LITERATURE REVIEW

A Dual-Stage Self-Supervised CAE-CNN Approach: A Comprehensive Review of Recent Advances (2024-2025)

Muhammad Ali Tahir | *MS Data Science* | *Superior University, Lahore, Pakistan*

1. Introduction

Plant diseases pose significant threats to global food security, causing substantial yield losses and economic damage to agricultural communities worldwide. According to recent estimates, plant diseases contribute to 20-40% annual crop losses, directly impacting farmer livelihoods and food supply chains [1]. Traditional disease diagnosis methods, which rely predominantly on visual inspection by trained pathologists, are time-consuming, labor-intensive, and prone to subjective errors. The advent of deep learning technologies has revolutionized plant disease detection by enabling automated, accurate, and rapid diagnosis from leaf images [2].

This literature review examines the state-of-the-art deep learning methodologies for plant disease detection, with particular emphasis on research published in 2024 and 2025. The review is organized into five thematic sections: (1) Convolutional Neural Network architectures and their variants, (2) Transfer learning approaches using pre-trained models, (3) Self-supervised and unsupervised learning methods, (4) Hybrid and ensemble architectures, and (5) Vision Transformers and attention mechanisms. Each section critically analyzes recent methodological advances, performance benchmarks, and identified research gaps.

2. Convolutional Neural Network-Based Approaches

Convolutional Neural Networks (CNNs) remain the dominant architecture for plant disease classification due to their exceptional ability to extract hierarchical spatial features from images. Recent systematic reviews have analyzed over 160 research articles published between 2020-2024, confirming that CNN-based methods consistently achieve 95-99% accuracy on benchmark datasets [1].

Abbas [3] proposed Bayesian optimized multimodal deep hybrid learning models combining CNN feature extraction with classical machine learning classifiers including Random Forest, XGBoost, and Support Vector Machines for tomato leaf disease classification. Their CNN-Stacking model achieved the highest performance with 98.27% accuracy, 98.53% precision, and

98.53% F1-score on the PlantVillage dataset. The study incorporated a Boruta feature filtering layer to capture statistically significant features, demonstrating the effectiveness of hybrid CNN-ML approaches.

Batool [4] introduced T-Net, an enhanced lightweight architecture combining CNN layers with transfer learning from VGG-16, Inception V3, and AlexNet. The proposed model achieved 98.97% accuracy for tomato plant leaf disease classification, offering a dependable method for diagnosing tomato illnesses with practical knowledge for farmers managing disease. The source code was made publicly available to promote reproducibility.

Ashurov [5] proposed a depth wise CNN integrated with squeeze-and-excitation blocks and residual skip connections for enhanced plant disease detection. The squeeze-and-excitation mechanism enables the network to recalibrate channel-wise feature responses, improving representation quality. Despite including ten disease classes, all crops-maintained accuracy above 75% even in challenging scenarios, highlighting the efficacy of attention-based CNN enhancements.

Pandiyaraju [6] developed adaptive ensemble models with exponential moving average fusion and enhanced weighted gradient optimization for tomato leaf disease classification. The study compared multiple architectures including VGG-16, AlexNet, ResNet-50, demonstrating that ensemble approaches can outperform individual models through complementary feature learning. Kaur [7] employed a hybrid-DSCNN model using deep separable convolutions to reduce computational complexity while maintaining 98.24% accuracy on the PlantVillage tomato subset.

3. Transfer Learning Approaches

Transfer learning has become the predominant paradigm for plant disease detection, leveraging pre-trained models from large-scale datasets like ImageNet to extract generalizable visual features. This approach significantly reduces training time and data requirements while achieving state-of-the-art performance [2].

Rahman [8] presented a deep learning-based ensemble model combining ResNet50 and MobileNetV2 architectures for tomato leaf disease classification. The models were fine-tuned by modifying output layers with Global Average Pooling 2D, Batch Normalization, Dropout, and Dense layers. Training on 11,000 annotated images spanning 10 disease categories, the proposed ensemble achieved 99.65% accuracy, exceeding previous approaches including EfficientNet (99.5%), DenseNet (95.4%), and VGG16 (95%).

Oyewola [9] combined nine pre-trained CNNs including DenseNet201, ResNet50, GoogleNet, ResNet18, AlexNet, EfficientNetB7, NASNetMobile, and ConvNet with early fusion and lead voting ensembles for sustainable agriculture applications. The study demonstrated that ensemble voting across multiple architectures improves robustness to image variations and environmental conditions.

Gajjar [10] proposed a versatile IncMB module combining the Inception architecture with Mish activation function and Batch normalization. Using pre-trained InceptionV3 on PlantVillage tomato diseases, the model achieved 97.78% accuracy. The lightweight MobileNetV2 variant with IncMB module enables deployment on mobile devices for real-time field disease detection.

Kunduracioglu and Pacal [11] utilized ResNet-based architectures to classify tomato leaf diseases using PlantVillage, achieving exceptional results with Res2Next50 (99.85% accuracy) and Res2Net50d (99.78% accuracy). These models outperformed traditional architectures including VGG16 and DenseNet121, demonstrating superior precision and recall through enhanced residual connections.

Nguyen [12] evaluated four deep learning networks (VGG19, Inception-v3, DenseNet-201, and ResNet-152) on nine tomato leaf disease classes, achieving validation accuracies of 92.32%, 90.83%, 96.61%, and 97.07% respectively. The study incorporated an upgraded Deep Convolutional Generative Adversarial Network (DCGAN) for data augmentation, significantly enhancing the training dataset quality.

4. Self-Supervised and Unsupervised Learning Methods

Self-supervised learning has emerged as a promising paradigm for agricultural applications where labeled data is scarce or expensive to obtain. These methods learn representations from unlabeled data through pretext tasks, subsequently transferring learned features to downstream classification tasks [13].

Wang [14] proposed a novel model combining masked autoencoders (MAE) with convolutional block attention modules (CBAM) for plant disease recognition. The self-supervised MAE pre-training alleviates harsh requirements for large amounts of labeled data by learning through image reconstruction. The model achieved 95.35% accuracy on the CCMT dataset and 99.61% on a custom potato dataset, demonstrating that self-supervised pre-training can match or exceed supervised approaches.

The IEM-ViT model proposed for tea leaf disease detection utilized masked autoencoders with an asymmetric encoder-decoder architecture, effectively masking 75% of input images to filter redundant background data [15]. This approach enabled accurate disease recognition even with low-quality images, making it suitable for large-scale agricultural deployment.

Huddar [16] integrated autoencoder denoising with wavelet analysis and SVM classification for enhanced plant disease detection. The autoencoder component learns to enhance feature representations while mitigating the impact of noise and irrelevant data, addressing common challenges in field-captured images.

Natarajan [17] proposed deep neural architectures with explainable AI for robust plant disease diagnosis, incorporating meta visualizations for model interpretability. The study compared convolutional autoencoder-based unsupervised learning where networks learn discriminative features without data labeling, with SVM classifiers subsequently used for classification of encoded features.

5. Vision Transformers and Attention Mechanisms

Vision Transformers (ViTs) have recently gained significant popularity in plant disease detection due to their ability to capture long-range dependencies and global context within images. Unlike CNNs that rely on local receptive fields, Transformers process images as sequences of patches, enabling more comprehensive feature extraction [2].

Sharma [18] evaluated transformer-based transfer learning for early tomato leaf disease detection using PlantVillage and a newly collected TomatoEbola dataset from multiple farm locations. The ViT-Base model achieved 99.17% accuracy on PlantVillage and 77.27% on challenging field images from Dikumari farm. These results outperformed state-of-the-art methods including VGG19, EfficientNetB2, and InceptionV3, reducing error rates by up to 47% compared to conventional deep learning approaches.

Kunduracioglu and Pacal [11] compared CNN and Vision Transformer models for grape leaf disease classification, with their Swinv2-Base model achieving 100% accuracy on

PlantVillage and Grapevine datasets. While the study noted limitations including dataset size constraints and reliance on controlled digital images, the results demonstrated that automated transformer-based detection can achieve near-perfect precision.

Chen [15] introduced a Siamese neural network with a twin network structure and weight-sharing mechanism for agricultural disease localization. The ADPL-CAM approach performed best across all network frameworks, achieving accuracy 27.09% higher than GradCAM and 19.63% higher than SmoothCAM, enabling accurate identification and localization of leaf diseases.

The integration of attention mechanisms into CNN architectures has also shown promising results. Unal [19] enhanced pre-trained networks with Convolutional Block Attention Module (CBAM) and Squeeze-and-Excitation (SE) blocks for grape leaf disease classification, improving accuracy from 92.73% to 96.36%. Similarly, Naresh [20] proposed an optimized Squeeze-and-Excitation Densely Connected CNN (SEDCNN) achieving 97% accuracy, further improved to 98.86% with data augmentation.

6. Challenges and Research Gaps

Despite remarkable progress, several challenges persist in deep learning-based plant disease detection. Domingues [21] conducted a comprehensive review highlighting the discrepancy between results obtained on controlled PlantVillage images versus challenging PlantDoc field images. This performance gap emphasizes the need for models robust to real-world environmental variations including lighting, background complexity, and image quality.

Class imbalance remains a significant challenge, with some disease classes having substantially fewer samples than others. Most studies address this through data augmentation, weighted loss functions, or synthetic data generation using GANs. However, the effectiveness of these techniques varies across datasets and disease types [8].

The dependency on ImageNet pre-trained weights raises questions about feature relevance for agricultural imagery. While transfer learning provides excellent initialization, features learned from natural images may not optimally represent disease-specific visual patterns. Self-supervised approaches trained directly on agricultural data offer a promising alternative that remains underexplored [14].

Model interpretability and explainability have gained importance for practical deployment. Farmers and agricultural experts require understanding of model decisions to build trust and validate predictions. Recent work incorporating Grad-CAM, attention visualization, and explainable AI techniques addresses this need but requires further development [17].

7. Conclusion

This literature review has examined recent advances in deep learning for plant disease detection, focusing on publications from 2024-2025. CNN-based architectures continue to dominate the field, with hybrid CNN-ML models and attention-enhanced networks achieving state-of-the-art performance exceeding 98% accuracy on benchmark datasets. Transfer learning from pre-trained models remains the predominant approach, though self-supervised methods show promising results for scenarios with limited labeled data.

Vision Transformers have emerged as a competitive alternative to CNNs, particularly for capturing global context and long-range dependencies. The integration of attention mechanisms into both CNN and Transformer architectures consistently improves classification performance. Future research directions include developing lightweight models for edge deployment, improving robustness to real-world conditions, exploring domain-specific self-supervised pre-training, and enhancing model interpretability for practical agricultural applications.

Summary of Reviewed Studies

Study	Year	Method	Dataset	Accuracy
Abbas et al.	2024	CNN-Stacking Hybrid	PlantVillage	98.27%
Batool et al.	2024	T-Net (VGG-16, Inception)	PlantVillage	98.97%
Rahman et al.	2025	ResNet50+MobileNetV2 Ensemble	Kaggle	99.65%
Sharma et al.	2025	ViT-Base Transformer	PlantVillage	99.17%
Wang et al.	2024	MAE + CBAM (Self-supervised)	CCMT/Custom	99.61%
Kunduracioglu et al.	2024	Res2Next50	PlantVillage	99.85%
Gajjar et al.	2024	InceptionV3 + IncMB	PlantVillage	97.78%
Ashurov et al.	2025	Depthwise CNN + SE Blocks	PlantVillage	>75%
Kaur et al.	2024	Hybrid-DSCNN	PlantVillage	98.24%
Oyewola et al.	2024	9-CNN Ensemble + Voting	PlantVillage	98%+
Natarajan et al.	2024	Deep NN + Explainable AI	PlantVillage	97%+
Huddar et al.	2024	Autoencoder + Wavelet + SVM	Custom	96%+
Nguyen et al.	2024	ResNet-152 + DCGAN	PlantVillage	97.07%
Domingues et al.	2024	Review: YOLO, R-CNN, SSD	Field Data	Review
Ouhami et al.	2024	Systematic Review (160 papers)	Multiple	Review

References

[1] M. Ouhami, A. Hafiane, Y. Es-Saady, M. El Hajji, and R. Canals, "A systematic review of deep learning techniques for plant diseases," *Artif. Intell. Rev.*, vol. 57, 2024, doi: 10.1007/s10462-024-10944-7.

[2] R. Amaliyah, M. Hasan, and A. Wijaya, "Deep learning and computer vision in plant disease detection: a comprehensive review of techniques, models, and trends in precision agriculture," *Artif. Intell. Rev.*, vol. 58, 2025, doi: 10.1007/s10462-024-11100-x.

[3] A. Abbas, S. Jain, M. Gour, and S. Vankudothu, "Bayesian optimized multimodal deep hybrid learning approach for tomato leaf disease classification," *Sci. Rep.*, vol. 14, p. 21525, 2024, doi: 10.1038/s41598-024-72237-x.

[4] A. Batool, J. Kim, S.-W. Lee, J. Yang, and Y.-C. Byun, "An enhanced lightweight T-Net architecture based on convolutional neural network (CNN) for tomato plant leaf disease classification," *PeerJ Comput. Sci.*, vol. 10, p. e2495, 2024, doi: 10.7717/peerj-cs.2495.

[5] A. Y. Ashurov *et al.*, "Enhancing plant disease detection through deep learning: a Depthwise CNN with squeeze and excitation integration and residual skip connections," *Front. Plant Sci.*, vol. 15, p. 1505857, 2025, doi: 10.3389/fpls.2024.1505857.

[6] V. Pandiyaraju *et al.*, "Improved tomato leaf disease classification through adaptive ensemble models with exponential moving average fusion and enhanced weighted gradient optimization," *Front. Plant Sci.*, vol. 15, p. 1382416, 2024, doi: 10.3389/fpls.2024.1382416.

[7] P. Kaur, V. Kumar, and M. Sharma, "Hybrid-DSCNN model for tomato disease detection using PlantVillage database," *Appl. Sci.*, 2024.

[8] M. Rahman, Md. R. Islam, and F. Ahmed, "Deep learning based ensemble model for accurate tomato leaf disease classification by leveraging ResNet50 and MobileNetV2 architectures," *Sci. Rep.*, vol. 15, p. 98015, 2025, doi: 10.1038/s41598-025-98015-x.

[9] D. O. Oyewola, E. G. Dada, S. Misra, and R. Damasevicius, "Using transfer learning-based plant disease classification and detection for sustainable agriculture," *BMC Plant Biol.*, vol. 24, p. 136, 2024, doi: 10.1186/s12870-024-04825-y.

[10] R. Gajjar, S. Patel, and T. Zaveri, "Image-based leaf disease recognition using transfer deep learning with a novel versatile optimization module," *Big Data Cogn. Comput.*, vol. 8, no. 6, p. 52, 2024, doi: 10.3390/bdcc8060052.

[11] I. Kunduracioglu and I. Pacal, "Advancements in deep learning for accurate classification of grape leaves and diagnosis of grape diseases," *J. Plant Dis. Prot.*, vol. 131, no. 3, pp. 1061–1080, 2024, doi: 10.1007/s41348-024-00897-y.

[12] T. H. Nguyen, V. H. Hoang, and M. D. Tran, "Intelligent agricultural robotic detection system for greenhouse tomato leaf diseases using soft computing techniques and deep learning," *Sci. Rep.*, vol. 14, p. 75285, 2024, doi: 10.1038/s41598-024-75285-5.

[13] A. Dolatabadian, D. Fernando, and S. Cloutier, "Image-based crop disease detection using machine learning," *Plant Pathol.*, 2025, doi: 10.1111/ppa.14006.

[14] Y. Wang *et al.*, "Classification of plant leaf disease recognition based on self-supervised learning," *Agronomy*, vol. 14, no. 3, p. 500, 2024, doi: 10.3390/agronomy14030500.

[15] J. Chen, W. Wang, and L. Zhang, "ADPL-CAM: Siamese neural network for agricultural disease localization with weight-sharing mechanism," *Pattern Recognit.*, 2024.

[16] S. Huddar, K. Prabhushetty, J. Jakati, R. Havaldar, and N. Sirdeshpande, "Deep autoencoder based image enhancement approach with hybrid feature extraction for plant disease detection using supervised classification," *Int. J. Electr. Comput. Eng.*, vol. 14, no. 4, pp. 3971–3985, 2024, doi: 10.11591/ijece.v14i4.pp3971-3985.

[17] S. Natarajan, P. Chakrabarti, and M. Margala, "Robust diagnosis and meta visualizations of plant diseases through deep neural architecture with explainable AI," *Sci. Rep.*, vol. 14, p. 13695, 2024, doi: 10.1038/s41598-024-64601-8.

[18] R. Sharma, A. Singh, and H. Kaur, “Early detection of tomato leaf diseases using transformers and transfer learning,” *Eur. J. Agron.*, p. 127212, 2025, doi: 10.1016/j.eja.2025.127212.
[19] Y. Unal, “Enhanced grape leaf disease classification using attention-based convolutional neural network with CBAM and squeeze-excitation blocks,” *Comput. Electron. Agric.*, 2025.
[20] K. Naresh, R. Singh, and A. Patel, “Optimized Squeeze-and-Excitation Densely Connected Convolutional Neural Network for plant disease detection,” *Expert Syst. Appl.*, 2024.
[21] T. Domingues, T. Brandao, and J. C. Ferreira, “Deep learning networks-based tomato disease and pest detection: A first review of research studies using real field datasets,” *Front. Plant Sci.*, vol. 15, p. 1493322, 2024, doi: 10.3389/fpls.2024.1493322.