Channel-spatial dual-attention for plant disease detection: CBAM-ECA integrated CNN models with visual explainability

Anton [a,b,1,*], Supriyadi Rustad [a,2], Guruh Fajar Shidik [a,3], Abdul Syukur [a,4]

[a] Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang 50131, Indonesia
[b] Faculty of Information Technology, Universitas Nusa Mandiri, Jakarta Timur 13620, Indonesia
[1] anton@nusamandiri.ac.id; [2] srustad@dsn.dinus.ac.id; [3] guruh.fajar@research.dinus.ac.id; [4] abah.syukur01@dsn.dinus.ac.id
* corresponding author

ARTICLE INFO

Article history
Received January 19, 2026
Revised March 28, 2026
Accepted March 28, 2026
Available Online August 31, 2026

Keywords
Plant disease detection
DenseNet121
Dual-attention
CBAM-ECA
Visual explainability

ABSTRACT

Early detection of plant leaf diseases is critical for minimizing crop losses and supporting precision agriculture. While Convolutional Neural Networks (CNNs) have demonstrated high accuracy in image-based diagnosis, conventional architectures may not optimally balance spatial localization and channel-wise feature refinement, particularly in multi-crop classification settings. This study proposes a redundancy-aware dual-attention architecture, termed ATSA-DenseNet, which integrates the spatial branch of the Convolutional Block Attention Module (CBAM-Spatial) with Efficient Channel Attention (ECA) within a DenseNet121 backbone. Unlike prior dual-attention frameworks that retain full CBAM and introduce channel-level redundancy, the proposed design isolates complementary spatial and channel mechanisms to improve representational efficiency without increasing computational complexity. The framework is evaluated on controlled multi-crop PlantVillage-derived datasets comprising tomato, potato, pepper, and maize. Across both 3-crop and 4-crop configuration, ATSA-DenseNet consistently outperforms baseline DenseNet121 and single-attention variants, achieving 99.94% accuracy and 0.9994 macro-F1 on the 4-crop setting while maintaining a lightweight footprint (6.96M parameters, 2.87G FLOPs). Grad-CAM visualizations indicate improved localization of disease-relevant regions compared to the baseline. While results are obtained under controlled imaging conditions, the findings demonstrate that redundancy-aware dual-attention enhances feature discrimination efficiency in multi-class agricultural classification tasks. Future work will extend validation to real-field datasets with natural variability.

1. Introduction

Agriculture remains foundational to global food security. Horticultural crops, tomato (Solanum lycopersicum), potato (S. tuberosum), pepper (Capsicum spp.), and maize (Zea mays) suffer 20–40% annual yield losses due to phytopathogens [1], [2]. Diseases like early/late blight, bacterial spot, and northern maize leaf blight destabilize supply chains, especially where diagnostic infrastructure is limited [3], [4], [5]. Manual inspection is subjective and infeasible at scale due to illumination shifts and subtle symptoms [6]. Deep learning, particularly CNNs, enables automated diagnosis with more than 90% accuracy [7], [8], [9]. Modern architectures such as DenseNet, EfficientNet, and MobileNet achieve more than 97% accuracy while enabling deployment on resource-constrained devices [10], [11], [12], [13]. In parallel, attention mechanisms have been introduced to enhance feature representation by guiding models to focus on disease-relevant regions [14], [15]. For instance, CBAM improves spatial localization in plant disease detection tasks [16], [17], [18], while Efficient Channel Attention (ECA) provides lightweight channel-wise refinement with minimal computational overhead [19], [20].

Table 1 provides a clearer overview of recent developments in attention-based approaches for plant disease detection. The comparison highlights the attention mechanisms, model architectures, datasets, target crops and diseases, and evaluation performance reported in previous research. This synthesis provides a concise view of current methodological trends and helps identify research gaps that motivate the proposed approach.

Table 1. Comparative Summary of Recent Studies on Attention Mechanisms for Plant Disease Detection.

Ref.	Base Model	Attention Mechanism	Application Domain / Dataset	Accuracy	Key Findings
[1]	MobileNet (Depthwise Separable CNN)	Augmented Attention Layer	Public Rice Disease dataset (4 classes)	94.65%	Attention enables efficient, mobile-compatible lesion detection.
[18]	Inception + ResNeXt Hybrid	CBAM-based Ternary Parallel Attention	Apple leaf diseases images	98.71%	Multi-scale attention ensures robust lesion detection.
[31]	DenseNet121	CBAM + Squeeze-and-Excitation (SE)	Grapevine dataset (3 classes: blister mite, downy mildew, healthy)	96.36%	Dual-attention optimizes DenseNet121.
[32]	ResNet-based CNN	Spatial–channel hybrid attention	Tomato & pepper leaf disease dataset	99%	Attention enables robust lesion localization.
[16]	MobileNetV2	CBAM (Channel + Spatial)	Citrus Huanglongbing dataset (3 severity levels, 6008 images)	98.75%	CBAM enhances fine-grained features and lightweight model accuracy.
[28]	YOLOv5	Hybrid Attention: CBAM + ECA	UAV imagery of Chinese fir seedlings; tiny-bud detection	Precision 95.55%, Recall 95.84%, F1-Score 96.54%	Attention boosts small-target accuracy by suppressing background interference.
[23]	DenseNet121	CBAM	PlantVillage: Tomato, Potato, Pepper, Maize (10 disease classes)	98.88% (4-crop), 98.23% (3-crop)	CBAM optimizes feature discrimination for multi-crop classification.
[33]	DenseNet (enhanced)	Attention Mechanism	Date Fruit Dataset (10 varieties)	98.05%	Attention surpasses CNNs in robustness.
[34]	Dilated DenseNet	Attention Module + Weighted Feature Fusion	Mendeley Medicinal Leaf Dataset (6 disease classes, 3838 images)	90.69%	Dilated attention enhances subtle texture classification.
[35]	Lightweight ResNet-style CNN (20 layers)	CBAM, SE, Self-Attention, Dual-Attention	PlantVillage Tomato (10 classes)	CBAM 99.69%, SA 99.34%	CBAM achieves an optimal accuracy-efficiency balance.
This work	ATSA-DenseNet (CBAM-Spatial + ECA)	Redundancy-aware dual-attention (decoupled CBAM)	Multi-crop PlantVillage dataset (Tomato, Potato, Pepper, Maize)	-	Introduces a redundancy-aware dual-attention design combining CBAM spatial attention with ECA, reducing redundant channel operations while preserving complementary feature refinement across multiple crop species.

Despite these advances, several limitations remain. First, many existing dual-attention frameworks combine full CBAM with additional channel attention mechanisms such as ECA, which may introduce redundant channel refinement and unnecessary computational overhead [21], [22], [23]. Second, most studies evaluate attention-enhanced models on single-crop datasets or without systematically isolating the contribution of attention modules under controlled backbone comparisons [24], [25]. Furthermore, prior works often employ either heavyweight detection architectures (e.g., YOLO-based models) [26], [27] or do not consider multi-crop generalization, limiting their applicability in practical agricultural settings [28], [29], [30].

To address these limitations, we propose ATSA-DenseNet, a redundancy-aware dual-attention framework that explicitly decouples CBAM by retaining only its spatial branch while integrating Efficient Channel Attention (ECA) for channel-wise feature recalibration. This design eliminates redundant channel operations while preserving complementary spatial and channel feature refinement pathways. By embedding ECA modules after each dense block and applying CBAM-Spatial at the final feature stage, the proposed architecture enhances feature discrimination without increasing model complexity [36].

Unlike prior studies that primarily emphasize accuracy improvements, this work focuses on representational efficiency, architectural minimalism, and systematic ablation under identical training conditions across multiple crop species. DenseNet121 is adopted as the backbone due to its dense connectivity [31], [37] which promotes feature reuse and stabilizes gradient propagation across diverse lesion patterns [38]. We evaluate the proposed model on two benchmark configurations: a 3-crop dataset (tomato, potato, pepper) and an extended 4-crop dataset that additionally includes maize. Experimental results show that the proposed model achieves 99.94% accuracy and a macro F1-score of 0.9994, outperforming baseline and single-attention variants without increasing the number of parameters or FLOPs [39], [40]. Comparative evaluation against MobileNetV2 [41], ResNet50 [32], and EfficientNetB0 [42] further demonstrates the proposed architecture's competitiveness. Grad-CAM visualizations indicate improved localization of disease-relevant regions [43], [44] while the lightweight design (6.96M parameters and 2.87G FLOPs) supports deployment in resource-constrained agricultural environments.

The main contributions of this work are summarized as follows:

- A lightweight redundancy-aware dual-attention architecture that integrates CBAM-Spatial and ECA to jointly improve spatial localization and channel-wise feature discrimination for multi-crop plant disease detection.
- Systematic cross-crop evaluation, demonstrating consistent performance across four agronomically distinct crop species with diverse disease characteristics.
- Efficiency-preserving performance improvement, where the proposed architecture achieves 99.94% accuracy and a macro F1-score of 0.9994 while maintaining identical parameter count and FLOPs compared to the baseline model.
- Interpretable and deployable model design, where Grad-CAM visualization provides qualitative insight into model attention behavior, and the lightweight architecture (6.96M parameters, 2.87G FLOPs) supports potential deployment on edge devices in agricultural monitoring systems.

These contributions collectively show that redundancy-aware dual-attention can improve feature discrimination efficiency in multi-crop plant disease classification while maintaining a lightweight, computationally efficient architecture.

2. Method

2.1. Dataset and Preprocessing

We utilize PlantVillage-derived datasets (Table 2) consisting of two configurations: (1) a 3-crop dataset including tomato, potato, and pepper (8 classes; 9,127 images), and (2) a 4-crop dataset that additionally includes maize (10 classes; 11,274 images). The data were partitioned using stratified random sampling into 70% training, 15% validation, and 15% testing subsets to preserve class distribution. No image overlap was allowed between training, validation, and test sets. Images were resized to 224×224×3 and normalized m using ImageNet statistics. Training augmentations included random rotation (±20°), horizontal flipping (p=0.5), shear (±10°), and color jitter (brightness ±0.3, contrast ±0.2). Validation and test images were resized and normalized only. The dataset distribution was further analyzed to provide descriptive statistics across classes. The number of images per class ranges from 152 (minimum) to 1909 (maximum), with a mean of 1127.4, median of 1000.0, and standard deviation of 469.907839664096. This indicates a moderately imbalanced class distribution across the dataset.

Table 2. Dataset Configuration and Data Partitioning.

Dataset	Crops	Classes	Train	Validation	Test	Total Images
3-crop	Tomato, Potato, Pepper	8	6.387	1.370	1.370	9.127
4-crop	Tomato, Potato, Pepper, Maize	10	7.890	1.692	1.692	11.274

2.2. Proposed ATSA-DenseNet Architecture

We propose ATSA-DenseNet, a lightweight dual-attention variant of DenseNet121 for efficient multi-crop plant disease classification. Our model integrates Efficient Channel Attention (ECA) and the spatial branch of CBAM to jointly refine channel and spatial features without redundancy. We insert ECA modules after each of the four dense blocks to capture cross-channel dependencies via lightweight 1D convolutions on global average-pooled features. While the original ECA uses an adaptive kernel size based on channel count, we adopt a fixed kernel size of 3, empirically found to yield comparable performance with a simpler implementation and consistent behavior across dense blocks. To avoid overlap with ECA, we exclude CBAM's channel attention and retain only its spatial module, placed once at the end of the feature extractor (after norm5). It suppresses background clutter by generating a spatial map from concatenated channel-wise average- and max-pooled features via a 7×7 convolution and sigmoid.

Formally, for input $X \in \mathbb{R}^{C\times H\times W}$:

$$X_{\text{spatial}} = X \otimes \sigma(f^{7x7}([\text{AvgPool}_c(X); \text{MaxPool}_c(X)])) \tag{1}$$

and each ECA module produces:

$$X_{\text{channel}} = X \otimes \sigma(\text{Conv1D}_3(\text{GAP}(X))), \tag{2}$$

where $\otimes$denotes broadcasting multiplication.

As illustrated in Fig. 1, the proposed ATSA-DenseNet framework follows a seven-stage pipeline designed to ensure robustness, reproducibility, and interpretability. The process begins with dataset collection (A), followed by preprocessing and data augmentation (B), and splitting the dataset into training, validation, and test sets (C). Model variants are then constructed for ablation analysis (D), including the DenseNet121 baseline, ECA-enhanced, CBAM-Spatial, and the proposed dual-attention configuration integrating ECA and CBAM-Spatial. All models are trained under identical optimization settings (E) using the Adam optimizer (learning rate = 1×10^{-4}), StepLR scheduling (γ = 0.5 every 15 epochs), and early stopping with a patience of 10 epochs. Grad-CAM is subsequently applied to visualize attention regions and provide qualitative interpretability (F). Finally, model performance is evaluated on the held-out test set (G). This pipeline ensures systematic comparison of architectural variants under identical training conditions.

ATSA-DenseNet integrates ECA after each dense block and applies the spatial branch of CBAM after the Norm5 layer (Fig. 2). The CBAM-Spatial module generates spatial attention by applying channel-wise average and max pooling, concatenating the results, and processing them through a 7×7 convolution followed by a sigmoid activation to produce a spatial attention map. The ECA module enhances channel relationships using a lightweight one-dimensional convolution across the channel dimension with an adaptively determined kernel size. The resulting feature maps are then aggregated via global average pooling, regularized with dropout (rate = 0.5), and classified with a Softmax layer.

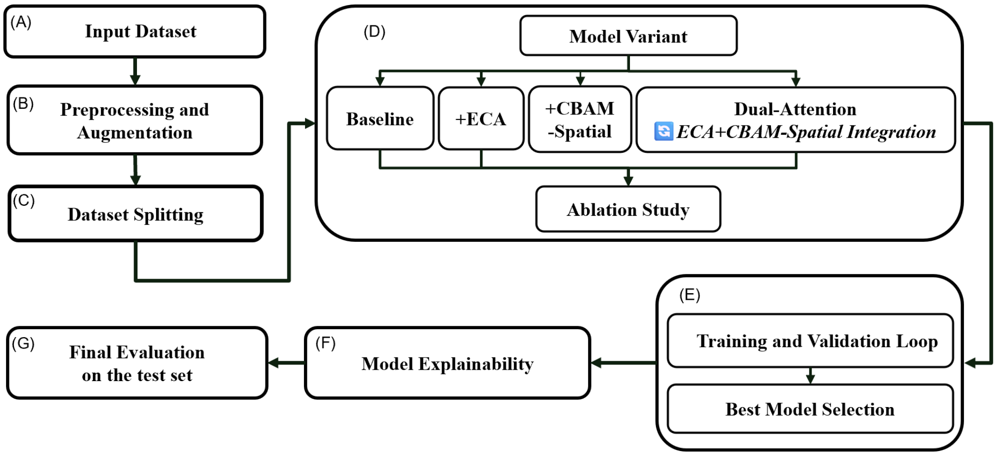


Fig. 1. Overall workflow of the proposed ATSA-DenseNet framework for multi-crop plant disease classification.

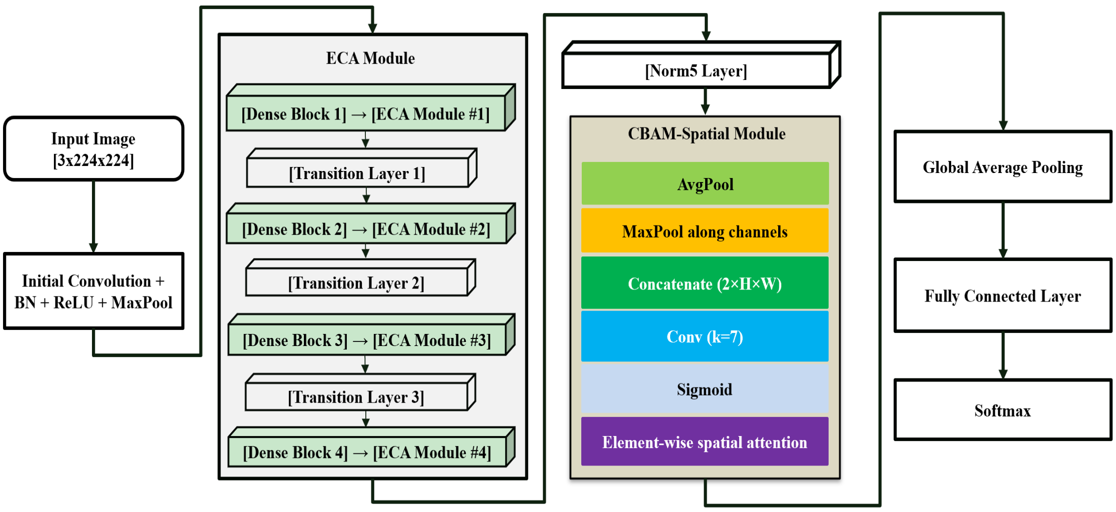


Fig. 2. Dual-Attention DenseNet121 Architecture Diagram.

2.3. Training and Optimization Protocol

All models were trained end-to-end under identical optimization settings to ensure fair comparison. We employed the Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and weight decay = 1×10^{-4}. The initial learning rate was set to 1×10^{-4} and reduced by a factor of 0.5 every 15 epochs using StepLR scheduling. Training was conducted for a maximum of 25 epochs with a batch size of 16. Early stopping (patience = 10) was applied based on validation loss to mitigate overfitting. A fixed random seed (42) was used to ensure reproducibility. All layers, including the ImageNet-pretrained DenseNet121 backbone and the inserted attention modules, are fine-tuned jointly throughout training. Ablation studies are performed on four DenseNet121 variants: baseline, +ECA, +CBAM-Spatial, and +ECA+CBAM-Spatial. We also benchmark MobileNetV2, ResNet50, and EfficientNetB0 under identical training conditions.

2.4. Evaluation Metrics

We evaluate model performance only on the held-out test set using the following methods [18]:

- Accuracy,

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

Proportion of correctly classified samples.

- Per-class and macro-averaged F1-score,

$$\text{F1} = 2 \times \frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \tag{4}$$

We report both per-class F1 and macro-averaged F1 (arithmetic mean across classes), which treats all classes equally, critical under class imbalance.

- Precision and recall per class,

$$\text{Precision}\chi = \frac{TP\chi}{TP\chi+FP\chi} \tag{5}$$

Reliability of positive predictions; critical for minimizing false alarms for specific diseases.

- $\text{Recall}_{\chi} = \frac{\text{TP}_{\chi}}{\text{TP}_{\chi}+FN_{\chi}}$ (6)

For each class c, we compute a binary Precision-Recall curve by varying the decision threshold on the softmax output. The Area Under the PR Curve (AUC-PR) is then calculated via the trapezoidal rule or interpolated average precision :

- $\text{AUC-PR}_c = \int_0^1 \text{Precision}_c\,(Recall)d(Recall)$ (7)

To ensure biological plausibility, we perform Grad-CAM interpretability analysis to visualize lesion-relevant regions driving predictions.

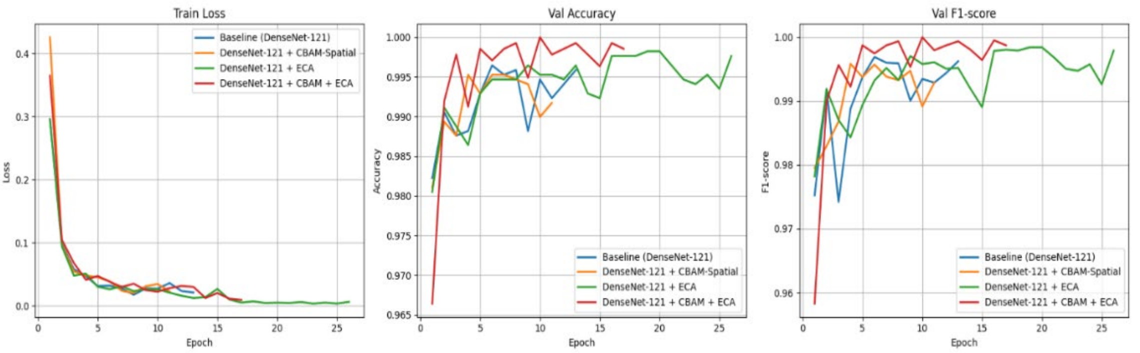
3. Results and Discussion

3.1. Results on the 4-Crop Dataset (10 Classes)

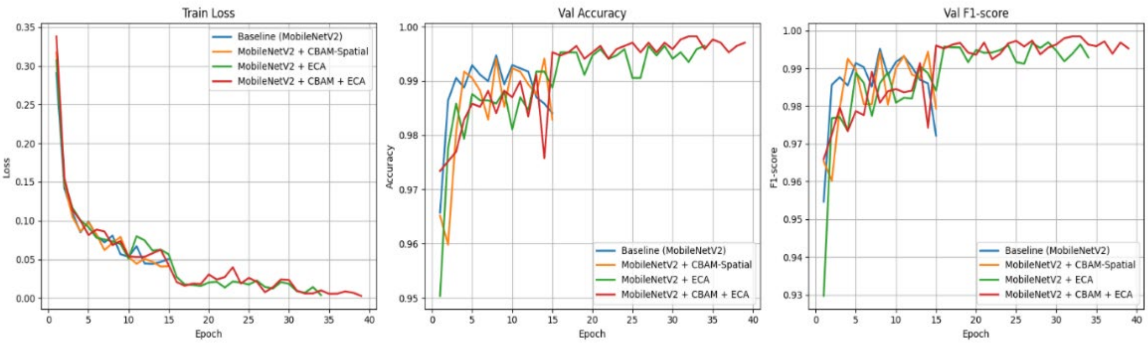
We evaluate the proposed ATSA-DenseNet framework on a 4-crop plant disease dataset comprising tomato, potato, pepper, and maize, with 10 disease classes and 1,692 test images. All backbone architectures were trained under identical optimization settings to ensure a fair comparison. As reported in Table 3, the proposed ATSA-DenseNet configuration achieves the best performance, reaching 99.94% accuracy and a macro-F1 score of 0.9994. Although the improvement over the baseline DenseNet121 (99.65%) is numerically modest, it is achieved without increasing parameters or FLOPs, indicating that the gain comes from architectural refinement rather than model scaling. Lightweight architectures such as MobileNetV2 and EfficientNet-B0 also demonstrate competitive results with substantially lower computational cost, highlighting their suitability for resource-constrained environments. However, DenseNet121 variants exhibit more stable convergence and stronger class discrimination, particularly for visually similar disease patterns, suggesting that dense connectivity provides more effective feature reuse for fine-grained plant disease classification.

Table 3. Performance comparison of dual-attention CNN architectures on the 4-crop plant disease dataset.

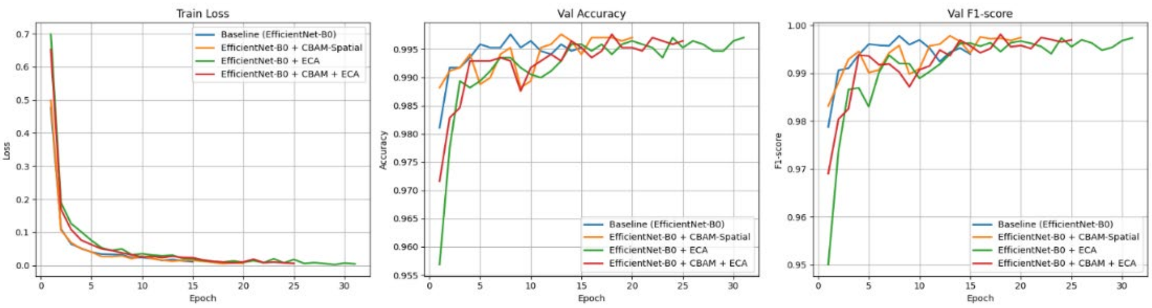
Model	Accuracy (%)	F1 (macro)	Precision	Recall	Params (M)	FLOPs (G)
DenseNet121+CBAM-spatial+ECA	99.94	0.9994	0.9994	0.9994	6.96	2.87
MobileNetV2+CBAM-spatial+ECA	99.82	0.9985	0.9982	0.9985	2.24	0.31
EfficientNet-B0+CBAM-spatial+ECA	99.76	0.9981	0.9976	0.9981	4.02	0.40
ResNet50+CBAM-spatial+ECA	99.70	0.9973	0.9970	0.9973	23.53	4.11



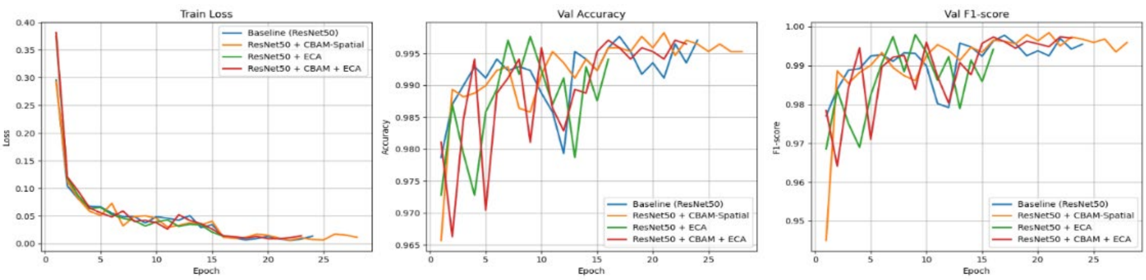
(a)



(b)



(c)



(d)

Fig. 3. Training and validation dynamics of four CNN architectures on the 4-crop dataset: (a) DenseNet121, (b) MobileNetV2, (c) EfficientNet-B0, and (d) ResNet50. Each panel shows training loss, validation accuracy, and validation F1-score across epochs for baseline, CBAM-Spatial, ECA, and dual-attention variants.

The plots in Fig. 3 show training loss, validation accuracy, and validation F1-score across epochs, illustrating how attention mechanisms affect convergence stability and final performance. DenseNet121 + CBAM-Spatial + ECA exhibits the best results, surpassing 99.94% validation accuracy and F1-score, indicating the efficacy of combined spatial and channel attention in improving model stability and generalization for plant disease classification.

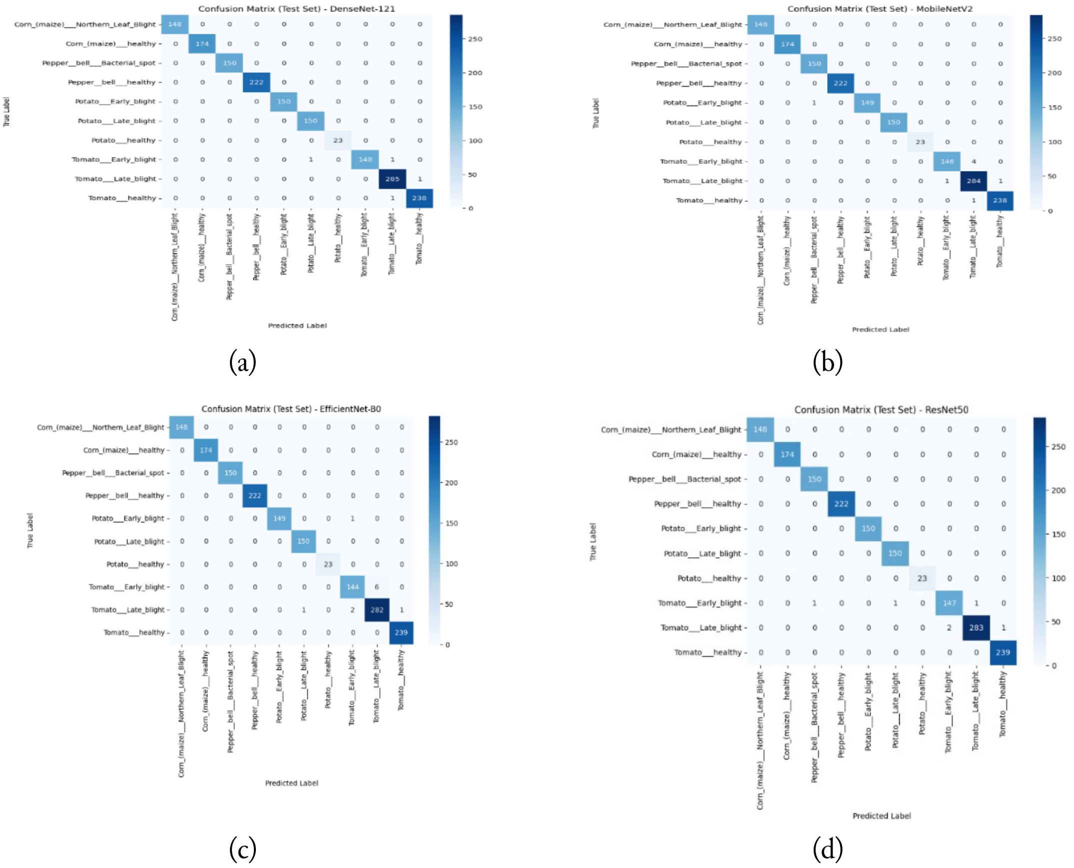


Fig. 4. Confusion Matrix: (a) DenseNet121, (b) MobileNetV2, (c) EfficientNet-B0, (d) ResNet50.

As shown in Fig. 4 (a), its confusion matrix reveals near-perfect diagonal dominance, with minimal off-diagonal errors even for visually similar pathologies such as tomato early/late blight and potato early/late blight - a critical advantage for reliable agricultural diagnostics.

Table 4. 5-Fold Cross-Validation Performance (Mean ± Std).

Model	Accuracy (%)	Std (%)	F1-Macro	Std F1
DenseNet121 (Baseline)	99.83	±0.15	0.9948	±0.0056
+ECA	99.82	±0.11	0.9965	±0.0016
+CBAM-spatial	99.80	±0.09	0.9968	±0.0008
ATSA-DenseNet (Proposed)	99.82	±0.07	0.9968	±0.0021

The dual-attention variants consistently converge faster and achieve higher performance. While the model achieved 99.94% accuracy on the held-out test set, cross-validation results show a mean accuracy of 99.82% ± 0.07 across folds (Table 4). The close agreement between these results and the held-out test accuracy indicates that the observed performance is stable and not dependent on a specific data split. Although slightly lower than the single split evaluation, the cross-validation results remain highly consistent, indicating robust model generalization across different data partitions.

To further investigate the generalization capability of the proposed model under real-world conditions, additional experiments were conducted on the PlantDoc dataset [45], which contains images captured in natural field environments with complex backgrounds, illumination variability, and occlusion. While the model demonstrates near-perfect performance under controlled conditions (99.94%) and stable performance across cross-validation (99.82 ± 0.07), a significant performance drop is observed under cross-dataset evaluation. However, fine-tuning on PlantDoc substantially improves performance, indicating strong transferability of the learned representations. A zero-shot evaluation was first performed by directly applying the model trained on PlantVillage to the PlantDoc dataset without retraining. The results show (Table 5) a significant performance drop, with an accuracy of 12.50% and a macro-F1 score of 0.0936, indicating a substantial domain shift between controlled and real-world data distributions. To further assess whether the learned feature representations are transferable, a fine-tuning experiment was conducted using a subset of PlantDoc data. After fine-tuning, the model achieved a significantly improved accuracy of 60.29% and a macro-F1 score of 0.4968. This corresponds to an absolute improvement of +47.79% in accuracy and +0.4032 in macro-F1 score compared to the zero-shot baseline. These results demonstrate that although the model trained on controlled datasets does not directly generalize to real-world conditions, the learned representations remain transferable and can be effectively adapted through fine-tuning. This finding highlights the importance of domain adaptation strategies when deploying deep learning models in practical agricultural environments.

Table 5. Cross-dataset adaptation results on PlantDoc.

Model	Training Setting	Accuracy (%)	F1-Macro
ATSA-DenseNet (Zero-shot)	PlantVillage only	12.50	0.0936
ATSA-DenseNet (Fine-tuned)	+ PlantDoc	60.29	0.4968

The substantial improvement after fine-tuning indicates that the proposed model captures transferable features, even under significant domain shift. The results also reveal that models trained solely on controlled datasets are insufficient for direct real-world deployment, emphasizing the need for domain adaptation and field data integration.

3.2. Results on the 3-Crop Dataset (8 Classes)

On the 3-crop dataset, the proposed ATSA-DenseNet achieves perfect classification performance, reaching 100% accuracy with a macro F1-score, precision, and recall of 1.0000 (Table 6). This result indicates near-perfect classification performance, with no misclassifications observed on the test set under controlled conditions. While such near-perfect performance should be interpreted cautiously given the controlled conditions of the dataset, the results indicate that the proposed attention integration enables highly discriminative feature representations for visually distinct plant disease patterns. Comparative evaluation across alternative backbones, including MobileNetV2, EfficientNet-B0, and ResNet50, equipped with the same dual-attention mechanism, further shows that DenseNet121 consistently provides stronger class separation. Although lightweight architectures such as MobileNetV2 achieve competitive performance with substantially lower FLOPs, ATSA-DenseNet maintains superior classification consistency while preserving a moderate computational footprint. This balance between predictive performance and computational efficiency supports its potential suitability for practical deployment in precision agriculture applications.

Table 6. Performance comparison of dual-attention CNN architectures on the 3-crop plant disease dataset.

Model	Accuracy (%)	F1 (macro)	Precision	Recall	Params (M)	FLOPs (G)
DenseNet121+CBAM-Spatial+ECA	100	1.0000	1.0000	1.0000	6.96	2.87
MobileNetV2+CBAM-spatial+ECA	99.93	0.9994	0.9993	0.9994	2.23	0.31
EfficientNet-B0+CBAM-spatial+ECA	99.85	0.9985	0.9985	0.9987	4.02	0.40
ResNet50+CBAM-spatial+ECA	99.78	0.9957	0.9978	0.9957	23.52	4.11

3.3. Architecture Comparison

To evaluate the robustness of the proposed architecture under increasing classification complexity, we compare the performance of ATSA-DenseNet across both 3-crop and 4-crop datasets. The 3-crop configuration contains eight disease classes, while the 4-crop dataset introduces additional crop diversity and two additional classes, resulting in a more challenging classification scenario. As summarized in Table 7, ATSA-DenseNet achieves 100.00% accuracy on the 3-crop dataset and 99.94% accuracy on the 4-crop dataset. Although a slight decrease in performance is observed when the number of crops and disease classes increases, the overall accuracy remains consistently high. This indicates that the proposed attention-based architecture maintains stable discriminative capability even as dataset complexity grows. The small performance difference between the two settings suggests that the dual-attention mechanism effectively enhances feature localization and channel-wise representation, allowing the network to capture disease patterns across different plant species. These results demonstrate that ATSA-DenseNet generalizes well across multiple crop types under controlled dataset conditions.

Table 7. Comparison of 3-Crop and 4-Crop Accuracies.

Model	3-Crop Accuracy (%)	4-Crop Accuracy (%)	Notes
DenseNet121 + CBAM-spatial + ECA	100.00	99.94	Best across all tasks
MobileNetV2 + CBAM-spatial + ECA	99.93	99.82	Largest drop when adding more classes
EfficientNet-B0 + CBAM-spatial + ECA	99.85	99.76	Stable and consistent
ResNet50 + CBAM-spatial + ECA	99.78	99.70	Slightly below DenseNet121

MobileNetV2, despite being lightweight, exhibited the greatest performance drop when scaling from the 3-crop to the 4-crop dataset, indicating limited robustness under increased class complexity. This suggests that plant disease detection, particularly across diverse crops and visually similar pathologies, benefits from architectures with richer feature reuse mechanisms, such as DenseNet121. EfficientNet-B0 and ResNet50 performed moderately but still lagged behind DenseNet121 in discriminating fine-grained disease symptoms. These outcomes reaffirm prior findings that dense connectivity and strong representational capacity are critical for multi-crop disease prediction as task complexity grows.

3.4. Ablation Study of Attention Modules

To isolate the contribution of each attention component, we conducted an ablation study using DenseNet121 as the backbone under identical training conditions. Table 8 presents the ablation analysis of the proposed attention configuration using DenseNet121 as the backbone architecture. The baseline model achieves 99.65% accuracy, while the addition of ECA provides a modest improvement to 99.70%, indicating that channel-wise feature recalibration contributes positively to representation learning. Introducing CBAM-Spatial results in a larger performance gain (99.88%), suggesting that spatial localization plays a critical role in identifying disease patterns on leaf surfaces. When both mechanisms are integrated in the proposed ATSA-DenseNet architecture, the model achieves the highest performance (99.94% accuracy and 0.9994 macro F1-score) without increasing the number of parameters or computational cost. This result indicates that the spatial and channel attention mechanisms provide complementary benefits, improving feature discrimination while maintaining the lightweight characteristics of the DenseNet backbone.

Table 8. Ablation study of attention modules.

Model Variant	ECA	CBAM-Spatial	Params (M)	FLOPs (G)	Accuracy (%)	F1-Macro
DenseNet121 (Baseline)	x	x	6.96	2.86	99.65	0.9949
+ECA	√	x	6.96	2.87	99.70	0.9976
+CBAM-spatial	x	√	6.96	2.86	99.88	0.9989
ATSA-DenseNet (Proposed)	√	√	6.96	2.87	99.94	0.9994

3.5. Visual Explainability (Grad-CAM)

To provide qualitative interpretability of the learned representations, Gradient-weighted Class Activation Mapping (Grad-CAM) was employed to visualize the regions that contribute most strongly to the model's predictions. Fig. 5. compares the Grad-CAM heatmaps produced by the baseline DenseNet121 and the proposed dual-attention variant. The baseline model tends to generate broader activation regions across the leaf surface, occasionally extending beyond the symptomatic areas. In contrast, the dual-attention model produces more localized activation patterns that concentrate around lesion-relevant regions. This behavior suggests improved spatial discrimination in the learned feature representations. These visualizations provide qualitative evidence that the proposed attention integration guides the model toward more relevant regions under controlled dataset conditions. Similar activation patterns were observed across other evaluated architectures. However, these visualizations should be interpreted as qualitative indicators of model attention rather than definitive confirmation of biological lesion detection.

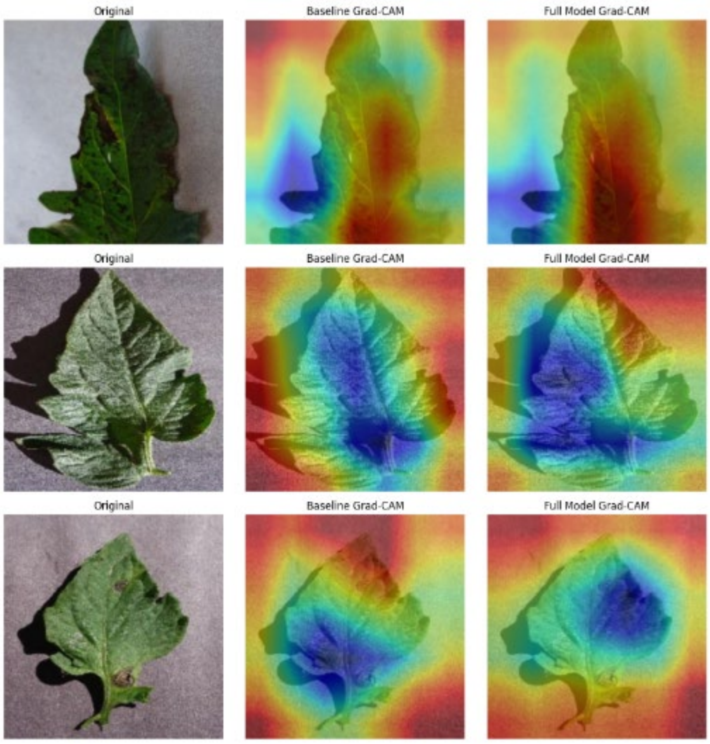


Fig. 5. Grad-CAM visualizations for DenseNet121.

3.6. Discussion of Findings

The experimental results provide several important observations regarding the effectiveness of the proposed architecture. First, the DenseNet121 backbone integrated with dual-attention consistently achieves the highest accuracy across both 3-crop and 4-crop evaluation settings. This indicates that combining spatial localization with channel-wise feature recalibration improves the model's ability to capture disease-relevant patterns.

Table 9. Comparison with recent studies on plant disease classification.

Study	Model	Dataset	Classes	Accuracy (%)	Notes
[32]	Resnet + Spatial–hybrid attention	Tomato & pepper (PlantVillage)	8	99	Attention improves lesion localization
[16]	MobileNetV2 + CBAM	Tomato diseases (PlantVillage)	10	98.75	CBAM enhances fine-grained features
[23]	*DenseNet121 + CBAM*	*PlantVillage (3- & 4-crops)*	*8 / 10*	*98.23 / 98.88*	*CBAM improves feature discrimination*
[33]	*CNN + LBP Fusion*	*Mixed crops (PlantVillage)*	*8*	*98.05*	*Hybrid texture features*
This Work	***Proposed ATSA-DenseNet***	***PlantVillage (3- & 4-crops)***	***8 / 10***	***100.00 / 99.94***	***Dual-attention improves feature localization and channel representation.***

Second, confusion matrix analysis shows strong class separation, particularly for visually similar diseases such as tomato early blight and late blight, suggesting improved discriminative feature learning. Third, the macro F1-scores indicate a balanced relationship between precision and recall, highlighting the model's stable performance across multiple disease categories. Finally, while lightweight architectures such as MobileNetV2 provide strong efficiency advantages, the DenseNet121-based configuration demonstrates more stable classification performance as the number of classes and crop diversity increases. These findings suggest that dense connectivity combined with attention-guided feature refinement can enhance representation learning for multi-crop disease classification tasks. The comparison indicates that the proposed dual-attention configuration achieves competitive performance relative to recent attention-based approaches while maintaining a lightweight architectural design.

3.7. Limitations and Future Work

Although the proposed ATSA-DenseNet framework demonstrates strong performance across both 3-crop and 4-crop datasets, several limitations should be acknowledged. First, the experiments are conducted on the PlantVillage dataset, which consists of images captured under relatively controlled conditions with homogeneous backgrounds. While this setting enables standardized benchmarking, real-world agricultural environments typically involve more complex scenarios, including varying illumination, background clutter, occlusion, and overlapping leaves. Second, the current evaluation primarily focuses on classification performance under a fixed dataset configuration. Although the proposed dual-attention mechanism improves discriminative capability without increasing model complexity, further validation on cross-domain datasets or field-acquired images is necessary to fully assess generalization capability. The cross-dataset evaluation results further confirm that models trained on controlled datasets do not directly generalize to real-world environments. This highlights the necessity of incorporating field-acquired data and domain adaptation strategies for practical deployment. Finally, while the proposed architecture maintains a relatively lightweight footprint suitable for edge deployment, additional optimization techniques, such as model pruning, quantization, or hardware-aware compression, can be investigated to further improve inference efficiency. These directions will be explored in future work to support scalable and real-time plant disease monitoring systems in precision agriculture.

4. Conclusion

This study introduces ATSA-DenseNet, a redundancy-aware dual-attention framework that integrates the spatial branch of CBAM with ECA within a DenseNet121 backbone for multi-crop plant disease classification. Unlike prior dual-attention approaches that retain full CBAM and potentially introduce channel-level redundancy, the proposed design isolates complementary spatial and channel mechanisms to enhance representational efficiency without increasing computational complexity. Experimental evaluation on controlled multi-crop PlantVillage-derived datasets demonstrates that the proposed architecture consistently outperforms baseline and single-attention variants. On the 4-crop configuration, ATSA-DenseNet achieves 99.94% accuracy and a macro-F1 score of 0.9994 while maintaining a lightweight footprint (6.96M parameters, 2.87G FLOPs). Cross-validation analysis further confirms stable performance with low variance, indicating that improvements are not dependent on a specific data split. While the near-perfect performance observed on the 3-crop dataset reflects strong intra-dataset generalization, the results should be interpreted within the context of controlled imaging conditions. Therefore, the contribution of this work lies not in claiming field-level robustness, but in demonstrating that redundancy-aware dual-attention improves feature discrimination efficiency under standardized multi-class evaluation. Grad-CAM analysis suggests that the proposed architecture promotes more localized activation on disease-relevant regions compared to the baseline, contributing to improved interpretability. These findings highlight the effectiveness of redundancy-aware dual-attention integration for improving feature discrimination while maintaining computational efficiency in multi-crop disease detection. Future work will focus on enhancing generalization and deployment readiness under real-world agricultural conditions.

Declarations

Author contribution. All authors participated in the development of the manuscript and approved its final version.

Funding statement. This research received no grant from any funding agency in the public, commercial, or sectoral sectors.
Conflict of interest. The authors declare no conflict of interest.
Additional information. No additional information is available for this paper.

References

[1] Y. Wang, H. Wang, and Z. Peng, "Rice diseases detection and classification using attention based neural network and bayesian optimization," *Expert Syst. Appl.*, vol. 178, no. September, p. 114770, Sep. 2021, doi: 10.1016/j.eswa.2021.114770.

[2] J. Zhao *et al.*, "A review of plant leaf disease identification by deep learning algorithms," *Front. Plant Sci.*, vol. 16, p. 1637241, Aug. 2025, doi: 10.3389/fpls.2025.1637241.

[3] M. S. Farooq, A. Kamran, S. A. Raza, M. F. Wasiq, B. Hassan, and N. J. Herzog, "Advancing Early Blight Detection in Potato Leaves Through ZeroShot Learning," *J. Imaging*, vol. 11, no. 8, p. 256, Jul. 2025, doi: 10.3390/jimaging11080256.

[4] P. Jha, D. Dembla, and W. Dubey, "Implementation of Machine Learning Classification Algorithm Based on Ensemble Learning for Detection of Vegetable Crops Disease," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 1, pp. 584–594, Jan. 2024, doi: 10.14569/IJACSA.2024.0150157.

[5] J. Feng, W. E. Ong, W. C. Teh, and R. Zhang, "Enhanced Crop Disease Detection With EfficientNet Convolutional Group-Wise Transformer," *IEEE Access*, vol. 12, pp. 44147–44162, 2024, doi: 10.1109/ACCESS.2024.3379303.

[6] M. Abdalla, O. Mohamed, and E. M. Azmi, "Adaptive Learning Model for Detecting Wheat Diseases," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 5, pp. 1287–1298, May 2024, doi: 10.14569/IJACSA.2024.01505130.

[7] M. B. Gulame, T. G. Thite, and K. D. Patil, "Plant disease prediction system using advance computational Technique," *J. Phys. Conf. Ser.*, vol. 2601, no. 1, p. 012031, Sep. 2023, doi: 10.1088/1742-6596/2601/1/012031.

[8] S. Karthikeyan, R. Charan, S. Narayanan, and L. Jani Anbarasi, "Enhanced plant disease classification with attention-based convolutional neural network using squeeze and excitation mechanism," *Front. Artif. Intell.*, vol. 8, p. 1640549, Aug. 2025, doi: 10.3389/frai.2025.1640549.

[9] W. Shafik, A. Tufail, C. De Silva Liyanage, and R. A. A. H. M. Apong, "Using transfer learning-based plant disease classification and detection for sustainable agriculture," *BMC Plant Biol.*, vol. 24, no. 1, p. 136, Feb. 2024, doi: 10.1186/s12870-024-04825-y.

[10] G. Suresh, V. L. Narla, G. Tummalapalli, A. S. Peddinti, and Q. M. Al-Mdallal, "AgroDetect Smart Plant Pathology From Leaf Images," *Eng. Reports*, vol. 7, no. 10, p. e70384, Oct. 2025, doi: 10.1002/eng2.70384.

[11] S. K. Shah, M. B. M. Su'ud, A. Khan, M. M. Alam, and M. Ayaz, "PLDC-Net: Potato Leaf Disease Classification Network Based on an Efficient Convolutional Neural Network," *Eng. Reports*, vol. 7, no. 7, p. e70178, Jul. 2025, doi: 10.1002/eng2.70178.

[12] S. K. Noon, M. Amjad, M. A. Qureshi, A. Mannan, and T. Awan, "An Improved Detection Method for Crop & Fruit Leaf Disease under Real-Field Conditions," *AgriEngineering*, vol. 6, no. 1, pp. 344–360, Feb. 2024, doi: 10.3390/agriengineering6010021.

[13] V. A and A. R V, "Transfer learning based deep learning model for classifying tomato plant leaf diseases," *Eng. Res. Express*, vol. 7, no. 2, p. 025250, Jun. 2025, doi: 10.1088/2631-8695/add6f5.

[14] D. R, D. S. Mythili, X. Rexeena, and D. Devadas, "Detection of Tomato Leaf Diseases Using Deep Learning and Spatial Attention Mechanisms," *Int. J. Res. Sci. Innov.*, vol. XII, no. VII, pp. 1144–1153, Aug. 2025, doi: 10.51244/IJRSI.2025.120700118.

[15] X. Ma, W. Chen, and Y. Xu, "ERCP-Net: a channel extension residual structure and adaptive channel attention mechanism for plant leaf disease classification network," *Sci. Rep.*, vol. 14, no. 1, p. 4221, Feb. 2024, doi: 10.1038/s41598-024-54287-3.

[16] S. Dou, L. Wang, D. Fan, L. Miao, J. Yan, and H. He, "Classification of Citrus Huanglongbing Degree Based on CBAM-MobileNetV2 and Transfer Learning," *Sensors*, vol. 23, no. 12, p. 5587, Jun. 2023, doi: 10.3390/s23125587.

[17] X. Sun and H. Huo, "Corn leaf disease recognition based on improved EfficientNet," *IET Image Process.*, vol. 19, no. 1, p. e13288, Jan. 2025, doi: 10.1049/ipr2.13288.

[18] E. Zhang, N. Zhang, F. Li, and C. Lv, "A lightweight dual-attention network for tomato leaf disease identification," *Front. Plant Sci.*, vol. 15, p. 1420584, Aug. 2024, doi: 10.3389/fpls.2024.1420584.

[19] T. Dong *et al.*, "Wheat disease recognition method based on the SC-ConvNeXt network model," *Sci. Rep.*, vol. 14, no. 1, p. 32040, Dec. 2024, doi: 10.1038/s41598-024-83636-5.

[20] J. Wang, J. Jia, Y. Zhang, H. Wang, and S. Zhu, "RAAWC-UNet: an apple leaf and disease segmentation method based on residual attention and atrous spatial pyramid pooling improved UNet with weight compression loss," *Front. Plant Sci.*, vol. 15, p. 1305358, Mar. 2024, doi: 10.3389/fpls.2024.1305358.

[21] J. Zhang, X. Yang, X. Fu, B. Wang, and H. Li, "LDL-MobileNetV3S: an enhanced lightweight MobileNetV3-small model for potato leaf disease diagnosis through multi-module fusion," *Front. Plant Sci.*, vol. 16, p. 1656731, Oct. 2025, doi: 10.3389/fpls.2025.1656731.

[22] wei Dai, F. Shi, X. Wang, H. Xu, L. Yuan, and X. Wen, "A Dense Residual Correlation Attention Multi-scale Network for Remote Sensing Scene Classification," Jan. 11, 2024. doi: 10.21203/rs.3.rs-3823437/v1.

[23] A. Gupta and A. Kaur, "Multi-Crop Disease Prediction Using Attention-Augmented Deep Learning: A DenseNet121-CBAM Model," in *2025 International Conference on Electronics, AI and Computing (EAIC)*, IEEE, Jun. 2025, pp. 1–6. doi: 10.1109/EAIC66483.2025.11101367.

[24] B. Sundararaman, S. Jagdev, and N. Khatri, "Transformative Role of Artificial Intelligence in Advancing Sustainable Tomato (Solanum lycopersicum) Disease Management for Global Food Security: A Comprehensive Review," *Sustainability*, vol. 15, no. 15, p. 11681, Jul. 2023, doi: 10.3390/su151511681.

[25] M. Arafath, A. A. Nithya, and S. Gijwani, "Tomato Leaf Disease Detection Using Deep Convolution Neural Network," in *Advances in Science and Technology*, Trans Tech Publications Ltd, Feb. 2023, pp. 236–245. doi: 10.4028/p-vph2n1.

[26] M. Muthulakshmi, N. Aishwarya, R. K. Vinesh Kumar, and B. Rakesh Thoppaen Suresh, "Potato Leaf Disease Detection and Classification With Weighted Ensembling of YOLOv8 Variants," *J. Phytopathol.*, vol. 172, no. 6, p. e13433, Nov. 2024, doi: 10.1111/jph.13433.

[27] M. Umar, S. Altaf, S. Ahmad, H. Mahmoud, A. S. N. Mohamed, and R. Ayub, "Precision Agriculture Through Deep Learning: Tomato Plant Multiple Diseases Recognition With CNN and Improved YOLOv7," *IEEE Access*, vol. 12, pp. 49167–49183, 2024, doi: 10.1109/ACCESS.2024.3383154.

[28] Z. Ye *et al.*, "Recognition of terminal buds of densely-planted Chinese fir seedlings using improved YOLOv5 by integrating attention mechanism," *Front. Plant Sci.*, vol. 13, p. 991929, Oct. 2022, doi: 10.3389/fpls.2022.991929.

[29] J. Wang, M. Li, C. Han, and X. Guo, "YOLOv8-RCAA: A Lightweight and High-Performance Network for Tea Leaf Disease Detection," *Agriculture*, vol. 14, no. 8, p. 1240, Jul. 2024, doi: 10.3390/agriculture14081240.

[30] W. Guo, S. Feng, Q. Feng, X. Li, and X. Gao, "Cotton leaf disease detection method based on improved SSD," *Int. J. Agric. Biol. Eng.*, vol. 17, no. 2, pp. 211–220, 2024, doi: 10.25165/J.IJABE.20241702.8574.

[31] Y. Unal, "Integrating <scp>CBAM</scp> and Squeeze-and-Excitation Networks for Accurate Grapevine Leaf Disease Diagnosis," *Food Sci. Nutr.*, vol. 13, no. 6, p. e70377, Jun. 2025, doi: 10.1002/fsn3.70377.

[32] H. E. David, K. Ramalakshmi, R. Venkatesan, and G. Hemalatha, "Tomato Leaf Disease Detection Using Hybrid CNN-RNN Model," in *Advances in Parallel Computing*, vol. 38, IOS Press, 2021, pp. 593–597. doi: 10.3233/APC210108.

[33] E. Hassan, S. A. Ghazalah, N. El-Rashidy, T. A. El-Hafeez, and M. Y. Shams, "DenseNet Model with Attention Mechanisms for Robust Date Fruit Image Classification," *Int. J. Comput. Intell. Syst.*, vol. 18, no. 1, p. 228, Sep. 2025, doi: 10.1007/s44196-025-00809-4.

[34] R. Leelavathi and M. Kalamani, "Medicinal plant leaf disease classification using optimal weighted features with dilated adaptive DenseNet and attention mechanism," *Sci. Rep.*, vol. 15, no. 1, p. 36852, Oct. 2025, doi: 10.1038/s41598-025-20629-y.

[35] A. Bhujel, N.-E. Kim, E. Arulmozhi, J. K. Basak, and H.-T. Kim, "A Lightweight Attention-Based Convolutional Neural Networks for Tomato Leaf Disease Classification," *Agriculture*, vol. 12, no. 2, p. 228, Feb. 2022, doi: 10.3390/agriculture12020228.

[36] P. Radočaj, D. Radočaj, and G. Martinović, "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, May 2024, doi: 10.3390/bdcc8060052.

[37] L. Alswilem and E. Asadov, "Deep Learning in Maize Disease Classification," *Artif. Intell. Appl. Sci.*, vol. 1, no. 1, pp. 20–27, Jul. 2025, doi: 10.69882/adba.ai.2025074.

[38] H. A. Santoso, B. Fandhi Safsalta, N. Febrianto, G. Wilujeng Saraswati, and S.-C. Haw, "Comparative analysis of convolutional neural network and DenseNet121 transfer learning in agriculture focusing on crop leaf disease identification," *Appl. Comput. Informatics*, vol. 18, no. 06, pp. 1–13, Jun. 2024, doi: 10.1108/ACI-03-2024-0132.

[39] M. K. A. Mazumder, M. M. Kabir, A. Rahman, M. Abdullah-Al-Jubair, and M. F. Mridha, "DenseNet201Plus: Cost-effective transfer-learning architecture for rapid leaf disease identification with attention mechanisms," *Heliyon*, vol. 10, no. 15, p. e35625, Aug. 2024, doi: 10.1016/j.heliyon.2024.e35625.

[40] M. S. Krishna, P. Machado, R. I. Otuka, S. W. Yahaya, F. Neves dos Santos, and I. K. Ihianle, "Plant Leaf Disease Detection Using Deep Learning: A Multi-Dataset Approach," Nov. 11, 2024, *Preprints*. doi: 10.20944/preprints202411.0732.v1.

[41] Y. Xu, D. Li, C. Li, Z. Yuan, and Z. Dai, "LiSA-MobileNetV2: an extremely lightweight deep learning model with Swish activation and attention mechanism for accurate rice disease classification," *Front. Plant Sci.*, vol. 16, p. 1619365, Aug. 2025, doi: 10.3389/fpls.2025.1619365.

[42] S. P. Ashwini and S. Uma, "A Hybrid Approach using ResNet 50 and EfficientNetB0 for Attention-enhanced Deep Learning for Early Detection of Apple Plant Diseases: A Review," *Agric. Sci. Dig. - A Res. J.*, no. Of, Oct. 2025, doi: 10.18805/ag.D-6375.

[43] S. M. Alhammad, D. S. Khafaga, W. M. El-hady, F. M. Samy, and K. M. Hosny, "Deep learning and explainable AI for classification of potato leaf diseases," *Front. Artif. Intell.*, vol. 7, p. 1449329, Feb. 2025, doi: 10.3389/frai.2024.1449329.

[44] Y. Li *et al.*, "Research on Potato Leaf Disease Recognition Method Based on Convolutional Neural Networks," *Concurr. Comput. Pract. Exp.*, vol. 37, no. 25–26, p. e70355, Nov. 2025, doi: 10.1002/cpe.70355.

[45] D. Singh, N. Jain, P. Jain, P. Kayal, S. Kumawat, and N. Batra, "PlantDoc: A dataset for visual plant disease detection," in *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*, New York, NY, USA: ACM, Jan. 2020, pp. 249–253. doi: 10.1145/3371158.3371196.