



## OPEN ACCESS

## EDITED BY

Qingxia (Jenny) Wang,  
University of Southern Queensland, Australia

## REVIEWED BY

Lijun Qian,  
Prairie View A&M University, United States  
Elio Romano,  
Centro di ricerca per l'Ingegneria e le  
Trasformazioni agroalimentari (CREA-IT), Italy

## \*CORRESPONDENCE

Ranbing Yang  
✉ yangranbing@163.com

RECEIVED 05 September 2024

ACCEPTED 30 December 2024

PUBLISHED 13 February 2025

## CITATION

Sun W, Xu Z, Xu K, Ru L, Yang R, Wang R and  
Xing J (2025) Ultra-lightweight tomatoes  
disease recognition method based on  
efficient attention mechanism  
in complex environment.  
*Front. Plant Sci.* 15:1491593.  
doi: 10.3389/fpls.2024.1491593

## COPYRIGHT

© 2025 Sun, Xu, Xu, Ru, Yang, Wang and Xing.  
This is an open-access article distributed under  
the terms of the [Creative Commons Attribution  
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that the  
original publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or reproduction  
is permitted which does not comply with  
these terms.

# Ultra-lightweight tomatoes disease recognition method based on efficient attention mechanism in complex environment

Wenbin Sun<sup>1</sup>, Zhilong Xu<sup>2</sup>, Kang Xu<sup>1</sup>, Lin Ru<sup>3</sup>, Ranbing Yang<sup>2\*</sup>,  
Rong Wang<sup>4</sup> and Jiejie Xing<sup>2</sup>

<sup>1</sup>College of Information and Communication Engineering, Hainan University, Haikou, China, <sup>2</sup>College of Mechanical and Electrical Engineering, Hainan University, Haikou, China, <sup>3</sup>College of Civil Engineering and Water Conservancy, Heilongjiang Bayi Agricultural University, Daqing, China, <sup>4</sup>Information Technology Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China

A variety of diseased leaves and background noise types are present in images of diseased tomatoes captured in real-world environments. However, existing tomato leaf disease recognition models are limited to recognizing only a single leaf, rendering them unsuitable for practical applications in real-world scenarios. Additionally, these models consume significant hardware resources, making their implementation challenging for agricultural production and promotion. To address these issues, this study proposes a framework that integrates tomato leaf detection with leaf disease recognition. This framework includes a leaf detection model designed for diverse and complex environments, along with an ultra-lightweight model for recognizing tomato leaf diseases. To minimize hardware resource consumption, we developed five inverted residual modules coupled with an efficient attention mechanism, resulting in an ultra-lightweight recognition model that effectively balances model complexity and accuracy. The proposed network was trained on a dataset collected from real environments, and 14 contrasting experiments were conducted under varying noise conditions. The results indicate that the accuracy of the ultra-lightweight tomato disease recognition model, which utilizes the efficient attention mechanism, is 97.84%, with only 0.418 million parameters. Compared to traditional image recognition models, the model presented in this study not only achieves enhanced recognition accuracy across 14 noisy environments but also significantly reduces the number of required model parameters, thereby overcoming the limitation of existing models that can only recognize single disease images.

## KEYWORDS

plant disease identification, image classification, attention mechanism, deep separable convolution, deep learning

## 1 Introduction

Tomatoes, as a widely cultivated and significant crop, possess considerable edible and medicinal value. The diagnosis of early diseases typically relies on the assessment of leaf damage by experts. However, the inability of these experts to provide real-time diagnoses often results in missed opportunities for timely prevention, which can lead to substantial economic losses. Consequently, the development of an automatic and efficient crop disease diagnosis system has become an urgent area of research.

In early research, scholars aimed to leverage computer vision technology for the automatic identification of diseases. [Gulhane and Gurjar \(2011\)](#) employed a Back Propagation Neural Network (BPNN) to identify cotton diseases by extracting relevant disease characteristics. [Sannakki et al. \(2010\)](#) utilized a fuzzy algorithm in conjunction with K-means clustering ([Bashish et al., 2011](#)) to facilitate the grading of leaf diseases. [Xie et al. \(2016\)](#) developed a diagnostic system for wheat leaf diseases based on an Android smartphone, which effectively reduced the computational complexity associated with automated algorithms. Additionally, [Siricharoen et al. \(2016\)](#) diagnosed plant diseases by analyzing leaf textures, colors, and shapes. [Qin et al. \(2016\)](#) employed K-means clustering, the Naive Bayes algorithm, regression trees, and various other supervised classification methods to identify alfalfa diseases, ultimately seeking the optimal method for disease identification in alfalfa. However, the reliance on manually designed features necessitates expert knowledge, which not only limits the degree of automation but also renders the system less adaptable to accommodate variations in growth conditions such as light intensity and complex backgrounds when only a few related features are utilized. Consequently, both recognition accuracy and the level of automation in complex environments require further enhancement.

With the development of computer vision technology and internet technology ([Keswani et al., 2019, 2020](#)), the application of deep learning algorithms ([Krizhevsky et al., 2017](#); [Simonyan and Zisserman, 2014](#); [He et al., 2016](#)) to crop disease recognition has shown great potential ([Dyrmann et al., 2016](#); [Mohanty et al., 2016](#)). [Jiang et al. \(2020\)](#) used deep learning methods to extract the disease characteristics of tomato leaves, such as spot blight, late blight, and yellow leaf curl. Their method predicted the category of each disease after continuous iterative learning, and the accuracy achieved on the training dataset and testing dataset were increased by 0.6% and 2.3%, respectively. [Lv et al. \(2020\)](#) proposed a framework to identify corn leaf diseases based on an improved AlexNet network and feature enhancement, thus improving the extraction performance of corn disease features in complex environments and increasing the accuracy of disease recognition. [Liu et al. \(2020\)](#) used DenseNet to train a generative adversarial network to generate images of 4 different leaf diseases and proposed a leaf disease recognition model based this network. It was further verified that data enhancement could not only effectively overcome the overfitting problem in disease recognition but also effectively improve the recognition accuracy. [Liang and Zhang \(2020\)](#) integrated three classifiers for plant disease image recognition, and the accuracy of

the proposed method on a segmentation testing dataset was close to 99.92%. [Huang et al. \(2019\)](#) proposed an end-to-end plant disease diagnosis model based on deep neural networks that could reliably identify plant types and diseases. [Sanida and Dasygenis \(2024\)](#) proposed a lightweight convolutional neural network for tomato leaf disease identification, achieving an accuracy of 97.63% and an AUC score of 98.51%. [Ni et al. \(2023\)](#) introduced a tomato leaf disease recognition model based on ResNet18, enhanced by the addition of a squeeze-and-excitation module, which attained an average recognition accuracy of 99.63% on the publicly available PlantVillage dataset. [Li et al. \(2023\)](#) developed a tomato leaf disease identification model utilizing a self-attention mechanism, achieving an accuracy of 99.97% with a parameter size of 27 MB on the PlantVillage dataset. Additionally, [Yang et al. \(2024\)](#) proposed a lightweight CNN model, which demonstrated a recognition accuracy of 95.54%. These studies have proven that data enhancement can improve the robustness of models in different environments, but such recognition methods still have some shortcomings. (1) These methods all improve the recognition accuracy of the developed models by sacrificing complexity, lacking focus on the model's complexity. (2) The input images of these end-to-end models can contain only one leaf. If the leaf size is small, the background is large, or multiple diseased leaves appear in one image, the models cannot recognize it. (3) Robustness studies regarding the data variability caused by image acquisition devices are scarce.

Therefore, to overcome these problems, this paper presents a framework that combines tomato leaf detection and leaf disease identification and is mainly divided into three parts. First, a tomato leaf detection model is designed to detect and crop tomato leaves in captured images to solve the problem of non-singular disease image recognition. Then, an effective data enhancement method is designed to improve the robustness of the model in complex environments. Finally, an ultra-lightweight tomato disease recognition model is designed to reduce required the number of model parameters while ensuring that the recognition accuracy remains unchanged and to balance the contradictory relationship between the complexity and recognition accuracy of the model.

## 2 Related work

To build a tomato leaf detection model in a complex environment, it is necessary to select a suitable object detection network. Mainstream object detection algorithms based on deep learning are mainly divided into two categories, and different types of object detection algorithms have different performances. The first category contains single-step detection networks that do not generate proposal regions but can directly convert an object frame positioning problem into a regression processing problem to achieve rapid object detection; such methods include You Only Look Once (YOLO) ([Redmon et al., 2016](#)), YOLOv3 ([Redmon & Farhadi, 2018](#)), YOLOv4 ([Bochkovskiy et al., 2020](#)), YOLOv6 ([Li et al., 2022](#)), YOLOv7 ([Wang et al., 2022](#)), YOLOv9 ([Wang et al., 2024b](#)), YOLOv10 ([Wang et al., 2024a](#)) and other networks, but

their disadvantage is that they are prone to missing and falsely detected objects. The second type includes two-step detection networks based on proposal regions, which have high detection accuracy and positioning accuracy, and the probabilities of missed detections and false detections are relatively small; such methods include the fast region-convolutional neural network (Fast R-CNN) (Girshick, 2015) (Fast Region-Convolutional Neural Network) and Faster R-CNN (Ren et al., 2017). The tomato leaf detection results directly affect the accuracy of the utilized disease recognition model, so how to accurately detect tomato leaves in a complex background environment is particularly important. In this paper, we choose Faster RCNN as the basis for constructing a tomato leaf detection network due to its higher accuracy.

The recognition accuracy of image classification algorithms based on deep learning is superior to that traditional algorithms. VGG (Simonyan and Zisserman, 2014), ResNet (He et al., 2016), and GoogLeNet (Szegedy et al., 2015) realize deep convolutional neural networks (DCNNs) by increasing the number of utilized convolutional layers and widening their network structures, resulting in stronger feature extraction capabilities but requiring many hardware resources. To reduce the number of model parameters, MobileNetV1 (Howard et al., 2017) uses deep separable convolution to significantly improve computational efficiency, while MobileNetV2 (Sandler et al., 2018) employs a resource-efficient block with an inverse residual and a linear bottleneck to extend it. MobileNetV3 (Howard et al., 2019) introduces an attention mechanism and modifies the number of extended layer filters. These MobileNet series networks have become mainstream lightweight models, as they greatly reduce the number of required model parameters under the premise of ensuring recognition accuracy. Although MobileNetV3 incorporates an attention mechanism into the lightweight network and achieves better recognition performance, the introduced attention mechanism uses two fully connected layers to perform dimensionality reduction and then upgrade operations on channel features, resulting in an increased number of parameters and greater feature losses. Therefore, this paper designs an ultra-lightweight tomato disease recognition model based on the MobileNet series of models.

Because the image features of early diseased leaves are not obvious and a large number of network parameters can be used to extract the texture features of leaves, few studies have been conducted on the use of lightweight models for crop disease identification. If a lightweight network is used for disease image recognition, it is necessary to improve the feature extraction ability of the model and the utilization of model parameters to extract the features of diseased leaves at different scales. An attention mechanism enables a network to pay more attention to the most effective information in the image and ignore the irrelevant information, so it is considered an effective module for the aggregation of enhanced features. The squeeze-and-excitation network (SE-Net) (Jie et al., 2017) first proposed network that included an effective mechanism for learning channel attention, and it has achieved good performance. A convolutional block attention module (CBAM) (Woo et al., 2018) uses both average pooling and maximum pooling to aggregate features. Most of the

attention modules proposed later, such as the Gram-Schmidt orthogonalization procedure (GSoP) (Gao et al., 2018) and Gather-excite (GE) (Hu et al., 2018), have high model complexity and can only be used in a single block or several convolutional blocks. Notable, all the above methods focus on developing complex attention modules to achieve better performance. An efficient attention mechanism aims to learn effective channel attention with low model complexity. The introduction of a high-efficiency attention mechanism into a tomato disease recognition model can reduce the impact of model weight on the resulting recognition accuracy.

Therefore, this paper defines a framework that combines tomato leaf detection and tomato disease recognition, divides tomato the disease recognition into three parts, and optimizes the method used for each part according to tomato disease characteristics to make it suitable for tomato disease recognition in real scenes. The main contributions of this article are as follows:

1. By using MobileNetV2 to improve the feature extraction module of Faster RCNN, a fast detection method for tomato leaves in a complex environment is proposed to cut out the complete leaves in the camera's field of view from the background.
2. The existing tomato disease dataset was formed by picking diseased leaves and taking their pictures in an ideal experimental environment, so it lacks disease images in real growth environments and complex environments. Therefore, this paper collects tomato leaf images in an unconstrained tomato planting environment and designs a set of data enhancement methods to enhance the robustness of the recognition model in complex environments.
3. An ultra-lightweight tomato disease recognition model based on an efficient attention mechanism is proposed; it balances the contradictory relationship between model complexity and accuracy.

### 3 Materials and methods

Data collection is an important part of crop disease identification. In this paper, tomato leaf images were collected at the tomato planting base of the Beijing Agricultural Information Technology Research Center. Under natural light, a Sony camera was used to capture tomato images with resolutions of 3024×4023 and 3024×3024. The disease image collection times were 6:00-8:00 am, 10:00-12:00 am and 2:00-6:00 pm. The collected disease category database includes powdery mildew, leaf mould and late blight, with a total of 2437 images. The images collected in an unconstrained environment are stored according to their disease categories. It is worth noting that to increase the diversity of the dataset, the photographed tomato disease images not only include tomato leaves, stems and fruits but also images collected at different angles, at different scales and with different background information for the same leaf. These complex environments, with characteristics such as light, dirt, healthy stems and leaves, the

ground, human hands, and other objects, affect the results of tomato disease recognition. Therefore, this paper designs a framework that integrates tomato leaf detection and leaf disease recognition and proposes a rapid leaf detection model and an ultra-lightweight tomato disease recognition model that is suitable for complex environments. As shown in Figure 1, the framework is mainly divided into two parts. The first part mainly uses an object detection model to obtain a dataset of tomato leaf diseases, and then enhances the data of tomato leaves. A lightweight leaf recognition model is constructed using the enhanced tomato leaf disease data, and the trained recognition model is used to recognize leaf diseases.

### 3.1 Tomato leaf detection in a complex environment

#### 3.1.1 Data preparation

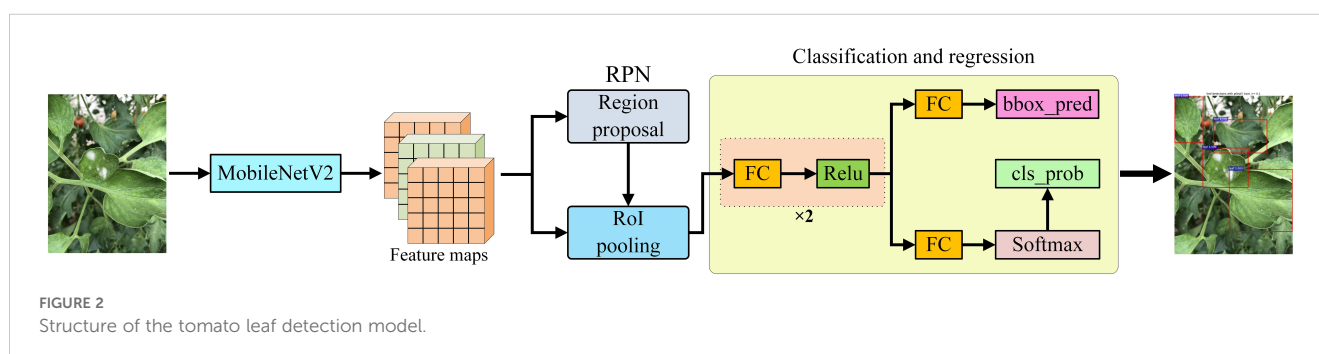
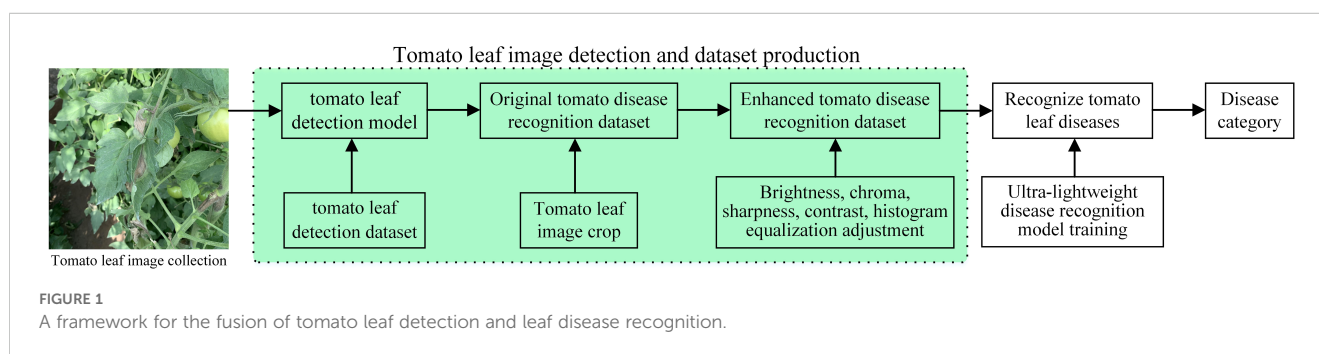
Pascal VOC2007 (Ren et al., 2017) is a public dataset in the field of object detection. To train the tomato leaf detection model, 800 images are extracted from among the collected images as the tomato leaf detection dataset and annotated according to the Pascal VOC2007 format. According to the standard that the target area accounts for more than 3/4 of the entire leaf area, it is considered a complete leaves. ImgLabel software is used to manually select complete healthy leaves and diseased leaves with clear pixels as the target areas, and fuzzy, incomplete leaves and other background information are uniformly defined as the image background and label labelled. The annotated images generate labels in Extensible Markup Language (XML) format as the tomato leaf detection dataset. (Xu et al., 2024), the dataset are randomly divided in an 8:1:1 ratio, resulting in 648 training images, 80 testing images, and 72 validation images.

#### 3.1.2 Tomato leaf detection model

The accuracy of tomato leaf detection is directly related to the accuracy of the disease identification model. Currently, mainstream target detection models are mainly divided into two categories: single-stage and two-stage. Two-stage target detection algorithms consist of two steps: first, generating candidate regions, and then applying a classifier to these regions. This method is more accurate than single-stage detection but slower. Considering that the detection effect of leaves significantly impacts subsequent results, this paper constructs a tomato leaf detection model based on the Faster RCNN, which has high accuracy in two-stage target detection networks, as shown in Figure 2.

The tomato leaf detection model includes a feature extraction network, region proposal network (RPN) and leaf classification and regression network. MobileNetV2 is used as the feature extraction network in Faster RCNN to extract tomato leaf features, reduce the number of model parameters, and then input the generated feature map into the RPN to search for a predefined number of suggested regions. The feature map output by MobileNetV2 and the region proposal output by the RPN are input into the region of interest (RoI) pooling layer, and after being adjusted to a fixed size, each leaf area is input into the leaf classification and regression network to detect and identify tomato leaves. The reasoning process of the tomato leaf detection model is shown in Figure 3.

Each image is scaled to a size of 600×1000 pixels and input into the tomato leaf detection model to detect all clear and complete tomato leaf coordinates in the image, and the output detection results are rescaled to fit the original image. When there are no leaves in the given image, the number of detected leaves is 0, and the image is not saved. The tomato leaf detection model accurately extracts all clear and complete leaf images in the camera's field of view, thereby overcoming the limitation of current disease



recognition research that the input image can only contain one leaf. This paper automatically detects images containing multiple tomato leaves and inputs the detected single leaves into the disease recognition network, which not only improves the accuracy and adaptability of tomato disease detection but also greatly reduces the workloads of researchers.

### 3.1.3 Tomato disease image enhancement

After the complete tomato leaves are cut out from the images taken by the camera, they are divided into four categories: powdery mildew, leaf mould, late blight and healthy leaves. These four kinds of disease data form the original tomato disease recognition dataset, which is divided into a training set and a test set at a ratio of 8:2. To enhance the robustness of the tomato disease recognition model in an unconstrained environment, this paper uses 13 data enhancement methods to expand the number of training images and test images. The enhancement results are shown in Figure 4.

Image enhancement methods include flipping an image horizontally or vertically and increasing or decreasing the brightness, chroma, contrast, or sharpness of the image. These methods can simulate special lighting noise in a real environment,

not only increasing the diversity of the training set but also enabling the design of comparative experiments to test the robustness of the developed model in a complex and changeable environment. In the image field, the Laplacian algorithm can highlight the edge information of an image, the gamma transform can perform different contrast enhancements according to different grey values, and contrast-limited adaptive histogram equalization (CLAHE) can enhance the local contrast of an image and increase the number of detailed features. The tomato disease dataset expanded using these methods is called the enhanced tomato disease recognition dataset, in which the brightness, chroma, contrast, etc. are increased to 50% of the original image or reduced to 50% of those of the original image.

### 3.2 Ultra-lightweight tomato leaf disease recognition model

At present, the mainstream recognition model has a relatively high number of parameters and consumes a relatively large amount of hardware resources. To balance the contradictory relationship

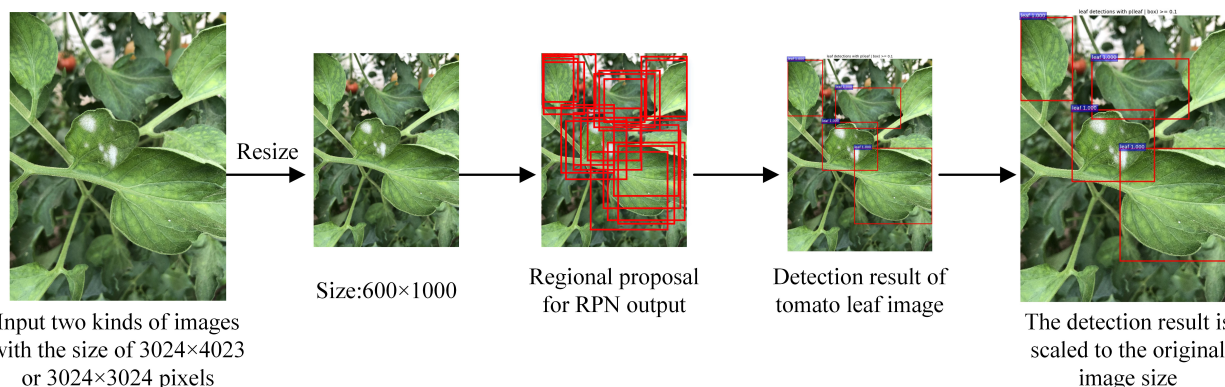


FIGURE 3 Schematic diagram of the inference process of the tomato leaf detection model.

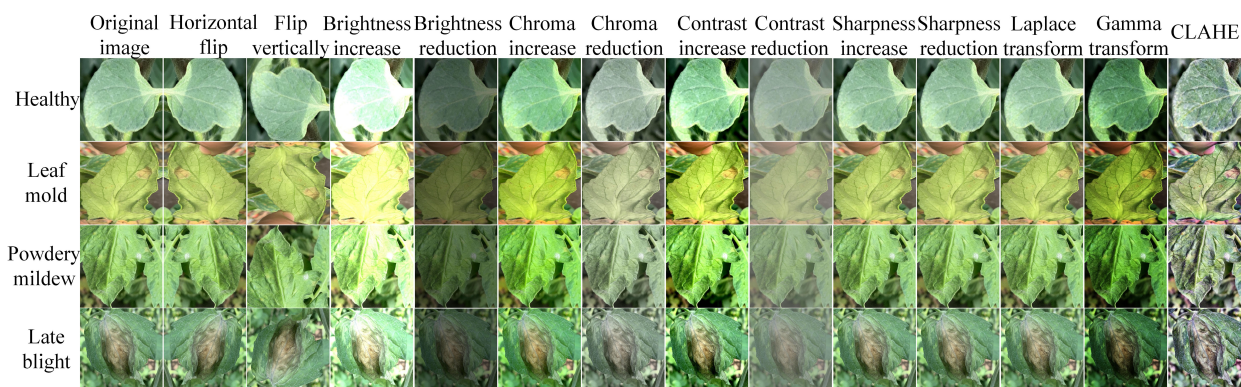


FIGURE 4 The results of tomato disease image enhancement.

between accuracy and complexity and to facilitate the deployment and application of the model, this paper uses deep separable convolution, an inverted residual structure and efficient channel attention (ECA) to design an ultra-lightweight tomato leaf disease recognition model. The specific architecture is shown in Figure 5.

To preserve more detailed diseased leaf features, five improved inverted residual structures are designed in the model, based on MobileNet v3 small, and the input image size of the model is set to 448×448 size. After a tomato leaf image feature is extracted by the case-based reasoning (CBR) module, the obtained feature map is input into the feature extraction network, which contains 5 improved inverted residual modules. Finally, the designed last-stage module uses a 1×1 convolution instead of a fully connected layer and uses a global pooling layer to transform 7×7 feature maps into 1×1 objects, which greatly reduces the number of required parameters. A dropout layer is added between the last two 1×1 convolutional layers of the model to prevent the model from overfitting. The parameters of each module are shown in Table 1.

The model uses rectified linear unit 6 (Relu6) as the non-linear activation function and limits the maximum value of the output to 6. The calculation method is shown in equation (1):

$$\text{Relu6}x = \min(\max(0, x), 6) \tag{1}$$

The deep network easily causes gradient degradation when extracting tomato disease characteristics. The traditional residual structure performs compression first and then conducts expansion, thus losing much of the effective information contained in the feature map. Therefore, this paper designs five inverted residual structures to perform feature extraction after upgrading the channel information, and the structures are shown in Figure 6.

To reduce the number of parameters required by the tomato disease recognition model, depth-wise separable convolution (DSC), which is composed of depth-wise (DW) convolution and point-wise (PW) convolution, is introduced. The use of DSC to replace the traditional standard convolution can reduce the numbers of model parameters and calculations. The inverted residual (IR) module uses PW convolution to upscale the channels and then uses DW convolution to extract features so

that the model can extract rich feature information. Finally, PW convolution is used to keep the dimensionality of the input channel consistent with that of the output channel, thus achieving cross-layer connections and lightweight models. Since tomato leaf disease is mainly manifested in the texture information of tomatoes, to enhance the ability of the model to extract detailed leaf image features, an ECA mechanism is incorporated into the inverted residual structure. The module in Figure 6A does not perform dimensional upscaling and directly inputs the given feature map into the DW convolution mechanism. The IR modules in Figure 6B and Figure 6C both use PW convolution, DW convolution, and more PW convolution operations to increase the dimensionality of the input features and then extract the features. When the DW convolution step length is 1, a cross-layer connection is adopted, and when the step length is 2, a cross-layer connection is not adopted. In the last PW convolution, only a batch normalization (BN) structure is used, and the non-linear activation function is not used.

The standard convolution input is a feature map  $L_i$  of size  $h_i \times w_i \times d_i$ , an output feature map  $L_j$  of size  $h_j \times w_j \times d_j$  is generated by the convolution kernel  $K \in R^{k \times k \times d_i \times d_j}$ , and the calculation cost is  $h_i \cdot w_i \cdot d_i \cdot d_j \cdot k \cdot k$ . The DSC mechanism uses two layers instead of the standard convolution operation. The first layer is a DW convolution layer, which performs lightweight filtering on each input channel, and the second layer is a 1×1 point-by-point convolution layer, which constructs new features by calculating linear combinations of the input channels. However, the effect of DSC is similar to that of standard convolution, but the computational cost, which is the sum of a DW convolution and a 1×1 PW convolution, is lower:

$$h_i \cdot w_i \cdot d_i(k^2 + d_j) \tag{2}$$

where the DW convolution in the lightweight model designed in this paper uses 3×3 and 5×5 convolution operations, and the computational cost is reduced by 9-25 times compared with that of traditional convolution.

Reducing the number of model parameters can easily reduce the resulting recognition accuracy. To improve the accuracy of tomato

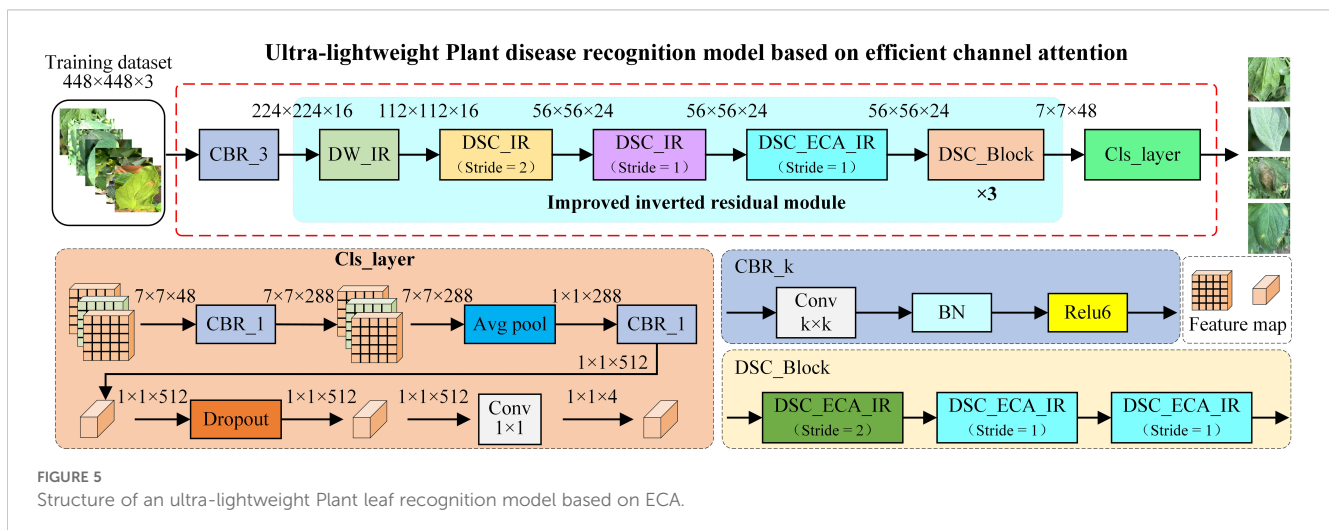


FIGURE 5 Structure of an ultra-lightweight Plant leaf recognition model based on ECA.

TABLE 1 Specific parameters of each module.

Input	Module	Expansion channel	Output channel	ECA	Step	Output size
448×448×3	Conv2d, 3×3	—	16		2	224×224×16
224×224×16	DW_IR, 3×3	16	16	√	2	112×112×16
112×112×16	DSC_IR, 3×3	72	24		2	56×56×24
56×56×24	DSC_IR, 3×3	88	24		1	56×56×24
56×56×24	DSC_ECA_IR, 3×3	88	24	√	1	56×56×24
56×56×24	DSC_ECA_IR, 5×5	96	40	√	2	28×28×40
28×28×40	DSC_ECA_IR, 5×5	240	40	√	1	28×28×40
28×28×40	DSC_ECA_IR, 5×5	240	40	√	1	28×28×40
28×28×40	DSC_ECA_IR, 5×5	120	48	√	2	14×14×48
14×14×48	DSC_ECA_IR, 5×5	144	48	√	1	14×14×48
14×14×48	DSC_ECA_IR, 5×5	288	48	√	1	14×14×48
14×14×48	DSC_ECA_IR, 5×5	288	48	√	2	7×7×48
7×7×48	DSC_ECA_IR, 5×5	288	48	√	1	7×7×48
7×7×48	DSC_ECA_IR, 5×5	288	48	√	1	7×7×48
7×7×48	Conv2d, 1×1	—	288	—	1	7×7×288
7×7×288	Avgpool, 7×7	—	—	—	1	1×1×288
1×1×288	Conv2d, 1×1	—	512	—	1	1×1×512
1×1×512	Dropout	—	—	—	—	1×1×512
1×1×512	Conv2d 1×1	—	4	—	1	1×1×4

The pointwise convolution channel in each module is called the expansion channel of the module. The '√' indicates whether the ECA module is used.

disease recognition, this paper introduces an ECA mechanism to improve the inverted residual modules, as shown in Figures 6A, C. The specific structure of the ECA mechanism is shown in Figure 7.

A feature map  $X$  is output after a convolution-based transformation as  $u_c(i, j) \in R^{W \times H \times C}$ , where  $W$ ,  $H$  and  $C$  are the width, height and channel dimensions, respectively. The attention module uses a compression operation to compress a feature map  $U$  into a  $1 \times 1 \times C$  format. This operation aggregates features across the spatial dimensions ( $H \times W$ ), generates channel descriptors, and obtains aggregate features  $F(u_c) \in R^C$  without dimensionality reduction. The channel attention can be learned by formula (3).

The channel attention can be learned by formula (3).

$$\omega = \sigma(MF(u_c)) \quad (3)$$

where  $\sigma$  is a sigmoid function, which is the activation function of the attention module,  $M$  is the convolution calculation:

$$\text{sigmoid}(z) = \frac{1}{1 + e^{-z}} \quad (4)$$

The ECA module uses global average pooling (GAP) to obtain an aggregate feature  $F_A(u_c)$ :

$$F_A(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (5)$$

When the convolution operation extracts features, the variance of the estimated value is easily increased due to the limited size of the neighborhood. GAP can reduce this error and retain more image background information.

To capture local cross-channel interactions and ensure the effectiveness of channel features, we can use a 1D convolution operation to make all channels share the same learning weight. The calculation of  $\omega_i$  is dependent on  $F_i(u_c)$  and its  $k$  neighbors, i.e.,

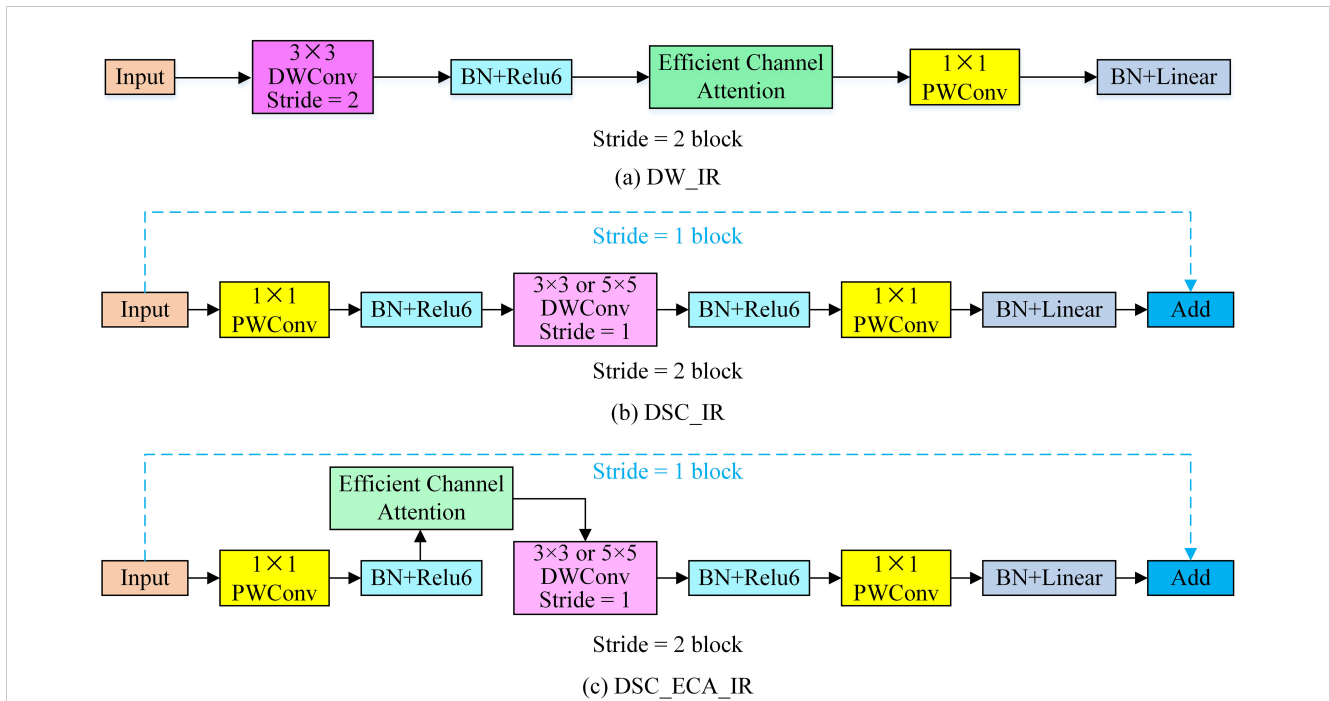
$$\begin{cases} \omega_i = \sigma(\sum_{j=1}^k \omega^j F_i^j(u_c)), F_i^j(u_c) \in \phi_i^k \\ \omega = \sigma(\text{Conv1D}_k(F_i(u_c))) \end{cases} \quad (6)$$

where  $\phi_i^k$  indicates the set of  $k$  adjacent channels of  $F_i^j(u_c)$ .  $\text{Conv1D}_k$  is a 1D convolution with a convolution kernel of  $k$ , and formula (6) involves only  $k$  parameters. Under the premise of low complexity, the efficiency and effectiveness of the model are ensured by appropriately capturing local cross-channel interactions.

Therefore, the  $M$  operation in formula (3) is a one-dimensional convolution operation, and the aggregate feature  $F_A(u_c)$  is input into the 1D convolution layer to filter the effective channel information and obtain  $\omega_c$ :

$$\omega_c = \sigma(\text{Conv1D}_k(F_A(u_c))) \quad (7)$$

The convolution kernel size  $k$  is set as an adaptive parameter. For a given number of channels  $C$ ,  $k$  is calculated as follows:



**FIGURE 6** The improved inverted residual module. **(A)** DW\_IR is an inverted residual module based on depth-wise convolution and an efficient attention mechanism. **(B)** is an inverted residual module based on depth-wise separable convolution (DSC\_IR). **(C)** is an inverted residual module based on depth-wise separable convolution and an efficient attention mechanism (DSC\_ECA\_IR).

$$k = \tau(C) = \left\lfloor \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right\rfloor_{\text{odd}} \tag{8}$$

where  $\gamma = 2$ ,  $b = 1$ ,  $\lfloor \alpha \rfloor_{\text{odd}}$  represents the nearest odd number of  $\alpha$  and  $\tau(C)$  is a nonlinear mapping.

$$x_c = F_{\text{scale}}(u_c, \omega_C) = \omega_C \cdot u_c \tag{9}$$

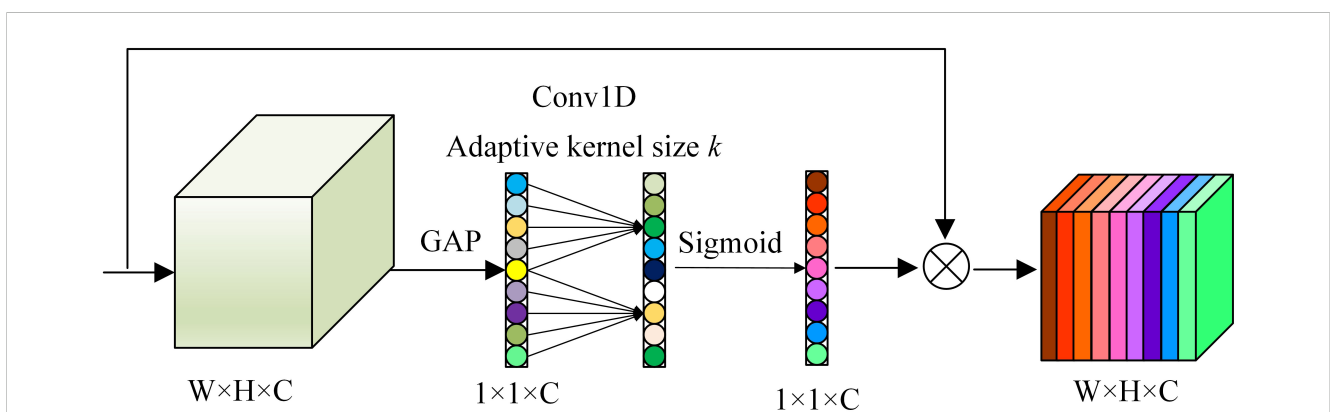
where  $X = [x_1, x_2, \dots, x_c]$  and  $F_{\text{scale}}(u_c, \omega_C)$  represent the channel multiplication relationship between the scalar  $\omega_C$  and the feature map. The ECA module does not reduce the dimensionality of the original channel features. High-dimensional channels have longer-range interactions, which are not limited to the local acceptance domain of the convolution response, and realize the enhancement and recalibration of important features in the former and latter layers.

## 4 Experimental results and analysis

### 4.1 Experimental configuration and data

The models are trained on one NVIDIA Tesla P100 GPU with 16 GB of RAM based on a 64-bit Ubuntu 16.04 operating system and the PyTorch framework, with Python version 3.7.6, PyTorch version 1.3.0, CUDA API version 10.0, and cuDNN version 7.5.1.

According to the data preprocessing procedure in Figure 1, the original tomato disease recognition dataset and the enhanced tomato disease recognition dataset are produced, as shown in Table 2. The original collected tomato disease dataset contains 3 kinds of diseases, and a total of 2,437 tomato images with two sizes,



**FIGURE 7** The specific structure of the ECA mechanism.



TABLE 2 Number of images in the tomato disease dataset.

Category	Captured image	Tomato leaf image crop	Original tomato disease recognition dataset		Image enhancement	Enhanced tomato disease recognition dataset	
			Training set	Testing set		Training set	Testing set
Powdery mildew	460	968	775	193	13552	10850	2702
Late blight	1231	2335	1868	467	32690	26152	6538
Leaf mould	746	1441	1153	288	20174	16142	4032
Healthy	0	1257	1006	251	17598	14084	3514
Total	2437	6001	4802	1199	84014	67228	16786

3024×4023 and 3024×3024. To train the tomato leaf detection model, 800 images are extracted from the collected images and manually labelled as the tomato leaf detection dataset. Finally, the overall dataset is divided into 80 testing images, 648 training images and 72 validation images. The trained tomato leaf detection model is used to crop the tomato leaves with powdery mildew, late blight, leaf mould and healthy tissues from the original image, and 6,001 leaf images are obtained; these form the original tomato disease recognition dataset and are divided into a training set and testing set at a ratio of 8:2. Thirteen kinds of image enhancement methods are used to expand the training set and testing set of the original tomato disease recognition, and 84,014 enhanced leaf images are obtained, forming the enhanced tomato disease recognition dataset.

## 4.2 Testing results of the tomato leaf detection model

The training process and detection results of the tomato leaf detection model are shown in Figure 8 and Figure 9, respectively. During the training process, the total loss function exhibits a downward trend, and the oscillation interval is stable between 0.15-0.20. After the training process is completed, the model can effectively detect the clear leaves in the original collected tomato images while ignoring the blurred leaves and incomplete leaves in the background. The calculated accuracy of the model is 93.7%, the recall is 85%, the weight of the model is 88MB, and which proves that the model has a good detection effect for tomato leaves.

At the same time, to verify the effectiveness of the proposed model, we conducted comparative experiments with YOLO v8x, YOLO v8l, and Faster R-CNN. The accuracy, recall rates, and model sizes are as follows: for the YOLO v8x model, the accuracy is 95%, the recall rate is 85.5%, and the model size is 130 MB; for the YOLO v8l model, the accuracy is 92.5%, the recall rate is 83.2%, and the model size is 84 MB; for the Faster R-CNN model, the accuracy is 94.5%, the recall rate is 85.9%, and the model size is 467 MB. Compared to the original Faster R-CNN, the detection performance of our proposed model is slightly lower, but the model size is only about one-fourth of that. In comparison with YOLO v8x, the model size is approximately two-thirds of that, and there is little difference in detection performance. However, when compared to YOLO v8l,

the detection performance is improved by about 2 percentage points, while the model size only increases by 4 MB.

## 4.3 Recognition results for tomato leaf diseases

To verify the effectiveness of the tomato leaf disease recognition model, the original tomato disease recognition training set and the enhanced tomato disease recognition training set are used to train the ultra-lightweight tomato disease recognition model based on the ECA mechanism proposed in this paper. The robustness of the model is tested on the testing set derived from 14 kinds of different environments. To further analyze the impact of the improved module on the recognition results, an ablation experiment is carried out on the ultra-lightweight tomato disease recognition model. Finally, the same parameter configuration and methods are used to train mainstream recognition networks, including the MobileNet series, VGG16, ResNet50 and AlexNet, for a comparison with the model proposed in this article in terms of recognition performance.

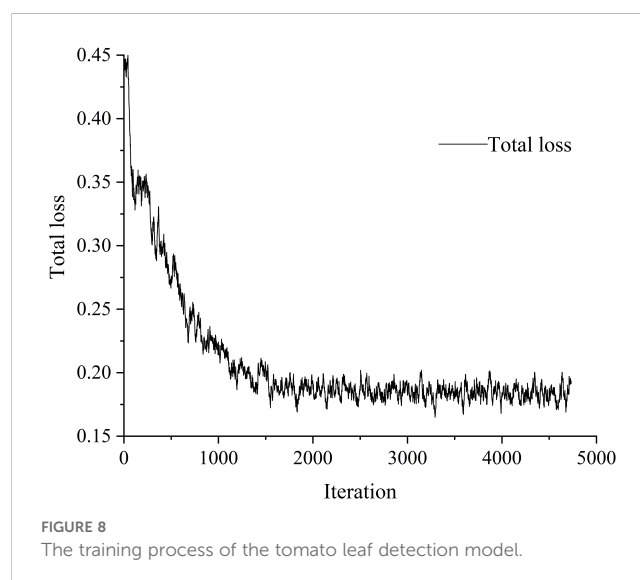


FIGURE 8 The training process of the tomato leaf detection model.

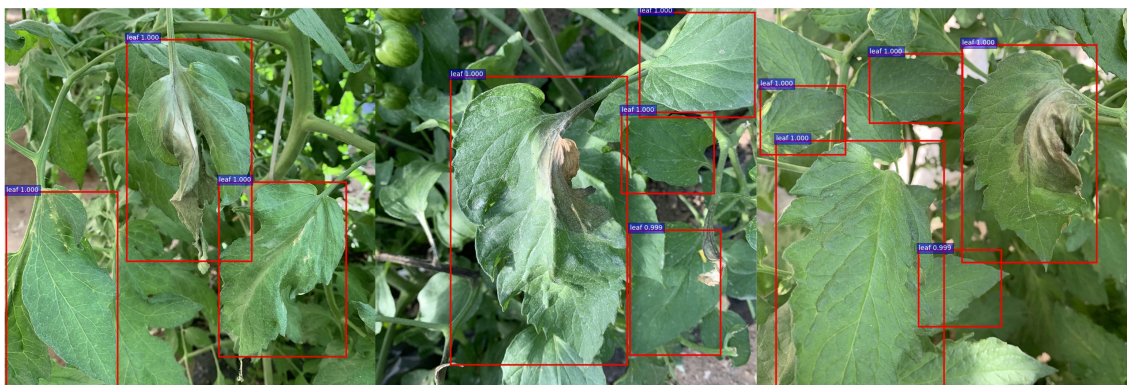


FIGURE 9 Test results of the tomato leaf detection models in different environments.

### 4.3.1 Training results of the ultra-lightweight tomato disease recognition model

The original tomato disease recognition training set and the enhanced tomato disease recognition training set are separately used to train the disease recognition model, and then the trained model is verified on the same enhanced testing set. The image size used for the two training sets is scaled to 448\*448 pixels. Figure 10 shows the change curve of the loss function and the accuracy achieved on the testing set during the training process. As the number of training epochs increases, the loss functions of the model for the two datasets first decrease and then stabilize, and the accuracy on the testing set also increases and then stabilizes. However, the loss function of the model trained on the enhanced tomato disease recognition dataset is lower, and the accuracy on the testing set reaches 97.84%. The accuracy of the model trained on the original tomato disease recognition dataset is only 77.09% on the same testing set, which is far lower than the recognition accuracy of the model trained on the enhanced tomato disease recognition dataset. It is proven that the 13 kinds of image enhancement

methods used in this paper can greatly improve the accuracy of the model with respect to the task of tomato disease recognition.

### 4.3.2 Comparison of the recognition results of each model

To further verify the influence of the model and enhancement method proposed in this paper on the recognition results, the current mainstream models are trained on the original tomato disease recognition training set and the enhanced tomato disease recognition training set and then uniformly tested on the enhanced testing set. The results are shown in Figure 11.

After conducting training on the enhanced tomato disease recognition training set, the recognition accuracies of the tomato disease recognition models in different environments are much higher than those of the models trained on the original tomato disease training set. This is because the enhanced testing set contains images of diseased leaves in a complex environment. Data enhancement can simulate diseased tomato leaves in a complex unconstrained environment, so training models with the

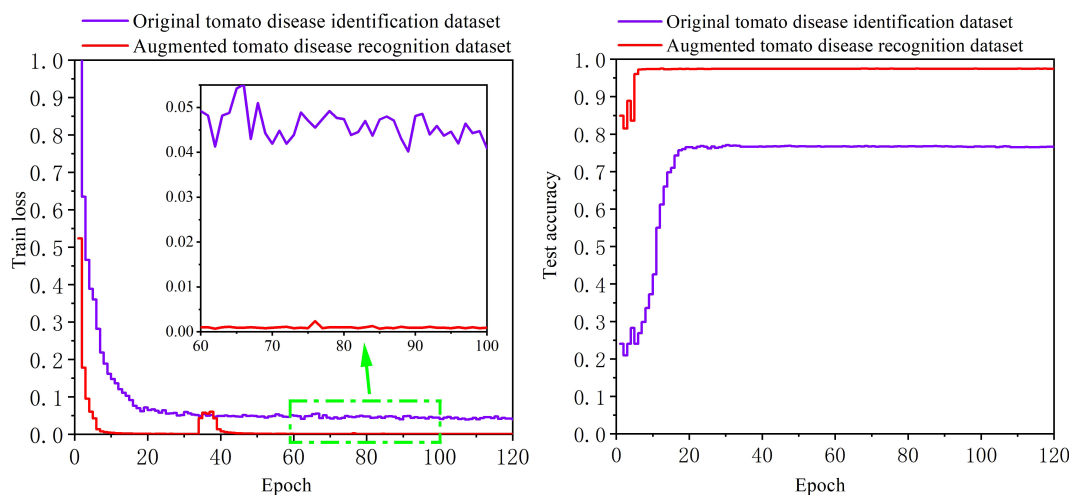


FIGURE 10 The training results of the ultra-lightweight tomato disease recognition model obtained with different datasets.

enhanced tomato disease dataset can make them more robust. In addition, although the proposed model trained on the original tomato disease dataset has a recognition accuracy of 77.09% on the testing set, which is lower than that of AlexNet, this value is much higher than the test results of other networks. However, the recognition accuracy of the proposed model on the testing set after training on the enhanced tomato disease recognition dataset is 97.84%, which is higher than the recognition accuracies of all other network models. The experimental results prove that the image enhancement method selected in this paper can effectively improve tomato recognition accuracy. When the number of datasets is expanded to 84,014, the shallow network AlexNet exhibits certain limitations and is unable to obtain higher accuracy, and the model proposed in this paper has the highest accuracy among all tested models.

To further study the influence of the utilized complex environment on the models, each type of network is trained on two datasets and then uniformly tested on 14 enhanced tomato disease recognition testing sets. The weight of the version of each model with the highest average accuracy on the testing set is saved, and then the influences of different environments on the accuracy of model recognition are analyzed. The test results obtained after training on the original tomato disease recognition dataset are shown in Table 3, and the test results obtained after training on the enhanced tomato disease recognition dataset are shown in Table 4.

Upon analyzing the recognition results of the models trained on the different sub-datasets of the enhanced tomato disease recognition testing set, it is found that the accuracy of each model on the original images of the testing set is very high. The image flip, sharpness transformation and Laplace transformation have the least influence on the recognition result, because sharpness transformation and Laplace transformation can enhance the details and edge information of the image, while flipping changes

the angle of the image and has little effect on its texture and shape. However, changes in the brightness, chroma, and contrast of an image or a gamma or CLAHE transformation reduce the recognition accuracy of the model to varying degrees, mainly because these transformations have a significant impact on the color, texture, and other aspects of the image, and these types of data are missing from the original data. The lack of these types of data in the original data results in the model lacking robustness to this type of noise. Meanwhile, based on the experimental results in Tables 3, 4, we can conclude that the number of model parameters does not necessarily represent the recognition performance of the model. Excessive model parameters may lead to the extraction of redundant features, thereby affecting the performance of the model. On the contrary, a small number of parameters may also reduce the feature extraction capability of the model. It can be seen from Table 3 that without image enhancement, the recognition accuracy of the model proposed in this paper is higher than that of most models in different environments. Through image enhancement, the model can be adapted to the changes in these tomato images, and the recognition accuracy of the model in different environments can be significantly improved. In Table 4, the accuracies of the model proposed in this paper is maintained at approximately 97% in different environments. The recognition accuracy in most environments is higher than that of other models, and the average accuracy is 97.84%, which is higher than that of other models. The experimental results verify that the model proposed in this paper not only has the highest recognition accuracy for tomato disease images in an unconstrained environment but also has better robustness to images with multiple types of noise interference.

The tomato disease recognition model designed in this paper not only has high accuracy and strong robustness in a variety of complex environments but is also ultra-lightweight. Table 5 compares the number of parameters utilized by each model. The

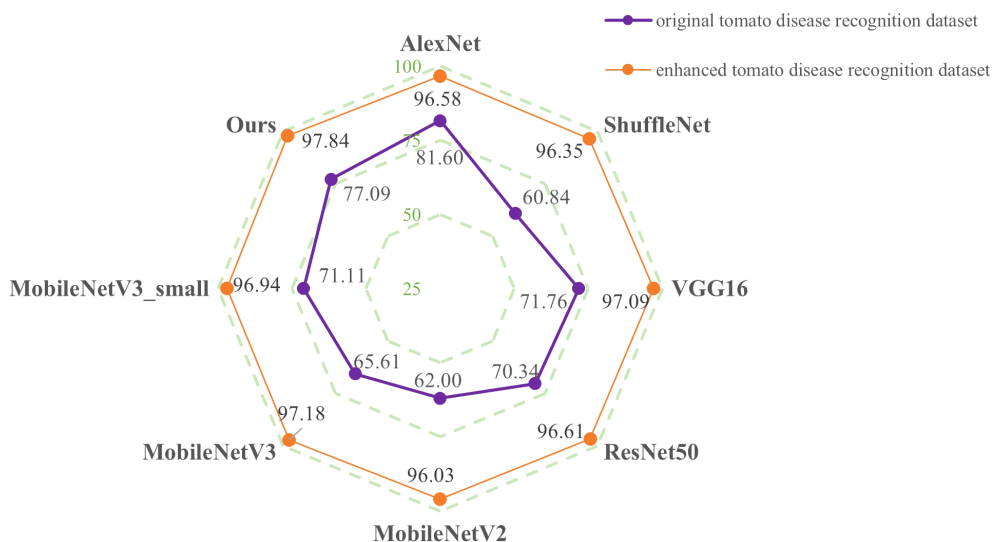


FIGURE 11  
Comparison of the accuracies of different models on the enhanced testing set.

TABLE 3 Test results obtained after training on the original tomato disease recognition dataset.

	AlexNet	ShuffleNet	VGG16	ResNet50	MobileNet			Ours
					V2	V3 large	V3 small	
Original image	94.58%	90.32%	91.24%	93.83%	85.24%	93.08%	92.16%	95.25%
Vertical flip	94.91%	90.53%	91.49%	93.08%	84.99%	93.66%	92.74%	96.08%
Horizontal flip	88.66%	82.35%	83.32%	87.24%	81.90%	90.16%	89.07%	91.16%
Brightness increase	68.56%	28.73%	50.21%	33.36%	31.61%	29.27%	48.96%	50.71%
Brightness reduction	35.70%	20.56%	50.29%	20.52%	21.52%	21.68%	40.95%	57.38%
Chroma increase	92.91%	73.13%	83.24%	83.32%	75.56%	81.73%	82.74%	87.57%
Chroma reduction	74.65%	40.12%	43.79%	72.39%	43.79%	39.12%	53.88%	58.97%
Contrast increase	91.58%	68.83%	81.32%	69.89%	72.98%	78.73%	79.32%	84.99%
Contrast reduction	79.32%	38.71%	42.95%	31.94%	46.62%	26.77%	29.86%	51.54%
Sharpness increase	94.50%	84.19%	91.08%	93.83%	85.49%	92.91%	92.08%	94.91%
Sharpness reduction	94.50%	84.36%	91.24%	93.91%	85.24%	93.08%	92.16%	95.08%
Laplace transform	94.33%	83.53%	90.16%	93.83%	85.32%	92.91%	91.91%	95.25%
Gamma transform	78.07%	47.39%	78.73%	62.05%	46.54%	54.88%	65.72%	67.31%
CLAHE	52.54%	19.02%	35.53%	55.55%	21.18%	30.61%	43.95%	53.13%
Average accuracy	81.06%	60.84%	71.76%	70.34%	62.00%	65.61%	71.11%	77.09%

weight of the model proposed in this paper is only 1.6 MB, and the number of parameters is 0.418 M, which is only one-tenth of that required by MobileNetV3. Although the tomato disease recognition accuracy listed in Table 3 is slightly lower than that of AlexNet, the

number of model parameters is reduced by 200 times. After training on the enhanced tomato disease recognition dataset in Table 4, the accuracy of the model in this paper ranks first, and the number of parameters is much smaller than that of VGG16 and ResNet50.

TABLE 4 Test results obtained after training on the enhanced tomato disease recognition dataset.

	AlexNet	ShuffleNet	VGG16	ResNet50	MobileNet			Ours
					V2	V3 large	V3 small	
Original image	97.41%	97.08%	97.58%	97.58%	97.00%	98.25%	97.83%	98.58%
Vertical flip	97.00%	97.12%	97.50%	97.08%	97.25%	97.91%	97.83%	98.33%
Horizontal flip	96.25%	95.63%	96.83%	95.91%	95.75%	97.08%	96.25%	97.58%
Brightness increase	94.66%	94.36%	95.16%	93.49%	92.66%	94.75%	94.16%	96.33%
Brightness reduction	97.08%	96.15%	97.16%	95.91%	95.58%	97.08%	96.41%	97.75%
Chroma increase	96.66%	96.02%	97.33%	97.25%	96.00%	97.50%	97.33%	97.50%
Chroma reduction	96.25%	95.94%	97.16%	96.25%	96.16%	96.83%	96.91%	98.17%
Contrast increase	96.58%	96.28%	97.16%	96.91%	96.00%	97.08%	97.08%	97.66%
Contrast reduction	96.91%	96.59%	97.08%	96.66%	96.25%	96.33%	96.33%	97.75%
Sharpness increase	97.41%	97.12%	97.58%	97.66%	96.91%	98.25%	97.75%	98.58%
Sharpness reduction	97.41%	97.16%	97.58%	97.58%	97.00%	98.25%	97.75%	98.58%
Laplace transform	97.25%	96.98%	97.50%	97.66%	97.00%	98.17%	97.66%	98.42%
Gamma transform	96.08%	96.63%	97.58%	96.41%	95.58%	96.66%	96.83%	97.08%
CLAHE	95.16%	95.86%	96.08%	96.16%	95.25%	96.33%	97.08%	97.41%
Average accuracy	96.58%	96.35%	97.09%	96.61%	96.03%	97.18%	96.94%	97.84%

Compared with the lightweight MobileNet series models proposed in recent years, the model proposed in this paper not only has a higher recognition accuracy but also has a lower model weight. For example, the weight of the model developed in this paper is one-fifth of that of MobileNetV2, but the test results in [Table 3](#) and [Table 4](#) show that the recognition accuracies of the model in this paper are 15.09% and 1.44% higher than those of MobileNetV2. Experiments show that the ultra-lightweight tomato disease recognition model based on the ECA module proposed in this paper balances the contradiction between tomato disease recognition accuracy and model complexity and uses the ultra-lightweight network to obtain disease recognition accuracy.

### 4.3.3 Results of an ablation experiment

In this paper, an ablation experiment is designed to further explore the effect of the high-efficiency channel attention mechanism and dropout layer on the obtained tomato disease recognition accuracy. After the ECA mechanisms in all modules are removed, the formed tomato disease recognition model is called No\_ECA. After the last dropout layer in the model proposed in this paper is removed, the resulting tomato disease recognition model is called No\_Dropout. These two models are trained on the enhanced tomato disease recognition dataset, and the models are tested according to the test method described in section 4.3.2. The test results are shown in [Table 6](#).

According to the results of the ablation experiment, when using the same dataset and training configuration to train No\_ECA and No\_Dropout, their recognition accuracies are lower than that of the model proposed in this paper. Although a large number of ECA

modules are used in the model proposed in this paper to improve the accuracy of tomato disease recognition, the number of model parameters increases by 33, which is only a slight increase. This proves that our model not only improves the accuracy of tomato disease recognition but also realizes ultralightweight characteristics. It balances the contradiction between model accuracy and complexity and has strong robustness to complex and unconstrained environments.

## 5 Discussion and conclusion

This paper addresses the limitations of existing tomato disease recognition models, which are typically capable of identifying only a single diseased leaf and struggle to reconcile the conflicting demands of accuracy and complexity. To overcome these challenges, we propose a comprehensive framework that integrates tomato leaf detection with leaf disease recognition. Initially, we develop a tomato leaf detection model designed to extract individual diseased leaves from tomato disease images captured in real-world, unconstrained environments. Subsequently, we introduce an effective data enhancement method that simulates image noise across various settings, thereby augmenting the model's robustness in complex environments. Finally, we present an ultra-lightweight tomato leaf disease recognition model tailored for operation in such challenging conditions. This model leverages a high-efficiency channel attention mechanism and incorporates five inverted residual modules to enhance accuracy while minimizing the number of parameters,

TABLE 5 Model size comparison and evaluation.

	AlexNet	ShuffleNet	VGG16	ResNet50	MobileNet			Ours
					V2	V3 large	V3 small	
Model weight	217MB	7.6MB	268MB	89.9MB	8.69MB	16.2MB	2.32MB	1.68MB
Parameter	57M	1.98M	70M	23M	2.23M	4.2M	0.588M	0.418M
Test results in <a href="#">Table 3</a>	81.06%	60.84%	71.76%	70.34%	62.00%	65.61%	71.11%	77.09%
Test results in <a href="#">Table 4</a>	96.58%	96.35%	97.09%	96.61%	96.03%	97.18%	96.94%	97.84%

TABLE 6 Comparison of the ablation experiment results for the model proposed in this paper.

	Original image	Vertical flip	Horizontal flip	Brightness increase	Brightness reduction	Chroma increase	Chroma reduction	Contrast increase
No_ECA	97.50%	97.66%	96.16%	94.91%	96.75%	97.33%	97.16%	97.41%
No_Dropout	97.83%	97.50%	97.25%	95.50%	97.16%	97.91%	97.83%	97.00%
Ours	98.58%	98.33%	97.58%	96.33%	97.75%	97.50%	98.17%	97.66%
	Contrast reduction	Sharpness increase	Sharpness reduction	Laplace Transform	Gamma transform	CLAHE	Average accuracy	Parameter
No_ECA	97.00%	97.58%	97.50%	97.66%	96.75%	96.41%	96.99%	418148
No_Dropout	97.58%	97.91%	97.83%	97.91%	97.25%	97.16%	97.40%	418181
Ours	97.75%	98.58%	98.58%	98.42%	97.08%	97.41%	97.84%	418181

effectively balancing model complexity with recognition accuracy and reducing hardware resource consumption. We utilize both the original tomato disease dataset and an enhanced version to train the model, subsequently evaluating its recognition accuracy across 14 distinct noise environments. Experimental results reveal that our ultra-lightweight tomato disease recognition model, based on high-efficiency channel attention, achieves an accuracy of 97.84% while maintaining only 0.418 million parameters. In comparison to traditional image recognition models such as AlexNet, VGG16, ResNet50, and the MobileNet series, the proposed model not only demonstrates superior accuracy across the 14 noisy environments but also significantly reduces the parameter count. Additionally, our model addresses the limitation of existing systems that can only recognize a single disease image, making it suitable for deployment on mobile devices for practical demonstration applications. At the same time, based on the framework proposed in this study, using disease images of other crops for model fine-tuning can also be applied to other crops. Although the existing models have achieved good results, there is still room for improvement in future research. Firstly, the existing image data mainly simulates complex environments through different data processing methods. In the future, more image acquisition devices can be used to obtain image data under different performance devices, so that the images are closer to the natural environment. At the same time, more complex leaf disease images can be collected to establish a more balanced leaf disease dataset. Secondly, based on this model, techniques such as compression pruning, knowledge distillation, and quantification can be used to further improve the lightweight process of the model.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## References

- Bashish, D. A., Braik, M., and Bani-Ahmad, S. (2011). Detection and classification of leaf diseases using k-means-based segmentation and neural-networks-based classification. *Inf. Technol. J.* 10, 267–275. doi: 10.3923/itj.2011.267.275
- Bochkovskiy, A., Wang, C. Y., and Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv*, 1–17. doi: 10.48550/arXiv.2004.10934
- Gao, Z., Xie, J., Wang, Q., and Li, P. (2018). Global second-order pooling convolutional networks. *arXiv preprint arXiv*, 1–9. doi: 10.48550/arXiv.1811.12006
- Girshick, R. (2015). Fast R-CNN. *arXiv preprint arXiv*, 1–9. doi: 10.48550/arXiv.1504.08083
- Gulhane, M., and Gurjar, A. A. (2011). Detection of diseases on cotton leaves and its possible diagnosis. *Int. J. Image Process.* 5, 590–598.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. *arXiv preprint arXiv*, 1–12. doi: 10.48550/arXiv.1512.03385
- Howard, A., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., et al. (2019). Searching for mobilenetv3. *arXiv preprint arXiv*, 1–11. doi: 10.48550/arXiv.1905.02244
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., et al. (2017). Mobilenets: efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv*, 1–19. doi: 10.48550/arXiv.1704.04861
- Hu, J., Shen, L., Albanie, S., Sun, G., and Vedaldi, A. (2019). “Gather-excite: Exploiting feature context in convolutional neural networks,” in *NeurIPS*.
- Huang, S., Liu, W., Qi, F., and Yang, K. (2019). “Development and validation of a deep learning algorithm for the recognition of plant disease,” in *2019 IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*. IEEE, 1951–1957. doi: 10.1109/HPCC/SmartCity/DSS.2019.00269
- Jiang, D., Li, F., Yang, Y., and Yu, S. (2020). “A tomato leaf diseases classification method based on deep learning,” in *2020 Chinese Control And Decision Conference (CCDC)*. (Singapore: Springer) 594, 1446–1450. doi: 10.1109/CCDC49329.2020.9164457
- Jie, H., Li, S., Gang, S., and Albanie, S. (2017). Squeeze-and-excitation networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 42, 2011–2023. doi: 10.1109/TPAMI.2019.2913372
- Keswani, B., Mohapatra, G. A., Keswani, P., Khanna, A., Gupta, D., and Rodrigues, J. (2020). Improving weather dependent zone specific irrigation control scheme in IoT and big data enabled self driven precision agriculture mechanism. *Enterprise Inf. Syst.* 14, 1494–1515. doi: 10.1080/17517575.2020.1713406
- Keswani, B., Mohapatra, A. G., Mohanty, A., Khanna, A., Rodrigues, J., Gupta, D., et al. (2018). Adapting weather conditions based iot enabled smart irrigation technique in precision agriculture mechanisms. *Neural Computing Appl.* 31, 277–292. doi: 10.1007/s00521-018-3737-1
- Krizhevsky, A., Sutskever, I., and Hinton, G. (2012). Imagenet classification with deep convolutional neural networks. *Commun. ACM* 60 (6), 84–90. doi: 10.1145/3065386

## Author contributions

WS: Conceptualization, Methodology, Visualization, Writing – original draft, Writing – review & editing. ZX: Data curation, Visualization, Writing – review & editing. KX: Data curation, Methodology, Writing – review & editing. LR: Data curation, Writing – review & editing. RY: Conceptualization, Data curation, Methodology, Project administration, Writing – review & editing. RW: Data curation, Methodology, Writing – review & editing. JX: Methodology, Supervision, Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was supported by National Talent Foundation Project of China (Grant No. T2019136).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., et al. (2022). YOLOv6: A single-stage object detection framework for industrial applications. *arXiv preprint arXiv*, 1–17. doi: 10.48550/arXiv.2209.02976
- Li, Z., Tao, W., Liu, J., Zhu, F., Du, G., and Ji, G. (2023). Tomato leaf disease recognition via optimizing deep learning methods considering global pixel value distribution. *Horticulturae* 9, 1034. doi: 10.3390/horticulturae9091034
- Liang, S., and Zhang, W. (2020). “Accurate image recognition of plant diseases based on multiple classifiers integration,” in *Proceedings of 2019 Chinese Intelligent Systems Conference. CISC 2019. Lecture Notes in Electrical Engineering*, Vol. 594, 103–113. doi: 10.1007/978-981-32-9698-5\_13
- Liu, B., Tan, C., Li, S., He, J., and Wang, H. (2020). A data augmentation method based on generative adversarial networks for grape leaf disease identification. *IEEE Access* 8, 102188–102198. doi: 10.1109/ACCESS.2020.2998839
- Lv, M., Zhou, G., He, M., Chen, A., and Hu, Y. (2020). Maize leaf disease identification based on feature enhancement and dms-robust alexnet. *IEEE Access* 8, 57952–57966. doi: 10.1109/ACCESS.2020.2982443
- Dyrmann, M., Karstoft, H., and Midtby, H. S. (2016). Plant species classification using deep convolutional neural network. *Biosyst. Eng.* 151, 72–80. doi: 10.1016/j.biosystemseng.2016.08.024
- Mohanty, S. P., Hughes, D. P., and Salathe, M. (2016). Using deep learning for image-based plant disease detection. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01419
- Ni, J., Zhou, Z., Zhao, Y., Han, Z., and Zhao, L. (2023). Tomato leaf disease recognition based on improved convolutional neural network with attention mechanism. *Plant Pathol.* 72, 1335–1344. doi: 10.1111/ppa.13745
- Qin, F., Liu, D., Sun, B., Liu, R., Ma, Z., Wang, H., et al. (2016). Identification of alfalfa leaf diseases using image recognition technology. *PLoS One* 11, e0168274. doi: 10.1371/journal.pone.0168274
- Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). “You only look once: unified, real-time object detection,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. (IEEE), 779–788. doi: 10.1109/CVPR.2016.91
- Redmon, J., and Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv*, 1–6. doi: 10.48550/arXiv.1804.02767
- Ren, S., He, K., Girshick, R., and Sun, J. (2017). Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 1137–1149. doi: 10.1109/TPAMI.2016.2577031
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L. C. (2018). “Inverted residuals and linear bottlenecks: mobile networks for classification, detection and segmentation,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. p. 4510–4520. doi: 10.1109/CVPR.2018.00474
- Sanida, T., and Dasygenis, M. (2024). MiniTomatoNet: a lightweight CNN for tomato leaf disease recognition on heterogeneous FPGA-SoC. *J. Supercomput* 80, 21837–21866. doi: 10.1007/s11227-024-06301-8
- Sannakki, S. S., Rajpurohit, V. S., Nargund, V. B., Kumar, A., and Yallur, P. S. (2010). Leaf disease grading by machine vision and fuzzy logic. *Int. J. Comput. Technol. Appl.* 2, 1709–1716.
- Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv*, 1–14. doi: 10.48550/arXiv.1409.1556
- Siricharoen, P., Scotney, B., Morrow, P., and Parr, G. (2016). A lightweight mobile system for crop disease diagnosis. In *Image Analysis and Recognition: 13th International Conference, ICIAR 2016, in Memory of Mohamed Kamel Póvoa de Varzim, Portugal: Springer International Publishing*. p. 7893–791. doi: 10.1007/978-3-319-41501-7\_87
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., and Rabinovich, A. (2014). “Going deeper with convolutions. IEEE Computer Society,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. (IEEE), 1–9. doi: 10.1109/CVPR.2015.7298594
- Wang, C. Y., Bochkovskiy, A., and Liao, H. Y. M. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv preprint arXiv*, 1–15. doi: 10.48550/arXiv.2207.02696
- Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., et al. (2024a). Yolov10: Real-time end-to-end object detection. *arXiv preprint arXiv*, 1–21. doi: 10.48550/arXiv.2405.14458
- Wang, C. Y., Yeh, I. H., and Liao, H. Y. M. (2024b). YOLOv9: Learning what you want to learn using programmable gradient information. *arXiv preprint arXiv*, 1–18. doi: 10.48550/arXiv.2402.13616
- Woo, S., Park, J., Lee, J. Y., and Kweon, I. S. (2018). “CBAM: convolutional block attention module,” in *Computer Vision—ECCV 2018: 15th European Conference*. (Munich, Germany: Springer International Publishing), 3–19. doi: 10.1007/978-3-030-01234-2\_1
- Xie, X., Zhang, X., Bing, H., Dong, L., and Huang, L. (2016). “A system for diagnosis of wheat leaf diseases based on Android smartphone,” in *Proc. SPIE 10155, Optical Measurement Technology and Instrumentation*, 1015526. doi: 10.1117/12.2246919
- Xu, K., Sun, W., Chen, D., Qing, Y., Xing, J., and Yang, R. (2024). Early sweet potato plant detection method based on YOLOv8s(ESPPD-YOLO): A model for early sweet potato plant detection in a complex field environment. *Agronomy* 14, 2650. doi: 10.3390/agronomy14112650
- Yang, S., Zhang, L., Lin, J., Cernava, T., Cai, J., Pan, R., et al. (2024). LSGNet: A lightweight convolutional neural network model for tomato disease identification. *Crop Prot.* 182, 106715. doi: 10.1016/j.cropro.2024.106715