JCTA Journal of Computing Theories and Applications ISSN:3024-9104

Research Article

EDANet: A Novel Architecture Combining Depthwise Separable Convolutions and Hybrid Attention for Efficient Tomato Disease Recognition

Yusuf Ibrahim [1,*], Muyideen O. Momoh [1], Kafayat O. Shobowale [2], Zainab M. Abubakar [1], and Basira Yahaya [1]

[1] Department of Computer Engineering, Ahmadu Bello University, Zaria 810103, Nigeria; e-mail : yibrahim@abu.edu.ng; momuyadeen@gmail.com; zmabubakar@abu.edu.ng; byahaya@abu.edu.ng
[2] Department of Mechatronic Engineering, Air Force Institute of Technology, Kaduna 800282, Nigeria; e-mail : kshobowale@afit.edu.ng
* Corresponding Author : Yusuf Ibrahim

Abstract: Tomato crop yields face significant threats from plant diseases, with existing deep learning solutions often computationally prohibitive for resource-constrained agricultural settings; to address this gap, we propose Efficient Disease Attention Network (EDANet), a novel lightweight architecture combining depthwise separable convolutions with hybrid attention mechanisms for efficient Tomato disease recognition. Our approach integrates channel and spatial attention within hierarchical blocks to prioritize symptomatic regions while utilizing depthwise decomposition to reduce parameters to only 104,043 (multiple times smaller than MobileNet and EfficientNet). Evaluated on ten tomato disease classes from PlantVillage, EDANet achieves 97.32% accuracy and exceptional (~1.00) micro-AUC, with perfect recognition of Mosaic virus (100% F1-score) and robust performance on challenging cases like Early blight (93.2% F1) and Target Spot (93.6% F1). The architecture processes 128×128 RGB images in ~23ms on standard CPUs, enabling real-time field diagnostics without GPU dependencies. This work bridges laboratory AI and practical farm deployment by optimizing the accuracy-efficiency tradeoff, providing farmers with an accessible tool for early disease intervention in resource-limited environments.

Keywords: Attention mechanisms; Computer vision; Deep learning; Depthwise separable convolutions; Lightweight models; Plant disease classification; Precision agriculture; Tomato disease detection.

Received: September, 3rd 2025
Revised: September, 23rd 2025
Accepted: September, 30th 2025
Published: October, 2nd 2025

1. Introduction

Plant diseases pose a major threat to food security worldwide. Pathogens (fungi, bacteria, viruses) and pests can destroy up to 20–40% of crop yields annually, causing economic losses [1]. Tomato (*Solanum lycopersicum*), the second most consumed vegetable globally, is especially susceptible. Outbreaks of diseases such as late blight, leaf curl virus, and bacterial spot can devastate harvests [2], with cumulative losses on the order of tens of billions of dollars over time. Climate change compounds these problems by expanding the range and activity of pathogens, and smallholder farms in developing regions are hardest hit due to a lack of rapid diagnostics.

Traditional plant disease diagnostics rely on visual scouting by experts or slow lab tests, e.g., Polymerase Chain Reaction (PCR), Enzyme-Linked Immunosorbent Assay (ELISA) [3]. These methods are highly specific but suffer from low throughput and delay: manual surveys typically cover <1% of farm areas at great labor cost, and lab assays take days to yield results. Untrained farmers can misdiagnose disease symptoms up to 30% of the time. In contrast, computer vision approaches using deep learning have shown promise: convolutional neural networks (CNNs) routinely exceed 95% accuracy on benchmark datasets like PlantVillage. For example, Mohanty et al. (2016) trained deep CNNs on >50,000 leaf images and achieved ~99% accuracy on many crops, including tomato [4]. State-of-the-art models (ResNet,

EfficientNet, Vision Transformers) can match or exceed human experts on curated images [5]–[7].

However, there is a critical gap between lab benchmarks and field deployment. Modern high-accuracy models are extremely resource-intensive: for instance, EfficientNet-B4 requires ~4 billion FLOPs per inference [8] and >8 GB GPU memory. Models often exceed tens of megabytes and are unsuitable for smartphones or embedded hardware. In practice, most published plant-disease models remain research prototypes because they fail to run on low-power CPUs [9]. In particular, standard CNNs cannot easily run real-time diagnostics on field devices. Moreover, most CNNs treat all regions of an image equally, ignoring the fact that diseases manifest in localized patterns (spots, lesions, chlorosis).

This work addresses the need for a lightweight, attention-guided architecture tailored to on-device plant disease detection. Our goal is an operational system that balances high accuracy, extreme efficiency (model size ~100K parameters), and real-time speed. To achieve this, we propose the Efficient Disease Attention Network (EDANet), which introduces three innovations:

- Hybrid Attention: We combine channel-wise attention (to highlight diagnostically important features) with spatial attention (SA) to localize lesion regions in each EDA block, thereby focusing the network on symptom-relevant areas.
- Depthwise Separable Convolutions: All convolutions are decomposed (depthwise + pointwise) to slash parameter count and computation (cf. MobileNet, [10]) while preserving representational power.
- CPU-Optimized Design: EDANet processes 128×128 RGB images with a small hierarchy of convolutional blocks. The entire model (~104K parameters) runs in only ~23 ms per image on an Intel i5 CPU, enabling real-time inference without GPUs.

The design of EDANet is motivated by the need to balance efficiency and discriminative power in plant disease recognition. Depthwise separable convolutions (DSConv) substantially reduce computational cost and parameter count, enabling deployment on low-power CPUs. However, convolution alone treats all features and spatial regions equally, which limits sensitivity to subtle pathological cues. To address this, we integrate Channel Attention (CA) to emphasize diagnostically relevant feature maps and Spatial Attention (SA) to focus on lesion-specific regions of the leaf. The combination of DSConv, CA, and SA provides a synergistic effect: DSConv ensures efficiency, CA enhances feature selection, and SA improves localization. Together, these components allow EDANet to achieve competitive accuracy while remaining lightweight and suitable for real-time use in resource-constrained agricultural settings. We rigorously benchmark EDANet on 10 tomato disease classes (PlantVillage dataset), reporting accuracy, precision, recall, F1-score, and AUC. It attains 97.3% accuracy with only 104K parameters. It fully separates some diseases (100% F1 for Mosaic virus) and dramatically improves on rare or subtle classes (e.g., Early Blight, Target Spot) compared to larger networks. Hence, EDANet bridges laboratory AI and practical farm deployment by capturing spatial-symptom context and maintaining compactness. The remainder of this paper is organized as follows. Section 2 reviews prior work in plant disease diagnostics and lightweight CNNs. Section 3 is the materials and methods, which details the PlantVillage dataset, EDANet's architecture, and training protocol. Section 4 presents experimental results and discussions. Section 5 concludes with key findings and future directions.

2. Related Work

Traditional plant disease diagnosis has relied on human expertise and laboratory tests. Field scouting by agronomists is time-consuming and error-prone; even experts achieve only about 60–80% accuracy under controlled conditions and suffer from fatigue. Lab-based assays (PCR, ELISA) offer high specificity but require specialized equipment and days for results. Spectral imaging (e.g., chlorophyll fluorescence sensors) can detect stress signals in leaves but demands expensive hardware and remains largely experimental. These methods lack the scalability needed to monitor large or remote farms in real time.

The advent of deep learning sparked many CNN-based approaches in plant pathology. For example, Imanulloh et al. [11] demonstrated the effectiveness of CNNs in plant disease recognition, achieving promising results using relatively simple architectures. Early works, e.g, [4] showed that standard deep learning architectures trained on the PlantVillage dataset can exceed 99% on some classes. However, high accuracy often came with enormous models:

Mohanty et al. used AlexNet (60M parameters) and GoogleNet with 5M parameters. Subsequent efforts incorporated newer networks: for instance,[12] applied EfficientNets (B0–B7) to crops, achieving high accuracy but still with multi-million-parameter models (EfficientNet-B0 ~5.3M). Vision Transformers [13] and self-attention models (e.g., [7], [14]) have demonstrated strong recognition on leaf images by modeling global context, but their large parameter counts and attention computations further restrict edge use.

More recent approaches have explored advanced architectures beyond standard CNNs. Rachman et al. [7] proposed an Enhanced Vision Transformer with transfer learning for rice disease recognition, reporting significant improvements in accuracy. Similarly, Firnando et al. [15] analyzed InceptionV3 and InceptionResNetV2 with data augmentation for rice leaf disease classification, further highlighting the role of augmentation and transfer learning in improving generalization.

To make models deployable on devices, researchers have explored lightweight CNNs in agriculture. MobileNetV2 [10] became a de facto choice: [16], [17] applied MobileNetV2 to tomato images, achieving ~95.3% accuracy with 3.4M parameters. ShuffleNetV2 [18]further reduces compute by channel shuffling, delivering real-time speed on phones but sometimes at the cost of 2–5% lower accuracy on complex diseases. Neural Architecture Search (NAS) approaches [17] optimize for speed, yet often struggle with detecting fine symptoms like leaf mold or minor lesions.

Attention mechanisms have been incorporated to improve symptom awareness. The Convolutional Block Attention Module (CBAM;[19], [20]) applies sequential channel/spatial attention and has been used in crop models to boost accuracy. Vision Transformers inherently apply global self-attention and have shown ~93.5% accuracy on Tomato disease classification tasks [11], [21], but they require even more computation.

Despite these advances, most attention-equipped models remain too heavy for the field. Existing attention modules add hundreds of thousands of parameters per block, and self-attention layers can dominate runtime. In summary, recent studies have demonstrated that both CNNs and attention-based architectures (e.g., CBAM, Vision Transformers) can achieve high accuracy in plant disease recognition. However, most of these models are computationally expensive, with parameter counts in the millions, and rely heavily on GPU acceleration. This creates a 'last-mile' challenge: while accurate in laboratory conditions, such models are often unsuitable for deployment in real-world agricultural environments where devices are CPU-limited and power-constrained. Addressing this gap, our work proposes EDANet, which combines depthwise separable convolutions with lightweight hybrid attention to deliver competitive accuracy at a fraction of the computational cost, enabling real-time diagnosis on standard CPUs.

3. Materials and Methods

This section describes the overall pipeline of the proposed EDANet framework, the dataset used, and the implementation details. Figure 1 shows a schematic diagram of the proposed pipeline.

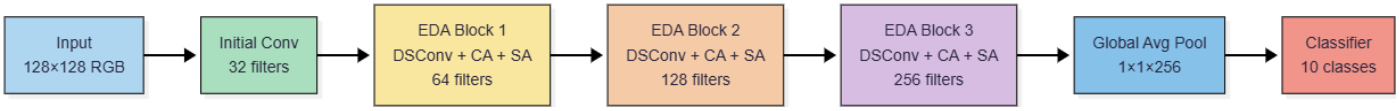

Figure 1. Schematic diagram of the proposed pipeline

3.1. Dataset Description

This work utilized the publicly available PlantVillage Tomato Dataset [4] containing 18,160 RGB images across 10 disease classes, as shown in Table 1.

3.2. Preprocessing Pipeline

The preprocessing pipeline employed in this study was designed for optimal CPU efficiency and robust model generalization. First, all input images were resized to a fixed resolution of 128×128 pixels, ensuring uniform input dimensions and compatibility with the lightweight architecture. Normalization was then applied on a per-channel basis using the mean values [0.485, 0.456, 0.406] and standard deviations [0.229, 0.224, 0.225], consistent with

ImageNet preprocessing standards. To address class imbalance, particularly for underrepresented categories such as Mosaic Virus and Leaf Mold, random oversampling was utilized to ensure equitable representation during training (only). Data augmentation was applied exclusively during the training phase using a sequence of randomized transformations [15]. These included random resized cropping (scale range 0.8–1.0), horizontal flipping (probability 0.5), small-angle rotations (±10°), and mild color jittering (brightness, contrast, and saturation adjustments of ±10%). These augmentations simulate real-world variability in leaf appearance due to environmental factors, while preserving key pathological features. Finally, all images were converted to tensors and normalized using the aforementioned channel statistics.

Table 1. Class distribution and pathology.

Disease Class	Samples	Key Visual Symptoms	Pathogen Type
Bacterial spot	2,127	Small water-soaked lesions, yellow halos	*Xanthomonas spp.*
Early blight	1,000	Concentric rings, target-like spots	*Alternaria solani*
Healthy	1,591	Uniform green coloration, no lesions	-
Late blight	1,909	Water-soaked margins, white sporulation	*Phytophthora infestans*
Leaf Mold	952	Yellow upper surface, grayish mold underside	*Fulvia fulva*
Mosaic virus	373	Mottled light/dark patterns, leaf distortion	*Tobacco mosaic virus*
Septoria leaf spot	1,771	Circular gray spots with dark borders	*Septoria lycopersici*
Spider mites	1,676	Stippling, webbing, bronze discoloration	*Tetranychus urticae*
Target Spot	1,404	Brown spots with concentric rings	*Corynespora cassiicola*
Tomato Leaf Curl Virus	5,357	Upward curling, chlorosis	*Begomovirus*

3.3. EDANet Architecture

3.3.1. Depthwise Separable Convolution Foundation

In EDANet, standard convolutional layers are replaced with depthwise separable convolutions (DSConv), which decompose the operation into two distinct stages to significantly reduce computational complexity and parameter count.

1. Depthwise Convolution performs spatial filtering independently for each input channel using a 3×3 kernel with a stride of 1 and padding of 1. This results in a parameter count of $K \times K \times C_{in}$, in contrast to standard convolutions, which require $\mathrm{K} \times \mathrm{K} \times \mathrm{C_{in}} \times C_{out}$ parameters.
2. Pointwise Convolution follows by applying a 1 × 1 kernel across all channels to combine the outputs of the depthwise stage. This operation learns linear combinations of features across channels and enables the integration of inter-channel information.

The entire transformation can be mathematically expressed as:

$$Output = PointwiseConv_{1\times1}\left(DepthwiseConv_{3\times3}(X)\right) \quad (1)$$

This decomposition yields a dramatic reduction in parameters, quantified by the ratio:

$$\frac{C_{\text{in}} \times K^2 + C_{\text{in}} \times C_{\text{out}}}{C_{\text{in}} \times K^2 \times C_{\text{out}}} = \frac{1}{C_{\text{out}}} + \frac{1}{K^2} \quad (2)$$

For instance, with $C_{\text{out}} = 64$ = and $K = 3$, this translates to an approximate 87% reduction in parameters compared to standard convolution, without significantly compromising representational capacity.

3.3.2. Hybrid Attention Module

To focus the model on salient features associated with tomato disease symptoms, EDANet employs a hybrid attention module composed of two complementary components: channel attention (CA), which emphasizes informative feature maps, and spatial attention,

which highlights relevant regions of the input. Together, they selectively enhance disease-relevant signals while suppressing irrelevant background patterns.

The Channel Attention Submodule recalibrates channel weights to emphasize disease-specific features (e.g., chlorosis in Tomato Leaf Curl). Its implementation is detailed in Algorithm 1 (ChannelAttention).

Algorithm 1. ChannelAttention(x, in_channels, reduction=8)

INPUT: Feature map $x \in \mathbb{R}^{B\times C\times H\times W}$
OUTPUT: Channel-wise refined feature map
$avg_{out} \leftarrow \text{AdaptiveAvgPool2D(x)} \rightarrow \text{shape: } \mathbb{R}^{B\times C\times 1\times 1}$
$max_{out} \leftarrow \text{AdaptiveMaxPool2D(x)} \rightarrow \text{shape: } \mathbb{R}^{B\times C\times 1\times 1}$
Flatten avg_{out} and max_{out} to shape $\mathbb{R}^{B\times C}$
Pass both through shared MLP:
    $h_{avg} \leftarrow \text{ReLU}(\text{Linear}(avg_{out}, \frac{C}{\text{reduction}}))$
    $h_{max} \leftarrow \text{ReLU}(\text{Linear}(max_{out}, \frac{C}{\text{reduction}}))$
    $out_{avg} \leftarrow \text{Linear}(h_{avg}, C)$
    $out_{max} \leftarrow \text{Linear}(h_{max}, C)$
$\text{attention} \leftarrow \text{sigmoid}(out_{avg} + out_{max}) \rightarrow \text{shape: } \mathbb{R}^{B\times C}$
Reshape attention to $\mathbb{R}^{B\times C\times 1\times 1}$
Return: $y = x \odot attention$

3.3.3. Spatial Attention Submodule

The Spatial Attention Submodule highlights symptomatic regions (e.g., concentric rings in Early Blight). Its implementation is detailed in Algorithm 2 (SpatialAttention).

Algorithm 2. SpatialAttention(x)

INPUT: Feature map $x \in \mathbb{R}^{B\times C\times H\times W}$
OUTPUT: Spatially refined feature map
$avg_{out} \leftarrow \text{Mean}(x, dim = channels) \rightarrow \text{shape: } \mathbb{R}^{B\times 1\times H\times W}$
$max_{out} \leftarrow \text{Max}(x, dim = channels) \rightarrow \text{shape: } \mathbb{R}^{B\times 1\times H\times W}$
$concat \leftarrow \text{Concatenate}(avg_{out}, max_{out})\ along\ channel\ dim \rightarrow shape: \mathbb{R}^{B\times 2\times H\times W}$
$\text{attention} \leftarrow \text{sigmoid}(Conv2D(concat, kernel = 7\times 7, padding = 3))$
Return: $y = x \odot attention$

3.4. Implementation Details

EDANet was implemented in PyTorch 2.4.1 and trained on a workstation equipped with an Intel i5 CPU, 16 GB RAM, and Windows 11 Pro. To emphasize the CPU-friendly design, no GPU was used during training, with 1e-3, β_1=0.9, β_2=0.999, and weight decay 1e-4. A ReduceLROnPlateau scheduler was employed to halve the learning rate whenever validation accuracy plateaued for 5 consecutive epochs.

A stratified split of the dataset was used: 70% for training (12,712 images), 15% for validation (2,724 images), and 15% for testing (2,724 images). The total number of trainable parameters in EDANet was 104,043, confirming its lightweight design.

Table 2. Training configuration and hyperparameters used for EDANet experiments.

Parameter	Value / Setting
Dataset	PlantVillage (Tomato)
Image size	128 × 128 RGB
Batch size	16
Epochs	50
Optimizer	Adam (β_1= 0.9, β_2=0.999))
Initial learning rate	0.001

Parameter	Value / Setting
Learning rate schedule	ReduceLROnPlateau (factor = 0.5, patience = 5, mode = max)
Loss function	Cross-Entropy
Weight decay	1e-4
Regularization	Dropout (0.3 in fully connected layer)
Data augmentation	RandomResizedCrop, Horizontal Flip, Rotation (±10°), Color Jitter (±0.1)
Normalization	Mean = [0.485, 0.456, 0.406], Std = [0.229, 0.224, 0.225]

4. Results and Discussion

4.1. Overall Classification Performance

To examine the convergence behavior of EDANet, we recorded the training and validation performance across all 50 epochs. Figure 2 shows the corresponding loss and accuracy curves, which provide insight into the model's optimization stability and generalization.

Figure 2. Training and validation loss (**a**) and accuracy (**b**) curves of EDANet over 50 epochs.

As illustrated in Figure 3, both training and validation accuracy increased steadily, while the loss decreased consistently, indicating effective optimization. Validation accuracy closely followed training accuracy throughout, with no major divergence, suggesting that the model generalized well without overfitting. These results confirm the stability of the training process and provide confidence in the reliability of the reported test performance.

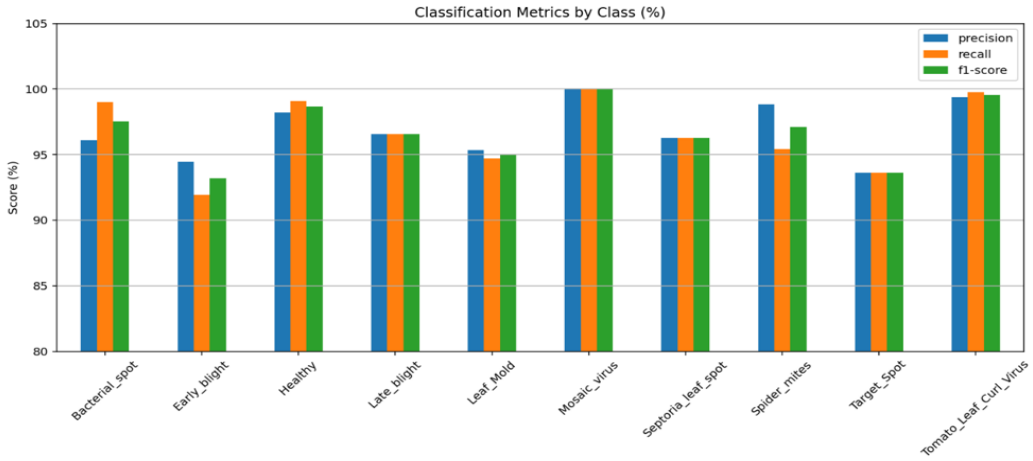

Figure 3. Bar chart showing per-class metrics.

EDANet achieved an overall classification accuracy of 97.30%, with macro-averaged precision, recall, and F1-score of 96.88%, 96.64%, and 96.76%, respectively. These values highlight the model's robustness and balanced performance across all 10 tomato disease classes. The micro-averaged AUC was computed at 0.9993, indicating excellent overall discriminative capacity.

Notably, Mosaic Virus was detected with a 100% F1-score, confirming perfect separation from other classes. Tomato Leaf Curl Virus was the second-best recognized disease, achieving a 99.56% F1-score and 99.75% recall, with only two instances misclassified. These results suggest that EDANet is highly effective at detecting both visually distinct viral patterns and subtle bacterial or fungal infections. The detailed per-class metrics are summarized in Table 3 and Figure 3, showing that all classes achieved F1-scores above 93%, except for minor drops in early blight and target spot, both of which exhibit overlapping visual symptoms such as concentric ring patterns.

Table 3. Class-wise precision, recall, F1-score, AUC, and support.

Class	Precision (%)	Recall (%)	F1-score (%)	AUC	Support
Bacterial_spot	96.10	99.00	97.53	0.999571	299
Early_blight	94.48	91.95	93.20	0.997587	149
Healthy	98.22	99.10	98.66	0.999948	223
Late_blight	96.56	96.56	96.56	0.999107	291
Leaf_Mold	95.33	94.70	95.02	0.999359	151
Mosaic_virus	100.00	100.00	100.00	1.000000	58
Septoria_leaf_spot	96.28	96.28	96.28	0.999606	269
Spider_mites	98.82	95.45	97.11	0.999544	264
Target_Spot	93.64	93.64	93.64	0.997856	220
Tomato_Leaf_Curl_Virus	99.38	99.75	99.56	0.999669	800
Accuracy			**97.32**		
Macro avg	**96.88**	**96.64**	**96.76**		
Weighted avg	**97.32**	**97.32**	**97.31**		**2724**

To further validate the efficiency of EDANet, we compared its performance with state-of-the-art lightweight architectures, namely MobileNetV2 and EfficientNet-B0, on the same PlantVillage tomato dataset. Table 4 summarizes accuracy, F1-score, parameter count, and CPU inference time.

Table 4. Performance of EDANet compared with MobileNetV2 and EfficientNet-B0 on the PlantVillage tomato dataset.

Model	Accuracy (%)	F1-score (%)	Parameters (M)	CPU Inference (ms)
MobileNetV2	99.08	99.08	2.24	32.41
EfficientNet-B0	99.01	99.01	4.02	36.42
EDANet	97.32	97.31	0.10	23

Although MobileNetV2 and EfficientNet-B0 achieved slightly higher accuracy (≈99%) compared to EDANet (97.3%), their parameter counts were more than 20–40× larger, and inference times were slower. By contrast, EDANet maintained competitive accuracy with only 0.10M parameters and the fastest CPU inference (23 ms), making it particularly suitable for real-time deployment in resource-constrained environments.

To further validate the design of EDANet, we conducted an ablation study to isolate the contribution of each architectural component. Four variants were trained for 10 epochs (for efficiency) under identical settings: (i) DSConv only, (ii) DSConv + Channel Attention (CA), (iii) DSConv + Spatial Attention (SA), and (iv) DSConv + CA + SA (full EDANet). Table 5 presents the results across accuracy, precision, recall, F1-score, and AUC. The baseline (DSConv only) achieved good performance, but the addition of either CA or SA improved class discrimination. The full combination of DSConv + CA + SA produced the highest performance, confirming the synergistic benefit of hybrid attention.

Table 5. Ablation study results after 10 epochs of training on the PlantVillage tomato dataset.

Model Variant	Test Accuracy (%)	Test Loss	AUC (Micro)	AUC (Macro)	Precision (Macro) (%)	Recall (Macro) (%)	F1-Score (Macro) (%)	Trainable Parameters
DSConv_only	91.45	0.2499	0.9962	0.9949	89.35	89.26	89.10	81,738
DSConv_CA	93.76	0.1918	0.9976	0.9969	92.64	91.53	91.85	103,746
DSConv_SA	93.39	0.1922	0.9976	0.9966	91.86	90.53	91.03	82,035
Full_EDANet	94.57	0.1676	0.9982	0.9967	93.01	93.30	93.13	104,043

From Table 5, the trends are clear and consistent: CA enhances channel-level feature weighting, SA emphasizes symptomatic spatial regions, and their combination maximizes discriminative capacity. This confirms the effectiveness of EDANet's hybrid attention design.

4.2. Confusion Matrix Analysis

The confusion matrix (Figure 4) reveals specific patterns of misclassification. The highest confusion was observed between Early Blight and Target Spot, with 12–14 misclassified samples for these two classes. This is consistent with prior findings that these diseases share similar ring-like lesion structures, making them difficult to differentiate even for human experts. Other minor misclassifications include:

- Leaf Mold occasionally confused with Septoria Leaf Spot due to shared speckling patterns.
- Tomato Leaf Curl Virus maintained a 99.00% precision, indicating very low false positives.

Although some misclassifications occurred, most classes recorded low error counts, and only a few exceeded 10, highlighting the model's robust generalization performance.

Figure 4. Heatmap of the confusion matrix with 10-class prediction layout.

4.3. ROC and AUC Analysis

To evaluate the model's class-separation capabilities beyond threshold-dependent metrics like accuracy, we computed ROC curves and AUC scores for each disease class (Figure

5). Remarkably, each class achieved an AUC of approximately 1.00 (see actual values in Table 3), indicating near-perfect separability between positive and negative instances in the one-vs-rest classification setting.

This exceptional performance demonstrates that the model consistently ranks true class instances higher than non-target ones, even in cases where final predictions may occasionally be incorrect. It confirms the model's high discriminative ability across all disease types.

The macro-average AUC (equal weighting of all classes) and the micro-average AUC (instance-weighted average) were both approximately 1.00, emphasizing EDANet's outstanding ability to distinguish between classes with near-perfect ranking precision. These results complement the confusion matrix findings and reinforce the model's robustness and generalization strength in real-world diagnostic scenarios.

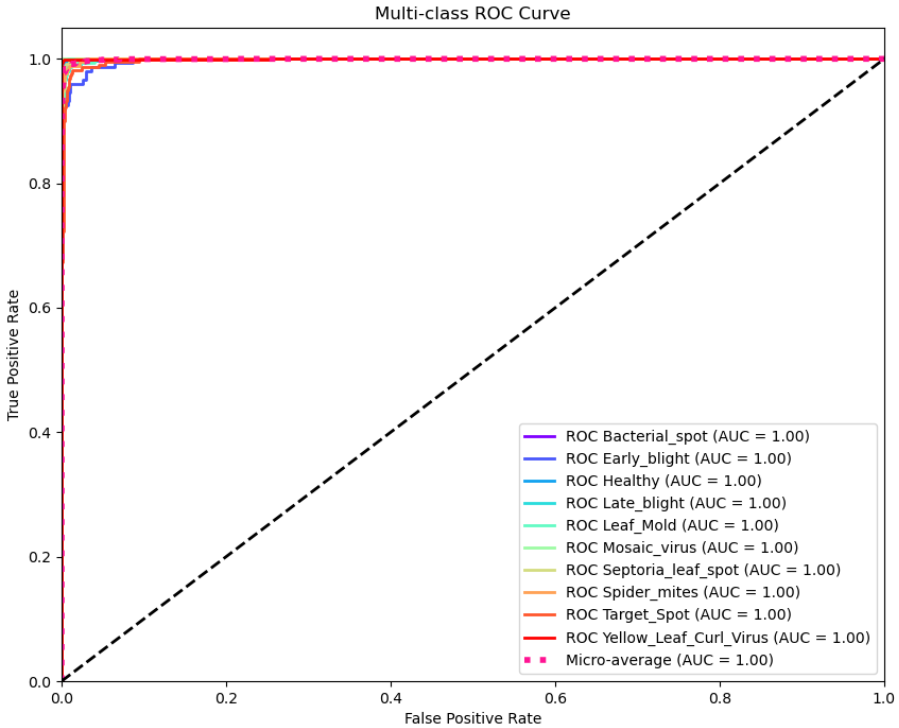

Figure 5. ROC curves for each class along with micro- and macro-average AUCs.

5. Conclusions

The objective of this study was to design a lightweight yet accurate model for tomato leaf disease recognition that could be deployed on CPU-only environments. To this end, we proposed EDANet, which integrates depthwise separable convolutions with channel and spatial attention mechanisms. The model achieved competitive performance while maintaining only 104,043 trainable parameters and the fastest inference speed among compared baselines, demonstrating its suitability for real-time applications in resource-constrained settings.

While EDANet shows promising results, this work has certain limitations. Evaluation was restricted to the PlantVillage tomato dataset, which, although widely used, represents controlled conditions rather than field environments. In addition, the study was limited to static images, without considering temporal dynamics that may provide richer diagnostic cues.

Future work will address these limitations by extending EDANet to other crops and multi-disease scenarios, validating performance on larger and more diverse real-world datasets, and exploring its integration with temporal or video-based monitoring systems for precision agriculture.

Author Contributions: Conceptualization: Y.I. and B.Y.; Methodology: Y.I.; Software: Y.I.; Validation: Y.I., B.Y., and Z.M.A.; Formal analysis: Y.I.; Investigation: Y.I.; Resources: K.O.S.; Data curation: Z.M.A.; Writing—original draft preparation: Y.I.; Writing—review

and editing: B.Y., Z.M.A., M.O.M., and K.O.S.; Visualization: Y.I.; Supervision: K.O.S.; Project administration: B.Y.; Funding acquisition: Y.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Nigeria Artificial Intelligence Research Scheme (NAIRS), an initiative of the Federal Ministry of Communications, Innovation, and Digital Economy, Federal Republic of Nigeria in collaboration with NITDA.

Data Availability Statement: The tomato disease image data used in this study were obtained from the publicly available PlantVillage dataset, which can be accessed at https://www.kaggle.com/datasets/emmarex/plantdisease. All data used were pre-existing and are openly available for research purposes.

Acknowledgments: The authors gratefully acknowledge the administrative and technical support provided by the Nigeria Artificial Intelligence Research Scheme (NAIRS) and Lagos Business School. Their resources were instrumental in facilitating this research on AI-driven crop disease detection. No external AI tools were used in the preparation of this manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] Y. Gai and H. Wang, "Plant Disease: A Growing Threat to Global Food Security," *Agronomy*, vol. 14, no. 8, p. 1615, Jul. 2024, doi: 10.3390/agronomy14081615.

[2] S. Panno *et al.*, "A Review of the Most Common and Economically Important Diseases That Undermine the Cultivation of Tomato Crop in the Mediterranean Basin," *Agronomy*, vol. 11, no. 11, p. 2188, Oct. 2021, doi: 10.3390/agronomy11112188.

[3] J.-T. Chen, M. Khan, A. Parveen, and J. K. Patra, Eds., *Molecular and Biotechnological Tools for Plant Disease Management*. Singapore: Springer Nature Singapore, 2025. doi: 10.1007/978-981-97-7510-1.

[4] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using Deep Learning for Image-Based Plant Disease Detection," *Front. Plant Sci.*, vol. 7, Sep. 2016, doi: 10.3389/fpls.2016.01419.

[5] D. Mohnish Kumaar and S. Palani, "ResNet50 Integrated Vision Transformer for Enhanced Plant Disease Classification," in *2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT)*, May 2024, pp. 1–6. doi: 10.1109/AIIoT58432.2024.10574771.

[6] Santosh Kumar Upadhyay and Rajesh Prasad, "Efficient-ViT B0Net: A high-performance light weight transformer for rice leaf disease recognition and classification," *J. Niger. Soc. Phys. Sci.*, p. 2940, Nov. 2025, doi: 10.46481/jnsps.2025.2940.

[7] R. K. Rachman, D. R. I. M. Setiadi, A. Susanto, K. Nugroho, and H. M. M. Islam, "Enhanced Vision Transformer and Transfer Learning Approach to Improve Rice Disease Recognition," *J. Comput. Theor. Appl.*, vol. 1, no. 4, pp. 446–460, Apr. 2024, doi: 10.62411/jcta.10459.

[8] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," *arXiv*, May 2019, [Online]. Available: http://arxiv.org/abs/1905.11946

[9] A. Musa, M. Hassan, M. Hamada, and F. Aliyu, "Low-Power Deep Learning Model for Plant Disease Detection for Smart-Hydroponics Using Knowledge Distillation Techniques," *J. Low Power Electron. Appl.*, vol. 12, no. 2, p. 24, Apr. 2022, doi: 10.3390/jlpea12020024.

[10] A. G. Howard *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *arXiv*, Apr. 2017, [Online]. Available: http://arxiv.org/abs/1704.04861

[11] S. B. Imanulloh, A. R. Muslikh, and D. R. I. M. Setiadi, "Plant Diseases Classification based Leaves Image using Convolutional Neural Network," *J. Comput. Theor. Appl.*, vol. 1, no. 1, pp. 1–10, Aug. 2023, doi: 10.33633/jcta.v1i1.8877.

[12] B. T. Hanh, H. Van Manh, and N.-V. Nguyen, "Enhancing the performance of transferred efficientnet models in leaf image-based plant disease classification," *J. Plant Dis. Prot.*, vol. 129, no. 3, pp. 623–634, Jun. 2022, doi: 10.1007/s41348-022-00601-y.

[13] A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," *arXiv*, Jun. 2021, [Online]. Available: http://arxiv.org/abs/2010.11929

[14] I. Haider *et al.*, "Crops Leaf Disease Recognition From Digital and RS Imaging Using Fusion of Multi Self-Attention RBNet Deep Architectures and Modified Dragonfly Optimization," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 17, pp. 7260–7277, 2024, doi: 10.1109/JSTARS.2024.3378298.

[15] F. M. Firnando, D. R. I. M. Setiadi, A. R. Muslikh, and S. W. Iriananda, "Analyzing InceptionV3 and InceptionResNetV2 with Data Augmentation for Rice Leaf Disease Classification," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 1, pp. 1–11, May 2024, doi: 10.62411/faith.2024-4.

[16] E. S. Kumar, A. Kumar, H. Vardhan, R. Prasanth, and P. P. Sai, "Classification of Healthy and Unhealthy Leaf Disease Using Modified MobileNetV2," in *2024 5th International Conference on Data Intelligence and Cognitive Informatics (ICDICI)*, Nov. 2024, pp. 482–488. doi: 10.1109/ICDICI62993.2024.10810763.

[17] A. O. Adedoja, P. A. Owolawi, T. Mapayi, and C. Tu, "Intelligent Mobile Plant Disease Diagnostic System Using NASNet-Mobile Deep Learning," *IAENG Int. J. Comput. Sci.*, vol. 49, no. 1, pp. 1–23, 2022, [Online]. Available: https://www.iaeng.org/IJCS/issues_v49/issue_1/IJCS_49_1_23.pdf

[18] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design," *J. Plant Dis. Prot.*, vol. 129, pp. 623–634, Jul. 2018, [Online]. Available: http://arxiv.org/abs/1807.11164

[19] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," *arXiv*. Jul. 18, 2018. [Online]. Available: http://arxiv.org/abs/1807.06521

[20] M. N. Kathiravan, K. G, G. Srilakshmi, G. N. Raju, D. Suganthi, and N. Gandhewar, "Leveraging IoT & CNN-CBAM for Early Blight Detection and Optimal Fungicide Application in Potato and Tomato Crops," in *2024 International Conference on Distributed Systems, Computer Networks and Cybersecurity (ICDSCNC)*, Sep. 2024, pp. 1–6. doi: 10.1109/ICDSCNC62492.2024.10939352.

[21] A. Gangwar, V. S. Dhaka, G. Rani, S. Khandelwal, E. Zumpano, and E. Vocaturo, "Time and Space Efficient Multi-Model Convolution Vision Transformer for Tomato Disease Detection from Leaf Images with Varied Backgrounds," *Comput. Mater. Contin.*, vol. 79, no. 1, pp. 117–142, 2024, doi: 10.32604/cmc.2024.048119.