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

Research Article

YOLOv9s with Region-Dispersion Channel Spatial Attention for Robust Chili Leaf Disease Detection

Miwan Kurniawan Hidayat [1,2,*], Jufriadif Na'am [1], and Ferda Ernawan [1,3]

[1] Faculty of Information Technology, Universitas Nusa Mandiri, Jakarta 13620, Indonesia; e-mail : miwan.hidayat@gmail.com; jufriadifnaam@nusamandiri.ac.id; ferda1902@gmail.com
[2] Faculty of Engineering and Informatics, Universitas Bina Sarana Informatika, Jakarta 10450, Indonesia; e-mail : miwan@bsi.ac.id
[3] Faculty of Computing, Universiti Malaysia Pahang Al-Sultan Abdullah, Pekan 26600, Pahang, Malaysia; e-mail : ferda@umpsa.edu.my
* Corresponding Author : Miwan Kurniawan Hidayat

Abstract: Abstract: Detecting chili leaf diseases remains challenging due to the non-uniform manifestation of symptoms, local discoloration, small lesion regions, and visual similarity between disease patterns and natural leaf background variations. Although YOLO-based detectors provide favorable computational efficiency, lightweight variants often struggle to distinguish subtle lesion characteristics, while conventional attention mechanisms such as CBAM primarily rely on global feature aggregation and may overlook regional activation variability. To address these limitations, this study proposes a YOLOv9s-based detection framework integrated with a Region-Dispersion Channel Spatial Attention (RDCSA) module. The proposed module incorporates regional dispersion statistics, namely mean, standard deviation, and range, as channel descriptors to capture inter-region feature variability before applying spatial attention refinement. Experiments were conducted on the COLD dataset containing 532 original images from five chili leaf condition categories using a split-before-augmentation protocol to ensure objective evaluation. RDCSA was integrated at the P5 feature level and evaluated through attention placement analysis, component-wise ablation, sensitivity analysis, stability assessment, and comparison with modern attention mechanisms. The proposed YOLOv9s + RDCSA model achieved an mAP@50 of 0.894, mAP@50–95 of 0.773, precision of 0.858, recall of 0.861, and an F1-score of 0.859 with only a marginal increase in model parameters. The results suggest that regional dispersion-based attention improves feature discrimination while preserving computational efficiency, particularly for disease symptoms characterized by heterogeneous spatial patterns. Nevertheless, performance remains influenced by visually ambiguous symptom categories, indicating that further validation across multiple datasets and field conditions is required. Overall, the proposed RDCSA module enhances detection capability without substantially increasing computational overhead, making it a promising attention mechanism for lightweight plant disease detection systems.

Keywords: Attention mechanism; Computer vision; Deep learning; Plant disease detection; Precision agriculture; Region-Dispersion Channel Spatial Attention; Sustainable agriculture; YOLOv9s.

Received: April, 26th 2026
Revised: May, 31st 2026
Accepted: June, 1st 2026
Published: June, 6th 2026

1. Introduction

Chili is an important horticultural commodity that contributes significantly to agricultural production and trade activities. However, chili productivity is frequently affected by diseases, pest infestations, nutritional deficiencies, and environmental stress. Leaf symptoms serve as primary visual indicators for assessing plant health, monitoring disease progression, and estimating potential yield losses. The magnitude of these impacts may vary depending on disease type, severity, crop management practices, environmental conditions, geographic location, and observation period. Consequently, disease detection systems require a clearly defined scope and evaluation protocols that are aligned with the disease categories represented in the dataset.

Although numerous chili diseases have been reported in the agricultural literature, this study focuses on the five leaf-condition classes available in the COLD dataset, namely Healthy, Cercospora, Mites and Thrips, Nutritional Deficiency, and Powdery Mildew [1]. These categories were selected because they represent diverse symptom manifestations, including healthy foliage, fungal infections, pest-induced damage, and nutritional disorders. The selected classes exhibit several challenging visual characteristics, such as non-uniform lesion distribution, local discoloration, small symptomatic regions, and visual similarity between disease symptoms and natural leaf texture variations. Such characteristics make reliable disease detection particularly challenging and provide a suitable benchmark for evaluating the discriminative capability of deep learning models. Focusing on these five categories enables a targeted assessment of both inter-class separability and the ability to recognize subtle disease symptoms under heterogeneous visual conditions.

Early detection of chili leaf diseases plays an important role in agronomic decision-making by enabling timely intervention, reducing excessive pesticide application, and supporting effective disease management strategies. Conventional manual inspection methods remain widely used but face several operational limitations, including lengthy inspection times, dependence on expert knowledge, and high labor requirements [2]. Their reliability may also be affected by variations in lighting conditions, plant growth stages, and observer subjectivity [3]. These limitations have encouraged the adoption of computer vision and deep learning approaches for automated disease detection, offering non-destructive analysis, faster processing, real-time responsiveness, and improved spatial precision [4].

Recent studies have demonstrated the effectiveness of convolutional neural networks and YOLO-based detectors for plant disease recognition tasks. Among these approaches, the YOLO architecture is particularly attractive for agricultural applications because localization and classification are performed within a single-stage detection framework. Nevertheless, several challenges remain. First, performance evaluation in previous studies often emphasizes overall accuracy or mAP values, whereas disease detection datasets frequently exhibit class imbalance and visual similarity between categories, requiring more comprehensive class-level analysis. Second, lightweight YOLO variants designed for computational efficiency may exhibit limited capability in distinguishing subtle lesion patterns from complex leaf textures and background variations. Third, conventional attention mechanisms such as CBAM rely primarily on global average pooling and global max pooling operations, which may not adequately capture regional activation variability associated with non-uniform disease symptoms. These limitations highlight the need for more adaptive feature representation mechanisms capable of preserving discriminative regional information while maintaining computational efficiency.

To address these challenges, this study proposes a YOLOv9s-based detection framework integrated with a Region-Dispersion Channel Spatial Attention (RDCSA) module. YOLOv9s was selected as the baseline architecture because it provides a favorable balance between detection accuracy and computational efficiency for lightweight object detection tasks. The proposed RDCSA module extends conventional channel attention by incorporating regional dispersion statistics, namely mean, standard deviation, and range, as channel descriptors. The underlying motivation is that non-uniform disease symptoms are not solely characterized by average activation intensity but also by the variability and contrast of activation patterns across spatial regions. By explicitly modeling regional dispersion information, the proposed module aims to enhance the representation of pathological features while preserving lightweight computational characteristics. Based on placement analysis, RDCSA is integrated at the P5 feature pyramid level to achieve a balance between detection performance and computational efficiency.

The use of the term robust in the title is intentionally limited to the scope of the experiments conducted in this study. Specifically, robustness refers to the ability of the proposed model to maintain consistent detection performance under the evaluation protocol, disease categories, and image conditions represented in the COLD dataset. It does not imply robustness across different datasets, acquisition devices, chili varieties, geographic regions, or field environments. Validation under more diverse operational conditions remains an important direction for future work. The main contributions of this study are summarized as follows:

- A RDCSA module is proposed to incorporate regional dispersion statistics as channel descriptors, enabling improved feature representation for non-uniform chili leaf disease symptoms.

- A data preparation protocol is implemented by performing dataset splitting before augmentation and restricting augmentation to the training subset, thereby reducing the risk of data leakage and supporting a more reliable evaluation process.
- The proposed framework is comprehensively evaluated through placement ablation analysis, component-wise ablation, region partition sensitivity analysis, comparison with modern attention mechanisms, repeated experiments, per-class evaluation, confusion matrix analysis, and feature activation analysis.

The remainder of this paper is organized as follows. Section 2 reviews related work on YOLO-based plant disease detection, lightweight detection architectures, and attention mechanisms. Section 3 presents the proposed YOLOv9s-RDCSA framework, including the network architecture and attention design. Section 4 reports the experimental setup, quantitative results, ablation studies, stability analysis, and comparative evaluations. Finally, Section 5 concludes the paper and outlines directions for future research..

2. Related Work

2.1. YOLO-Based Plant Disease Detection

YOLO-based object detectors have been widely adopted in agricultural computer vision due to their ability to perform object localization and classification within a single-stage detection framework [5]. Compared with two-stage detectors such as Faster R-CNN, YOLO architectures generally offer faster inference and lower computational complexity, making them well suited for real-time agricultural applications. For example, YOLOv11 achieved an inference time of 13.5 ms and an mAP@0.5 of 0.935 in weed detection, outperforming Faster R-CNN, which required 63.8 ms and achieved an mAP@0.5 of 0.821 [6]. These characteristics make YOLO-based models attractive for plant disease detection tasks that require rapid identification of infected tissues and lesion localization under field conditions [7].

Despite their success, the performance of YOLO-based disease detection systems remains influenced by factors such as dataset size, image quality, class imbalance, symptom similarity, background complexity, and object scale variation [8], [9]. Furthermore, many studies primarily report aggregate metrics such as accuracy or mAP, which may not fully reflect model reliability when disease categories exhibit visual overlap or uneven class distributions [10]. In such scenarios, additional evaluations using per-class precision, recall, F1-score, and confusion matrix analysis are necessary to provide a more comprehensive assessment of detection performance [11], [12]. These observations suggest that improvements in agricultural disease detection should focus not only on overall accuracy but also on the robustness of feature discrimination across different disease categories.

2.2. Lightweight Object Detection for Agricultural Applications

Lightweight object detection has become increasingly important for agricultural applications because deployment often targets edge devices, mobile platforms, robots, and drones with limited computational resources [13]. To address these constraints, various optimization strategies have been explored, including model pruning, quantization, knowledge distillation, and lightweight backbone design. Several studies have demonstrated that such approaches can substantially reduce computational requirements while maintaining competitive detection performance. For instance, MobileNetv3-YOLOv4 achieved high detection accuracy with significantly fewer parameters, while lightweight detection frameworks have also demonstrated promising results in crop monitoring and fruit maturity assessment tasks [14], [15]. Similar findings have been reported across a variety of agricultural detection scenarios, highlighting the practical feasibility of lightweight deep learning models for field deployment [5].

However, improving computational efficiency often introduces a trade-off between model compactness and feature representation capability. Reducing model complexity may lead to the loss of fine-grained information required to distinguish subtle disease symptoms, particularly when lesions are small, irregularly distributed, or visually similar to surrounding leaf textures [14]. In addition, cross-dataset generalization remains challenging because model performance often degrades under varying crop types, environmental conditions, and imaging settings [16]. Agricultural environments also introduce additional challenges such as foliage occlusion, illumination variation, low-light conditions, and dense object distributions [15], [17]–[19]. Consequently, lightweight detection models require mechanisms that can selectively

enhance informative features while preserving computational efficiency, motivating the integration of lightweight attention modules into modern detection frameworks.

2.3. Attention Mechanisms in Plant Disease Detection

Attention mechanisms have been extensively incorporated into deep learning models to improve feature selectivity and enhance the representation of relevant visual patterns [20]. Representative attention modules include Squeeze-and-Excitation (SE), Efficient Channel Attention (ECA), Coordinate Attention, SimAM, EMA Attention, and the Convolutional Block Attention Module (CBAM), all of which have demonstrated effectiveness across a wide range of computer vision applications, including plant disease recognition [21]–[24]. Among these approaches, CBAM is one of the most widely adopted modules because it sequentially combines channel attention and spatial attention. Channel attention emphasizes informative feature channels, whereas spatial attention highlights important regions within feature maps.

The integration of attention mechanisms has been shown to improve detection performance in various agricultural applications. Previous studies reported mAP improvements ranging from 7.7% to 11.8% through the incorporation of CBAM into YOLO-based architectures, while hybrid attention frameworks have achieved disease classification accuracies exceeding 94% [25], [26]. These findings demonstrate the effectiveness of attention mechanisms in enhancing feature discrimination and suppressing irrelevant information.

Nevertheless, conventional channel attention mechanisms rely primarily on global average pooling and global max pooling to summarize channel responses. Although effective, these descriptors may not fully capture regional activation variability associated with non-uniform disease symptoms. Chili leaf diseases frequently exhibit scattered lesions, localized discoloration, irregular boundaries, and heterogeneous texture distributions. Such characteristics suggest that informative cues may be contained not only in global activation magnitudes but also in the variation and contrast of activations across different spatial regions. Therefore, incorporating regional information into channel attention may provide additional discriminative cues for disease recognition.

The need for more adaptive attention mechanisms is further reinforced by practical agricultural challenges, including complex backgrounds, symptom similarity across disease categories, and the requirement for simultaneous multi-disease detection [21]. Existing studies also highlight limitations related to dataset availability, geographic diversity, imaging quality, and the differentiation between healthy and infected plants [27], [28]. Furthermore, insufficient feature representation capacity may restrict disease identification performance even when attention mechanisms are incorporated [29]. These observations indicate that attention mechanisms capable of preserving regional symptom variability while maintaining computational efficiency remain an important research direction for plant disease detection.

2.4. Research Gap and Position of This Study

Previous studies have demonstrated the effectiveness of YOLO-based detectors, lightweight architectures, and attention mechanisms for plant disease recognition. Nevertheless, several limitations remain. First, many studies emphasize aggregate performance metrics while providing limited class-level analysis for datasets characterized by class imbalance or visually similar disease categories. Second, the contribution of attention mechanisms is frequently assessed through baseline comparisons alone, with limited investigation of the individual components responsible for performance improvements. Third, conventional channel attention descriptors based on global pooling operations may not adequately capture regional dispersion patterns associated with non-uniform disease symptoms.

To address these limitations, this study introduces the RDCSA module, which incorporates regional dispersion statistics in the form of mean, standard deviation, and range as channel descriptors. Unlike standard CBAM, which summarizes channel responses through global pooling operations, RDCSA partitions feature maps into regional grids and explicitly models inter-region activation variability. This design enables the attention mechanism to capture both central activation tendencies and regional contrasts that may characterize heterogeneous disease manifestations while maintaining lightweight computational characteristics.

Recent studies further emphasize the importance of balancing computational efficiency with effective feature representation. EDANet combines depthwise separable convolutions and hybrid attention mechanisms to achieve efficient and accurate tomato disease recognition

[30], while studies based on InceptionV3 and InceptionResNetV2 demonstrate the benefits of data augmentation strategies for improving rice leaf disease classification performance [31]. Collectively, these studies highlight the need for detection frameworks that preserve discriminative disease features while maintaining practical computational requirements for real-world deployment.

Accordingly, the proposed YOLOv9s-RDCSA framework aims to bridge the gap between lightweight detection efficiency and the ability to represent non-uniform disease symptoms through regional dispersion-aware attention. Beyond introducing a new attention descriptor, the framework is evaluated through placement ablation analysis, component-wise ablation, sensitivity analysis, repeated experiments, per-class evaluation, and feature activation analysis to provide a comprehensive assessment of its contribution to chili leaf disease detection. Based on these observations, the architecture and implementation details of the proposed YOLOv9s-RDCSA framework are presented in the following section.

3. Proposed Method

3.1 Dataset

3.1.1 Data Acquisition

This study employed the Chili Leaf Disease Dataset (COLD), which contains 532 images collected directly from real agricultural environments [1]. The dataset comprises five leaf-condition categories, namely Healthy, Cercospora, Mites and Thrips, Nutritional Deficiency, and Powdery Mildew, representing diverse manifestations of chili leaf health and disease symptoms.

Image acquisition was conducted under varying field conditions, including both sunny and rainy weather, allowing the dataset to capture environmental variability commonly encountered in agricultural practice. Images were acquired using a Canon EOS Mark II APS-C camera with a resolution of 72 dpi, a 24-bit color depth, and an integrated autofocus system. These specifications provide sufficient visual quality to preserve fine-grained leaf textures and disease symptom characteristics, supporting detailed analysis of lesion morphology and symptom distribution. Representative examples from the five disease categories are presented in Figure 1.

Figure 1. Representative samples from the COLD dataset. (**a**) Healthy; (**b**) Cercospora; (**c**) Mites and Thrips; (**d**) Nutritional Deficiency; and (**e**) Powdery Mildew.

3.1.2 Data Splitting

Prior to model development, all images underwent quality verification to assess image sharpness, illumination consistency, and annotation validity. Disease labels were manually verified by plant pathology experts to ensure dataset reliability. In addition, class distribution analysis was performed to identify potential imbalance among disease categories. Object annotations were subsequently converted into YOLO format to support training and evaluation within the proposed detection framework. To minimize the risk of data leakage, dataset partitioning was performed before any augmentation procedure was applied. The original dataset was divided into training, validation, and testing subsets using a ratio of 70%, 15%, and 15%, respectively. The splitting procedure preserved the class distribution across all subsets to maintain representative sampling for each disease category.

As summarized in Table 1, the partitioning process was conducted exclusively on the original images prior to augmentation or oversampling operations [32]. This protocol ensures complete isolation between the training and evaluation stages, thereby improving the

reliability of the reported performance metrics and reducing the possibility of information leakage between subsets.

Table 1. Dataset splitting configuration.

Disease Class	Total (100%)	Train (70%)	Validation (15%)	Test (15%)
Healthy	68	48	10	10
Cercospora	152	106	23	23
Mites and Thrips	107	75	16	16
Nutritional Deficiency	101	71	15	15
Powdery Mildew	104	72	16	16
Total	532	372	80	80

3.1.3. Dataset Augmentation

Data augmentation was applied exclusively to the training subset to increase visual diversity and improve the model's ability to generalize to unseen samples. Following the split-before-augmentation protocol, the validation and test subsets remained unchanged throughout the experimental process, ensuring an unbiased evaluation procedure [33]. The augmentation strategy consisted of four geometric transformations: 90° clockwise rotation, 90° counterclockwise rotation, horizontal flipping, and vertical flipping. These operations were selected because they preserve disease characteristics while increasing the variability of leaf orientation and spatial presentation. The augmentation configuration for each disease class is summarized in Table 2.

Starting from 372 original training images, the augmentation process generated an additional 828 samples, resulting in a final training set containing 1,200 images. By increasing the diversity of symptom appearance and spatial orientation, the augmentation strategy improves feature coverage during training while maintaining evaluation consistency through the use of untouched validation and test datasets [34].

Table 2. Dataset augmentation results.

Class	Initial Train Data	90° CW Rotation	Horizontal Flip	Vertical Flip	90° CCW Rotation	Total Augmentation
Healthy	48	48	48	48	48	192
Cercospora	106	67	67	0	0	134
Mites and Thrips	75	55	55	55	0	165
Nutritional Deficiency	71	55	55	55	4	169
Powdery Mildew	72	55	55	55	3	168
Total	372	280	280	213	55	828

3.2 YOLOv9s Baseline Configuration

YOLOv9s was selected as the baseline architecture because it provides a favorable balance between detection accuracy and computational efficiency, making it suitable for lightweight agricultural vision applications [35]. Compared with larger model variants, YOLOv9s offers lower computational requirements while maintaining competitive detection performance, which is advantageous for deployment on resource-constrained platforms commonly used in precision agriculture [36], [37]. In this study, all input images were resized to 256 × 256 pixels before being processed by the network. The architecture follows the standard YOLO detection pipeline, consisting of a Backbone for hierarchical feature extraction, a Neck for multi-scale feature fusion, and a Detection Head for object localization and classification.

To enhance feature representation, the proposed RDCSA module was integrated into the Neck stage at the P5 feature level. The P5 layer was selected because it contains high-level semantic information while preserving sufficient spatial context for disease symptom localization. Furthermore, its larger receptive field enables the network to capture broader disease patterns and lesion distributions that are often observed in infected chili leaves.

The integration strategy was intentionally designed to preserve the original YOLOv9s architecture. Neither the Backbone nor the Detection Head was modified; instead, RDCSA was inserted as a lightweight feature enhancement module within the feature fusion stage. This design allows the network to improve its sensitivity to subtle disease characteristics and heterogeneous symptom distributions without introducing substantial computational overhead. To justify the selected placement, additional ablation experiments were conducted by integrating RDCSA at the P3, P4, and P5 feature levels, and the comparative results are presented in the experimental section.

3.3. Overview of the Proposed Network

The proposed framework extends the YOLOv9s baseline architecture by integrating the RDCSA module within the feature fusion stage. The overall architecture follows the standard YOLO pipeline consisting of three sequential components: Backbone, Neck, and Head. This design provides a unified workflow from image input to disease localization and classification while explicitly illustrating the integration point of the proposed attention mechanism. The complete architecture is presented in Figure 2.

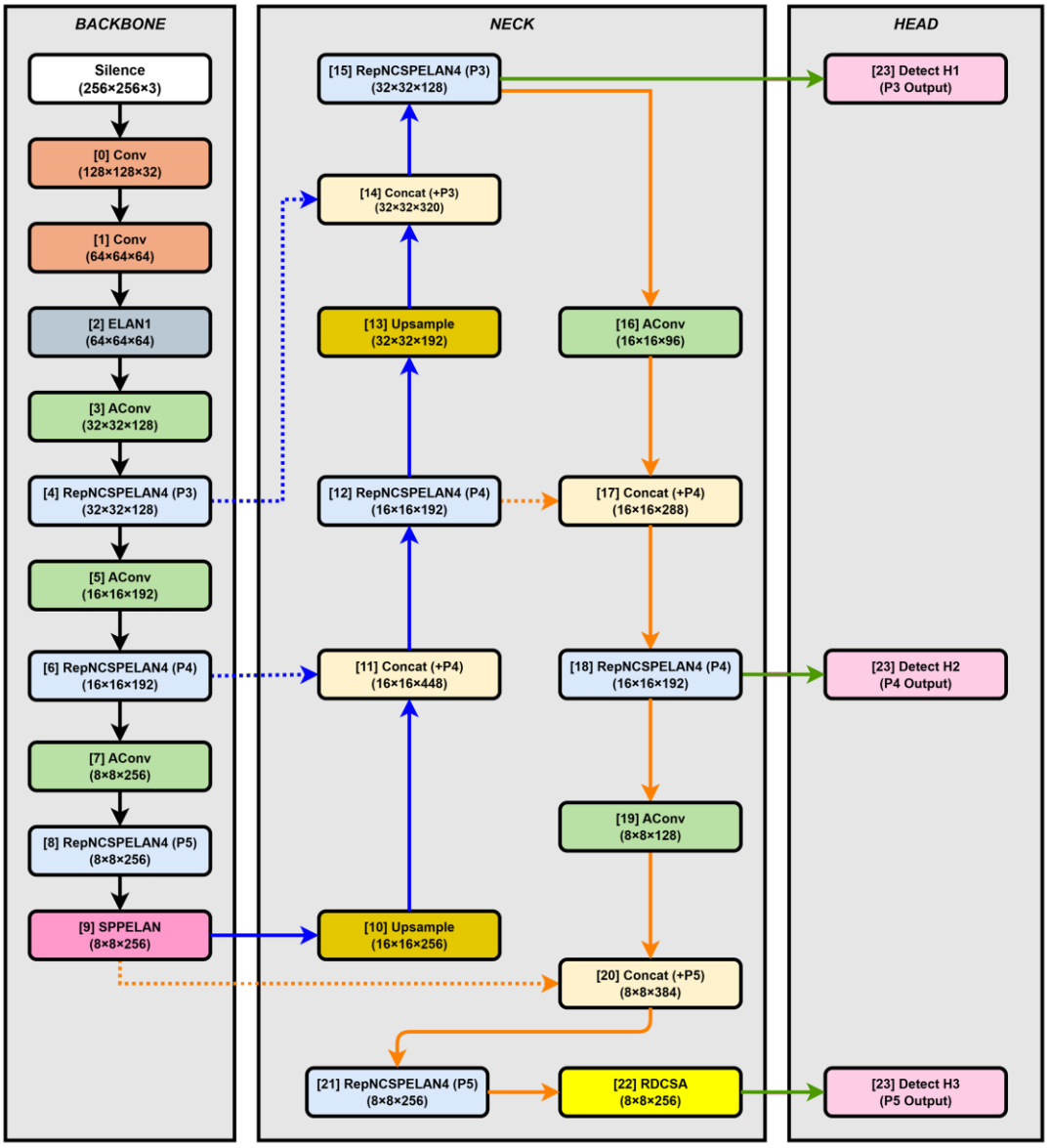

Figure 2. Overall architecture of the proposed YOLOv9s-RDCSA framework with RDCSA integrated at the P5 feature level.

Input images are first processed by the Backbone (Blocks [0]–[9]), which extracts hierarchical feature representations through convolutional and RepNCSPELAN modules. During this stage, spatial resolution is progressively reduced while channel depth increases, enabling the network to learn multi-scale contextual information. The extracted features are then forwarded to the Neck (Blocks [10]–[22]), which adopts a Path Aggregation Network (PANet) structure to perform feature fusion across multiple scales through upsampling and cross-level aggregation. Finally, the fused representations are delivered to the Detection Head (Block [23]), where predictions are generated at the P3, P4, and P5 scales for disease localization and classification.

The principal architectural modification is the integration of RDCSA at the P5 feature level within the Neck stage, immediately before the detection process. This placement was selected because P5 features contain rich semantic information while retaining sufficient spatial context for disease localization. Moreover, the larger receptive field available at this level facilitates the representation of disease patterns that extend across broader leaf regions. To assess the influence of attention placement, additional ablation experiments were conducted by integrating attention modules at the P3, P4, and P5 feature levels.

Unlike approaches that modify multiple network components, the proposed framework preserves the original Backbone and Detection Head of YOLOv9s. RDCSA is introduced solely as a lightweight feature enhancement module within the feature fusion stage, allowing the network to improve its sensitivity to heterogeneous symptom patterns while maintaining computational efficiency. This design enables a direct evaluation of the contribution of regional dispersion-aware attention without introducing substantial architectural changes.

3.4. Convolutional Block Attention Module

The Convolutional Block Attention Module (CBAM) is adopted as the reference attention mechanism because it sequentially combines channel attention and spatial attention within a lightweight architecture [38]. In the channel attention stage, CBAM summarizes feature responses using global average pooling and global max pooling operations. The resulting descriptors are processed through a shared multi-layer perceptron (MLP) to generate channel attention weights according to Equation (1).

$$M_c(F) = \sigma\left(\mathrm{MLP}\big(\mathrm{AvgPool}(F)\big) + \mathrm{MLP}\big(\mathrm{MaxPool}(F)\big)\right) \tag{1}$$

The generated attention map is then multiplied element-wise with the input feature map to produce a channel-refined representation, as shown in Equation (2).

$$F' = F \otimes M_c(F) \tag{2}$$

Subsequently, the spatial attention stage applies average pooling and max pooling along the channel dimension. The pooled features are concatenated and processed using a 7×7 convolution followed by a sigmoid activation function to generate a spatial attention map, as formulated in Equation (3).

$$M_s(F') = \sigma\big(\mathrm{conv}^{7\times7}([\mathrm{AvgPool}_{chan}(F'); \mathrm{MaxPool}_{chan}(F')])\big) \tag{3}$$

The final refined feature representation is obtained through element-wise multiplication between the spatial attention map and the channel-refined feature map, as expressed in Equation (4).

$$F'' = F' \otimes M_s(F') \tag{4}$$

In this study, CBAM serves both as a benchmark attention mechanism and as the structural foundation for the proposed RDCSA module. The primary distinction lies in the channel attention stage, where conventional global pooling descriptors are replaced with regional dispersion descriptors, while the spatial attention mechanism remains unchanged. This design enables a controlled evaluation of the contribution of regional dispersion information to feature discrimination. The CBAM architecture is illustrated in Figure 3.

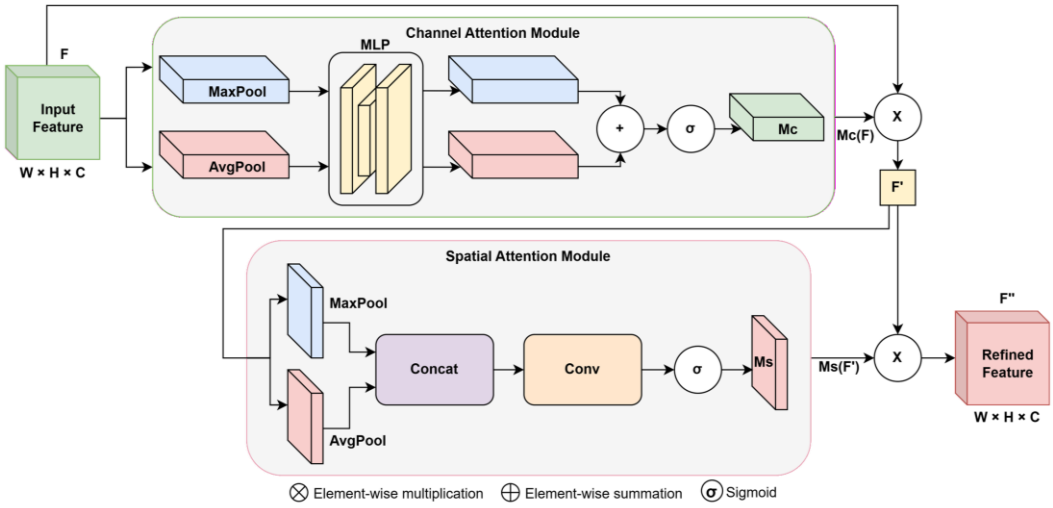

Figure 3. Architecture of the Convolutional Block Attention Module (CBAM).

3.5. Region-Dispersion Channel Spatial Attention (RDCSA)

The proposed RDCSA module extends the conventional CBAM framework by replacing the standard global pooling-based channel descriptors with regional dispersion statistics. The principal objective of this modification is to capture spatial activation variability associated with heterogeneous disease symptoms while preserving the lightweight characteristics of the original attention mechanism. Unlike standard CBAM, which summarizes channel responses using global average pooling and global max pooling, RDCSA explicitly models the distribution of activations across spatial regions before channel attention is generated.

As illustrated in Figure 4, the RDCSA module consists of two sequential stages: Region-Dispersion Channel Attention (RDCA) and Spatial Attention (SA). The methodological contribution of this study is concentrated in the RDCA stage, whereas the spatial attention component remains identical to the CBAM formulation presented in Equations (3)–(4).

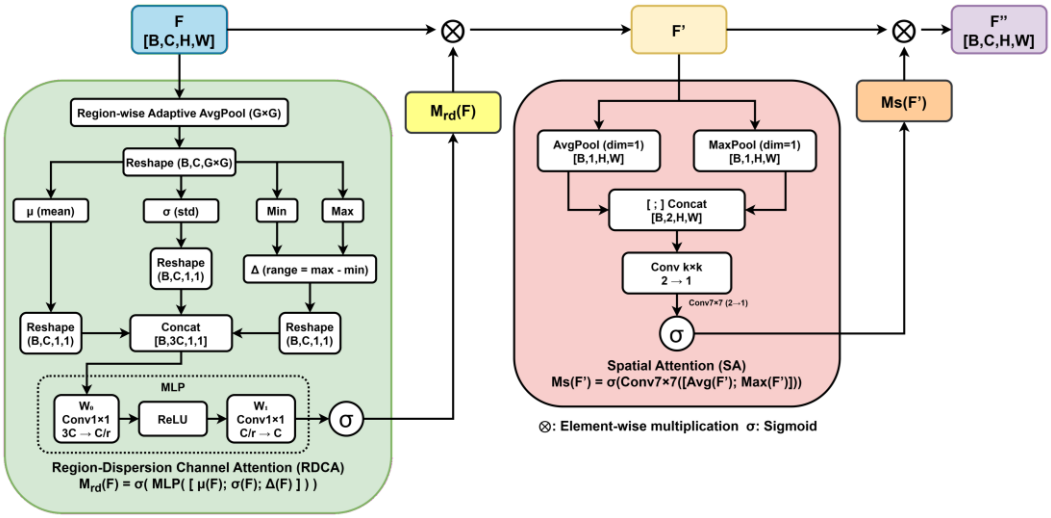

Figure 4. Architecture of the proposed RDCSA module.

Given an input feature map $F \in \mathbb{R}^{B\times C\times H\times W}$, the RDCA stage first partitions the feature map into a $G \times G$ regional grid using adaptive average pooling. The resulting regional representation is reshaped into $B \times C \times G^2$, enabling the extraction of statistical descriptors across spatial regions. Three complementary statistics are then computed for each channel: mean (μ), standard deviation (σ), and range (Δ). These descriptors respectively represent the central tendency, regional variability, and maximum activation contrast of the feature responses.

These descriptors are concatenated to construct a regional dispersion representation D, as defined in Equation (5).

$$D = \text{Concat}([\mu, \sigma, \Delta], dim = 1) \tag{5}$$

The resulting descriptor is processed through a lightweight bottleneck transformation consisting of two 1×1 convolutional layers with channel reduction and expansion operations. A ReLU activation function is applied between the two layers, followed by a sigmoid function to generate the regional dispersion channel attention map M_{rd}, as expressed in Equation (6).

$$M_{rd}(F) = \sigma\left(\text{Conv}_{1\times1}^{(up)}\left(\text{ReLU}\left(\text{Conv}_{1\times1}^{(down)}(D)\right)\right)\right) \tag{6}$$

The generated attention map is then applied to the input feature map through element-wise multiplication, producing a channel-refined representation F':

$$F' = F \otimes M_{rd}(F) \tag{7}$$

After channel refinement, the resulting feature representation is processed using the same spatial attention mechanism employed by CBAM, as described in Equations (3)–(4). This design choice enables a controlled evaluation of the contribution of regional dispersion descriptors while preserving the spatial attention formulation and overall computational structure. The complete processing pipeline can therefore be summarized as follows:

1. Regional partition of the input feature map.
2. Extraction of mean, standard deviation, and range descriptors.
3. Construction of the regional dispersion representation.
4. Bottleneck-based channel attention generation.
5. Channel feature refinement using regional dispersion attention.
6. Spatial attention refinement using the standard CBAM mechanism.
7. Generation of the final attention-enhanced feature representation.

By explicitly modeling inter-region activation variability, RDCSA enables the network to capture non-uniform disease symptoms that may not be adequately represented by conventional global descriptors. This characteristic is particularly relevant for chili leaf disease detection, where lesions frequently exhibit scattered distributions, localized discoloration, and irregular boundaries. Consequently, RDCSA enhances feature discrimination while maintaining the lightweight computational profile required for efficient deployment within the YOLOv9s framework..

4. Results and Discussion

4.1. Experimental Setup

All experiments were conducted in a cloud-computing environment equipped with an x86_64 CPU architecture (six physical cores, twelve logical cores, 2200.18 MHz), an NVIDIA L4 GPU with 23.66 GB of dedicated memory and Compute Capability 8.9 support, and 56.86 GB of system memory. The software environment consisted of CUDA Driver 580.82.07, CUDA Toolkit 12.8, Python 3.12.12, PyTorch 2.10.0, and Ultralytics YOLO 8.3.219. The complete experimental workflow was executed using Google Colab.

Model training was initialized from scratch using a deterministic configuration with a fixed random seed of 21 to improve reproducibility. The network was trained for 100 epochs using the AdamW optimizer with a batch size of 32, an initial learning rate of 0.001111, a momentum coefficient of 0.9, and an L2 weight decay of 0.0005 applied to the weight parameters. The overall objective function combines localization, classification, and distribution focal loss components as follows:

$$L_{total} = \lambda_{box}L_{box} + \lambda_{cls}L_{cls} + \lambda_{dfl}L_{dfl} \tag{8}$$

where $\lambda_{box} = 7.5$, $\lambda_{cls} = 0.5$, and $\lambda_{dfl} = 1.5$. This weighting scheme prioritizes localization accuracy while maintaining reliable classification and bounding-box distribution modeling.

Training stability and computational efficiency were further enhanced through the use of Automatic Mixed Precision (AMP), cosine annealing learning-rate scheduling, gradient scaling, and DFL layer freezing. Input images were resized to 256 × 256 pixels and processed using four parallel workers with disk caching enabled. The augmentation pipeline included Mosaic augmentation and HSV-V exposure variation (±0.4), while all additional geometric transformations and image-mixing strategies were disabled during training.

A fixed training duration of 100 epochs was adopted to ensure a fair and consistent comparison across all evaluated configurations. This setting is consistent with recent benchmarking studies involving YOLOv8, YOLOv9, YOLOv10, and YOLOv11, which commonly employ 100 training epochs under standardized experimental settings [39]. A controlled comparison requires identical training conditions, including dataset partitioning, optimization strategy, learning-rate scheduling, augmentation configuration, and training duration [40]. Consequently, early stopping was intentionally not applied, as the objective of this study was to evaluate the relative contribution of different attention mechanisms under a uniform training protocol rather than to optimize an individual stopping criterion for each model.

4.2. Evaluation Metrics

Model performance was assessed using five primary evaluation metrics: Precision, Recall, F1-score, mean Average Precision (mAP), and GFLOPs (Giga Floating-Point Operations). Precision and Recall quantify the proportion of correct detections relative to predicted and actual instances, respectively, whereas the F1-score provides a balanced measure of both metrics. Detection accuracy was further evaluated using mAP, which summarizes performance across confidence thresholds and target classes.

In addition to predictive performance, computational efficiency was evaluated using inference latency and Frames Per Second (FPS). Latency includes preprocessing, inference, and post-processing times, while FPS measures the number of images processed per second. To provide a normalized assessment of efficiency, two additional indicators were calculated: EfficiencyFLOP and EfficiencyParam, defined as the ratio between mAP@50–95 and GFLOPs, and between mAP@50–95 and the number of model parameters (in millions), respectively. These indicators quantify the relative accuracy achieved per unit of computational resource.

4.3. Main Performance Evaluation

This section evaluates the proposed framework through comparison with the YOLOv9s baseline and attention placement analysis at different feature pyramid levels. All experiments followed the split-before-augmentation protocol described in Section 3, where only the training subset was augmented. For a controlled comparison, a single attention module was independently inserted at either the P3, P4, or P5 feature level while all other architectural and training settings remained unchanged. The quantitative results are summarized in Table 3.

Table 3. Main performance evaluation and attention placement analysis.

Model	Precision	Recall	F1-Score	mAP@50	mAP@50–95	Parameters	GFLOPs	Latency (ms)	FPS
YOLOv9s	0.746	0.849	0.794	0.830	0.714	7,169,023	26.7	7.4	135
YOLOv9s + CBAM (P3)	0.876	0.845	0.860	0.885	0.776	7,171,305	26.7	7.8	128
YOLOv9s + CBAM (P4)	0.868	0.796	0.830	0.827	0.730	7,173,933	26.7	7.9	127
YOLOv9s + CBAM (P5)	0.842	0.803	0.822	0.839	0.737	7,177,585	26.8	7.9	127
YOLOv9s + RDCSA (P3)	0.830	0.852	0.841	0.873	0.741	7,173,353	26.7	7.7	130
YOLOv9s + RDCSA (P4)	0.788	0.801	0.794	0.821	0.688	7,178,541	26.7	7.8	128
YOLOv9s + RDCSA (P5)	0.858	0.861	0.859	0.894	0.773	7,185,777	26.8	7.6	132

The results demonstrate that the effectiveness of attention integration is strongly influenced by the feature level at which the module is applied. Among all evaluated configurations, YOLOv9s + RDCSA (P5) achieved the highest mAP@50 and mAP@50–95 while introducing only a marginal increase in parameter count and computational cost relative to the baseline model. These results indicate that the proposed regional dispersion descriptors can improve feature discrimination without substantially affecting model efficiency.

The placement analysis also reveals distinct characteristics across the feature pyramid hierarchy. The P3 layer preserves high spatial resolution and is therefore effective for capturing fine-grained local details. However, its limited semantic abstraction may restrict its ability to represent broader disease patterns. Conversely, the P5 layer contains richer semantic information and a larger receptive field, enabling the network to better model non-uniform symptom distributions, dispersed lesions, and contextual variations among disease categories. The intermediate P4 layer provides a compromise between spatial detail and semantic abstraction; however, the results suggest that this balance was less effective for the characteristics of the evaluated dataset.

This observation is consistent with the design objective of RDCSA, which relies on regional activation variability to guide channel attention. Such information becomes more informative when extracted from high-level semantic representations that already encode disease-related contextual patterns. Consequently, integrating RDCSA at the P5 feature level provides the most favorable trade-off between representational capability and computational overhead. Based on these findings, YOLOv9s + RDCSA (P5) was selected as the reference configuration for the subsequent analyses. The observed improvements should be interpreted as incremental gains obtained under the evaluated dataset and experimental conditions rather than as evidence of universal superiority across different datasets or deployment environments.

4.4. Component-Wise Ablation of RDCSA

To investigate the contribution of individual RDCSA components, a component-wise ablation study was conducted by selectively enabling the mean, standard deviation, and range descriptors within the RDCA module, with and without spatial attention refinement. All configurations were evaluated using the same training and evaluation protocol to ensure a fair comparison. The results are presented in Table 4.

Table 4. Component-wise ablation of RDCSA.

Model Variant	Mean	Std	Range	Spatial Attention	Precision	Recall	F1-Score	mAP@50	mAP@50–95
YOLOv9s	–	–	–	–	0.746	0.849	0.794	0.830	0.714
RDCA-Mean	✓	–	–	–	0.851	0.781	0.814	0.856	0.784
RDCA-Std	–	✓	–	–	0.849	0.831	0.840	0.883	0.774
RDCA-Range	–	–	✓	–	0.900	0.858	0.878	0.871	0.790
RDCA Mean+Std	✓	✓	–	–	0.817	0.744	0.779	0.819	0.724
RDCA Mean+Range	✓	–	✓	–	0.797	0.791	0.794	0.852	0.754
RDCA Std+Range	–	✓	✓	–	0.905	0.761	0.827	0.862	0.766
RDCA Full	✓	✓	✓	–	0.794	0.844	0.818	0.847	0.767
RDCSA Full	✓	✓	✓	✓	0.858	0.861	0.859	0.894	0.773

The ablation results indicate that each regional dispersion descriptor contributes differently to feature discrimination. The range descriptor produced the strongest individual response, suggesting that activation contrast is particularly informative for distinguishing lesion boundaries and visually distinctive disease regions. The standard deviation descriptor also contributed positively by capturing inter-region variability associated with heterogeneous symptom distributions. In contrast, the mean descriptor alone provided limited discrimination capability because it primarily reflects average activation intensity.

Interestingly, combining multiple descriptors did not consistently improve performance when spatial attention was absent. This observation suggests that descriptor fusion alone may introduce redundant responses that require additional spatial filtering. The full RDCSA configuration achieved the most balanced overall performance, indicating that regional dispersion-based channel selection and spatial attention refinement play complementary roles. Together, these components enable the network to emphasize informative disease patterns while suppressing irrelevant background responses.

4.5. Sensitivity Analysis of the Region Partition Parameter

A sensitivity analysis was conducted to evaluate the influence of the region partition parameter G on RDCSA performance. The parameter G determines the granularity of regional partitioning used during dispersion statistic extraction and therefore directly affects how regional information is represented. Four partition settings ($G = 2,4,7,$ and 14) were evaluated while maintaining identical architectural and training configurations. The results are summarized in Table 5.

Table 5. Sensitivity analysis of the region partition parameter G.

G Value	Precision	Recall	F1-Score	mAP@50	mAP@50–95	Parameters	GFLOPs
2	0.858	0.861	0.859	0.894	0.773	7,185,777	26.8
4	0.837	0.888	0.862	0.893	0.768	7,185,777	26.8
7	0.877	0.832	0.854	0.887	0.786	7,185,777	26.8
14	0.906	0.803	0.851	0.902	0.800	7,185,777	26.8

The results demonstrate that the partition parameter influences the trade-off between feature selectivity and symptom coverage. Smaller partition values generate broader regional representations that preserve sensitivity to diverse symptom manifestations, whereas larger values emphasize finer local structures and increase selectivity toward highly distinctive patterns. Although all configurations maintain identical parameter counts and computational costs, their detection behavior differs substantially. Larger partition values generally improve precision and mAP but reduce recall, indicating a tendency toward more selective feature activation. Considering the balance between precision, recall, and overall detection performance, $G = 2$ was selected as the default configuration. This setting provides sufficient regional context while retaining sensitivity to subtle symptom variations commonly observed in chili leaf diseases.

4.6. Comparison with Modern Attention Modules

To further assess the effectiveness of RDCSA, the proposed module was compared with several widely used attention mechanisms, including Squeeze-and-Excitation (SE), Coordinate Attention (CoordAtt), SimAM, and CBAM. Each attention module was integrated into the YOLOv9s baseline using an equivalent configuration and evaluated under the same experimental protocol. The comparative results are presented in Table 6.

Table 6. Comparison with modern attention modules.

Model	Attention Module	Precision	Recall	F1-Score	mAP@50	mAP@50–95	Parameters	GFLOPs
YOLOv9s	None	0.746	0.849	0.794	0.830	0.714	7,169,023	26.7
YOLOv9s + SE	SE	0.756	0.870	0.809	0.846	0.741	7,177,215	26.7
YOLOv9s + CoordAtt	CoordAtt	0.865	0.802	0.832	0.843	0.721	7,175,183	26.7
YOLOv9s + SimAM	SimAM	0.846	0.831	0.838	0.848	0.736	7,169,023	26.7
YOLOv9s + CBAM	CBAM	0.842	0.803	0.822	0.839	0.737	7,177,585	26.8
YOLOv9s + RDCSA	RDCSA	0.858	0.861	0.859	0.894	0.773	7,185,777	26.8

The comparison highlights that different attention mechanisms emphasize different aspects of feature learning. SE improved detection coverage through higher recall, whereas CoordAtt increased selectivity at the expense of recall. SimAM and CBAM provided moderate improvements while maintaining lightweight computational characteristics. Among the evaluated methods, RDCSA achieved the most balanced overall performance. This improvement appears to stem from its ability to model inter-region activation variability through regional dispersion descriptors, rather than solely relying on global channel statistics. The results therefore suggest that explicitly incorporating regional activation patterns can provide additional discriminative information for detecting heterogeneous disease symptoms. Nevertheless, these findings should be interpreted within the scope of the evaluated COLD dataset and do not imply universal superiority across all datasets or deployment conditions.

4.7. Stability Analysis

To assess the consistency of the proposed approach, repeated training and evaluation experiments were conducted under identical settings. Results are reported as mean ± standard deviation to quantify performance variability arising from stochastic training factors such as parameter initialization and data ordering. The results are presented in Table 7.

Table 7. Stability analysis using repeated experiments.

Model	Precision	Recall	F1-Score	mAP@50	mAP@50–95
YOLOv9s	0.815 ± 0.050	0.841 ± 0.058	0.825 ± 0.023	0.857 ± 0.030	0.742 ± 0.031
YOLOv9s + CBAM	0.817 ± 0.047	0.819 ± 0.055	0.816 ± 0.016	0.865 ± 0.031	0.760 ± 0.030
YOLOv9s + RDCSA	0.860 ± 0.046	0.828 ± 0.027	0.843 ± 0.020	0.879 ± 0.020	0.762 ± 0.016

The repeated experiments reveal different variability patterns across the evaluated configurations. The baseline YOLOv9s model exhibited the largest fluctuations in precision and recall, indicating greater sensitivity to stochastic training effects. CBAM produced the lowest standard deviation in F1-score, reflecting stable balancing of precision and recall across runs. The proposed RDCSA configuration achieved the highest mean precision, F1-score, and mAP values while maintaining relatively low variability. These results suggest that the observed improvements are reproducible and are not solely attributable to random training fluctuations. The findings further indicate that regional dispersion descriptors contribute to more stable feature discrimination when processing heterogeneous disease symptoms. Nevertheless, validation on larger datasets and more diverse acquisition conditions remains necessary to determine whether similar stability characteristics can be maintained under broader operational scenarios.

4.8. Per-Class Performance and Confusion Matrix Analysis

To assess class-level reliability, the proposed YOLOv9s + RDCSA model was evaluated using per-class precision, recall, F1-score, and mAP@50 metrics. Such analysis is particularly important because aggregate metrics may conceal weaknesses in specific categories, especially when disease symptoms exhibit different levels of visual ambiguity. The quantitative results are summarized in Table 8.

Table 8. Per-class performance of YOLOv9s + RDCSA.

Class	Precision	Recall	F1-Score	mAP@50
Healthy	0.802	0.769	0.785	0.773
Cercospora	0.794	0.917	0.851	0.919
Mites and Thrips	0.839	0.869	0.854	0.934
Nutritional Deficiency	0.919	0.750	0.826	0.854
Powdery Mildew	0.934	1.000	0.966	0.991

The results indicate that detection performance varies according to symptom distinctiveness. Powdery Mildew achieved the highest performance across all metrics, suggesting that its characteristic white powder-like lesions provide strong visual contrast against the leaf surface. Cercospora and Mites and Thrips also exhibited reliable detection performance due to the presence of recognizable lesion structures and texture variations. In contrast, Healthy and Nutritional Deficiency remained more challenging categories. These classes often contain subtle visual cues, low-contrast symptom patterns, or physiological variations that resemble normal leaf textures. Consequently, disease-related features are less distinguishable from surrounding background information, increasing the likelihood of misclassification. The corresponding confusion matrix is presented in Figure 5.

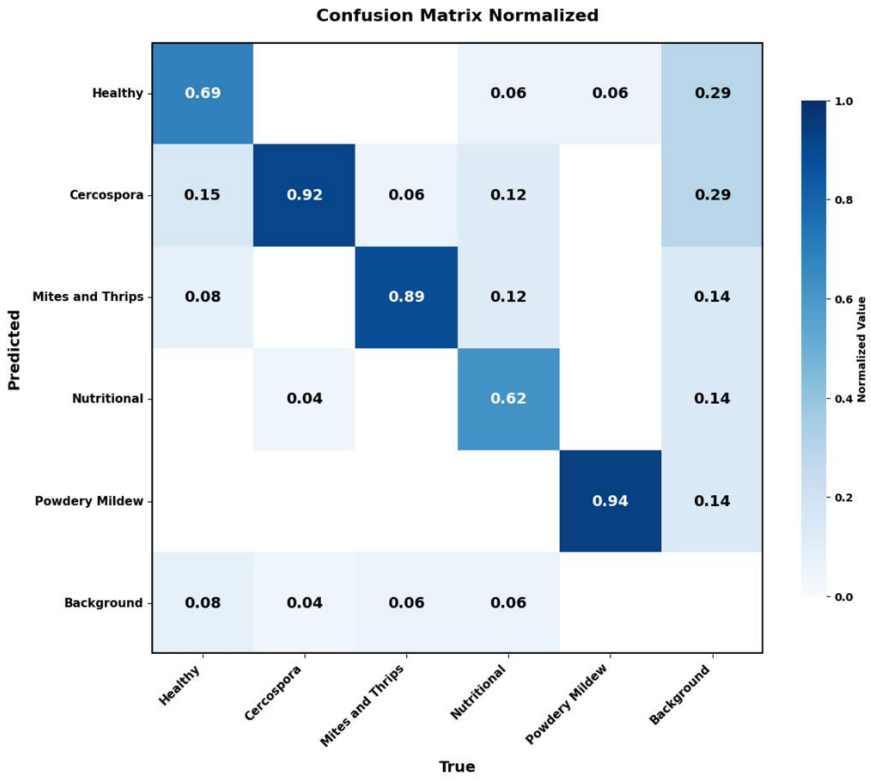

Figure 5. Normalized confusion matrix of the proposed YOLOv9s + RDCSA model.

Most classification errors are concentrated in the Healthy and Nutritional Deficiency categories, where a considerable proportion of samples are confused with the background class. This observation suggests that early-stage symptoms and low-contrast visual manifestations remain difficult to separate from natural leaf appearance and environmental noise. A limited degree of overlap between Nutritional Deficiency and Cercospora is also observed, indicating partial similarity in color and texture characteristics at certain symptom stages. Overall, the class-level analysis confirms that YOLOv9s + RDCSA performs particularly well on disease categories with distinctive visual manifestations while remaining more susceptible to ambiguity in subtle symptom classes. These findings highlight the importance of evaluating disease detection systems beyond aggregate metrics alone.

4.9. Attention and Feature Map Visualization

To better understand the behavior of the proposed attention mechanism, feature activation analysis was performed at the P5 feature level, where RDCSA was integrated into the final architecture. Feature tensors before and after the attention module were extracted through a forward-hook mechanism implemented in PyTorch and the Ultralytics YOLO framework. The analysis employed activation-based indicators including Before Mean, After Mean, Diff Mean, Diff Std, Suppression (%), Enhancement (%), and Reduction (%). These indicators were used exclusively as interpretability measures and were not treated as direct evidence of improved detection performance without support from quantitative evaluation and ablation experiments. The quantitative results are summarized in Table 9.

The activation statistics reveal distinct modulation behaviors across the evaluated configurations. The baseline YOLOv9s architecture generally amplified feature responses, reflecting the absence of an explicit attention-guided filtering mechanism. In contrast, CBAM introduced systematic suppression of less informative activations through its sequential channel and spatial attention operations. RDCSA produced a more selective activation pattern characterized by stronger suppression and greater variation across disease categories. This behavior is consistent with the design objective of regional dispersion-based attention, which explicitly models inter-region activation variability rather than relying solely on global feature statistics. Such modulation is particularly relevant for chili leaf disease detection, where

symptoms frequently appear as irregular lesions, scattered discoloration, and heterogeneous texture changes. The corresponding feature map visualization is presented in Figure 6.

Table 9. Comparative analysis of feature activation at the P5 feature level.

Model	Classes	Before Mean	After Mean	Diff Mean	Diff Std	Suppression (%)	Enhancement (%)	Reduction (%)
YOLOv9s (P5)	Healthy	0.519	4.714	4.195	3.209	9.50	90.50	-808.11
	Cercospora	0.458	0.594	0.136	0.584	67.75	32.25	-29.64
	Mites and Thrips	0.452	0.739	0.286	0.749	60.25	39.75	-63.33
	Nutritional Def.	0.501	1.958	1.457	0.988	5.00	95.00	-290.83
	Powdery Mildew	0.512	1.563	1.051	0.863	14.25	85.75	-205.24
	Average	0.489	1.914	1.425	1.279	31.35	68.65	-279.43
YOLOv9s +CBAM (P5)	Healthy	4.054	2.795	-1.259	0.800	100.00	0.00	31.05
	Cercospora	2.443	1.680	-0.763	0.395	100.00	0.00	31.22
	Mites and Thrips	1.376	0.924	-0.453	0.284	100.00	0.00	32.90
	Nutritional Def.	3.021	2.047	-0.974	0.473	100.00	0.00	32.24
	Powdery Mildew	4.080	2.803	-1.277	0.621	100.00	0.00	31.30
	Average	2.995	2.050	-0.945	0.515	100.00	0.00	31.74
YOLOv9s +RDCSA (P5)	Healthy	2.725	0.967	-1.757	1.684	100.00	0.00	64.50
	Cercospora	4.401	1.547	-2.854	2.251	100.00	0.00	64.85
	Mites and Thrips	3.311	1.056	-2.255	1.901	100.00	0.00	68.11
	Nutritional Def.	0.874	0.300	-0.574	0.472	100.00	0.00	65.63
	Powdery Mildew	2.739	0.903	-1.836	1.653	100.00	0.00	67.02
	Average	2.810	0.955	-1.855	1.592	100.00	0.00	66.02

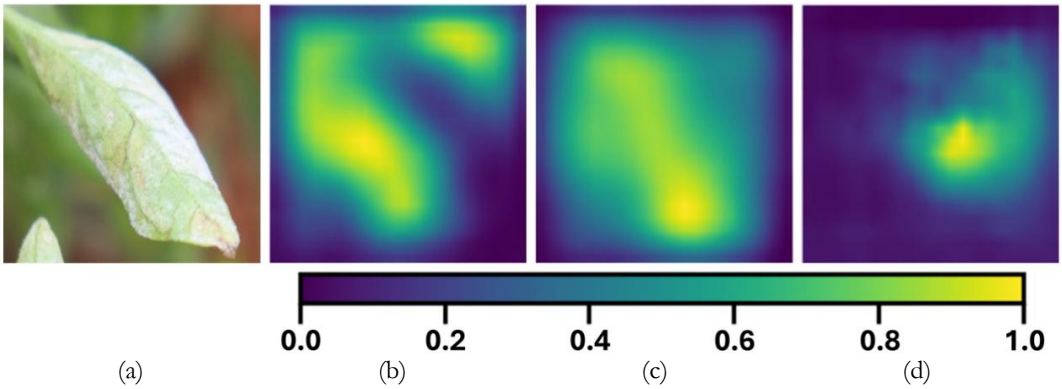

Figure 6. Feature map visualization of YOLOv9s, CBAM, and RDCSA at the P5 feature level (**a**) Powdery Mildew; (**b**) Yolov9s; (**c**) Yolov9s+CBAM; (**d**) Yolov9s+RDCSA.

Visual inspection further supports the quantitative observations. Compared with the baseline and CBAM configurations, RDCSA exhibits a stronger tendency to concentrate activation within symptom-relevant regions while suppressing surrounding background responses. This suggests that regional dispersion descriptors help guide attention toward informative disease patterns and improve feature selectivity. Nevertheless, these visualizations should be interpreted as complementary evidence rather than standalone proof of performance improvement. The observed activation patterns are most meaningful when considered together with the ablation studies, stability analysis, and class-level evaluation presented in the preceding sections.

4.10. Comparative Benchmark with YOLO Variants and Recent COLD Studies

To position the proposed method within the broader object detection literature, a comparative benchmark was conducted using recent YOLO variants, including YOLOv10s and YOLOv11s, alongside the YOLOv9s baseline. All models were trained and evaluated using an identical protocol to ensure a fair comparison. The results are reported in Table 10.

Table 10. Comparative benchmark results.

Model	Precision	Recall	F1-Score	mAP@50	mAP@50–95	Parameters	GFLOPs	Latency (ms)	FPS
YOLOv9s	0.746	0.849	0.794	0.830	0.714	7,169,023	26.7	7.4	135
YOLOv10s	0.735	0.769	0.752	0.820	0.713	7,219,935	21.4	4.5	222
YOLOv11s	0.853	0.775	0.812	0.845	0.744	9,414,735	21.3	7.3	137
YOLOv9s + RDCSA (P5)	**0.858**	**0.861**	**0.859**	**0.894**	**0.773**	7,185,777	26.8	7.6	132

The benchmark results demonstrate different trade-offs between detection accuracy and computational efficiency. YOLOv10s achieved the fastest inference speed but delivered lower detection accuracy, indicating that computational efficiency alone may not be sufficient for distinguishing subtle disease symptoms. YOLOv11s improved detection performance relative to YOLOv10s but required a substantially larger parameter budget. The proposed YOLOv9s + RDCSA configuration achieved the most balanced overall performance, improving detection metrics while introducing only a marginal increase in model complexity relative to the baseline. These findings suggest that regional dispersion-based attention can enhance feature discrimination without requiring substantial architectural expansion.

To further contextualize the results, an additional comparison was performed against recent deep learning models evaluated on the COLD dataset [41]. Although these methods were developed for image classification rather than object detection, they provide useful reference points for assessing overall predictive capability on the same dataset.

Table 11. Comparative performance of deep learning models evaluated on the COLD dataset (%).

Method	Precision	Recall	F1-Score	Accuracy / mAP@50
ViT	81.77	85.40	82.86	84.58
DenseNet121	78.32	81.02	79.02	82.82
InceptionV3	74.30	75.53	76.27	75.53
EfficientNetB3	78.28	82.18	79.67	82.38
EfficientNetB4	79.63	84.27	81.32	83.26
CLAHEEfficientNetB4	79.30	83.39	80.95	83.70
LGEfficientNetB4	80.51	84.69	82.07	84.14
EfficientNetB4CA	81.75	85.36	83.13	85.46
LGENetB4CA	82.57	87.29	84.42	85.90
YOLOv9s + RDCSA (P5)	85.80	86.10	85.90	89.40

As shown in Table 11, YOLOv9s + RDCSA achieved competitive performance relative to recent COLD-based deep learning models. The proposed framework obtained the highest F1-score and maintained strong precision and recall values, suggesting that regional dispersion-based attention contributes positively to feature discrimination in chili leaf disease recognition. However, this comparison should be interpreted with caution because the referenced studies addressed image classification, whereas the proposed framework performs object detection. Consequently, the reported metrics are not directly equivalent, and the comparison is intended primarily to provide contextual positioning rather than a strict head-to-head evaluation.

5. Conclusions

This study proposed a YOLOv9s-based chili leaf disease detection framework enhanced with the RDCSA module. The proposed approach extends conventional channel attention by incorporating regional dispersion statistics, namely mean, standard deviation, and range, to represent inter-region activation variability. By combining regional dispersion-aware channel selection with spatial attention refinement, the framework improves the representation of heterogeneous disease symptoms while preserving the lightweight characteristics of the underlying YOLOv9s architecture.

Comprehensive evaluation demonstrated that the proposed YOLOv9s + RDCSA (P5) configuration provides a favorable balance between detection performance and

computational efficiency. The ablation studies, sensitivity analysis, comparison with modern attention mechanisms, stability evaluation, and per-class assessment collectively indicate that regional dispersion information contributes to more discriminative feature representations, particularly for disease symptoms characterized by irregular lesion distribution, local discoloration, and heterogeneous visual patterns. The results further highlight the importance of class-level and interpretability-oriented analyses for understanding model behavior beyond aggregate detection metrics.

Despite these promising findings, the scope of the present study remains limited to the COLD dataset, five chili leaf condition categories, and static image-based evaluation. Therefore, the reported robustness should be interpreted within the context of the evaluated experimental setting rather than as evidence of universal generalization. Future work should focus on validating the proposed attention mechanism across more diverse datasets, acquisition conditions, and agricultural environments, as well as investigating deployment on edge devices and mobile agricultural monitoring platforms. Further exploration of adaptive attention strategies for low-contrast and visually ambiguous disease symptoms may also contribute to improving reliability under real-world field conditions.

Author Contributions: Conceptualization: M.K.H. and J.N.; Methodology: M.K.H.; Software: M.K.H.; Validation: M.K.H., J.N. and F.E.; Formal analysis: M.K.H.; Investigation: M.K.H.; Resources: J.N. and F.E.; Data curation: M.K.H.; Writing—original draft preparation: M.K.H.; Writing—review and editing: M.K.H., J.N. and F.E.; Visualization: M.K.H.; Supervision: J.N. and F.E.; Project administration: M.K.H.; Funding acquisition: M.K.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data supporting the findings of this study are publicly available in the COLD dataset at https://doi.org/10.1016/j.dib.2024.110524. The dataset comprises 532 chili leaf images and was utilized for model training, validation, and testing without restriction.

Acknowledgments: The authors gratefully acknowledge Universitas Bina Sarana Informatika, Jakarta, Indonesia, for providing the computing resources and technical support that enabled this research. The authors thank colleagues for constructive discussions that refined the experimental design and manuscript. AI tools were employed exclusively for language editing and grammatical correction. All scientific content, experimental design, data analysis, interpretation of results, and conclusions were developed and verified entirely by the authors.

Conflicts of Interest: The authors declare no conflicts of interest. This study was conducted independently, free from any commercial, financial, or personal affiliations that could potentially influence the research design, data interpretation, or publication decisions.

References

[1] M. P. Aishwarya and A. P. Reddy, "Dataset of chilli and onion plant leaf images for classification and detection," *Data Br.*, vol. 54, p. 110524, Jun. 2024, doi: 10.1016/j.dib.2024.110524.

[2] A. Jafar, N. Bibi, R. A. Naqvi, A. Sadeghi-Niaraki, and D. Jeong, "Revolutionizing agriculture with artificial intelligence: plant disease detection methods, applications, and their limitations," *Front. Plant Sci.*, vol. 15, Mar. 2024, doi: 10.3389/fpls.2024.1356260.

[3] T. Nyawose, R. C. Maswanganyi, and P. Khumalo, "A Review on the Detection of Plant Disease Using Machine Learning and Deep Learning Approaches," *J. Imaging*, vol. 11, no. 10, p. 326, Sep. 2025, doi: 10.3390/jimaging11100326.

[4] A. Upadhyay *et al.*, "Deep learning and computer vision in plant disease detection: a comprehensive review of techniques, models, and trends in precision agriculture," *Artif. Intell. Rev.*, vol. 58, no. 3, p. 92, Jan. 2025, doi: 10.1007/s10462-024-11100-x.

[5] K. Kanna S, K. Ramalingam, P. P, J. R, and P. P.C., "YOLO deep learning algorithm for object detection in agriculture: a review," *J. Agric. Eng.*, vol. 55, no. 4, Dec. 2024, doi: 10.4081/jae.2024.1641.

[6] A. Sharma, V. Kumar, and L. Longchamps, "Comparative performance of YOLOv8, YOLOv9, YOLOv10, YOLOv11 and Faster R-CNN models for detection of multiple weed species," *Smart Agric. Technol.*, vol. 9, p. 100648, Dec. 2024, doi: 10.1016/j.atech.2024.100648.

[7] 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.

[8] Y. Alhwaiti, M. Khan, M. Asim, M. H. Siddiqi, M. Ishaq, and M. Alruwaili, "Leveraging YOLO deep learning models to enhance plant disease identification," *Sci. Rep.*, vol. 15, no. 1, p. 7969, Mar. 2025, doi: 10.1038/s41598-025-92143-0.

[9] M. Jelali, "Deep learning networks-based tomato disease and pest detection: a first review of research studies using real field datasets," *Front. Plant Sci.*, vol. 15, Oct. 2024, doi: 10.3389/fpls.2024.1493322.
[10] C. Gupta *et al.*, "Deep vision in agriculture: assessing the function of YOLO in the classification of plant leaf diseases (PLDs)," *BioData Min.*, vol. 18, no. 1, p. 91, Nov. 2025, doi: 10.1186/s13040-025-00497-y.
[11] J. S. Aguilar-Ruiz and M. Michalak, "Classification performance assessment for imbalanced multiclass data," *Sci. Rep.*, vol. 14, no. 1, p. 10759, May 2024, doi: 10.1038/s41598-024-61365-z.
[12] H. N. Ngugi, A. A. Akinyelu, and A. E. Ezugwu, "Machine Learning and Deep Learning for Crop Disease Diagnosis: Performance Analysis and Review," *Agronomy*, vol. 14, no. 12, p. 3001, Dec. 2024, doi: 10.3390/agronomy14123001.
[13] P. Mittal, "A comprehensive survey of deep learning-based lightweight object detection models for edge devices," *Artif. Intell. Rev.*, vol. 57, no. 9, p. 242, Aug. 2024, doi: 10.1007/s10462-024-10877-1.
[14] I.-A. Lin, Y.-W. Cheng, and T.-Y. Lee, "Enhancing Smart Agriculture With Lightweight Object Detection: MobileNetv3-YOLOv4 and Adaptive Width Multipliers," *IEEE Sens. J.*, vol. 24, no. 23, pp. 40017–40028, Dec. 2024, doi: 10.1109/JSEN.2024.3478810.
[15] H. Yu, C. Qian, Z. Chen, J. Chen, and Y. Zhao, "Ripe-Detection: A Lightweight Method for Strawberry Ripeness Detection," *Agronomy*, vol. 15, no. 7, p. 1645, Jul. 2025, doi: 10.3390/agronomy15071645.
[16] D. Lu and Y. Wang, "MAR-YOLOv9: A multi-dataset object detection method for agricultural fields based on YOLOv9," *PLoS One*, vol. 19, no. 10, p. e0307643, Oct. 2024, doi: 10.1371/journal.pone.0307643.
[17] K. Vinoth and S. P, "Lightweight object detection in low light: Pixel-wise depth refinement and TensorRT optimization," *Results Eng.*, vol. 23, p. 102510, Sep. 2024, doi: 10.1016/j.rineng.2024.102510.
[18] J. Kerec, A. L. Machidon, and O. M. Machidon, "Deployment-Aware NAS for Lightweight UAV Object Detectors in Precision Agriculture Crop Monitoring," *AgriEngineering*, vol. 8, no. 2, p. 43, Feb. 2026, doi: 10.3390/agriengineering8020043.
[19] Y. Liu *et al.*, "FEGW-YOLO: A Feature-Complexity-Guided Lightweight Framework for Real-Time Multi-Crop Detection with Advanced Sensing Integration on Edge Devices," *Sensors*, vol. 26, no. 4, p. 1313, Feb. 2026, doi: 10.3390/s26041313.
[20] Z. Ullah, M. Hong, T. Mahmood, and J. Kim, "Systematic Integration of Attention Modules into CNNs for Accurate and Generalizable Medical Image Classification," *Mathematics*, vol. 13, no. 22, p. 3728, Nov. 2025, doi: 10.3390/math13223728.
[21] S. Duhan *et al.*, "Investigating attention mechanisms for plant disease identification in challenging environments," *Heliyon*, vol. 10, no. 9, p. e29802, May 2024, doi: 10.1016/j.heliyon.2024.e29802.
[22] A. El Hanafy, A. Hessane, and Y. Farhaoui, "Enhancing Deep Learning Models with Attention Mechanisms for Interpretable Detection of Date Palm Diseases and Pests," *Technologies*, vol. 13, no. 12, p. 596, Dec. 2025, doi: 10.3390/technologies13120596.
[23] 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, Aug. 2025, doi: 10.3389/frai.2025.1640549.
[24] P. Nasra *et al.*, "Optimized ReXNet variants with spatial pyramid pooling, CoordAttention, and convolutional block attention module for money plant disease detection," *Discov. Sustain.*, vol. 6, no. 1, p. 391, May 2025, doi: 10.1007/s43621-025-01241-6.
[25] Y. Wang, P. Zhang, and S. Tian, "Tomato leaf disease detection based on attention mechanism and multi-scale feature fusion," *Front. Plant Sci.*, vol. 15, Apr. 2024, doi: 10.3389/fpls.2024.1382802.
[26] J. Liu, X. Wang, Q. Zhu, and W. Miao, "Tomato brown rot disease detection using improved YOLOv5 with attention mechanism," *Front. Plant Sci.*, vol. 14, Nov. 2023, doi: 10.3389/fpls.2023.1289464.
[27] E. Yilmaz, S. C. Bocekci, C. Safak, and K. Yildiz, "Advancements in smart agriculture: A systematic literature review on state-of-the-art plant disease detection with computer vision," *IET Comput. Vis.*, vol. 19, no. 1, Jan. 2025, doi: 10.1049/cvi2.70004.
[28] M. Shoaib *et al.*, "An advanced deep learning models-based plant disease detection: A review of recent research," *Front. Plant Sci.*, vol. 14, Mar. 2023, doi: 10.3389/fpls.2023.1158933.
[29] S. Duhan, P. Gulia, N. S. Gill, and E. Narwal, "RTR_Lite_MobileNetV2: A lightweight and efficient model for plant disease detection and classification," *Curr. Plant Biol.*, vol. 42, p. 100459, Jun. 2025, doi: 10.1016/j.cpb.2025.100459.
[30] Y. Ibrahim, M. O. Momoh, K. O. Shobowale, Z. Mukhtar Abubakar, and B. Yahaya, "EDANet: A Novel Architecture Combining Depthwise Separable Convolutions and Hybrid Attention for Efficient Tomato Disease Recognition," *J. Comput. Theor. Appl.*, vol. 3, no. 2, pp. 160–170, Oct. 2025, doi: 10.62411/jcta.14620.
[31] F. M. Firnando, D. R. I. M. Setiadi, A. R. Muslikh, and S. W. Iriananda, "Analyzing InceptionV3 and InceptionResNetV2 with Data Augmentation for Rice Leaf Disease Classification," *J. Futur. Artif. Intell. Technol.*, vol. 1, no. 1, pp. 1–11, May 2024, doi: 10.62411/faith.2024-4.
[32] X. Chi, M. Wang, Y. Gao, and Z. Ge, "Deep learning-based assessment of periapical radiographic image quality," *Sci. Rep.*, vol. 16, no. 1, p. 5047, Jan. 2026, doi: 10.1038/s41598-026-35100-9.
[33] R. N. Asif *et al.*, "Brain tumor detection empowered with ensemble deep learning approaches from MRI scan images," *Sci. Rep.*, vol. 15, no. 1, p. 15002, Apr. 2025, doi: 10.1038/s41598-025-99576-7.
[34] A. Raza, F. Hanif, and H. A. Mohammed, "Clinical validation of lightweight CNN architectures for reliable multi-class classification of lung cancer using histopathological imaging techniques," *Sci. Rep.*, vol. 16, no. 1, p. 6512, Jan. 2026, doi: 10.1038/s41598-026-36652-6.
[35] Y. Tian, Q. Ye, and D. Doermann, "YOLOv12: Attention-Centric Real-Time Object Detectors," *ArXiv*. Feb. 18, 2025. [Online]. Available: http://arxiv.org/abs/2502.12524
[36] R. Hakani and A. Rawat, "Edge Computing-Driven Real-Time Drone Detection Using YOLOv9 and NVIDIA Jetson Nano," *Drones*, vol. 8, no. 11, p. 680, Nov. 2024, doi: 10.3390/drones8110680.
[37] Ultralytics, "YOLOv9: A Leap Forward in Object Detection Technology," *Ultralytics Docs*, 2024. https://docs.ultralytics.com/models/yolov9/ (accessed Dec. 22, 2025).
[38] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," 2018, pp. 3–19. doi: 10.1007/978-3-030-01234-2_1.

[39] L. T. Ramos, E. Casas, C. Romero, F. Rivas-Echeverría, and E. Bendek, "A study of YOLO architectures for wildfire and smoke detection in ground and aerial imagery," *Results Eng.*, vol. 26, p. 104869, Jun. 2025, doi: 10.1016/j.rineng.2025.104869.
[40] A. Aldubaikhi and S. Patel, "Advancements in Small-Object Detection (2023–2025): Approaches, Datasets, Benchmarks, Applications, and Practical Guidance," *Appl. Sci.*, vol. 15, no. 22, p. 11882, Nov. 2025, doi: 10.3390/app152211882.
[41] H. T. Van, G. Van Vu, T. Thanh Tuan, B. Vo, and Y. S. Chung, "LGENetB4CA: A novel deep learning approach for chili germplasm Differentiation and leaf disease classification," *Comput. Electron. Agric.*, vol. 233, p. 110149, Jun. 2025, doi: 10.1016/j.compag.2025.110149.