

Article

ShuffleNetV2-hSimKD: A Lightweight Network for Plant Disease Detection

Qiuxin Si ¹, Yoojeong Song ² and Sang Ik Han ^{2,*}¹ Department of Artificial Intelligence, Zibo Polytechnic University, Zibo 255000, China; 12459@zbpu.edu.cn² Department of Computer Software Engineering, Soonchunhyang University, Asan-si 31538, Republic of Korea; yoojeong@sch.ac.kr

* Correspondence: sihan@sch.ac.kr

Abstract

Early and accurate plant disease detection is essential for reducing crop losses and supporting sustainable agricultural management. Although deep learning-based approaches have achieved strong performance in plant disease analysis, many existing models require substantial computational resources, which limits their use in resource-constrained agricultural environments. This study proposes ShuffleNetV2-hSimKD, a lightweight integration framework for plant disease detection. It adopts ShuffleNetV2 as the backbone and incorporates the parameter-free SimAM attention mechanism to enhance disease-related feature representation without introducing additional learnable parameters. In addition, the standard ReLU activation function is replaced with h-swish to improve nonlinear feature extraction and preserve informative feature responses. A hybrid knowledge distillation strategy is further employed to transfer both output-level and feature-level knowledge from a high-capacity teacher model to the lightweight student network during training. Unlike previous studies that apply these techniques in isolation, ShuffleNetV2-hSimKD synergistically integrates parameter-free SimAM, h-swish optimization, and hybrid KD to overcome the representation limitations of lightweight backbones in subtle disease symptom detection. The proposed framework was evaluated on a balanced subset of the PlantVillage dataset, in which leaf images were categorized as healthy or diseased. ShuffleNetV2-hSimKD achieved an accuracy of 90.41% with only 1.4M parameters and 151M FLOPs. Compared with representative lightweight Convolutional Neural Networks (CNNs), the proposed model achieved improved accuracy and recall while maintaining low computational complexity. These results demonstrate that ShuffleNetV2-hSimKD provides an effective balance between detection performance and computational efficiency, highlighting its potential as a lightweight candidate for plant disease detection in resource-constrained agricultural scenarios.



Academic Editor: Xanthoula Eirini Pantazi

Received: 6 July 2026

Revised: 2 August 2026

Accepted: 4 August 2026

Published: 6 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: ShuffleNetV2; parameter-free attention; hybrid knowledge distillation; edge inference; agricultural image analysis

1. Introduction

Plant diseases are a major threat to crop productivity, food security, and sustainable agricultural development. Plant pathogens are estimated to cause annual yield losses of approximately 20–40% in major food crops worldwide, creating substantial economic and social burdens, particularly in regions where access to agricultural experts and diagnostic facilities is limited [1]. Conventional plant disease diagnosis commonly relies

on visual inspection by agronomists or laboratory-based pathological analysis. However, these approaches are labor-intensive, time-consuming, and often unsuitable for timely large-scale crop monitoring [2]. Delayed diagnosis may result in inappropriate pesticide application, reduced crop quality, and further yield loss. Therefore, automated plant disease detection has become an important component of precision agriculture, enabling rapid, non-destructive, and scalable crop health monitoring [3,4].

Among various sensing modalities, RGB imaging has received considerable attention because visible disease symptoms frequently appear on leaves as spots, lesions, chlorosis, discoloration, or texture abnormalities. With the rapid development of deep learning, convolutional neural networks (CNNs) have become widely used for image-based plant disease detection. Studies using public datasets, such as PlantVillage, have reported high classification performance under controlled imaging conditions [5,6]. In particular, deep CNN architectures including AlexNet, VGG, ResNet, Inception, DenseNet, and EfficientNet have demonstrated strong capability for extracting discriminative visual features from plant images [7–9]. More recently, Vision Transformer-based models have also been explored to capture long-range dependencies and fine-grained visual patterns in plant disease images [10].

Despite their high predictive performance, many deep neural networks require a large number of parameters and floating-point operations (FLOPs), resulting in substantial memory consumption and inference latency. These requirements restrict their use in practical agricultural environments where computational resources, battery capacity, and network connectivity may be limited. In particular, agricultural applications may involve smartphones, portable imaging devices, drones with onboard processors, and low-power IoT nodes deployed in greenhouses or orchards [11,12]. In such environments, cloud-based inference may not always be reliable because of intermittent connectivity, communication delay, or limited bandwidth. Accordingly, lightweight plant disease detection models that can provide efficient inference with limited computational resources are increasingly important for smart agriculture [13].

To reduce computational complexity, several lightweight CNN architectures have been developed, including MobileNetV2, MobileNetV3, ShuffleNetV2, and EfficientNet-Lite [14–16]. These networks employ efficient design strategies, such as depthwise separable convolution, channel shuffle, and optimized channel scaling, to reduce model size and FLOPs while preserving reasonable predictive performance. Among these architectures, ShuffleNetV2 was adopted as the baseline backbone because its hardware-friendly design reduces memory access cost (MAC), which is a critical factor for edge CPU deployment [15]. Nevertheless, its compact structure may limit feature representation for small or low-contrast disease symptoms, and the standard ReLU activation may suppress informative negative responses. These limitations motivate the proposed integration of SimAM, h-swish, and hybrid knowledge distillation.

Attention mechanisms have been introduced to address this limitation by enabling neural networks to emphasize informative regions and suppress less relevant features. In plant disease detection, attention can help a model focus on disease-related regions, such as lesion boundaries, discolored areas, and abnormal texture patterns [17,18]. Previous studies have incorporated attention modules into lightweight CNNs to improve plant disease classification performance. For example, Ref. [19] integrated the squeeze-and-excitation (SE) module into ShuffleNetV2, while other studies applied channel or spatial attention mechanisms to MobileNet-based architectures [20]. Although these approaches can improve feature representation, many conventional attention modules, including SE and convolutional block attention module (CBAM), introduce additional learnable parameters and computational overhead [21,22]. Such overhead may reduce the advantage of lightweight networks when they are used in highly resource-constrained environments.

A promising alternative is the parameter-free Simple Attention Module (SimAM), which generates three-dimensional attention weights directly from feature map statistics without introducing additional learnable parameters [23]. SimAM estimates the importance of individual neurons based on their separability from neighboring neurons and can therefore enhance informative feature responses while preserving model compactness. This characteristic makes SimAM appropriate for lightweight vision models intended for computationally constrained applications. Previous studies have also demonstrated the potential of ShuffleNetV2 and SimAM in agricultural image analysis. For example, Ref. [24] proposed a ShuffleNetV2-based corn leaf disease model incorporating SimAM and reported improved recognition performance. These findings suggest that parameter-free attention can be an effective means of strengthening the representation capability of lightweight CNNs without substantially increasing model complexity.

In addition to attention design, activation functions influence the feature extraction capability of lightweight CNNs. The ReLU activation function is computationally efficient but suppresses all negative responses, which may limit feature diversity in compact networks. In contrast, h-swish provides a smoother nonlinear transformation while maintaining low computational cost [25]. Because h-swish preserves partial negative responses within a limited range, it can improve nonlinear feature representation without significantly increasing computational burden. Therefore, replacing ReLU with h-swish may further enhance the ability of a lightweight model to capture disease-related visual patterns.

Although lightweight architectural improvements can enhance inference efficiency, compact models still have limited learning capacity compared with larger CNNs. Knowledge distillation (KD) provides an effective training strategy to address this limitation without increasing the complexity of the final deployed model [26]. In knowledge distillation, a lightweight student network learns not only from ground-truth labels but also from the outputs or intermediate features of a larger teacher network. This additional supervision enables the student model to acquire richer class relationships and feature representations. Recent studies have shown that combining response-based and feature-based distillation can improve the performance of compact CNNs in image classification tasks [27,28]. Therefore, knowledge distillation is a suitable complementary strategy for improving lightweight plant disease detection models while preserving their low inference cost.

Despite these advances, the combined use of an efficient lightweight backbone, parameter-free attention, activation optimization, and hybrid knowledge distillation remains insufficiently explored for plant disease detection. In addition, many previous studies primarily emphasize classification accuracy on controlled datasets without jointly analyzing predictive performance and computational efficiency. For practical smart agriculture applications, model size, computational complexity, and inference time should be considered together with detection performance. Although Ref. [24] applied SimAM to ShuffleNetV2 for corn leaf disease detection and other studies have used knowledge distillation separately, the joint integration of parameter-free attention, h-swish activation, and hybrid knowledge distillation for low-parameter plant disease detection remains underexplored. This gap motivates the proposed framework.

In this study, we propose ShuffleNetV2-hSimKD, a lightweight integration framework for plant disease detection. ShuffleNetV2 is adopted as the backbone because of its hardware-efficient design. SimAM is integrated into selected ShuffleNetV2 stages to strengthen disease-related feature representation without introducing additional trainable parameters. In addition, h-swish is used in place of ReLU to improve nonlinear feature extraction. During training, a hybrid knowledge distillation strategy transfers both output-level and feature-level knowledge from an EfficientNetB3 teacher model to the lightweight student model. The proposed framework is evaluated using a balanced subset of the

PlantVillage dataset, where leaf images are categorized as healthy or diseased. The main contributions of this study are summarized as follows:

- **Lightweight plant disease detection architecture:** We propose ShuffleNetV2-hSimKD, a lightweight CNN-based framework that combines ShuffleNetV2, parameter-free attention, activation optimization, and knowledge distillation for efficient plant disease detection.
- **Parameter-free feature enhancement:** We integrate SimAM into selected ShuffleNetV2 blocks to emphasize disease-related feature responses without increasing the number of learnable parameters.
- **Improved nonlinear feature extraction:** We replace the standard ReLU activation with h-swish to improve feature representation while preserving the computational efficiency required for lightweight CNNs.
- **Hybrid knowledge distillation strategy:** We employ both output-level and feature-level distillation from an EfficientNetB3 teacher model to improve the predictive performance of the lightweight student network during training without increasing inference time complexity.
- **Performance and efficiency evaluation:** We evaluate the lightweight integration framework on a healthy-versus-diseased plant disease detection task using PlantVillage leaf images and compare it with representative lightweight CNNs in terms of accuracy, precision, recall, F1 score, parameters, FLOPs, and inference time.

The remainder of this paper is organized as follows: Section 2 reviews the fundamentals of ShuffleNetV2. Section 3 describes the proposed ShuffleNetV2-hSimKD framework, including the SimAM attention mechanism, h-swish activation design, and hybrid knowledge distillation framework. Section 4 presents the experimental setup, results, and performance analysis. Finally, Section 5 concludes the paper.

2. Research Background

2.1. ShuffleNetV2 Architecture and Design Principles

ShuffleNetV2 is a hardware-efficient lightweight CNN designed to provide efficient inference on resource-constrained hardware [29]. Unlike conventional CNN designs that primarily focus on reducing FLOPs, ShuffleNetV2 considers both theoretical computational complexity and practical hardware efficiency, including MAC [15].

In plant disease detection, leaf images often exhibit high spatial redundancy, where healthy tissue occupies large and relatively uniform regions, whereas disease symptoms may appear as small, sparse lesions with subtle texture and color variations. ShuffleNetV2's design principles, including balanced channel allocation, reduced branch fragmentation, and fewer element-wise operations, are well aligned with these imaging characteristics. These properties make ShuffleNetV2 an appropriate backbone for lightweight plant disease detection models intended for deployment on resource-constrained agricultural devices. The overall architecture of ShuffleNetV2 is illustrated in Figure 1.

2.2. Group Convolution and Channel Shuffle

Group convolution was introduced to reduce the computational cost of conventional convolution. In standard convolution, each output channel is generated by convolving all input channels with a corresponding kernel. For an input with C_{in} channels, an output with C_{out} channels, and a kernel size of $k \times k$, the number of parameters is expressed as

$$k^2 C_{in} C_{out}. \quad (1)$$

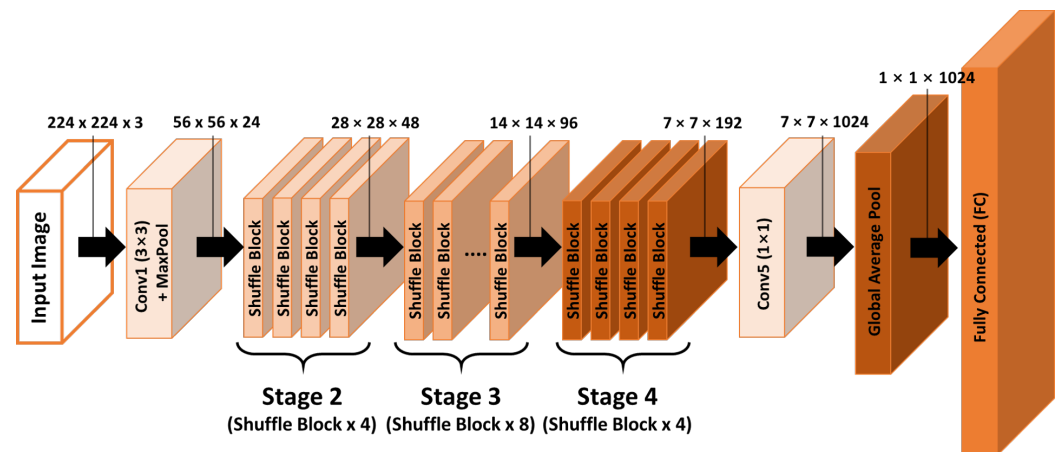


Figure 1. Overall architecture of ShuffleNetV2.

When the input and output channels are divided into g groups, convolution is independently performed within each group. The number of parameters is then reduced to

$$\frac{k^2 C_{in} C_{out}}{g}. \quad (2)$$

As illustrated in Figure 2, group convolution can substantially reduce the number of parameters and computational operations by limiting convolution to channel groups rather than processing all channels jointly [27].

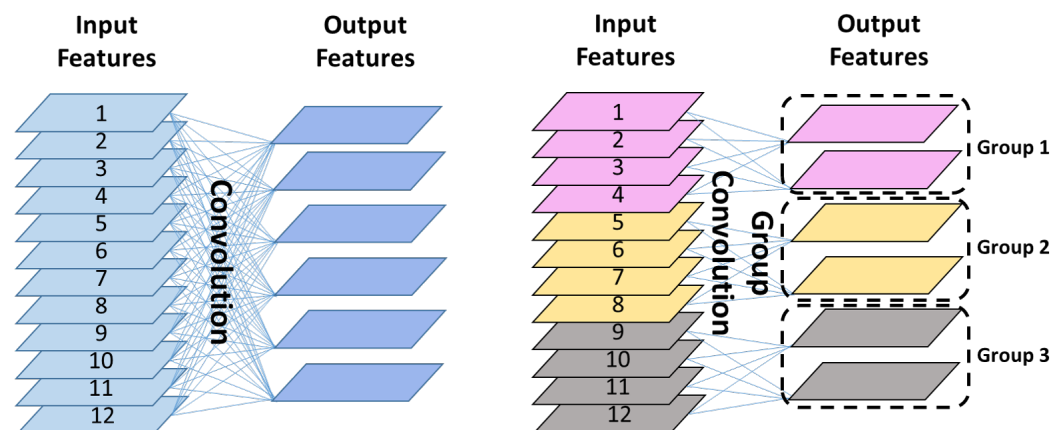


Figure 2. Illustration of group convolution.

However, group convolution restricts information exchange across different channel groups. If multiple group convolution layers are stacked without additional processing, each group continues to process only a limited subset of the available features. Channel shuffle was introduced to address this limitation. As shown in Figure 3, channel shuffle permutes the output channels after group convolution so that subsequent groups receive feature information from multiple preceding groups [30]. This operation enables cross-group information exchange while preserving the computational advantage of grouped processing.

In the context of plant disease detection, group convolution can reduce redundant computation over relatively uniform leaf-background regions, whereas channel shuffle supports the integration of complementary feature cues across channel groups. For example, color-related responses associated with chlorosis and edge- or texture-related responses associated with lesion boundaries can be more effectively combined through cross-group information exchange. This property is useful for recognizing visually subtle disease symptoms while maintaining the computational efficiency of a compact CNN.

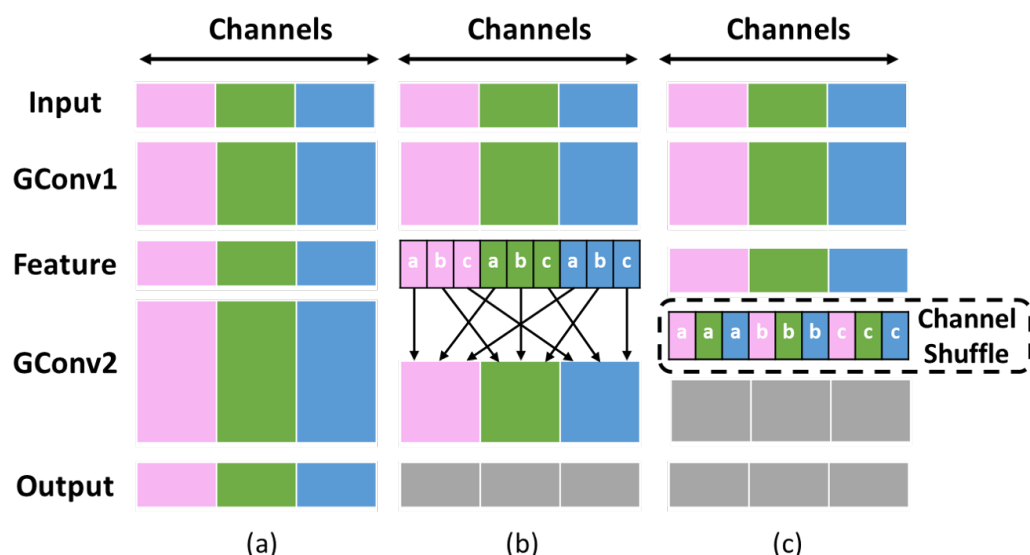


Figure 3. Channel shuffle for cross-group feature interaction: (a) group convolution, (b) channel permutation and (c) channel shuffle.

2.3. ShuffleNetV2 Blocks

The ShuffleNetV2 architecture is constructed by stacking two types of Shuffle Blocks, as illustrated in Figure 4. Structure (a) is used when the spatial resolution and the number of channels are preserved, whereas structure (b) is used for spatial downsampling.

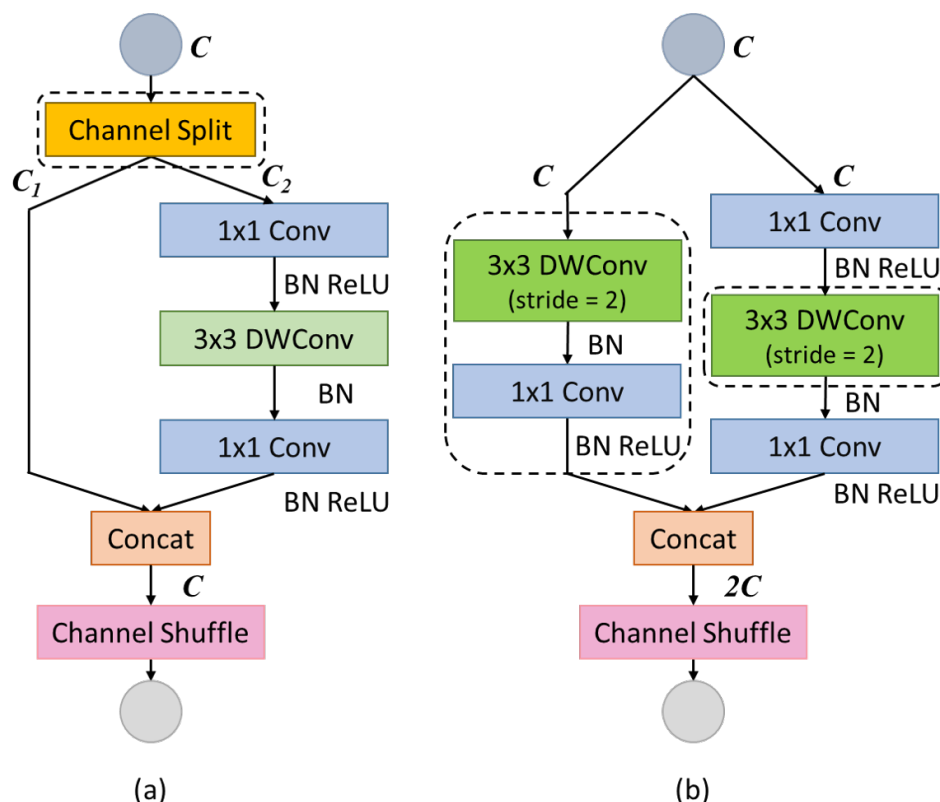


Figure 4. ShuffleNetV2 blocks: (a) stride-1 block and (b) stride-2 downsampling block.

In structure (a), the input feature map is first divided equally into two channel groups through a channel split operation. One group is directly propagated through the shortcut branch without additional computation. The other group is processed by a 1×1 pointwise convolution, a 3×3 depthwise convolution, and another 1×1 pointwise convolution. The outputs of the two branches are then concatenated and channel-shuffled before being

passed to the next block. This design reduces redundant computation by allowing part of the input feature map to bypass convolutional processing.

Structure (b) is used when spatial downsampling is required. Unlike structure (a), all input channels participate in convolutional processing because excessive channel splitting during downsampling may lead to information loss. The two branches independently perform downsampling operations, and their outputs are concatenated to reduce the spatial resolution while increasing the number of output channels. This design enables efficient feature extraction while maintaining sufficient information flow across network stages.

ShuffleNetV2 was designed according to four practical principles for efficient CNN implementation [15]:

1. **Balanced channel allocation:** Maintaining comparable numbers of input and output channels reduces MAC. For leaf-image analysis, this design helps preserve informative disease-related regions while keeping computation over relatively uniform background areas tractable.
2. **Limited use of group convolution:** Avoiding excessive group convolution reduces memory latency, which is important for real-time inference on edge devices, such as smartphones and drones, used in field monitoring.
3. **Reduced branch fragmentation:** Minimizing fragmented parallel branches decreases synchronization and memory-access overhead, thereby supporting more efficient inference under resource-constrained agricultural deployment conditions.
4. **Fewer element-wise operations:** Replacing element-wise addition with channel concatenation reduces memory traffic, enabling high-resolution leaf images to be processed efficiently while maintaining detection performance. As illustrated in Figure 5, the concatenation, channel shuffle, and subsequent channel split operations can be efficiently implemented as a unified operation.

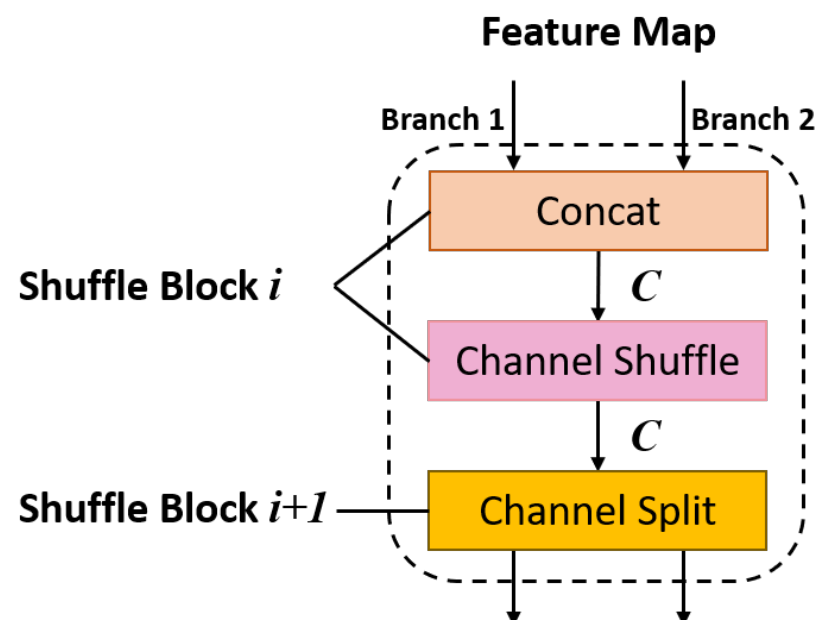


Figure 5. Fusion of concatenation, channel shuffle, and channel split as an efficient element-wise operation.

2.4. Overall ShuffleNetV2 Architecture

The overall ShuffleNetV2 architecture consists of a 3×3 convolution layer, a max-pooling layer, three main stages, a 1×1 convolution layer, global average pooling, and a fully connected classification layer. The three stages, denoted as Stage 2, Stage 3, and Stage 4, are formed by stacking Shuffle Blocks. In the standard ShuffleNetV2 configuration, Stage 2, Stage 3, and Stage 4 contain four, eight, and four Shuffle Blocks, respectively.

Within each stage, the first block uses structure (b) with a stride of 2 to reduce the spatial resolution of the feature map and increase the number of channels. The remaining blocks use structure (a) with a stride of 1 to preserve the spatial resolution while refining feature representations. After Stage 4, the 1×1 convolution layer further transforms the high-level feature map before global average pooling and final classification.

Through this hardware-aware design, ShuffleNetV2 achieves a favorable trade-off between representational capacity and computational efficiency. However, when directly applied to plant disease detection, its compact architecture may under-represent subtle and spatially sparse disease symptoms. To address this limitation, the proposed framework enhances ShuffleNetV2 through three complementary modifications: integration of the parameter-free SimAM attention mechanism (Section 3.2), replacement of ReLU with h-swish activation (Section 3.3), and hybrid knowledge distillation (Section 3.4). These enhancements are designed to improve sensitivity to fine-grained disease-related features while preserving the lightweight inference characteristics of ShuffleNetV2.

3. Proposed ShuffleNetV2-hSimKD Model

ShuffleNetV2 is an efficient lightweight CNN architecture. In this study, we adopt the ShuffleNetV2 $0.5 \times$ variant [15] as the baseline backbone. This variant consists of three stages of Shuffle Blocks containing 4, 8, and 4 blocks, respectively. However, its compact design may limit feature representation capability when disease symptoms are small, weak, or visually similar to healthy leaf regions. To improve plant disease detection performance while preserving computational efficiency, this study progressively enhances ShuffleNetV2 in three stages. First, the parameter-free SimAM attention mechanism is incorporated into ShuffleNetV2 to emphasize informative disease-related features. Second, the ReLU activation function is replaced with h-swish to improve nonlinear feature representation. Finally, hybrid knowledge distillation is applied during training to transfer knowledge from a high-capacity teacher network to the lightweight student network. The resulting models are denoted as ShuffleNetV2-Sim, ShuffleNetV2-hSim, and ShuffleNetV2-hSimKD, respectively.

3.1. Simple Attention Module

Attention mechanisms can improve CNN performance by emphasizing informative features and suppressing less relevant responses. In plant disease detection, attention is useful because disease symptoms may appear only in localized regions of leaf images, such as lesion boundaries, discolored areas, or texture abnormalities.

The SimAM is a parameter-free attention mechanism that estimates the importance of individual neurons directly from feature statistics [23]. Unlike conventional attention mechanisms that introduce additional convolutional or fully connected layers, SimAM computes attention weights without adding trainable parameters. Therefore, it is suitable for lightweight CNN architectures intended for computationally constrained environments.

Let the input feature map be denoted as $X \in \mathbb{R}^{C \times H \times W}$. For the c^{th} channel, the corresponding feature map is represented as $X_c \in \mathbb{R}^{H \times W}$ and reshaped into a vector $x = [x_1, x_2, \dots, x_N]$, where $N = H \times W$. SimAM evaluates the importance of each neuron according to its separability from other neurons within the same channel.

The mean and variance of the c^{th} channel are calculated as

$$\mu_c = \frac{1}{N} \sum_{i=1}^N x_i, \quad (3)$$

$$\sigma_c^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu_c)^2. \quad (4)$$

Based on these statistics, the energy value of the i^{th} neuron x_i is obtained as

$$e_i^* = \frac{4(\sigma_c^2 + \lambda)}{(x_i - \mu_c)^2 + 2\sigma_c^2 + 2\lambda}, \quad (5)$$

where λ is a regularization parameter. A lower energy value indicates that the neuron is more distinguishable from its surrounding neurons and should receive greater attention. The attention weight is computed as

$$w_i = \text{sigmoid}\left(\frac{1}{e_i^*}\right), \quad (6)$$

and the weighted output feature map is expressed as

$$\tilde{X}_c = W_c \odot X_c, \quad (7)$$

where W_c denotes the attention weight matrix for the c^{th} channel and $\odot$ represents element-wise multiplication. The weighted outputs from all channels are then combined to produce the final output feature map $\tilde{X}$.

Unlike channel-attention methods such as SE and spatial-attention methods such as CBAM, SimAM assigns an importance weight to each neuron in the feature map. Thus, it can refine channel-wise and spatial feature responses without introducing additional learnable parameters. The structure of SimAM is illustrated in Figure 6.

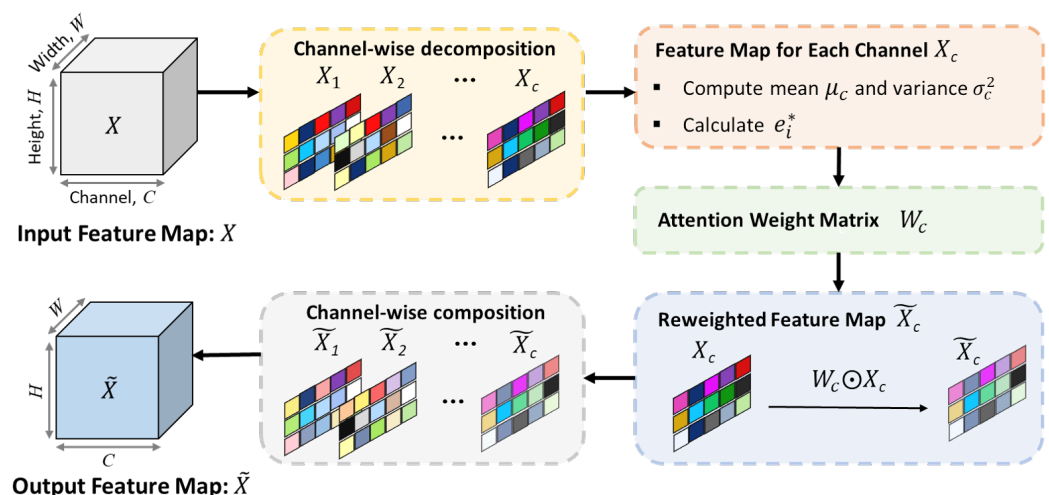


Figure 6. Structure of the parameter-free SimAM attention module.

3.2. SimAM-Enhanced ShuffleNetV2 (ShuffleNetV2-Sim)

To improve the feature representation capability of ShuffleNetV2, SimAM is inserted after the channel shuffle operation at the end of selected Shuffle Blocks. At this point, feature information from the main and shortcut branches has already been concatenated, and the channel shuffle operation has redistributed information across channel groups. Therefore, SimAM can evaluate neuron importance using more comprehensive feature representations. To preserve the lightweight characteristics of ShuffleNetV2, SimAM is applied only to the stride-1 Shuffle Blocks in Stage 2 and Stage 3. These stages retain sufficient spatial resolution to refine disease-related visual patterns, including lesion regions, discoloration, and local texture abnormalities. In contrast, stride-2 blocks are primarily used for spatial downsampling and computational reduction. Therefore, SimAM is not inserted into the downsampling blocks. The resulting modified block is referred to as the Shuffle-Sim block, as illustrated in Figure 7. The ShuffleNetV2 backbone enhanced with SimAM is denoted as ShuffleNetV2-Sim.

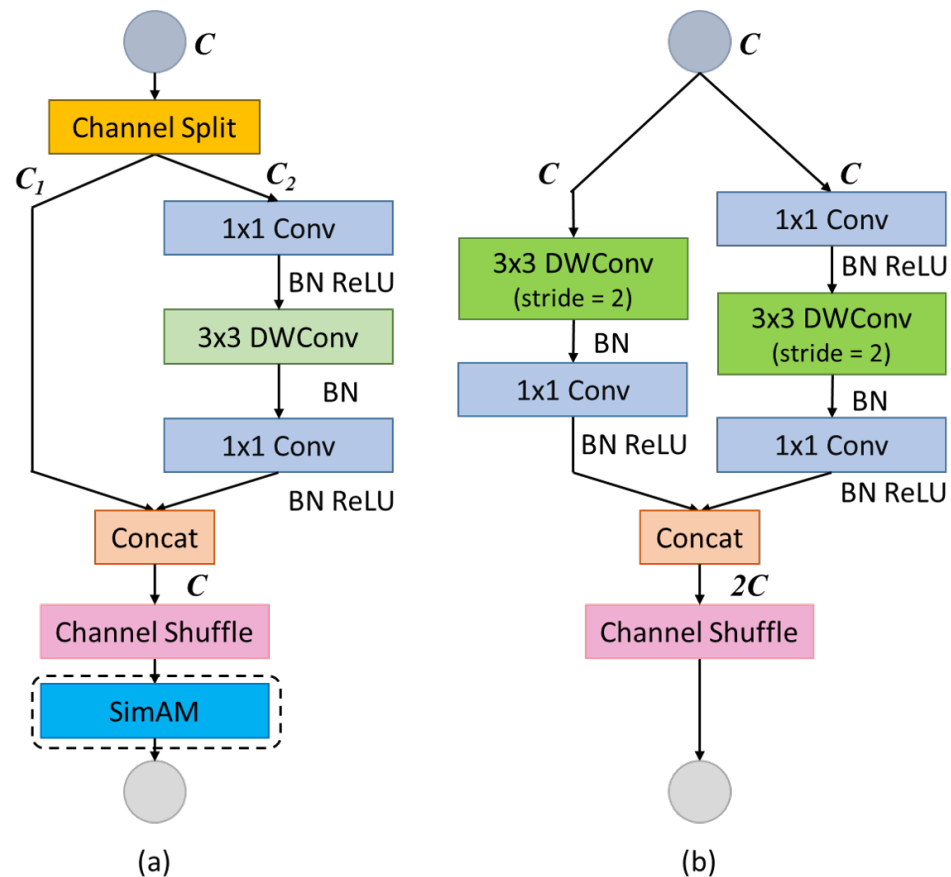


Figure 7. Shuffle-Sim blocks: (a) stride-1 block and (b) stride-2 downsampling block.

3.3. Activation Design with H-Swish

Although SimAM improves the feature representation capability of ShuffleNetV2, the original ReLU activation function may still limit the diversity of learned feature responses. In ShuffleNetV2, the pointwise convolution is followed by the rectified linear unit (ReLU), which is defined as

$$\text{ReLU}(x) = \max(0, x). \quad (8)$$

ReLU is computationally efficient; however, it suppresses all negative responses to zero. In lightweight networks, this behavior may reduce the preservation of weak but informative feature responses. To address this limitation, h-swish is adopted as the activation function [31]. It is defined as

$$\text{h-swish}(x) = x \cdot \frac{\text{ReLU6}(x + 3)}{6}, \quad (9)$$

where

$$\text{ReLU6}(x) = \min(\max(0, x), 6). \quad (10)$$

Compared with ReLU, h-swish retains nonzero responses over part of the negative input range while maintaining low computational cost. This characteristic can improve nonlinear feature representation without substantially increasing the number of FLOPs [32].

In the proposed framework, h-swish complements SimAM. SimAM emphasizes informative neurons according to intra-channel feature statistics, whereas h-swish preserves useful activation responses that would otherwise be entirely removed by ReLU. This combination can improve the representation of disease-related visual patterns, including small lesion regions, mild discoloration, and local texture abnormalities. Accordingly, the ReLU activation following the 1×1 pointwise convolution in the Shuffle-Sim block is replaced

with h-swish. The resulting block, termed the Shuffle-hSim block, is illustrated in Figure 8. The corresponding network is denoted as ShuffleNetV2-hSim.

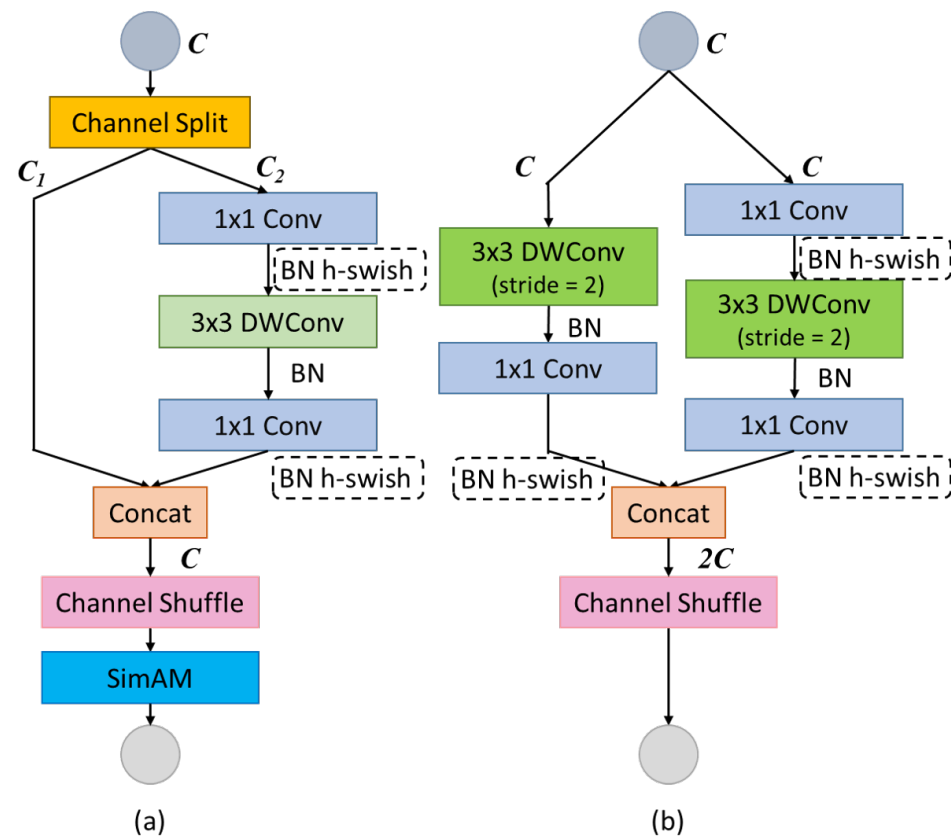


Figure 8. Shuffle-hSim blocks with h-swish activation: (a) stride-1 block and (b) stride-2 downsampling block.

3.4. Hybrid Knowledge Distillation

Although ShuffleNetV2-hSim improves feature representation through SimAM and h-swish, its model capacity remains limited compared with larger CNN architectures. Therefore, hybrid knowledge distillation is applied during training to further improve the predictive performance of the lightweight student network without increasing inference-time complexity.

In the proposed framework, EfficientNetB3 is used as the teacher network, and ShuffleNetV2-hSim is used as the student network. The hybrid knowledge distillation strategy combines logit-based distillation and feature-based distillation. Logit-based distillation transfers class-level knowledge from the output distribution of the teacher network, whereas feature-based distillation transfers intermediate feature representations. The final model trained through this procedure is denoted as ShuffleNetV2-hSimKD.

3.4.1. Logit-Based Distillation

For a classification problem with M classes, let z_t and z_s denote the logits generated by the teacher and student networks, respectively. The temperature-scaled output probabilities are calculated as

$$p_t^{(i)} = \frac{\exp(z_t^{(i)} / T)}{\sum_{j=1}^M \exp(z_t^{(j)} / T)}, \quad (11)$$

$$p_s^{(i)} = \frac{\exp(z_s^{(i)} / T)}{\sum_{j=1}^M \exp(z_s^{(j)} / T)}, \quad (12)$$

where T is the temperature parameter. A larger value of T produces a softer probability distribution and enables the student network to learn class relationships represented by the teacher output.

The logit distillation loss is calculated using Kullback–Leibler divergence:

$$L_{KD} = T^2 \sum_{i=1}^M p_t^{(i)} \log \left(\frac{p_t^{(i)}}{p_s^{(i)}} \right). \quad (13)$$

The student network is also supervised using the ground-truth labels. The cross-entropy loss is defined as

$$L_{CE} = - \sum_{i=1}^M y_i \log \left(\hat{p}_s^{(i)} \right), \quad (14)$$

where y_i denotes the ground-truth label and $\hat{p}_s^{(i)}$ denotes the standard student output probability calculated using Softmax with $T = 1$. The logit-based distillation loss is

$$L_{LD} = \alpha L_{KD} + \beta L_{CE}, \quad (15)$$

where α and β balance soft and hard supervision [33], and its process is illustrated in Figure 9.

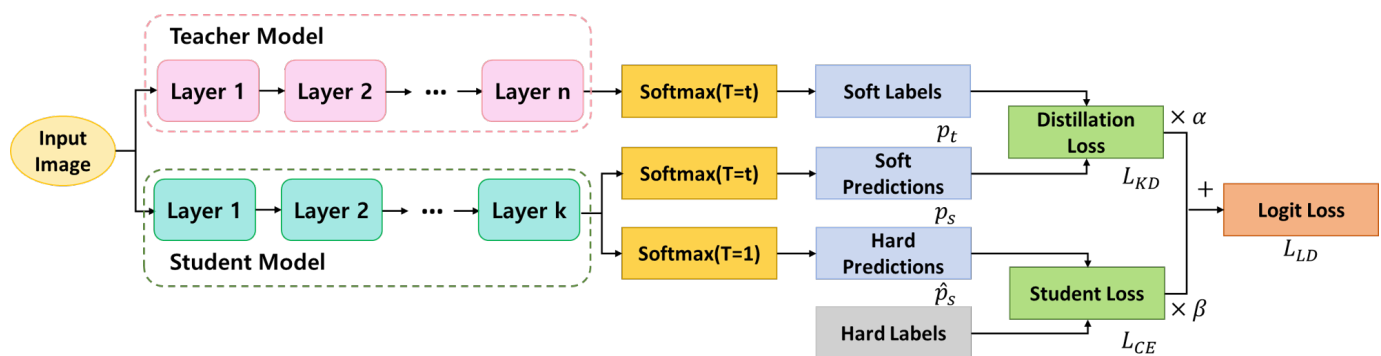


Figure 9. Logit-based knowledge distillation process.

3.4.2. Feature-Based Distillation

While logit-based distillation transfers class-level knowledge from the teacher output, feature-based distillation guides the student network to learn intermediate feature representations from the teacher network [34,35]. This additional supervision helps the lightweight student model capture richer spatial and semantic information that may not be sufficiently conveyed by output-level distillation alone.

In the proposed framework, EfficientNetB3 is employed as the teacher network, whereas ShuffleNetV2-hSim serves as the student network. In this study, an input resolution of 300×300 is used to better preserve fine-grained lesion textures that may occupy only a small portion of the leaf image. Under this setting, candidate intermediate feature maps from the teacher and student networks were examined to determine an appropriate feature distillation layer. Aligning an earlier teacher feature map with the corresponding student feature map preserves higher spatial resolution, but such shallow representations mainly contain low-level visual cues, such as color and edge information, and may provide limited semantic guidance for disease-related feature learning. In contrast, deeper teacher feature maps contain stronger semantic information but may be spatially compressed, which can reduce their ability to represent small or dispersed disease regions. Therefore, we selected an intermediate layer that provides a balance between semantic abstraction and spatial detail. Specifically, the Stage 4 output of EfficientNetB3 and the Stage 3 output of ShuffleNetV2-hSim were used for feature-based distillation. Under the 300×300 input

resolution, these feature maps have the same spatial resolution of 19×19 and the same channel dimension of 96. This direct alignment avoids additional spatial resizing or channel projection, thereby reducing computational overhead and preventing possible distortion of the feature distribution during knowledge transfer.

Because the selected teacher and student feature maps have the same spatial resolution and channel dimension in the proposed configuration, no additional spatial resizing or channel projection is required. Let F_t and F_s denote the selected feature maps of the teacher and student networks, respectively. The feature-based distillation loss is calculated using mean squared error (MSE):

$$L_{FD} = \frac{1}{BCHW} \|F_s - F_t\|_2^2, \quad (16)$$

where B , C , H , and W denote the batch size, channel dimension, height, and width of the aligned feature maps, respectively.

Because the selected feature map of the student network is refined by SimAM, feature-based distillation directly supervises the attention-enhanced representation learned by the student model. This enables the student network to learn intermediate feature characteristics that are consistent with the teacher network while preserving the lightweight inference structure. The feature-based knowledge distillation process is illustrated in Figure 10.

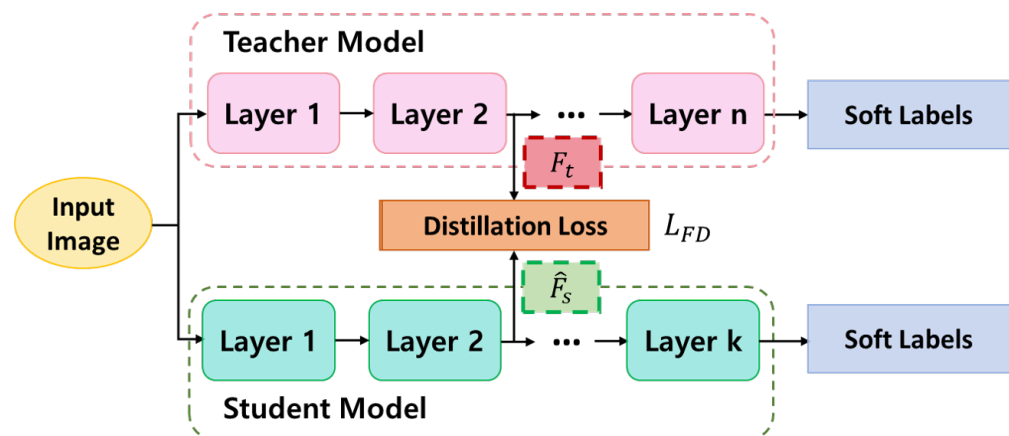


Figure 10. Feature-based knowledge distillation between the teacher and student networks.

3.4.3. Overall Training Objective

The final training objective combines hard-label supervision, logit-based distillation, and feature-based distillation:

$$L_{\text{Total}} = \alpha L_{KD} + \beta L_{CE} + \gamma L_{FD}, \quad (17)$$

where α , β , and γ control the contribution of the three loss terms. The temperature parameter T and the loss weights are specified in the experimental settings. The training pipeline of the proposed hybrid knowledge distillation framework is shown in Figure 11.

The overall hybrid knowledge distillation process is illustrated in Figure 12. The pre-trained and frozen teacher network and feature alignment operation are employed exclusively during training. During inference, only ShuffleNetV2-hSimKD is retained; therefore, knowledge distillation does not increase the inference time complexity of the proposed lightweight model.

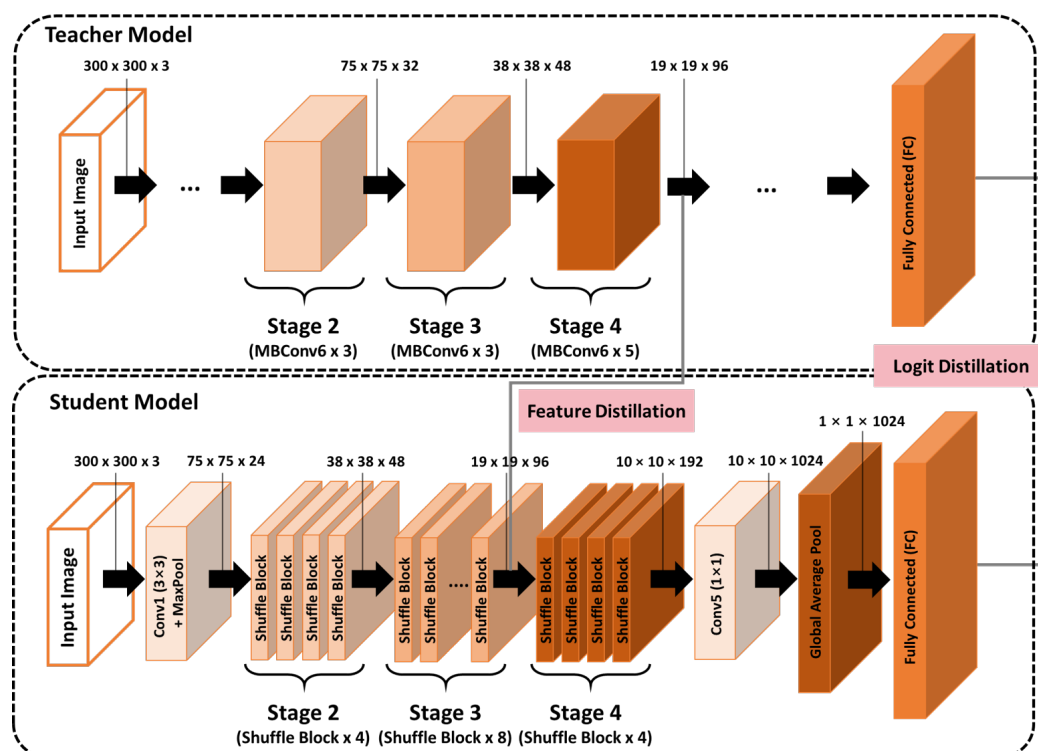


Figure 11. Training pipeline of the proposed hybrid knowledge distillation framework.

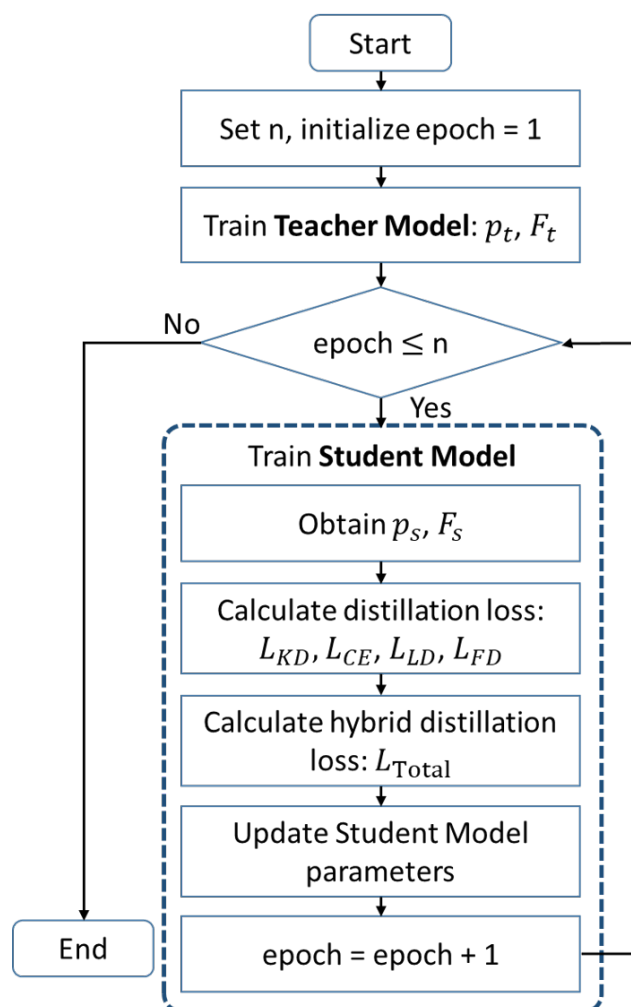


Figure 12. Overall training process of hybrid knowledge distillation.

4. Experiments

4.1. Dataset

The experiments were conducted using a binary version of the publicly available PlantVillage dataset [5]. The original PlantVillage dataset contains 54,306 leaf images collected from 14 crop species and organized into 38 classes, including healthy leaf classes and disease-related classes. For this study, the original classes were reorganized into a binary plant disease detection task. Classes containing the suffix “_healthy” were assigned to the healthy category, whereas all disease-related classes were assigned to the diseased category. This configuration enables the proposed model to distinguish between healthy and diseased leaf images.

To construct a balanced experimental dataset, 1000 healthy leaf images and 1000 diseased leaf images were randomly sampled from the original PlantVillage dataset. All images were resized to 300×300 pixels and normalized using ImageNet normalization statistics. During training, data augmentation was applied only to the training set. The augmentation procedures included grayscale conversion, random horizontal flipping, random vertical flipping, and random rotation. The validation and test sets were processed only through resizing and normalization. Representative examples of healthy and diseased leaf images are shown in Figure 13.



Figure 13. Representative healthy and diseased leaf images from the experimental dataset.

The balanced binary setting was adopted to reduce class imbalance bias and to focus on evaluating the healthy-versus-diseased discrimination capability of the proposed lightweight model. This setting can serve as a foundational first-stage screening task in smart agriculture pipelines, where abnormal leaf images are first separated from healthy ones before more detailed diagnosis is performed. The sampled subset preserves intra-class variations in leaf morphology and symptom severity, enabling a controlled evaluation of SimAM-based feature enhancement and knowledge distillation without confounding effects caused by severe class imbalance.

Nevertheless, the binary setting simplifies the broader plant disease recognition problem because it does not evaluate the model's ability to distinguish among specific disease categories. In addition, although PlantVillage is a widely used benchmark, its images are mainly captured under controlled conditions and may not fully represent real-world field variability. Therefore, the results should be interpreted as evidence of lightweight binary plant disease detection performance under a controlled benchmark setting rather than definitive validation of fine-grained field disease recognition. Moreover, the observed failure patterns, particularly for early-stage symptoms, indicate that visually subtle disease cases remain challenging even under controlled conditions.

4.2. Evaluation Metrics

The evaluated models were compared in terms of plant disease detection performance and computational efficiency. Detection performance was assessed using Accuracy, Precision, Recall, and F1 Score [36]. Computational efficiency was evaluated using the number of parameters, FLOPs, and average inference time per image.

For the binary detection task, diseased leaf images were treated as the positive class. Accuracy represents the proportion of correctly classified samples among all test samples:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%, \quad (18)$$

where TP , TN , FP , and FN denote true positives, true negatives, false positives, and false negatives, respectively.

Precision represents the proportion of correctly detected diseased leaf images among all images predicted as diseased:

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\%. \quad (19)$$

Recall represents the proportion of diseased leaf images correctly detected by the model:

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\%. \quad (20)$$

The F1 Score is the harmonic mean of Precision and Recall:

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\%. \quad (21)$$

The number of parameters indicates the storage complexity of a model, whereas FLOPs represent the estimated computational cost of a single forward pass. Both metrics are reported in millions (M).

In addition, inference time was measured to compare the practical execution efficiency of the evaluated models. The reported value represents the average execution time required for a single forward pass with a batch size of one after the input image had been prepared. To reduce the effect of initialization overhead and short-term timing variation, the reported inference time was obtained by averaging repeated inference runs on the same platform. Accordingly, the reported value should be interpreted as model-level inference time rather than end-to-end application latency.

4.3. Training Setup

Table 1 summarizes the training hyperparameters. All models were trained using the Adam optimizer [37] with an initial learning rate of 1×10^{-3} and a weight decay of 1×10^{-5} [38]. The batch size was set to 16, and all models were trained for 100 epochs.

Model training was performed on an NVIDIA GeForce RTX 4090 GPU with 24 GB memory under Windows 10 using Python 3.8 and PyTorch 1.13.1. Inference time was measured separately on an Intel Core i7-10750H processor using a batch size of one.

Table 1. Training hyperparameters.

Parameter	Setting
Optimizer	Adam
Initial learning rate	1×10^{-3}
Weight decay	1×10^{-5}
Batch size	16
Number of epochs	100

4.4. Teacher Model Selection for Knowledge Distillation

To select an appropriate teacher model for knowledge distillation, three representative CNN architectures were evaluated: ResNet50, Xception, and EfficientNetB3. These models were selected because they have been widely applied to image classification tasks and provide different trade-offs between predictive performance and computational complexity. The comparative results are summarized in Table 2. Among the candidate models, EfficientNetB3 achieved the highest Accuracy, Precision, Recall, and F1 Score while requiring fewer parameters and lower computational cost than ResNet50 and Xception. Therefore, EfficientNetB3 was selected as the teacher network for the proposed hybrid knowledge distillation framework. EfficientNetB3 is based on mobile inverted bottleneck convolution (MBConv) blocks and incorporates SE-based channel recalibration [39,40]. These characteristics enable the model to generate discriminative intermediate feature representations, which are beneficial for feature-based distillation.

Table 2. Performance comparison of candidate teacher models.

Model	Accuracy	Precision	Recall	F1 Score	Params (M)	FLOPs (M)
ResNet50	89.33%	87.70%	85.30%	87.50%	23.5	7810
Xception	91.68%	87.83%	86.46%	87.11%	22.9	8450
EfficientNetB3	94.67%	92.57%	90.45%	91.50%	10.8	1850

Regarding the training hyperparameter configuration, all models were trained using the Adam optimizer with an initial learning rate of 1×10^{-3} and a weight decay of 1×10^{-5} . To stabilize training and improve convergence, we adopted a cosine annealing learning rate schedule, in which the learning rate was gradually reduced from 1×10^{-3} to 1×10^{-5} over 100 training epochs. A linear warm-up phase of 5 epochs was also applied at the beginning of training. The batch size was set to 16, and the total number of training epochs was 100.

In the hybrid knowledge distillation process, EfficientNetB3 pre-trained on the ImageNet-1k dataset was employed as the teacher network. During the distillation training phase, all parameters of the teacher network were frozen and were not updated by backpropagation. Therefore, the teacher network provided a fixed and stable supervisory signal for training the ShuffleNetV2-hSim student network. For logit-based distillation, the temperature parameter was set to $T = 1$ according to the validation setting used in this study. Although this setting does not introduce additional temperature smoothing, it allows the student to learn from the teacher's probabilistic output distribution while preserving a sharp decision boundary for the binary classification task. For feature-based distillation, the Stage 4 feature map of the EfficientNetB3 teacher and the Stage 3 feature map of the ShuffleNetV2-hSim student were selected. Under the adopted input resolution, these feature maps have the same spatial resolution of 19×19 and the same channel di-

mension of 96, enabling direct feature alignment without additional bilinear interpolation, pooling, or channel projection. The detailed feature alignment procedure is described in Section 3.4.2. The hard-label supervision weight was fixed at $\beta = 1$, whereas the weights of the logit-based and feature-based distillation losses were adjusted to determine an appropriate balance among the loss components. Based on the validation results presented in Table 3, the final loss weights were set to $\alpha = 0.1$, $\beta = 1$, and $\gamma = 0.01$.

Table 3. Validation results for distillation weight selection.

α	β	γ	Accuracy	Precision	Recall	F1 Score
0.05	1	0.05	86.33%	85.70%	84.30%	86.50%
0.01	1	0.10	87.62%	84.53%	82.62%	84.23%
0.01	1	0.01	88.62%	87.21%	86.53%	86.53%
0.10	1	0.01	90.41%	88.19%	88.51%	88.31%
0.10	1	0.10	88.62%	86.33%	84.62%	87.53%

4.5. Results and Analysis

The proposed ShuffleNetV2-hSimKD model was compared with the baseline ShuffleNetV2, the intermediate ShuffleNetV2-hSim model, and representative lightweight architectures, including MobileNetV3, MobileViT, and EfficientNet-Lite. Figure 14 presents the validation accuracy curves over 100 epochs. All models showed stable convergence after approximately 40 epochs. ShuffleNetV2-hSimKD achieved the highest validation accuracy during the later stages of training, indicating that the proposed attention, activation, and distillation strategies improved the learning capability of the lightweight student network.

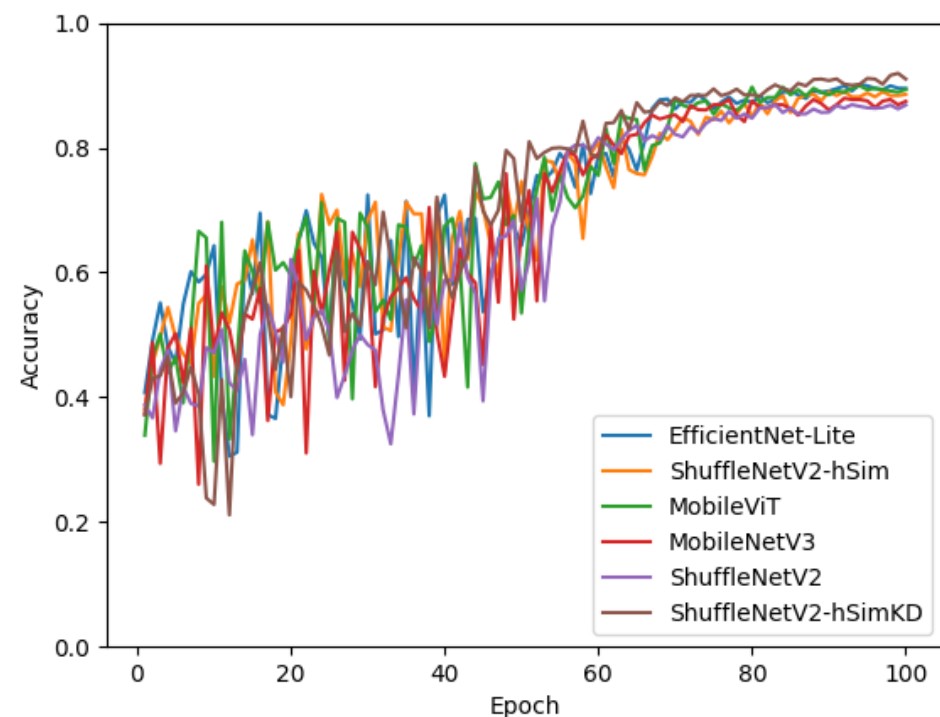


Figure 14. Validation accuracy curves of the compared models.

To evaluate the reproducibility of the observed performance improvements, all core experiments were repeated five times using different random seeds. As summarized in Table 4, ShuffleNetV2-hSimKD achieved mean values of $90.41 \pm 0.13\%$ for Accuracy, $88.19 \pm 0.19\%$ for Precision, $88.51 \pm 0.17\%$ for Recall, and $88.31 \pm 0.20\%$ for F1 Score. In comparison, the baseline ShuffleNetV2 achieved $86.84 \pm 0.14\%$ Accuracy and $83.01 \pm 0.18\%$ Recall.

These results indicate that the proposed model consistently outperformed the baseline across repeated runs, suggesting that the observed improvements are reproducible and not solely attributable to random training fluctuations. For the other comparative lightweight models, including MobileNetV3 and EfficientNet-Lite, we report the average performance obtained under the same experimental setup. ShuffleNetV2-hSimKD achieved the highest Accuracy, Recall, and F1 Score among the evaluated lightweight models while requiring only 1.4 M parameters and 151M FLOPs.

Table 4. Comparison with representative lightweight models.

Model	Accuracy	Precision	Recall	F1 Score	Params (M)	FLOPs (M)	Inference Time (ms/Image)
ShuffleNetV2-hSimKD	90.41 ± 0.13	88.19 ± 0.19	88.51 ± 0.17	88.31 ± 0.20	1.4	151	2.5
ShuffleNetV2-hSim	88.52 ± 0.24	87.78 ± 0.24	84.62 ± 0.24	86.10 ± 0.24	1.4	150	12.4
ShuffleNetV2	86.84 ± 0.14	85.10 ± 0.19	83.01 ± 0.18	84.01 ± 0.19	1.4	125	10.2
EfficientNet-Lite	89.53 ± 0.18	88.90 ± 0.18	85.20 ± 0.18	87.02 ± 0.18	4.7	391	22.1
MobileNetV3	87.42 ± 0.24	85.20 ± 0.24	81.51 ± 0.24	83.31 ± 0.24	2.9	225	15.3
MobileViT	89.33 ± 0.24	84.81 ± 0.24	84.70 ± 0.24	84.75 ± 0.24	1.3	338	19.8

Compared with the baseline ShuffleNetV2, ShuffleNetV2-hSimKD improved Accuracy by 3.57%, Recall by 5.5%, and F1 Score by 4.3%. Compared with ShuffleNetV2-hSim, the proposed distilled model improved Accuracy by 1.89% and Recall by 3.89% without changing the deployed student architecture. Although EfficientNet-Lite achieved competitive Accuracy, it required approximately 3.4 times more parameters and 2.6 times more FLOPs than ShuffleNetV2-hSimKD. MobileViT had a slightly smaller parameter count, but its FLOPs and inference time were substantially higher than those of the proposed model. These results indicate that ShuffleNetV2-hSimKD achieves an effective balance between plant disease detection performance and computational efficiency. To compare practical execution efficiency, the average inference time was measured using a batch size of one. As shown in Table 4, ShuffleNetV2-hSimKD required an average of 12.5 ms per image, whereas EfficientNet-Lite required 22.1 ms per image. This result indicates that the proposed model achieves lower model-level inference time than the compared larger lightweight models.

Compared with Ref. [19], which enhanced ShuffleNetV2 with the parameter-intensive SE attention module for edible fungi disease recognition, our model leverages parameter-free SimAM combined with h-swish to maintain hardware efficiency while improving feature representation. Unlike Ref. [24], which introduced SimAM to ShuffleNetV2 for maize leaf disease detection without activation optimization or distillation, our framework further integrates hybrid knowledge distillation to transfer rich teacher knowledge, resulting in a 3.57% accuracy improvement over the baseline under a comparable computational budget. These distinctions underscore the idea that our contribution lies not in isolated components but in their synergistic integration tailored for low-resource plant disease detection.

It should be noted that the reported measurement represents model inference time only. End-to-end execution time in an agricultural application may additionally depend on image acquisition, preprocessing, data transfer, postprocessing, and device-specific software overhead. Nevertheless, the low parameter count, computational cost, and inference time suggest that ShuffleNetV2-hSimKD holds promise for edge-oriented deployment, although real-world field validation remains warranted. The overall model generation process is illustrated in Figure 15.

For qualitative error analysis, we additionally report a representative confusion matrix from one held-out test run in Table 5. The model achieved 90.25% accuracy on this test set, with 20 false positives and 19 false negatives, corresponding to 39 misclassified samples in

total. This relatively balanced error distribution indicates that the model does not exhibit a severe bias toward either the healthy or diseased class.

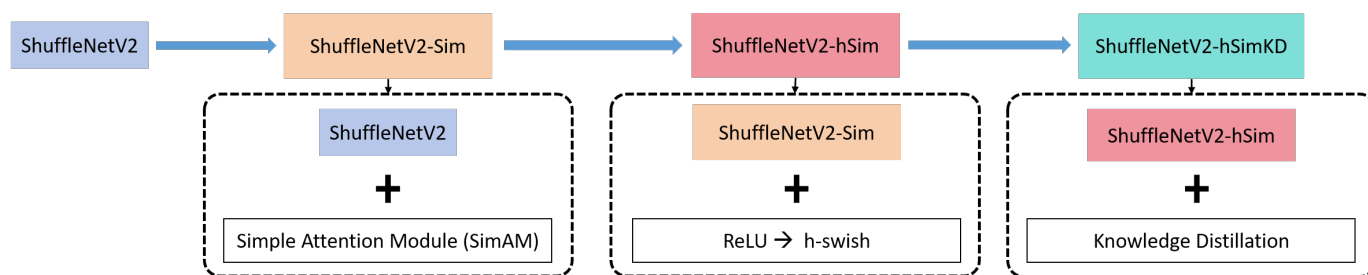


Figure 15. Overall construction process of the proposed ShuffleNetV2-hSimKD model.

Table 5. Confusion Matrix for ShuffleNetV2-hSimKD on the Test Set.

Actual \ Predicted	Healthy	Diseased	Total
Healthy	180 (TN)	20 (FP)	200
Diseased	19 (FN)	181 (TP)	200
Total	199	201	400

The FN cases mainly involve early-stage infections characterized by faint lesions occupying less than 5% of the leaf area. These subtle lesions often exhibit low contrast against surrounding healthy tissue, making them difficult for the attention mechanism to emphasize effectively. In contrast, the FP cases are mostly healthy leaves with natural physiological variations, such as yellowing, insect damage, or texture irregularities, that visually resemble disease symptoms. These observations suggest that, although the proposed model performs competitively under the evaluated binary PlantVillage setting, its sensitivity to incipient disease symptoms and robustness against natural leaf variations require further improvement. Future work will investigate contrast-enhancing preprocessing, enriched healthy-class samples, and uncertainty estimation to mitigate these errors. In addition, extending this analysis to the full 38-class PlantVillage dataset or field-oriented datasets would enable a more detailed investigation of disease-specific confusion patterns.

4.6. Ablation Study

An ablation study was conducted to evaluate the individual contributions of SimAM, h-swish, and knowledge distillation. Starting from the baseline ShuffleNetV2, SimAM and h-swish were progressively incorporated, followed by hybrid knowledge distillation. The results are summarized in Table 6.

Table 6. Ablation study of the proposed components.

Model	SimAM	h-swish	KD	Accuracy	Recall	FLOPs (M)	Inference Time (ms/Image)
ShuffleNetV2	No	No	No	86.84	83.01	125	10.2
ShuffleNetV2-Sim	Yes	No	No	87.85	84.08	126	11.1
ShuffleNetV2-hSim	Yes	Yes	No	88.52	84.62	150	12.4
ShuffleNetV2-hSimKD	Yes	Yes	Yes	90.41	88.51	151	2.5

First, the addition of SimAM improved Accuracy from 86.84% to 87.85%, corresponding to an improvement of 1.01%. Recall also increased from 83.01% to 84.08%. This result indicates that the parameter-free attention mechanism improved the ability of ShuffleNetV2 to emphasize disease-related feature responses while introducing only a small increase in computational cost. Second, replacing ReLU with h-swish further improved Accuracy

from 87.85% to 88.52% and Recall from 84.08% to 84.62%. Although h-swish increased the computational cost of the student network, the resulting ShuffleNetV2-hSim architecture achieved improved plant disease detection performance while retaining a low parameter count. Finally, hybrid knowledge distillation improved Accuracy from 88.52% to 90.41% and Recall from 84.62% to 88.51%. The improvement was obtained without changing the deployed student architecture because the teacher network and feature-level supervision were used only during training. Therefore, ShuffleNetV2-hSimKD retained the same average inference time as ShuffleNetV2-hSim.

Overall, the ablation results demonstrate that SimAM, h-swish, and hybrid knowledge distillation provide complementary benefits. SimAM improves disease-related feature refinement, h-swish improves nonlinear feature representation, and knowledge distillation transfers additional supervisory information from the EfficientNetB3 teacher network. Their integration produces the strongest overall performance while maintaining the computational efficiency of a lightweight CNN.

5. Conclusions

This paper proposed ShuffleNetV2-hSimKD, a lightweight network for plant disease detection. The proposed model progressively enhances the ShuffleNetV2 backbone by integrating the parameter-free SimAM attention mechanism, replacing the original ReLU activation with h-swish, and applying hybrid knowledge distillation during training. SimAM improves disease-related feature refinement without introducing additional trainable parameters, whereas h-swish enhances nonlinear feature representation with limited computational overhead. In addition, the hybrid knowledge distillation framework transfers both output-level and intermediate feature-level knowledge from an EfficientNetB3 teacher network to the lightweight ShuffleNetV2-hSim student network.

Experiments were conducted on a balanced binary subset of the PlantVillage dataset consisting of healthy and diseased leaf images. The proposed ShuffleNetV2-hSimKD achieved an Accuracy of 90.41%, a Recall of 88.51%, and an F1 Score of 88.31%, while requiring only 1.4 M parameters and 151 M FLOPs. Compared with the baseline ShuffleNetV2, the proposed model improved Accuracy by 3.57 percentage points and Recall by 5.50 percentage points. The results from five independent runs with different random seeds further indicate that the proposed model provides stable and reproducible performance improvements over the baseline. The ablation results also demonstrated that SimAM, h-swish, and hybrid knowledge distillation provide complementary benefits. In particular, knowledge distillation improved detection performance without increasing the inference complexity of the deployed student network. The proposed model also achieved an average inference time of 12.5 ms per image, indicating a favorable balance between plant disease detection performance and computational efficiency.

Despite its promising performance, this study has several limitations. Although the PlantVillage dataset is a widely used benchmark, it is primarily composed of images captured under controlled conditions with relatively uniform backgrounds. Consequently, the reported performance may not fully extrapolate to real-world field conditions characterized by complex backgrounds, variable illumination, occlusion, overlapping leaves, and device-dependent image quality. Furthermore, this study focuses on a binary classification task, namely healthy-versus-diseased classification, which simplifies the challenge of differentiating among multiple specific disease types. These limitations regarding dataset heterogeneity and task scope should be considered when interpreting the generalizability of the results. Regarding deployment, inference time was measured on a desktop CPU to provide a model-level efficiency assessment; however, comprehensive validation on dedicated agricultural edge hardware, such as Raspberry Pi or Jetson Nano platforms, remains necessary.

To address these limitations, future work will focus on three main directions. First, the proposed framework will be extended to multi-class plant disease recognition using the full PlantVillage dataset or more challenging datasets such as PlantDoc to assess performance across diverse disease categories and imaging conditions [41]. Second, rigorous on-device benchmarking will be conducted on embedded platforms to quantify latency, memory footprint, and energy consumption during actual deployment. Third, misclassification analysis and uncertainty estimation will be incorporated to improve model reliability in heterogeneous field environments. In addition, domain adaptation techniques will be investigated to reduce the gap between controlled laboratory images and unstructured field data, thereby facilitating the transition from proof-of-concept evaluation to practical smart agriculture applications.

Author Contributions: Conceptualization, Q.S. and S.I.H.; methodology, Y.S.; software, Q.S.; validation, Q.S., Y.S. and S.I.H.; formal analysis, Q.S.; investigation, Y.S.; writing—original draft preparation, Q.S.; writing—review and editing, S.I.H.; visualization, Y.S.; supervision, S.I.H.; project administration, S.I.H.; funding acquisition, S.I.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Soonchunhyang University Research Fund.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article..

Acknowledgments: During the preparation of this manuscript, the authors used ChatGPT-5.6 Sol (OpenAI) solely for English language editing. The authors reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ma, C.-S.; Wang, B.-X.; Wang, X.-J.; Lin, Q.-C.; Zhang, W.; Yang, X.-F.; van Baaren, J.; Bebbler, D.P.; Eigenbrode, S.D.; Zalucki, M.P.; et al. Crop pest responses to global changes in climate and land management. *Nat. Rev. Earth Environ.* **2025**, *6*, 264–283. [\[CrossRef\]](#)
2. Upadhyay, A.; Chandel, N.S.; Singh, K.P.; Chakraborty, S.K.; Nandede, B.M.; Kumar, M.; Subeesh, A.; Upendar, K.; Salem, A.; Elbeltagi, A. Deep learning and computer vision in plant disease detection: A comprehensive review of techniques, models, and trends in precision agriculture. *Artif. Intell. Rev.* **2025**, *58*, 92. [\[CrossRef\]](#)
3. Padhiary, M.; Saha, D.; Kumar, R.; Sethi, L.N.; Kumar, A. Enhancing precision agriculture: A comprehensive review of machine learning and AI vision applications in all-terrain vehicle for farm automation. *Smart Agric. Technol.* **2024**, *8*, 100483. [\[CrossRef\]](#)
4. Gamage, A.; Gangahagedara, R.; Subasinghe, S.; Gamage, J.; Guruge, C.; Senaratne, S.; Randika, T.; Rathnayake, C.; Hameed, Z.; Madhujith, T.; et al. Advancing sustainability: The impact of emerging technologies in agriculture. *Curr. Plant Biol.* **2024**, *40*, 100420. [\[CrossRef\]](#)
5. Pacal, I.; Kunduracioglu, I.; Alma, M.H.; Deveci, M.; Kadry, S.; Nedoma, J.; Slany, V.; Martinek, R. A systematic review of deep learning techniques for plant diseases. *Artif. Intell. Rev.* **2024**, *57*, 304. [\[CrossRef\]](#)
6. Tan, L.; Lu, J.; Jiang, H. Tomato leaf diseases classification based on leaf images: A comparison between classical machine learning and deep learning methods. *AgriEngineering* **2021**, *3*, 542–558. [\[CrossRef\]](#)
7. Saleem, M.H.; Potgieter, J.; Arif, K.M. Plant disease classification: A comparative evaluation of convolutional neural networks and deep learning optimizers. *Plants* **2020**, *9*, 1319. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
9. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning*; PMLR: Long Beach, CA, USA, 2019; pp. 6105–6114.
10. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16 × 16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
11. Abetu, M.T.; Abebe, T.U.; Tune, K.K. Efficient and Robust Deep Learning for Field-Level Wheat Disease Classification on Resource-Constrained Devices. *Smart Agric. Technol.* **2026**, *14*, 102091. [\[CrossRef\]](#)

12. Majdalawieh, M.; Martins, C.; Radi, M.; Alaraj, M.; Khan, S. Precision agriculture in the age of AI: A systematic review of machine learning methods for crop disease detection. *Smart Agric. Technol.* **2025**, *12*, 101491. [\[CrossRef\]](#)
13. Abiri, R.; Rizan, N.; Balasundram, S.K.; Shahbazi, A.B.; Abdul-Hamid, H. Application of digital technologies for ensuring agricultural productivity. *Heliyon* **2023**, *9*, e22601. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Howard, A.; Sandler, M.; Chu, G.; Chen, L.-C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; et al. Searching for MobileNetV3. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1314–1324.
15. Ma, N.; Zhang, X.; Zheng, H.-T.; Sun, J. Shufflenet v2: Practical guidelines for efficient cnn architecture design. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 116–131.
16. Koonce, B. EfficientNet. In *Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization*; Apress: Berkeley, CA, USA, 2021; pp. 109–123.
17. Si, Y.; Xu, H.; Zhu, X.; Zhang, W.; Dong, Y.; Chen, Y.; Li, H. SCSA: Exploring the synergistic effects between spatial and channel attention. *Neurocomputing* **2025**, *634*, 129866. [\[CrossRef\]](#)
18. Xu, W.; Wan, Y.; Zhao, W. ELA: Efficient location attention for deep convolution neural networks. *J. Real-Time Image Process.* **2025**, *22*, 140. [\[CrossRef\]](#)
19. Xu, X.; Zhang, Y.; Cao, H.; Yang, D.; Zhou, L.; Yu, H. Recognition of edible fungi fruit body diseases based on improved ShuffleNetV2. *Agronomy* **2023**, *13*, 1530. [\[CrossRef\]](#)
20. Rashid, R.; Aslam, W.; Aziz, R.; Aldehim, G. A modified mobileNetv3 coupled with inverted residual and channel attention mechanisms for detection of tomato leaf diseases. *IEEE Access* **2025**, *13*, 52683–52696. [\[CrossRef\]](#)
21. Jin, X.; Xie, Y.; Wei, X.-S.; Zhao, B.-R.; Chen, Z.-M.; Tan, X. Delving deep into spatial pooling for squeeze-and-excitation networks. *Pattern Recognit.* **2022**, *121*, 108159. [\[CrossRef\]](#)
22. Sapkota, R.; Flores-Calero, M.; Qureshi, R.; Badgujar, C.; Nepal, U.; Poullose, A.; Zeno, P.; Vaddevolu, U.B.P.; Khan, S.; Shoman, M.; et al. YOLO advances to its genesis: A decadal and comprehensive review of the You Only Look Once (YOLO) series. *Artif. Intell. Rev.* **2025**, *58*, 274. [\[CrossRef\]](#)
23. Yang, L.; Zhang, R.-Y.; Li, L.; Xie, X. SimAM: A simple, parameter-free attention module for convolutional neural networks. In Proceedings of the 38th International Conference on Machine Learning, PMLR, Virtual Event, 18–24 July 2021; pp. 11863–11874.
24. Zhou, H.; Su, Y.; Chen, J.; Li, J.; Ma, L.; Liu, X.; Lu, S.; Wu, Q. Maize leaf disease recognition based on improved convolutional neural network ShuffleNetV2. *Plants* **2024**, *13*, 1621. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Xu, J.; Jia, D.; Lin, Z.; Zhou, T. PSFNet: A Deep Learning Network for Fake Passport Detection. *IEEE Access* **2022**, *10*, 123337–123348. [\[CrossRef\]](#)
26. Qu, G.; Chen, Q.; Wei, W.; Lin, Z.; Chen, X.; Huang, K. Mobile edge intelligence for large language models: A contemporary survey. *IEEE Commun. Surv. Tutor.* **2025**, *27*, 3820–3860. [\[CrossRef\]](#)
27. Cong, S.; Zhou, Y. A review of convolutional neural network architectures and their optimizations. *Artif. Intell. Rev.* **2023**, *56*, 1905–1969. [\[CrossRef\]](#)
28. Gou, J.; Yu, B.; Maybank, S.J.; Tao, D. Knowledge distillation: A survey. *Int. J. Comput. Vis.* **2021**, *129*, 1789–1819. [\[CrossRef\]](#)
29. Zhu, S.; Gao, H. MC-ShuffleNetV2: A lightweight model for maize disease recognition. *Egypt. Inform. J.* **2024**, *27*, 100503. [\[CrossRef\]](#)
30. Chattopadhyay, S.; Dey, A.; Singh, P.K.; Sarkar, R. DRDA-Net: Dense residual dual-shuffle attention network for breast cancer classification using histopathological images. *Comput. Biol. Med.* **2022**, *145*, 105437. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Glegoła, W.; Karpus, A.; Przybyłek, A. MobileNet family tailored for Raspberry Pi. *Procedia Comput. Sci.* **2021**, *192*, 2249–2258. [\[CrossRef\]](#)
32. Chen, L.; Li, S.; Bai, Q.; Yang, J.; Jiang, S.; Miao, Y. Review of image classification algorithms based on convolutional neural networks. *Remote Sens.* **2021**, *13*, 4712. [\[CrossRef\]](#)
33. Rao, J.; Meng, X.; Ding, L.; Qi, S.; Liu, X.; Zhang, M.; Tao, D. Parameter-efficient and student-friendly knowledge distillation. *IEEE Trans. Multimed.* **2023**, *26*, 4230–4241. [\[CrossRef\]](#)
34. Wang, L.; Yoon, K.-J. Knowledge distillation and student-teacher learning for visual intelligence: A review and new outlooks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 3048–3068. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Ye, F.; Bors, A.G. Dynamic self-supervised teacher-student network learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 5731–5748. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Si, Q.; Han, S.I. RepVGG-MEM: A Lightweight Model for Garbage Classification Achieving a Balance Between Accuracy and Speed. *IEEE Access* **2025**, *13*, 36451–36469. [\[CrossRef\]](#)
37. Reyad, M.; Sarhan, A.M.; Arafa, M. A modified Adam algorithm for deep neural network optimization. *Neural Comput. Appl.* **2023**, *35*, 17095–17112. [\[CrossRef\]](#)
38. Wang, Y.; Xiao, Z.; Cao, G. A convolutional neural network method based on Adam optimizer with power-exponential learning rate for bearing fault diagnosis. *J. Vibroeng.* **2022**, *24*, 666–678. [\[CrossRef\]](#)

39. Batool, A.; Byun, Y.-C. Lightweight EfficientNetB3 model based on depthwise separable convolutions for enhancing classification of leukemia white blood cell images. *IEEE Access* **2023**, *11*, 37203–37215. [[CrossRef](#)]
40. Alhichri, H.; Alswayed, A.S.; Bazi, Y.; Ammour, N.; Alajlan, N.A. Classification of remote sensing images using EfficientNet-B3 CNN model with attention. *IEEE Access* **2021**, *9*, 14078–14094. [[CrossRef](#)]
41. Krishna, M.S.; Machado, P.; Otuka, R.I.; Yahaya, S.W.; Neves dos Santos, F.; Ihianle, I.K. Plant leaf disease detection using deep learning: A multi-dataset approach. *J* **2025**, *8*, 4. [[CrossRef](#)]

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