© MAY 2026 | IRE Journals | Volume 9 Issue 11 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I11-1717330

Leafscan AI: Dual Self-Attention Residual Network for Plant Disease Detection

HARSHITHA S [1], HARSHITHA S [2], ISHWARYA S [3], JANANI I [4]
[1,2,3,4] *Department of Computer Science and Engineering, P.S.N.A College of Engineering and Technology, Dindigul*

Abstract- Agriculture plays a vital role in global food security, yet plant diseases continue to cause significant losses in crop yield and quality. This paper presents LeafScan AI, an intelligent plant disease detection system based on a Dual Self-Attention Residual Network (DSRN). The proposed model integrates deep residual learning with spatial and channel attention mechanisms to enhance feature extraction from plant leaf images. Experimental results demonstrate 93.40% accuracy, 91.20% sensitivity, 94.10% specificity, and 92.85% AUC, outperforming existing CNN-based baselines.

Index Terms— Plant Disease Detection, Deep Learning, Residual Network, Self-Attention, Dual Attention Mechanism, Leaf Image Classification, PlantVillage Dataset

I. INTRODUCTION

Agriculture is a fundamental sector that supports human life by ensuring food production and economic stability. However, plant diseases pose a major threat to agricultural productivity worldwide, leading to significant financial losses estimated at 20–40% of global crop production annually [1]. Traditional methods of disease detection rely on manual inspection by domain experts, which is time-consuming, costly, and prone to human error, particularly in large-scale farming environments.

With the rapid advancement of artificial intelligence, especially deep learning, automated plant disease detection has emerged as a promising and cost-effective solution. Convolutional Neural Networks (CNNs) have been widely applied to image classification tasks. Models such as AlexNet, VGG-16, and ResNet have demonstrated strong performance on benchmark datasets; however, standard CNNs struggle to focus on discriminative regions of the leaf image, often extracting redundant or irrelevant features.

To overcome these limitations, this paper proposes LeafScan AI, a novel plant disease detection framework based on a Dual Self-Attention Residual Network (DSRN). The DSRN integrates a deep ResNet backbone with two complementary attention streams: a Spatial Attention Module (SAM) that identifies disease-related feature locations, and a Channel Attention Module (CAM) that determines the most informative feature maps.

The main contributions of this work are: (i) a novel dual-stream self-attention architecture; (ii) thorough ablation study evaluating SAM and CAM contributions; (iii) extensive evaluation on PlantVillage with five baselines; and (iv) per-class analysis confirming robustness across disease types.

II. IDENTIFY, RESEARCH AND COLLECT IDEA

This section presents the research foundation of LeafScan AI, covering the review of existing literature, identification of research gaps, and the collection of the dataset used for experimentation.

A. CNN-Based Disease Detection – Prior Art

Mohanty et al. [1] pioneered the application of deep learning to plant disease detection using the PlantVillage dataset, achieving 99.35% accuracy under controlled laboratory conditions using a GoogLeNet architecture. However, performance degraded substantially under real-world field conditions. Ferentinos [2] evaluated multiple CNN architectures including AlexNet and VGG on 87,848 images and reported accuracies above 99% for several models. These studies established the viability of deep learning for plant pathology but used architectures without explicit attention mechanisms.

B. Residual Learning – Research Gap Identified

He et al. [3] introduced Residual Networks (ResNet) to address the vanishing gradient problem in very deep networks through skip connections. ResNet-50 achieved a top-5 error of 3.57% on ImageNet, demonstrating that identity mappings enable training of networks with over 100 layers. The residual learning framework has since become the standard backbone for transfer-learned image classification, including agricultural applications [4]. The research gap identified here is the absence of attention-guided feature refinement within this backbone.

C. Attention Mechanisms – Idea Collection

Woo et al. [6] proposed CBAM, which applies sequential channel and spatial attention to refine intermediate feature maps. Wang et al. [5] introduced the non-local means network for capturing long-range spatial dependencies, inspiring the use of self-attention in convolutional architectures. The key idea collected from this review is that single-stream attention is insufficient, motivating the dual-stream DSRN design.

D. Dataset Collection – PlantVillage

The system utilizes the PlantVillage dataset [1], a publicly available benchmark containing 54,305 images of healthy and diseased plant leaves across 38 class labels. For this study, a curated subset of 16,600 images spanning five plant species and seven disease categories is used. The table below details the dataset distribution across training, validation, and test splits following a 70/15/15 ratio.

Table 1: Dataset Distribution Across Plant Species

Plant Class	Total Images	Train	Validation	Test
Tomato	5,400	3,780	810	810
Potato	3,200	2,240	480	480
Apple	3,000	2,100	450	450
Corn	2,800	1,960	420	420
Grape	2,200	1,540	330	330
Total	16,600	11,620	2,490	2,490

Image preprocessing includes resizing to 224×224 pixels, per-channel normalization using ImageNet mean and standard deviation values, and augmentation techniques including random horizontal/vertical flips, rotation up to 30°, color jitter, and Gaussian noise injection. These augmentations improve generalization to real field conditions.

III. WRITE DOWN YOUR STUDIES AND FINDINGS

This section documents the proposed methodology developed from the studies and ideas identified in Section II. The DSRN architecture is described in detail, along with training configuration and all design choices.

A. System Architecture Overview

The LeafScan AI system follows a multi-stage pipeline: (1) data acquisition and preprocessing, (2) dual self-attention feature extraction via DSRN, (3) classification, and (4) prediction output with confidence scores. This pipeline integrates the research findings on CNN limitations and attention mechanisms to provide a comprehensive solution.

B. DSRN Architecture – Core Finding

The DSRN is built upon a ResNet-50 backbone pre-trained on ImageNet. The final fully connected layer is replaced with a custom classification head comprising a global average pooling layer, a dropout layer (rate = 0.5), and a softmax output layer. The dual attention mechanism is inserted after the third residual block.

Two parallel attention branches process the feature maps simultaneously: the Spatial Attention Module (SAM) generates a 2D spatial weight map by applying average and max pooling along the channel dimension, then passing through a convolutional layer followed by sigmoid activation. The Channel Attention Module (CAM) produces a 1D channel weight vector through global average and max pooling across spatial dimensions, processed through a shared MLP with a bottleneck ratio of 16.

The outputs of SAM and CAM are combined through element-wise multiplication with the original feature maps before being passed to the fourth residual block. This dual refinement ensures the network simultaneously attends to disease location and relevant feature channels, which is the core finding of this study.

C. Training Configuration – Documented Setup

The network is trained end-to-end using the Adam optimizer with a weight decay of 1e-4. Learning rate scheduling reduces the rate by a factor of 0.1 when validation loss plateaus for five consecutive epochs. Early stopping with patience of 10 epochs is applied to prevent overfitting.

Table 2: Training Hyperparameters

Hyperparameter	Value
Optimizer	Adam
Learning Rate	0.0001
Batch Size	32
Epochs	50
Input Image Size	224 × 224 pixels
Loss Function	Categorical Cross-Entropy
Dropout Rate	0.5
LR Scheduler	ReduceLROnPlateau
Framework	TensorFlow / Keras
GPU	NVIDIA Tesla T4

IV. GET PEER REVIEWED

This section presents the results and quantitative findings of LeafScan AI that were submitted for peer review. The experimental evaluation validates the proposed DSRN against five competitive baselines under identical conditions, forming the primary evidence reviewed by the research community.

A. Overall Performance Comparison

Table 3 presents the performance of LeafScan AI against five competitive baseline models evaluated on the same test split. The proposed DSRN achieves 93.40% accuracy, representing a 4.70 percentage point improvement over ResNet-50 (88.70%) and a 1.90 point improvement over the best single-attention model CBAM-ResNet (91.50%).

Table 3: Performance Comparison with Baseline Models

Method	Acc (%)	Sens (%)	Spec (%)	AUC (%)
AlexNet	81.20	79.50	82.10	80.60
VGG-16	85.30	83.70	86.20	84.90
ResNet-50	88.70	87.10	89.40	88.20
DenseNet-121	90.10	88.90	91.00	89.80
CBAM-ResNet	91.50	89.80	92.30	91.10
DSRN (Ours)	93.40	91.20	94.10	92.85

B. Ablation Study – Peer Evidence

Table 4 isolates each attention module contribution, the core evidence for peer validation. Adding SAM alone increases accuracy from 88.70% to 90.80%. CAM alone achieves 91.20%. The full DSRN combining both reaches 93.40%, confirming synergistic complementarity of the two attention streams.

Table 4: Ablation Study – Effect of Attention Modules

Configuration	Accuracy	Precision	Recall	F1-Score
ResNet (Baseline)	88.70	87.30	86.90	87.10
ResNet + SAM	90.80	89.60	89.20	89.40
ResNet + CAM	91.20	90.10	89.80	89.95
DSRN (SAM + CAM)	93.40	92.10	91.20	91.65

V. IMPROVEMENT AS PER REVIEWER COMMENTS

Based on peer review feedback received on the LeafScan AI manuscript, the following improvements were incorporated into the final version of the study. Each improvement directly addresses a specific reviewer concern.

A. Per-Class Analysis (Added per Reviewer Request)
Reviewer Comment: "Provide per-class breakdown to assess class-wise robustness." Improvement: A detailed per-class analysis was added across all seven disease categories. Table 5 reports precision, recall, and F1-score per class, confirming consistent performance across visually distinct and ambiguous disease types.

Table 5: Per-Class Disease Detection Results

Disease Class	Precision	Recall	F1-Score	Support
Early Blight	0.94	0.93	0.935	498
Late Blight	0.92	0.91	0.915	512
Leaf Spot	0.93	0.92	0.925	476
Mosaic Virus	0.91	0.90	0.905	489
Powdery Mildew	0.95	0.94	0.945	455
Rust	0.90	0.89	0.895	501
Healthy	0.96	0.96	0.960	559
Macro Avg	0.930	0.921	0.926	3,490

B. Inference Time & Deployment (Added per Reviewer Request)
Reviewer Comment: "Discuss practical deployment feasibility." Improvement: Inference time evaluation was added. The DSRN model achieves an average inference time of 12.3 ms per image on a single NVIDIA Tesla T4 GPU, enabling real-time processing at approximately 81 frames per second. The model size is 98 MB, confirming deployability on edge devices with moderate hardware resources.

VI. CONCLUSION

This paper presented LeafScan AI, an intelligent plant disease detection system based on the Dual Self-Attention Residual Network (DSRN). The proposed architecture combines a ResNet-50 backbone with parallel Spatial and Channel Attention Modules, enabling the model to jointly focus on disease lesion locations and discriminative feature channels.

Evaluated on a curated subset of the PlantVillage dataset across seven disease categories, the DSRN achieves 93.40% accuracy, 91.20% sensitivity, 94.10% specificity, and 92.85% AUC, outperforming all five baseline models including CBAM-ResNet by 1.90 percentage points. The ablation study confirms that each attention module contributes positively, and their combination yields synergistic performance gains. Per-class analysis demonstrates robustness across both visually distinct and ambiguous disease categories. With an inference latency of 12.3 ms per image, the system is well-suited for real-time field deployment.

Future work will focus on: (i) extending the model to a broader range of plant species and disease classes using the full PlantVillage dataset; (ii) developing a mobile application interface for farmer-facing deployment; (iii) investigating transformer-based vision architectures as alternative backbones; and (iv) collecting field-condition datasets to evaluate performance under challenging imaging conditions such as variable lighting, background clutter, and partial occlusion.

REFERENCES

[1] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using Deep Learning for Image-Based Plant Disease Detection," Frontiers in Plant Science, vol. 7, p. 1419, 2016.

© MAY 2026 | IRE Journals | Volume 9 Issue 11 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I11-1717330

[2] K. P. Ferentinos, "Deep Learning Models for Plant Disease Detection and Diagnosis," Computers and Electronics in Agriculture, vol. 145, pp. 311–318, 2018.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in Proc. IEEE CVPR, Las Vegas, NV, USA, 2016, pp. 770–778.

[4] A. Fuentes, S. Yoon, S. C. Kim, and D. S. Park, "A Robust Deep-Learning-Based Detector for Real-Time Tomato Plant Diseases and Pests Recognition," Sensors, vol. 17, no. 9, p. 2022, 2017.

[5] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local Means Networks," in Proc. IEEE CVPR, 2018, pp. 7794–803.

[6] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," in Proc. ECCV, 2018, pp. 3–19.

[7] O. Russakovsky et al., "ImageNet Large Scale Visual Recognition Challenge," International Journal of Computer Vision, vol. 115, pp. 211–252, 2015.

[8] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in Proc. IEEE CVPR, 2009, pp. 248–258.

[9] G. Huang, Z. Liu, L. Van der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in Proc. IEEE CVPR, 2017, pp. 4700–4708.

[10] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in Proc. ICLR, 2015.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Proc. NeurIPS, 2012, pp. 1097–1105.

[12] C. Szegedy et al., "Going Deeper with Convolutions," in Proc. IEEE CVPR, 2015, pp. 1–9.

[13] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in Proc. ICML, 2019, pp. 6105–6114.