Artificial Intelligence Review (2023) 56:7985–8002
https://doi.org/10.1007/s10462-022-10374-3

Weakly-supervised learning method for the recognition of potato leaf diseases

Junde Chen[1,4,5] · Xiaofang Deng[2] · Yuxin Wen[1] · Weirong Chen[3] · Adnan Zeb[4,6] · Defu Zhang[4]

Accepted: 10 December 2022 / Published online: 21 December 2022

Abstract
As a crucial food crop, potatoes are highly consumed worldwide, while they are also susceptible to being infected by diverse diseases. Early detection and diagnosis can prevent the epidemic of plant diseases and raise crop yields. To this end, this study proposed a weakly-supervised learning approach for the identification of potato plant diseases. The foundation network was applied with the lightweight MobileNet V2, and to enhance the learning ability for minute lesion features, we modified the existing MobileNet-V2 architecture using the fine-tuning approach conducted by transfer learning. Then, the atrous convolution along with the SPP module was embedded into the pre-trained networks, which was followed by a hybrid attention mechanism containing channel attention and spatial attention submodules to efficiently extract high-dimensional features of plant disease images. The proposed approach outperformed other compared methods and achieved a superior performance gain. It realized an average recall rate of 91.99% for recognizing potato disease types on the publicly accessible dataset. In practical field scenarios, the proposed approach separately attained an average accuracy and specificity of 97.33% and 98.39% on the locally collected image dataset. Experimental results present a competitive performance and demonstrate the validity and feasibility of the proposed approach.

Keywords Potato crop diseases · Image recognition · Atrous convolution · SPP module · Lightweight network

1 Introduction

In the global economy, agriculture is critical, and with population growth as well as the COVID-19 pandemic, the agricultural system faces more strain. After wheat and rice, potato is currently the third most important food crop in the world, and global production of potatoes which are regarded by over a billion people as the primary stable exceeds

✉ Junde Chen
jundchen@chapman.edu

✉ Yuxin Wen
yuwen@chapman.edu

Extended author information available on the last page of the article

300 million metric each year (Oppenheim et al. 2019). In addition to being a considerable source of calories for humanity, potatoes are also widely utilized as industrial materials. However, the potato crop is susceptible to being infected by diverse diseases too. Early detection and diagnosis have a positive impact on suppressing the epidemic of potato plant diseases, while the traditional approach of visual observations requires constant supervision of plants. It is undoubtedly inefficient, intuitive, labor-intensive, and cannot be transplanted in a broad range (Marino et al. 2019a; Al-Hiary et al. 2011). Thence, there is a great need and significant realistic importance to seek a simple, quick, and effective tool for automatically recognizing potato plant diseases.

In the latest literature, new methods of plant disease identification are being proposed with the rapid advancement of digital cameras and calculational capacity. More and more attention has been paid to the research and application of machine learning (ML) and image processing techniques, which are becoming attractive alternative approaches for the continuous monitoring of plant diseases (Chen et al. 2021a). For instance, by integrating image processing and ML techniques, Islam et al. (2017) introduced a potato disease recognition model and successfully identified over 300 images. They achieved a 95% recognition accuracy. Using a hyperspectral imaging technique, Ji et al. (2019) recognized the bruised potatoes through discrete wavelet transform, and they attained the highest recognition accuracy of 99.82% for the damaged potatoes. Gassoumi et al. (2000) recommended an artificial neural network (ANN) based method for the identification of insect pests in cotton ecosystems. Their method was implemented with good stability and achieved 90% accuracy. Using ANN, random forest (RF), and support vector machine (SVM) methods, Patil et al. (2017) executed a comparative analysis for identifying potato disease images. In their experiments, the RF realized an accuracy of 79%, the SVM achieved an accuracy of 84%, and the ANN gained the highest accuracy of 92%. Although impressive results are reported in the literature, the conventional ML methods also have some demerits, such as the dependence on hand-crafted features, complicated image processing procedures, and low robustness. Recently, a novel ML technology named deep learning (DL), explicitly convolutional neural network (CNN), has been introduced to address the most challenging tasks associated with image identification and classification (Junde et al. 2021; Pattnaik et al. 2020; Cristin et al. 2020; Shrivastava et al. 2019). It has also been applied in the field of plant disease recognition. For example, using 2250 potato leaf images on the PlantVillage dataset, Al-Amin et al. (2019) trained a CNN model to identify different potato diseases. Their model realized the best recognition accuracy of 98.33%. By applying the method of transfer learning (TL), Islam et al. (2019) recognized three categories of potato leaf images, including 1000 late blight leaves, 1000 early blight leaves, and 152 normal leaves. Their experimental results revealed that TL outperformed the compared methods and they reached a 99.43% test accuracy with a 4:1 ratio for splitting the training and test sets. Marino et al. (2019b) inferred a CNN model to locate the regions of potato defects, which is realized by a heat map output. They performed the classification of potato defects and realized an average F1-score of 0.94. Besides, applying an ensemble CNN model, Nanni et al. (2020) performed the detection of plant insect pests and attained an advanced accuracy of 92.43%. Based on the squeeze-and-excitation (SE) MobileNet, Chen et al. (2021b) proposed a method to identify paddy diseases and they attained an average identification accuracy of 99.78% for recognizing paddy disease types on the publicly accessible dataset, etc. Generally speaking, there are two varieties of DL methods, including strongly supervised and weakly-supervised approaches for crop disease recognition. The strongly-supervised method primarily adopts object detection techniques based on more manually-annotated information like coordinate data, bounding boxes, and crucial points of the

target objects. Undoubtedly, this approach is tedious and labor-intensive to gain numerous annotation data for training models. Instead, the other alternative approach is a weakly-supervised scheme, which only requests the label information of images, e.g., the plant disease images with the same disease type are stored in the same folder, and the detailed annotation information is not needed. Therefore, more and more research has focused on fine-grained image classification using a weakly-supervised learning strategy, which is also adopted in our work. On another front, despite reasonably good findings reported in the literature, deep CNN (DCNN) based methods need a great number of annotated samples to train the model, which poses a challenging problem for DCNN models. Particularly, the great volume of deep CNN models also limits the portable device deployment for plant disease identification models than can run offline in practical applications. As a consequence, this study puts forward a lightweight network architecture for recognizing potato diseases. The pre-trained MobileNet-V2 was chosen as the basis feature extractor of the network, and to enhance the learning ability of minute plant lesion characteristics, we modified the classical MobileNet-V2 architecture. The atrous convolution along with the SPP module was incorporated into the network, which was followed by a hybrid attention module including sequential channel-wise attention and spatial attention mechanisms. In this manner, the features of inter-channel dependencies and spatial points are grasped, thereby improving the accuracy of the model. Overall, our work makes the following specific contributions:

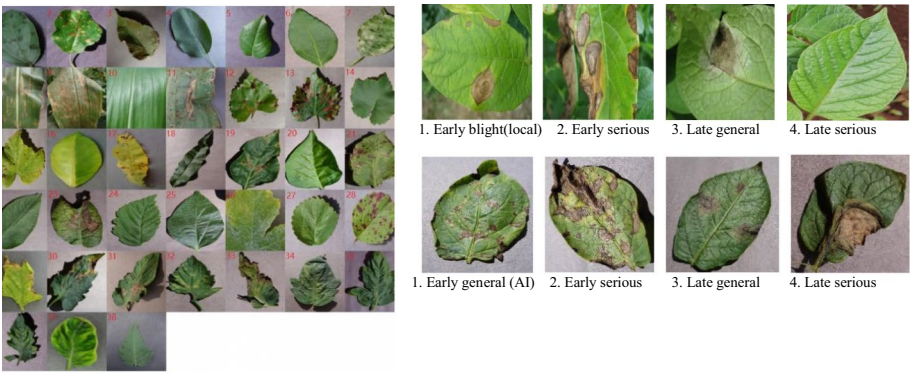
- A lightweight MobS_Net model was proposed for recognizing potato plant diseases with an accuracy of 97.73%, and it attained increasing effectiveness compared with other state-of-the-art methods.
- The traditional MobileNet-V2 was modified and the atrous SPP was incorporated into the pre-trained network, which was connected by a hybrid attention mechanism for improving the capability of feature extraction.
- We enhanced the Focal-Loss (FL) function to make it address multi-classification tasks. To alleviate the imbalance of data problems and make the model keep more attention to positive samples, we utilized the enhanced Focal-Loss (EFL) function to substitute the traditional Cross-Entropy one.

The remainder of this writing is organized below. Section 2 introduces the used materials and the proposed approaches. This section importantly discusses the methodology. Section 3 dedicates to the experimental analysis, and extensive experiments are performed in this section along with an ablation study. Section 4 concludes this paper with a summary and specific recommendations for further work.

2 Materials and methods

2.1 Materials

We have collected the materials from diverse sources and many images are derived from the open-access dataset of the 2018 AI Challenger Contest (www.challenger.ai, AI dataset), which is a wide acquisition of plant leaf images employed for ML algorithm test of plant disease identification. It is essential to emphasize that the potato plant leaf images of this dataset are sourced from the PlantVillage repository, where the samples are taken under controlled conditions, both in background and brightness. This potato image



(a) Open-access PlantVillage dataset. (b) Potato disease sample images.

Fig. 1 The samples of plant disease images

Table 1 The details of the dataset

Plant disease types	No. of samples			
	No. of original images	No. of training and validation images	No. of augmented training and validation samples	No. of testing samples
Potato healthy	1634	1430	1430	204
Early_blight	207	200	1000	7
Early blight fungus serious	583	510	1000	73
Early blight fungus general	232	203	1000	29
Late blight fungus serious	510	446	1000	64
Late blight fungus general	287	251	1000	36
Total	3246	2840	5430	406

dataset contains 3276 potato leaf images categorized into five classes: early blight fungus serious, early blight fungus general, late blight fungus serious, late blight fungus general, and health. In other terms, the species is composed of several disease types and a healthy category, and each disease type includes two varieties of severity levels such as serious and general types. Additionally, the other images are from the local dataset, which is collected under practical field wild scenarios with complicated backgrounds and uneven lighting intensities. Including 363 healthy samples, 207 early blight samples, 109 late blight samples, and 133 potato virus disease samples, a total of 812 image samples are collected in this local dataset. Note that these images have a wide variety, which means that some of the images taken have less noise background, while others have high noise interference. The potato plant images belonging to the same disease type are stored in the same folder. Only the category information is labeled for each folder, and the detailed annotation data are not required by the weakly-supervised methods. By this means, the sample dataset is formed and employed for the test of potato disease recognition. Figure 1 shows the partial sample images, and the details of these samples are summarized in Table 1.

2.2 Related work

2.2.1 MobileNet-V2

Mobile-nets, which can be deployed on portable devices for image recognition and classification, is a series of lightweight healthy depending on the depth-wise separable convolution (DSC) and streamlined structure (Sifre and Mallat 2014). Among them, DSC splits a standard convolution into a depth-wise convolution (DC) and a point-wise convolution (PC), respectively. DC executes the convolution operation on each channel with one filter for input maps, and PC performs the 1×1 convolving on the output of DC. The formulas of DC and PC are calculated in Eqs. (1, 2), respectively.

$$DC(\theta, x)_{(i,j)} = \sum_{w=0}^{W} \sum_{h=0}^{H} \theta_{(w,h)} \cdot x_{(i+w,j+h)} \tag{1}$$

$$PC(\theta, x)_{(i,j)} = \sum_{k=0}^{K} \theta_k \cdot x_{(i,j)}, \tag{2}$$

where H and W separately stand for the height and width of the input feature map, θ is the weights of filters, (i, j) index position of input x. By this means, DC does not modify the channel number and PC unifies the outputs of the DC, as expressed in Eq. (3).

$$DSC(\theta_p, \theta_d, y)_{(i,j)} = PC_{(i,j)}(\theta_p, DC_{(i,j)}(\theta_d, y)) \tag{3}$$

Consequently, the output results of DSC can be gained, and in addition to DSC, MobileNets also embed batch normalization (BN) behind the convolution layer to alleviate the problem of disappearing gradient in the back-propagation (BP) procedure. On the ground of this, MobileNet V2 (Sandler et al. 2018) introduces the concepts of linear bottleneck framework along with inverted residual block to address the risk of vanishing gradients and attains some advancement over V1.

2.2.2 Atrous spatial pyramid pooling

Spatial pyramid pooling (or SPP in short) (He et al. 2015a) is a pooling method that maps local characteristics to diverse dimension spaces and merges them. Except for generating fixed-size feature vectors, SPP can make the CNN architecture adapt the image input with different dimensions and extract multi-scale feature information of plant diseases or pests. The module of SPP accepts features extracted from the backbone network and executes convolution operations at multiple scales for extracting global contextual information. Suppose w and h separately represent the width and height of an input feature map, and thus with a $G \times G$ grid size of SPP, the size of the convolution kernel represented by $f = f_h = f_w$ can be computed using $f_h = [h/G]$ and $f_w = [w/G]$, where $[\cdot]$ symbolizes the ceiling operation. However, because of increasing parameters and computational loads, the convolution kernel cannot be as large as desired, and the normal convolution kernel has the demerit that the spatial resolution of the feature map is halved at each step. Therefore, atrous SPP was designed to alleviate this challenge and the hyper-parameter of rate $r = 2$ was set for the atrous convolution. Compared with

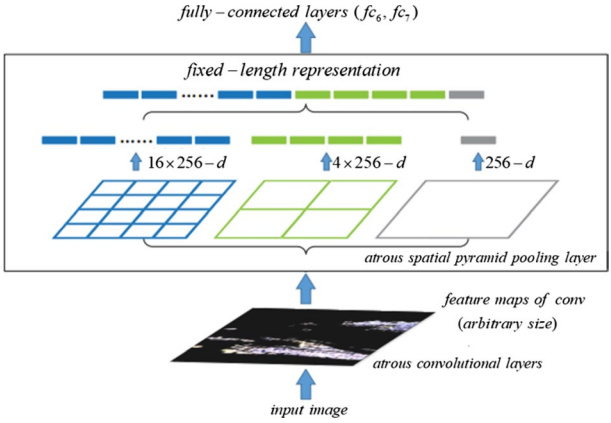


Fig. 2 The structure of Atrous SPP

traditional SPP, the atrous SPP increases receptive fields of convolution while remaining the same computational cost (Fig. 2).

2.2.3 Channel-wise and spatial attention

Similar to the human visual attention mechanism, the attention module in deep learning can help the model focus on useful features while suppressing unwanted information. For this purpose, researchers have introduced many attention mechanisms, which can be classified as channel-wise attention (e.g. SE-block) (Hu et al. 2018), spatial attention (Wang et al. 2019), time attention mechanisms (Woo et al. 2018), etc. Among them, channel-wise attention is prominent in capturing the desirable objects in multi-scale feature maps while spatial attention is positive in locating the object regions in feature maps. In this study, unlike a single attention network such as channel-wise or spatial attention networks used in recent research, we incorporated a hybrid attention mechanism that combined the merits of channel-wise or spatial attention into the plant disease identification model.

Suppose a feature map $f \in R^{W \times H \times C}$ is input into the attention module, the channel-wise attention shrinks the feature map using a global average pooling (GAP) to form a statistic z, and thus an excitation operation is executed to grasp the features of channel dependency with the information accumulated in the shrinking phase, as calculated in Eq. (4).

$$s = F_{ex}(W, z) = \sigma(g(W, z)) = \sigma\left(W_2 \delta\left(W_1 z\right)\right), \tag{4}$$

where σ symbolizes the Sigmoid function, δ is the ReLU function (He et al. 2015b), $W_2 \in \mathrm{R}^{C \times c/r}$, $W_1 \in \mathrm{R}^{(c/r) \times C}$, and r is a hyper-parameter of reduction ratio. In particular, W_1 and W_2 are inferred by two completely linked layers around the non-linearity, and thus the results of the channel-wise attention module can be gained through rescaling u_c using the activations s:

$$\tilde{x}_c = F_{scale}\left(s_c, u_c\right) = u_c \cdot s_c \tag{5}$$

where $\left[\tilde{x}_1, \tilde{x}_2, \cdots, \tilde{x}_c\right]$ indicates the output $\tilde{X}$. Subsequently, the spatial attention module executes the pooling for the input feature map, thereby gaining the spatial attention map. The formula is expressed in

$$F_S(\tilde{X}) = sigmoid\left(c^{7\times7}\left(\left[GMP(\tilde{X});GAP(\tilde{X})\right]\right)\right), \tag{6}$$

where *GMP* and *GAP* denote the global maximum pooling and global average pooling, respectively. $c^{7\times7}$ implies a 7×7 convoluting. After that, the channel-wise attention (*CWA*) and spatial attention (*SPA*) is concatenated using a sequential cascade manner in our network, as written by

$$F_{att} = CWA(f) + SPA(f) = F_C(f) * f + f * F_s(f) \tag{7}$$

In Eq. (7), f denotes the input feature map, and * symbolizes a dot product operation.

2.3 Proposed approach

2.3.1 MobS_Net

To the best of our knowledge, the DL-based models, which usually involve a great number of parameters and have large volumes, require big computational memories for training models. Therefore, they are not suited to be deployed in mobile phone applications because of the limited capacities and computation capability of mobile smartphones. In this study, we select the lightweight CNNs as the backbone network, and the MobileNet-V2 is selected as the backbone extractor in our model for recognizing plant disease types. To improve the learning capability of minute plant disease features, we altered the classical architecture of the MobileNet-V2 using the fine-tuning approach conducted by transfer learning. The atrous convolution along with the SPP module was embedded into the pre-trained network, which was followed by a hybrid attention mechanism containing channel-wise and spatial attention submodules to efficiently extract high-dimensional features of plant disease images.

More specifically, the atrous convolution layer designed in this study consisted of 512 convolution kernels with the size of 3×3, and the atrous rate was assigned as $r=2$, which was used to increase the convolutional receptive fields. Then, following the BN layer for alleviating the vanishing gradient problem, the SPP module was integrated into the network to generate fixed-size feature vectors and make the CNN adapt to the input images with different sizes, thereby extracting multiple-scale features of images efficiently. Other than that, a hybrid attention mechanism that concatenated channel-wise and spatial attention in a cascade manner was embedded into the network, which makes the network infer the interdependence between channels and the importance of spatial points for intermediate features. Ultimately, the completely linked (CL) layer was substituted by a GAP layer, and a new CL Softmax layer with the actual number of classes was embedded as the classification layer of the network. By doing this, the newly formed network, which we termed the MobS_Net, was utilized to execute the task of potato disease recognition. It is noteworthy that the initial parameters of the network were injected by the following (He et al. 2015b).

Figure 3 portrays the network architecture of the proposed MobS_Net, where MobileNet-V2 pre-trained on ImageNet is utilized as the bottom convolution layers, and the atrous SPP module is incorporated into the network for multi-scale feature extraction. Plus, a hybrid attention module comprised of channel-wise and spatial attention is introduced into the network to maximize the reuse of inter-channel relations and infer the importance of spatial points, thereby recalibrating the channel-wise and spatial features. In this manner, we aim to realize a trade-off between the memory requirement and recognition accuracy in the CNN model, i.e., the model volume was compressed while the

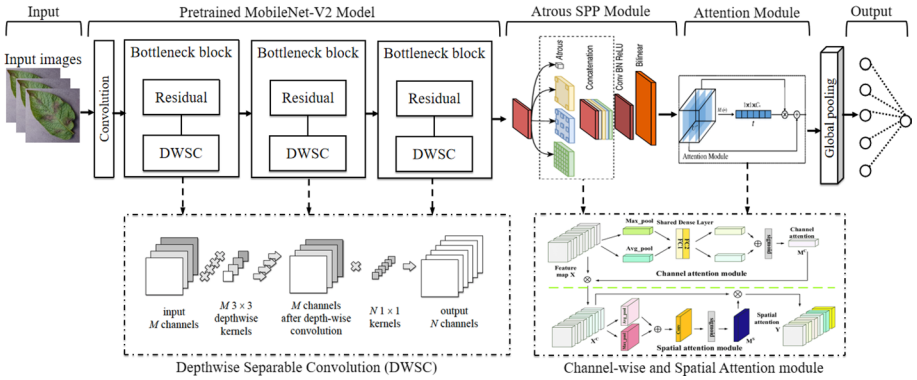


Fig. 3 The structure of MobS_Net

Table 2 The major parameters of MobS_Net

Types (blocks)	Input sizes	Extension factors	Output sizes	Repeat times	Strides
Inputs	224 × 224 × 3	–	224 × 224 × 3	1	–
Conv1_pad	224 × 224 × 3	–	225 × 225 × 3	1	–
Conv1:Conv2d	225 × 225 × 3	–	112 × 112 × 32	1	2
Bottleneck	112 × 112 × 32	1	112 × 112 × 16	1	1
Bottleneck	112 × 112 × 16	6	56 × 56 × 24	2	2
Bottleneck	56 × 56 × 24	6	28 × 28 × 32	3	2
Bottleneck	28 × 28 × 32	6	14 × 14 × 64	4	2
Bottleneck	14 × 14 × 64	6	14 × 14 × 96	3	1
Bottleneck	14 × 14 × 96	6	7 × 7 × 160	3	2
Bottleneck	7 × 7 × 160	6	7 × 7 × 320	1	1
Conv1:Conv2d	7 × 7 × 320	–	7 × 7 × 1280	1	1
Atrous Conv	7 × 7 × 1280	–	7 × 7 × 512	1	1
BatchNormalization	7 × 7 × 512	–	7 × 7 × 512	1	–
PyramidPoolingModule	7 × 7 × 512	–	7 × 7 × 516	1	–
Channel-wise attention	7 × 7 × 512	–	7 × 7 × 516	1	–
Spatial attention	7 × 7 × 516	–	7 × 7 × 516	1	–
Multiply layer: Multiply	7 × 7 × 516, 7 × 7 × 1	–	7 × 7 × 516	1	–
global avg pooling	7 × 7 × 516	–	516	1	–
Visualized layer: dense	516	–	k	1	1

accuracy was improved as much as possible. Table 2 summarizes the relevant parameters of the MobS_Net.

2.3.2 Loss function

Generally speaking, the Cross-Entropy (CE) loss function is frequently employed in CNN models, and the formula of CE loss can be expressed by

$$L(p_k) = -\sum_{k=1}^{C} y_k \log(p_k), \tag{8}$$

where C signifies the class number, and y_k is an indicator variable. If k is the same as the true class of the sample, then $y_k = 1$, otherwise $y_k = 0$. p_k denotes the predicted probability that the observed sample belongs to class k. Because the class loss weights of positive and negative samples are considered the same for the CE loss function, Reference (Lin et al. 2017) reports an FL function to alleviate this unbalanced sample issue. The formula of the FL function is presented in Eq. (9).

$$FL(p_k) = -(1 - p_k)^{\gamma} \theta_k \log(p_k), \tag{9}$$

where γ symbolizes a hyper-parameter of the modulating factor, and θ_k is the weighting factor when the class is 1. It is worth pointing out that the classical FL function is developed to handle the tasks of binary classification in object detection. However, the recognition of plant diseases belongs to a multi-classification task, and thus we modified the FL function and employed the enhanced FL (EFL) in place of the traditional CE function in the plant disease recognition model, as expressed in Eq. (10).

$$EFL(p_k) = -\sum_{k=1}^{C} \theta_k (1 - p(k|x))^{\gamma} y_k \log(p(k)) \tag{10}$$

$$w_k = count(x)/count(x \in k) \tag{11}$$

$$y_k = \begin{cases} 1, & k = actual\_class \\ 0, & k \neq actual\_class \end{cases}, \tag{12}$$

where x denotes the sample.

3 Experimental analysis and results

3.1 Experimental setup

Apart from the image pre-processing task performed by the software of Photoshop, the major algorithms were executed by Python 3.6, where the frequently-used libraries like OpenCV3, Tensorflow, and Keras were employed and accelerated by GPU. The experimental hardware configuration includes GeForce RTX 2080 Graphics Card, 64 GB RAM, and E5-2620V4 CPU, which are utilized for algorithm operation.

3.2 Model training

As mentioned in Sect. 2.1, the potato leaf disease images are utilized in our experiments. Considering the imbalanced samples and the number limitation of sample images, we utilized the data augmentation scheme to enrich the dataset. The commonly-used augmentation methods like color jetting, random rotation, flipping, translation, and other geometric transformation were executed to augment the dataset. Note that color jittering is

altering the contrast, saturation, and brightness of color with a random adjustment variable in (0, 3.1), the rotation range is in [0,360°], the translation range is in ± 20%, and the scale is changed from 0.9 to 1.1. Except for preserving some original images to assess the effect of the model, the proportion of the sample images randomly assigned to the validation and training sets was 1:4. Besides, to compare the proposed approach with other advanced methods, the five influential deep DCNNs including Xception, VGGNet-19, DenseNet-121, ResNet-50, and MobileNet-V2 were chosen as the benchmarks to implement the comparison experiments. Using transfer learning (TL), the original classification layers of the networks were truncated and the new CL layers with Softmax activation functions were embedded into the networks for the classification, where the class number was set as the actual number of potato plant disease types.

With this method, the diverse DCNN models were built and the weights were initialized with the parameters pre-trained on ImageNet (Russakovsky et al. 2015). The hyper-parameters of model training were assigned as a learning rate of 1×10^{-3}, a mini-batch size of 64, 30 epochs, and a stochastic gradient descent (SGD) optimizer. There is standard measure param for image recognition to check the efficiency of the network. These are *Accuracy* (*Acc.*), *Sensitivity* (*Recall*, *Rec.*), *Specificity* (*Spe.*), *FPR* (*false positive rate*, *Fpr.*), and *F1-Score* (F_1). The formulas of these evaluation metrics are expressed in Eqs. (13–17).

$$Acc. = \frac{TN + TP}{TN + FN + TP + FP} \tag{13}$$

$$Rec. = \frac{TP}{TP + FN} \tag{14}$$

$$Spe. = \frac{TN}{TN + FP} \tag{15}$$

$$FPR = \frac{FP}{TN + FP} \tag{16}$$

$$F1 = \frac{2TP}{2TP + FP + FN}, \tag{17}$$

where *TP* represents the number of correct recognition for positive samples. *FN* is the reverse, which denotes the number of mistakenly recognized. *FP* is the number of wrong-identified samples. *TN* implies the number of properly-recognized negative samples. Table 3 summarizes the training results and the performance of diverse methods is depicted in Fig. 4.

From Table 3, it can be visualized that the proposed approach has delivered an increasing performance relative to other advanced methods. After training for 10 and 30 epochs, the proposed MobS_Net has attained the training *Acc.* of 98.63% and 99.87%, respectively. Especially, after 30 epochs of training, the proposed approach realizes a validation accuracy of 94.57%, which is the top performance of all the algorithms. The crucial explanation for the substantial efficiency of the proposed method is that the atrous SPP coupled with a hybrid attention mechanism is embedded into the network, which enhances the capability to extract multi-dimensional features and maximize the reuse of inter-channel relation and spatial point characteristics. Moreover, the TL and enhanced Focal Loss function are applied in the model training, which makes the network gain the optimum weights and

Table 3 The accuracy of diverse methods

Pre-trained models	Training for 10 epochs				Training for 30 epochs				
	Training Acc. %	Validation Acc. %	Training loss	Validation loss	Training Acc. %	Validation Acc. %	Training loss	Validation loss	Run time (min)
Xception	91.99	89.72	2.8887	2.6725	96.25	92.99	1.3123	2.9794	09:16
VGGNet-19	89.07	86.92	0.2787	0.3526	95.02	91.59	0.1346	0.2474	06:17
DenseNet-121	98.61	92.52	0.5956	2.1573	99.86	90.65	0.1652	2.6765	11:03
ResNet-50	94.56	93.46	0.1573	0.1928	98.26	93.93	0.0759	0.1521	06:26
MobileNet-V2	98.06	91.12	0.7973	2.8657	99.72	91.52	0.3204	2.3299	06:24
This study	98.63	91.86	0.4688	2.2855	99.87	94.57	0.1299	1.8525	06:30

alleviates the problem of data imbalance, thereby improving the performance of the model. By comparison, the other methods are single networks and don't attain the ideal performance, although the TL and fine-tuned approach are utilized in model training. Additionally, the running time of the proposed method is 6.30 min, which is the competitive time-consuming of all the compared methods.

3.3 Ablation study

We implemented the ablation study on our model, where we analyzed the efficacy of Atrous SPP and hybrid attention modules on the test dataset of potato disease images. In the first ablation experiment, we separately removed the modules of Atrous SPP and hybrid attention in the network to investigate the performance of the model training. We notice a minor decrease in the results of the ablated model, where the validation accuracy of removing Atrous SPP and hybrid attention modules drop to 92.76% (decrease by 1.80%) and 93.67% (decrease by 0.90%), respectively. Although the effectiveness of ablated models is still better than that of the baseline model, it suffers a minor decline compared with the proposed architecture of MobS_Net. Therefore, this ablation experiment indicates that both the Atrous SPP and attention modules contribute to the performance gain of the proposed approach, and relatively, removing the Atrous SPP has a significant impact on the accuracy compared to the MobS_Net. In the second ablation experiment, we evaluate the effect of the optimized loss function in the recognition of plant diseases on the potato leaf image dataset. To do so, we substitute the enhanced Focal Loss function with the existing Cross-Entropy (CE) one, and we notice a minor decrease in the results of the ablated model, where the validation accuracy drops to 92.41% (decrease by 2.16%). This ablation experiment demonstrates that the enhanced Focal Loss (EFL) function delivers slightly better results than that of the CE loss function used in our model for potato plant disease identification. Table 4 summarizes the comparison results of ablation experiments.

3.4 Recognition results

Therefore, using the trained model of the MobS_Net, we further performed the identification of potato plant diseases on new unseen samples (test set), where unseen samples in

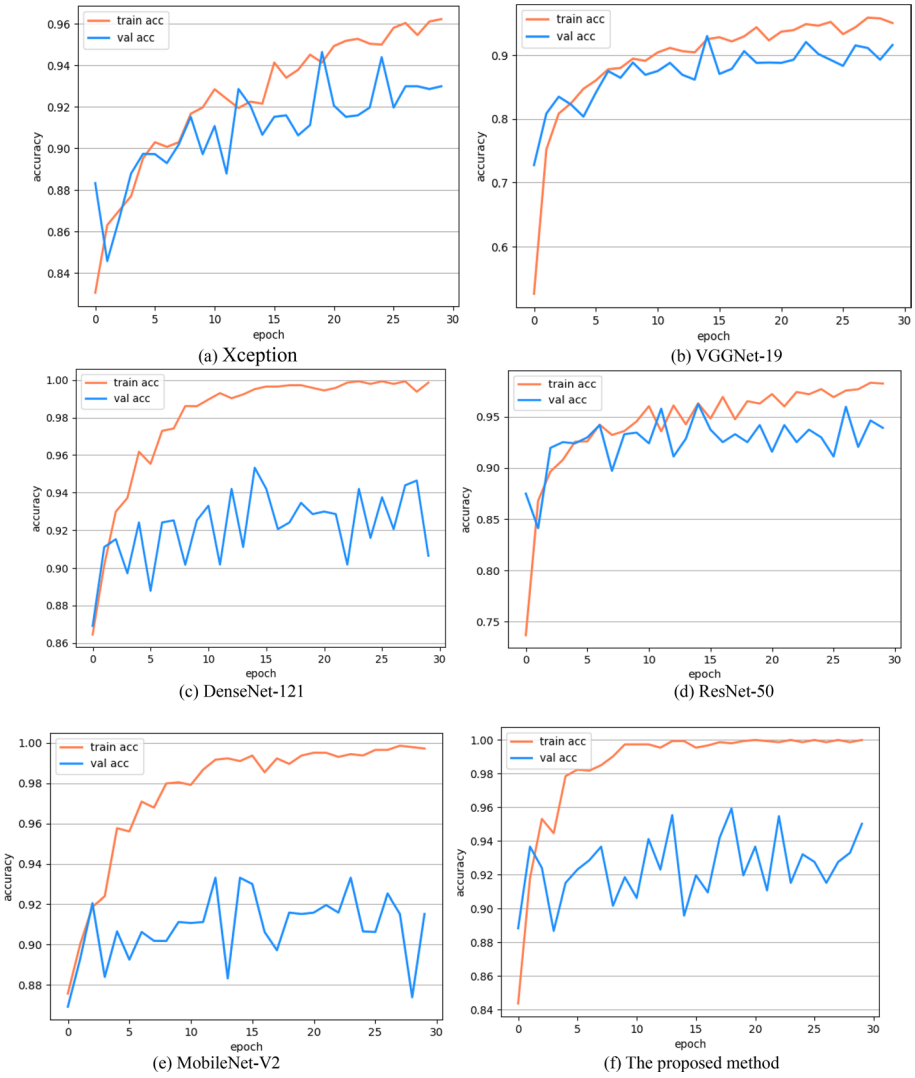


Fig. 4 The training performance of the models

this context indicate potato leaf images that have never been used by the model during the training and validation. Figure 5 is the identification results depicted by the Receiver Operating Characteristic (ROC) curve and confusion matrix. The related measurement metrics are summarized in Table 5.

As seen in Fig. 5a, the curves of most classes are close to the upper left corner of the figure, which exhibits the satisfied operating points of the ROC curve. Plus, it can be also observed from the confusion matrix (Fig. 5b) that MobS_Net has accurately identified most of the sample images. The sample images in the category of Potato healthy and Early_Blight have all been properly identified by the proposed method. For the category

of early blight fungus general, 18 samples have been correctly recognized in 29 samples. Also, 67 images are properly recognized by the proposed method in 73 early blight fungus serious samples, and the accuracy rate reached 98.05%. In summary, a total of 319 images have been properly identified in 413 test samples, and the average recognition *Acc.* attains 97.33%. The average *Rec.* and *Spe.* have also reached no less than 91.99% and 98.39% respectively, as presented in Table 5.

Furthermore, the comparative analysis of our experimental results with that of some latest literature has been summarized in Table 6, where most of the experimental materials are sourced from the potato leaf images of the PlantVillage dataset. As mentioned in Sect. 2.1, the public dataset we tested also comes from the PlantVillage repository, which is the same as the materials used by other methods. In addition, we have identified some local potato disease images with cluttered backgrounds and uneven illumination intensity, which undoubtedly increases the difficulty of potato disease recognition. Nevertheless, a competitive performance has been achieved by our method. The comparison results demonstrate the validity of the proposed approach compared with other state-of-the-art methods.

On the contrary, there are also several misidentified samples, such as 7 misdetections in the category of "Early blight fungus general", which are incorrectly recognized as the type of "Early blight fungus serious". Despite individual misidentifications, most of the sample images have been accurately recognized and the misidentifications are primarily the misclassification of disease severity instead of the potato disease types. Consequently, this reveals the proposed MobS_Net has a certain ability to recognize potato plant diseases. Figure 6 presents the samples of recognized potato disease types.

As shown in Fig. 6, the samples in the top layer are the raw potato disease images, the samples in the middle layer are the disease regions exhibited by visual technology of classification activation map (CAM), and the samples in the bottom layer are the images recognized by the proposed method. It can be observed from Fig. 6 that the recognized classes of most samples are compatible with their actual disease types. For example, the real disease type of Fig. 6a is early serious and this sample is accurately identified by the proposed method with a probability of 0.8513. Likewise, the sample of Fig. 6b is properly identified by the MobS_Net with a high probability of 0.9052. The other samples, such as Fig. 6d and e, have also been properly identified by the proposed method. Despite the impressive performance, there are also some misidentified instances, including the sample of Fig. 6c, which belongs to the category "potato early general" but is wrong classified as category "potato early serious". A certain ambiguity for the severity division of plant disease images may result in this issue. Additionally, the irregular lightweight intensities, which affect the feature extraction of plant disease images, can lead to the misidentification of plant diseases too. Though individual samples were incorrectly recognized, most of the samples have been correctly identified by the proposed approach and the misidentified samples are primarily for the severity level rather than the detailed disease types. Moreover, the predicted probability of misidentification is also relatively low, such as the 0.3609 of the sample in Fig. 6c. Consequently, depending upon the experimental findings, it can be assumed that the proposed approach has successfully performed the identification of potato plant diseases and can also be applied to other domains.

Table 4 The comparison results of ablation experiments

Ablation approach	Training for 10 epochs				Training for 30 epochs				
	Training Acc.%	Validation Acc. %	Training loss	Validation loss	Training Acc. %	Validation Acc. %	Training loss	Validation loss	Run time (min)
Delete atrous SPP	98.83	91.86	0.4150	1.3623	99.93	92.76	0.1152	2.0425	06:12
Delete attention	99.67	91.52	0.2867	2.6520	99.87	93.67	0.1045	2.8042	06:02
CE loss function	87.17	90.50	0.4896	0.4110	94.86	92.41	0.2822	0.2919	05:23
This study	98.63	91.86	0.4688	2.2855	99.87	94.57	0.1299	1.8525	06:30

4 Conclusions

Various plant diseases can result in a disastrous impact on crop growth and food security. To guarantee an adequate supply of foods, the timely and effective identification of plant diseases is of great realistic significance. The latest development in DL has delivered an impressive alternative approach for the automatic identification of plant diseases instead of the traditional manual approaches. Among them, DCNNs are the most popular methods because they can extract features of images automatically and implement the classification. However, due to the great number of parameters and large volumes, the classical DCNNs are not suitable to be deployed on portable device applications and require a large number of annotated images to train models, which is undoubtedly a challenging problem. To this end, this study proposes a novel lightweight network architecture named MobS_Net and uses transfer learning to implement the recognition of potato plant diseases. The pre-trained MobileNet-V2 was chosen as the backbone network of the model, and to enhance the learning ability of minute plant lesion characteristics, we altered the classical architecture of MobileNet-V2 by incorporating

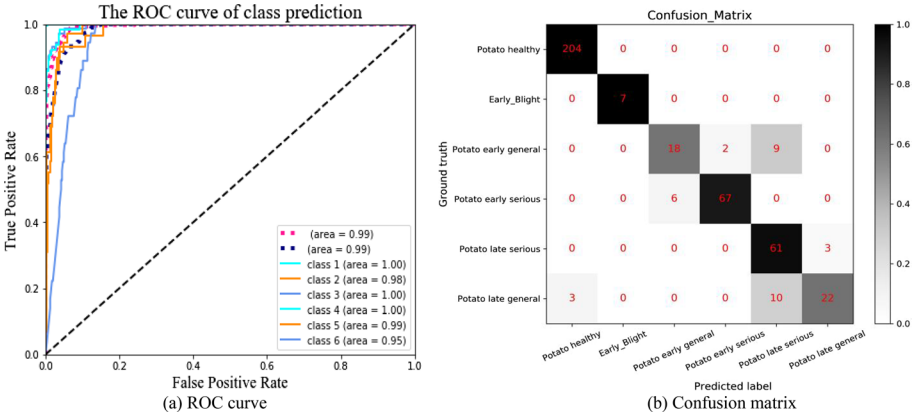


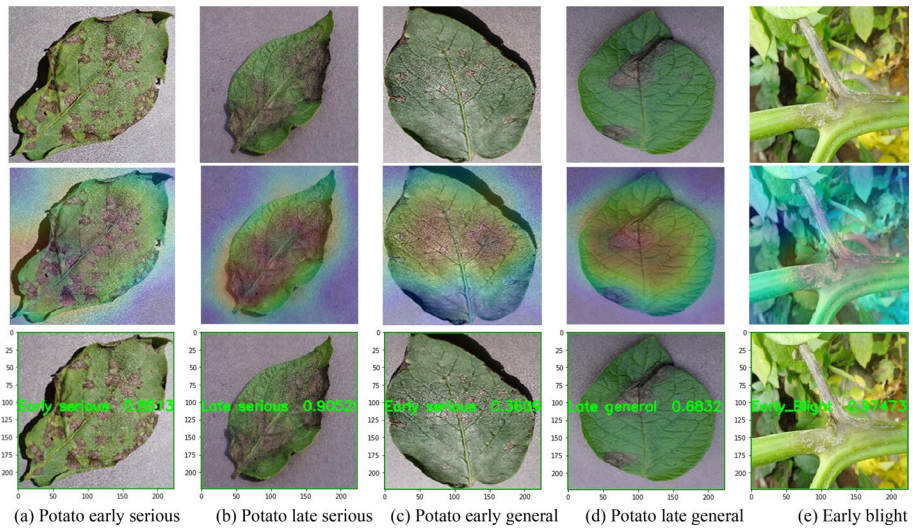
Fig. 5 The visualization of the tested results

Table 5 The metrics assessment for the recognition results

ID	Potato plant types	Identified sample no.	Correct sample no.	*Acc.* (%)	*Rec.* (%)	*Spe.* (%)	*FPR*	*F1* (%)
1	Potato healthy	204	204	99.27	100.00	98.55	0.01442	99.27
2	Early_blight	7	7	100.00	100.00	100.00	0.00000	100.00
	Early blight fungus serious	73	67	98.05	91.78	99.41	0.0059	94.36
3	Early blight fungus general	29	18	95.87	62.07	98.43	0.01566	67.92
5	Late blight fungus serious	64	61	94.66	95.31	94.54	0.05459	84.72
6	Late blight fungus general	36	22	96.12	62.85	99.20	0.0079	73.33
–	*Average*	–	–	97.33	91.99	98.39	0.0160	91.99

Table 6 Comparison with recent work

ID	References	Year	Description	Accuracy (%)
1	Athanikar and Badar (2016)	2016	K-means clustering + image segmentation + backpropagation ANN	92.00
2	El Massi et al. (2017)	2017	K-means clustering + image segmentation + feed-forward ANN	95.30
3	Khalifa et al. (2021)	2019	Deep CNN	94.80
4	Sholihati et al. (2020)	2020	VGGNet-16	91.31
5	Barman et al. (2020)	2020	Self-build CNN (SBCNN)	96.98
6	This study	2021	MobOca_Net	97.73



(a) Potato early serious (b) Potato late serious (c) Potato early general (d) Potato late general (e) Early blight

Fig. 6 The samples of recognized potato disease types

the atrous convolution along with the SPP module into the network. Further, a hybrid attention module containing the channel-wise attention and spatial attention submodules sequentially was embedded into the network to grasp the features of inter-channel dependencies and the significance of spatial points. Experimental findings demonstrate the effectiveness of the proposed method. In future development, we want to assign the model on mobile devices to monitor broader ranges of crop disease information. Moreover, we would like to transplant the model to other domains like online failure detection, computer-aided diagnosis, and virtual defect assessment, etc.

Acknowledgements This study is partially supported by the Fundamental Research Funds for the Central Universities with Grant No. of 20720181004. The authors also wish to appreciate all the judges and editors for their helpful suggestions.

References

Al-Amin M, Karim DZ, Bushra TA (2019) Prediction of rice disease from leaves using deep convolution neural network towards a digital agricultural system. In: 2019 22nd international conference on computer and information technology (ICCIT). IEEE

Al-Hiary H et al (2011) Fast and accurate detection and classification of plant diseases. Int J Comput Appl 17(1):31–38

Athanikar G, Badar P (2016) Potato leaf diseases detection and classification system. Int J Comput Sci Mob Comput 5(2):76–88

Barman U et al (2020) Comparative assessment of deep learning to detect the leaf diseases of potato based on data augmentation. In: 2020 international conference on computational performance evaluation (COMPE). IEEE

Chen J et al (2021a) Attention embedded lightweight network for maize disease recognition. Plant Pathol 70(3):630–642

Chen J et al (2021b) Identification of plant disease images via a squeeze-and-excitation MobileNet model and twice transfer learning. IET Image Process 15(5):1115–1127

Cristin R et al (2020) Deep neural network based Rider-Cuckoo Search Algorithm for plant disease detection. Artif Intell Rev 53(7):4993–5018

El Massi I et al (2017) Automatic recognition of vegetable crops diseases based on neural network classifier. Int J Comput Appl 975:8887

Gassoumi H, Prasad NR, Ellington JJ (2000) Neural network-based approach for insect classification in cotton ecosystems. In: International conference on intelligent technologies

He K et al (2015a) Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans Pattern Anal Mach Intell 37(9):1904–1916

He K et al (2015b) Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. In: Proceedings of the IEEE international conference on computer vision

Hu J, Shen L, Sun G (2018) Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition

Islam M et al (2017) Detection of potato diseases using image segmentation and multiclass support vector machine. In: 2017 IEEE 30th Canadian conference on electrical and computer engineering (CCECE). IEEE

Islam F, Hoq MN, Rahman CM (2019) Application of transfer learning to detect potato disease from leaf image. In: 2019 IEEE international conference on robotics, automation, artificial-intelligence and internet-of-things (RAAICON). IEEE

Ji Y et al (2019) Detection of bruised potatoes using hyperspectral imaging technique based on discrete wavelet transform. Infrared Phys Technol 103:103054

Junde C, Zhang D, Zeb A, Nanehkaran YA (2021) Identification of rice plant diseases using lightweight attention networks. Expert Syst Appl 169:114514

Khalifa NEM et al (2021) Artificial intelligence in potato leaf disease classification: a deep learning approach. In: Machine learning and big data analytics paradigms: analysis, applications and challenges. Springer, Cham, pp 63–79

Lin T-Y et al (2017) Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision
Marino S, Beauseroy P, Smolarz A (2019a) Weakly-supervised learning approach for potato defects segmentation. Eng Appl Artif Intell 85:337–346
Marino S, Smolarz A, Beauseroy P (2019b) Potato defects classification and localization with convolutional neural networks. In: Fourteenth international conference on quality control by artificial vision, vol 11172. International Society for Optics and Photonics
Nanni L, Maguolo G, Pancino F (2020) Insect pest image detection and recognition based on bio-inspired methods. Ecol Inform 57:101089
Oppenheim D et al (2019) Using deep learning for image-based potato tuber disease detection. Phytopathology 109(6):1083–1087
Patil P, Yaligar N, Meena SM (2017) Comparison of performance of classifiers-SVM, RF and ANN in potato blight disease detection using leaf images. In: 2017 IEEE international conference on computational intelligence and computing research (ICCIC). IEEE
Pattnaik G, Shrivastava VK, Parvathi K (2020) Transfer learning-based framework for classification of pest in tomato plants. Appl Artif Intell 34(13):981–993
Russakovsky O et al (2015) ImageNet large scale visual recognition challenge. Int J Comput Vision 115(3):211–252
Sandler M et al (2018) MobileNetV2: inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition
Sholihati RA et al (2020) Potato leaf disease classification using deep learning approach. In: 2020 international electronics symposium (IES). IEEE
Shrivastava VK, Pradhan MK, Minz S, Thakur MP (2019) Rice plant disease classification using transfer learning of deep convolution neural network. Int Arch Photogramm Remote Sens Spat Inf Sci 3(6):631–635
Sifre L, Mallat S (2014) Rigid-motion scattering for image classification. PhD Thesis
Wang W et al (2019) Driver action recognition based on attention mechanism. In: 2019 6th international conference on systems and informatics (ICSAI). IEEE
Woo S et al (2018) CBAM: convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV)

Authors and Affiliations

Junde Chen[1,4,5] · Xiaofang Deng[2] · Yuxin Wen[1] · Weirong Chen[3] · Adnan Zeb[4,6] · Defu Zhang[4]

Xiaofang Deng
lgyxshqk@163.com

Weirong Chen
cwrndnu@163.com

Adnan Zeb
adnanzeb@sustech.edu.cn

Defu Zhang
dfzhang@xmu.edu.cn

[1] Dale E. and Sarah Ann Fowler School of Engineering, Chapman University, Orange, CA 92866, USA

[2] National Academy of Forestry and Grassland Administration, Beijing 102600, China

[3] Department of Information and Electrical Engineering, Ningde Normal University, Ningde 352100, China

[4] School of Informatics, Xiamen University, Xiamen 361005, China

[5] Department of Electronic Commerce, Xiangtan University, Xiangtan 411105, China

[6] College of Engineering, Southern University of Science and Technology, Shenzhen 518000, China