Journal of Computer Science and Intelligent Systems
Volume 1, Issue 1 | July - September 2026 | www.journalindiapublication.com | 2026 Journal India Publication

Hybrid Vision Mamba Convnext V2 for Sesame Leaf Disease Classification

*K Sivasankari * B Pazhaniraj ***

ABSTRACT

Purpose

Sesame (Sesamum indicum L.) is an economically important oilseed crop whose productivity is significantly affected by foliar diseases. This study aims to develop an automated and accurate deep learning framework, VMCAF-Net, for early identification of sesame leaf diseases, supporting crop health monitoring and precision agriculture.

Design/Methodology/Approach

The proposed VMCAF-Net integrates Vision Mamba and ConvNeXt V2 to capture complementary global contextual and local texture features. An enhanced preprocessing pipeline incorporating CLAHE, gamma correction, image resizing, and data augmentation is employed to improve image quality and model robustness. A Cross-Attention Feature Fusion module combines the extracted representations, while a Label-Smoothed Softmax classifier performs disease classification. Model performance is evaluated using Accuracy, Precision, Recall, F1-score, ROC-AUC, Cohen's Kappa, and Matthews Correlation Coefficient.

Findings

Experimental evaluation demonstrates that VMCAF-Net outperforms existing deep learning models in sesame leaf disease identification. The integration of Vision Mamba and ConvNeXt V2 effectively captures both local disease characteristics and broader contextual patterns, while cross-attention fusion enhances discriminative feature representation. The results indicate that the proposed framework provides an accurate, robust, and computationally efficient approach for automated disease detection.

Practical Implications

The proposed framework can support farmers, agricultural researchers, and precision-agriculture practitioners by enabling rapid and automated identification of sesame leaf diseases. Its potential for real-time deployment can facilitate timely disease management, reduce crop losses, and support data-driven agricultural decision-making.

---

* *Head & Assistant Professor, School of computing E.S Arts and Science College (Co-ed), Annamalai University, Tamilnadu, India, [Email: shivashankariphd@gmail.com]*

* * *Assistant Professor, School of computing, E.S Arts and Science College (Co-ed), Annamalai University, Tamilnadu, [Email: pazhaniraj508@gmail.com]*

Originality/Value

The study proposes a novel hybrid architecture that combines Vision Mamba, ConvNeXt V2, cross-attention feature fusion, and enhanced image preprocessing for sesame leaf disease identification. The integration of complementary feature representations with comprehensive performance evaluation provides a potentially effective framework for intelligent and scalable crop disease diagnosis.

Keywords: Sesame Leaf Disease; Vision Mamba; ConvNeXt V2; Cross-Attention Fusion; Deep Learning; Precision Agriculture.

JEL Codes: O33; C45; Q16.

---

1. Introduction

Agriculture remains one of the most significant contributors to global food security and economic sustainability. Among oilseed crops, sesame (*Sesamum indicum* L.) occupies an important position due to its high nutritional value, medicinal properties, and commercial demand. Sesame seeds are rich in proteins, antioxidants, vitamins, minerals, and high-quality edible oil. India is one of the major producers of sesame, where the crop supports the livelihood of millions of farmers. Despite its economic importance, sesame production is continuously threatened by numerous fungal, bacterial, and viral diseases that considerably reduce crop yield and quality. Leaf diseases represent one of the major challenges in sesame cultivation. Diseases including Alternaria leaf spot, Cercospora leaf spot, Powdery mildew, Bacterial leaf blight, Anthracnose, and Phyllody damage leaf tissues, reduce photosynthetic efficiency, and eventually decrease seed production. Early disease detection enables timely intervention through appropriate pesticide application and agronomic management, minimizing crop losses and improving productivity.Recent developments in deep learning have revolutionized image classification by enabling automatic hierarchical feature learning. Convolutional Neural Networks (CNNs), including AlexNet, VGG16, ResNet, DenseNet, EfficientNet, and ConvNeXt, have demonstrated remarkable success in plant disease recognition. CNNs effectively capture local texture information and spatial patterns; however,

Journal of Computer Science and Intelligent Systems
JOURNAL India Publication
Volume 1, Issue 1 | July - September 2026 | www.journalindiapublication.com | 2026 Journal India Publication

their receptive fields remain relatively limited, making it difficult to model long-range dependencies between distant image regions.

Transformer-based architectures have recently emerged as powerful alternatives for computer vision tasks. Vision Transformer (ViT), Swin Transformer, DeiT, and related models utilize self-attention mechanisms to learn global contextual relationships across entire images. Although transformers achieve excellent performance, they generally require large training datasets and substantial computational resources, restricting their applicability in resource-constrained agricultural systems. The recently introduced Vision Mamba architecture represents a significant advancement in computer vision. Unlike conventional transformers, Vision Mamba employs selective state-space models to efficiently capture long-range dependencies while maintaining linear computational complexity. This enables Vision Mamba to process high-resolution agricultural images more efficiently than conventional transformer architectures. Meanwhile, ConvNeXt V2 combines the strengths of modern convolutional neural networks with transformer-inspired architectural improvements. Its enhanced convolutional blocks provide highly discriminative local feature representations while maintaining computational efficiency suitable for practical deployment. Despite these advances, existing approaches typically employ either CNN-based local feature extraction or transformer-based global representation independently. Consequently, valuable complementary information remains underutilized. Furthermore, simple feature concatenation strategies often fail to model complex interactions between heterogeneous feature representations. To overcome these limitations, this research proposes a novel Vision Mamba and ConvNeXt V2 Hybrid Network with Cross-Attention Feature Fusion (VMCAF-Net) for sesame leaf disease identification. The proposed framework combines the local texture learning capability of ConvNeXt V2 with the global contextual modeling ability of Vision Mamba. A Cross-Attention Feature Fusion module adaptively integrates complementary representations, allowing the network to focus on disease-specific visual characteristics. Finally, adaptive average pooling and a Label-Smoothed Softmax classifier perform robust disease classification.

2. Literature Review

Woo et al. [1] proposed ConvNeXt V2, integrating Masked Autoencoders and Global Response Normalization to improve feature representation and scalability. The model achieved superior image classification performance with enhanced feature diversity. Dosovitskiy et al. [2] introduced the Vision Transformer (ViT), which processes image patches using self-attention to capture global contextual information. The architecture achieved state-of-the-art performance on large-scale image datasets. Liu et al. [3] proposed the Swin Transformer, employing shifted-window self-attention for hierarchical feature extraction. The model effectively balances computational efficiency and classification accuracy. He et al. [4] introduced ResNet, which uses residual connections to overcome degradation problems in deep neural networks. The architecture significantly improved image recognition accuracy. Huang et al. [5] developed DenseNet, where dense connectivity promotes feature reuse and efficient gradient propagation. The model reduces parameters while improving classification performance. Tan and Le [6] proposed EfficientNet, introducing compound scaling to optimize network depth, width, and resolution simultaneously. The architecture achieved high accuracy with fewer parameters. Selvaraju et al. [7] proposed Grad-CAM, an explainable AI technique that visualizes important image regions influencing predictions. It enhances the interpretability of deep learning models.Chen et al. [8] reviewed recent deep learning techniques for plant image recognition, highlighting CNNs, transformers, and attention mechanisms. The study emphasized the importance of transfer learning and data augmentation.Nibret et al. [9] developed a deep CNN model for sesame plant disease classification. The model achieved promising accuracy for identifying multiple sesame leaf diseases.Zhang et al. [10] proposed VMamba for plant leaf disease identification using Vision Mamba. The model efficiently captured long-range dependencies with reduced computational complexity.Mamun et al. [11] introduced ConMamba, combining Vision Mamba with contrastive learning for robust plant disease detection. The framework improved feature discrimination under challenging field conditions.Zhang and Liu [12] proposed a multi-scale Vision Mamba architecture that integrates hierarchical feature representations. The model improved robustness for plant disease classification. Zhang et al. [13] presented a hybrid Mamba–Transformer

Journal of Computer Science and Intelligent Systems
JOURNAL India Publication
Volume 1, Issue 1 | July - September 2026 | www.journalindiapublication.com | 2026 Journal India Publication

framework for smart agriculture. The hybrid architecture enhanced global feature representation and disease classification accuracy. ConvGeM-Next [14] introduced a hybrid convolution-based framework with advanced feature fusion for plant disease detection. The architecture achieved high classification performance while maintaining computational efficiency.Bedi and Gole [15] proposed a hybrid Convolutional Autoencoder-CNN model for plant disease detection. The framework improved feature learning and disease classification accuracy.Bedi et al. [16] developed PDSE-Lite**,** a lightweight framework combining convolutional autoencoders and few-shot learning. The model efficiently estimates plant disease severity with reduced computational cost.

3. Research Gap

Despite significant advancements in deep learning-based plant disease classification, several research gaps remain. Conventional CNNs effectively capture local texture features but fail to model long-range contextual information. Although Vision Transformers learn global dependencies, their high computational complexity limits real-time deployment. Moreover, existing hybrid models often use simple feature concatenation, which does not effectively integrate complementary features. Their performance also degrades under varying field conditions due to illumination changes and background complexity. In addition, the lack of interpretability reduces the reliability of disease prediction systems. To overcome these limitations, the proposed VMCAF-Net integrates Vision Mamba and ConvNeXt V2 with a Cross-Attention Feature Fusion module, enhanced preprocessing techniques, and Grad-CAM visualization to achieve accurate, robust, and interpretable sesame leaf disease classification.

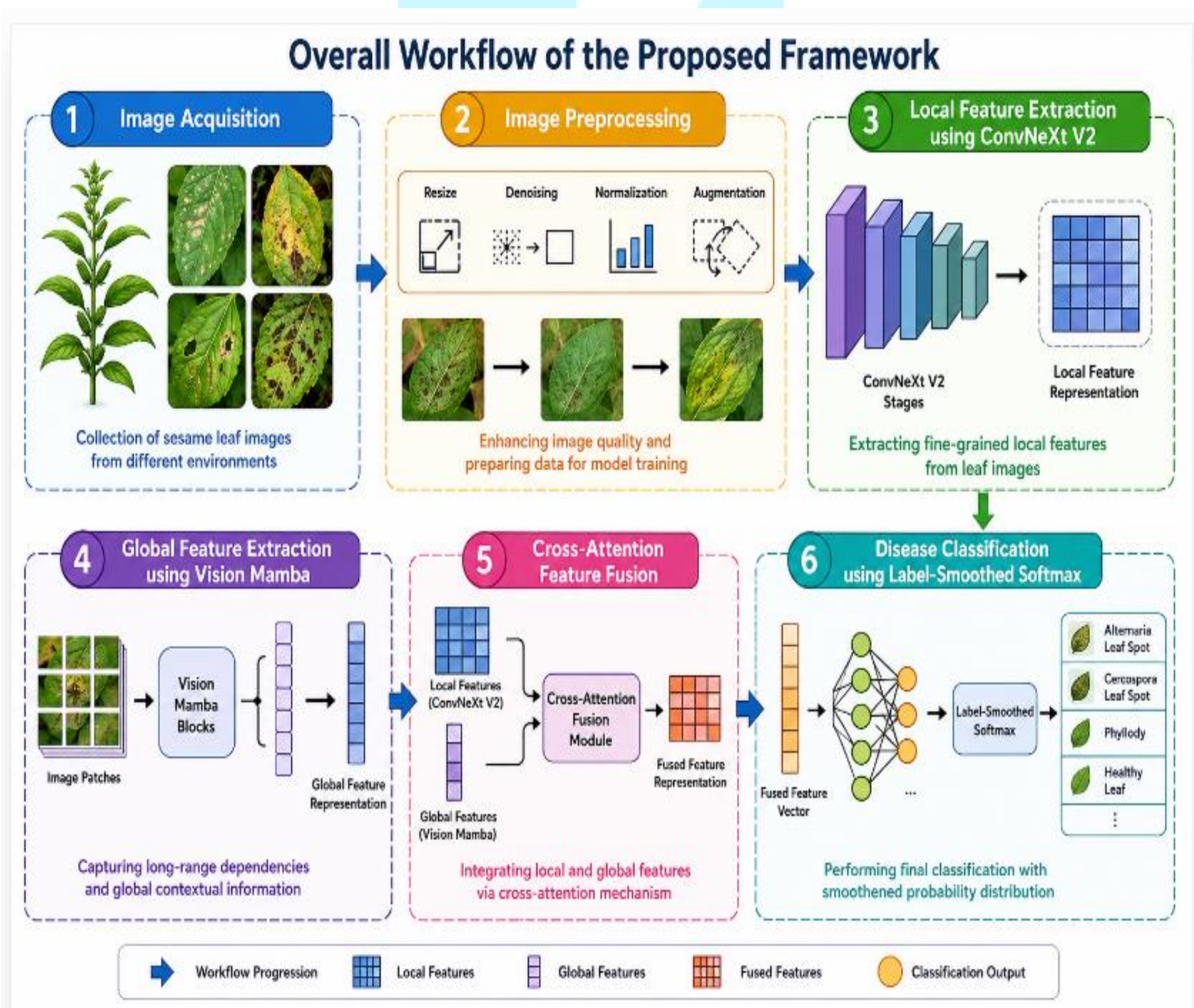
4. Motivation

The existing literature indicates that no single deep learning architecture effectively captures both fine-grained local lesion characteristics and long-range contextual information while maintaining computational efficiency for precision agriculture. Although Vision Mamba excels at modeling global dependencies, ConvNeXt V2 provides superior local feature extraction. Therefore, this

JOURNAL India Publication
Journal of Computer Science and Intelligent Systems
Volume 1, Issue 1 | July - September 2026 | www.journalindiapublication.com | 2026 Journal India Publication

work proposes a hybrid framework, VMCAF-Net, which integrates Vision Mamba and ConvNeXt V2 through a Cross-Attention Feature Fusion module to exploit complementary feature representations. The proposed framework aims to improve sesame leaf disease classification accuracy, reduce computational complexity compared with transformer-only models, provide interpretable predictions using Grad-CAM, and enable efficient deployment on edge devices for real-time precision agriculture applications.

5. Proposed Methodology



Journal of Computer Science and Intelligent Systems
JOURNAL India Publication
Volume 1, Issue 1 | July - September 2026 | www.journalindiapublication.com | 2026 Journal India Publication

FIG 1 : OVERALL WORKFLOW OF THE PROPOSED FRAMEWORK

PROPOSED ALGORITHIM: **PROPOSED VMCAF-NET**

Input: Preprocessed sesame leaf image I

Output: Predicted disease class $\hat{y}$

1. Load the input image I.
2. Apply preprocessing (CLAHE, Gamma Correction, Resize, and Data Augmentation).
3. Extract local features using **ConvNeXt V2**:

   F_c=ConvNeXtV2(I)

4. Extract global features using **Vision Mamba**:

   F_m=VisionMamba(I)

5. Fuse local and global features using Cross-Attention:

   F_{fusion}=Attention(Q,K,V)+F_c

6. Apply Adaptive Average Pooling to obtain compact feature vectors.
7. Classify the pooled features using an MLP and Softmax.
8. Compute the Label-Smoothed Cross-Entropy Loss:
9. Update model parameters using the AdamW optimizer until convergence.
10. Return the predicted disease class$\hat{y}$

A. OVERVIEW

This research proposes a novel hybrid deep learning framework called **VMCAF-Net (Vision Mamba–ConvNeXt V2 with Cross-Attention Fusion Network)** for automatic sesame leaf disease identification. The framework is designed to overcome the limitations of conventional convolutional neural networks and transformer-based architectures by simultaneously learning

Journal of Computer Science and Intelligent Systems
Volume 1, Issue 1 | July - September 2026 | www.journalindiapublication.com | 2026 Journal India Publication

local texture features and global contextual representations.Unlike traditional CNN-based approaches, which mainly focus on local receptive fields, and transformer-based methods, which require substantial computational resources, the proposed VMCAF-Net combines the complementary strengths of ConvNeXt V2 and Vision Mamba using a Cross-Attention Feature Fusion (CAFF) module.

The overall workflow of the proposed framework consists of six major stages:

1. Image Acquisition
2. Image Preprocessing
3. Local Feature Extraction using ConvNeXt V2
4. Global Feature Extraction using Vision Mamba
5. Cross-Attention Feature Fusion
6. Disease Classification using Label-Smoothed Softmax

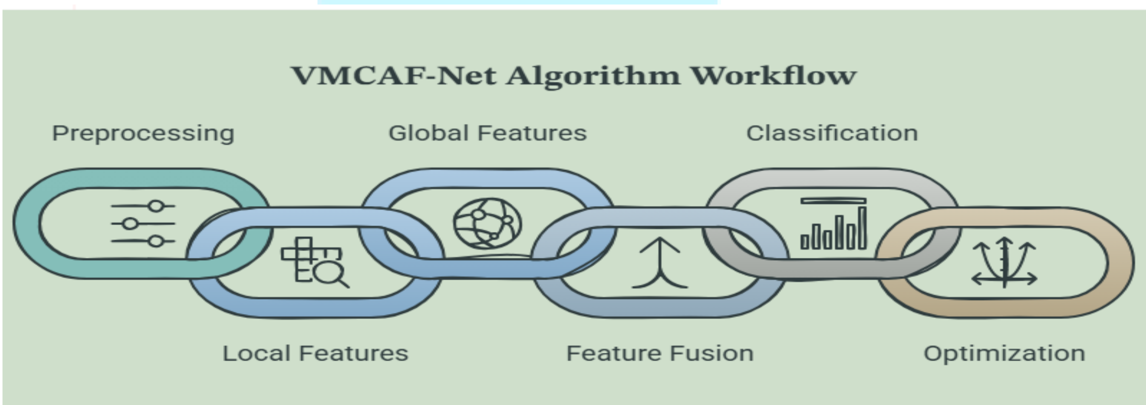


Fig 2 : VMCAF – Net Algorithm Workflow

B. IMAGE PREPROCESSING

Image preprocessing is a vital stage in the proposed VMCAF-Net framework, as the quality of input images directly affects disease classification performance. Sesame leaf images collected from field environments often contain variations in illumination, shadows, background clutter, dust, and noise, which can obscure disease symptoms. To address these issues, a preprocessing

Journal of Computer Science and Intelligent Systems
Volume 1, Issue 1 | July - September 2026 | www.journalindiapublication.com | 2026 Journal India Publication

pipeline consisting of Contrast Limited Adaptive Histogram Equalization (CLAHE), Gamma Correction, Image Resizing, and Data Augmentation is employed. CLAHE enhances local contrast, making disease symptoms such as lesions and discoloration more prominent, while Gamma Correction normalizes image brightness under varying lighting conditions. The processed images are then resized to 224 × 224 × 3 pixels to ensure compatibility with both ConvNeXt V2 and Vision Mamba architectures. Finally, data augmentation techniques, including flipping, rotation, cropping, MixUp, CutMix, brightness adjustment, and Gaussian noise addition, increase dataset diversity, reduce overfitting, and improve the robustness and generalization capability of the proposed model for accurate sesame leaf disease classification.

C. LOCAL FEATURE EXTRACTION USING CONVNEXT V2

The first branch of the proposed VMCAF-Net utilizes ConvNeXt V2 as the local feature extraction network. ConvNeXt V2 is a modern convolutional neural network architecture that combines the strengths of conventional convolutional operations with several design improvements inspired by transformer-based models. It provides excellent capability for learning detailed texture information while maintaining computational efficiency.Leaf diseases usually appear as localized visual symptoms such as small lesions, fungal spots, chlorosis, necrotic regions, and vein discoloration. ConvNeXt V2 effectively captures these fine-grained texture patterns through hierarchical convolutional operations. The network consists of depthwise convolution, pointwise convolution, layer normalization, GELU activation, Global Response Normalization (GRN), and residual connections.Depthwise convolution extracts spatial information independently from each feature channel, while pointwise convolution combines the extracted information to generate more discriminative feature representations. Layer normalization stabilizes the training process, and the GELU activation function introduces non-linearity, enabling the model to learn complex disease characteristics. Global Response Normalization further improves feature diversity and enhances the representational capability of the network. Residual connections facilitate efficient gradient propagation and enable the training of deeper architectures without performance degradation.The output of ConvNeXt V2 represents the local texture information of sesame leaves and provides detailed descriptions of disease symptoms, including lesion shape, color variation, and texture abnormalities.

D. GLOBAL FEATURE EXTRACTION USING VISION MAMBA

Although convolutional neural networks effectively learn local features, they are often limited in capturing long-range relationships between different regions of an image. To overcome this limitation, the proposed framework incorporates Vision Mamba as the global feature extraction network.Vision Mamba is a recently developed deep learning architecture based on Selective State Space Models (SSMs). Unlike transformer-based architectures that rely on computationally expensive self-attention mechanisms, Vision Mamba efficiently captures long-range dependencies while maintaining linear computational complexity. This enables the model to analyze high-resolution images with lower memory consumption and faster inference.Vision Mamba processes the image as a sequence of visual tokens and learns the relationships among different leaf regions. Consequently, it captures global contextual information such as disease spread, overall leaf deformation, and spatial interactions that may not be fully represented by convolutional networks alone.The extracted global feature representation complements the local texture information obtained from ConvNeXt V2 and provides a more comprehensive understanding of disease characteristics.

E. CROSS-ATTENTION FEATURE FUSION

The proposed VMCAF-Net employs a Cross-Attention Feature Fusion (CAFF) module to adaptively integrate the complementary local features extracted by ConvNeXt V2 and the global contextual features learned by Vision Mamba, generating a more discriminative representation while suppressing irrelevant background information. The fused feature maps are then processed using Adaptive Average Pooling, which reduces spatial dimensions and converts them into compact fixed-length feature vectors with minimal information loss, thereby reducing computational complexity. These feature vectors are passed to a Multi-Layer Perceptron (MLP) classifier, where the final disease prediction is performed using a Label-Smoothed Softmax layer that improves model calibration, reduces overconfidence, and enhances generalization for visually similar disease classes. The network is trained using the Label-Smoothed Cross-Entropy Loss function, which minimizes the difference between predicted probabilities and ground-truth labels while improving robustness against noisy labels. To optimize the model, the AdamW optimizer is employed with a learning rate of 0.0001, a batch size of 32, and 100 training epochs, where decoupled weight decay regularization ensures faster convergence, improved stability, reduced overfitting, and enhanced classification performance.Overall, the combination of an effective preprocessing pipeline, dual feature extraction networks, Cross-Attention Feature Fusion, adaptive pooling, label-smoothed classification, and AdamW optimization enables the proposed VMCAF-Net to achieve robust and accurate sesame leaf disease identification under real-world agricultural conditions.

6. Conclusion

This paper proposed VMCAF-Net, a novel hybrid deep learning framework that integrates Vision Mamba**,** ConvNeXt V2**,** and Cross-Attention Feature Fusion for accurate sesame leaf disease identification. The preprocessing pipeline, including CLAHE, gamma correction, image resizing, and data augmentation, enhanced image quality and improved feature learning. ConvNeXt V2 effectively extracted local texture features, while Vision Mamba captured global contextual information. The Cross-Attention module efficiently fused these complementary features, and the Label-Smoothed Softmax classifier improved disease prediction. In addition, Grad-CAM provided interpretable visual explanations of the classification results. Overall, the proposed framework offers a robust, computationally efficient, and reliable solution for intelligent sesame leaf disease detection, making it suitable for real-time precision agriculture and edge-based deployment.

Limitations & Future Work

Future research will focus on expanding the sesame leaf disease dataset by collecting images from diverse geographical regions and varying environmental conditions to improve model generalization. Multi-modal learning integrating environmental factors such as temperature,

Journal of Computer Science and Intelligent Systems
Volume 1, Issue 1 | July - September 2026 | www.journalindiapublication.com | 2026 Journal India Publication

humidity, rainfall, and soil moisture will be explored to enhance disease prediction. Lightweight model optimization using pruning, quantization, and knowledge distillation will enable deployment on smartphones, drones, and edge devices. Self-supervised learning approaches, such as DINOv2 and MAE, will also be investigated to reduce the need for large, annotated datasets. Furthermore, integrating disease severity estimation, treatment recommendation, and yield prediction with advanced Explainable AI techniques will facilitate the development of intelligent real-time precision agriculture systems.

Reference

"ConvGeM-Next: A Deep Learning Framework for Plant Disease Detection," *Frontiers in Plant Science*, 2026.

A. A. Mamun, M. Zhang, D. Ahmedt-Aristizabal, Z. Hayder, and M. Awrangjeb, "ConMamba: Contrastive Vision Mamba for Plant Disease Detection," *Pattern Recognition*, vol. 176, 113177, 2026.

A. Bera, D. Bhattacharjee, and O. Krejcar, "PND-Net: Plant Nutrition Deficiency and Disease Classification Using Graph Convolutional Network," *Scientific Reports*, vol. 14, 15537, 2024.

A. Dosovitskiy *et al.*, "An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale," *International Conference on Learning Representations (ICLR)*, 2021.

A. S. Abade, P. A. Ferreira, and F. B. Vidal, "Plant Diseases Recognition on Images Using Convolutional Neural Networks: A Systematic Review," *arXiv preprint arXiv:2009.04365*, 2020.

D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *International Conference on Learning Representations (ICLR)*, 2015.

E. A. Nibret, A. M. Mequanenit, A. M. Ayalew, K. Kusrini, and R. Martínez-Béjar, "Sesame Plant Disease Classification Using Deep Convolution Neural Networks," *Applied Sciences*, vol. 15, no. 4, 2124, 2025.

G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," *Proceedings of CVPR*, 2017.

H. Zhang, S. Li, J. Xie, Z. Chen, J. Chen, and J. Guo, "VMamba for Plant Leaf Disease Identification: Design and Experiment," *Frontiers in Plant Science*, 2025.

Journal of Computer Science and Intelligent Systems
JOURNAL India Publication
Volume 1, Issue 1 | July - September 2026 | www.journalindiapublication.com | 2026 Journal India Publication

I. Loshchilov and F. Hutter, "Decoupled Weight Decay Regularization," *International Conference on Learning Representations (ICLR)*, 2019.

J. Chen, J. Guo, H. Zhang, Z. Liang, and S. Wang, "Weakly Supervised Localization Model for Plant Disease Based on Siamese Networks," *Frontiers in Plant Science*, 2024.

J. Deng *et al.*, "ImageNet: A Large-Scale Hierarchical Image Database," *Proceedings of CVPR*, 2009.

K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *Proceedings of CVPR*, 2016.

M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," *Proceedings of ICML*, 2019.

P. Bedi and P. Gole, "Plant Disease Detection Using Hybrid Model Based on Convolutional Autoencoder and Convolutional Neural Network," *Artificial Intelligence in Agriculture*, vol. 5, pp. 90–101, 2021.

P. Bedi, P. Gole, and S. Marwaha, "PDSE-Lite: Lightweight Framework for Plant Disease Severity Estimation Based on Convolutional Autoencoder and Few-Shot Learning," *Frontiers in Plant Science*, 2024.

Q. Zhang *et al.*, "Toward Smart Agriculture: A Hybrid Mamba-Transformer Vision Framework for Plant Disease Detection," *Frontiers in Plant Science*, 2026.

R. R. Selvaraju *et al.*, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," *Proceedings of ICCV*, 2017.

S. Woo, S. Debnath, R. Hu, X. Chen, Z. Liu, I. S. Kweon, and S. Xie, "ConvNeXt V2: Co-designing and Scaling ConvNets with Masked Autoencoders," *arXiv preprint arXiv:2301.00808*, 2023.

S. Zhang and R. Liu, "Multi-scale Feature Fusion-Based Vision Mamba for Robust Plant Disease Image Classification on Field-Acquired PlantDoc Data," *Frontiers in Plant Science*, 2026.

Y. Borhani, J. Khoramdel, and E. Najafi, "A Deep Learning Based Approach for Automated Plant Disease Classification Using Vision Transformer," *Scientific Reports*, vol. 12, 11554, 2022.

Y. Chen *et al.*, "Plant Image Recognition with Deep Learning: A Review," *Computers and Electronics in Agriculture*, vol. 212, 108072, 2023.

Y. Yue and Z. Li, "MedMamba: Vision Mamba for Medical Image Classification," *arXiv preprint arXiv:2403.03849*, 2024.

Z. Liu *et al.*, "A ConvNet for the 2020s," *Proceedings of CVPR*, 2022.

Z. Liu *et al.*, "Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows," *Proceedings of ICCV*, 2021.