A lightweight convolutional neural network for tea leaf disease and pest recognition

Xiaojie Wen[1,2†], Qi Liu[1,2†], Xuanyuan Tang[1,2], Fusheng Yu[1,2] and Jing Chen[1,2*]

Abstract

The tea industry plays a vital role in China's green economy. Tea trees (*Melaleuca alternifolia*) are susceptible to numerous diseases and pest threats, making timely pathogen detection and precise pest identification critical requirements for agricultural productivity. Current diagnostic limitations primarily arise from data scarcity and insufficient discriminative feature representation in existing datasets. This study presents a new tea disease and pest dataset (TDPD, 23-class taxonomy). Five lightweight convolutional neural networks (LCNNs) were systematically evaluated through two optimizers, three learning rate configurations and six distinct scheduling strategies. Additionally, an enhanced MnasNet variant was developed through the integration of SimAM attention mechanisms, which improved feature discriminability and increased the accuracy of tea leaf disease and pest classification. Model validation employs both our proprietary TDPD dataset and an open-access dataset, with performance evaluation metrics including average accuracy, F1 score, recall, and parameter size. The experimental results demonstrated the superior classification performance of the model, which achieved accuracies of 98.03% based on TDPD and 84.58% based on the public dataset. This research outlines an effective paradigm for automated tea disease and pest detection, with direct applications in precision agriculture through integration with UAV-mounted imaging systems and mobile diagnostic platforms. This study provides practical implementation pathways for intelligent tea plantation management.

Keywords Tea, Diseases and pests, Lightweight convolutional neural networks (LCNNs), Attention mechanism, MnasNet, SimAM

[†]Xiaojie Wen and Qi Liu are contributed equally to this work.

*Correspondence:
Jing Chen
chenj@xjau.edu.cn
[1]Key Laboratory of Pest Monitoring and Safety Control of Crops and Forests of the Xinjiang Uygur Autonomous Region, College of Agronomy, Xinjiang Agricultural University, Urumqi 830052, China
[2]Key Laboratory of Prevention and Control of Invasive Alien Species in Agriculture and Forestry of the North-Western Desert Oasis, Ministry of Agriculture and Rural Affairs, Urumqi 830052, China

Introduction

China is the world's leading producer and consumer of tea. As of 2020, tea plantations covered 3.17 million hectares, representing 62.10% of the global cultivation area and 47.60% of the worldwide production output [1]. Yunnan Province constituted the majority of the plantation area in China. In 2023, Yunnan Province alone covered 535,000 hectares, representing approximately one-sixth of the national coverage area and generating an output valuated at ¥150.42 billion, solidifying its position as China's primary tea-producing region [2]. This economic prominence underscores the critical role of tea cultivation in Yunnan Province and rural development.

Wen *et al. Plant Methods*   (2025) 21:129

Page 2 of 12

Increasing climate variability has introduced unprecedented challenges to tea ecosystems. Rising temperatures and irregular precipitation patterns have increased pest proliferation rates and disease susceptibility in tea tree cultivation, directly threatening yield stability [3].

Conventional diagnostic methods relying on visual symptomology and morphological analysis exhibit critical limitations, including operator subjectivity, spatial–temporal constraints, and inefficiency in large-scale plantations [4]. These methodological shortcomings impede rapid response mechanisms, which are essential for modern precision agriculture.

To address these limitations, researchers have focused on the use of computer vision techniques and machine learning algorithms for the onsite diagnosis of tea diseases and pests, facilitating the development of timely countermeasures. In the field of machine learning, deep learning methods have become a focal point of computer image recognition. Convolutional neural networks (CNNs), recognized for hierarchical feature learning abilities and translation invariance [5–8], have demonstrated exceptional performance in crop disease and pest recognition through automated feature extraction [9–11]. Li et al. developed a hybrid framework combining Mask R-CNN with wavelet transforms and F-RNet, achieving a detection rate of 98.7% [12]. The innovative architecture was AX-RetinaNet, which integrates multiscale feature fusion (X-module) and channel attention mechanisms, achieving mean average precision (mAP) of 93.83% and an F1 score of 0.954, surpassing conventional models such as SSD, RetinaNet, YOLO-v3, YOLO-v4, Center-Net, M2det, and EfficientNet through adaptive feature weighting. Chen et al. constructed TeaViTNet, a hybrid transformer-CNN model incorporating EMA-PANet and RFBNet modules, which achieved an accuracy of 89.1% in multiclass pest/disease differentiation [13].

Despite advances in computer vision techniques, critical bottlenecks persist in CNN deployment for disease and pest recognition. For example, dataset quality, taxonomic diversity, and annotation consistency constrain model efficacy. Second, deep networks necessitate extensive computational resources and risk overfitting without sufficient training data. However, collecting images of diseases and pests remains a significant challenge. Finally, the structural design of CNNs determines their ability to extract relevant data features. A deeper network architecture can facilitate higher-level feature extraction but requires larger datasets and substantial computational resources, whose models are prone to gradient dynamics (e.g., vanishing/exploding gradients) and hyperparameter sensitivity (e.g., learning rate scheduling, optimizer selection), which critically govern model convergence and generalizability.

The training method determines the model's ability to effectively learn data features. The optimizer controls the process of updating the model's weights. Different optimizers (e.g., stochastic gradient descent (SGD) and adaptive moment estimation (Adam)) employ distinct update strategies, which impact the model's convergence speed and accuracy. The learning rate is one of the most critical hyperparameters in training, as it determines how updates are applied and influences both the model's convergence speed and accuracy. An appropriate learning rate facilitates rapid convergence to the optimal solution, whereas an inappropriate learning rate can result in training failure. Consequently, the choice of optimizer and learning rate plays a pivotal role in determining the stability and convergence rate of model training. Thus, developing efficient CNNs remains a pressing challenge.

On the basis of prior research and addressing current challenges in using CNNs to identify tea leaf diseases and pests, this study introduces a lightweight CNN framework optimized for tea disease and pest recognition under data-constrained conditions. The contributions of this study are summarized as follows:

a. A tea disease and pest dataset (TDPD) was established, comprising 22 taxonomic classes (7 diseases, 14 pests and healthy tea leaf images). All images were captured in the natural environment of a tea garden and augmented via geometric/radiometric transformations to mitigate overfitting.
b. MnasNet-SimAM integrates SimAM attention mechanisms into three terminal inverted residual blocks and is trained with transfer learning, enabling dual disease-pest recognition without structural redundancy.
c. The generalizability of MnasNet-SimAM was evaluated based on public datasets through multiclass metrics such as average accuracy, F1 score and parameter size.

Materials and methods
Image acquisition and preprocessing
Field imagery of tea diseases and pests was acquired across tea plantations in Yunnan Province using Cannon PowerShot G12 digital camera under natural illumination conditions. The initial dataset comprised 3678 annotated images (2388 pest specimens spanning 14 taxa; 1290 disease samples across 7 pathologies) validated by entomologists and phytopathologists. To address class imbalance and dataset limitations, a multimodal augmentation approach was implemented. This approach included geometric transformations, such as random rotations (90°, 180°, and 270°) and horizontal/vertical flipping and radiometric adjustments, such as brightness variance and Gaussian blur. Post augmentation, the tea disease

Wen *et al. Plant Methods*   (2025) 21:129

Page 3 of 12

and pest dataset (TDPD) contained 20,854 images, which were partitioned into training (70%), validation (20%), and test (10%) sets via stratified sampling. All the images were normalized and scaled to 224×224 pixels (Table 1 and Fig. 1).

Lightweight convolutional neural networks

Five state-of-the-art lightweight convolutional neural networks (LCNNs) were evaluated for computational efficiency and discriminative capacity:

EfficientNet This network employs compound scaling to balance model depth, width, and resolution, thereby improving performance without substantially increasing computational cost [14]. This scaling methodology mitigates the resource inefficiencies encountered when these dimensions are adjusted separately in traditional models. The EfficientNet series includes EfficientNetV2, which incorporates Fused-MBConv into the search space and integrates an adaptive mechanism for regularization strength adjustment. These enhancements collectively result in superior performance and expedited training times [15].

MnasNet Mobile neural architecture search is a deep learning model optimized for mobile devices. At its core, it adopts reinforcement learning techniques to explore and identify optimal neural network architectures auton-omously within an extensive search space. It incorporates latency constraints during the search process to ensure that the identified architecture operates within the hardware limitations of mobile devices [16].

GhostNet GhostNet includes an innovative ghost module, which is capable of generating numerous "ghost" feature maps through linear transformations of a portion of the original feature map [17]. This approach captures information akin to the original feature map, thereby reducing computational complexity. The network architecture of GhostNet is constructed by stacking multiple ghost modules The model is lightweight and achieved high performance. Updates to the GhostNet family, such as GhostNetv2 [18] and GhostNetv3 [19], continue to be released to further improve the model's performance and efficiency.

MobileNet The core innovation of MobileNet is the introduction of depthwise separable convolution. Conventional convolution is divided into depthwise convolution and pointwise convolution. Depthwise convolution employs a distinct convolution kernel for each input channel, whereas pointwise convolution uses 1×1 convolution kernels to integrate the outputs of the depthwise convolution. This approach significantly reduces the computational complexity and parameter count while maintaining effective classification performance. The MobileNet family has evolved into several versions, including Mobile-

Table 1 TDPD comprising seven diseases and 14 pests used in this study

Number	Type	Original	Training	Validation	Testing
1	White spot	108	453	130	65
2	Bird's eye spot	108	453	130	65
3	Red leaf spot	107	449	128	65
4	Gray blight	115	482	138	70
5	Anthracnose	155	651	1855	94
6	Brown blight	129	541	155	78
7	Algal leaf spot	72	302	86	44
8	Healthy	496	1736	496	248
9	*Spilarctia subcarnea* Walker	199	696	199	100
10	*Rikiosatoa vandervoordeni* Prout	351	1228	351	176
11	*Arctonis alba* Bremer	35	122	35	18
12	*Eterusia aedea* Linnaeus	43	150	43	22
13	*Euproctis pseudoconspersa* Strand	394	1379	393	198
14	*Homona coffearia* Meyrick	148	518	147	75
15	*Culcula panterinaria* Bremer et Gray	74	259	73	38
16	*Ricania speculum* Walker	160	560	159	81
17	*Euricania ocellus* Walker	76	266	75	39
18	*Amata germane* Felder	52	182	51	27
19	*Scopula subpunctaria* Herrich-Schaeffer	236	826	236	118
20	*Ricania sublimbata* Jacobi	397	1389	397	199
21	*Ceratonoros transiens* Walker	160	560	159	81
22	*Spilosoma menthastri* Esper	63	220	63	32
	Total	3678	13,422	5499	1933

Fig. 1 Typical example images of tea diseases and pests used in the study

NetV2 [20] and MobileNetV3. MobileNetV2 introduces inverted residual blocks and linear bottlenecks to improve the model's expressive capacity and feature reuse capabilities. MobileNetV3 employs a neural architecture search (NAS) approach to determine the network structure and introduces the squeeze-and-excitation (SE) module to improve the model's channel feature representation capabilities. These advancements enable the MobileNet family to further improve model performance and increase accuracy while remaining lightweight.

ShuffleNet The primary innovation of ShuffleNet is the introduction of channel shuffling and group convolution techniques to achieve efficient feature representation [21].

In pointwise group convolution, feature maps are divided into distinct groups, and each group convolves separately, thereby reducing the computational load and model training parameters. Channel shuffling addresses the issue of information noncommunication between groups following group convolution. By rearranging the channel order, channel shuffling facilitates information interaction within groups, thereby mitigating information and precision loss. The ShuffleNet family includes ShuffleNetV2 [22], which introduces four lightweight network design principles, channel width balance, moderate group convolution, network fragmentation reduction, and consideration of the impact of elementwise operations, to further increase network efficiency and accuracy.

SimAM attention mechanism integration

The attention mechanism represents a significant breakthrough in deep learning, particularly in natural language processing and computer vision. Its fundamental principle involves mimicking human attention: focusing on the most relevant parts of extensive information to improve model performance. Traditional deep learning models often process all parts of an input date with the same weight. In contrast, the attention mechanism assigns varying weights to different components, allowing the model to dynamically prioritize the most relevant information and then improving both the performance and efficiency of the models.

To identify tea leaf diseases and pests, the model must first distinguish between them and then differentiate specific types within these categories. The attention mechanism aids in capturing subtle distinctions. This integration helps mitigate data imbalance, highlight key features, and minimize noise. Although attention mechanisms are essential in CNNs for directing models toward key regions, they typically require additional parameters, which increase model complexity and computational cost. The introduction of parameter-free attention mechanisms addresses this challenge. Yang et al. proposed a straightforward parameter-free attention mechanism, which is distinct from conventional channel and spatial attention mechanisms [23]. Three-dimensional attention weights are derived within this module for feature maps without introducing additional parameters. This approach relies on the local similarity of images, wherein adjacent pixels in lower-level images exhibit strong similarities, whereas distant pixels exhibit weaker correlations.

In this study, the parameter-free attention mechanism SimAM was embedded into MnasNet's terminal three inverted residual blocks without modifying its backbone structure and trained with pretrained weights based on ImageNet. The modified MnasNet-SimAM is illustrated in Fig. 2. MnasNet-SimAM was subsequently trained based on the TDPD, and its performance was assessed. Furthermore, MnasNet was optimized with alternative attention mechanisms, such as the SE, CA, CBAM, and ECA modules, and the resulting performance was compared.

Evaluation metrics

Model performance was quantified using formulas [24]. Per-class metrics were computed from confusion matrices. Statistical significance was assessed using analysis of variance ($\alpha = 0.05$).

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \quad (1)$$

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \quad (2)$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) \quad (3)$$

$$\text{F1} - \text{score} = 2 / [(1/\text{Recall}) + (1/\text{Precision})] \quad (4)$$
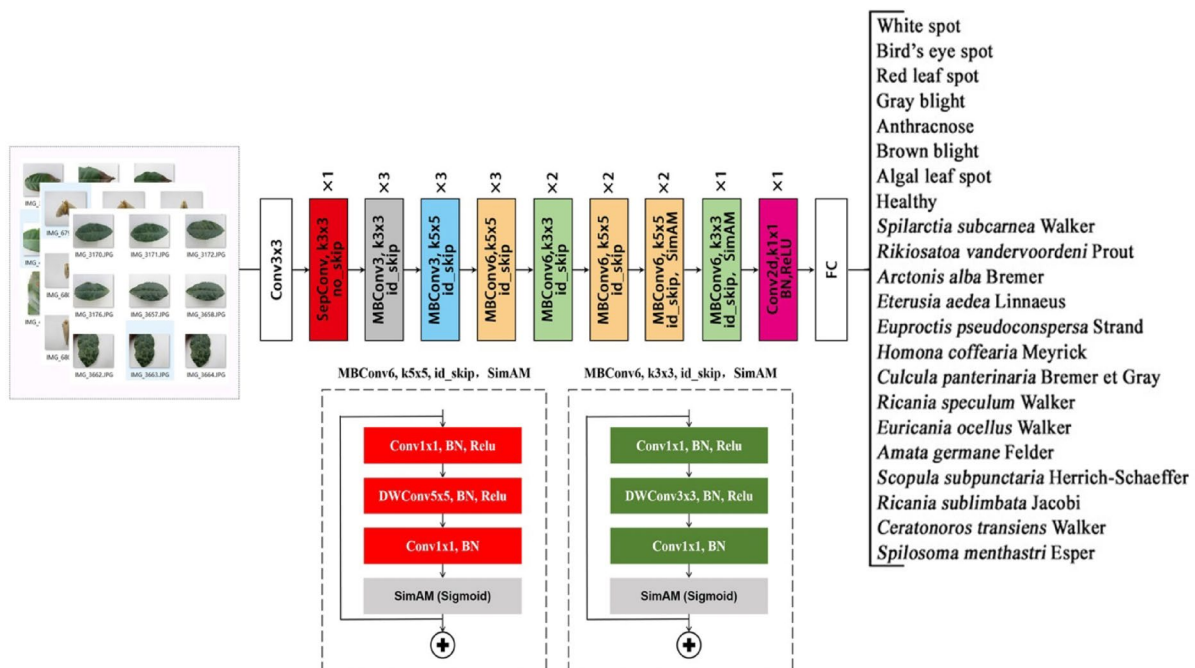

Fig. 2 Architecture of the modified MnasNet-SimAM

where TP stands for true positive samples, FP for false positive samples, TN for true negative samples, and FN for false negative samples.

Implementation environment and training protocol

The models were trained on a Linux operating system and an NVIDIA RTX 3080 Ti (12 GB) using PyTorch 2.0 with CUDA 11.8 acceleration. Optimization was performed using either Adam or SGD (momentum = 0.9); the learning rates were as follows: warm up (0.01), cosine annealing (10^{-5}), and weight decay ($5e^{-4}$); and the batch configuration was 32 samples/batch, 50 epochs.

Results

Five LCNNs comparative performance

In this study, different combinations of initial learning rates and optimizers were used to compare their effects on five LCNNs, with the aim of selecting the optimal

initial learning rate and optimizer. The accuracy and loss values were recorded, and the top three weights with the highest accuracy based on the validation set were selected for calculating the mean accuracy and standard error, as shown in Table 2. After the appropriate initial learning rate and optimizer were selected, different training strategies were compared, and the accuracy and loss values were recorded. The results are shown in Table 3.

The classification accuracy and loss values of the five LCNNs after 10 training epochs are shown in Fig. 3. Both the accuracy and loss values stabilized, and the validation set accuracy exceeded 98%, demonstrating strong training performance. Tables 2 and 3 show that at a significance level of $p < 0.05$, variations in initial learning rates and optimizers significantly influence model accuracy, while different learning rate scheduling strategies also have notable impacts. Consequently, appropriate initial learning rates, optimizers, and training strategies

Table 2 ACC and LOSS under different initial learning rates and optimizers

Model	Optimizer	Initial learning rate	Average train accuracy (%)	Average verification accuracy (%)	Average test accuracy (%)	Maximum test accuracy (%)
EfficientNetV2s	SGD	0.01	99.66 ± 0.01	99.57 ± 0.01	99.07 ± 0.08 a	99.22
		0.001	99.70 ± 0.01	99.53 ± 0.00	98.83 ± 0.02 a	98.86
		0.0001	99.27 ± 0.02	99.38 ± 0.01	98.83 ± 0.02 a	98.86
	Adam	0.01	97.23 ± 0.21	95.85 ± 0.12	39.88 ± 1.22 b	42.16
		0.001	99.45 ± 0.05	98.91 ± 0.04	90.62 ± 0.97 a	92.40
		0.0001	99.59 ± 0.01	99.57 ± 0.01	98.60 ± 0.03 a	98.65
MnasNet	SGD	0.01	99.71 ± 0.02	99.39 ± 0.01	96.36 ± 0.12 a	96.59
		0.001	99.70 ± 0.01	99.21 ± 0.02	94.64 ± 0.12 b	94.88
		0.0001	98.67 ± 0.04	98.82 ± 0.08	87.67 ± 0.12 c	87.89
	Adam	0.01	99.65 ± 0.01	99.25 ± 0.06	97.15 ± 0.18 a	97.36
		0.001	99.69 ± 0.02	99.33 ± 0.05	96.62 ± 0.05 a	96.69
		0.0001	95.05 ± 0.25	96.04 ± 0.14	81.24 ± 0.07 d	81.38
GhostNet	SGD	0.01	99.69 ± 0.01	99.48 ± 0.02	95.29 ± 0.18 a	95.50
		0.001	99.61 ± 0.00	99.41 ± 0.03	77.82 ± 0.12 b	78.01
		0.0001	98.07 ± 0.06	98.84 ± 0.02	52.97 ± 0.39 c	53.75
	Adam	0.01	96.95 ± 0.22	97.15 ± 0.11	52.72 ± 1.82 c	55.77
		0.001	99.21 ± 0.03	99.27 ± 0.07	94.02 ± 0.45 a	94.88
		0.0001	99.63 ± 0.01	99.47 ± 0.03	54.08 ± 2.04 c	57.42
MobileNetV3	SGD	0.01	99.71 ± 0.01	99.45 ± 0.00	94.97 ± 0.20 a	95.24
		0.001	99.69 ± 0.01	99.27 ± 0.01	86.86 ± 0.17 c	87.17
		0.0001	98.47 ± 0.07	98.59 ± 0.02	80.29 ± 0.11 d	80.44
	Adam	0.01	94.19 ± 0.13	89.51 ± 0.64	62.41 ± 0.87 e	64.05
		0.001	99.24 ± 0.03	99.07 ± 0.03	90.52 ± 0.24 b	91.00
		0.0001	99.64 ± 0.01	99.42 ± 0.02	91.10 ± 0.51 b	91.83
ShuffleNetV2	SGD	0.01	99.71 ± 0.02	99.41 ± 0.02	91.17 ± 2.30 a	95.76
		0.001	99.69 ± 0.00	99.32 ± 0.00	85.08 ± 0.92 cd	86.91
		0.0001	98.79 ± 0.02	98.98 ± 0.03	79.48 ± 0.36 bc	80.08
	Adam	0.01	98.81 ± 0.03	97.72 ± 0.12	57.54 ± 0.87 d	58.82
		0.001	99.30 ± 0.02	99.16 ± 0.02	74.91 ± 1.11 c	76.41
		0.0001	99.66 ± 0.00	99.48 ± 0.02	87.03 ± 0.93 a	88.88

Different lowercase letters indicate statistically significant differences ($p < 0.05$)

Wen *et al. Plant Methods*   (2025) 21:129

Page 7 of 12

Table 3 ACC and LOSS under different training strategies

Model	Scheduling strategy	Average train accuracy (%)	Average verification accuracy (%)	Average Test accuracy (%)	Maximum test accuracy (%)
EfficientNetV2s	SGD	99.66 ± 0.01	99.57 ± 0.01	99.07 ± 0.08 a	99.22
	SGD + StepLR	99.70 ± 0.01	99.57 ± 0.01	98.93 ± 0.08 a	99.07
	SGD + Warmup + COS	99.72 ± 0.01	99.43 ± 0.01	97.03 ± 0.12 a	97.26
	Adam	99.59 ± 0.01	99.57 ± 0.01	98.60 ± 0.03 a	98.65
	Adam + StepLR	99.63 ± 0.01	99.57 ± 0.04	98.93 ± 0.06 b	99.02
	Adam + Warmup + COS	99.69 ± 0.01	99.44 ± 0.01	95.62 ± 0.68 c	96.95
MnasNet	SGD	99.71 ± 0.02	99.39 ± 0.01	96.36 ± 0.12 b	96.59
	SGD + StepLR	99.71 ± 0.00	99.45 ± 0.02	96.88 ± 0.15 ab	97.05
	SGD + Warmup + COS	99.72 ± 0.00	99.37 ± 0.02	93.57 ± 0.02 c	93.59
	Adam	99.65 ± 0.01	99.25 ± 0.06	97.15 ± 0.18 a	97.36
	Adam + StepLR	99.69 ± 0.00	99.36 ± 0.03	96.88 ± 0.05 ab	96.95
	Adam + Warmup + COS	99.74 ± 0.01	99.42 ± 0.03	96.32 ± 0.41 b	97.15
GhostNet	SGD	99.69 ± 0.01	99.48 ± 0.02	95.29 ± 0.18 a	95.50
	SGD + StepLR	99.71 ± 0.01	99.52 ± 0.03	75.25 ± 0.39 c	76.00
	SGD + Warmup + COS	99.71 ± 0.01	99.36 ± 0.01	95.12 ± 0.09 ab	95.29
	Adam	99.21 ± 0.03	99.27 ± 0.07	94.02 ± 0.45 ab	94.88
	Adam + StepLR	99.61 ± 0.01	99.50 ± 0.02	92.42 ± 0.52 b	93.33
	Adam + Warmup + COS	99.60 ± 0.01	99.20 ± 0.11	54.08 ± 2.04 d	57.42
MobileNetV3	SGD	99.71 ± 0.01	99.45 ± 0.00	94.97 ± 0.20 a	95.24
	SGD + StepLR	99.71 ± 0.00	99.45 ± 0.01	94.69 ± 0.11 a	94.83
	SGD + Warmup + COS	99.74 ± 0.01	99.38 ± 0.01	95.17 ± 0.17 a	95.50
	Adam	99.64 ± 0.01	99.42 ± 0.02	91.10 ± 0.51 b	91.83
	Adam + StepLR	99.68 ± 0.02	99.44 ± 0.02	90.50 ± 0.25 b	91.00
	Adam + Warmup + COS	99.71 ± 0.02	99.40 ± 0.00	88.31 ± 0.08 c	88.46
ShuffleNetV2	SGD	99.71 ± 0.02	99.41 ± 0.02	91.17 ± 2.30 b	95.76
	SGD + StepLR	99.73 ± 0.01	99.46 ± 0.02	87.79 ± 0.18 c	88.10
	SGD + Warmup + COS	99.75 ± 0.02	99.42 ± 0.01	94.65 ± 0.15 a	94.93
	Adam	99.66 ± 0.00	99.48 ± 0.02	87.03 ± 0.93 c	88.88
	Adam + StepLR	99.67 ± 0.00	99.45 ± 0.00	89.02 ± 0.11 bc	89.24
	Adam + Warmup + COS	99.74 ± 0.01	99.36 ± 0.01	89.25 ± 0.42 bc	89.91

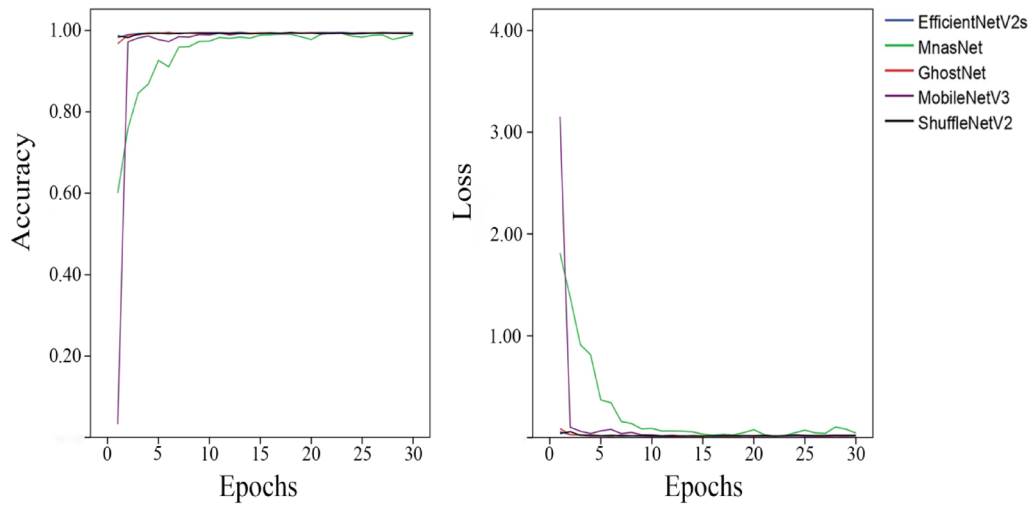Different lowercase letters indicate statistically significant differences ($p < 0.05$)

Fig. 3 The accuracy and loss values of five LCNNs based on the validation set

Wen *et al. Plant Methods*   (2025) 21:129

Page 8 of 12

Table 4 Influence of different λ values on model accuracy ($p < 0.05$)

λ Value	Average Test Accuracy(%)	Maximum Test Accuracy (%)
10^{-4}	96.90 ± 0.19 b	97.26
10^{-5}	97.79 ± 0.15 a	98.03
10^{-6}	96.38 ± 0.29 b	96.95
10^{-7}	96.76 ± 0.17 b	96.95

Different lowercase letters indicate statistically significant differences ($p < 0.05$)

should be carefully selected on the basis of specific tasks. EfficientNetV2s achieved the highest test set accuracy of 99.22%, followed by MnasNet, with an accuracy of 97.36%. Considering the fourfold difference in model weight size with minimal accuracy variation, the more lightweight MnasNet, with a size of 19.16 MB, should be considered.

SimAM optimization analysis

The SimAM module does not introduce additional parameters for altering weights, with only λ serving as an important variable for model normalization. In this study, the recognition accuracy ranged from 10^{-4} to 10^{-7} [25]– [26], and a comparative analysis of the recognition accuracy for different λ values was conducted. The results for the test set are shown in Table 4. Each λ value was measured three times to obtain the mean and standard error. According to Table 4, the model achieved optimal recognition performance when λ was set to 10^{-5}. The results indicated that smaller λ values improved the model's ability to focus on detailed sample features, which are critical for accurately identifying diseases and pests. Through the analysis of local sample features, the model can more precisely distinguish between various diseases and pest types. However, when λ was reduced to 10^{-7}, the model's accuracy decreased because smaller λ values limit the search ranges, focusing solely on local features while disregarding global information.

The optimal SimAM module was selected for comparison with other attention mechanisms (CA, ECA, SE, and CBAM) on the TDPD dataset, and the results are shown in Table 5. Integrating the SimAM module into the Mnas-Net network yielded the highest accuracy of 98.03%,

representing a 0.67% improvement over that of the original model. In contrast, the integration of other attention mechanisms led to varying degrees of accuracy reduction compared with that of the original model. Visualization of the final residual network layers of MnasNet and MnasNet-SimAM is shown in Fig. 4. The SimAM enabled the model to more effectively focus on the features of tea plant diseases and pests and improve model performance in a relatively lightweight architecture.

Classification of the model on the test set

The confusion matrix for 22 categories with MnasNet-SimAM is shown in Fig. 5. The diagonal of the confusion matrix represents the true positive samples (TPs) for each class. Therefore, a darker diagonal color indicates higher recognition accuracy. The model demonstrated strong recognition performance across other categories except for *Arctonis alba* Bremer. Specifically, for eight categories, the model achieved 100% recognition accuracy. Table 6 presents the evaluation metrics for each category using MnasNet-SimAM. The accuracy exceeded 90% across all 22 categories. From the comprehensive evaluation metrics, except *Arctonis alba* Bremer, the model achieved scores above 90% for all categories, demonstrating its strong recognition ability.

Cross-dataset generalization

To verify the generalizability of the improved model, a publicly available dataset of tea tree diseases and pests was downloaded and used to evaluate the accuracy of the model [27]. The comparative experimental results are shown in Table 7, and the confusion matrix is depicted in Fig. 6.

For external validation, MnasNet-SimAM achieved an accuracy of 84.58%, and the F1 scores surpassed those of F-RNet by 18.6% and 4.7% in the classification of red leaf spot and tea coal, respectively. However, these F1 scores were 4.4% and 26.3% lower than those of F-RNet in classifying *Aapolygus lucorum* and gray blight, respectively. Figure 6 shows that the model performs best in classifying *Aapolygus lucorum* and tea coal, followed by red leaf

Table 5 Influence of different attention mechanisms on MnasNet ($p < 0.05$)

Network	AverageTest Accuracy (%)	MaximumTest Accuracy (%)	Parameter Size (MB)	Number of layers	Training time(Min)	Prediction time(Min)
MnasNet	97.15 ± 0.18 b	97.36	19.19	171	165.501	1.81
MnasNet-SimAM	97.79 ± 0.15 a	98.03	19.19	171	70.063	1.26
MnasNet-CA	94.12 ± 0.02 e	94.15	19.31	171	70.856	1.79
MnasNet-ECA	96.72 ± 0.24 b	94.05	19.19	171	71.557	1.58
MnasNet-CBAM	95.28 ± 0.09 d	95.40	19.36	171	71.332	1.27
MnasNet-SE	95.74 ± 0.08 c	95.86	19.27	171	71.693	1.85

Different lowercase letters indicate statistically significant differences ($p < 0.05$)
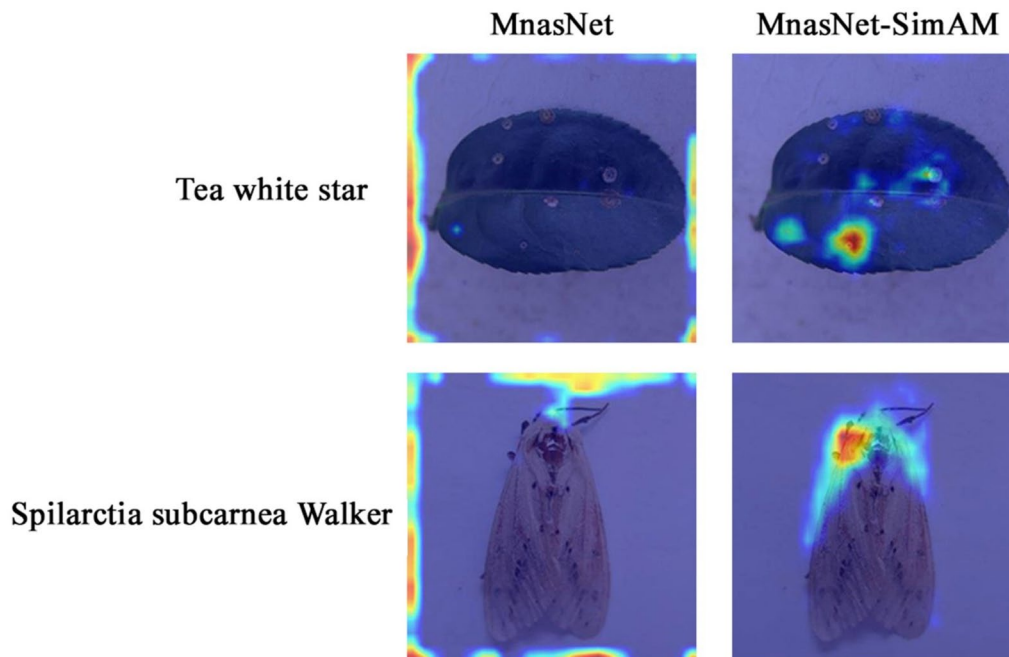
Fig. 4 Class activation map of tea diseases and pests

spot. The model achieved the lowest accuracy for gray blight, with 23 samples misclassified as red leaves.

Discussion

This study presents a systematic evaluation pf various LCNN architectures to identify optimal configurations for tea disease and pest recognition. Comprehensive assessments of different model structures while exploring synergistic combinations of optimizers, hyperparameter settings, and training methodologies were conducted. Subsequent architectural refinements focused on models with high recognition accuracy, with rigorous performance evaluations conducted across public datasets to validate model generalizability.

Attention mechanism efficacy

The parameter-free SimAM integration increased accuracy by 0.67% ($p < 0.05$) without parameter inflation, whereas SE/CBAM-induced overfitting (2.17–3.98% decrease in accuracy). The findings aligned with those of Li et al.'s in maize disease recognition, with a reported 1.5% increase in accuracy through attention mechanism integration [27]. Parameter optimization revealed peak performance at $\lambda = 10^{-5}$, supporting Yang et al.'s findings for the SimAM configuration [23]. The above results are consistent with those in established research demonstrating the efficacy of attention mechanisms in improving model performance.

Attention mechanism selection

Notably, our experimental observations paralleled those reported in Li et al.'s study of convergence instability and overfitting risks in Yunnan tea disease and pest recognition using attention-enhanced transfer learning [12]. Three critical insights emerged from our mechanism selection process: (1) Parameter consideration: Most attention modules introduce trainable parameters. Our results suggest that SimAM's parameter-free architecture prevents the overfitting tendencies observed with parametric alternatives, which is particularly beneficial when the model capacity approaches saturation. (2) Task-specific compatibility: Attention mechanism efficacy appears contingent on source−target task alignment. Ambiguity from incompatible feature emphasis can degrade performance, necessitating careful mechanism selection. (3) Architectural integration: Optimal placement within network hierarchies were found to be crucial. Constrained by pretrained backbone preservation, we positioned SimAM at the terminal inverted residual layer. Inappropriate positioning risks feature map overgeneralization, which is particularly evident in gray blight vs. red leaf spot differentiation challenges.

Performance validation

Despite classification difficulties between morphologically similar pathologies (gray blight/red leaf spot), compounded by dataset limitations such as suboptimal intraclass variance in gray blight samples, the optimized MnasNet-SimAM architecture achieved accuracies of 98.03% based on the TDPD and 84.58% based on a
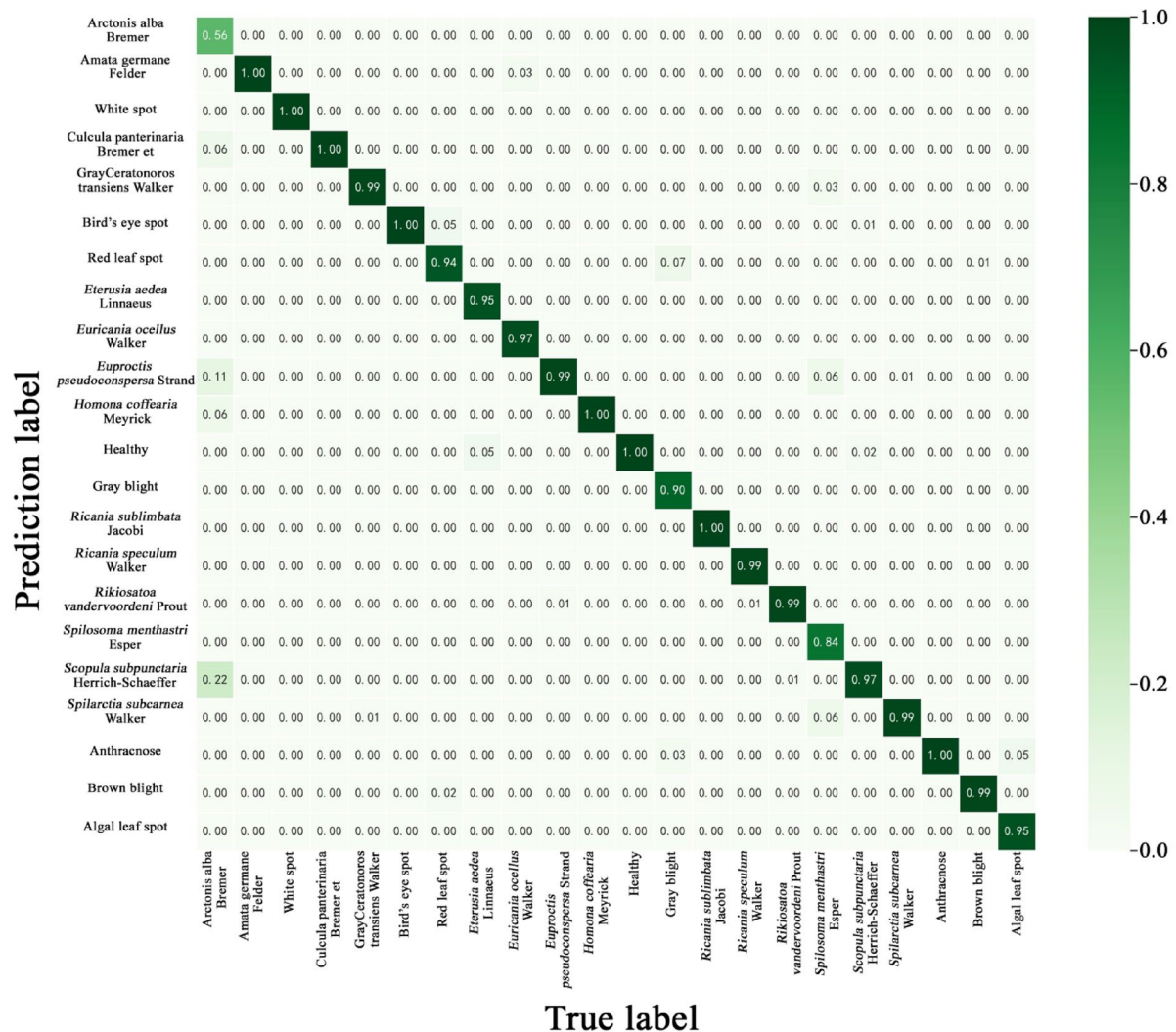
Fig. 5 Confusion matrix of MnasNet-SimAM based on the TDPD dataset

public dataset. Moreover, MnasNet-SimAM achieved accuracies of 95.14% based on the proprietary wheat disease dataset and 92% based on the public wheat disease dataset [28]. Cross-dataset validation confirmed the model's generalizability, validating its capacity to extract domain-agnostic feature representations. These results not only confirm the viability of lightweight CNNs for resource-conscious agricultural diagnostics but also highlight the unique equilibrium between architectural minimalism (19.16 MB) and computational efficacy of MnasNet-SimAM. MnasNet-SimAM's edge compatibility positions it as a transformative solution for real-time crop monitoring systems, particularly in UAV-mounted edge devices deployed across heterogeneous cultivation environments.

Despite the promising performance of the constructed tea tree pest and disease recognition model in the experimental scenario and public dataset, it still exhibits several limitations that need to be addressed in subsequent research. First, the model's generalization ability is constrained by the diversity of its training data. This study focused on 22 major types of tea tree pests and diseases that have high incidence rates in southwest China. However, there are more than 300 known pests and diseases that can harm tea trees globally, including emerging or low-incidence ones. Due to the lack of sufficient training samples for these rare types, the model currently cannot recognize them effectively, which restricts its practical application in large-scale and comprehensive tea garden pest management. Second, the model lacks robustness to complex field interference factors. In the controlled environment, the collected images have clear backgrounds, which help the model extract effective feature information. However, in actual tea garden scenarios, the model frequently encounters challenges such as overlapping tea leaves, fluctuating light intensities, and the presence of non-target objects (e.g., weeds, dew droplets). To address these limitations, future work will involve collecting tea

Table 6 Classification results of MnasNet-SimAM based on the TDPD dataset

Type	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
White spot	98.03	97.01	100.00	98.48
Bird's eye spot		98.48	100.00	99.24
Red leaf spot		95.31	93.85	94.57
Gray blight		96.92	90.00	93.33
Anthracnose		94.00	100.00	96.91
Brown blight		96.25	98.72	97.47
Algal leaf spot		100.00	95.45	97.67
Healthy		99.60	99.60	99.60
Spilarctia subcarnea Walker		97.06	99.00	98.02
Rikiosatoa vandervoordeni Prout		96.15	99.43	97.77
Arctonis alba Bremer		100.00	55.56	71.43
Eterusia aedea Linnaeus		100.00	95.45	97.67
Euproctis pseudoconspersa Strand		98.50	99.49	98.99
Homona coffearia Meyrick		98.68	100.00	99.34
Culcula panterinaria Bremer et Gray		100.00	100.00	100.00
Ricania speculum Walker		100.00	98.77	99.38
Euricania ocellus Walker		100.00	97.44	98.70
Amata germane Felder		96.43	100.00	98.18
Scopula subpunctaria Herrich-Schaeffer		97.46	97.46	97.46
Ricania sublimbata Jacobi		100.00	100.00	100.00
Ceratonoros transiens Walker		97.56	98.77	98.16
Spilosoma menthastri Esper		100.00	84.38	91.53

tree pest and disease images from geographically diverse regions. This effort aims to enrich both the variety and quantity of samples, including coverage of low-incidence types. Additionally, we will simulate complex field interference factors during data collection—such as variable lighting (dawn, dusk, rainy days), leaf overlaps, and non-target objects—to construct a more realistic and robust training dataset. These measures are expected to enhance the model's generalization capability and expand its recognition coverage.
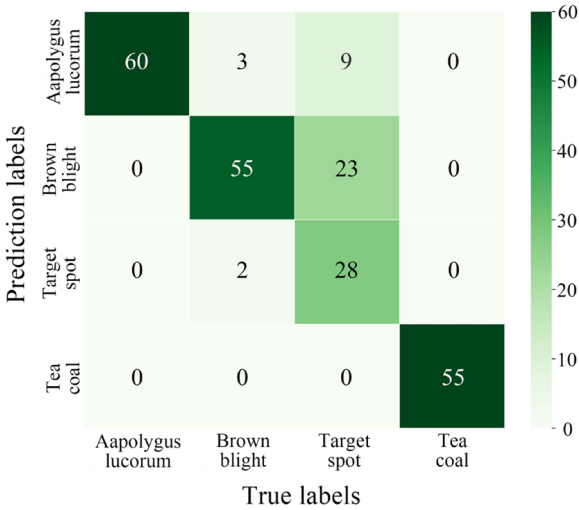

Fig. 6 Confounding matrix of MnasNet-SimAM based on public data

Conclusion

This study introduces MnasNet-SimAM as a state-of-the-art solution for tea leaf disease and pest recognition. This model achieved accuracies of 98.03% based on the TDPD and 84.58% based on cross-data. Parameter-free SimAM integration has significant advantages over conventional attention modules in edge-computing environments and exhibits robust generalizability. Moreover, the results substantiate the operational robustness and practical viability of MnasNet-SimAM for automated tea tree protection systems.

Abbreviations

CNNs	Convolutional neural networks
LCNNs	Lightweight convolutional neural networks
TDPD	Tea disease and pest dataset
UAV	Unmanned aerial vehicle
MnasNet	Mobile neural architecture search
SimAM	Simple attention module
SGD	Stochastic gradient descent
Adam	Adaptive moment estimation
StepLR	Step learning rate
COS	Cosine

Table 7 Classification results of the two models on the publicly available dataset

Model	Type	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Parameter-Size (MB)	Number of layers	Training time (Min)	Prediction time(Min)
MnasNet-SimAM	*Aapolygus lucorum*	84.58	83.3	100.0	90.9	19.19	171	11.796	0.027
	Red leaf spot		70.5	91.7	79.7	19.19	171	17.815	0.093
	Gray blight		93.3	46.7	62.2	19.19	171	18.342	0.116
	Tea coal		100.0	100.0	100.0	19.19	171	15.866	0.043
F-RNet	*Aapolygus lucorum*	88.00	92.3	98.5	95.3	0.06	19	11.208	0.031
	Red leaf spot		50.1	78.3	61.1	0.06	19	17.582	0.107
	Gray blight		89.4	87.2	88.3	0.06	19	18.060	0.109
	Tea coal		92.3	98.5	95.3	0.06	19	15.243	0.052

Data availability
No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate
Not applicable.

Competing interests
The authors declare no competing interests.

References
1. Mao ZQ, Zeng Z, Shi JY. Comparative study on regional competitiveness of tea industry in China. J Agron. 2024;14(09):75–85. https://doi.org/10.11923/j.issn.2095-4050.cjas2023-0201.
2. Chen LJ, Fan ZP, Li XY. Analysis and reflection on the status quo of tea standardization in Yunnan Province. China Tea. 2024;46(11):23–32. https://link.cnki.net/urlid/33.1117.S.20241114.1716.006.
3. Skendžić S, Zovko M, P Živković I, Lešić V, Lemić D. The impact of climate change on agricultural insect pests. Insects. 2021;12(5):440. https://doi.org/10.3390/insects12050440.
4. Panchbhai KG, Lanjewar MG. Enhancement of tea leaf diseases identification using modified SOTA models. Neural Comput Appl. 2025;37(4):2435–53. https://doi.org/10.1007/s00521-024-10758-2.
5. Lanjewar MG, Panchbhai KG. Convolutional neural network based tea leaf disease prediction system on smart phone using Paas cloud. Neural Comput Appl. 2023;35(3):2755–71. https://doi.org/10.1007/s00521-022-07743-y.
6. Rawat W, Wang Z. Deep convolutional neural networks for image classification: a comprehensive review. Neural Comput. 2017;29(9):2352–449. https://doi.org/10.1162/NECO_a_00990.
7. Liu YH. Feature extraction and image recognition with convolutional neural networks. In: Journal of Physics: Conference Series (Vol.1087, p.062032). 2018. IOP Publishing. https://doi.org/10.1088/1742-6596/1087/6/062032
8. Tian Y. Artificial intelligence image recognition method based on convolutional neural network algorithm. IEEE Access. 2020;8:125731–44. https://doi.org/10.1109/ACCESS.2020.3006097.
9. Rahman CR, Arko PS, Ali ME, Khan MAI, Apon SH, Nowrin F, Wasif A. Identification and recognition of rice diseases and pests using convolutional neural networks. Biosyst Eng. 2020;194:112–20. https://doi.org/10.1016/j.biosystemseng.2020.03.020.
10. Turkoglu M, Yanikoğlu B, Hanbay DP. Disease net: convolutional neural network ensemble for plant disease and pest detection. Signal Image Video P. 2022;16(2):301–9. https://doi.org/10.1007/s11760-021-01909-2.
11. Shafik W, Tufail A, Liyanage CDS, Apong RA. A. H. M. Using a novel convolutional neural network for plant pests detection and disease classification. J Sci Food Agric, 2023:103(12), 5849–61. https://doi.org/10.1002/jsfa.12700
12. Li H, Shi H, Du A, Mao Y, Fan K, Wang Y, Ding Z. Symptom recognition of disease and insect damage based on Mask R-CNN, wavelet transform, and F-RNet. Front Plant Sci 2022:13, 922797. https://doi.org/10.3389/fpls.2022.922797.
13. Chen Z, Zhou H, Lin H, Bai D, TeaViTNet. Tea disease and pest detection model based on fused multiscale attention. Agronomy. 2024;14(3):633. https://doi.org/10.3390/agronomy14030633.
14. Tan M, Le Q, Efficientnet. Rethinking model scaling for convolutional neural networks. In International conference on machine learning, PMLR 2019:97: 6105–6114. https://proceedings.mlr.press/v97/tan19a/tan19a.pdf
15. Tan M, Le Q. Efficientnetv2: Smaller models and faster training. In International conference on machine learning. PMLR 2021:139:10096–10106. https://proceedings.mlr.press/v139/tan21a/tan21a.pdf
16. Tan M, Chen B, Pang R, Vasudevan V, Sandler M, Howard A, Le QV, Mnasnet. Platform-aware neural architecture search for mobile. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: pp. 2820–2828. https://openaccess.thecvf.com/content_CVPR_2019/papers/Tan_MnasNet_Platform-Aware_Neural_Architecture_Search_for_Mobile_CVPR_2019_paper.pdf
17. Han K, Wang Y, Tian Q, Guo J, Xu C, Xu C, Ghostnet. More features from cheap operations. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020:pp. 1580–1589. https://openaccess.thecvf.com/content_CVPR_2020/html/Han_GhostNet_More_Features_From_Cheap_Operations_CVPR_2020_paper.html
18. Tang YH, Han K, Guo JY, Xu C, Xu C, Wang Y. GhostNetv2: enhance cheap operation with long-range attention. Adv Neural Inf Process Syst. 2022;35:9969–82. https://proceedings.neurips.cc/paper_files/paper/2022/file/40b60852a4abdaa696b5a1a78da34635-Paper-Conference.pdf.
19. Liu Z, Hao Z, Han K, Tang Y, Wang Y. GhostNetV3: exploring the training strategies for compact models. arXiv:2404.11202.2024. https://doi.org/10.48550/arXiv.2404.11202
20. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 4510–4520. https://openaccess.thecvf.com/content_cvpr_2018/html/Sandler_MobileNetV2_Inverted_Residuals_CVPR_2018_paper.html
21. Zhang X, Zhou X, Lin M, Sun J, Shufflenet. An extremely efficient convolutional neural network for mobile devices. In Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 6848–6856. https://openaccess.thecvf.com/content_cvpr_2018/papers/Zhang_ShuffleNet_An_Extremely_CVPR_2018_paper.pdf
22. Ma N, Zhang X, Zheng HT, Sun J. Shufflenet v2: Practical guidelines for efficient cnn architecture design. In Proceedings of the European conference on computer vision (ECCV). 2018: 116–131. https://openaccess.thecvf.com/content_ECCV_2018/papers/Ningning_Light-weight_CNN_Architecture_ECCV_2018_paper.pdf
23. Yang L, Zhang RY, Li L, Xie XSAM. A Simple, Parameter-Free Attention Module for Convolutional Neural Networks. International Conference on Machine Learning, PMLR. 2021. https://proceedings.mlr.press/v139/yang21o.html
24. Panchbhai KG, Lanjewar MG, Naik AV. Modified MobileNet with leaky ReLU and LSTM with balancing technique to classify the soil types. Earth Sci Inf. 2025;18(1):77. https://doi.org/10.1007/s12145-024-01521-1.
25. Sharma A, Patel RK, Pranjal P, Panchal B, Chouhan SS. Computer vision-based smart monitoring and control system for crop. Applications of computer vision and drone technology in agriculture 4.0. Singapore: Springer; 2024. pp. 65–82. https://doi.org/10.1007/978-981-99-8684-2_5.
26. Chouhan SS, Singh UP, Jain S. Performance evaluation of different deep learning models used for the purpose of healthy and diseased leaves classification of Cherimoya (Annona Cherimola) plant. Neural Comput Appl. 2025;37(6):4531–44. https://doi.org/10.1007/s00521-024-10830-x. https://link.springer.com/article/.
27. Li H, Qi M, Du B, Li Q, Gao H, Yu J, Bi C, Yu H, Liang M, Ye G, Tang Y. Maize disease classification system design based on improved ConvNeXt. Sustainability. 2023;2015. https://doi.org/10.3390/su152014858.
28. Wen X, Maimaiti M, Liu Q, Yu F, Gao H, Li G, Chen J. MnasNet-SimAM: an improved deep learning model for the identification of common wheat diseases in complex Real-Field environments. Plants. 2024;1613. https://doi.org/10.3390/plants13162334.

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