

Article

Not peer-reviewed version

REA-CNN: A Region-Aware & Enhanced Attention CNN for Robust Leaf Disease Classification

[M. Ammad Uddin](#)*, [Muhammad Ayaz](#), H.M. Aggoune, [Ali Mansour Mansour](#), [Ahmad Hani El Fawal](#), Manimurugan Shanmuganathan

Posted Date: 9 January 2026

doi: 10.20944/preprints202601.0718.v1

Keywords: plant leaf health detection; machine learning; deep learning; CNN; CBAM



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

REA-CNN: A Region-Aware & Enhanced Attention CNN for Robust Leaf Disease Classification

M. Ammad Uddin ^{1,*}, Muhammad Ayaz ², H.M. Aggoune ², Ali Mansour Mansour ¹, Ahmad Hani El Fawal ^{1,3} and Manimurugan Shanmuganathan ⁴

¹ LabSTICC, UMR 6285 CNRS, ENSTA, Institut Polytechnique de Paris, 29806 Brest, France

² Artificial Intelligence and Sensing Technologies (AIST) Research Center, University of Tabuk, Saudi Arabia

³ School of Computer and Applied Sciences, Modern University for Business and Science, Damour, Lebanon

⁴ Faculty of Computers and Information Technology, University of Tabuk

* Correspondence: mohammad.ammad-uddin@ensta-bretagne.org; Tel.: +33769300994

Abstract

Automated plant leaf disease detection using deep learning has achieved high accuracy on benchmark datasets; however, its performance often degrades when applied to real-world agricultural images affected by background clutter, illumination variability, and partial occlusions. These factors limit the reliability of conventional convolutional neural network (CNN)-based models trained under controlled conditions. To address this limitation, this paper proposes a Region-Aware and Enhanced Attention Convolutional Neural Network (REA-CNN) for robust classification of plant leaf diseases. The proposed framework integrates explicit region-aware pre-processing for background suppression with an attention-enhanced CNN backbone, enabling the model to focus on disease-relevant visual patterns. Unlike conventional two-stage CNN-SVM pipelines, REA-CNN is trained in a fully end-to-end manner, allowing joint optimization of feature extraction, attention refinement, and classification. Experimental results show that the proposed approach achieves higher classification accuracy and improved generalization on heterogeneous and real-world images compared to existing methods. These results demonstrate the effectiveness of combining region awareness and attention-guided learning for developing practical and deployable decision-support systems in precision agriculture.

Keywords: plant leaf health detection; machine learning; deep learning; CNN; CBAM

1. Introduction

Accurate detection of plant leaf diseases is crucial for enhancing crop productivity and facilitating timely agricultural interventions. In recent years, convolutional neural networks (CNNs), particularly when combined with transfer learning, have demonstrated strong performance in plant disease classification. However, most existing CNN-based methods are evaluated primarily on controlled benchmark datasets such as PlantVillage, where images are captured under uniform lighting conditions with clean and uncluttered backgrounds. As a result, their performance often degrades when applied to real agricultural images that exhibit background clutter, illumination variability, occlusions, and partial leaf visibility. In our previous work, we investigated hybrid CNN-CBAM-SVM architectures to improve classification performance by incorporating attention mechanisms and margin-based classifiers. Although this approach demonstrated that attention can enhance disease-relevant feature discrimination, its two-stage training strategy introduces important limitations. In particular, decoupling feature extraction from classification prevents end-to-end optimization, increases system complexity, and restricts the influence of the classifier on learned feature representations. In addition, background interference is not explicitly addressed, which limits robustness when models trained on laboratory-style datasets are deployed in real-world environments.

These limitations motivate the development of a unified learning framework that explicitly accounts for both background suppression and feature refinement. To this end, we propose a Region-Aware and Enhanced Attention Convolutional Neural Network (REA-CNN), designed as an end-to-end architecture for real-world plant leaf disease classification. The core idea of REA-CNN is to suppress irrelevant background information while enhancing disease-specific visual patterns within the leaf region. This is achieved by integrating region-aware pre-processing prior to feature extraction and embedding an attention mechanism directly into the CNN backbone, without requiring pixel-level lesion annotations.

Unlike conventional pre-processing techniques that apply uniform transformations across the entire image, region-aware processing enables the model to focus on leaf-centric visual cues, reducing dataset bias and improving robustness to environmental variations. Furthermore, end-to-end training ensures uninterrupted gradient flow from the classification layer to early convolutional stages, allowing joint optimization of feature extraction, attention refinement, and classification. Consequently, the proposed REA-CNN provides a robust and deployable framework that addresses the limitations of conventional CNN-based and hybrid plant disease classification systems under real-world imaging conditions.

In recent years, numerous hybrid artificial intelligence models have been proposed for plant leaf disease detection, combining convolutional neural networks with complementary techniques such as attention mechanisms, optimization algorithms, transformers, and recurrent neural networks. These hybrid designs aim to enhance feature representation and improve classification accuracy. Most reported results are obtained on benchmark datasets captured under controlled conditions, characterized by uniform backgrounds and consistent illumination. Although these approaches demonstrate high accuracy on benchmark datasets, many hybrid frameworks process the full image and rely primarily on implicit mechanisms such as attention layers, transformer blocks, or feature optimization strategies to suppress irrelevant information. While effective under controlled conditions, these mechanisms do not explicitly address background clutter, illumination variability, or environmental noise. Consequently, their performance often degrades when applied to real-world field images containing complex backgrounds, shadows, partial occlusions, and unconstrained lighting conditions.

In addition, several hybrid models employ multi-stage learning pipelines that separate feature extraction, optimization, and classification into distinct components. Although such designs can improve classification accuracy in specific settings, they increase system complexity and limit end-to-end optimization. In these architectures, the classifier has limited influence on earlier feature extraction stages, restricting the model's ability to adapt learned representations to task-specific and real-world variations. In contrast, the proposed Region-Aware and Enhanced Attention Convolutional Neural Network (REA-CNN) adopts a unified design that explicitly integrates region-aware background suppression with attention-enhanced feature learning within a single end-to-end framework. By isolating leaf regions and enhancing disease-relevant visual features before deep feature extraction, the proposed approach guides the learning process toward semantically meaningful information while reducing environmental interference.

Unlike recent hybrid models that primarily increase architectural depth or complexity, REA-CNN emphasizes input-level refinement combined with joint optimization of feature extraction, attention refinement, and classification. This design improves robustness, interpretability, and suitability for practical agricultural deployment. While a limited number of studies incorporate segmentation or detection, these components are typically applied as independent pre-processing steps and are not jointly optimized with attention mechanisms and classification. A detailed comparison with recent studies published in 2025 is summarized in Table 1.

Table 2 provides a comparative overview of recent state-of-the-art approaches for plant leaf disease detection across multiple crops, including tomato, potato, apple, pepper, and grape. As summarized, most existing studies rely on conventional convolutional neural networks, transfer-learning-based architectures, or classical machine-learning pipelines, with performance commonly

reported using overall classification accuracy. Although some methods incorporate region-based processing, such as segmentation or object detection, or attention mechanisms, these components are typically applied independently rather than within a unified learning framework.

In contrast, the proposed REA-CNN integrates explicit region-aware pre-processing for background suppression with attention-enhanced feature learning in a single end-to-end architecture. This joint design enables more effective focus on disease-relevant visual patterns while reducing sensitivity to background clutter and illumination variability.

Overall, the comparison in Table 2 indicates that REA-CNN combines multiple complementary components within a unified framework, extending beyond incremental architectural modifications observed in prior approaches and supporting its practical relevance for plant leaf disease classification.

For image-based classification tasks, AI system architectures can be interpreted in terms of multiple functional stages, depending on task complexity and deployment constraints. In this work, we adopt a level-based design perspective that organizes such systems into up to four functional levels (Table 3). A basic two-level system typically consists of limited input preparation followed by direct feature extraction and classification using a convolutional neural network. While effective for clean and well-controlled datasets, such architectures often exhibit limited robustness under real-world conditions.

A three-level system extends this structure by introducing an intermediate refinement stage, such as attention mechanisms, feature optimization strategies, or transfer learning, to improve discriminative capability. Although this additional stage enhances feature representation, it may remain insufficient when images are affected by background clutter, illumination variability, or environmental noise.

For more challenging real-world scenarios, an advanced four-level system design can be employed. This configuration incorporates explicit input pre-processing and augmentation, deep feature extraction, intelligent feature refinement, and a final decision layer. By decomposing the learning process into distinct functional stages, each level addresses specific challenges such as background suppression, illumination variation, or feature ambiguity, thereby improving robustness, interpretability, and deployment suitability. This level-based perspective provides a practical framework for analysing existing approaches and for positioning the proposed REA-CNN within a unified and application-oriented system design, as summarized in Tables 3 and 4.

Table 1. COMPARISON OF REA-CNN WITH RECENT PUBLISHED ARTICLES.

Aspect	[1]	[2]	[3]	[4]	[5]	[6]	[7]	RA/EA-CNN
Publication venue	Biology MDPI	AEEE Journal	Elsevier SciDirect	Springer	Springer	IEEE Access	Springer	—
Application domain	Plant diseases	Face Recognition	Plant diseases	Rice diseases	Tea leaf diseases	Plant diseases	Plant leaf diseases	Plant diseases
Input type	Leaf images	Face images	Leaf images	Leaf images	Leaf images	Leaf image sequences	Leaf images	Leaf images
Background removal	No	No	No	segmentation	Partial	No	No	Explicit
Region awareness	No	No	No	Yes	Partial	No	Partial (ViT attention)	Yes
Image enhancement	No	No	No	Yes	CLAHE	No	Indirect	Explicit

Feature extraction	Hand-crafted (2D-DWT)	PCA / LBP / CNN	CAE latent space	Hybrid fusion	CNN + Transformer	VGG16 CNN	CNN ensemble + ViT	Region-guided CNN
Temporal modeling	No	No	No	No	No	MBi-LSTM	No	No
Optimization strategy	Flower Pollination Algorithm	Manual	Implicit	African Vultures Optimization	Gray Wolf Optimization	None	None	Not required
Classifier / backbone	SVM + CNN	CNN / KNN	CNN	Depth-Separable NN	CNN-Transformer	CNN + Bi-LSTM	CNN + Vision Transformer	CNN
End-to-end learning	Partial	Mixed	Semi	Semi	Semi	Semi	Semi	Full
Model complexity	High	Medium	Low	Moderate-High	Moderate-High	High	High	Moderate
Explainability	Weak	Weak	Weak	Moderate	Moderate	Moderate	Moderate	Strong
Real-time suitability	Possible	Limited	Yes	Limited	Moderate	Limited	Limited	Yes
Generalizability	oderate	Low	oderate	Moderate	Moderate	Moderate	Moderate	High
Robust to real-world images	No	No	Limited	Limited	Limited	Limited	Limited	High

Table 2. PUBLISHED HYBRID AI SYSTEMS AND THEIR RELEVANCE WITH OUR PROPOSED ONE.

Title	Crop Studied	AI Model Used	Accuracy (%)	Relevance with respect to REA-CNN
<i>Tomato Leaf Disease Detection System Based on FC-SNDPN [8]</i>	Tomato	FC-SNDPN (FCN + Dual-Path + Switchable Norm)	97.59	Partial (segmentation, no explicit attention)
<i>An Improved CNN Model for Diagnosis of Diseases in Tomato Crop [9]</i>	Tomato	Hybrid CNN (VGG + Inception)	99.17	No
<i>Tomato Leaf Disease Detection Using Deep CNN[10]</i>	Tomato	Deep CNN (Batch Norm + Dropout)	98.00	No
<i>Classification of Plant Leaf Disease Using Deep Learning [11]</i>	Multi (incl. Tomato)	CNN / AlexNet / MobileNet	84.24 / 91.19 / 97.33	No
<i>Detection of Tomato Leaf Diseases... PCA DeepNet [12]</i>	Tomato	PCA DeepNet + GAN + Faster R-CNN	99.60	Partial (region detection, no attention)
<i>MMDGAN : Fusion Augmentation + B-ARNet [13]</i>	Tomato	GAN-based + B-ARNet	97.12	No
<i>Tomato Plant Disease Classification... Adaptive Attention [14]</i>	Tomato	Multilevel fusion + attention	≈99.83	No

<i>PLDPNet: Hybrid CNN Framework [15]</i>	Potato / Tomato	Hybrid CNN	94.25 (Tomato)	No
<i>Multi-Level Deep Learning Model [16]</i>	Potato / Tomato	Multi-level CNN	96.71 (Tomato)	No
<i>Potato Leaf Disease Detection Using CNN [17]</i>	Potato	CNN	98.07	No
<i>Efficient DenseNet-Based Potato Leaf Disease Detection [18]</i>	Potato	Efficient DenseNet-201	97.20	No
<i>Disease Detection in Potato Leaves Using DL + SVM [19]</i>	Potato	Enhanced DL + SVM	99.42	No
<i>Leaf Diseases Classification in Bell Pepper Using VGGNet [20]</i>	Bell Pepper	VGG16 / VGG19	97 / 96	No
<i>Bell-Pepper Leaf Bacterial Spot Detection [21]</i>	Bell Pepper	AlexNet / VGG-16	97.87	No
<i>Detection of Bacterial Spot Disease Using YOLOv3 [22]</i>	Bell Pepper	YOLOv3 (detection)	90	Partial (region detection only)
<i>Optimized GSAtt-CMNetV3 for Pepper Leaf Disease [23]</i>	Pepper	GSAtt-CMNetV3 + Os-OA	97.87	Partial (attention, no RA pre-processing)
<i>CNN Based Apple Leaf Disease Detection [24]</i>	Apple	GoogleNet (pretrained)	99.79	No
<i>Apple Leaf Disease Detection using Color + DWT + RF [25]</i>	Apple	Color + DWT + Random Forest	98.63	No
<i>Recognizing Apple Leaf Diseases using Mask R-CNN [26]</i>	Apple	Mask R-CNN + Transfer Learning	96.6	Partial (segmentation, no attention)
<i>Efficient Deep Learning Approach for Apple Leaf Diseases</i>	Apple	CNN (transfer learning)	97.82	No
<i>A Novel Framework for Grape Leaf Disease Detection [27]</i>	Grape	Hybrid segmentation + ensemble classifier	98.7	Partial
<i>Deep Learning Precision Farming: Grapes & Mango Leaf Diseases [28]</i>	Grapes / Mango	AlexNet (pretrained)	99 (Grapes) 89 (Mango)	No

Table 3. AI SYSTEM DEVELOPMENT DEPTH/COMPLEXITY.

AI System Type	Level-1: Input & Preparation	Level-2: Feature Extraction	Level-3: Feature Refinement / Intelligence	Level-4: Decision / Classification
2-Level AI System	Preprocessing / Normalization / Augmentation	CNN / Feature Extractor	—	—
3-Level AI System	Preprocessing / Augmentation / Normalization	CNN / Transfer Learning	Attention / Optimization / Feature Selection	—

4-Level AI System (Advanced / Hybrid)	Preprocessing	+	CNN	CBAM / Attention /	Dense / Softmax
	Augmentation + Normalization		Backbone	Hybrid Intelligence	/SVM

Table 4. AI TECHNOLOGY INVOLVED AT EACH LAYER OF THE AI SYSTEM.

Level	Possible Techniques
Level-1 (Input & Preparation)	Pre-processing, normalization, data augmentation, CLAHE, background removal, segmentation
Level-2 (Feature Extraction)	CNN, ResNet, VGG, DenseNet, EfficientNet, Autoencoder, Transfer Learning
Level-3 (Refinement / Intelligence)	CBAM, SE-block, Attention, ViT, PCA, GWO, PSO, LSTM
Level-4 (Decision Layer)	Dense + Softmax, SVM, KNN, Random Forest, XGBoost

Based on the literature survey, each level of a level-based AI system can be implemented using a range of techniques or their combinations, as summarized in Table 4. The input level commonly includes image pre-processing, normalization, and data augmentation to improve visual consistency and reduce data variability. The feature extraction level is typically realized using convolutional neural networks or transfer-learning architectures capable of learning hierarchical and discriminative representations.

The feature refinement level further enhances these representations by emphasizing informative features and suppressing redundancy, often through attention mechanisms, optimization algorithms, or transformer-based modules. The decision level produces the final classification output and is commonly implemented using fully connected layers with softmax activation or, in hybrid systems, classical machine-learning classifiers such as SVM or k-NN. Together, the techniques listed in Table 4 illustrate the modular and flexible nature of multi-level AI system design, enabling architectures to be adapted to task complexity, data characteristics, and deployment requirements.

Based on this categorization, the proposed REA-CNN corresponds to a four-level AI system: region-aware pre-processing and enhancement at Level-1, CNN-based deep feature extraction at Level-2, CBAM-based attention refinement at Level-3, and dense softmax classification at Level-4.

2. Proposed System

Conventional convolutional neural networks (CNNs) learn discriminative representations directly from raw input images. Given an RGB input image:

$$I \in \mathbb{R}^{H \times W \times 3}$$

a standard CNN relies on hierarchical feature learning to implicitly suppress irrelevant visual information. While effective under controlled imaging conditions, this assumption is often violated in real agricultural environments, where leaf images exhibit complex backgrounds, illumination variability, shadows, occlusions, and viewpoint changes. Under such conditions, convolutional filters may respond strongly to background textures rather than disease-related patterns, resulting in degraded generalization. In our earlier work, this limitation was partially addressed using a hybrid CNN-CBAM-SVM framework, in which attention-enhanced deep features were classified using a separate SVM. Although this approach improved class separability, it exhibits two fundamental limitations:

1. The absence of explicit background suppression at the input level, and
2. The decoupling of feature learning and classification, which prevents end-to-end optimization and limits task-driven feature adaptation.

To overcome these limitations, we propose a Region-Aware and Enhanced Attention Convolutional Neural Network (REA-CNN), an end-to-end learning framework that integrates explicit region-aware pre-processing with attention-guided feature refinement. By suppressing

background interference prior to convolutional feature extraction and embedding attention directly within the CNN backbone, the proposed REA-CNN learns disease-relevant representations that are more robust to real-world imaging variability. This unified design improves robustness and consistency compared to conventional CNNs and the previously proposed CNN-CBAM-SVM model, as summarized in Table 5.

Table 5. COMPARISON OF THE REA-CNN SYSTEM WITH CNN AND THE HYBRID-1 SYSTEM.

CNN Model	Hybrid-1 [29]	Proposed Hybrid
Input Leaf Image ↓ Image Pre-processing Resize, normalization, augmentation ↓ CNN Backbone ResNet50 pretrained ↓ Global Feature Vector Average Pooling ↓ Fully Connected Layer ↓ Softmax ↓ Disease / Healthy Label	Input Leaf Image ↓ Image Pre-processing Resize, normalize , augmentation ↓ CNN Backbone ResNet50 pretrained ↓ CBAM (Channel + Spatial Attention) ↓ Deep Feature Vector ↓ SVM Classifier ↓ Disease / Healthy Label	Input Leaf Image ↓ Region-Aware Processing Leaf-focused, BG Suppression GrabCut-based leaf extraction ↓ CNN Backbone ResNet50 pretrained ↓ CBAM Attention Module ↓ Global Average Pooling ↓ Fully Connected Layer ↓ Softmax ↓ Disease / Healthy Label
	Stage 1: CNN + CBAM trained using Softmax loss Stage 2: CNN frozen and SVM trained on extracted deep features	Single-stage, end-to-end training
Joint optimization of feature extraction and classification	Extract strong features, then classify robustly.	Focus on the right regions, attend to disease patterns, and learn everything end-to-end.
<ul style="list-style-type: none"> ✓ End-to-end training ✓ Single loss function ✗ No explicit attention ✗ No region isolation Best for: Clean benchmark datasets (e.g., PlantVillage) Weakness: Sensitive to background and dataset bias 	<ul style="list-style-type: none"> ✗ Not end-to-end ✓ Attention focuses on disease regions ✓ SVM improves margin-based generalization Best for: Small datasets, controlled environments Weakness: More complex training, harder deployment 	<ul style="list-style-type: none"> ✓ End-to-end ✓ Attention-enhanced ✓ Region-focused ✓ Robust to background & noise ✓ Simple deployment Best for: Real-world images and practical deployment Weakness: Slightly more complex than CNN-only

2.1. REA-CNN: Architecture and Learning Framework

REA-CNN is an end-to-end deep learning framework that integrates region-aware processing and attention-enhanced feature learning within a unified CNN architecture. The primary objective of this design is to suppress irrelevant background information while emphasizing disease-specific visual patterns, enabling reliable classification under diverse and unconstrained imaging conditions.

Unlike conventional CNN pipelines that operate directly on full images, REA-CNN explicitly guides the learning process toward leaf-centric and disease-relevant regions. This design improves generalization beyond controlled benchmark datasets and enhances suitability for real agricultural deployment.

As illustrated in Figure 1, the REA-CNN architecture is implemented as a sequence of five processing stages: region-aware pre-processing, CNN-based feature extraction, attention-based feature refinement, feature aggregation, and classification. These stages represent functional processing blocks and are collectively mapped onto the four-level system design introduced earlier. Specifically, feature aggregation and classification are treated as part of a single decision level in the

level-based framework, while being shown as separate stages here for architectural clarity. Formally, the proposed model can be expressed as a composite function $f(I; \theta)$:

$$f(I; \theta) = f_{cls}(f_{gap}(f_{ea}(f_{cnn}(f_{ra}(I))))))$$

where

$f_{ra}(\cdot)$ denotes region-aware pre-processing

$f_{cnn}(\cdot)$ is the CNN backbone

$f_{ea}(\cdot)$ represents the CBAM attention module

$f_{gap}(\cdot)$ denotes global average pooling

$f_{cls}(\cdot)$ is the classification head and

$\theta = \{\theta_{ra}, \theta_{cnn}, \theta_{ea}, \theta_{cls}\}$ is everything learned by the network during a training

As illustrated in **Figure 1**, the REA-CNN pipeline consists of five sequential stages. Here we are taking GAP and dense layer separately, otherwise it will be equal to 4 layers as mentioned in Table 3:

$I \rightarrow RA \rightarrow CNN \text{ Backbone} \rightarrow CBAM (EA) \rightarrow GAP \rightarrow Dense+Softmax$

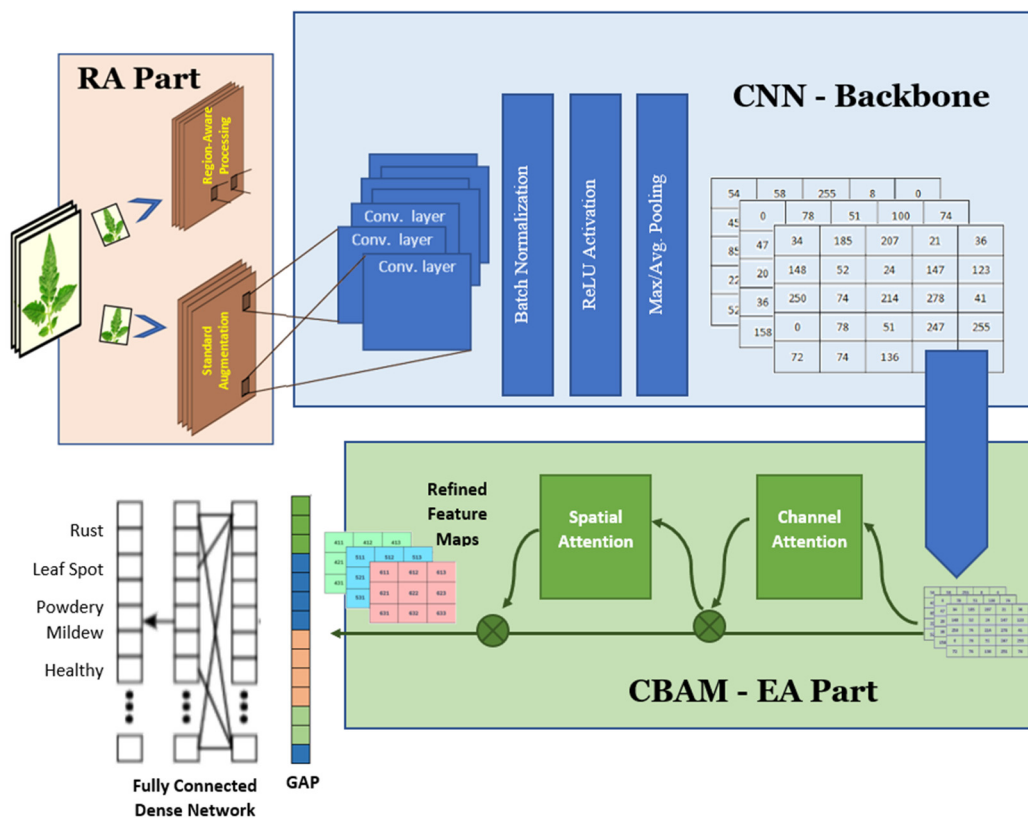


Figure 1. RA/EA-CNN SYSTEM ARCHITECTURE.

2.2. Stage I – Region-Aware Processing (RA)

The objective of the region-aware processing stage is to suppress irrelevant background information and emphasize the leaf region before deep feature extraction. By isolating the leaf and enhancing its visual characteristics, this stage reduces background interference and illumination variability, providing a refined input for the subsequent CNN backbone. Let the input RGB image be denoted as

$$I \in \mathbb{R}^{H \times W \times 3}$$

where H and W represent the image height and width, respectively. To isolate the leaf region, a binary foreground mask, one can define the mask as follows:

$$M \in \{0,1\}^{H \times W}$$

The mask M is estimated using mask-based or segmentation-driven background removal techniques. In this mask, $M(x, y) = 1$ indicates leaf pixels, while $M(x, y) = 0$ corresponds to background regions. The region-aware image is obtained through element-wise multiplication of the input image with the binary mask:

$$I_{RA}(x, y, c) = I(x, y, c) M(x, y)$$

where (x, y) denotes the spatial pixel location and $c \in \{R, G, B\}$ represents the color channel. This operation preserves pixel values within the leaf region and suppresses background pixels across all channels, resulting in a leaf-centric image representation.

Although background suppression isolates the leaf region, variations in illumination, contrast, and color distribution may still affect visual consistency. To address these issues, classical image enhancement operations are applied to the region-aware image I_{RA} .

A. Contrast normalization: First, contrast normalization is performed using *min-max* normalization:

$$I' = \frac{I_{RA} - \min(I_{RA})}{\max(I_{RA}) - \min(I_{RA}) + \varepsilon}$$

where the minimum and maximum values are combined over all pixel intensities of I_{RA} , and $\varepsilon = 10^{-8}$ is a small constant added to ensure numerical stability. This operation scales pixel values to a normalized range and reduces the effect of illumination bias.

B. Color normalization and brightness/saturation adjustment: Next, color normalization and brightness–saturation adjustment is applied to improve chromatic consistency and highlight disease-relevant patterns:

$$I_{ENH} = N(I')$$

where $N(\cdot)$ denotes the combined color normalization and enhancement function. These operations improve visual clarity, enhance symptom visibility, and reduce intra-class variability caused by lighting conditions.

C. Output of Region-Aware Processing: The final output of this stage, I_{ENH} constitutes a refined, leaf-centric, and visually enhanced image representation. As illustrated in Figure 2, this pipeline includes foreground–background separation, background suppression, and final image enhancement, producing an input that emphasizes disease-relevant visual features while minimizing environmental noise. This enhanced representation is then forwarded to the CNN backbone for robust feature extraction.

Figure 2 shows the region-aware enhancement pipeline, including background–foreground separation, background suppression, and final image enhancement, resulting in a refined leaf-centric representation that emphasizes disease-relevant visual features.

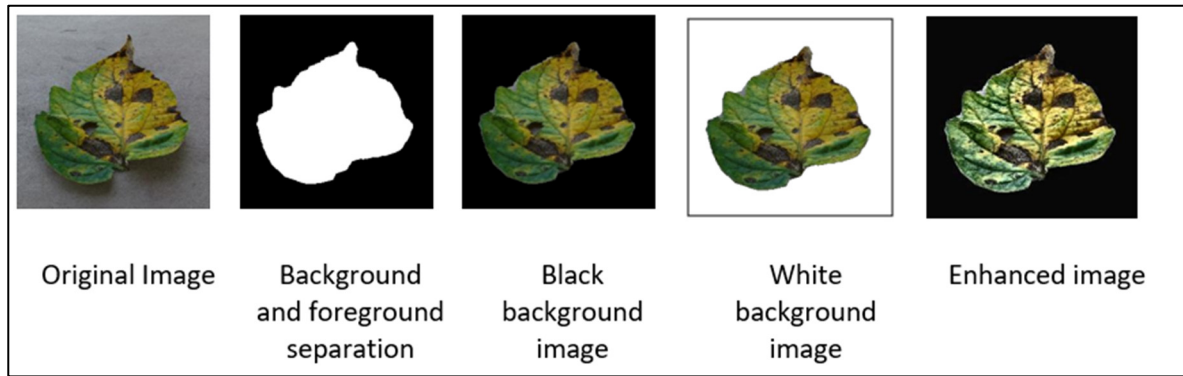


Figure 2. IMAGE ENHANCEMENT.

2.3. Stage II – CNN Backbone (Feature Extraction)

The enhanced leaf-centric image obtained from the region-aware processing stage, denoted as I_{ENH} , is forwarded to a pretrained convolutional neural network (CNN) backbone for hierarchical feature extraction. In this work, a ResNet50 architecture initialized with ImageNet pretrained weights is employed due to its strong representation capability and proven generalization performance.

The CNN backbone progressively transforms the input image into high-level feature representations through a sequence of convolution, normalization, nonlinearity, and pooling operations.

A. Convolutional Feature Extraction: At the l^{th} convolutional layer, the feature maps are computed as:

$$F^{(l)} = W^{(l)} * F^{(l-1)} + b^{(l)}$$

where $F^{(l-1)}$ and $F^{(l)}$ denote the input and output feature maps of layer l , respectively, $W^{(l)}$ represents the convolutional filter weights, $b^{(l)}$ is the bias term, and the convolution product is denoted by “*”. Through this process, local spatial patterns such as edges, textures, and disease-related visual cues are captured at increasing levels of abstraction.

B. Batch Normalization: To stabilize training and accelerate convergence, batch normalization is applied to the convolutional outputs:

$$\hat{F}^{(l)} = \frac{F^{(l)} - \mu_B}{\sigma_B + \varepsilon}$$

$$F_{BN}^{(l)} = \gamma \hat{F}^{(l)} + \beta$$

where μ_B and σ_B denote the mean and the standard deviation computed over the mini-batch, ε is a small constant added for numerical stability, and γ and β are learnable scaling and shifting parameters. This normalization reduces internal covariate shift and improves optimization stability.

C. Nonlinearity and Pooling: Nonlinear activation is introduced using the Rectified Linear Unit (ReLU):

$$F_{ReLU}^{(l)} = \max(0, F_{BN}^{(l)})$$

which enables the network to model complex nonlinear relationships and prevents vanishing gradients. To further enhance robustness and spatial invariance, spatial pooling is applied:

$$F_{pool}(i, j) = \max_{(m, n) \in \Omega(i, j)} F_{ReLU}^{(l)}(m, n)$$

where $\Omega(i, j)$ defines the local pooling neighbourhood. Pooling reduces spatial resolution while preserving dominant activations, allowing the network to focus on salient disease-related structures.

D. Output of CNN Backbone: After multiple stacked convolutional blocks, the CNN backbone produces a high-level feature tensor:

$$F \in \mathbb{R}^{\hat{H} \times \hat{W} \times C}$$

where \hat{H} and \hat{W} are the spatial dimensions of the extracted feature maps and C denotes the number of feature channels. This tensor encodes rich semantic information about leaf structure and disease patterns and serves as the input to the subsequent attention-enhancement stage.

2.4. Stage III – Enhanced Attention via CBAM (EA)

To further refine the feature representations extracted by the CNN backbone, the proposed REA-CNN employs a **Convolutional Block Attention Module (CBAM)**. CBAM sequentially applies **channel attention** followed by **spatial attention**, allowing the network to emphasize informative feature channels and spatial locations while suppressing irrelevant or noisy responses. This attention-guided refinement enables more discriminative learning of disease-specific visual patterns. Let the high-level feature tensor denoted by:

$$F \in \mathbb{R}^{\hat{H} \times \hat{W} \times C}$$

Channel Attention Mechanism

The channel attention module focuses on identifying **which feature channels are more important** for disease classification.

- A. Channel Descriptor Generation:** Global information is first aggregated from the feature tensor using both global average pooling, $GAP(\cdot)$, and global max pooling, $GMP(\cdot)$:

$$z_{avg} = GAP(F)$$

$$z_{max} = GMP(F)$$

where $z_{avg}, z_{max} \in \mathbb{R}^C$ summarize channel-wise statistics and capture complementary contextual information.

- B. Channel Weight Estimation:** The pooled descriptors are passed through a shared multi-layer perceptron (MLP) to compute channel attention weights:

$$s = \sigma(W_2 \delta(W_1 z))$$

where W_2 and W_1 are learnable weight matrices, $\delta(\cdot)$ denotes the ReLU activation, and $\sigma(\cdot)$ is the sigmoid function. The resulting vector $s \in \mathbb{R}^C$ assigns an importance weight to each feature channel.

- C. Channel-wise Feature Refinement:** The original feature tensor is refined by channel-wise multiplication:

$$F_{CA} = F \otimes s$$

where \otimes denotes element-wise multiplication with broadcasting across spatial dimensions. This operation amplifies disease-relevant channels while suppressing less informative ones.

2.5. Spatial Attention Mechanism

While channel attention emphasizes *what* to focus on, spatial attention identifies *where* important features are located within the image.

- A. Spatial Descriptor Computation:** Spatial descriptors are generated by applying average and max pooling across the channel dimension:

$$F_{avg} = \text{mean}_c(F_{CA})$$

$$F_{max} = \text{max}_c(F_{CA})$$

where both descriptors capture complementary spatial cues related to disease symptoms.

B. Spatial Attention Map Generation: The pooled feature maps are concatenated and passed through a convolutional layer to compute the spatial attention map:

$$A_s = \sigma \left(\text{Conv}_{7 \times 7} (F_{avg}, F_{max}) \right)$$

where $\text{Conv}_{7 \times 7}(\cdot)$ denotes a convolution with a 7×7 kernel and $\sigma(\cdot)$ is the sigmoid activation. The resulting map $A_s \in \mathbb{R}^{\hat{H} \times \hat{W}}$ highlights spatial locations that are most relevant to disease patterns.

C. Spatial Feature Refinement: The final attention-enhanced feature tensor is obtained as:

$$F_{EA} = F_{CA} \otimes A_s$$

which selectively emphasizes disease-relevant regions while suppressing background or irrelevant spatial responses.

D. Output of Enhanced Attention Stage: The output F_{EA} represents an attention-refined feature tensor that integrates both channel-wise and spatial importance. By explicitly modelling *what* features and *where* they matter, the CBAM module significantly improves feature discriminability and robustness, particularly for visually similar plant diseases. This refined representation is subsequently forwarded to the feature aggregation and classification stages.

2.6. Stage IV – Feature Aggregation

The attention-refined feature tensor produced by the CBAM module,

$$F_{EA} \in \mathbb{R}^{\hat{H} \times \hat{W} \times C}$$

still retains spatial dimensions. To convert these spatial features into a compact and discriminative representation suitable for classification, **Global Average Pooling (GAP)** is applied.

Global Average Pooling: For each channel c , GAP computes the average response over all spatial locations:

$$v_c = \frac{1}{\hat{H}\hat{W}} \sum_{i=1}^{\hat{H}} \sum_{j=1}^{\hat{W}} F_{EA}(i, j, c)$$

where \hat{H} and \hat{W} denote the height and width of the feature maps, respectively. By aggregating spatial information in this manner, GAP produces a compact feature vector:

$$v \in \mathbb{R}^C$$

Role of Feature Aggregation: Global Average Pooling acts as a **structural regularizer**, significantly reducing the number of trainable parameters and mitigating overfitting. Moreover, since GAP directly aggregates the attention-refined feature maps, it preserves the semantic consistency enforced by the preceding channel and spatial attention mechanisms. The resulting feature vector v therefore encodes global, disease-relevant information in a compact and robust form.

2.7. Stage V – Classification Head

The final classification stage maps the aggregated feature vector v to class scores corresponding to the plant disease categories.

Fully Connected Layer: A fully connected layer is used to compute the logits:

$$o = W_{fc}v + b_{fc}$$

where W_{fc} and b_{fc} denote the learnable weights and bias of the classification layer, respectively.

Softmax Classification: The logits are transformed into normalized class probabilities using the

$$P(y = k|I) = \frac{e^{o_k}}{\sum_{i=1}^K e^{o_i}}$$

where K is the number of disease classes and o_k represents the logit corresponding to class k .

Output Interpretation: The predicted class corresponds to the category with the highest posterior probability. By combining attention-refined features, global aggregation, and end-to-end optimization, this classification head enables reliable and robust decision-making for plant leaf disease recognition under real-world imaging conditions.

2.8. End-to-End Learning Strategy

The proposed REA-CNN is trained in an **end-to-end manner** using a single objective function, allowing all network components from region-aware processing to final classification to be jointly optimized.

Loss Function: Let K denote the number of disease classes. The model is optimized using the **categorical cross-entropy loss**, defined as:

$$L = - \sum_{k=1}^K y_k \log P(y = k|I)$$

Where $y_k \in \{0,1\}$ is the ground-truth label for class k , and $P(y = k|I)$ denotes the predicted probability for class k obtained from the softmax output of the network. This loss encourages the model to assign high probability to the correct disease class while penalizing incorrect predictions.

Gradient Propagation and Joint Optimization: During training, gradients of the loss function are propagated backward through the entire network:

$$\frac{\partial L}{\partial \theta} \rightarrow \text{Classification Head} \rightarrow \text{GAP} \rightarrow \text{CBAM} \rightarrow \text{CNN Backbone} \rightarrow \text{Region - Aware Input}$$

where Θ denotes the set of all trainable parameters of the model. This uninterrupted gradient flow ensures that:

- The **classification objective directly influences feature learning**,
- The attention weights are dynamically adapted based on task relevance, and
- The convolutional filters learn representations that are aligned with disease-specific visual patterns.

Significance of End-to-End Learning: Unlike multi-stage or hybrid pipelines that decouple feature extraction and classification, the proposed REA-CNN benefits from **joint optimization across all stages**. This unified learning strategy enables task-driven adaptation of region awareness, attention mechanisms, and deep feature representations simultaneously, resulting in improved robustness, faster convergence, and better generalization to real-world plant leaf images. Key Advantages:

1. *Explicit background suppression reduces input entropy*
2. *Attention enhances semantic and spatial discriminability*
3. *GAP improves generalization and parameter efficiency*
4. *Fully end-to-end differentiability ensures robustness under domain shift*

2.9. Comparison with Previous Hybrid Models

Unlike CNN+CBAM+SVM architectures:

- REA-CNN does **not** freeze feature extractors
- The classifier influences feature learning directly
- Attention mechanisms adapt dynamically during training

This results in improved robustness, generalization, and deployment simplicity.

3. Experimental Setup

Experiments were conducted on the PlantVillage dataset and its region-aware enhanced variant (PlantVillage_BG_ENH), comprising approximately 32,000 leaf images across multiple disease classes. No pixel-level lesion annotations were used; region awareness was achieved using classical unsupervised background suppression. The enhanced dataset (PlantVillage_BG_ENH) was generated by applying region-aware background suppression and image enhancement to the original PlantVillage images. Images were resized to 224×224 pixels and normalized using ImageNet pre-processing. The dataset was split into 70% training, 15% validation, and 15% testing using class-stratified sampling. The proposed REA-CNN employed a ResNet50 backbone pretrained on ImageNet, followed by a CBAM attention module and a global average pooling-based classification head. Training was performed end-to-end using the Adam optimizer with an initial learning rate of 10^{-4} , categorical cross-entropy loss, and early stopping based on validation accuracy.

3.1. Evaluation Metrics

The performance of the proposed REA-CNN model was evaluated using Top-1 accuracy, Top-5 accuracy, and categorical cross-entropy loss. Given the balanced nature of the dataset, these metrics provide a reliable and comprehensive assessment of classification correctness, ranking quality, and prediction confidence.

Top-1 Accuracy: Top-1 accuracy measures the proportion of test samples for which the predicted class with the highest probability matches the ground-truth label. It is defined as:

$$TOP - 1: Accuracy = \frac{1}{N} \sum_{i=1}^N 1(\operatorname{argmax}_{k \in \{1, \dots, K\}} (P(y = k | I_i)) = y_i)$$

where N denotes the total number of samples, y^i is the true class label of sample i , I_i is the corresponding input image, $\operatorname{argmax}(\cdot)$ provides the argument that maximizes the function, and $1(\cdot)$ is the indicator function. Top-1 accuracy directly reflects the model's classification precision under strict decision criteria. To analyze model behavior at different stages of training, three accuracy values are reported:

- **acc:** training accuracy,
- **val_acc:** validation accuracy,
- **test_acc:** final test accuracy, representing the true generalization performance.

Top-5 Accuracy: Top-5 accuracy considers a prediction correct if the ground-truth class appears among the five most probable predicted classes:

$$TOP - 5: Accuracy = \frac{1}{N} \sum_{i=1}^N 1(y_i \in \operatorname{Top5}(\{P(y = k | I_i)\}_{k=1}^K))$$

This metric is particularly relevant for plant disease classification, where visually similar symptoms can lead to ambiguity between multiple disease categories. A high Top-5 accuracy indicates that the model effectively captures the underlying disease structure even when fine-grained distinctions are challenging.

3.2. Result Analysis

The confusion matrix shown in Figure 3 indicates strong overall classification performance of the proposed REA-CNN, with clear diagonal dominance across most classes, reflecting high true-positive rates. Classes such as *Healthy*, *Target Spot*, *Yellow Leaf Curl Virus*, and *Spider Mite (Two-Spotted)* exhibit particularly high correct classification rates, suggesting that region-aware background suppression and attention-enhanced feature learning effectively emphasize disease-relevant visual patterns. Misclassifications are primarily observed between visually similar disease categories, notably *Early Blight* and *Late Blight*, as well as between *Bacterial Spot* and *Late Blight*. These confusions are expected due to

overlapping texture and color characteristics in leaf symptoms. Minor confusion is also observed between *Septoria Leaf Spot* and *Leaf Mold*, indicating residual inter-class similarity despite attention-based refinement. Overall, the confusion matrix analysis shows that the proposed REA-CNN reduces inter-class confusion across most disease categories while maintaining consistent performance, supporting its applicability to plant disease classification under complex background conditions.

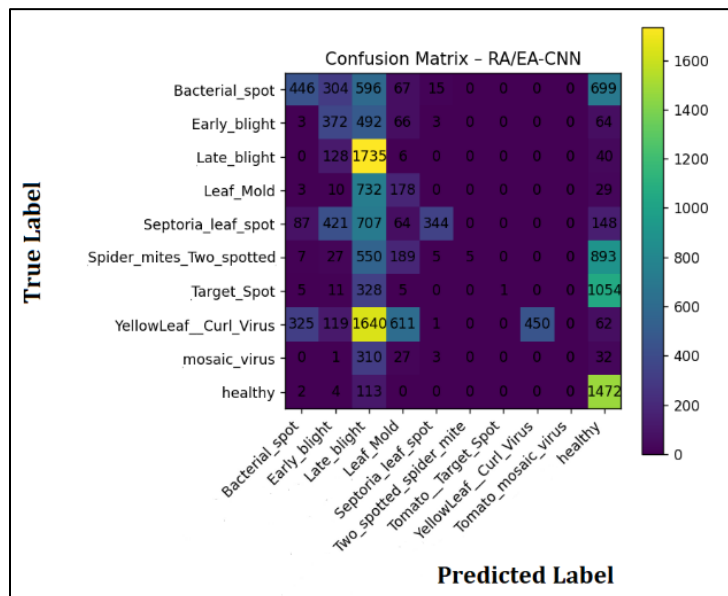


Figure 3. CONFUSION MATRIX FOR REA-CNN MODEL.

The proposed REA-CNN model was evaluated against the Hybrid-1 (CNN+CBAM+SVM) architecture over 30 training epochs, as shown in Figure 4, where the x-axis represents the number of epochs and the y-axis denotes training accuracy (%). The REA-CNN exhibits faster convergence, achieving high training accuracy within the initial epochs, whereas the Hybrid-1 model demonstrates a more gradual learning trend. Throughout training, REA-CNN maintains consistently higher accuracy, approaching approximately 99% by the final epochs, compared to around 96–97% for Hybrid-1. This behaviour indicates that the integration of region-aware pre-processing and attention-enhanced learning facilitates more efficient feature learning during early training stages, contributing to improved convergence and overall training performance.

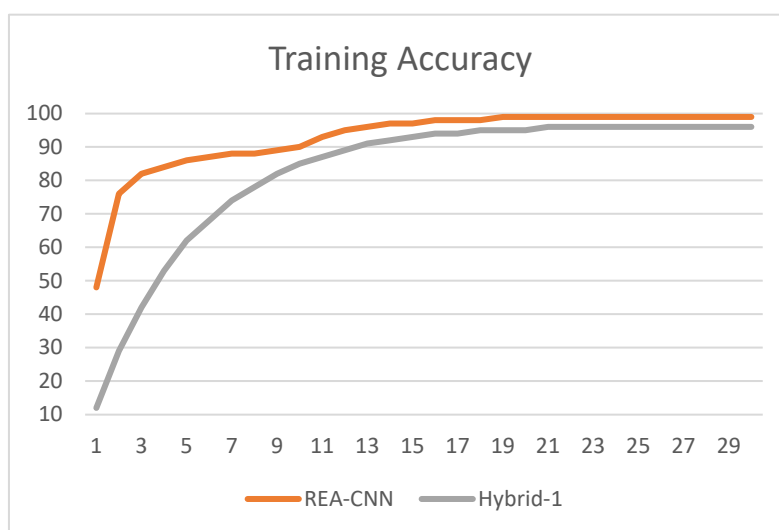


Figure 4. COMPARISON OF TRAINING ACCURACY OF THE PROPOSED MODEL VS HYBRID-1.

Training accuracy is reported to illustrate convergence behaviour, whereas validation accuracy is used to assess generalization performance. Figure 5 compares the validation accuracy of the proposed REA-CNN with the Hybrid-1 (CNN+CBAM+SVM) model. The REA-CNN achieves higher validation accuracy during the early training epochs, indicating faster stabilization on unseen data. Although the Hybrid-1 model gradually improves and approaches comparable validation accuracy in later epochs, it requires a larger number of training iterations to do so. In contrast, the REA-CNN validation curve remains consistently high with limited fluctuations, suggesting improved stability during training. Overall, the observed validation behaviour indicates that the proposed framework achieves more efficient convergence and maintains consistent generalization performance, supporting its applicability to plant disease classification under challenging imaging conditions.

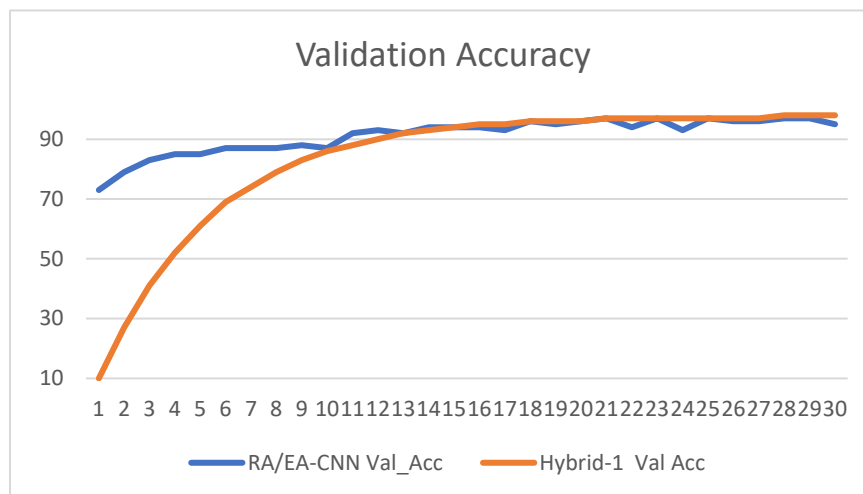


Figure 5. COMPARISON OF VALIDATION ACCURACY OF THE PROPOSED MODEL VS HYBRID-1.

Figure 6 illustrates the relationship between overall validation accuracy (Top-1) and Top-5 validation accuracy of the proposed REA-CNN model over 30 training epochs, where the x-axis represents epochs and the y-axis denotes accuracy (%). The Top-5 accuracy rapidly approaches nearly 100% within the initial epochs and remains saturated throughout training, indicating that the ground-truth class is consistently included among the model's top predictions. In contrast, the overall validation accuracy exhibits a more gradual and stable improvement, converging to approximately 95–97% in later epochs. The observed gap between Top-5 and Top-1 accuracy reflects the presence of fine-grained visual similarities among certain disease classes, which can lead to occasional misclassification in the top prediction. Overall, the comparison between Top-1 and Top-5 validation accuracy indicates strong class separability and effective feature representation learning by the proposed REA-CNN, supporting its suitability for multi-class plant disease classification tasks.

To examine the behavior of the proposed REA-CNN in realistic conditions, we conducted a small-scale experiment using 12 tomato leaf images representing different disease types, including target spot, mosaic virus, and spider mite (Figure 7). These images were intentionally selected to reflect real-world variability, with noticeable differences in image resolution, leaf size and shape, background clutter, and lighting conditions. Unlike laboratory datasets, the test samples were not standardized or visually optimized, allowing the evaluation to focus on the model's generalization capability under practical conditions. Model performance was evaluated using Top-1 accuracy, Top-5 accuracy, and average prediction confidence. The proposed REA-CNN correctly identified 9 out of 12 images at Top-1, while the remaining samples were correctly included within the Top-5 predictions, resulting in complete Top-5 coverage. In addition, the model maintained a consistently high confidence level, exceeding 70%, indicating stable and reliable predictions even in the presence of background noise and illumination variation.

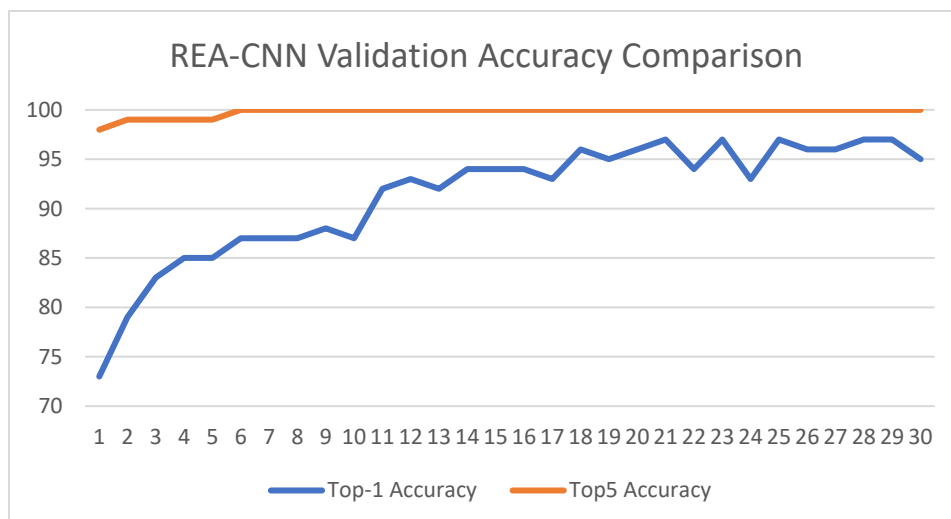


Figure 6. COMPARISON OF OVERALL VALIDATION ACCURACY VS TOP-5 ACCURACIES OF THE PROPOSED MODEL.

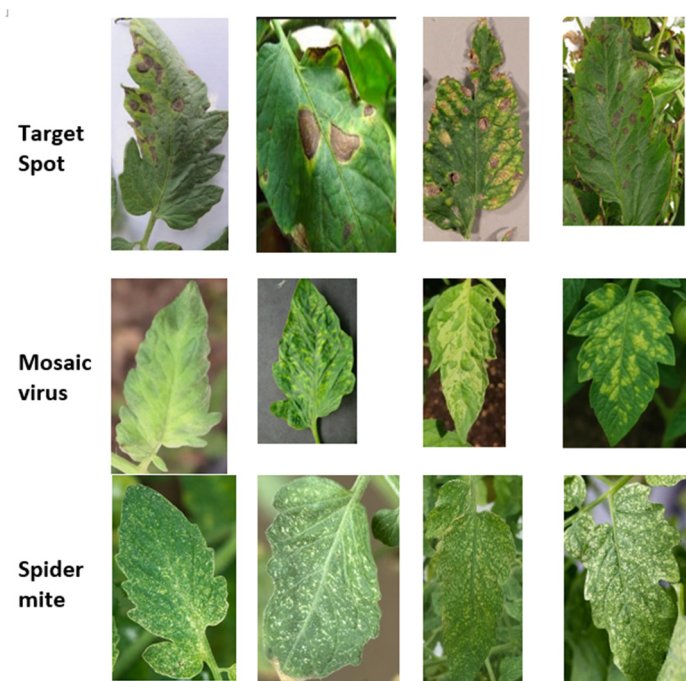


Figure 7. DATA SET OF REAL IMAGES.

For comparison (Figure 8), the existing Hybrid-1 (CNN+CBAM+SVM) framework showed limited robustness on the same test set, correctly identifying only one image with a very low confidence score of 1.25%. This comparison suggests that the explicit incorporation of region awareness and attention-based feature enhancement enables REA-CNN to better focus on disease-relevant visual patterns, leading to improved accuracy and confidence in real-world scenarios. Overall, the experiment demonstrates the suitability of the proposed approach for deployment beyond controlled laboratory environments.

Metric	REA-CNN	Hybrid-1
Top-1 Correct (out of 12)	9 (75%)	1 (8.5%)
Top-5 Correct (out of 12)	12 (100%)	5 (41%)
Average Confidence (%)	70+	1.25

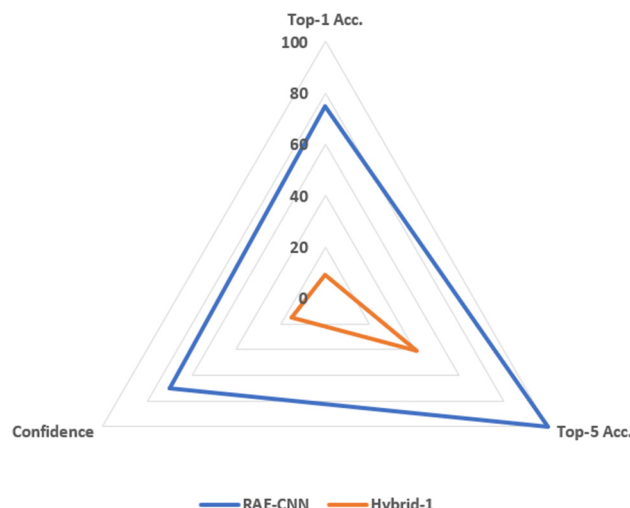


Figure 8. HIGH ACCURACY AND CONFIDANCE IS OBSERVED FOR PROPOSED REA-CNN MODEL.

4. Conclusion

This paper presented REA-CNN, a region-aware and attention-enhanced convolutional neural network for plant leaf disease classification under real agricultural imaging conditions. Unlike conventional CNN-based and hybrid pipelines that rely primarily on implicit feature learning or multi-stage optimization, the proposed framework integrates explicit region-aware pre-processing and CBAM-based attention within a single end-to-end architecture. This design enables the model to focus on leaf-centric and disease-relevant representations while maintaining uninterrupted gradient flow during training. Experimental evaluation shows that REA-CNN exhibits improved convergence behaviour and consistent classification performance compared to conventional CNNs and CNN–CBAM–SVM hybrid systems. The confusion matrix and accuracy trends indicate reduced inter-class confusion, particularly for visually similar disease categories, while attention-based feature refinement contributes to improved discriminative learning. These observations suggest that combining input-level refinement with feature-level attention can be effective for improving robustness beyond controlled benchmark conditions.

Overall, the proposed REA-CNN provides a unified and practical framework for plant disease recognition that is better suited to real-world deployment scenarios. Future work will focus on extending the framework to larger and more diverse field datasets, exploring lightweight backbone architectures for edge deployment, and incorporating temporal and multispectral information to further enhance early disease detection in precision agriculture applications.

References

1. I. Yağ and A. Altan, "Artificial Intelligence-Based Robust Hybrid Algorithm Design and Implementation for Real-Time Detection of Plant Diseases in Agricultural Environments," *Biology (Basel)*, vol. 11, no. 12, p. 1732, Nov. 2022, doi: 10.3390/biology11121732.
2. P. Kamencay, M. Benco, T. Mizdos, and R. Radil, "A New Method for Face Recognition Using Convolutional Neural Network," *Advances in Electrical and Electronic Engineering*, vol. 15, no. 4, Nov. 2017, doi: 10.15598/aeer.v15i4.2389.
3. P. Bedi and P. Gole, "Plant disease detection using hybrid model based on convolutional autoencoder and convolutional neural network," *Artificial Intelligence in Agriculture*, vol. 5, pp. 90–101, 2021, doi: 10.1016/j.aiaa.2021.05.002.
4. N. Hashmi and M. Haroon, "A Hybrid AI-Based Approach for Early and Accurate Rice Disease Detection," *Fusion: Practice and Applications*, vol. 21, no. 1, 2026, doi: 10.54216/FPA.210107.

5. M. F. Kabir, I. S. Rahat, C. Beverley, R. Uddin, and S. Kant, "Tealeafnet-gwo: an intelligent CNN-Transformer hybrid framework for tea leaf disease detection using gray wolf optimization," *Discover Artificial Intelligence*, vol. 5, no. 1, p. 377, Dec. 2025, doi: 10.1007/s44163-025-00686-y.
6. M. Khalid and MD. A. Talukder, "A Hybrid Deep Multistacking Integrated Model for Plant Disease Detection," *IEEE Access*, vol. 13, pp. 116037–116053, 2025, doi: 10.1109/ACCESS.2025.3583793.
7. S. Aboelenin, F. A. Elbasheer, M. M. Eltoukhy, W. M. El-Hady, and K. M. Hosny, "A hybrid Framework for plant leaf disease detection and classification using convolutional neural networks and vision transformer," *Complex & Intelligent Systems*, vol. 11, no. 2, p. 142, Feb. 2025, doi: 10.1007/s40747-024-01764-x.
8. X. Huang, A. Chen, G. Zhou, X. Zhang, J. Wang, N. Peng, N. Yan, and C. Jiang, "Tomato Leaf Disease Detection System Based on FC-SNDPN," *Multimed Tools Appl*, vol. 82, no. 2, pp. 2121–2144, Jan. 2023, doi: 10.1007/s11042-021-11790-3.
9. P. Baser, J. R. Saini, and K. Kotecha, "TomConv: An Improved CNN Model for Diagnosis of Diseases in Tomato Plant Leaves," *Procedia Computer Science*, vol. 218, pp. 1825–1833, 2023, doi: 10.1016/j.procs.2023.01.160.
10. M. Arafath, A. A. Nithya, and S. Gijwani, "Tomato Leaf Disease Detection Using Deep Convolution Neural Network," *Procedia Computer Science*, Feb. 2023, pp. 236–245. doi: 10.4028/p-vph2n1.
11. K. Indira and H. Mallika, "Classification of Plant Leaf Disease Using Deep Learning," *Journal of The Institution of Engineers (India): Series B*, vol. 105, no. 3, pp. 609–620, Jun. 2024, doi: 10.1007/s40031-024-00993-5.
12. K. Roy *et al.*, "Detection of Tomato Leaf Diseases for Agro-Based Industries Using Novel PCA DeepNet," *IEEE Access*, vol. 11, pp. 14983–15001, 2023, doi: 10.1109/ACCESS.2023.3244499.
13. L. Zhang, G. Zhou, C. Lu, A. Chen, Y. Wang, L. Li, and W. Cai, "MMDGAN: A fusion data augmentation method for tomato-leaf disease identification," *Applied Soft Computing*, vol. 123, p. 108969, Jul. 2022, doi: 10.1016/j.asoc.2022.108969.
14. C.K. Sunil, C.D. Jaidhar, and N. Patil, "Tomato plant disease classification using Multilevel Feature Fusion with adaptive channel spatial and pixel attention mechanism," *Expert Syst Appl*, vol. 228, p. 120381, Oct. 2023, doi: 10.1016/j.eswa.2023.120381.
15. F. Arshad, M. Mateen, S. Hayat, M. Wardah, Z. Al-Huda, Y. H. Gu, and M. A. Al-antari, "PLDPNet: End-to-end hybrid deep learning framework for potato leaf disease prediction," *Alexandria Engineering Journal*, vol. 78, pp. 406–418, Sep. 2023, doi: 10.1016/j.aej.2023.07.076.
16. J. Rashid, I. Khan, G. Ali, S. H. Almotiri, M. A. AlGhamdi, and K. Masood, "Multi-Level Deep Learning Model for Potato Leaf Disease Recognition," *Electronics (Basel)*, vol. 10, no. 17, p. 2064, Aug. 2021, doi: 10.3390/electronics10172064.
17. P. Khobragade, A. Shriwas, S. Shinde, A. Mane, and A. Padole, "Potato Leaf Disease Detection Using CNN," *International IEEE Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)*, Dec. 2022, pp. 1–5. doi: 10.1109/SMARTGENCON56628.2022.10083986.
18. R. Mahum, H. Munir, Z. Mughal, M. Awais, F. Sher Khan, M. Saqlain, S. Mahmud, and I. Tlili "A novel framework for potato leaf disease detection using an efficient deep learning model," *Human and Ecological Risk Assessment: An International Journal*, vol. 29, no. 2, pp. 303–326, Feb. 2023, doi: 10.1080/10807039.2022.2064814.
19. B. Goyal, A. Kumar Pandey, R. Kumar, and M. Gupta, "Disease Detection in Potato Leaves Using an Efficient Deep Learning Model," *International IEEE Conference on Data Science and Network Security (ICDSNS)*, Jul. 2023, pp. 01–05. doi: 10.1109/ICDSNS58469.2023.10245369.
20. P. Kumar Das, "Leaf Disease Classification in Bell Pepper Plant using VGGNet," *Journal of Innovative Image Processing*, vol. 5, no. 1, pp. 36–46, Mar. 2023, doi: 10.36548/jiip.2023.1.003.
21. K. Kapoor, S. Singh, N. P. Singh, and Priyanka, "Bell-Pepper Leaf Bacterial Spot Detection Using AlexNet and VGG-16," 2023, pp. 507–519. doi: 10.1007/978-981-99-0838-7_44.
22. T. Y. Mahesh and M. P. Mathew, "Detection of Bacterial Spot Disease in Bell Pepper Plant Using YOLOv3," *IETE J Res*, vol. 70, no. 3, pp. 2583–2590, Mar. 2024, doi: 10.1080/03772063.2023.2176367.

23. S. S. A. Begum and H. Syed, "GSAtt-CMNetV3: Pepper Leaf Disease Classification Using Osprey Optimization," *IEEE Access*, vol. 12, pp. 32493–32506, 2024, doi: 10.1109/ACCESS.2024.3358833.
24. S. Fatima, R. Kaur, A. Doegar, and K. G. Srinivasa, "CNN Based Apple Leaf Disease Detection Using Pre-trained GoogleNet Model," 2023, pp. 575–586. doi: 10.1007/978-981-99-0047-3_49.
25. S. Hasan, S. Jahan, and Md. I. Islam, "Disease detection of apple leaf with combination of color segmentation and modified DWT," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 9, pp. 7212–7224, Oct. 2022, doi: 10.1016/j.jksuci.2022.07.004.
26. Z. U. Rehman, M. A. Khan, F. Ahmed, R. Damasevicius, S. R. Naqvi, W. Nisat, and K. Javed, "Recognizing apple leaf diseases using a novel parallel real-time processing framework based on MASK RCNN and transfer learning: An application for smart agriculture," *IET Image Process*, vol. 15, no. 10, pp. 2157–2168, Aug. 2021, doi: 10.1049/ipr2.12183.
27. N. Kaur and V. Devendran, "A novel framework for semi-automated system for grape leaf disease detection," *Multimed Tools Appl*, vol. 83, no. 17, pp. 50733–50755, Nov. 2023, doi: 10.1007/s11042-023-17629-3.
28. U. Sanath Rao, R. Swathi, V. Sanjana, L. Arpitha, K. Chandrasekhar, Chimayi, and P. K. Naik, "Deep Learning Precision Farming: Grapes and Mango Leaf Disease Detection by Transfer Learning," *Global Transitions Proceedings*, vol. 2, no. 2, pp. 535–544, Nov. 2021, doi: 10.1016/j.gltp.2021.08.002.
29. M. Altalak, M. A. Uddin, A. Alajmi, and A. Rizg, "A Hybrid Approach for the Detection and Classification of Tomato Leaf Diseases," *Applied Sciences*, vol. 12, no. 16, p. 8182, Aug. 2022, doi: 10.3390/app12168182.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.