ORIGINAL ARTICLE

SLIF-Tomato: Inverted Residual Convolutional Block Attention Module for In-field Tomato Leaf Disease Recognition

Romiyal George · Selvarajah Thuseethan · Roshan G. Ragel · Kandiah Pakeerathan · Vaithehi Sivalingam · Thurairatnam Mithuran

Received: 10 September 2025 / Accepted: 11 May 2026 / Published online: 27 June 2026
© The Author(s) 2026

Abstract
Tomato cultivation represents a critical component of nutrition, economic development, and public health, yet it is increasingly compromised by foliar diseases that diminish yield and intensify dependency on hazardous agrochemicals. Although deep learning models have demonstrated strong capabilities for automated disease recognition, existing benchmark datasets exhibit limited real-world utility, primarily due to the absence of in-field imagery and precise annotations of diseased regions. A novel attention mechanism, **I**nverted **R**esidual **C**onvolutional **B**lock **A**ttention **M**odule (**IR-CBAM**), is proposed, combining inverted residual blocks with the CBAM Module, and is specifically tailored to address challenges posed by in-field image variability, such as complex backgrounds and inconsistent lighting. Furthermore, this study introduces **SLIF-Tomato**, the **S**ri **L**ankan **I**n-**F**ield **Tomato** leaf disease dataset, which is the first complete in-field dataset comprising class labels and bounding box annotations collected under diverse real-world conditions. The proposed approach achieved 99.66% and 99.91% accuracy rates on two curated versions of the SLIF-Tomato dataset. Subsequently, the YOLOv12-large model is employed to detect diseased regions, which obtained an average precision score of 88.5%. These contributions advance the development of accurate, efficient and field-adaptable diagnostic systems for tomato leaf disease management in precision agriculture.

Keywords Tomato leaf disease · In-field dataset · Deep learning · Attention · Disease recognition

1 Introduction

Vegetables are essential for sustaining a nutritious and healthy lifestyle worldwide [1]. The *Solanum genus*, a nutritionally and economically significant group within the *Solanaceae* family, comprises tomatoes, brinjal, potatoes, etc [2]. Among them, tomato is a widely consumed vegetable cultivated globally in fields, greenhouses and net houses [3]. Tomatoes provide abundant phytochemicals and essential nutrients such as lycopene, potassium,



https://doi.org/10.1007/s00521-026-12271-0

iron, folate and vitamin C [4]. In Sri Lanka, tomatoes are widely utilised in culinary *applications*[1], such as fresh consumption in salads, processing into juice drinks, jams and sauces, and as a key ingredient in various types of curries. The tomato plant is adaptable to diverse agroclimatic conditions in Sri Lanka and can be cultivated in all regions except the up-country wet zone [5]. According to the IASRI-2018 report, tomato yield in Sri Lanka averages 18.9 tons/ha, contributing to national income and export revenue. It also generates employment, boosts household income, and supports national nutrition and public health [6].

Meanwhile, tomato plants are susceptible to foliar diseases, which significantly hinder their growth. Consequently, farmers encounter multiple challenges that substantially reduce crop yield [7]. A major challenge in vegetable cultivation globally and in Sri Lanka is the high incidence of diseases, which drives increased chemical usage and poses risks to human health. Integrating smart agriculture with early disease detection using deep learning frameworks promotes environmentally sustainable management practices by reducing the reliance on harmful pesticide applications [8, 9]. Tomato cultivation is affected by various diseases, such as Early and Late Blight, Cercospora, Wilt, Powdery Mildew, Septoria, Curly Top and Tomato Mosaic in Sri Lanka [5]. Disease identification in Sri Lanka is limited to a few research stations, which lack systematic screening facilities [6]. Moreover, diseases show location-specific variations, limiting regional generalisation, as evidenced by inter-dataset evaluations [10]. Therefore, region-specific datasets are essential for advancing research in this domain.

Although several benchmark datasets exist for tomato disease recognition, the PlantVillage dataset [11], one of the largest in plant pathology, comprises 22,930 tomato images across ten classes. The Taiwan Tomato Leaves dataset [12] is another widely used resource for tomato disease classification, initially comprising 622 original images across six classes. These datasets have two key limitations that hinder their real-world applicability: (1) the lack of samples collected in field conditions and (2) the absence of bounding box annotations for disease area detection [13]. The absence of bounding box annotations in both datasets limits the applicability of widely used object detection algorithms, such as You Only Look Once (YOLO), for accurate disease area localisation. The PlantVillage dataset comprises images collected in a controlled laboratory environment, whereas the Taiwan Tomato Leaves dataset includes some in-field data. Given the growing emphasis on in-field datasets to improve accuracy and applicability, developing new in-field datasets is essential [14–17].

While in-field datasets enhance the practical applicability of tomato leaf disease recognition, developing accurate recognition frameworks is equally crucial. Although deep learning models demonstrate strong classification accuracy in tomato leaf disease recognition, lightweight neural networks offer greater adaptability for real-world applications [18, 19]. In general, attention mechanisms, such as the convolutional block attention module (CBAM) [20] and squeeze-and-excitation (SE) [21], have been effective at enhancing classification accuracy [22]. However, existing attention mechanisms often lack the specificity needed for in-field conditions, where challenges such as complex backgrounds, variable lighting and occlusions are common [23]. Therefore, there is a need to design attention mechanisms tailored specifically for in-field tomato leaf disease recognition to improve the performance in in-field deployments. The key contributions of this paper are summarised as follows:

1. **A novel state-of-the-art attention mechanism for recognition:** The **I**nverted **R**esidual **CBAM** (**IR-CBAM**), as illustrated in Fig. 1, has been proposed for integration with the lightweight deep learning architectures. To the best of our knowledge, this is the first study to combine inverted residual blocks with CBAM, which enables enhanced feature learning through sequential channel and spatial attention for in-field tomato leaf disease recognition.
2. **A novel benchmark tomato leaf disease dataset:** A new **S**ri **L**ankan **I**n-**F**ield **Tomato** (**SLIF-Tomato**) dataset is constructed to validate in-field deep learning disease recognition and detection approaches. The SLIF-Tomato is the first tomato leaf disease dataset captured under diverse in-field conditions, which offers (1) disease class labels and (2) bounding box annotations for accurate disease area detection.

[1] https://asianfarmers.org/sri-lanka-tomato-production-in-dodamkumbura-gets-a-boost-through-mtcp2/

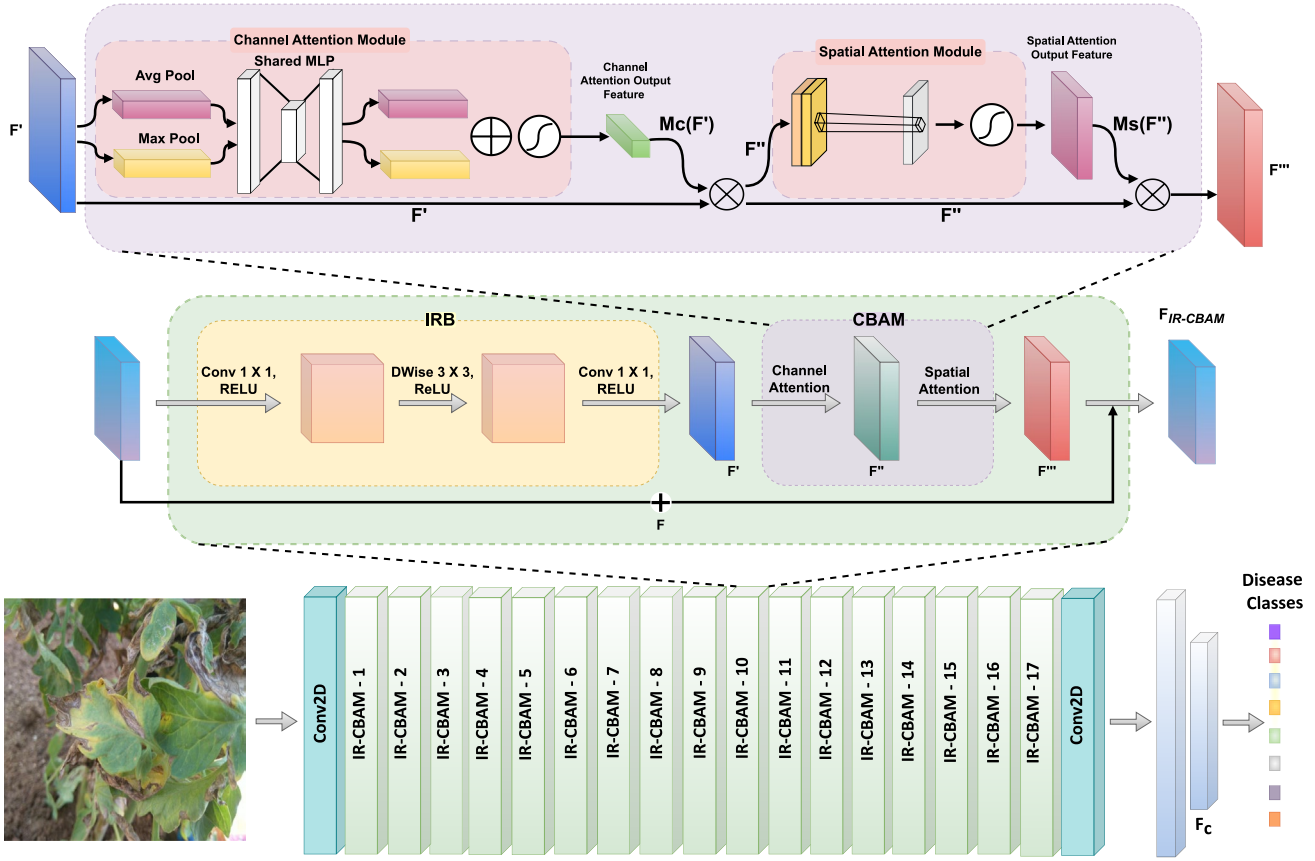


Fig. 1 The overall architecture of the proposed IR-CBAM with MobileNetV2

3. **Diseased region detection:** A diseased region detection approach is introduced that targets the identification of affected areas in tomato leaves. A variant of the YOLO architecture is employed to ensure precise and efficient detection, which supports real-time analysis and allows for targeted interventions to reduce crop loss.

The structure of the rest of the article is organised as follows. Section 2 presents a comparison of existing datasets and a review of key studies on tomato leaf disease recognition. Section 3 presents an overview, the dataset construction process and the statistical summary of the novel SLIF-Tomato dataset. The proposed method is presented in Sect. 4 and the experimental results are discussed in Sect. 5. Finally, Sect. 6 summarises the conclusions and highlights the future work.

2 Related work

Automated tomato leaf disease recognition has attracted growing interest, with research focused on improving both benchmark datasets and recognition algorithms, which are interdependent for reliable classification. This related work section reviews recent work on (1) benchmarking tomato leaf disease recognition methods and (2) evaluating existing tomato leaf disease datasets. For broader context, see recent surveys [24] and [25].

2.1 Evaluation of existing tomato leaf disease datasets

Numerous tomato leaf disease datasets have been utilised in prior research, with the tomato subset of the PlantVillage dataset [11] serving as one of the most widely used benchmarks in the field. This dataset consists of ten classes, including one representing healthy leaves and nine corresponding to different disease conditions, comprising a total of 18,162 samples. Collected under fully controlled laboratory conditions, it is primarily utilised for tomato leaf disease recognition tasks. Although several notable studies have investigated segmentation tasks using this dataset [26], they have relied on manual masking techniques that are not readily available. Detection approaches have also been applied to this dataset [27], supported by the supplementary effort of manually generated annotations.

Another widely utilised dataset is the Taiwan tomato dataset [12], comprising 622 images of tomato leaves categorised into six classes, including five disease categories and one representing healthy leaves. This dataset was collected under a combination of controlled laboratory conditions and real-world environmental settings. Numerous studies have employed this dataset for tomato leaf disease recognition tasks using various techniques, such as distillation learning [28] and metric learning [10]. Its hybrid acquisition setting improves the dataset's ability to capture and represent real-world variability to a certain extent. Despite its relatively limited sample size, the dataset remains a valuable benchmark for evaluating lightweight and few-shot learning methods.

In addition to the two most widely used tomato datasets, a newly introduced resource is the CCMT: Dataset for Crop Pest and Disease Detection [29], which provides both raw and augmented images. The raw dataset comprises 24,881 samples spanning four major crop types: cashew (6,549 images), cassava (7,508 images), maize (5,389 images), and tomato (5,435 images). The tomato subset includes images of both healthy leaves and those exhibiting symptoms of common diseases such as Leaf Blight, Curl and Septoria. Although this dataset is predominantly collected under real-world conditions, it also contains images obtained in controlled laboratory environments. Another notable dataset, the FieldPlant dataset [15], comprises 6,334 samples, including 2,363 images of corn, 1,993 images of cassava, and 1,978 images of tomato. Collected in real-world agricultural environments, this dataset provides a realistic representation of plant conditions, making it valuable for practical disease recognition applications. This dataset exhibits a high level of class imbalance and includes a relatively small number of classes.

In summary, the PlantVillage dataset is collected entirely under controlled laboratory conditions, which ensures consistency in image quality and background uniformity. However, there has been a growing interest among researchers, application developers, and agricultural stakeholders in developing models suited for deployment in real-world agricultural environments. Similar limitations are observed in the Taiwan Tomato dataset, which also lacks environmental variability reflective of field conditions. In contrast, the CCMT and Field Plant datasets are steps toward addressing this issue by providing data collected under real-world conditions. However, they still suffer from limited diversity of disease classes. Additionally, all of these datasets often suffer from other constraints, including limited sample size and imbalanced class distributions. Such limitations can significantly affect the robustness and generalisation capacity of deep learning models, especially when applied to complex and dynamic real-world agricultural settings [30, 31].

2.2 Tomato leaf disease recognition benchmarks

Early studies mainly employed conventional CNNs that process all image regions equally, without selectively focusing on diagnostically relevant areas. However, while these approaches have demonstrated effectiveness with laboratory-acquired tomato leaf datasets, their performance is often limited when applied to images captured under real-world field conditions. Consequently, with the growing shift toward in-field deployment, several studies have suggested that future research should focus on enhancing feature extraction capabilities by incorporating attention-based mechanisms into the model architecture [32, 33]. This limitation arises because non-attention-based approaches often struggle to extract salient features from images with complex backgrounds,

leading to suboptimal feature representation and reduced model accuracy [34, 35]. Conversely, attention mechanisms enhance feature extraction by selectively emphasising salient regions through adaptive feature weighting, as demonstrated in models such as SE-Net [21], ECA-Net [36], and CBAM [20]. These techniques have shown substantial improvements in various computer vision tasks involving images with complex and cluttered backgrounds.

Attention-integrated architectures have demonstrated superior performance in real-world agricultural imaging tasks. The baseline MobileNet-v2 model often directs attention toward background features, leading to suboptimal feature extraction. In contrast, MobileNet-v3 incorporates a channel attention mechanism that enhances the focus on diagnostically relevant regions while suppressing background noise, resulting in improved feature learning. The ECA-SNet_1.0 model, which employs an interactive attention mechanism, achieved a classification accuracy of 99.66%, which illustrates the effectiveness of attention-enhanced architectures in tomato leaf disease recognition [37]. Similarly, SE-ResNet50 achieved 96.81% accuracy, outperforming the baseline ResNet50's 92.56%, which demonstrates the advantages of channel attention modules. In a related study, Ding et al. [34] applied the attention-based RFCA-ResNet to an in-field apple leaf disease dataset, achieving 89.61% accuracy, which surpassed the 89.07% of the baseline ResNet18 model. These results highlight the critical role of attention mechanisms in enhancing model performance under challenging, real-world conditions.

As the novel tomato leaf disease dataset constructed in this study reflects in-field conditions, it is essential to explore attention-based networks that can selectively focus on relevant regions, improving feature representation and performance in complex real-world scenarios.

3 SLIF dataset

This section presents a comprehensive overview of the newly developed SLIF-Tomato dataset, including the need analysis, construction methodology, and statistical summary.

The SLIF-Tomato dataset represents a novel and comprehensive collection of in-field images captured across the entire growing cycle of tomato plants, covering the period from January to December, in 2024. Table 1 presents an overview of the key specifications of the newly constructed SLIF-Tomato dataset. To ensure environmental and geographical diversity within the dataset, leaf images were collected from various agro-climatic zones in Sri Lanka, including the wet zone (Kandy) and the dry zone (Trincomalee and Jaffna), as shown in Fig. 2. These regions in Sri Lanka were strategically selected to capture various climatic conditions, soil types and disease manifestations, enhancing the dataset's generalizability for real-world agricultural applications. The images were captured under natural lighting conditions, with minimal interference, to reflect in-field scenarios encountered by

Table 1 Dataset specifications with details

Specification	Details
Cultivar	*Roma* tomato
Broad subject area	Computer science, agricultural science, machine learning and deep learning
Specific subject area	Image processing and plant disease recognition
Acquisition method	Captured using various high-resolution camera devices. See Table 2 for camera specifications
Data format	Raw and augmented
Annotation	Class labels and disease annotations were provided by pathologists
Region of interest (RoI)	Leaf
Collection parameters	Disease and healthy leaf images of tomato plants
Geographical location	Country: Sri Lanka
	Zones: wet zone (Kandy) and dry zone (Trincomalee and Jaffna)
Data Accessibility	Repository Name and URL: `SLIF-Tomato` DOI: https://doi.org/10.34740/kaggle/ds/8911633

 https://doi.org/10.1007/s00521-026-12271-0

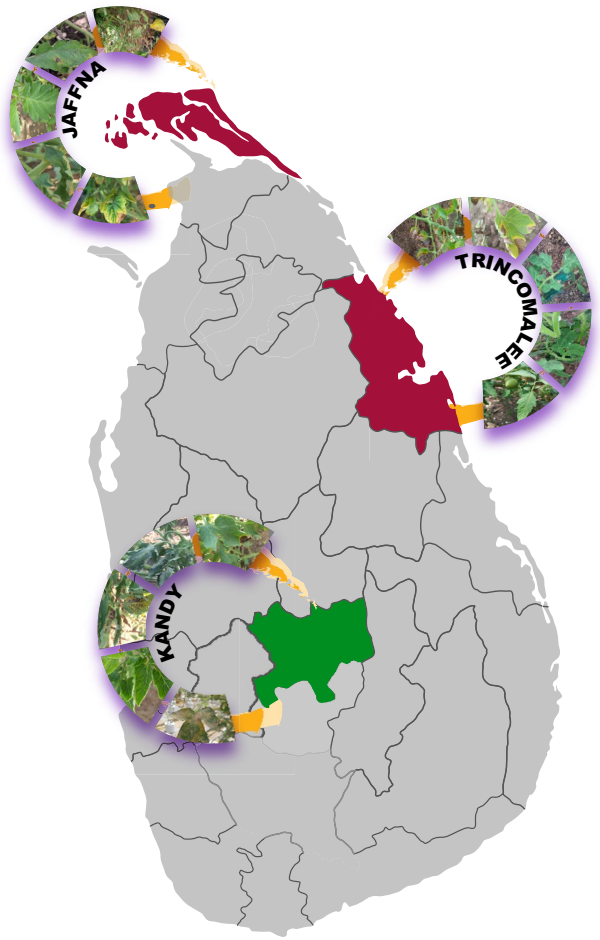


Fig. 2 Sampling regions in Sri Lanka for the SLIF-Tomato dataset. The red regions (Jaffna and Trincomalee) represent the dry zone, while the green region (Kandy) represents the wet zone

farmers and agricultural practitioners. The SLIF-Tomato dataset thus provides a valuable resource for developing and benchmarking automated plant disease diagnosis systems in precision agriculture.

3.1 Need and value of the dataset

In recent years, there has been a significant increase in the demand for real-world precision agriculture solutions, driven by the collaborative efforts of researchers, practitioners and policymakers who aim to improve agricultural productivity and environmental sustainability. Existing tomato leaf disease datasets fall short in supporting these advancements, as they are predominantly collected under controlled laboratory conditions and lack representation from diverse agro-climatic zones. Such limitations hinder the generalizability of machine and deep learning models when deployed in real-world farming environments. As a result, the development of in-field datasets has become critical for enabling context-aware and practically deployable solutions, particularly in the context of precision agriculture and effective plant disease management under real-world conditions [17].

The newly constructed SLIF-Tomato dataset provides high-quality images for prominent tomato disease classes, supporting precise disease classification and region detection. Leveraging this dataset enables researchers to advance disease classification and region detection techniques and contribute to the broader development of data-driven approaches in modern agriculture. Below are the key highlights of the SLIF-Tomato dataset.

1. Collected under in-field conditions across multiple agro-climatic zones, which ensures environmental diversity and real-world applicability.

2. Consists of high-resolution images categorised into eight classes, comprising seven distinct tomato leaf diseases and one class representing healthy tomato leaves.
3. Includes class labels and annotated diseased regions, which support a range of applications requiring disease classification and infected region segmentation.

3.2 Dataset creation

The dataset creation process encompasses multiple stages, such as data acquisition, annotation and augmentation. Figure 3 provides a comprehensive overview of the entire creation process of the SLIF-Tomato dataset.

3.2.1 Acquisition

The acquisition of the SLIF-Tomato dataset was carried out through a well-defined and systematic procedure, consisting of several critical steps to ensure both environmental diversity and dataset representativeness. To capture the natural variability found in in-field agricultural settings, a variety of imaging devices were employed, and image acquisition was conducted at different times throughout the day. This approach facilitated the inclusion of diverse lighting and shading conditions and enhanced the generalizability of the dataset. In addition, images were collected from multiple geographically distinct locations, such as wet and dry zones. This geographical diversity ensures the dataset captures a wide range of environmental conditions.

Table 2 outlines the technical specifications of the imaging devices utilised for data acquisition. The dataset was acquired using five mobile devices equipped with high-resolution cameras, namely the `Samsung A04s`, `Samsung A30`, `Samsung A72`, `Samsung A10s`, `Nokia 6.1 Plus` and `Oppo A37`. The use of multiple imaging devices was necessitated by the collaborative nature of the data collection process, involving several individuals across different locations. Consequently, the dataset incorporates a variety of image resolutions and sensor characteristics, thereby increasing its heterogeneity and enhancing its applicability to diverse real-world scenarios. Additionally, variations in device specifications and capture angles resulted in images with differing dimensions. This diversity in image characteristics reflects real-world agricultural practices, where farmers and researchers commonly utilise smartphones with a wide range of technical capabilities for image capture. Incorporating such heterogeneity aligns with practical deployment scenarios and contributes to the dataset's robustness,

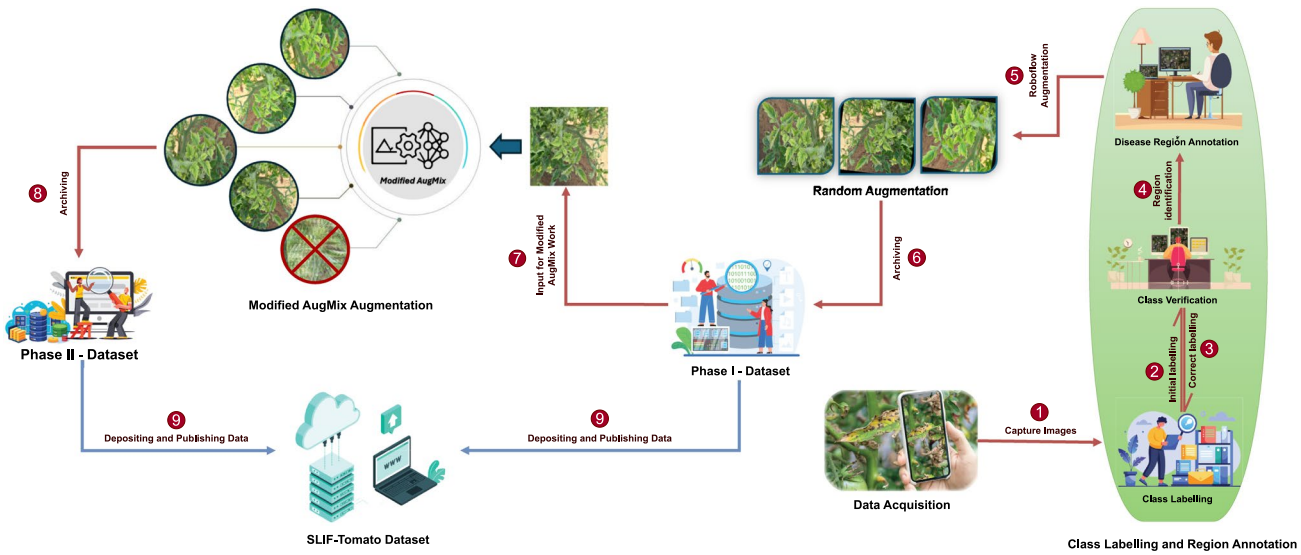


Fig. 3 SLIF-Tomato dataset creation process

 https://doi.org/10.1007/s00521-026-12271-0

Table 2 Device specifications used in data acquisition of SLIF-Tomato dataset

MP - Megapixel, PDAF - Phase detection auto focus and AF - Auto focus

Device	Image resolution details
Samsung A04s	50 MP, f/1.8 (wide), PDAF 2 MP, f/2.4 (macro) 2 MP, f/2.4 (depth)
Samsung A30	16 MP, f/1.7, 27 mm (wide), PDAF 5 MP, f/2.2, 12 mm (ultrawide)
Samsung A72	64 MP, f/1.8, 26 mm (wide), PDAF 5 MP, f/2.4 (macro)
Samsung A10s	13 MP, f/1.8, 28 mm (wide), AF 2 MP, f/2.4 (depth)
Nokia 6.1 Plus	Dual-sensor rear camera: 16 MP (primary) + 5 MP (depth)
Oppo A37	8 MP, f/2.0, 1/3.2", 1.4 µm, AF

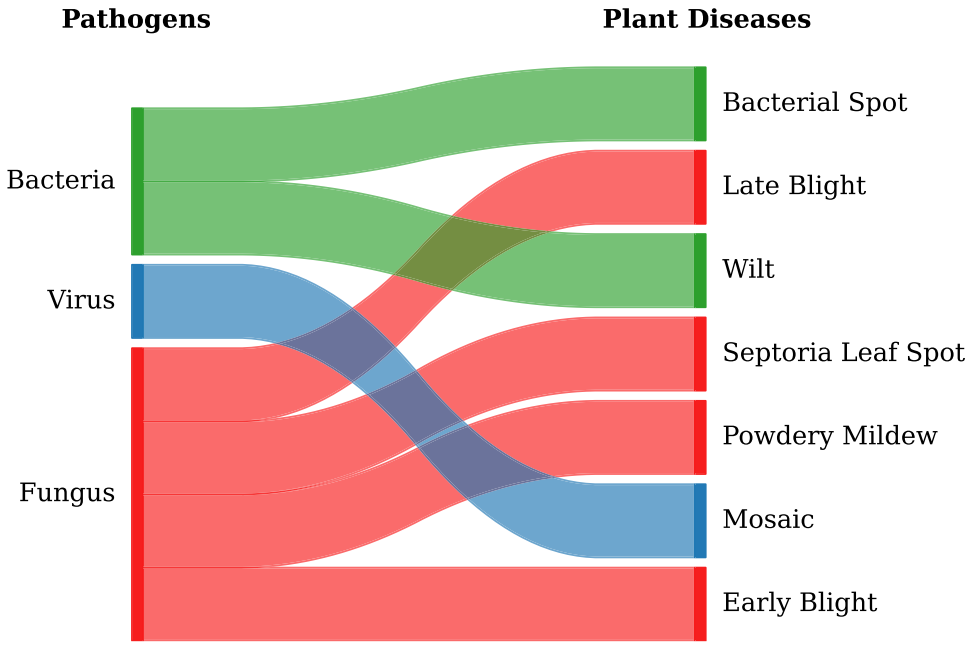


Fig. 4 Illustration of the relationships between plant pathogens and associated diseases

enabling the development of machine learning models that are more resilient to variations in image acquisition conditions and better suited for real-world applications.

3.2.2 Class labelling and region annotation

Following the data acquisition process, the dataset underwent comprehensive (1) class labelling and (2) diseased region annotation steps. Both of these steps are conducted in collaboration with agricultural pathology experts to ensure the accurate identification of plant diseases.

First, the class labelling is carried out by agricultural pathologists, involving the categorisation of images into diseased and healthy tomato leaf classes. To ensure the reliability of the dataset, a cross-verification process was implemented wherein multiple experts independently reviewed and validated the labelled images, thereby minimising potential errors and inconsistencies. The dataset comprises seven disease classes, each attributed to one of three major plant pathogen types: bacteria, viruses and fungi. Figure 4 illustrates the plant diseases represented in the dataset along with their corresponding plant pathogen types. By capturing diverse pathogen-induced disease manifestations and expert-labelled classes, the dataset supports robust training for accurate and generalisable in-field plant disease recognition.

After the class labelling and verification process, the annotation of diseased regions is conducted using the *Roboflow*[2] software. This enables the efficient annotation of disease-affected regions with precise bounding boxes. Roboflow software is selected for its standardised annotation tools, which promote consistency and scalability

[2] https://roboflow.com/

in dataset preparation. Figure 5 presents a sample image with bounding box annotations highlighting diseased regions, accompanied by their corresponding class labels. For samples from each disease class, each image contains multiple leaves with various diseased regions, which results in numerous bounding box annotations. Even for healthy class samples, each image contains multiple leaves, all of which are annotated with several bounding boxes. This characteristic closely mirrors the real-world conditions of tomato fields, reflecting the complexity of in-field imagery. These annotations assist researchers in the precise identification of tomato leaf diseases and the accurate localisation of affected regions.

3.2.3 Augmentation

Following the labelling and annotation process, data augmentation is applied to address the class imbalance problem. A two-phase augmentation strategy is implemented to increase the overall dataset size and enhance the representation of minority classes. Prior to the first phase of augmentation, the images are resized to 640 × 640 pixels to standardise input dimensions.

The first phase involves applying random augmentation across all classes to increase the dataset size. In this phase, two augmented variants are generated for each original image, resulting in the creation of the *Phase I - Dataset*. Each original sample is augmented using all five augmentation schemes listed below, with randomly selected parameters applied within each scheme to generate two augmented images.

1. **Flip:** horizontal and vertical
2. **Crop:** 0% minimum zoom, 17% maximum zoom
3. **Rotation:** between −11° and +11°
4. **Saturation:** adjusted between −26% and +26%
5. **Brightness:** adjusted between −12% and +12%

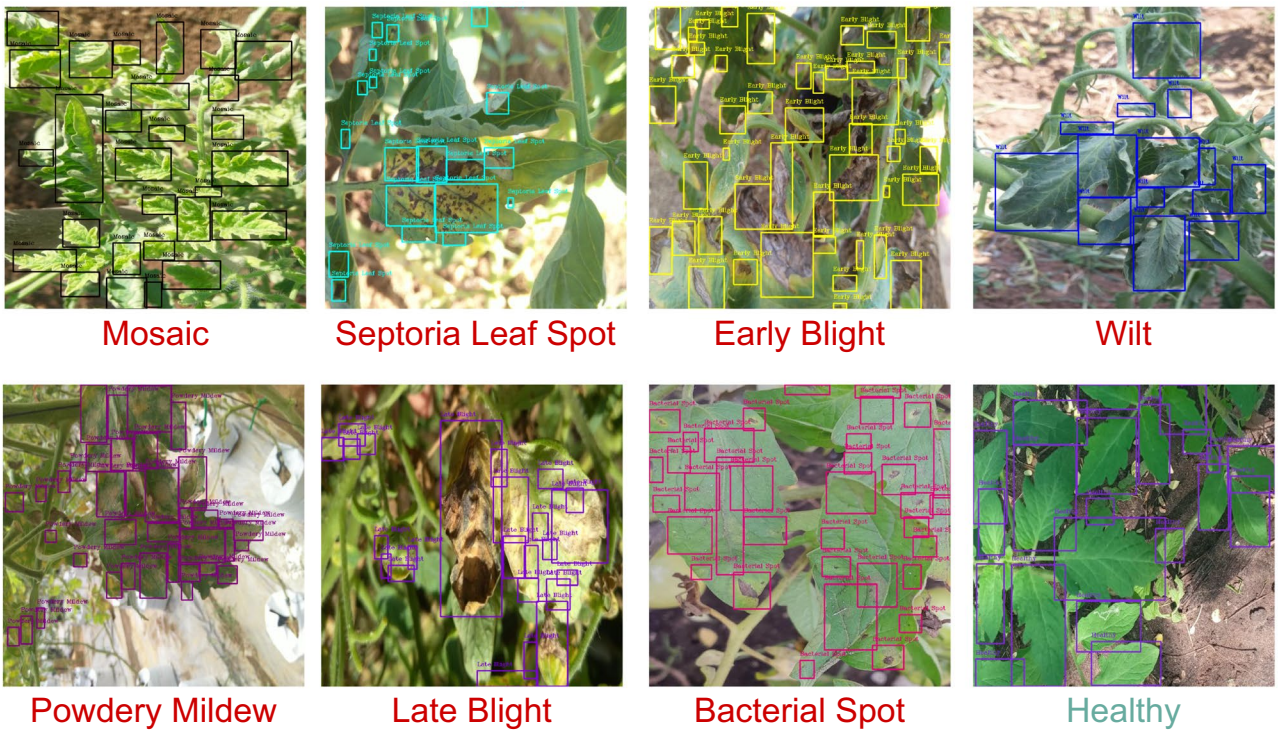


Fig. 5 Bounding box annotations applied to diseased regions in disease class samples and to healthy leaf regions in healthy class samples

The second phase of augmentation, however, adopts a more structured approach by modifying and applying the AugMix [38] algorithm, specifically to enhance the sample size of minority classes. Figure 6 illustrates the proposed modified AugMix augmentation algorithm. The proposed modified AugMix algorithm employed various augmentation schemes, such as `autocontrast`, `equalize`, `posterize`, `rotate`, `solarize`, `shear_x`, `shear_y`, `translate_x`, `translate_y`, `color`, `contrast`, `brightness` and `sharpness`. Typically, augmentation selection is performed randomly, with one to three augmentation schemes applied during each iteration across three iterations. To develop the proposed modified AugMix, augmentation schemes from the original AugMix algorithm that heavily distort the tomato leaf disease images are carefully disabled. A manual review of the dataset is performed to remove such images, ensuring the augmented dataset remains realistic and useful for training. The following constraints are imposed:

- If `shear_x` or `shear_y` is selected, then `translate_x` and `translate_y` are not selected in the same iteration, and vice versa.
- If `translate_x` is selected, then `translate_y` is not selected in the same iteration, and vice versa.
- If `shear_x` is selected, then `shear_y` is not selected in the same iteration, and vice versa.

In this phase, the proposed modified AugMix augmentation algorithm is applied to all classes except early blight, which already has a substantially large number of samples. This results in the creation of the *Phase II - Dataset*, which addresses and eliminates the class imbalance observed in the *Phase I - Dataset*.

3.3 Dataset statics

A total of 890 raw images were collected for the dataset, consisting of eight distinct classes: Bacterial Spot, Early Blight, Healthy, Mosaic, Powdery Mildew, Septoria, Wilt, and Late Blight. After the first phase augmentation (i.e., random), the *Phase I - Dataset* consists of 2,670 images. Following the second phase of augmentation, the dataset was expanded to 8,934 images, resulting in the Phase II - Dataset.

A detailed overview of the data statistics for each phase in the SLIF-Tomato dataset is provided in Table 3. The SLIF-Tomato Dataset exhibits a structured and methodical development across its three phases, progressively addressing key challenges inherent to real-world agricultural datasets. Initially, the Raw Dataset displays substantial class imbalance, with classes such as Early Blight and Mosaic Virus being significantly overrepresented compared to classes like Septoria and Powdery Mildew. This imbalance is partially mitigated in the *Phase I - Dataset* through targeted data augmentation. However, the most notable advancement occurs in the *Phase II - Dataset*, where a modified AugMix augmentation algorithm expands the dataset to 8,934 images. More importantly, a

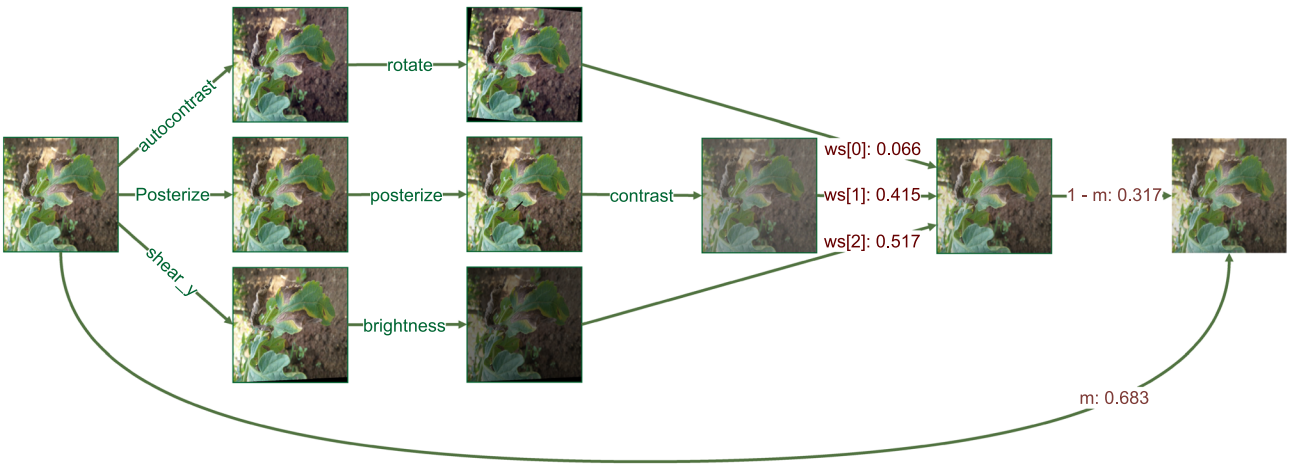


Fig. 6 The modified AugMix augmentation used in the second phase augmentation

Table 3 Description of tomato SLIF-Tomato dataset in different phases

Disease	Number of images		
	Raw dataset	Phase I - dataset	Phase II - dataset
Bacterial spot	21	63	917
Early blight	513	1539	1539
Mosaic	198	594	1544
Powdery Mildew	23	69	1011
Septoria	12	36	698
Wilt	45	135	1028
Late blight	34	102	954
Healthy	44	132	1243
Total	**890**	**2670**	**8934**

comparison between the *Phase I - Dataset* and the *Phase II - Dataset* demonstrates a substantial increase in the number of minority class samples, effectively addressing the class imbalance issue. Although some variation in class sizes persists, the overall distribution is markedly improved, which is crucial for training robust deep learning models capable of generalising across all disease categories.

4 Proposed method

This section presents an overview of the proposed **IR-CBAM** in detail. Designed to enhance tomato leaf disease recognition, IR-CBAM is an advanced attention mechanism that can be integrated into a lightweight backbone architecture to improve its ability to focus on relevant features in complex in-field images. Figure 1 illustrates the overall architecture integrating the proposed IR-CBAM with MobileNetV2 [39] for tomato leaf disease recognition. The architecture comprises two 2-dimensional convolutional layers followed by seventeen IR-CBAM blocks. MobileNetV2 was selected over other lightweight deep models due to its efficient Inverted Residual (IR) structure, which naturally complements the integration of the proposed IR-CBAM module. The architecture's linear bottlenecks and depthwise separable convolutions offer ideal insertion points for IR-CBAM. The IR-CBAM retains the original IR Block (IRB) structure while incorporating a CBAM mechanism, facilitating adaptive attention across channel and spatial dimensions.

4.1 IR-CBAM block

The proposed IR-CBAM block integrates an IRB with a CBAM scheme that includes both channel and spatial attention modules. This integrated design enhances feature representation by enabling the model to capture both channel-wise and spatial dependencies more effectively.

4.1.1 IRB

The IR function, originally introduced in MobileNetV2, is a lightweight architectural component designed to enhance computational efficiency without compromising performance. In plant disease recognition from in-field images, which are often high-resolution and captured under varying conditions, the IR function offers clear advantages. The IR function combines depthwise separable convolutions and linear bottlenecks to reduce computational complexity, which enables real-time inference on resource-limited devices. The IRB in the proposed IR-CBAM employs depthwise separable convolutions to enable efficient feature extraction and is composed of three primary layers as follows:

1. A pointwise 1×1 convolution that expands the dimensionality of the input channels.
2. A depthwise 3×3 convolution that operates independently on each channel, capturing spatial correlations.

3. A second pointwise 1×1 convolution that projects the feature maps back to the original channel dimensionality.

This configuration is typically succeeded by batch normalisation and the ReLU6 activation function. The operation of the IRB can be formally represented as:

$$\mathbf{F}' = \mathrm{IRB}(\mathbf{F}) \tag{1}$$

where $\mathbf{F}$ is the input feature map, and $\mathbf{F}'$ is the output.

4.1.2 CBAM

The CBAM is a lightweight attention mechanism that refines feature representations by sequentially applying *channel* and *spatial* attention. CBAM refines feature maps by sequentially applying channel and spatial attention to emphasise important features and suppress irrelevant ones. Its modular design enables easy integration into CNNs with minimal overhead.

The **channel attention** scheme determines which information is important by assigning attention weights to the feature channels, highlighting the most relevant features. To generate the channel attention map, global average pooling and global max pooling are applied across the spatial dimensions to extract two complementary context descriptors. These descriptors are then passed through a shared multi-layer perceptron (MLP) with a single hidden layer. The resulting outputs are combined through element-wise summation and subsequently passed through a sigmoid activation function to produce the final channel attention map. These operations are defined by the following equations.

$$\mathbf{M}_c(\mathbf{F}') = \sigma\left(\mathrm{MLP}(\mathrm{AvgPool}(\mathbf{F}')) + \mathrm{MLP}(\mathrm{MaxPool}(\mathbf{F}'))\right) \tag{2}$$

$$\mathbf{F}'' = \mathbf{M}_c(\mathbf{F}') \otimes \mathbf{F}' \tag{3}$$

where σ denotes the sigmoid activation function, and the MLP employs shared weights for processing both pooled descriptors. The resulting channel attention map $\mathbf{M}_c(\mathbf{F}')$ is subsequently applied to the input feature map $\mathbf{F}'$ via element-wise multiplication, enabling dynamic recalibration of feature responses along the channel dimension.

The **spatial attention** scheme identifies the spatial locations of the most informative regions within the feature map. It aggregates spatial information by applying average pooling and max pooling operations along the channel axis, yielding two 2D feature maps. These maps are then concatenated along the channel dimension and passed through a convolutional layer with a 7×7 kernel to generate the spatial attention map. The implementation of the spatial attention scheme is defined as follows.

$$\mathbf{M}_s(\mathbf{F}'') = \sigma\left(f^{7\times7}\left([\mathrm{AvgPool}(\mathbf{F}''); \mathrm{MaxPool}(\mathbf{F}'')]\right)\right) \tag{4}$$

where [;] denotes channel-wise concatenation, $f^{7\times7}$ represents a convolution with a 7×7 kernel and σ denotes the sigmoid activation function.

The resulting spatial attention map $\mathbf{M}_s(\mathbf{F}'')$ is then applied to the input feature map through element-wise multiplication to emphasise salient spatial regions:

$$\mathbf{F}''' = \mathbf{M}_s(\mathbf{F}'') \otimes \mathbf{F}'' \tag{5}$$

In IR-CBAM, after applying the IRB, the channel and spatial attention mechanisms are subsequently integrated sequentially. A residual connection is incorporated to improve gradient flow and enhance the stability of feature learning. The final output of the proposed IR-CBAM block is computed as:

$$\mathbf{F}_{\text{IR-CBAM}} = \mathbf{F}''' + \mathbf{F} \tag{6}$$

where $\mathbf{F}$ is the input feature map.

5 Experiments

This section presents the experimental results obtained on the newly introduced SLIF-Tomato datasets, along with the corresponding experimental settings. A central focus of this study is the creation and validation of a novel dataset, with particular emphasis on assessing its suitability for in-field disease classification using a lightweight network enhanced by the proposed IR-CBAM module. However, the experiments conducted in this work are twofold: (1) plant disease recognition and (2) disease region detection.

5.1 Implementation details

All models were trained on a high-performance computing server equipped with dual Intel Xeon 4215R processors, two NVIDIA RTX A6000 GPUs (48 GB each), 128 GB DDR4 2933 MHz RAM (8 × 16 GB modules), a 512 GB HP Z Turbo Quad Pro SSD, and two 4 TB SATA HDDs. The training process was implemented using the PyTorch[3] deep learning framework. This computational infrastructure facilitated efficient model training and contributed to a significant reduction in overall experimentation time.

For the tomato leaf disease recognition experiments, a five-fold cross-validation strategy is employed using the following hyperparameter configuration: a learning rate of 0.0001, 20 training epochs, a batch size of 16, and the Adam optimiser. Both the proposed lightweight architecture integrated with the IR-CBAM module and the baseline MobileNetV2 architecture, as well as additional lightweight models such as SqueezeNet [40], ShuffleNet-V2 (1.0 version) [41], and MobileNet-V3 (small version) [42], are trained under these settings. Fine-tuning is applied across all layers of the models.

For disease region localisation and identification, the YOLOv12l [43], YOLOv11l [44], and YOLOv10l [45] models, initialised with pretrained weights on the COCO dataset, were utilised. Model training was conducted using the following hyperparameters: 500 epochs, a patience value of 50, a batch size of 32, an input image resolution of 640×640, an initial learning rate of 0.001, and a final learning rate of 0.1. The *Phase I – Dataset* was partitioned into training (70%), validation (20%), and testing (10%) subsets. This experiment was designed to evaluate the models' ability to accurately detect disease-affected regions on tomato leaves under diverse environmental conditions.

5.2 Results

The results of the two core experiments, namely tomato leaf disease recognition and tomato leaf disease area detection, are presented. Subsection 5.2 details the results and discussion of the recognition experiments performed on the *Phase I - Dataset* and *Phase II - Dataset*. The outcomes of the disease area detection experiments on the *Phase I* dataset, together with their corresponding analyses, are discussed in Subsect. 5.2.1.

Table 4 summarises the disease recognition results of the proposed model integrated with the IR-CBAM block, in comparison to the baseline MobileNetV2 and additional lightweight models as SqueezeNet, ShuffleNet-V2_1.0, and MobileNet-V3_S, evaluated on both the *Phase I - Dataset* and *Phase II - Dataset*. The proposed model consistently outperforms the baseline and other comparative models across all evaluation metrics, including loss, accuracy (denoted as Accu.), precision (denoted as Prec.), recall, and F1 score. On the imbalanced *Phase I - Dataset*, the proposed model achieves a validation loss of 0.0188 and an accuracy of 99.66%, outperforming the baseline, which reports a higher loss of 0.0273 and a lower accuracy of 99.21%. More importantly, the proposed model yields significant improvements in precision (99.79% vs. 98.48%) and F1 score (99.61% vs.

[3] https://pytorch.org/

Table 4 Performance comparison across Phase I and Phase II datasets, evaluating the proposed model integrated with IR-CBAM alongside other lightweight models

Dataset	Model	Loss	Accu. (%)	Prec. (%)	Recall (%)	F1 (%)
Phase I	SqueezeNet	0.1368	95.77	96.71	89.51	91.75
	ShuffleNet-V2_1.0	0.1059	96.93	81.92	83.83	82.57
	MobileNet-V3_S	0.0298	98.84	97.66	98.69	98.03
	MobileNet-V2	0.0273	99.21	98.48	98.55	98.42
	Proposed	**0.0188**	**99.66**	**99.79**	**99.45**	**99.61**
Phase II	SqueezeNet	0.0368	99.01	99.11	99.14	99.12
	ShuffleNet-V2_1.0	0.0111	99.60	99.66	99.68	99.67
	MobileNet-V3_S	0.0098	99.66	99.69	99.72	99.71
	MobileNet-V2	0.0046	99.85	99.87	99.89	99.88
	Proposed	**0.0034**	**99.91**	**99.93**	**99.92**	**99.92**

98.42%), indicating its enhanced ability to maintain both sensitivity and specificity even under class-imbalanced conditions.

The baseline model and other comparative models showed improved accuracy on the *Phase II - Dataset*, likely due to the correction of class imbalance through targeted augmentation. Similarly, the proposed model demonstrates improved performance in the balanced *Phase II - Dataset*, though the margins are narrower. It achieves a loss of 0.0034 and an accuracy of 99.91%, compared to the baseline's 0.0046 loss and 99.85% accuracy. The proposed model improves or maintains precision, recall, and F1 score at near-perfect levels (above 99.9%). These results show the effectiveness of the IR-CBAM integration.

The proposed model presents a highly optimised balance between computational efficiency and superior representational capacity. With 2.58 million (M) parameters and 1,614.8 million (M) FLOPs, it incurs a minimal computational increase compared to the baseline model (2.55 M parameters, 1,600.1 M FLOPs). This marginal cost results in a slightly longer CPU inference time (175.46 ms vs. the baseline's 173.24 ms). Critically, this incremental overhead is directly attributable to the integration of the enhanced channel–spatial attention mechanism, which significantly bolsters the model's feature extraction capability. This architectural enhancement yields superior performance (accuracy, precision, and F1 Score) compared to the baseline and other models.

Figure 7 presents the confusion matrices obtained for the baseline MobileNetV2 model and the proposed lightweight model integrated with the IR-CBAM block on the *Phase II - Dataset*. Both models exhibit strong overall performance across all eight disease classes, with predictions predominantly aligned along the diagonal, reflecting high classification accuracy. However, the baseline model exhibits minor misclassifications, including two instances from the Early Blight class misclassified as Mosaic and Septoria, and one instance from the Mosaic class misclassified as Early Blight. These confusions suggest challenges in differentiating between visually similar disease patterns using the baseline feature representations.

In contrast, the proposed model integrated with the IR-CBAM block achieves near-perfect classification, with only a single misclassification: one instance of Late Blight incorrectly classified as Mosaic. All remaining classes are correctly classified, even those that were previously misclassified by the baseline model. This improvement highlights the effectiveness of the integrated attention mechanism in enhancing feature discriminability for subtle and complex visual symptoms. Overall, the confusion matrices demonstrate that the proposed model substantially improves class-level precision and robustness.

The t-SNE visualisation of feature embeddings, as shown in Fig. 8, underscores marked differences in the representational effectiveness between the baseline MobileNetV2 and the proposed model. The embeddings generated by MobileNetV2 demonstrate limited class separability, as evidenced by substantial overlap and diffuse clustering among several classes such as Bacterial Spot, Mosaic and Powdery Mildew. Furthermore, the features associated with Early Blight and Septoria are widely dispersed, indicating weak intra-class cohesion and insufficient inter-class discrimination. These observations suggest that the baseline model is unable to capture fine-grained and disease-specific patterns.

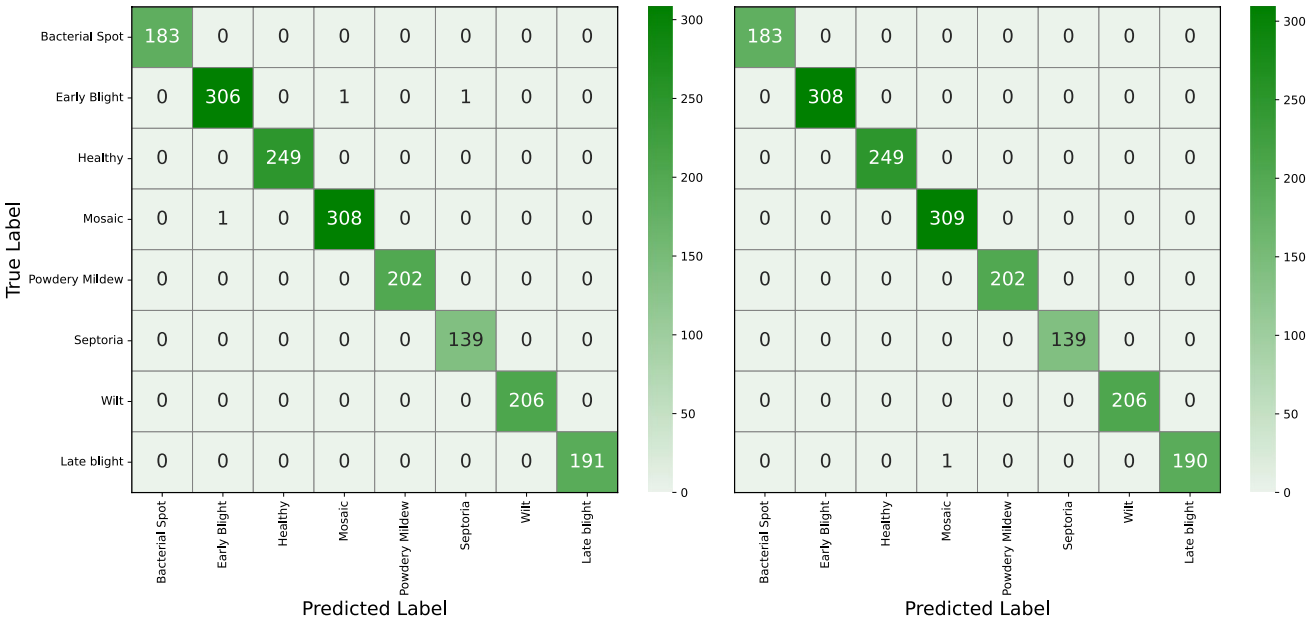


Fig. 7 Confusion matrices obtained for the (left) baseline model and the (right) proposed model with IR-CBAM

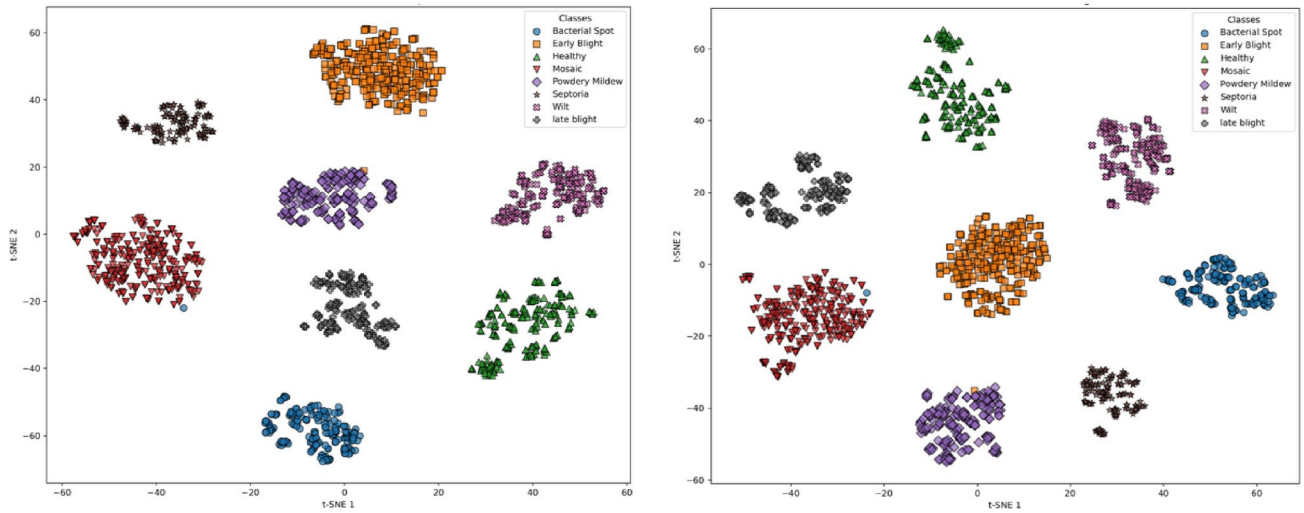


Fig. 8 t-SNE visualisations for the baseline MobileNetV2 and the proposed IR-CBAM

In contrast, the proposed model integrated with the IR-CBAM block produces a more structured and discriminative feature space, with disease categories forming compact and well-separated clusters. These clusters exhibit clearer boundaries and significantly reduced overlap compared to the baseline model. Notably, classes such as Bacterial Spot, Healthy, Wilt and Powdery Mildew are distinctly isolated, which demonstrates enhanced separability. This improvement can be attributed to the integration of the IR-CBAM block, which augments the model's capacity to focus on salient spatial and channel-wise features.

Table 4 clearly demonstrates that the attention mechanism becomes especially critical under class imbalance. In the imbalanced Phase I dataset, the proposed IR-CBAM model reduced significant class imbalance and feature confusion, resulting in substantial performance improvement in Phase I, but with a smaller gain in Phase II. This reduced impact is further supported by Fig. 8, which indicates that the Baseline model in the Phase II dataset is already smoothly separated, though some limitations remain. While IR-CBAM further improves the feature

space by refining boundary regions and increasing cluster compactness, it performs well, resulting in a small increment over the Baseline in Phase II.

Figure 9 presents Grad-CAM visualisations for eight classes, allowing for a direct comparison by displaying the original image alongside the class activation maps generated by the baseline MobileNetV2 (middle column) and the proposed approach integrated with IR-CBAM (right column). Across all class samples, the proposed model integrated with the IR-CBAM module consistently enhances attention toward pathologically relevant regions, resulting in more focused and high-contrast activations precisely over the diseased areas. In contrast, the baseline MobileNetV2 frequently produces diffuse activations, with attention often extending into non-relevant and background regions. These findings highlight the superior spatial localisation and improved interpretability of attention patterns achieved by the IR-CBAM module.

For instance, in the Bacterial Spot class, the proposed IR-CBAM integrated model precisely localises the lesion sites, with sharply focused attention on the dark brown spots and surrounding yellow halos, which are represented as intense red regions. Additionally, the model effectively suppresses responses over healthy tissue and background areas, enhancing the focus on relevant features. In contrast, the baseline MobileNetV2 displays weaker and less discriminative activations, with light red highlights over symptomatic regions and significant attention directed towards irrelevant background areas. This highlights the superior ability of the IR-CBAM module to capture pathologically significant features while minimising irrelevant activations.

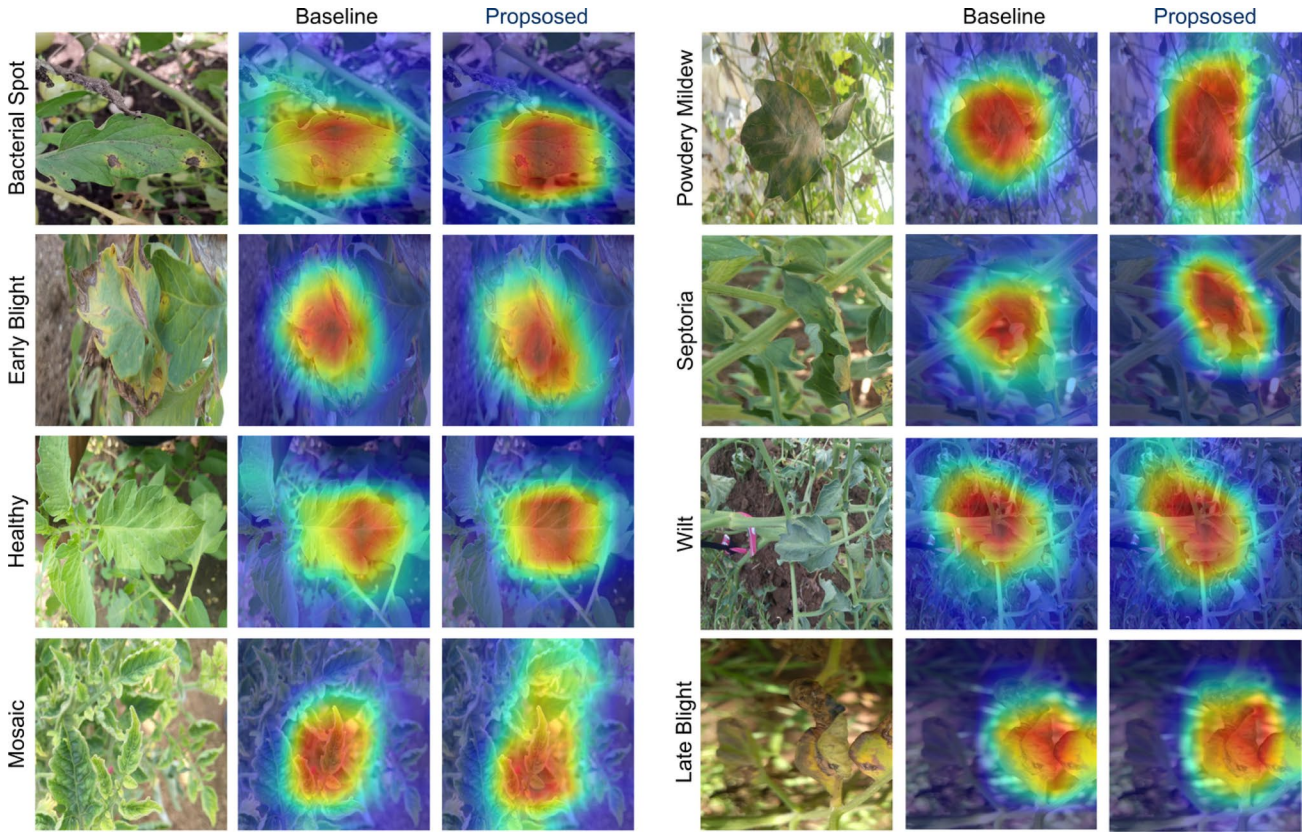


Fig. 9 Grad-CAM visualisation of sample class activations for baseline MobileNetV2 and the proposed IR-CBAM integrated model

Table 5 Performance comparison of YOLO versions for tomato leaf disease area detection

Model	Precision (%)	Recall (%)	mAP@50 (%)	mAP@50:95 (%)
YOLOv10l	80.1	59.8	67.3	45.6
YOLOv11l	80.1	64.5	70.6	47.1
YOLOv12l	**88.5**	**72.8**	**80.1**	**57.2**

Table 6 The tomato leaf disease area detection results obtained for the YOLOv12l model

Class	Precision (%)	Recall (%)	mAP@50 (%)	mAP@50:95 (%)
Bacterial Spot	78.3	67.3	72.2	42.6
Early Blight	87.7	65.1	71.9	47.9
Healthy	93.0	90.9	95.9	77.4
Mosaic	91.7	72.2	81.2	60.8
Powdery Mildew	88.7	61.0	72.5	50.0
Septoria	91.3	80.9	87.7	58.8
Wilt	90.6	63.9	76.3	57.0
Late Blight	86.9	81.0	83.4	63.2
All	**88.5**	**72.8**	**80.1**	**57.2**

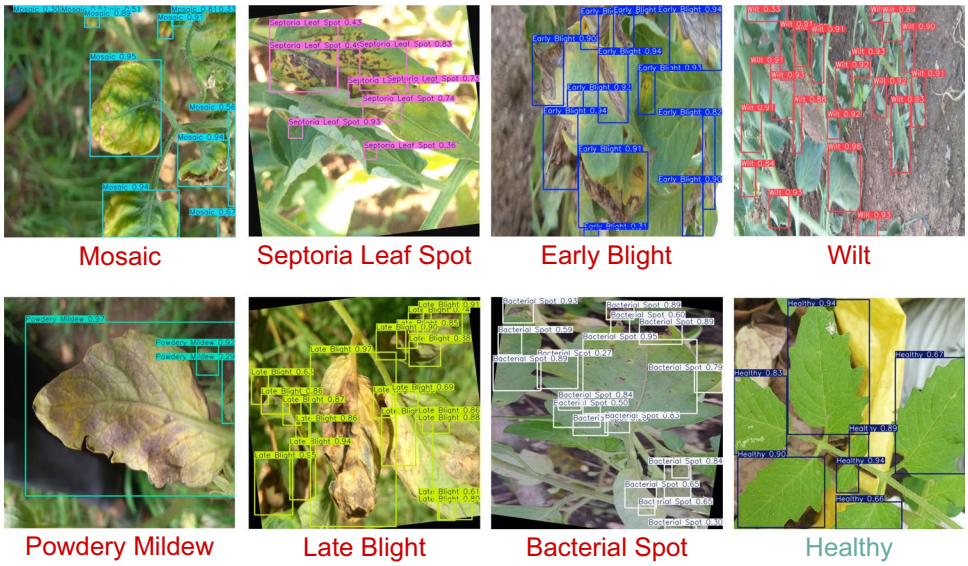


Fig. 10 Visualisation of detection results by YOLOv12l model

5.2.1 Tomato leaf disease area detection

Table 5 presents a performance comparison of the different YOLO versions: YOLOv10l, YOLOv11l, and YOLOv12l, for tomato leaf disease area detection. The YOLOv12l model demonstrates superior performance compared to the other versions. Table 6 summarises the detailed results of tomato leaf disease area detection obtained using the YOLOv12l model.

Figure 10 presents the disease area detection samples for each class. Overall, the model achieves strong performance with an average precision of 88.5%, recall of 72.8% and mAP@50 of 80.1%, indicating its effectiveness in detecting diseased and healthy leaf regions. The high precision scores across most classes, particularly for Healthy (93.0%) and Mosaic (91.7%), suggest that the model is proficient at reducing false positives. However, a notable disparity between precision and recall in several disease classes (e.g., Early Blight and Powdery Mildew) indicates that the model tends to be conservative in its detections, possibly missing a significant number of true disease instances.

The mAP@50:95 values provide a more stringent measure of localisation performance, and the results here are more modest, with an overall average of 57.2%. This drop reflects a challenge in the model's ability to precisely

localise disease-affected areas, especially in more complex or visually subtle classes such as Early Blight (47.9%) and Bacterial Spot (42.6%). Furthermore, metrics, including Precision (78.3%), are lower for Bacterial Spot, primarily due to the lesion's inherent complexity. The symptoms of disease halos and coalescing lesions cause necrosis, and indistinct, low-contrast boundaries complicate detection. This increased complexity directly results in higher false-positive bounding boxes and a reduction in the class-specific metric [46]. Conversely, the Healthy class achieves the highest mAP@50:95 (77.4%), highlighting the model's greater confidence and consistency in identifying non-diseased leaves. These findings suggest that while YOLOv12l performs reliably in distinguishing between classes, improvements are still needed in enhancing the granularity of spatial localisation, particularly for diseases with more variable presentation or overlapping visual features.

6 Conclusion

This paper presents the novel benchmark SLIF-Tomato dataset to advance in-field recognition of tomato leaf diseases. As the first in-field dataset of its kind, the SLIF-Tomato dataset comprises leaf images collected directly from field conditions. This dataset provides disease class labels for accurate recognition and bounding box annotations for precise detection of disease-affected regions. A new attention mechanism, IR-CBAM, combining inverted residual blocks with the CBAM scheme, was proposed to handle complex backgrounds and variable lighting in in-field agricultural settings. To enable in-field tomato disease recognition, the proposed model, integrated with the IR-CBAM module, significantly enhances feature learning, leading to improved recognition of tomato leaf diseases under real-world conditions. Experimental results on the SLIF-Tomato dataset showed that the proposed framework achieved 99.66% and 99.91% accuracy on two dataset variants. Additionally, the YOLOv12-large model was employed for diseased region detection, achieving an average precision score of 88.5%, which demonstrates its efficacy for precise, real-time disease diagnosis. These findings highlight the robustness of the SLIF-Tomato dataset and the effectiveness of the proposed recognition and disease area detection methods in real-world agricultural environments. Future research will expand the SLIF dataset by including additional high-priority crop types such as brinjal, chilli and onion, thereby increasing its applicability across Sri Lankan agricultural systems. The model will also be further improved to effectively handle various levels of disease severity and diverse environmental conditions. For real-world applications, the framework is planned to be deployed on mobile devices with clearly defined performance targets, fast inference times, and minimal computational load to enable smooth, on-site diagnosis. A thorough field evaluation will also be conducted, comparing the model's predictions with farmers' manual disease identification to assess practical reliability and establish a clear pathway for effective deployment within agricultural communities.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions. This research received no external funding.

Data availability The dataset constructed in this study is available at `SLIF-Tomato Dataset`.

Declarations

Ethical No ethical clearance is required.

Competing interest We have no competing interest to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your

intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Knez M, Ranić M, Gurinović M (2024) Underutilized plants increase biodiversity, improve food and nutrition security, reduce malnutrition, and enhance human health and well-being. Let's put them back on the plate. Nutr Rev 82(8):1111–1124
2. Rubatzky, V.E., Yamaguchi, M., Rubatzky, V.E., Yamaguchi, M.: Tomatoes, peppers, eggplants, and other solanaceous vegetables. World Vegetables: Principles, Production, and Nutritive Values, 532–576 (1997)
3. Banoo A, Hussain S, Hussain N, Hussain A, Khan FS, Dar SR (2024) Tomato performance in a protected structure: a review. Adv Res 25(5):29–37
4. Tufail T, Bader Ul Ain H, Noreen S, Ikram A, Arshad MT, Abdullahi MA (2024) Nutritional benefits of lycopene and beta-carotene: a comprehensive overview. Food Sci Nutr 12(11):8715–8741
5. Department of Agriculture, Sri Lanka: Horticultural crop - tomato (2025)
6. Fonseka HH (2020) Vegetable breeding in sri lanka in retrospect. Agricultural Research for Sustainable Food Systems in Sri Lanka: Volume 1: A Historical Perspective, pp 117–147
7. Zhang D, Huang Y, Wu C, Ma M (2023) Detecting tomato disease types and degrees using multi-branch and destruction learning. Comput Electron Agric 213:108244
8. Pakeerathan K, Vaishnavi K (2022) Smart crop protection in organic agriculture and policy recommendations for the sri lankan context. Environment Sustenance and Food Safety Need for More Vibrant Policy Initiatives for Sri Lanka 132:132
9. Sankaran S, Mishra A, Ehsani R, Davis C (2010) A review of advanced techniques for detecting plant diseases. Comput Electron Agric 72(1):1–13
10. Thuseethan S, Vigneshwaran P, Charles J, Wimalasooriya C (2024) Siamese network-based lightweight framework for tomato leaf disease recognition. Computers 13(12):323
11. Hughes D, Salathé M (2015) An open access repository of images on plant health to enable the development of mobile disease diagnostics.arXiv:1511.08060 arXiv preprint
12. Huang M-L, Chang Y-H (2020) Dataset of tomato leaves Mendeley Data 1:2020
13. Joshi K, Hooda S, Sharma A, Sonah H, Deshmukh R, Tuteja N, Gill SS, Gill R (2025) Precision diagnosis of tomato diseases for sustainable agriculture through deep learning approach with hybrid data augmentation. Current Plant Biology 41:100437
14. Gunarathna M, Rathnayaka R (2020) Experimental determination of cnn hyper-parameters for tomato disease detection using leaf images. 2020 2nd International Conference on Advancements in Computing (ICAC), vol 1. IEEE, pp 464–469
15. Moupojou E, Tagne A, Retraint F, Tadonkemwa A, Wilfried D, Tapamo H, Nkenlifack M (2023) Fieldplant: A dataset of field plant images for plant disease detection and classification with deep learning. IEEE Access 11:35398–35410
16. Al-gaashani MS, Shang F, Muthanna MS, Khayyat M, Abd El-Latif AA (2022) Tomato leaf disease classification by exploiting transfer learning and feature concatenation. IET Image Proc 16(3):913–925
17. Wang X, Liu J (2024) An efficient deep learning model for tomato disease detection. Plant Methods 20(1):61
18. Das A, Pathan F, Jim JR, Ouishy MR, Kabir MM, Mridha M (2025) Xltldisnet: a novel and lightweight approach to identify tomato leaf diseases with transparency. Heliyon. https://doi.org/10.1016/j.heliyon.2025.e42575
19. Albattah W, Nawaz M, Javed A, Masood M, Albahli S (2022) A novel deep learning method for detection and classification of plant diseases. Complex & Intelligent Systems 8(1):507–524
20. Woo S, Park J, Lee JY, Kweon IS (2018) Cbam: Convolutional block attention module. Proceedings of the European Conference on Computer Vision (ECCV). pp 3–19
21. Hu J, Shen L, Sun G (2018) Squeeze-and-excitation networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp 7132–7141
22. Takalkar MA, Thuseethan S, Rajasegarar S, Chaczko Z, Xu M, Yearwood J (2021) Lgattnet: Automatic micro-expression detection using dual-stream local and global attentions. Knowl-Based Syst 212:106566
23. Janarthan S, Thuseethan S, Rajasegarar S, Lyu Q, Zheng Y, Yearwood J (2024) Liran: A lightweight residual attention network for in-field plant pest recognition. IEEE Transactions on AgriFood Electronics
24. Thangaraj R, Anandamurugan S, Pandiyan P, Kaliappan VK (2022) Artificial intelligence in tomato leaf disease detection: a comprehensive review and discussion. J Plant Dis Prot 129(3):469–488
25. George R, Thuseethan S, Ragel RG, Mahendrakumaran K, Nimishan S, Wimalasooriya C, Alazab M (2025) Past, present and future of deep plant leaf disease recognition: A survey. Comput Electron Agric 234:110128
26. Shoaib M, Hussain T, Shah B, Ullah I, Shah SM, Ali F, Park SH (2022) Deep learning-based segmentation and classification of leaf images for detection of tomato plant disease. Front Plant Sci 13:1031748

27. Nawaz M, Nazir T, Javed A, Masood M, Rashid J, Kim J, Hussain A (2022) A robust deep learning approach for tomato plant leaf disease localization and classification. Sci Rep 12(1):18568
28. Liu B, Wei S, Zhang F, Guo N, Fan H, Yao W (2024) Tomato leaf disease recognition based on multi-task distillation learning. Front Plant Sci 14:1330527
29. Mensah PK, Akoto-Adjepong V, Adu K, Ayidzoe MA, Bediako EA, Nyarko-Boateng O, Boateng S, Donkor EF, Bawah FU, Awarayi NS et al (2023) CCMT: Dataset for crop pest and disease detection. Data Brief 49:109306
30. Deshpande R, Patidar H (2022) Tomato plant leaf disease detection using generative adversarial network and deep convolutional neural network. Imaging Sci J 70(1):1–9
31. George R, Thuseethan S, Ragel RG (2024) Comparative analysis of pre-trained deep neural networks for plant disease classification. 2024 21st International Joint Conference on Computer Science and Software Engineering (JCSSE). IEEE, pp 179–186
32. Hu WJ, Fan J, Du YX, Li BS, Xiong N, Bekkering E (2020) Mdfc-resnet: An agricultural iot system to accurately recognize crop diseases. IEEE Access 8:115287–115298
33. Yi J, Krusenbaum L, Unger P, Hüging H, Seidel SJ, Schaaf G, Gall J (2020) Deep learning for non-invasive diagnosis of nutrient deficiencies in sugar beet using RGB images. Sensors Basel 20(20):5893
34. Ding J, Zhang C, Cheng X, Yue Y, Fan G, Wu Y, Zhang Y (2023) Method for classifying apple leaf diseases based on dual attention and multi-scale feature extraction. Agriculture 13(5):940
35. Zhang E, Zhang N, Li F, Lv C (2024) A lightweight dual-attention network for tomato leaf disease identification. Front Plant Sci 15:1420584
36. Wang Q, Wu B, Zhu P, Li P, Zuo W, Hu Q (2020) Eca-net: Efficient channel attention for deep convolutional neural networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp 11534–11542
37. Wang P, Niu T, Mao Y, Liu B, Yang S, He D, Gao Q (2021) Fine-grained grape leaf diseases recognition method based on improved lightweight attention network. Front Plant Sci 12:738042
38. Hendrycks D, Mu N, Cubuk ED, Zoph B, Gilmer J, Lakshminarayanan B (2019) Augmix: A simple data processing method to improve robustness and uncertainty.arXiv:1912.02781 arXiv preprint
39. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC (2018) Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 4510–4520 (Mobilenetv 2: Inverted residuals and linear bottlenecks)
40. Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K (2016) Squeezenet: Alexnet-level accuracy with 50x fewer parameters and< 0.5 mb model size.arXiv:1602.07360 arXiv preprint
41. Ma N, Zhang X, Zheng HT, Sun J (2018) Shufflenet v2: Practical guidelines for efficient cnn architecture design. Proceedings of the European Conference on Computer Vision (ECCV), pp 116–131
42. Howard A, Sandler M, Chu G, Chen LC, Chen B, Tan M, Wang W, Zhu Y, Pang R, Vasudevan V et al (2019) Searching for mobilenetv3. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1314–1324
43. Tian Y, Ye Q, Doermann D (2025) Yolov12: Attention-centric real-time object detectors.arXiv:2502.12524 arXiv preprint
44. Jocher G, Qiu, J Ultralytics YOLO11. https://github.com/ultralytics/ultralytics
45. Ao Wang LLea Hui Chen (2024) Yolov10: Real-time end-to-end object detection. arXiv preprint arXiv:2405.14458
46. Potnis N, Timilsina S, Strayer A, Shantharaj D, Barak JD, Paret ML, Vallad GE, Jones JB (2015) Bacterial spot of tomato and pepper: diverse x anthomonas species with a wide variety of virulence factors posing a worldwide challenge. Mol Plant Pathol 16(9):907–920

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Romiyal George[1] · Selvarajah Thuseethan[2] · Roshan G. Ragel[1] · Kandiah Pakeerathan[3] · Vaithehi Sivalingam[4] · Thurairatnam Mithuran[5]

✉ Selvarajah Thuseethan
thuseethan.selvarajah@cdu.edu.au

Romiyal George
romiyalg@eng.pdn.ac.lk

Roshan G. Ragel
roshanr@eng.pdn.ac.lk

Kandiah Pakeerathan
pakeerathank@univ.jfn.ac.lk

Vaithehi Sivalingam
v.kalaianushan@stud.uni-goettingen.de

Thurairatnam Mithuran
it22311740@my.sliit.lk

[1] Department of Computer Engineering, University of Peradeniya, Peradeniya 20400, CP, Sri Lanka

[2] Faculty of Science and Technology, Charles Darwin University, Casuarina 0810, NT, Australia

[3] Faculty of Agriculture, University of Jaffna, Kilinochchi 44000, NP, Sri Lanka

[4] Department of Agricultural Economics and Rural Development, Georg-August-Universität Göttingen, Göttingen 37073, Germany

[5] School of Computing, Sri Lanka Institute of Information Technology, Jaffna 44000, NP, Sri Lanka