IET The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH PAPER

Identification of plant disease images via a squeeze-and-excitation MobileNet model and twice transfer learning

Junde Chen[1] | **Defu Zhang**[1] | **Md Suzauddola**[1] | **Yaser Ahangari Nanehkaran**[1] | **Yuandong Sun**[2]

[1] School of Informatics, Xiamen University, Xiamen, China

[2] Spira Inc, Los Angeles, California, USA

Correspondence
Defu Zhang, School of Informatics, Xiamen University, Xiamen 361005, China.
Email: dfzhang@xmu.edu.cn

Abstract

Crop diseases have a devastating effect on agricultural production, and serious diseases can lead to harvest failure entirely. Recent developments in deep learning have greatly improved the accuracy of image identification. In this study, we investigated the transfer learning of deep convolutional neural networks and modified the network structure to improve the learning capability of plant lesion characteristics. The MobileNet with squeeze-and-excitation (SE) block was selected in our approach. Integrating the merits of both, the pre-trained MobileNet and SE block were fused to form a new network, which we termed the SE-MobileNet, and was used to identify the plant diseases. In particular, the transfer learning was performed twice to obtain the optimum model. The first phase trained the model for the extended layers while the bottom convolution layers were frozen with the pre-trained weights on ImageNet; by loading the model trained in the first phase, the second phase retrained the model using the target dataset. The proposed procedure provides a significant increase in efficiency relative to other state-of-the-art methods. It reaches an average accuracy of 99.78% in the public dataset with clear backdrops. Even under multiple classes and heterogeneous background conditions, the average accuracy realises 99.33% for identifying the rice disease types. The experimental findings show the feasibility and effectiveness of the proposed procedure.

1 | INTRODUCTION

Food, cloth, housing, transportation, and education are the basic needs of every human being. All these five are essential for a healthy life. Among them, food security also gets the most priority. To secure a riskless food supply, we should focus on smoothly growing crops. Different types of plant diseases are the main threat to the spontaneous food supply [1]. Notably, the main food crops such as rice, which are the primary staple for over half the earth's population, are very important for ensuring enough food supply and security. Timely prediction and warning play a critical role in the prevention and control of plant diseases. It can reduce the costs of plant diseases and the unnecessary usages of pesticides. Therefore, obtaining the information of real-time plant diseases is highly desired in modern agriculture, and the machine learning-based image processing techniques alleviate the challenge of identifying plant diseases

in time before large-scale outbreaks [2]. Until now, however, plant disease monitoring in many regions, especially in developing countries, still primarily relies on the conventional manual approach. When a plant disease outbreak occurs somewhere, plant disease specialists or technicians from different agriculture research institutions appointed by the government will visit the place and advise the farmers. Nevertheless, compared with the number of farmers, there are not adequate plant disease specialists or technicians in many areas. Particularly, affected by the social and economic environment, few people are engaged in plant protection formerly. Thus, there are high needs and important realistic significances for seeking fast, readily available, and less expensive tools to identify plant diseases automatically.

Computer vision technology is an attractive alternative approach for the task of automatic crop observation since it is low-cost, visualised, and operates in a non-contact manner [3]. Much recent literature has concerned the image

recognition, and a distinctive classifier was employed to identify the plants into normal or different disease types [4–9]. Yao et al. [6] adopted the support vector machine (SVM) for the identification of rice diseases, and the SVM method was used to classify three types of rice diseases, including bacterial leaf blight, rice sheath blight, and rice blast. They classified these disease spots to an accuracy of 97.2%. Kahar et al. [7] used the artificial neural network (ANN) to identify rice plant diseases such as rice leaf blast, rice leaf blight, and sheath blight, and they achieved the accuracy of 74.21%. Kai et al. [8] applied a radial basis kernel function in SVM to identify the maize diseases. In their study, 262 images were divided as training and test sets for the samples. They identified three categories and realised the best accuracy of 89.6%. Zhang et al [9] adopted a genetic algorithm (GA)-based SVM for the identification of maize diseases and their average accuracy was at a peak of 90.25%. Recently, developments in deep-learning-based convolutional neural networks (CNN), which have greatly improved the accuracy of image identification, has become the leading methods for overcoming the challenges of digital image processing [10]. Mohanty et al. [11] proposed a deep CNN to identify 14 species and 26 plant diseases (or absence thereof). On a held-out test set, their trained model achieved an accuracy of 99.35%. Kawasaki et al. [12] introduced a CNN-based system to identify cucumber leaf diseases, and they reached an accuracy of 94.9%. Zhang et al. [13] proposed an enhanced deep CNN to identify the maize plant diseases, achieving an average accuracy of 98.8%. Though reasonably good findings were published in the literature, limited diversity of image databases has been used for research. Most of the photographic resources include images only in experimental (research lab) environments and rarely in the area of agriculture under practical wild scenarios [10]. Additionally, most studies were performed to identify the plant diseases based on leaf images instead of various parts. In reality, the images were usually taken under a variety of environments and the plant diseases may occur in any part of the plants, whether it is the leaf, stem, grain and so forth. Despite these limitations, the recent research successfully indicated the potential of deep-learning-based CNN algorithms to identify the plant diseases. In this study, we performed transfer learning the deep CNN and varied the network structure to increase the learning capability of plant lesion characteristics. The SE block [14] was added behind the pre-trained MobileNet [15] on ImageNet [16], which was followed by an additional $3 \times 3 \times 512$ convolutional layer for high-dimensional feature extraction. Then, the full connection layer was substituted for a global average pooling layer, and a new completely linked ReLU layer with 1024 neutrons was introduced into the network to enhance the learning ability. After that, the top layer comprised of a completely linked Softmax layer with the actual category number was ultimately used for the classification. Finally, considering the multi-classification tasks and different loss weights of positive and negative samples, an enhanced Focal-loss function was used in the network instead of the traditional cross-entropy loss function. Thus, the pre-trained MobileNet paired with SE block was fused to form a new network, namely, SE-MobileNet, which was utilised to identify the plant diseases.

Particularly, the transfer learning was performed twice for model training: In the first phase, only the parameters of new extended layers were inferred from scratch while the weights of convolutional layers pre-trained from ImageNet were frozen; the second phase retrained the weight parameters on the target dataset by loading the pre-trained model in the first phase. In this way, the yielded optimum model was used to identify crop diseases.

The rest of the study is organised as follows. In Section 2, following an introduction of the image dataset collection, a general overview is described, and this section primarily discusses the methodology along with related studies. Later in Section 3, experiments for investigating the efficiency of the proposed procedure are presented; multiple experiments are performed in this section and the results are evaluated through comparative analysis. This study is ultimately concluded in Section 4.

2 | MATERIALS AND METHODS

2.1 | Image dataset

In this study, we have used two datasets, including the public PlantVillage [17] and our collected rice disease image dataset, to perform the experiments.

PlantVillage is an international general image dataset used for the machine-learning algorithm test of plant disease identification. It has 38 categories that are composed of 26 diseases and 12 healthy plants for 14 plant species. PlantVillage contains 54,306 plant leaf images, which are captured under controlled conditions, both in grayscale and colour. All images are photographed in simple backdrops and uniform lighting strengths. Note that the number of samples in each category is not consistent. Besides, some of the different images are taken from the same plant leaf in diverse directions.

Approximately 600 rice plant disease images were captured from real-life agricultural fields. These collected natural images were all tagged with labels based on the domain experts' knowledge. The Fujian Institute of Subtropical Botany played a significant helping role, and the images of the plants used in this study were all in JPG format. The context conditions of rice plant images are complicated, and the lighting strengths are inconsistent. For later computations, photoshop software first processes these images evenly into the Red-Green-Blue (RGB) pattern and then adjusts the scale of the images. The rice disease types mainly comprise rice stackburn, leaf smut, leaf scald, white tip, bacterial leaf streak and so forth, and the key features of some rice disease have been described in Table 1. Partial sample images are also presented in Figure 1.

2.2 | Overview

As shown in Figure 2, a brief description of our approach is provided as follows. First, the sample images of plant diseases are captured and tagged based on the domain expert's knowledge.

TABLE 1 The key features of rice plant diseases

| | The disease characteristics | | |
Disease name	Shape	Colour	Location
Rice stackburn	Circular or ellipse	Yellowish-white macular	Leaf
Leaf smut	Short strip	Black	Leaf
Leaf scald	Watery stain form or strip	Offwhite or yellow	Leaf
White tip	Strip	Centre: Ashen; edge: Brown	Leaf tip
Rice stem rot	Long strip (stem)	Black	Stem
Rice straighthead	Crescent-like shape	Darker green	Grain
Rice brown spot	Ellipse of sesame seed size	Centre: Brown to ashen; periphery: Dim	Leaf
Rice ear blight	Imperfect, upright ears	Dark brown	Rice ears
Rice stripe blight	The strip parallel with vein	Yellow	Leaf
Rice sheath spot	Strip (sheath)	Centre: Off-white to brown	Stem
Rice sheath rot	Long strip (sheath)	Brown	Sheath
Bacterial leaf streak	Long strip	Brown or russet	Leaf
Rice false smut	Spores (grain)	Black	Rice ears
Rice blast	Circular punctate	Brown	Leaf, Stem
Rice sheath blight	Clouding form	Centre: Ashen; edge: Brown	Leaf
Grain spotting and peck	Circular (grain)	Centre: Off-white to brown	Grain
Rice kernel smut	Spores (kernel)	Black	Grain

(a) Rice stackburn (b) Rice leaf smut (c) Rice leaf scald (d) Rice blast
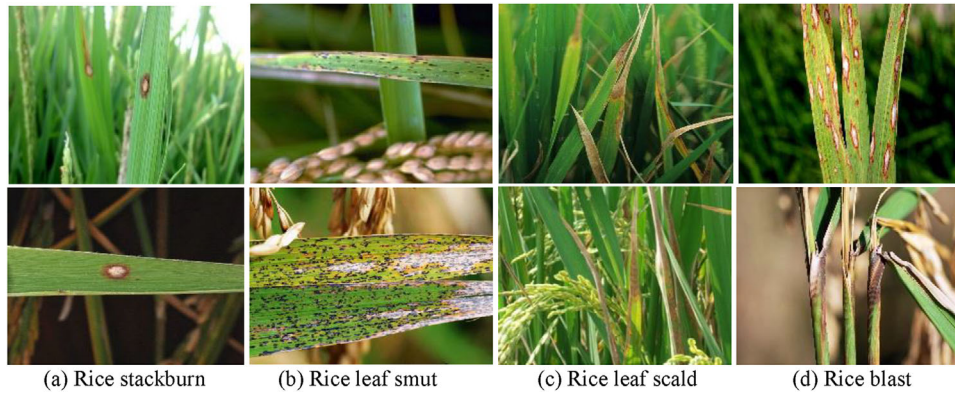
FIGURE 1 Examples of rice disease images (a) rice stackburn, (b) rice leaf smut, (c) rice leaf scald, (d) rice blast

The image pre-processing techniques such as image sharpening, filtering, resizing, and edge-filling are conducted on the collected images. Then, the image dataset is enriched using the data augmentation scheme; both the enhanced generative adversarial networks (GAN) [18] and the traditional methods, including colour-jittering, rotation, translation, and scale transformation, are implemented to synthesise the new images. Lastly, the pre-processed images are input into the SE-MobileNet for training the model, and the generated optimum model is employed to predict the types of plant diseases. In this way, the final identification results can be obtained accordingly, and the specific illustration of this phase is presented in subsequent sections.

2.3 | Deep neural network architecture

2.3.1 | Depthwise separable convolution (DWSC)

DWSC is a form of factorable convolution which splits a standard convolution into two steps: A depthwise convolution (DWC) and a 1 × 1 convolution namely pointwise convolution (PWC) [19]. In the first step, each channel is conducted with the convolutional operation for the intermediate feature map with one filter, as depicted in Figure 3(a). Where D_f is input length and width, M is the number of channel input and the number of convolution kernels. The length and width of the output are
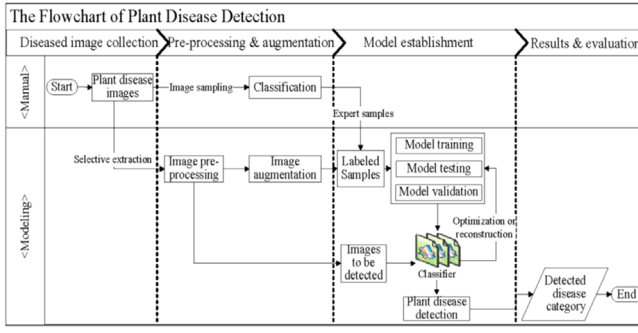
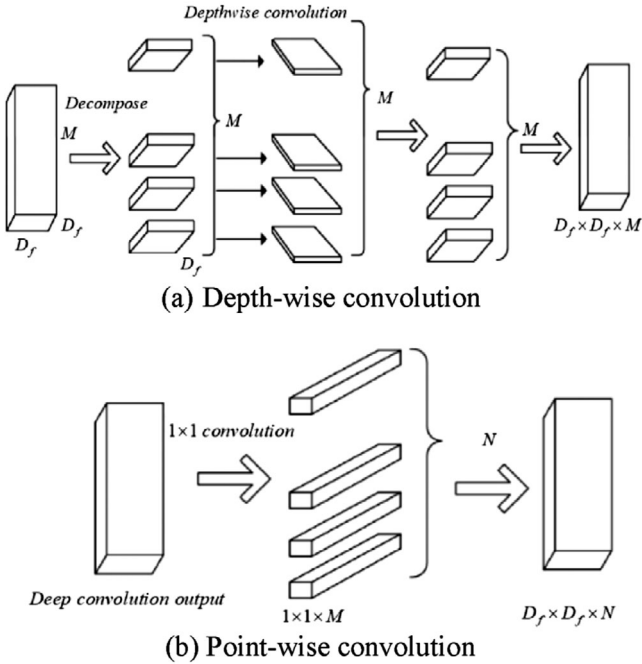**FIGURE 2** The flowchart of plant disease identification

FIGURE 3 The computation process of DWSC (a) DWC, (b) PWC

the same as the input. In the second step, PWC uses the results of DWC to conduct a 1 × 1 convolution kernel operation, as shown in Figure 3(b). Where M is the number of input channels, D_f is input length and width, convolution kernel size is 1 × 1, and N is the number of output channels. Mathematically, these formulas of the above processes can be separately defined in Equations (1) and (2).

$$DWC(K, F)_{i,j} = \sum_{h,n,d}^{H,W,D} K_{h,n,d} \times F_{h+i,w+j,d}^{l} \tag{1}$$

$$PWC(K, F)_{i,j} = \sum_{d}^{D} K_d \times \sum_{h}^{H} \sum_{w}^{W} K_{h,w} \times F_{h+i,w+j}^{l} \tag{2}$$

where F represents an intermediate feature map tensor, K denotes the filter (convolutional kernel), K_d is a 1 × 1 convolution kernel, and (i, j) index position of the feature map; H, W, and D are the height, width, and depth, respectively, of
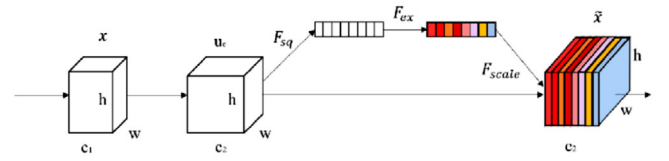

FIGURE 4 The squeeze-and-excitation (SE) building block [14]

the input for layer l in the network. DWSC has been explicitly incorporated in MobileNets [15], which are a family of mobile-first models considering the resource constraints for portable devices. Setting the size of input feature map as $D_f \times D_f$, the number of channels as M, the size of the convolutional kernel as $D_k \times D_k$, and the number of output channels as N, the computation cost of standard convolution (C_1) and the DWSC (C_2) can be expressed in Equations (3) and (4), respectively.

$$C_1 = D_k^2 NMD_f^2 \tag{3}$$

$$C_2 = \left(D_k^2 + N\right) D_f^2 M \tag{4}$$

As a result, the ratio of the computational cost for the DWSC to the standard convolution is

$$R_{C_2/C_1} = \frac{1}{D_k^2} + \frac{1}{N} \tag{5}$$

Particularly, the 3 × 3 convolution kernel is often used for DWSC networks in practical application scenarios, whereas the number of output channels N is big. Therefore, the computational cost of DWSC is approximately 1/9 of that of the standard convolution network.

2.3.2 | SE network

To improve the representation quality of a network by explicitly modelling the interdependencies between the channels of convolutional features, the SE network was first introduced by Hu et al. [14] in 2017, and it won the ImageNet challenge competition with outstanding performance. The core component of the SE network is the SE block, which primarily consists of two operations: Squeeze and excitation. The squeeze operation is responsible for shrinking feature maps $U \in R^{w \times h \times c2}$ through spatial dimensions ($w \times h$), which is achieved by using global average pooling to generate channelwise statistics as the F_{sq} depicted in Figure 4. Here, x denotes the input images, $U \in R^{w \times h \times c2}$ is the feature maps by a given transformation, for example, a standard convolution. F_{sq} is the squeeze operation and F_{ex} represents the excitation operation. $Fscale$ denotes a scale operation to weight the features of each channel. Subsequently, each feature map is represented by a statistic z as calculated in Equation (6).

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i, j) \tag{6}$$

where z_c is the cth statistic value, u_c represents the cth feature map of previous convolution operation, W and H denote the width and height of u_c separately. The excitation operation is to fully capture the dependencies of channels by making use of the information aggregated in the squeeze, where a simple gating mechanism is adopted with a sigmoid activation:

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z)) \quad (7)$$

Here, δ refers to the rectified linear unit (ReLU) [20] function, $W_1 \in R^{\frac{c}{r} \times C}$ $W_1 \in R^{c/r \times C}$ and $W_2 \in R^{C \times c/r}$; r is a reduction ration initialised with a constant. The parameters W_1 and W_2 are learned by two fully connected layers around the non-linearity.

Thus, the output of the SE block can be obtained by rescaling u_c with the activations s:

$$\tilde{x}_c = F_{scale}(u_c, s_c) = s_c \cdot u_c \quad (8)$$

where $[\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_c]$ devotes the final output $\tilde{X}$.

2.3.3 | Transfer learning principle

Transfer learning is the technique by which the knowledge gained during training in one type of problem is used to train in other related tasks or domains [21, 22]. Giving a source domain D_s with a corresponding source task T_s and a target domain D_t with a learning task T_t, the objective of transfer learning is to improve the performance of predictive function $f_T(\bullet)$ for learning task T_t by discovering and transferring latent knowledge from D_s and T_s, where $D_s \neq D_t$ or $T_s \neq T_t$ [22]. Transfer learning is particularly important in CNNs. The algorithms of deep learning require a large number of labelled data to train the models, and collecting numerous labelled data in a domain is undoubtedly expensive, time-consuming, and labour-intensive. Thus, the main challenge is that the training images are often not enough to adjust all the parameters without incurring losses in overfitting, and the scheme of transfer learning is naturally employed to address this type of problem.

2.3.4 | SE-MobileNet network mode

MobileNet is a variety of lightweight CNNs based on depth-wise separable convolution and has shown outstanding capability in processing large-scale and small-scale problems [19, 23]. Besides, the SE block makes good use of the channel relationship and can perform dynamic channelwise feature recalibration. Thus, inspired by the performance, the MobileNet paired with SE block was selected in our approach, and the pre-trained MobileNet with SE block was fused to generate a new network, which we termed the SE-MobileNet, which was utilised to identify the plant disease types. The network structure of traditional MobileNet was modified to enhance the capability of learning the tiny disease spot features for plant disease images. The original classification layer at the tail of MobileNet was abandoned

and the SE block was added behind the pre-trained MobileNet, which was followed by an additional $3 \times 3 \times 512$ convolutional layer for high-dimensional feature extraction. Then, the full connection layer was replaced using a global average pooling layer, and a new completely linked layer with 1024 neutrons and ReLU activation function was introduced to enhance the ability to identify tiny disease spot features. Ultimately, a top layer comprised of a completely linked Softmax layer with the actual category number was used for the classification. More importantly, the twice transfer learning was performed for model training. In the first step, only the parameters of newly extended layers were learned from scratch while the bottom convolutional layers were kept frozen with the pre-trained weights on ImageNet. The second step fine-tuned all the weight using the target dataset by loading the model pre-trained in the first phase. Concretely, the model was obtained by conducting the processes below.

1. An adaption was conducted for the network. To address new identification tasks, the adaption was accomplished through freezing the weight on the pre-trained layers, while the auxiliary layers were opened and trained with the images on the target dataset. In this process, Adam [24], which is a stochastic optimisation algorithm, is utilised to update the weights.

$$w_k = w_{k-1} - \eta * \hat{b}_k / \left(\sqrt{\hat{s}_k} + \varepsilon\right) \quad (9)$$

where η denotes the learning rate, w is the weight matrix, k represents the index of classes, $\hat{b}_k$ and $\hat{s}_k$ are the bias-corrected first and second moment estimations of the gradient, respectively

2. The second step retrained (fine-tuned) the model. Using the target dataset, the model was retrained by loading the weight trained in the first step, and all the layers were trained with the new images. Similarly, stochastic gradient descent [25] optimiser was employed to update the weight as defined in Equation (10).

$$w_k = w_{k-1} - \eta * (\partial L(w) / \partial w) \quad (10)$$

where $L(\cdot)$ indicates the loss function.

Particularly, considering the tiny disease spot symptoms of plant disease images and multi-classification tasks, the traditional Focal-loss function [26] was enhanced and employed in the network as calculated in Equation (11).

$$L_{mult} = -\sum_{c=1}^{C} \alpha_c (1 - p_c)^{\gamma} y_c \log(p_c) \quad (11)$$

$$y_c = \begin{cases} 1 & c = actual\_class \\ 0 & c \neq actual\_class \end{cases} \quad (12)$$

where p_c represents the predicted probability distribution, C is the number of classes, and a_c denotes the weighting factor. Figure 5 shows the network architecture, and related parameters are listed in Table 2.
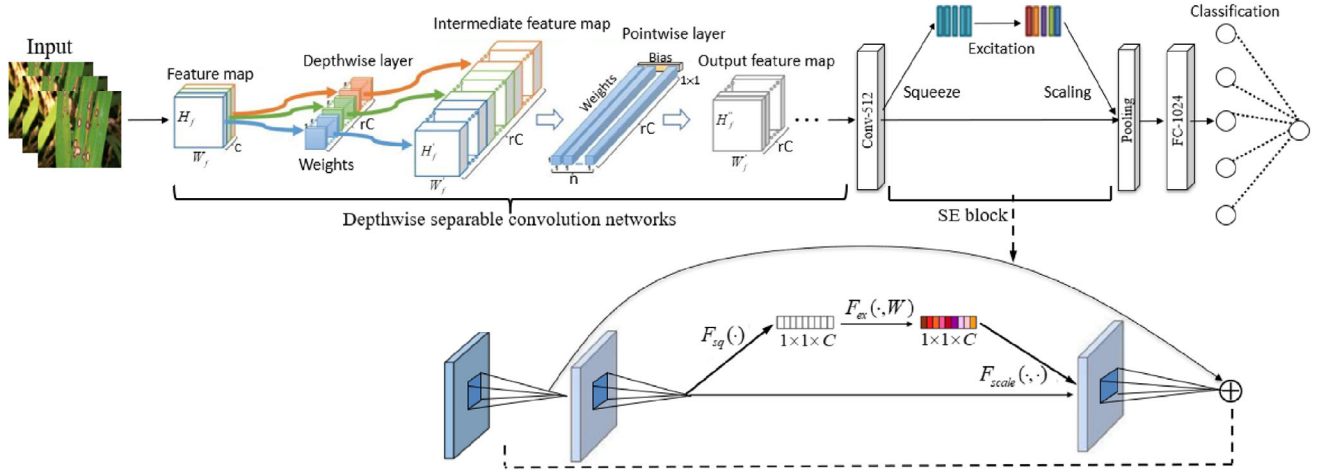
FIGURE 5 The network architecture of SE-MobileNet

TABLE 2 The related parameters of SE-MobileNet

No. of layers	Module (type)	Stride	Filtre shape	Input size
1	Convolution (Conv)	s2	$3 \times 3 \times 3 \times 32$	$224 \times 224 \times 3$
2	Conv depthwise (dw)	s1	$3 \times 3 \times 32$ dw	$112 \times 112 \times 32$
3	Conv	s1	$1 \times 1 \times 32 \times 64$	$112 \times 112 \times 32$
4	Conv dw	s2	$3 \times 3 \times 64$ dw	$112 \times 112 \times 64$
5	Conv	s1	$1 \times 1 \times 64 \times 128$	$56 \times 56 \times 64$
6	Conv dw	s1	$3 \times 3 \times 128$ dw	$56 \times 56 \times 128$
7	Conv	s1	$1 \times 1 \times 128 \times 128$	$56 \times 56 \times 128$
8	Conv dw	s2	$3 \times 3 \times 128$ dw	$56 \times 56 \times 128$
9	Conv	s1	$1 \times 1 \times 128 \times 256$	$28 \times 28 \times 128$
10	Conv dw	s1	$3 \times 3 \times 256$ dw	$28 \times 28 \times 256$
11	Conv	s1	$1 \times 1 \times 256 \times 256$	$28 \times 28 \times 256$
12	Conv dw	s2	$3 \times 3 \times 256$ dw	$28 \times 28 \times 256$
13	Conv	s1	$1 \times 1 \times 256 \times 512$	$14 \times 14 \times 256$
14~18	$5 \times$ Conv dw	s1	$3 \times 3 \times 256$ dw	$14 \times 14 \times 512$
19~23	$5 \times$ Conv	s1	$1 \times 1 \times 512 \times 512$	$14 \times 14 \times 512$
24	Conv dw	s2	$3 \times 3 \times 512$ dw	$14 \times 14 \times 512$
25	Conv	s1	$1 \times 1 \times 512 \times 1024$	$7 \times 7 \times 512$
26	Conv dw	s2	$3 \times 3 \times 1024$ dw	$7 \times 7 \times 1024$
27	Conv	s1	$1 \times 1 \times 1024 \times 1024$	$7 \times 7 \times 1024$
28	Conv	s1	$7 \times 7 \times 512$	$7 \times 7 \times 512$
29	SE block	s1	$7 \times 7 \times 512$	$7 \times 7 \times 512$
30	Avg Pool	s1	Pool 7×7	$7 \times 7 \times 512$
31	FC	s1	1024	$1 \times 1 \times 1024$
32	Softmax	s1	Classifier	$1 \times 1 \times 25$

3 | EXPERIMENTAL RESULTS AND ANALYSIS

We have primarily conducted the experiments using the software of Anaconda3 (Python 3.6), where the Tensorflow, OpenCV-python3, and Keras [27] libraries are utilised and accelerated by graphics processin unit (GPU). The hardware resources consist of Intel® Xeon(R) E5-2620 CPU, GeForce RTX 2080 graphics card (CUDA 10.2) [28], and 64-GB memory, which are used for the running program.

(a) Potato early blight (b) Potato late blight (c) Grape black rot (d) Grape healthy
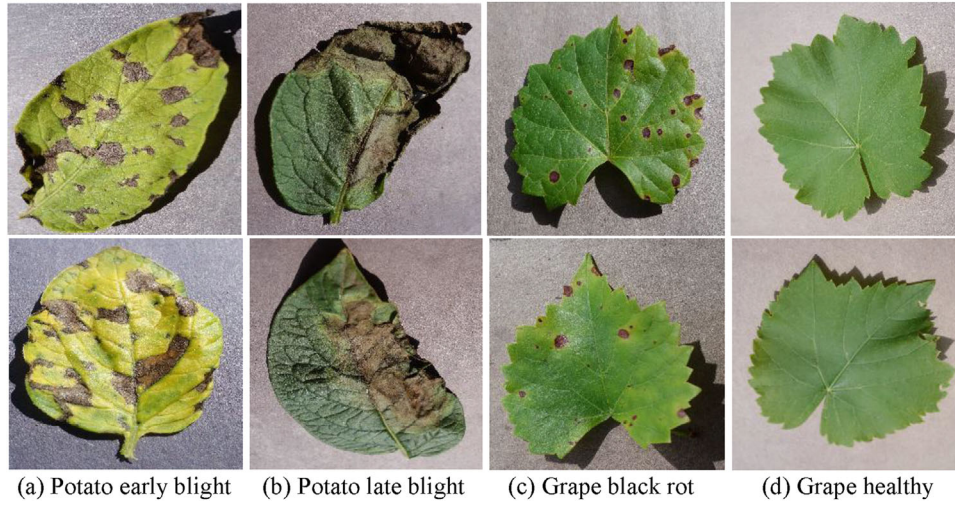
FIGURE 6 Examples of the plant image dataset (a) potato early blight, (b) potato late blight, (c) grape black rot, (d) grape healthy

TABLE 3 The training and validation accuracies on the PlantVillage database

| Pre-trained models	10 epochs			30 epochs			
	Training accuracies (%)	Validation accuracies (%)	Training losses	Training accuracies (%)	Validation accuracies (%)	Training losses	Validation losses
Inception V3	99.64	98.08	0.1227	99.92	98.41	0.0268	1.0182
VGGNet-19	96.16	93.94	0.2282	97.52	94.10	0.1866	0.3224
DenseNet-121	99.81	98.74	0.0806	99.95	98.89	0.0154	0.6133
NASNetMobile	99.59	97.23	0.1594	99.84	97.73	0.0482	1.2886
MobileNet-V2	98.57	98.74	1.3626	99.40	98.97	0.7897	1.0179
Proposed method	99.00	98.78	0.8387	99.65	99.04	0.5741	0.8146

3.1 | Experiment on a common dataset

As mentioned in Section 2.1, PlantVillage is a comprehensive global collection of data for the machine learning algorithm test to identify and classify crop diseases. To investigate the efficiency of the proposed procedure, we perform multiple experiments on this common dataset. The 38 categories of plant images, including apple, maize, grape, tomato, and potato, are downloaded from the PlantVillage database. Note that some same plant leaf images are taken in diverse orientations and the distributions of these sample images are not balanced. Some categories have a large number of samples while others have a small number of samples. The dimensions of all the images are uniformly adjusted to 256 × 256 pixels, and these sample images are taken in simple backdrops as displayed in Figure 6.

Based on the method introduced in Section 2.3.4, we conduct the model training and test on the plant disease dataset. To ensure full training and avoid overfitting problems as much as possible, the training and validation sets are split with the ratio of 8:2, which is chosen referring Too et al. [21]. Besides, a one-hot encoding of the categorical variable is done for the model training. In particular, to understand how the proposed procedure implements on the unseen samples, a certain proportion of the original images are reserved to verify the validity of the models (20% of the dataset, 10,861 images). We take five influential CNNs such as VGGNet (Visual Geometry Group Network) [29], Inception V3 [30], DenseNet [31], NASNetMobile (Neural Architecture Search Network Mobile) [32] and MobileNet V2 [33] into account to compare models. By applying the approach of transfer learning, these networks are built and the pre-trained models are loaded from ImageNet [34]. The tail layers of the networks are discarded by setting a new completely linked Softmax layer with the actual number of categories. Thus, these diverse CNN models are trained and multiple experiments are conducted on the common dataset. The training and validation accuracies of different methods are listed in Table 3.

From Table 3, it can be observed that the proposed procedure shows superior performance compared with the other commonly used methods on the experimental dataset. The key explanation for this is that the proposed procedure not only utilises the general features of images extracted from the Ima-
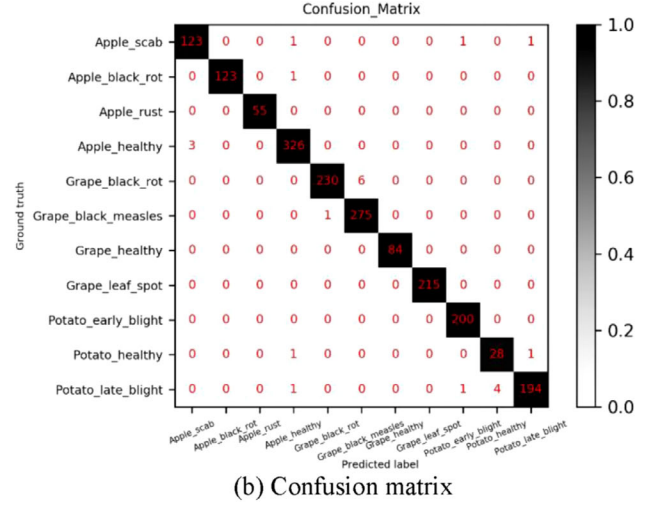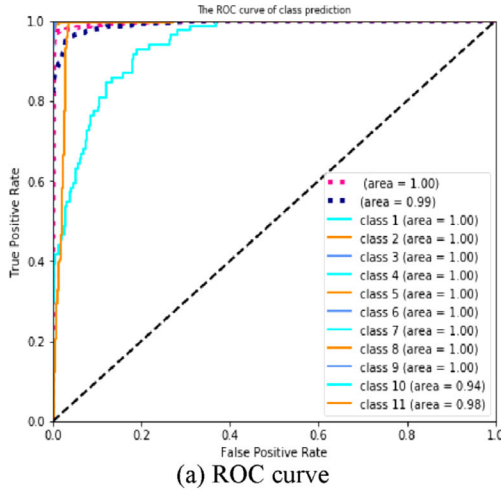
FIGURE 7 The receiver-operating characteristic (ROC) curve and confusion matrix of identification. (a) ROC curve, (b) confusion matrix

TABLE 4 The identified results of different plant disease types

No.	Category names	No. of identified samples	No. of correct samples	Accuracies (%)	Sensitivities (%)	Specificities (%)
1	Apple scab	126	123	99.68	97.62	99.83
2	Apple black rot	124	123	99.95	99.19	100.00
3	Apple rust	55	55	100.00	100.00	100.00
4	Apple healthy	329	326	99.63	99.09	99.74
5	Grape black rot	236	230	99.63	97.46	99.94
6	Grape black measles	276	275	99.63	99.64	99.63
7	Grape healthy	84	84	100.00	100.00	100.00
8	Grape leaf spot	215	215	100.00	100.00	100.00
9	Potato early blight	200	200	99.89	100.00	99.88
10	Potato late blight	30	28	99.68	93.33	99.78
11	Potato healthy	200	194	99.57	97.00	99.88
—	*Average*	—	—	99.78	98.83	99.88

geNet dataset by MobileNet but also takes advantage of the specific features extracted from the target dataset by newly extended layers. Especially, the model was trained in two phases, which made the optimum weights obtained for the network. In contrast, owing to the individual networks, the other models did not achieve the optimal results although the weights were initialised using pre-trained parameters rather than inferring from scratch. Therefore, the pre-trained MobileNet was fused with the SE block, which integrated the merits of both and achieved the best performance in experiments on the common dataset. Moreover, the optimum model obtained by the proposed procedure was selected to identify the types of plant diseases, and the corresponding identification results are depicted in Figure 7. Taking into account the statistics of accurate identifications (true positive (TP)), misidentification (false negatives (FN)), false positive (FP), and true negatives (TN), we may verify the efficiency of the models with the metrics like *Specificity*,

Sensitivity (*Recall*), and *Accuracy* as displayed in Equations (13)–(15).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (13)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (14)$$

$$Specificity = \frac{TN}{FP + TN} \quad (15)$$

where *TP* denotes the number of correctly identified samples; *FN* represents the number of failed identifications; *FP* represents the number of false identifications; *TN* denotes the number of samples that they are not in such disease and accurately identified by the classifier. Table 4 presents the actual analysis of the results which are assessed in the form of different indicators.

TABLE 5 Compared with the latest literature results on the public PlantVillage database

Literature	Year	Models	No. of classes	Accuracies
[11]	2016	GoogLeNet	26	99.35%
[35]	2017	LeNet	3	98.61%
[36]	2017	GoogLeNet	9	99.18%
[37]	2017	AlexNet	10	95.65%
[38]	2017	VGG16	4	90.40%
[39]	2018	VGG16	6	96.19%
[21]	2018	DenseNets	38	99.75%
[40]	2019	PlantDiseaseNet	18	93.67%
Proposed approach	2020	SE-MobileNet	38	99.78%

From Figure 7(a), we can see that the receiver-operating characteristic (ROC) curve of each class is close to the upper-left corner except for class 10, and the areas of ROC curves for most classes are even equal to 1, indicating the effectiveness of the proposed procedure. Besides, as seen in Figure 7(b) and Table 4, most of the samples in each class are correctly detected by the SE-MobileNet model, and the average prediction *Accuracy* of the experimental results reaches 99.78% on this common dataset, which indicates that the proposed procedure has an impressive ability to identify the plant diseases in simple backdrop conditions. Particularly, in addition to identifying whether the plant is healthy or diseased, the proposed approach also distinguishes the specific disease types of plant disease images, and the average *Sensitivity* and *Specificity* reach 98.83% and 99.88%, respectively.

Additionally, Table 5 presents the results of experiments conducted on the PlantVillage database for some researches since 2016. The highest accuracy rate and corresponding method for each study are listed in the table. Most of the research does not use all the datasets of the PlantVillage, so the number of classifications is much smaller than the 38 classes of this database. As is well known, the smaller the number of categories is, the easier it is to achieve higher accuracy. This study considers all the species and disease types in the experiments, and the number of categories is 38. In the case of more difficult classification, a higher accuracy rate is achieved for the proposed procedure, which shows a promising performance relative to other state-of-the-art methods.

3.2 | Experiment on the collected images

Similar to Section 3.1, the proposed procedure was further tested on the collected rice plant images, which were captured from real-life agricultural fields with heterogeneous background conditions and inconsistent lighting intensities. For example, the backgrounds of some images are the surroundings of the field, and in some other images, the backgrounds are the grasses or soils of different colours. Weather conditions are also diverse at different times such as images taken in sunny, cloudy or over-cast weather and so forth. In addition to retaining some images to verify the validity of the models, the original images were split into the training and validation sets according to the proportion of 8:2. In particular, to avoid the overfitting problems and make images diverse, the data-augmentation scheme was applied to ensure that no less than 200 sample images were in every category for the dataset. The detailed procedures of data augmentation are described as follows.

1. Colour-jittering: By adjusting the saturation, intensity, and contrast of the images, the different light conditions of the sample images were simulated.
2. Random rotation: By rotating the sample images by random angles in the interval (90, 180°), we would have five new sample images of each original image after this procedure.
3. Flipping and translation: Along the vertical axis, the flipping was performed for the original images, which were randomly moved to a certain distance in the x or y direction (or both directions).
4. Cropping: Considering the size of the original images is 256 × 256, we used a window with a size of 224 × 224 to randomly cut images. Also, the cropping transformation in x-direction was conducted. Let $(1, -tan\alpha; 0,1)$ be the transformation matrix in x-direction. For a pixel point $P(x_1, y_1)$: x_1, $y_1 \in R$, the transformed pixel point was $P'(x_2, y_2) = P'(x_1 - y_1 \cdot tan\alpha, y_1)$.
5. Scale transform: The original images were randomly zoomed in or out with the scale changed from 0.9 to 1.1.
6. GAN generation: The GAN was enhanced and assigned the input image dimensions as 224 × 224 pixels to generate new images for ensuring the quantity and diversity of images.

Different from the most used input size of 64 × 64 pixels, we modified the assignment of input image size for GAN as 224 × 224 pixels and employed a styled convolutional block of 128 × 64 × 3 following a 32 × 3 convolutional block as the last layer of the generator module. Likewise, for the discriminator module, the input shape was set as 224 × 224 × 3, and the corresponding 32 × 3 convolutional block followed by a 64 × 128 × 3 convolutional block was used in the module. In this way, applying the data-augmentation scheme, the size of the dataset is enlarged and the diversity of training sample images is improved, which is useful for the robustness of the model and avoids the overfitting problem. Figure 8 displays the partial augmented samples.

Based on the approach proposed in Section 2.3, the model training and test were conducted on the rice image dataset, and all the rice plant images were evenly adjusted to a fixed size to fit the model. To learn how the approach performs on new unknown images, we particularly reserved some images outside the modelling to validate the efficiency. Likewise, the five most-used CNNs including Inception V3, VGGNet, DenseNet, NASNetMobile, and MobileNet-V2 were selected for the comparison analysis. Using the transfer learning approach, we established these networks and loaded related weight parameters with the weights pre-trained on ImageNet. The initial top layers of the networks were replaced by a new completely linked layer with the actual category number and Softmax function. In this

| (a) Original images | (b) Color jittering | (c) Random rotation | (d) EGAN generating |
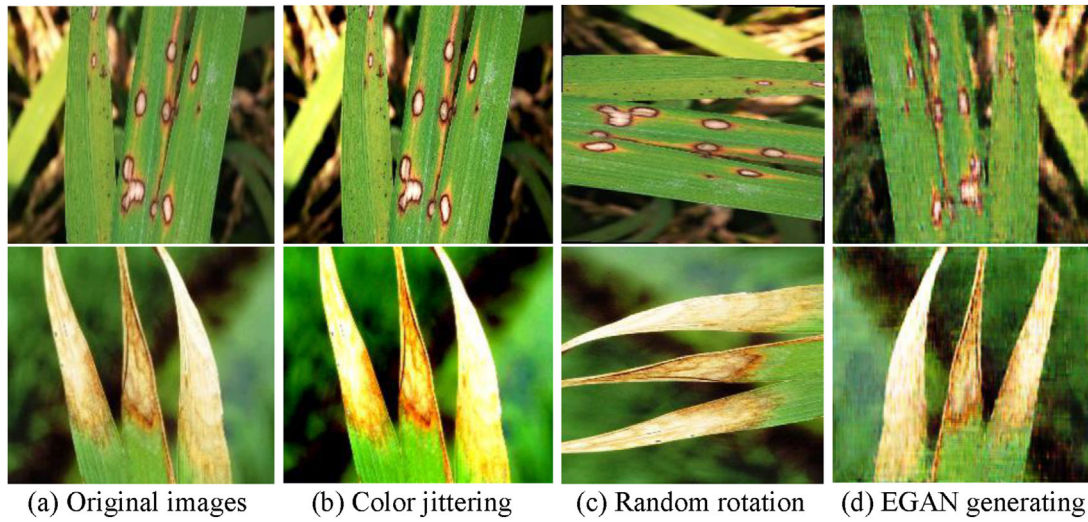
FIGURE 8 Partial augmented images of rice plant diseases (a) original images, (b) colour-jittering, (c) random rotation, (d) enhanced generative adversarial networks generation

TABLE 6 The training and validation accuracies on the rice plant images

Pre accuracies (%)	10 epochs			30 epochs				
	Training accuracies %	Validation accuracies %	Training accuracies %	Training accuracies %	Validation accuracies %	Training losses	Validation losses	Training time
Inception V3	98.30	87.50	0.4924	99.88	87.95	0.0886	18.6597	9.47 min
VGGNet-19	92.36	86.16	4.7447	100.00	86.61	0.0153	28.4182	12.10 min
DenseNet-121	99.25	89.29	0.2924	99.96	87.95	0.0521	20.7355	13.43 min
NASNetMobile	98.22	84.09	0.7830	99.41	87.50	0.3627	13.7815	20.12 min
MobileNet-V2	97.43	89.29	2.5631	99.01	85.27	1.8308	21.0415	9.01 min
Proposed method	98.58	90.45	5.9572	98.93	91.07	5.7155	19.5713	7.46 min

way, the model training of these networks was performed and many experiments were implemented for the identification of rice plant diseases. Table 6 shows the results of model training.

Table 6 reveals that the proposed SE-MobileNet outperforms other state-of-the-art methods that retrained the networks utilising the experimental dataset, even though the optimum classifier is implemented.

When trained for 10 and 30 epochs, the training accuracies of the proposed approach have achieved 98.58% and 98.93%, and the validation accuracies reach 90.45% and 91.07, respectively. More than that, the proposed approach is memory-efficient. It exhibits a fast running speed and substantial efficiency compared with other state-of-the-art methods. Further, applying the model trained by the introduced procedure, the new unseen images were used to conduct the identification of rice plant diseases. Figure 9 displays the identified results, and the evaluation indicators have been counted in Table 7.

From Figure 9(a), we can observe that the identification results of the proposed procedure show ideal operation points in the left-top corner of the ROC curve. Most identified rice disease types have a higher rate of TP while a lower rate of false-positive (FPR), which is also demonstrated by the confusion matrix of Figure 9(b). For example, the correct quantity is 13 for the identification of "rice stackburn" in 14 samples, and 18 samples are accurately identified in the category of "rice leaf smut" except for one misidentified sample. For the "white leaf streak," all the four samples are correctly identified by the proposed approach. Therefore, the average *Accuracy* achieves 99.33% and the average *Sensitivity* reaches no less than 87.87% for identifying the rice plant diseases, as displayed in Table 7. Conversely, there are also some misidentified samples such as 4 failed identifications in the category of "Rice leaf scald," which is caused by some rice diseases like the "Rice leaf scald" and "Rice white tip" occurring in the same plant images. Moreover, the cluttering wild backgrounds and irregular lighting strengths may also impact the results of identification. Figure 10 displayed the partially identified sample images.

As seen in Figure 10, the identified categories of most rice plant images are compatible with the true types of these samples and the proposed procedure accurately identifies most rice

(a) ROC curve
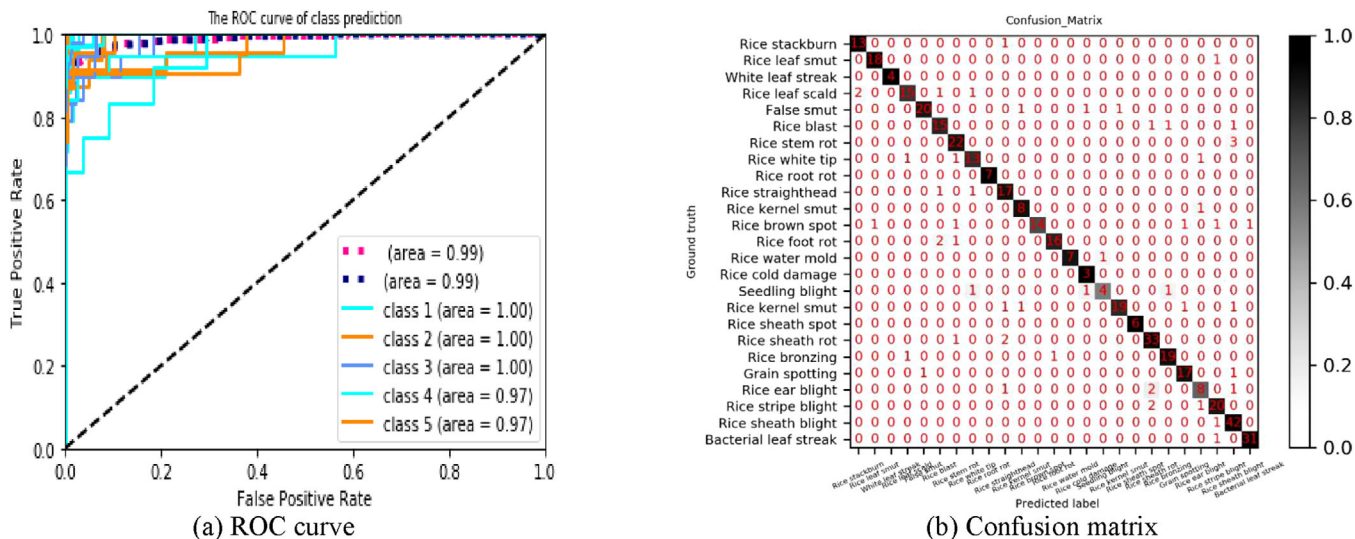
(b) Confusion matrix

FIGURE 9 The identification results on the collected rice plant disease dataset (a) ROC curve, (b) confusion matrix

TABLE 7 The identified results of different rice disease classes

No.	Rice disease types	No. of identified samples	No. of correct samples	Accuracies (%)	Sensitivities (%)	Specificities (%)
1	Rice stackburn	14	13	99.32	92.86	99.53
2	Rice leaf smut	19	18	99.68	94.74	99.84
3	White leaf streak	4	4	100.00	100.00	100.00
4	Rice leaf scald	19	15	99.04	78.95	99.67
5	Rice false smut	23	20	99.36	86.96	99.83
6	Rice blast	18	15	98.88	83.33	99.34
7	Rice stem rot	25	22	98.88	88.00	99.33
8	Rice white tip	16	13	99.04	81.25	99.51
9	Rice root rot	7	7	100.00	100.00	100.00
10	Rice straighthead	19	17	98.88	89.47	99.17
11	Rice kernel smut	9	8	99.52	88.89	99.68
12	Rice brown spot	19	14	99.20	73.68	100.00
13	Rice foot rot	19	16	99.36	84.21	99.84
14	Rice water mold	8	7	99.84	87.50	100.00
15	Rice cold damage	3	3	99.84	100.00	99.84
16	Rice seedling blight	7	4	99.36	57.14	99.84
17	Rice kernel smut	23	19	99.20	82.61	99.83
18	Rice sheath spot	6	6	100.00	100.00	100.00
19	rice sheath rot	36	33	98.88	91.67	99.32
20	Rice bronzing	21	19	99.36	90.48	99.67
21	Grain spotting and peck	19	17	99.36	89.47	99.67
22	Rice ear blight	12	8	98.88	66.67	99.51
23	Rice stripe blight	23	20	98.88	83.33	99.50
24	Rice sheath blight	43	42	98.72	97.67	98.79
25	Bacterial leaf streak	32	31	99.68	96.87	99.83
—	*Average*	—	—	99.33	87.87	99.67
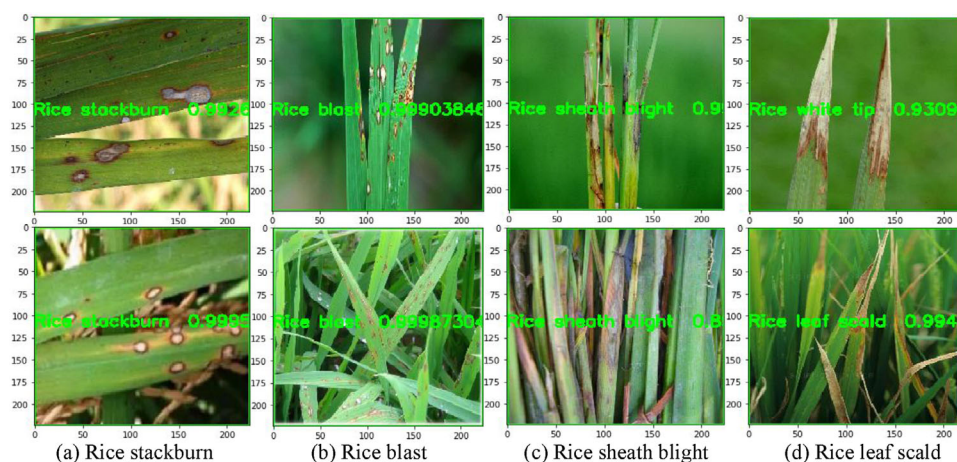
FIGURE 10 The identified examples of rice plant diseases (a) rice stackburn (b) rice blast (c) rice sheath blight (d) rice leaf scald

disease types. For example, the true disease type of Figure 10(a) is 'rice stackburn', which is accurately identified with a probability higher than 0.99. Similarly, the other samples, including Figures 10(b) and (c), are all identified accurately by the proposed procedure.

By contrast, some different types of rice diseases occurring in the same plant may lead to individual misclassification. Also, the extreme clutter field backdrops and irregular lighting strengths, which affect the feature extraction of plant lesion images, can cause some false identification as displayed in Figure 10(d). In spite of individual misidentifications, most of the rice plant diseases have been correctly identified by the proposed procedure. Therefore, based on the experimental analysis, it can be assumed that the proposed procedure is successful in identifying rice plant diseases and can also be transplanted to other related fields.

4 | CONCLUSIONS

To ensure the production of agricultural products, timely and successful identification of plant diseases is crucial, and therefore the quest for a quick, automated, less expensive, and reliable system for identifying crop diseases is of great practical importance. Deep-learning methods, especially CNNs, have successfully overcome much of the technological challenges correlated with image identification and classification. Hence, a new deep-learning structure called SE-MobileNet for identification of plant diseases is implemented in this study. The pre-trained MobileNet paired with SE block was used in our approach. By modifying the network structure with new extension layers and optimised loss function, the SE-MobileNet was created to enhance the learning capability of tiny lesion characteristics for plant disease images.

Moreover, the transfer learning was performed twice for model training: The first phase only inferred parameters from scratch for new extended layers while the convolutional layers were frozen with the parameters trained on ImageNet; the sec-

ond phase retrained (fine-tuned) the network using the target dataset by loading the pre-trained model in the first phase. In this way, the yielded optimum model was used to identify plant disease types. The experimental findings showed the model with substantial efficacy on both the open dataset and our captured rice plant images. We are planning to deploy it on portable devices to track and automatically identify the wide-ranging knowledge regarding plant diseases in future production. In the meantime, this procedure can be applied to other similar fields such as online defect assessment, molecular cell recognition, position detection from dissimilar images and much more.

REFERENCES

1. Picon, A., et al.: Crop conditional convolutional neural networks for massive multi-crop plant disease classification over cell phone acquired images taken on real field conditions. Comput. Electron. Agric. 167, 105093 (2019)
2. Wang, X., et al.: Automatic detection of rice disease using near infrared spectra technologies. J. Indian Soc. Remote Sens. 45, 785–794 (2017)
3. Wells, W.M. III: Medical image analysis–past, present, and future. Med. Image Anal. 33, 4–6 (2016)
4. Kaur, S., et al.: Semi-automatic leaf disease detection and classification system for soybean culture. IET Image Proc. 12, 1038–1048 (2018)
5. Chen, J., et al.: Detection of rice plant diseases based on deep transfer learning. J. Sci. Food Agric. 100, 3246–3256 (2020)
6. Yao, Q., et al.: Application of support vector machine for detecting rice diseases using shape and color texture features. In: 2009 international conference on engineering computation, Hong Kong, China (2009)
7. Kahar, M.A., et al.: Early detection and classification of paddy diseases with neural networks and fuzzy logic. In: Proceedings of the 17th International Conference on Mathematical and Computational Methods in Science and Engineering, MACMESE, Kuala Lumpur, Malaysia (2015)
8. Kai, S., et al.: Corn leaf disease recognition based on suport vector machine method. Trans. Chinese Soc. Agric. Eng. 23(1), 155–157 (2007)
9. Zhang, Z.Y., et al.: Image recognition of maize leaf disease based on ga-svm. Chem. Eng. Trans. 46, 199–204 (2015)

10. Barbedo, J.G.A.: Factors influencing the use of deep learning for plant disease recognition. Biosyst. Eng. 172, 84–91 (2018)
11. Mohanty, S.P., et al.: Using deep learning for image-based plant disease detection. Front. Plant Sci. 7, 1419 (2016)
12. Kawasaki, Y., et al.: Basic study of automated diagnosis of viral plant diseases using convolutional neural networks. In: International Symposium on Visual Computing. Springer, Cham (2015)
13. Zhang, X., et al.: Identification of maize leaf diseases using improved deep convolutional neural networks. IEEE Access 6, 30370–30377 (2018)
14. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, Salt Lake City, Utah (2018)
15. Howard, A.G., et al.: MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861 (2017)
16. Russakovsky, O., et al.: Imagenet large scale visual recognition challenge. Int. J. Comput. Vision 115, 211–252 (2015)
17. Hughes, D., Salathé, M.: An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv preprint arXiv:1511.08060 (2015)
18. Goodfellow, I., et al.: Generative adversarial nets. In: Advances in Neural Information Processing Systems, vol. 2, pp. 2672–2680. Curran, Red Hook, New York (2014)
19. Sifre, L., Mallat, S.: Rigid-motion scattering for image classification. Ecole Polytechnique, CMAP. PhD Dissertation (2014)
20. Glorot, X., et al.: Deep sparse rectifier neural networks. In: Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, Fort Lauderdale, Florida (2011)
21. Too, E.C., et al.: A comparative study of fine-tuning deep learning models for plant disease identification. Comput. Electron. Agric. 161, 272–279 (2019)
22. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 22(10), 1345–1359 (2009)
23. Li, X., et al.: DAC: Data-free automatic acceleration of convolutional networks. In: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). Waikoloa Village, Hawaii (2019)
24. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
25. Ghazi, M.M., et al.: Plant identification using deep neural networks via optimization of transfer learning parameters. Neurocomputing 235, 228–235 (2017)
26. Lin, T.-Y., et al.: Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision, Venice, Italy (2017)
27. Keras-GPU. https://anaconda.org/an-aconda/keras-gpu. Accessed 17 June 2019
28. GeForce GTX 1060. https://nvidia.com/enus/geforce/products/10series/geforce-gtx-1060/specifications. Accessed 17 June 2019
29. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
30. Szegedy, C., et al.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, Nevada (2016)
31. Huang, G., et al.: Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, Honolulu, Hawaii (2017)
32. Zoph, B., et al.: Learning transferable architectures for scalable image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, Salt Lake City, Utah (2018)
33. Sandler, M., et al.: MobileNetV2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, Salt Lake City, Utah (2018)
34. ImageNet. https://github.c- om/fchollet/deep-learning-models/releases
35. Amara, J., et al.: A deep learning-based approach for banana leaf diseases classification. In: Datenbanksysteme für Business, Technologie und Web (BTW 2017)-Workshopband, Stuttgart, Germany (2017)
36. Brahimi, M., et al.: Deep learning for tomato diseases: Classification and symptoms visualization. Appl. Artif. Intell. 31, 299–315 (2017)
37. Durmuş, H., et al.: Disease detection on the leaves of the tomato plants by using deep learning. In: 2017 6th International Conference on Agro-Geoinformatics, Fairfax, Virginia (2017)
38. Wang, G., et al.: Automatic image-based plant disease severity estimation using deep learning. Comput. Intell. Neurosci. 2017 (2017) https://doi.org/10.1155/2017/2917536.
39. Rangarajan, A.K., et al.: Tomato crop disease classification using pre-trained deep learning algorithm. Procedia Comput. Sci. 133, 1040–1047 (2018)
40. Arsenovic, M., et al.: Solving current limitations of deep learning based approaches for plant disease detection. Symmetry 11, 939 (2019)

How to cite this article: Chen J, Zhang D, Suzauddola M, Nanehkaran YA, Sun Y. Identification of plant disease images via a squeeze-and-excitation MobileNet model and twice transfer learning. *IET Image Processing.* 2020;1–13. https://doi.org/10.1049/ipr2.12090