.2026.v10i03.032 ISSN: 2581-4621

During training, the loss curve showed steady convergence without significant fluctuations, indicating stable optimization. The validation accuracy remained close to the training accuracy, demonstrating that the model did not suffer from severe overfitting.

The self-attention mechanism enabled the model to focus on disease-affected regions of the leaf images. Compared to traditional convolution-based approaches, the transformer architecture demonstrated improved contextual understanding across the entire image.

Results were evaluated using standard multi-class classification metrics, including accuracy and loss. The experimental findings confirm that the Vision Transformer effectively distinguishes between healthy and diseased leaf categories.

2.2. Discussion

The experimental results highlight the effectiveness of transformer-based architectures in agricultural image classification tasks. The ability of the Vision Transformer to capture long-range dependencies allows it to analyze subtle texture variations and color changes associated with different plant diseases.

Although the achieved validation accuracy of 84% indicates strong performance, further improvements can be achieved through hyperparameter tuning, larger datasets, and extended training epochs. Additionally, incorporating transfer learning from large-scale pretrained transformer models may further enhance classification capability.

The proposed framework demonstrates practical applicability for real-world precision agriculture systems. With proper deployment, the model can assist farmers in early disease diagnosis, reduce crop losses, and support sustainable farming practices.

CONCLUSION

This research presents a Vision Transformer-based framework for automated plant leaf disease detection. The proposed system leverages self-attention mechanisms to capture global contextual relationships across image patches, enabling effective multi-class classification of 33 plant leaf categories.

Experimental evaluation achieved 87% training accuracy and 84% validation accuracy, demonstrating strong generalization capability. The results confirm that transformer-based architectures can successfully identify subtle disease patterns and distinguish between visually similar leaf conditions.

The developed model provides a scalable and efficient solution for precision agriculture. By enabling early detection of plant diseases, the system can assist farmers in reducing crop losses, minimizing excessive pesticide usage, and improving overall agricultural productivity.

Future enhancements may include expanding the dataset, optimizing hyperparameters, and deploying the model in mobile or cloud-based agricultural monitoring systems.

ACKNOWLEDGEMENTS

The authors sincerely express their gratitude to the Department of Computer Science and Engineering, Sri Ranganathar Institute of Engineering and Technology (SRIET), Coimbatore, Tamil Nadu, India, for providing the necessary support and infrastructure to carry out this research work. The authors also thank the Associate Professor for valuable technical guidance and continuous encouragement throughout the project development.

REFERENCES

[1]. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., & Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. International Conference on Learning Representations.

[2]. Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. Frontiers in Plant Science, 7, 1419.

[3]. Too, E. C., Yujian, L., Njuki, S., & Yingchun, L. (2019). A comparative study of fine-tuning deep learning models for plant disease identification. Computers and Electronics in Agriculture, 161,

DOI:10.46647/ijetms.2026.v10i03.032 ISSN: 2581-4621
272–279.
[4]. Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A., & Jégou, H. (2021). Training data-efficient image transformers and distillation through attention. International Conference on Machine Learning, 139, 10347–10357.
[5]. Birari, H. P., Lohar, G. V., & Joshi, S. L. (2023). Advancements in machine vision for automated inspection of assembly parts: A comprehensive review. International Research Journal on Advanced Science Hub, 5(10), 365–371. doi:10.47392/IRJASH.2023.065.
[6]. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A. L., & Zhou, Y. (2021). TransUNet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*.
[7]. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin Transformer: Hierarchical vision transformer using shifted windows. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 10012–10022.
[8]. Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture*, 145, 311–318.