Global Journal of Engineering Innovations & Interdisciplinary Research

Original Article


# Agri Vision: Deep Learning Based Plant Disease Detection Using Swin Transformer

D. Anil[1], Yenduru Sai Jaswanth[2], Parakandla Bharath Kumar[2], Rajanala Naga Surendra[2], Vishnumolakala Venkat Sai Ram[2]

[1]Associate Professor, Department of CSE, KKR & KSR Institute of Technology and Sciences, Guntur, Andhra Pradesh, India
[2]B.Tech Student, Department of CSE, KKR & KSR Institute of Technology and Sciences, Guntur, Andhra Pradesh, India

Correspondence

**D Anil**

Associate Professor, Department of CSE, KKR & KSR Institute of Technology and Sciences, Guntur, Andhra Pradesh, India

- Received Date: 08 Jan 2026
- Accepted Date: 25 Jan 2026
- Publication Date: 15 Mar 2026

Keywords

Plant disease detection, Swin Transformer, deep learning, transfer learning, image classification, precision agri-culture, PlantVillage



## Abstract

*Plant disease detection plays a major role in preci-sion agriculture because delayed or inaccurate diagnosis directly affects crop quality and productivity. Manual inspection of leaves by farmers or agricultural experts is time-consuming, subjective, and difficult to scale for large cultivation areas. This paper presents Agri Vision, a deep learning based plant disease detection system that uses a pre-trained Swin Trans-former model for classifying plant leaf images into healthy or diseased categories. The proposed pipeline includes dataset loading, image preprocessing, augmentation, transfer learning, transformer-based feature learning, evaluation, and deployment through a web interface. In contrast to conventional machine learning approaches that rely on handcrafted features and CNN-only systems that primarily emphasize local patterns, the Swin Transformer can model both local and global characteristics of leaf symptoms. The overall framework supports fast, automated, and reliable disease prediction for smart agriculture applications.*

## Introduction

In agriculture, early disease detection is essential for pre-venting crop loss and improving productivity. Traditional diagnosis is often performed through manual inspection of leaves by farmers or domain experts. Although expert-based diagnosis can be effective, it becomes difficult when crop areas are large, when expert access is limited, or when symptoms are visually similar across diseases. These limitations motivate the development of automated image-based diagnosis systems. Recent advances in machine learning and deep learning have enabled practical computer vision systems for agricul-ture. Earlier methods relied on handcrafted color, texture, and shape features with classifiers such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Decision Trees. Later, deep convolutional neural networks improved performance by learning features directly from images. More recently, transformer-based vision models such as the Vision Transformer and Swin Transformer have demonstrated strong capability in modeling long-range dependencies and hierarchical image representations [4,5].

This work focuses on a plant leaf disease detection system that takes a plant leaf image as input, preprocesses it, and predicts whether the leaf is healthy or affected by a disease. The design uses transfer learning with a pre-trained Swin Transformer to improve classification effectiveness while re-ducing training time.

## Objectives

The main objectives of the proposed work are as follows:

- To build an automated plant leaf disease detection sys-tem using a Swin Transformer model.
- To classify leaf images into healthy and diseased cate-gories with improved reliability.
- To apply transfer learning for better accuracy and lower computational cost.
- To preprocess plant leaf images through resizing, nor-malization, and augmentation.
- To evaluate the trained model using accuracy, precision, recall, and F1-score.
- To provide a simple web-based interface for real-time prediction with confidence score.

## Problem Statement

Plant diseases are one of the major causes of crop damage and reduced agricultural output. Manual disease inspection is labor-intensive, slow, and prone to inconsistency. In rural or resource-constrained settings, timely access to agricultural expertise is also limited. Although several machine learning and CNN-based systems have been proposed, many of them

**Citation**: Anil D, Yenduru SJ, Parakandla BK, Rajanala NS, Vishnumolakala VSR. Agri Vision: Deep Learning Based Plant Disease Detection Using Swin Transformer. GJEIIR. 2026;6(3):216.

focus mainly on local image features or operate well only under controlled conditions. Therefore, there is a need for an efficient, scalable, and reliable automated disease detec-tion system that can learn richer visual representations from plant leaf images and support practical use through a simple interface.

## Literature Review

Image-based plant disease detection has been studied extensively in recent years. A widely used foundation for this area is the PlantVillage image repository introduced by Hughes and Salathé [3]. Mohanty et al. showed that deep learning models can achieve high accuracy on controlled plant leaf datasets [1]. Ferentinos further demonstrated the effectiveness of deep neural networks for plant disease diagnosis across multiple crops and disease types [2].

Fine-tuning strategies for deep learning models in plant disease identification were studied by Too et al. [8]. Ramcharan et al. extended image-based disease detection to cassava, highlighting the practical relevance of deep learning in agriculture [11]. Sladojevic et al. also demonstrated plant disease recognition using deep neural networks for leaf image classification [15].

On the vision backbone side, CNN architectures such as AlexNet [9], VGG [13], ResNet [12], and Inception [14] laid the foundation for modern image classification. Transformer-based vision models later emerged from the success of at-tention mechanisms in sequence modeling [17]. Vision Trans-former (ViT) showed that transformers can perform strongly on image recognition [4]. Swin Transformer improved this direction by introducing hierarchical representation learning and shifted windows, making it better suited for vision tasks with multi-scale structure [5]. These developments motivate the use of Swin Transformer for plant disease classification.

## Methodology

The proposed methodology follows the sequence of operations required for automated plant leaf disease recognition. The overall system begins with dataset acquisition and proceeds through preprocessing, transfer learning, classification, evaluation, and deployment.

### Dataset Collection

Plant leaf images are collected from the PlantVillage dataset, which contains labeled examples of healthy and diseased leaves [3]. These images are used for training, validation, and testing the model.

### Image Preprocessing

All images are resized to a uniform resolution to make them compatible with the model input size. Normalization is applied to stabilize training. Data augmentation techniques such as rotation and flipping are also used to improve generalization and reduce overfitting.

### Transfer Learning

A pre-trained Swin Transformer model is adopted as the backbone for feature extraction and classification. The final classification layer is modified according to the number of disease categories in the dataset. Fine-tuning allows the model to adapt its learned visual features to the plant disease domain.

### Transformer-Based Feature Learning

The Swin Transformer learns hierarchical visual representations using window-based self-attention and shifted-window mechanisms [5]. This allows the model to capture both local disease details and broader global patterns across the leaf surface.

**TABLE I:** Summary of Proposed Pipeline Modules

| | |
|---|---|
| Dataset Loading | Loads labeled healthy and diseased leaf images |
| Preprocessing | Resizes, normalizes, and augments images |
| Dataset Split | Separates data into train, validation, and test sets |
| Backbone Model | Loads pre-trained Swin Transformer |
| Fine-Tuning | Adapts model to plant disease classes |
| Evaluation | Computes classification metrics |
| Prediction | Accepts uploaded image and performs inference |
| Result Display | Shows disease label and confidence score |

### Model Evaluation

The trained model is evaluated using standard classification metrics such as accuracy, precision, recall, and F1-score. A confusion matrix can also be used to examine inter-class prediction behavior.

### Deployment

The trained model is integrated into a web-based application. Users upload a plant leaf image, and the system returns the predicted disease category along with a confidence score.

## System Architecture

The plant leaf disease detection system is designed as a sequential deep learning pipeline. The architecture includes the following modules:

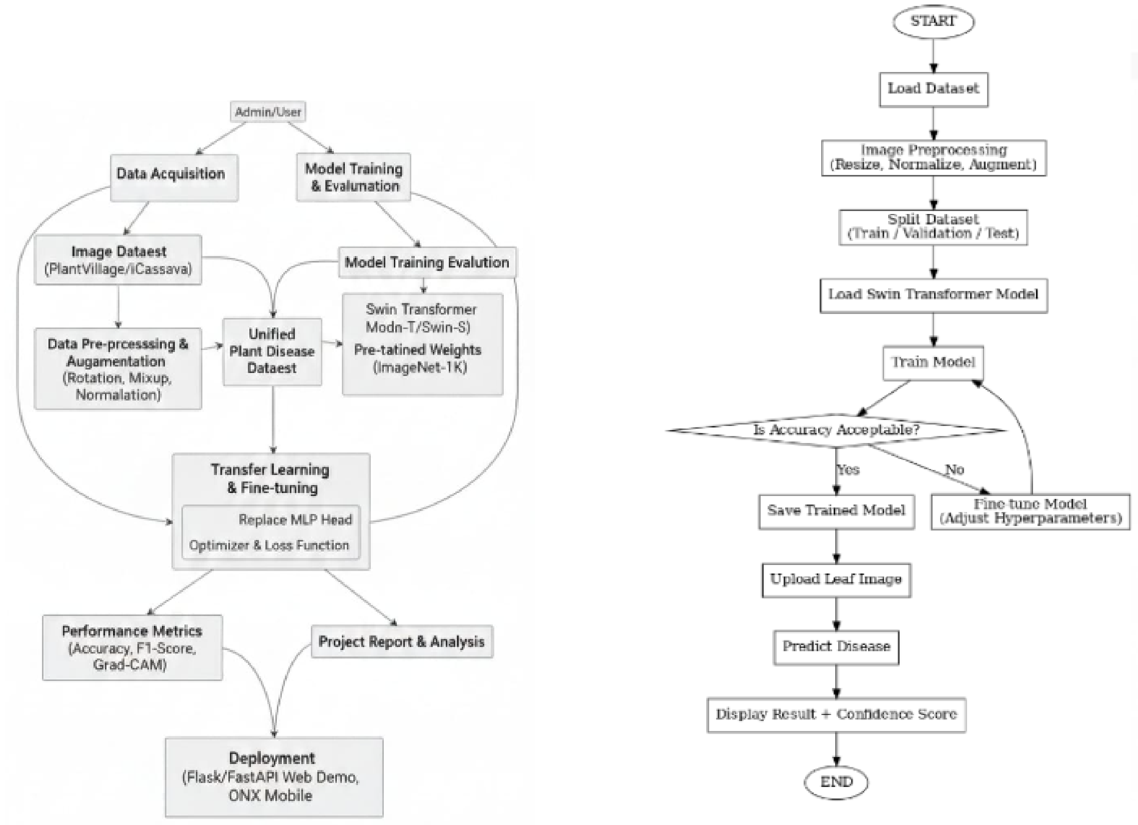
1. **Dataset Loading Module:** Loads plant leaf images and their labels.
2. **Image Preprocessing Module:** Resizes and normalizes images and applies augmentation.
3. **Dataset Splitting Module:** Separates data into training, validation, and testing partitions.
4. **Swin Transformer Model Module:** Loads a pre-trained model and replaces the classification head.
5. **Model Training and Fine-Tuning Module:** Learns task-specific patterns from labeled images.
6. **Model Saving Module:** Stores the trained model for future inference.
7. **Prediction Module:** Accepts new uploaded leaf images and predicts their category.
8. **Result Display Module:** Displays the predicted disease name and confidence score.

The architectural flow can be summarized as:

*Start → Load Dataset → Image Preprocessing → Dataset Split → Load Swin Transformer → Train/Fine-Tune → Save Model → Upload Image → Predict Disease → Display Result → End*

## Experimental Flow And Evaluation Criteria

The model is tested using plant leaf images from multiple disease and healthy classes. The evaluation is based on standard supervised image classification. The following metrics are used




(a) System architecture (b) Workflow representation

***Fig. 1:*** *Architecture and workflow figures from the source manuscript.*

*TABLE II: Evaluation Metrics Used in the System*

| Metric | Meaning |
|---|---|
| Accuracy | Overall fraction of correct predictions |
| Precision | Ratio of correct positive predictions to total predicted positives |
| Recall | Ratio of correct positive predictions to actual positives |
| F1-score | Harmonic mean of precision and recall |
| Confusion Matrix | Class-wise distribution of correct and incorrect predictions |

## Results

The proposed Swin Transformer based plant leaf disease detection system was tested on a dataset containing healthy and diseased leaf images across multiple classes. The outcome shows that the system correctly identified different disease cat egories and produced predictions with associated confidence levels. The model performed consistently across testing samples, and image preprocessing together with transfer learning improved the classification capability.

Compared with traditional machine learning and CNN-based methods, the Swin Transformer improved reliability by learning both local and global characteristics of plant leaves. The prediction process also completed quickly, making the approach efficient for practical use.

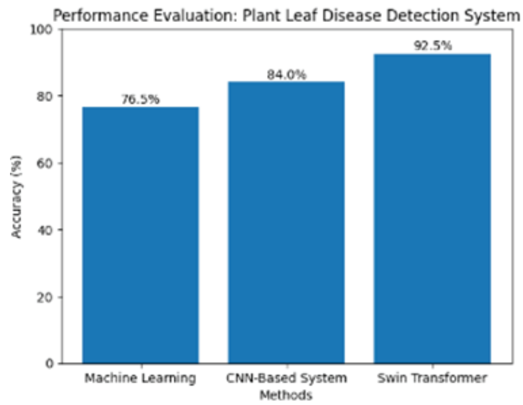



*Fig. 2: Experimental result figure from the source manuscript*

**Table 3.** Qualitative Comparison of Detection Approaches

| Approach | Automation | Local Features | Global Context |
|---|---|---|---|
| Manual Inspection | Low | Human-dependent | Human-dependent |
| Traditional ML | Medium | Handcrafted | Limited |
| CNN-based Models | High | Strong | Moderate |
| Swin Transformer | High | Strong | Strong |

## Discussion

The discussion highlights the suitability of transformer-based models for agricultural image classification. By combining preprocessing, transfer learning, and attention-based feature extraction, the system can represent both fine-grained leaf symptoms and broader disease patterns. This is important because many plant diseases are not identified only by local spots; they may also involve spatial distribution, texture variation, and overall leaf condition.

A major strength of the system is the use of a pre-trained Swin Transformer backbone, which reduces training time while still enabling powerful feature learning. Another advan-tage is the deployment-oriented design. Instead of remaining only as a training experiment, the model is connected to a web-based interface so that users can upload images and obtain real-time predictions.

At the same time, the paper does not provide detailed numerical comparisons, confusion matrices, or ablation studies. Therefore, future work should include stronger quantitative reporting and more extensive real-world validation.

## Implementation And Deployment Considerations

For practical deployment, the proposed system requires a trained model, an image preprocessing routine, and an interface for prediction. The user workflow is straightforward: upload a leaf image, preprocess it to match the model input requirements, run inference, and display the predicted class with a confidence score.

In deployment scenarios, the following considerations are relevant:

- The model should be validated on field images, not only controlled datasets.
- Lightweight deployment options may be considered for mobile or low-resource environments.
- Proper class coverage is needed to avoid overconfident predictions on unseen disease types.

## Limitations And Future Scope

The current system is primarily described using a controlled dataset-based workflow. Since the paper does not report large-scale field testing, there may be a gap between laboratory-style images and real agricultural environments. Variations in lighting, background clutter, leaf occlusion, disease stage, and mixed symptoms can affect practical performance.

Future scope includes:

1) testing on larger real-world datasets,
2) extending the system to additional crop categories,
3) optimizing the model for mobile deployment,
4) integrating treatment recommendation support, and
5) improving explainability through attention visualization or saliency methods.

## Conclusion

This paper presented a plant leaf disease detection system based on Swin Transformer. The system automates plant disease identification from leaf images using image pre-processing, transfer learning, hierarchical transformer-based feature learning, and web-based deployment. The system is accurate, efficient, and reliable for classification of healthy and diseased plant leaves. The use of Swin Transformer provides an effective way to capture local as well as global disease patterns. Overall, the work offers a practical AI-based direction for plant disease detection in smart agriculture and provides a foundation for future deployment on broader real-world data.

## References


1. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779–788.
2. J. Redmon and A. Farhadi, “YOLO9000: Better, Faster, Stronger,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2017, pp. 6517–6525.
3. J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement,”arXiv preprint arXiv:1804.02767, 2018.
4. A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” arXiv preprint arXiv:2004.10934, 2020.
5. R. Smith, “An Overview of the Tesseract OCR Engine,” in Proc. Int. Conf. Document Analysis and Recognition (ICDAR), 2007, pp. 629–633.
6. G. Bradski, “The OpenCV Library,” Dr. Dobb’s Journal of Software Tools, vol. 25, no. 11, pp. 120–126, 2000.
7. X. Zhou, C. Yao, H. Wen, Y. Wang, S. Zhou, W. He, and J. Liang, “EAST: An Efficient and Accurate Scene Text Detector,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2017, pp. 5551–5560.
8. Y. Baek, B. Lee, D. Han, S. Yun, and H. Lee, “Character Region Awareness for Text Detection,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2019, pp. 9365–9374.
9. S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real- Time Object Detection with Region Proposal Networks,” in Advances in Neural Information Processing Systems (NeurIPS), 2015, pp. 91–99.
10. W. Liu et al., “SSD: Single Shot MultiBox Detector,” in Proc. European Conf. Computer Vision (ECCV), 2016, pp. 21–37.
11. G. Jaume, H. Ekenel, and J.-P. Thiran, “FUNSD: A Dataset for Form Understanding in Noisy Scanned Documents,” in Proc. Int. Conf. Document Analysis and Recognition Workshops (ICDARW), 2019, pp. 1–6.
12. S. Park et al., “CORD: A Consolidated Receipt Dataset for Post- OCR Parsing,” in Proc. Int. Conf. Document Analysis and Recognition Workshops (ICDARW), 2019, pp. 97–102.
13. D. Katti et al., “Chargrid: Towards Understanding 2D Documents,” in Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP), 2018, pp. 4459–4469.
14. R. B. Palm, F. Laws, and O. Winther, “CloudScan—A Configuration- Free Invoice Analysis System Using Recurrent Neural Networks,” in Proc. Int. Conf. Document Analysis and Recognition (ICDAR), 2017,pp. 406–413.
15. Y. Xu, M. Li, L. Cui, S. Huang, F. Wei, and M. Zhou, “LayoutLM: Pre-training of Text and Layout for Document Image Understanding,” in Proc. ACM SIGKDD Int. Conf. Knowledge Discovery & Data Mining (KDD), 2020, pp. 1192–1200.
16. Y. Xu et al., “LayoutLMv2: Multi-modal Pre-training for Visually-Rich Document Understanding,” in Proc. Annual Meeting of the Association for Computational Linguistics (ACL-IJCNLP), 2021, pp. 2579–2591.
17. A. Appalaraju et al., “DocFormer: End-to-End Transformer for Docu- ment Understanding,” in Proc. IEEE/CVF Int. Conf. Computer Vision Workshops (ICCVW), 2021, pp. 993–1003.
18. G. Kim et al., “OCR-Free Document Understanding Transformer,” in Proc. European Conf. Computer Vision (ECCV), 2022, pp. 498–517.
19. J. Liang, D. Doermann, and H. Li, “Camera-Based Analysis of Text and Documents: A Survey,” International Journal on Document Analysis and Recognition, vol. 7, no. 2–3, pp. 84–104, 2005.
20. R. C. Gonzalez and R. E. Woods, Digital Image Processing, 4th ed. New York, NY, USA: Pearson, 2018