Recognition of grape leaf diseases using MobileNetV3 and deep transfer learning

Xiang Yin[1], Wenhua Li[1], Zhen Li[2], Lili Yi[1*]

(1. *School of Agricultural Engineering and Food Science, Shandong University of Technology, Zibo 255000, Shandong, China*; 2. *School of Artificial Intelligence, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China*)

Abstract: Timely diagnosis and accurate identification of grape leaf diseases are decisive for controlling the spread of disease and ensuring the healthy development of the grape industry. The objective of this research was to propose a simple and efficient approach to improve grape leaf disease identification accuracy with limited computing resources and scale of training image dataset based on deep transfer learning and an improved MobileNetV3 model (GLD-DTL). A pre-training model was obtained by training MobileNetV3 using the ImageNet dataset to extract common features of the grape leaves. And the last convolution layer of the pre-training model was modified by adding a batch normalization function. A dropout layer followed by a fully connected layer was used to improve the generalization ability of the pre-training model and realize a weight matrix to quantify the scores of six diseases, according to which the Softmax method was added as the top layer of the modified networks to give probability distribution of six diseases. Finally, the grape leaf diseases dataset, which was constructed by processing the image with data augmentation and image annotation technologies, was input into the modified networks to retrain the networks to obtain the grape leaf diseases recognition (GLDR) model. Results showed that the proposed GLD-DTL approach had better performance than some recent approaches. The identification accuracy was as high as 99.84% while the model size was as small as 30 MB.

Keywords: grape leaf diseases, real-time recognition, deep transfer learning, MobileNetV3

DOI: 10.25165/j.ijabe.20221503.7062

Citation: Yin X, Li W H, Li Z, Yi L L. Recognition of grape leaf diseases using MobileNetV3 and deep transfer learning. Int J Agric & Biol Eng, 2022; 15(3): 184–194.

1 Introduction

Grape is one of the most popular fruits as well as a source of red wine around the world. However, grape leaf diseases significantly decrease the yield and quality of the grape. It is urgent for farmers to investigate grape leaf diseases efficiently by taking various measures throughout the growth time[1].

Different grape leaf diseases show in the phenotype with different image symptoms. Based on this, researchers make automatic diagnoses using image data and identification algorithms to realize large-scale investigation in a more efficient way. Compared with traditional manual identification methods, automatic identification has higher accuracy and is more efficient. Machine learning (ML) has attracted more and more attention from researchers to apply it to do automatic disease identification of grapes and other crops. Some remarkable research and development results have been published[2-5]. Zhang et al.[2] proposed a *K*-means-based clustering technology to identify cucumber leaf diseases. Zhang et al.[3] proposed an apple leaf disease recognition method that is based on image processing and pattern recognition. Ramcharan et al.[4] applied transfer learning to train convolutional neural networks (CNNs) to identify three different leaf diseases and two types of pest damage. Lu et al.[5] proposed a novel rice diseases identification method based on deep CNNs. These research results are developed by traditional ML or deep learning (DL) methods in ML.

Since the 1980's, different ML approaches, such as database, decision support systems, rule-based systems, fuzzy logic, support vector machine (SVM)[6], discriminant analysis, and *K*-means[2] have been applied to solve disease recognition problems in agriculture. Due to the variety of agricultural diseases and subjective perception, it was difficult to identify leaf diseases at the very early stage to give warnings to avoid disease outbreak using these traditional ML methods. To overcome this problem, image pattern recognition with DL algorithms developed rapidly in leaf disease recognition. For example, a citrus leaf image system[7] was developed based on a scalable vocabulary tree approach. This approach was used to identify healthy, typical Huanglongbing (HLB), alleviated HLB, yellowed, and zinc deficiency. Average identification accuracy achieved as high as 95%.

In the past decades, image processing technology also made a breakthrough in the field of crop leaf disease recognition[8-12]. Rather than manually selecting features to feed traditional ML classification methods, DL[13-15] methods provided end-to-end pipelines to automatically extract advanced robust features and thus significantly improved the usability of agricultural disease identification. This improvement has greatly improved the automatic process of leaf disease identification with sufficient training images and powerful computing resources. However, it was difficult to collect large-scale target disease image data to support DL to establish a classification model. Thus, insufficient training images and limited computing resources were the main

Received date: 2021-09-10 **Accepted date:** 2022-01-20

Biographies: Xiang Yin, PhD, Associate Professor, research interest: agricultural automation and autonomy, Email: 666513@163.com; **Wenhua Li**, Postgraduate, research interest: agricultural automation, Email: L17864306386@163.com; **Zhen Li**, PhD, Professor, research interest: computer vision, Email: lizh0019@gmail.com.

***Corresponding author: Lili Yi**, PhD, Associate Professor, research interest: agricultural automation and autonomy. Room 6-903, School of Agricultural Engineering and Food Science, Shandong University of Technology, Zibo 255000, Shandong, China. Tel: +86-18553308656, Email: yili0001@sdut.edu. cn.

factors that blocked the further improvement in the identification accuracy of crop leaf diseases. Besides that, the independent identically distribution of the training data and the test data is also required, which is also difficult to obtain from practice data. To overcome these shortages, transfer learning (TL)[16-20] is proposed.

TL methods make use of the knowledge learned from the big data to help small-scale datasets to build the target domain model. This method greatly reduced the requirement of dataset scale in the target domain. That was why TL is developed rapidly in many research and engineering fields, especially in the field of crop leaf disease recognition. Researchers have carried out relevant research work[21-24] using the disease images data of some existing datasets with large-scale to assist in establishing the target disease images identification model with small sample data. Long et al.[21] used the deep learning network of TL to conduct image identification research on four kinds of camellia diseases. Arnal et al.[22] studied how the size and variety of the dataset impact the effectiveness of DL techniques applied to plant pathology. Lumini et al.[23] proposed an automatic plankton identification system based on the fusion of different DL methods. In the year 2020, Chen et al.[24] studied TL of the deep CNNs to identify plant leaf diseases. They also used the pre-training model learned from the typical massive dataset, and then transferred it to the specific task trained by their own data.

Among these studies, TL methods have been used extensively in the field of crop disease identification with satisfactory results. However, object identification has not been applied to the real-time monitoring of grape leaf diseases, which has high practical value in the field of agriculture. Another drawback is that the identification methods used to evaluate the quality of the dataset and to reduce computational resources are rarely improved. To overcome these weaknesses of the existed algorithms, a novel grape leaf disease recognition algorithm was proposed in this study.to evaluate the quality of the dataset and to reduce computing resources.

The objective of this research was to incorporate MobileNetV3[25] network model and deep TL (DTL)[26,27] to enhance the learning ability of small sample grape leaf diseases model with decreasing the computational complexity. It provided a sufficient and effective model for grape leaf diseases image with the grape leaf diseases dataset collected by the internet. The proposed algorithm also integrated the image definition detection method to eliminate the fuzzy images, improve the robustness of the CNNs model, and avoid the influence of the image definition on the training accuracy. After that, the image data augmentation[28,29] was also implemented in the proposed algorithm to process the diseased grape images and to generate enough training images to avoid over fitting problems caused by unbalanced sample distribution. In addition, four different processing steps were proposed to improve the learning ability of tiny lesion symptoms alone while decreasing the computational complexity. The processing steps were summarized as follows.

1) A pre-training model[30] was obtained by training MobileNetV3 using the ImageNet dataset to extract common features of the grape leaves. And the last convolution layer of the pre-training model was modified by adding a batch normalization function;

2) The convolution layer was added followed by the dropout layer, which was used to mitigate the occurrence of over fitting, and followed by the fully connected layer, which was used to fuse the output features of the convolution layer for feature classification;

3) The fully connected Softmax layer was added as the top layer of the modified networks to give probability distribution of six diseases;

4) The grape leaf diseases dataset was sent to the modified networks for retraining, and the GLDR model was obtained.

Experiment results demonstrated that the proposed approach achieved excellent performance with a mean average identification accuracy of 99.84%, while the model size was merely 30 MB which was only a small fraction of most existing TL models. In addition, the proposed GLDR model also exhibited higher robustness than state-of-the-art TL methods.

2 Dataset preparation and the proposed algorithm

2.1 Data acquisition

A total of 4344 images of original grape leaf diseases (O-GLD) were obtained, including Black rot, Esca measles, Leaf spot, Downey mildew, Phylloxera, and Healthy. Six different kinds of grape leaf conditions were selected since they are visually identified from leaves while they are responsible for substantial yield reductions in the grape industry. Figure 1 gives representative images of the grape leaves in the dataset.

Figure 1 Six different kinds of grape leaf conditions

Diversity among six grape leaves is easily observed in Figure 1. Healthy leaves are supposed to be light green. Leaves of Leaf spot are composed of rust-yellow spots with brown pinhead-sized points in the center. Leaves of Phylloxera are composed of rhombic or bird head nodules. Downey mildew leaves are composed of polygonal yellowish-brown spots. Leaves of Black rot are composed of relatively small and brown spots. Leaves of Esca measles are composed of dark red spots.

2.2 Image quality filter

To avoid wasting training effort on the original dataset, a filter was introduced to remove low-quality images. The filter was based on the detection of image clarity. Seven different clarity detection algorithms were introduced here.

Grape Brenner gradient (G-BG) function is the simplest gradient evaluation function. It simply calculates a square of gray difference between two adjacent pixels. The clearer the image is, the larger the output value of the Brenner gradient function is. The equation is defined as follows:

$$D_{\text{G-BG}} = \sum_{y}\sum_{x} | f(x+2, y) - f(x, y) |^2 \quad (1)$$

where, x and y are ordinates of pixel points, respectively; $f(x, y)$ is a gray value of pixel (x, y) corresponding to the image f.

Image sharpness is evaluated by calculating values of multiplying two gray differences in each pixel field and then accumulating them one by one. Generally, the larger the value is, the clearer the image is. The equation of grape gray variance

product (G-GVP) is defined as follows:

$$D_{\text{G-GVP}} = \sum_{y}\sum_{x} | f(x,y) - f(x+1,y) | \cdot | f(x,y) - f(x,y+1) | \quad (2)$$

where, $f(x, y)$ is a gray value of pixel (x, y) corresponding to the image f.

Image sharpness is evaluated by values of horizontal and vertical gradients, which are calculated by a Laplacian gradient grape (LG-G) function. The equation is defined as follows:

$$D_{\text{LG-G}} = \sum_{y}\sum_{x} | G(x,y) | \quad (G(x,y) > T) \quad (3)$$

where, $G(x, y)$ is a convolution of the Laplacian operator at pixel (x, y) while T is a given threshold of edge detection; $D_{\text{LG-G}}$ is a final output value. The higher the value of $D_{\text{LG-G}}$ is, the higher the definition of the image.

Because a clear focus image has a greater gray difference than a blurred image, Grape variance (G-VA) function can be used as an evaluation function. The formula is defined as shown in Equation (4).

$$D_{\text{G-VA}} = \sum_{y}\sum_{x} | f(x,y) - u |^2 \quad (4)$$

where, u is the average gray value of the whole image; $D_{\text{G-VA}}$ is the average gray value of the whole image. Similarly, the higher the value of $D_{\text{G-VA}}$ is, the higher the definition of the image.

Grape energy gradient (G-EG) function evaluates image sharpness by calculating gradient values of all pixels, where the gradient value is calculated by a sum of squares of the gray value of adjacent pixels in x and y directions. The calculation process can be expressed by Equation (5).

$$D_{\text{G-EG}} = \sum_{y}\sum_{x} [| f(x+1,y) - f(x,y) |^2 + | f(x,y+1) - f(x,y) |^2] \quad (5)$$

where, $f(x, y)$ is a gray value of the pixel (x, y) corresponding to the image f.

The function of grape vollath (G-V) is defined as shown in Equation (6).

$$D_{\text{G-V}} = \sum_{y}\sum_{x} [f(x,y) \cdot f(x+1,y) - M \cdot N \cdot u^2] \quad (6)$$

where, u is an average gray value of the whole image; M and N are the width and height of the image, respectively.

Grape entropy (G-E) function is based on the entropy function of statistical features to measure the richness of image information and further test the clarity of the image. According to information theory, the clarity of the image is evaluated by calculating the information entropy of the image. And the calculation formula of information entropy is shown in Equation (7).

$$D_{\text{G-E}} = \sum_{i=0}^{L-1} p_i \ln(p_i) \quad (7)$$

where, p_i is the probability of pixels with gray value i; L is the total number of gray levels. According to Shannon's information theory, the maximum value of entropy means the most information. Applying this principle to the focusing process, the higher the value of $D_{\text{G-E}}$ is, the clearer the image is.

To make use of this feature of the images, a python program was implemented with a certain threshold to select the images to achieve the requirements. The threshold for each function was obtained by experiments.

Take Esca measles-infected leaves as an example. The image shown in Figure 2 is used as input for the detector. Parameters and thresholds of the detector are listed in Table 1. When any one of the return values of differential equations was lower than the threshold, the image was automatically removed.

Figure 2 Image example of Esca measles-infected leaf

Table 1 Detection value and threshold of the O-GLD dataset definition detection for different functions

Function	Detection value	Threshold
$D_{\text{G-BG}}$	58 273 342	20 000 000
$D_{\text{G-GVP}}$	26 701 242	1 000 000
$D_{\text{LG-G}}$	5409	2000
$D_{\text{G-VA}}$	222 593 147	50 000 000
$D_{\text{G-EG}}$	81 879 921	30 000 000
$D_{\text{G-V}}$	192 571 933	10 000 000
$D_{\text{G-E}}$	5	3

2.3 Data augmentation

After selecting and sorting by the proposed detector, the number of grape leaf images set was reduced from 4344 to 2277. Among those 2277 images, there were 500 images with Black rot disease, 500 Esca measles images, 500 Leaf spot images, 104 Phylloxera images, 173 Downey mildew images, and 500 healthy leaf images. Details of the dataset are presented in Table 2.

Table 2 Proportion of grape leaf images with different diseases

Label	Class	Number	Proportion/%
0	Phylloxera	104	4.57
1	Black rot	500	21.96
2	Esca measles	500	21.96
3	Leaf spot	500	21.96
4	Downey mildew	173	7.59
5	Healthy	500	21.96

From Table 2, it was easy to see that the dataset had unbalanced sample distribution, which made it difficult to extract the classification features with a small sample size because of overfitting and reduced accuracy and robustness of the model.

The problems of the unbalanced distribution of samples in the training stage of CNNs can be overcome by data augmentation. For data augmentation, several algorithms were implemented, including disturbances of contrast, sharpness, brightness, rotation transformations, and flip. Inverse filtering noise processing operation was also applied. Besides that, online data augmentation was also adopted in this step to improve the computational efficiency of data augmentation. After the above operations, six new images were generated from each image, as shown in Figure 3.

2.4 The proposed approach

Inspired by the MobileNetV3 and DTL, a novel grape leaf diseases recognition approach GLD-DTL was proposed. The proposed approach was composed of two parts as shown in Figure 4. The first part was the pre-training module, which was used as a basic feature extractor, and the second part was a classifier that made use of feature maps for detection. The lightweight MobileNetV3 with complementary search technology was incorporated into the proposed approach to make the whole system as light as possible. The pre-training model was obtained by distillation training of MobileNetV3 model on the ImageNet dataset. Then, the last

convolution layer of the pre-training model was modified by adding a batch normalization function. A dropout layer followed by a fully connected layer was used to improve the generalization ability of the pre-training model and to realize a weight matrix to quantify the scores of six diseases. After that, the Softmax method was added as the top layer of the modified networks to give the probability distribution of six diseases. The detailed descriptions of the proposed approach are illustrated as follows.

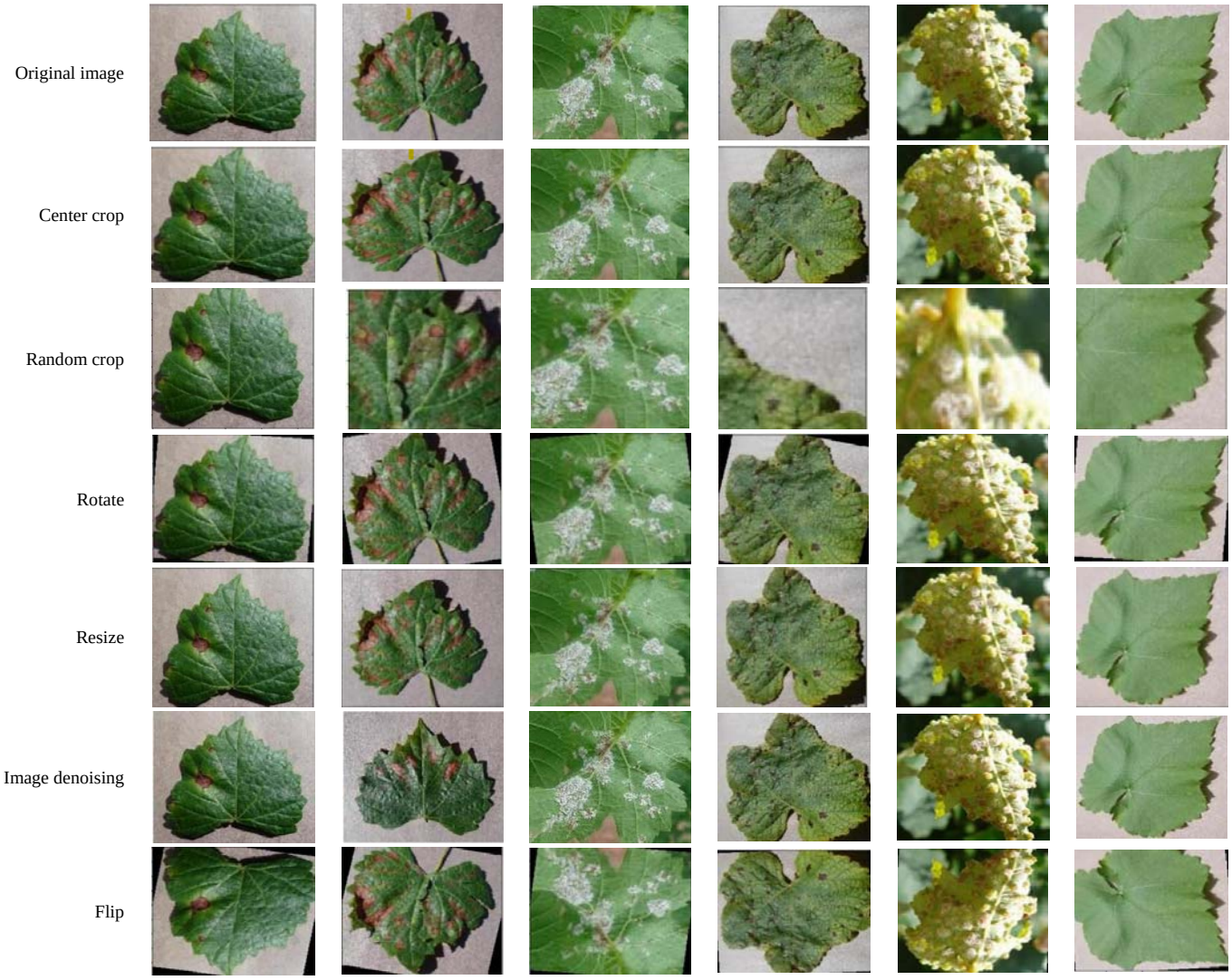


Figure 3 Data augmentation of grape leaf diseases images

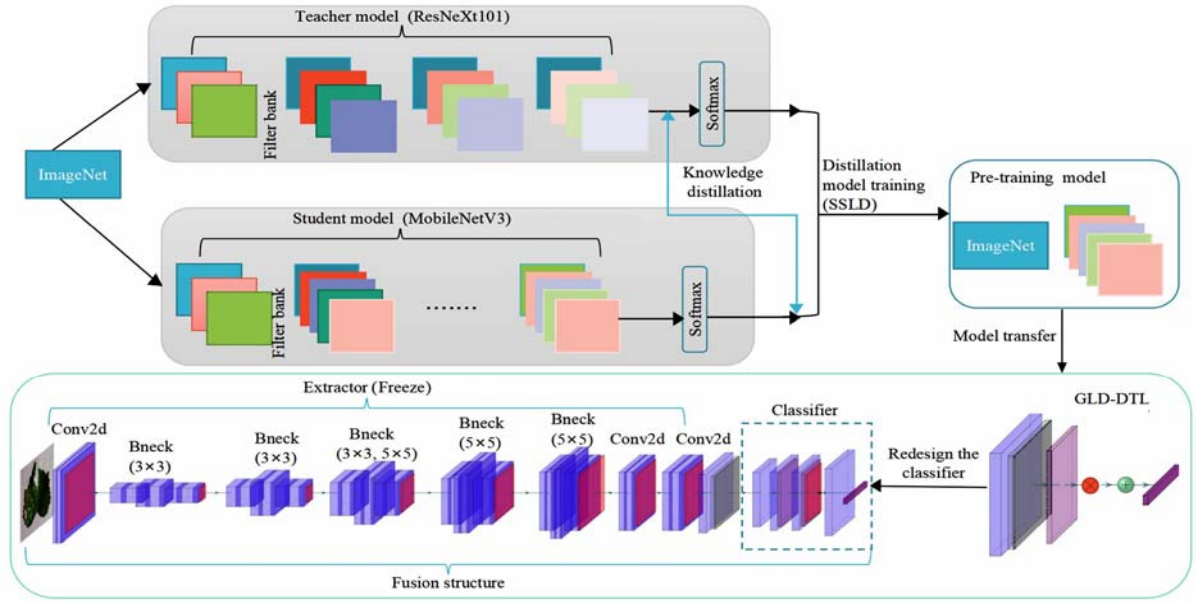


Figure 4 Overall structure of GLD-DTL approach

2.5 MobileNetV3 network model

MobileNetV3 is a lightweight convolutional neural network model with fewer network parameters, high accuracy, fast recognition, and training speed. Therefore, the lightweight MobileNetV3 with complementary search technology was selected as the basic model in the proposed approach.

2.6 Knowledge distillation

Knowledge distillation[31] is a model compression method based on the training of "teacher-student network thought". Because of its simplicity and efficiency, it is widely used in industrial applications. The Knowledge distillation technology was also adopted in this study to compress the pre-training model to reduce computational complexity. Semi-supervised label knowledge distillation (SSLD) algorithm was also used in the proposed approach, which used prediction information of ResNeXt101[32] as a label to supervise the learning of MobileNetV3 model. The specific process is shown in Figure 4. Firstly, the ImageNet dataset was input into ResNeXt101 model and MobileNetV3 model, respectively. Then, the output of ResNeXt101 model was taken as the real value to measure the loss between the output of MobileNetV3 model and itself. Finally, the weight of the MobileNetV3 model was updated by gradient descent, which made the output of the MobileNetV3 model closer to the ResNeXt101 model, to achieve the effect of using a small model to fit a large model. The pre-training model of compressed MobileNetV3 was obtained in this way.

2.7 Model transfer

By the above operations, MobileNetV3 was determined to be a basic network in the proposed approach. And the pre-training model was obtained by training the MobileNetV3 with the ImageNet dataset and the knowledge distillation method. Common features of grape leaves were extracted by this pre-training model. Based on this, the pre-training model needed to be used to allocate network weights in this stage. After that, the bottom layer of the pre-training model was modified to establish a new network structure. Finally, the new neural network was trained using grape leaf diseases dataset to get the final recognition model of grape leaf diseases. Details are described as follows.

Considering that the first several layers of MobileNetV3 pre-training model usually extracted color and corner features, which are common to grape leaf diseases dataset and ImageNet dataset. Therefore, all convolution layers of the pre-training model were preserved, and the last layer of the pre-training model was replaced with an extended convolution layer of $576\times1\times1$ in which hard-Swish was used as the activation function instead of the ReLU. The hard-Swish activation function was followed by a dropout layer, which was used to alleviate the occurrence of overfitting in the process of over model training. The fully connected layer was added after the dropout layer and followed by a Softmax classifier. Figure 5 depicts the network structure added by the proposed approach. A new network model was obtained in this way.

Reasons for this research adding each layer in Figure 5 to our new network model are described below. In the new network model, the convolution layer was added to the global average pooling layer for smaller feature maps at a faster speed. The hard-Swish activation function was added to the convolution layer to improve the numerical stability and speed up the calculation. The dropout function was added to the extended structure to avoid over fitting during model training. The matrix multiplication and the matrix addition were carried out in the fully connected layer with weight (Cls_out_w) of [1280×6] and bias (Cls_out_b) [6×1] to obtain the feature classification results. The Softmax classifier was added to the fully connected layer to calculate the final classification results.

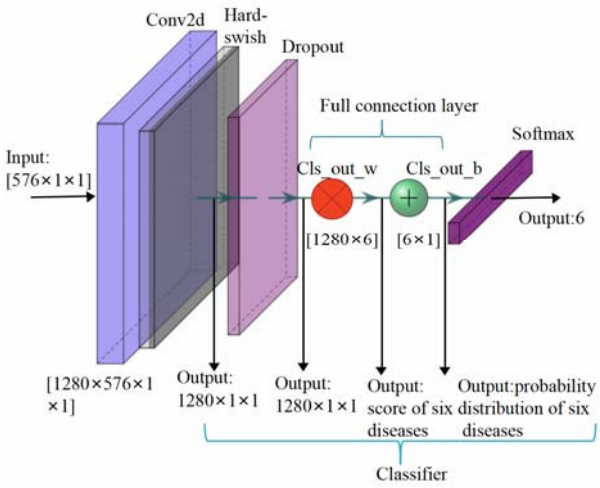


Figure 5 New extended network architecture

2.8 TL steps

Since the convolution blocks before the added TL module serving as backbone image feature extractor was pre-trained from ImageNet while the added TL module was initialized with random weights, a direct fine-tune of the whole network on the grape dataset in this study will probably impose a negative effect on backbone when TL module learns from random weights. In view of this, a three-stage training strategy of the proposed TL was used:

1) Freezed the backbone module so that the gradients in the image feature extractor was not backward propagated. Perform backward propagation on the TL module with a normal learning rate, and train the network until convergence.

2) Unlock the backbone convolution blocks and perform 1/10 of the normal learning rate on the whole network including backbone module and the TL module, and train the network until convergence.

3) Train the whole network with a normal learning rate until convergence.

3 Results and discussion

3.1 Experimental setup

This experiment was conducted on Windows server with an Intel (R) Core (TM) 7-9750H CPU @ 2.60GHz that was accelerated by an NVIDIA RTX2060 GPU (6 GB memory). In addition, the PaddlePaddle DL framework was used to implement the proposed model. Additional configuration parameters and training hyper parameters are listed in Table 3.

Table 3 Software and hardware environment for experiment

Configuration item	Description
CPU	Intel® Core (TM) i7-9750H
GPU	NVIDIA RTX2060 6 GB
Hard disk	1TB
Operating system	Windows10
Python	3.7.8
Batch size	8
Optimizer	Adam
Learning rate	0.0002

3.2 Dataset

In this phase, the method proposed in Section 2.2 was used to automatically delete the blurred images in the O-GLD dataset to

ensure the quality of the dataset. In addition, at least 500 images of each category were ensured by the data augmentation technique in Section 2.3. The images of grape leaf diseases were uniformly resized to the fixed-dimension of 224×224 pixels to fit the model, and 500 pre-processed images of each type were selected to construct the final grape leaf diseases (F-GLD) dataset. To perform the experiments, 80% of the F-GLD dataset was used for training and the other 20% was used for testing.

3.3 Evaluation metrics

In order to better evaluate the performance of the model of this study, accuracy, recall, false positive rate (FPR), true positive rate (TPR), precision, and the harmonic mean (F1) were taken as the evaluation metrics. Definitions of these metrics are shown as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{8}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{9}$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \tag{10}$$

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{11}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{12}$$

$$\text{F1} = \frac{2\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{13}$$

where, True positive (TP) is the number of instances that belongs to the class and are correctly identified by the classifier, False negative (FN) is on the contrary, which is the number of instances that belong to the class but are incorrectly classified. False positive (FP) is the number of instances that do not belong to the class but are mistakenly identified as this classification. True negative (TN) is the number of instances that are not in the class, and they are correctly identified.

3.4 Evaluation of image definition detection

To evaluate the proposed filter based on image definition detection, the dataset was divided into the F-GLD dataset after image definition detection and the O-GLD dataset without image definition detection. The proposed network model was used to train and test the two datasets. The experimental results showed that under the same training conditions, the recognition accuracy of the model trained with the F-GLD dataset was 99.84%, and the recognition accuracy of the model trained with the O-GLD dataset was 73.20%. Details of the experimental results are illustrated as follows.

It could be seen from the curve shown in Figure 6 that the accuracy of the model trained by the F-GLD dataset rose rapidly in the test process. The curve also had little fluctuation, and the curve tended to be stable after about 250 iterations. In contrast, the accuracy of the model trained by the O-GLD dataset rose relatively slowly. The fluctuation of the curve was larger than the former one, and it did not reach a stable state as shown in Figure 6.

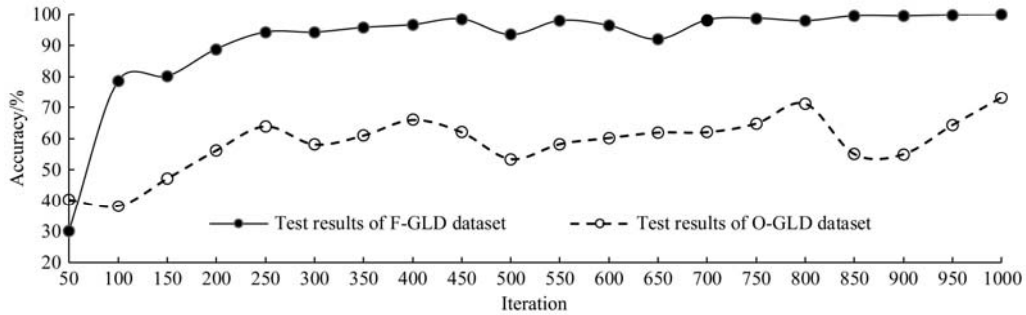


Figure 6 Effects of F-GLD and O-GLD datasets model recognition accuracy

In addition, it could be seen from Table 4 that the model size of the F-GLD dataset after training was 30 MB, and the training time of the model was 1.15 min. While the model size of the O-GLD dataset after training was 55 MB, and the training time of the model was 3.13 min. Obviously, the size of the model trained with the F-GLD dataset was smaller, while the training time of the model was also shorter.

Table 4 Performance of two datasets

Parameter	Dataset	
	F-GLD	O-GLD
Picture number	2277	4344
Training time/min	1.15	3.13
Identification accuracy/%	99.84	73.20
Model size/MB	30	55

Overall, the proposed model training with the F-GLD dataset constructed by an image sharpness detector had higher accuracy, smaller model size, and shorter training time than the model training with the O-GLD dataset. Therefore, the proposed filter based on image definition detection could effectively improve the performance of the model.

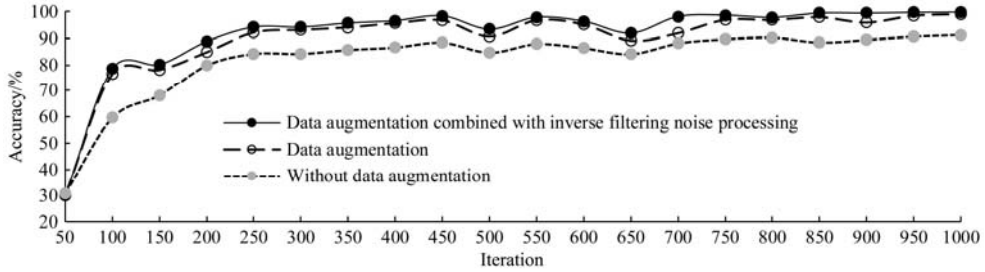
3.5 Evaluation of data augmentation

To solve the problems caused by the uneven distribution of samples and to prevent the occurrence of overfitting, a variety of digital image processing technologies, such as rotation transformations and flip, were proposed to simulate the real acquisition environment, and increase the diversity, and quantity of grape leaf training images. In addition, the inverse filtering noise processing operation was also applied to improve image clarity.

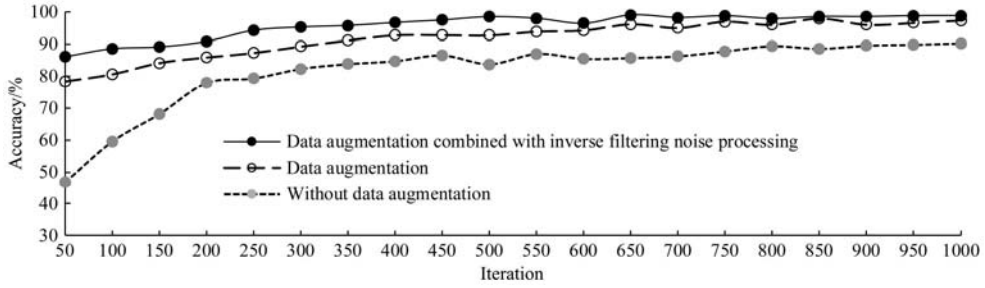
To analyze the performance of the image augmentation strategy proposed in Section 2.4, 80% of standard datasets were used for the non-data augmentation test, data augmentation test, and inverse filtering noise processing data augmentation test, respectively. 20% of standard datasets were used for verification. Experiment results are shown in Figure 7.

From Figure 7, it could be seen that the model without data augmentation had high training cost and test loss, and low training and recognition accuracy. The final training accuracy was 91.32% and the final testing accuracy was 90.15%. By data augmentation, the loss of model training and testing was reduced, while the accuracy of training and recognition was improved. The final training accuracy was as high as 99.12% and the final testing accuracy was 97.33%. Compared with the previous two

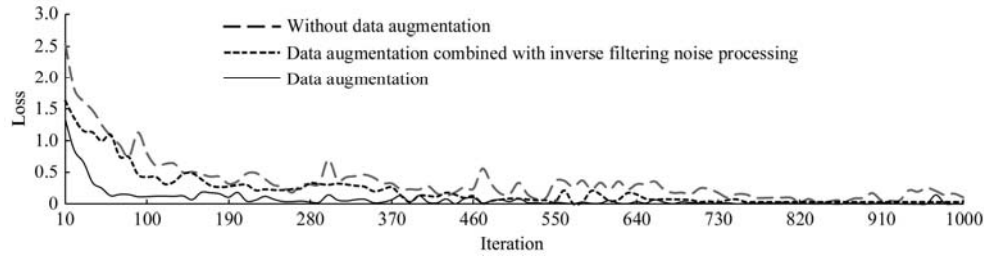
experiments, the loss of model training and testing was significantly reduced after adding the data augmentation of inverse filtering noise processing. The accuracy of training and recognition was also greatly improved, while the final training accuracy was as high as 99.84%, which was 8.42% higher than the training accuracy without data augmentation, and the final testing accuracy was as high as 98.78%, which was 8.63% higher than the testing accuracy without data augmentation. The experimental results showed that under the condition of ensuring the image quality, the scale of the training dataset had an impact on the accuracy of the model. The model would accumulate noise in the learning process when the accumulation reached a certain level. This accumulated noise would have a significant impact on the recognition results. The data augmentation strategy proposed in this study not only expanded the scale of the dataset, but also reduced the noise of the dataset.



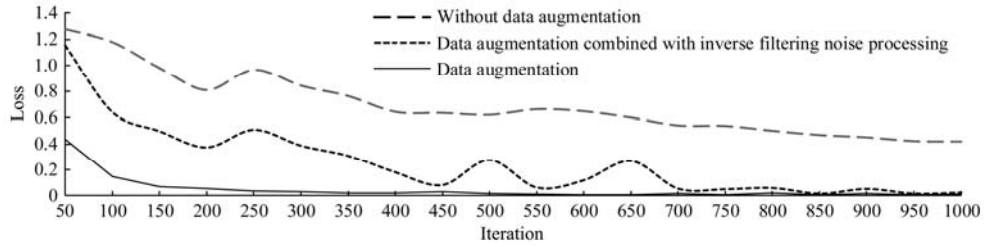
a. Training accuracy



b. Testing accuracy



c. Training loss



d. Testing loss

Figure 7 Effects without data augmentation, with data augmentation and data augmentation combined with inverse filtering noise processing on the accuracy and loss in the training and testing model

3.6 Evaluation of GLD-DTL approach

For comparison, several deep convolution networks including Xception65, Xception71[33], SE_ResNeXt50, ResNet50, VGG13, VGG16, VGG19, and MobileNetV3 were trained and tested. In the training process, stochastic gradient descent (SGD) strategy was used to calculate weights and bias sets of the neural network to minimize the loss function. To solve the problem of high variance parameter updating and unstable convergence when SGD strategy directly adopts single sample training, the batch size was set to 8 and the learning rate was set to 0.002. All convolution layers used a batch normalization layer. To determine how the converge speed of the SGD strategy, the momentum, which serves as an additional factor, was set to 0.9.

Performances of various models are summarized in Table 5. The model sizes of MobileNetV3, Xception71, Xception65, SE_ResNeXt50, and ResNet50 were much smaller than VGG13, VGG16, and VGG19. At the same time, the model with a high recognition accuracy of the traditional neural network model on the F-GLD dataset was not the ResNet50 model with a multi-level network, but the VGG19. Although the VGG model had higher test precision than other traditional neural network models, it had a large size, which directly increased the computational cost. Compared with the VGG19, the accuracy of the GLDR model was 1.26% higher, and the size of the model was 700 MB less. It was

indicated that the GLDR model in the GLD-DTL approach achieved the best performance in terms of final accuracy and model size in this comparative experiment, which was higher than the accuracy of the existing model based on TL.

Gradient-weighted class activation mapping (Grad-CAM) can provide a good visual basis for the classification results of the model. To further analyze the reasons for the difference in recognition accuracy, this study extracted the Grad-CAM of some test images in the comparative experiment of each network module. The test results are shown in Figure 8.

Table 5 Performance of nine models using the F-GLD dataset

Models	Identification accuracy/%	Model size/MB
Xception71	92.51	377.0
Xception65	87.52	384.0
SE_ResNeXt50	95.43	105.0
ResNet50	96.55	98.2
VGG13	94.32	616.0
VGG16	98.43	697.0
VGG19	98.58	778.0
MobileNetV3	98.49	34.2
GLDR (GLD-DTL)	99.84	30.0

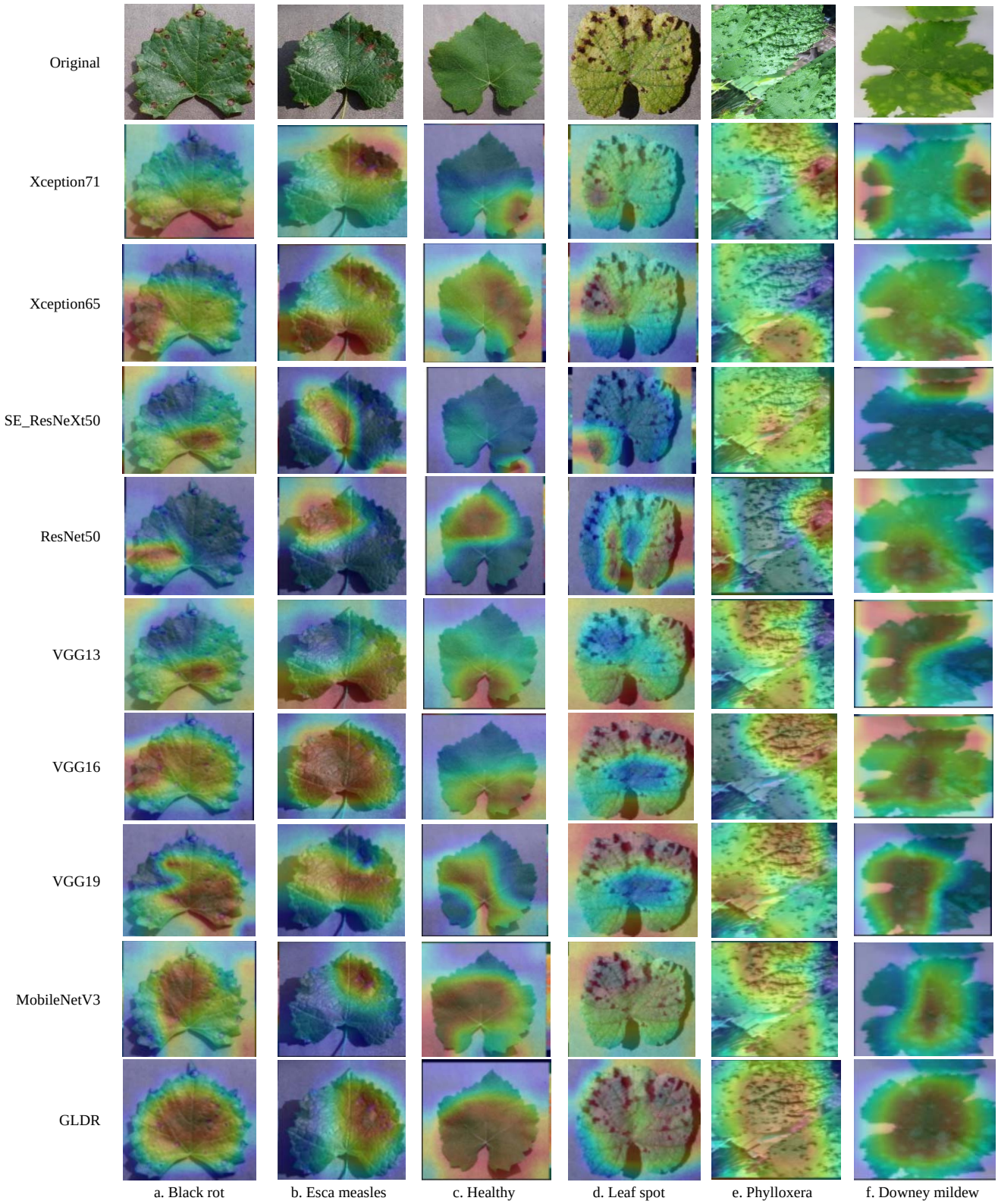


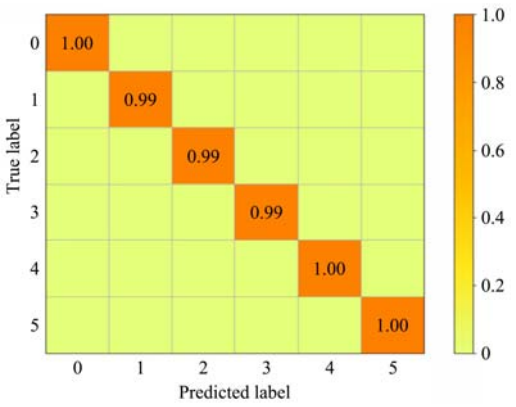
Figure 8 Grad-CAM comparison for nine models

It could be observed from Figure 8 that the GLDR model performed better to locate and select the pest area of leaves than other models. It also had a judgment effect on correct classification.

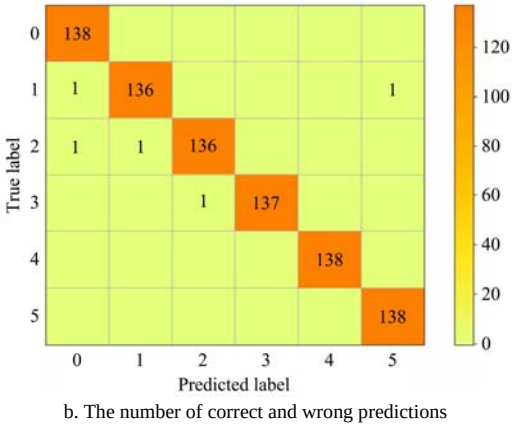
3.7 Confusion matrix

The performance of classifiers is crucial to the final classification effect of the model. However, when facing multiple similar shape classes, classifiers may be confused. Moreover, the images of infected grape leaves in different periods and different backgrounds may also lead to the high complexity of the patterns displayed in the same class. Meanwhile, the performance of the classifier will be degraded. To further evaluate the performance of the classifier to determine the accuracy of classification, a confusion matrix was introduced.

The confusion matrix is shown in Figure 9. The deeper the color in the visualization results, the higher the prediction accuracy of the model in the corresponding class. All correct predictions were on the diagonal and all incorrect predictions are off the diagonal. Hence, the confusion matrix shown in Figure 9a can be used to intuitively observe and evaluate the performance of the model in various categories. In addition, according to the evaluation metrics in Section 3.3, the accuracy rate and recall rate of each category corresponding to the model could be calculated using the values in Figure 9b. The detailed description of this experiment is illustrated as follows.



a. Accuracy of prediction



b. The number of correct and wrong predictions

Figure 9 Confusion matrix of the GLDR model identification results

According to the confusion matrix in Figure 9a and the analysis of six samples in Section 2.1, the features of Phylloxera and Downey mildew diseases and Healthy differed substantially from those of other diseases and reached recognition rates of 100%. Moreover, compared with other classes, the identification was more prone to confusion in distinguishing Black rot from Phylloxera and Healthy, and in distinguishing Esca measles from Black rot and Phylloxera. Among the tested 138 images with Black dots, only two images were wrongly identified. One was wrongly identified as Phylloxera and the other one was wrongly identified as Healthy. Among the tested 138 images with Esca measles, also two images were wrongly identified. One was wrongly identified as Black rot while the other one was wrongly identified as Phylloxera. Among the 138 Leaf spot images, only one image was wrongly identified as Esca measles. The reason caused this phenomenon was that the geometric characteristics of the diseases were very similar to each other. Except for these two, the remaining classes were well differentiated.

Although the confusion matrix provided an explanation for the low recognition accuracy of several types of experiments, the statistics in the confusion matrix is the number, which could not directly measure the advantages and disadvantages of the model. For further analysis of the performance of the GLDR model, four indexes in Table 6 and Figure 10 are added according to the statistical results of the confusion matrix shown in Figure 9b.

Table 6 Four metrics of the GLDR model identification results

Label	Class	Metrics			
		Precision/%	Recall/%	F1/%	Accuracy/%
0	Phylloxera	99.0	100.0	99.0	99.4
1	Black rot	99.0	99.0	99.0	
2	Esca measles	99.0	99.0	99.0	
3	Leaf spot	100.0	99.0	100.0	
4	Downey mildew	100.0	100.0	100.0	
5	Healthy	99.0	100.0	100.0	

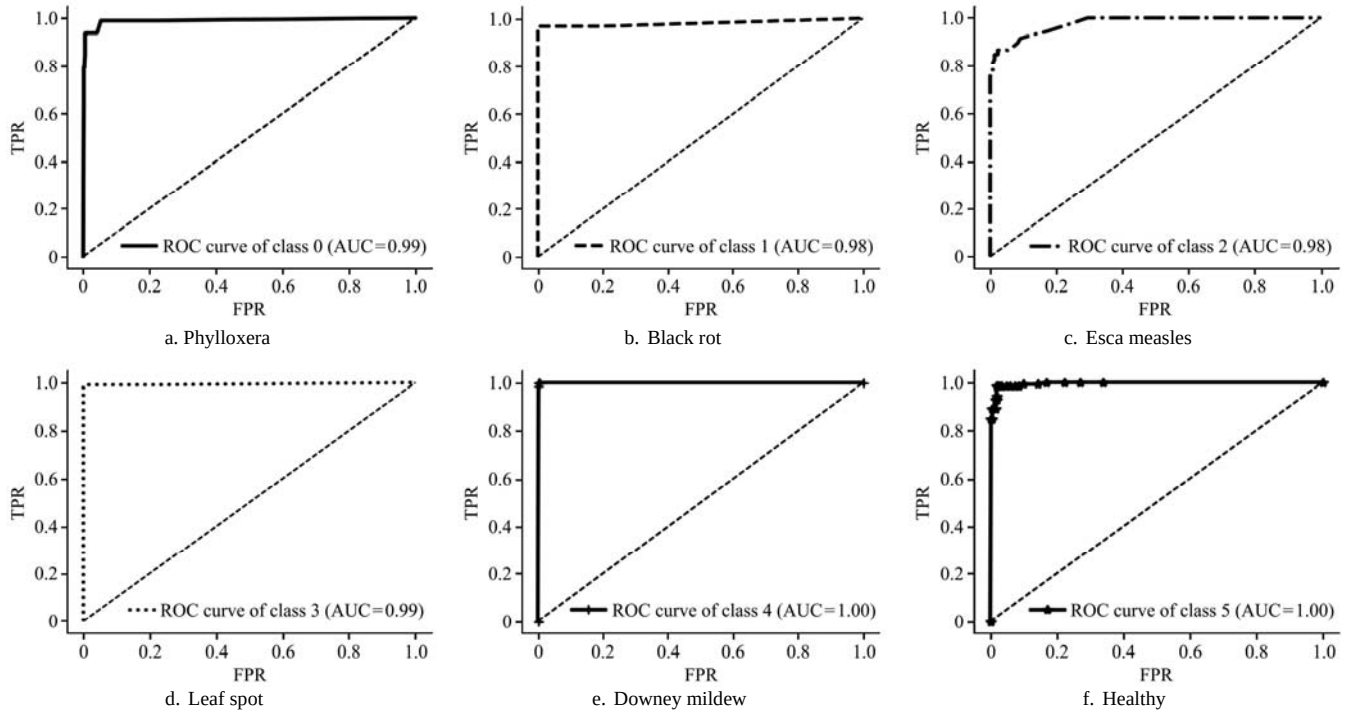
Accuracy is an evaluation of the overall accuracy of the classifier summarized in Table 6. Precision is an evaluation of the accuracy of the classifier's prediction for a certain category. The proportions of the samples that are positive in all the samples are predicted to be positive. A recall is an evaluation of the proportion that is predicted to be positive in all samples that are positive. F1 is a harmonic mean of precision and recall.

From results of precision and recall, Downey mildew had the best recognition effect, with the precision and recall values reaching 100%, followed by Leaf spot. In addition, it could be seen from Table 6 that the accuracy, recall, precision, and F1 of the six samples were extremely high. While the difference among them was small. The above results showed that the proposed GLDR model has good performance and high classification accuracy.

Considering the problem that the distribution of training samples was balanced but the distribution of test samples was unbalanced, the experiments of receiver operating characteristic (ROC) curve and area under ROC curve (AUC) were added to further evaluate the performance of the model, to avoid misjudgment of the experimental results caused by this problem.

As shown in Figure 10, the abscissa of ROC curve is FPR, and the ordinate is TPR. The curve is based on the confusion matrix, which is used to measure the overall performance of the classification algorithm under any distribution or any error cost. In the case of unbalanced samples, the ROC curve can still evaluate the performance of the classifier better. AUC is the area under the ROC curve, and the area under the curve is used as the evaluation index of the model. Generally, the larger the AUC value, the better the performance of the classifier.

From Figure 10, it was easy to see that the distance between each disease sample and the upper left corner was very close, and its value range was 0.98-1, which proved that the GLDR model had high accuracy and strong generalization performance.



a. Phylloxera b. Black rot c. Esca measles

d. Leaf spot e. Downey mildew f. Healthy

Note: TPR: True positive rate; FPR: False positive rate; ROC: Receiver operating characteristic; AUC: Area under the ROC curve.

Figure 10 AUC of the GLDR model identification results

4 Conclusions

In this study, an approach to grape leaf disease recognition based on DTL, namely, GLD-DTL, was proposed by introducing MobileNetV3 initial structure. This approach could automatically extract the discriminative features of the diseased grape leaf images and detect the six common types of grape leaf diseases with high accuracy. To ensure good generalization performance of the proposed model and a sufficient grape leaf disease images dataset, 4344 images of grape leaf diseases with uniform and complex background were collected in the network to form the O-GLD dataset. Then, the O-GLD dataset was processed by an image sharpness detector to eliminate the unqualified images automatically. Finally, the O-GLD dataset was further processed by data augmentation technology to solve the problem of uneven distribution of samples in the dataset and generate the F-GLD dataset.

In addition, the pre-training model was trained by training MobileNetV3 using the ImageNet dataset and the SSLD algorithm to extract common features of the grape leaves. And the last convolution layer of the pre-training model was modified by adding a batch normalization function. A dropout layer followed by a fully connected layer was used to improve the generalization ability of the pre-training model and realize a weight matrix to quantify the scores of six diseases, according to which the Softmax method was used to give probability distribution of six diseases. Finally, the F-GLD dataset was input into the improved network for retraining, and the GLDR model was obtained.

Results showed that the proposed GLD-DTL approach provided a solution with an accuracy of 99.84% using the GLDR model for the diagnosis of grape leaf diseases at an early time. And compared with the VGG19 model, the accuracy of GLDR model was improved by 1.26%. Meanwhile, the size of the model was as small as 30 MB. Hence, the proposed approach was fully capable of identifying grape leaf diseases. The results demonstrate that the proposed GLDR model could recognize the six common types of grape leaf diseases with high accuracy and provides a feasible solution for the rapid identification of grape leaf diseases.

Acknowledgements

The authors acknowledge that this work was financially supported by the National Natural Science Foundation of China (Grant No. 32171910); the Natural Science Foundation of Shandong Province (Grant No. ZR2020MC085); the Key R&D Project of Shandong Province (Grant No. 2019JZZY010734); the Key R&D Project of Zibo City, Shandong Province, China (Grant No. 2019ZBXC143).

[References]

[1] Ji M M, Zhang L, Wu Q F. Automatic grape leaf diseases identification via United Model based on multiple convolutional neural networks. Information Processing in Agriculture, 2020; 7(3): 418–426.

[2] Zhang S, Wu X, You Z, Zhang L. Leaf image based cucumber disease recognition using sparse representation classification. Computers and Electronics in Agriculture, 2017; 134: 135–141.

[3] Zhang C L, Zhang S W, Yang J C, Shi Y C, Chen J. Apple leaf disease identification using genetic algorithm and correlation based feature selection method. Int J Agric & Biol Eng, 2017; 10(2): 74–83.

[4] Ramcharan A, Baranowski K, McCloskey P, Ahmed B, Legg J, Hughes D P. Deep learning for image-based Cassava disease detection. Frontiers in Plant Science, 2017; 8: 1852. doi: 10.3389/fpls.2017.01852.

[5] Lu Y, Yi S, Zeng N, Liu Y, Zhang Y. Identification of rice diseases using deep convolutional neural networks. Neurocomputing, 2017; 267(6): 378–384.

[6] Bailey T L. MD-SVM: A novel SVM-based algorithm for the motif discovery of transcription factor binding sites. Bioinformatics, 2019; 28(1): 56–62.

[7] Deng X L, Li Z, Hong T S. Citrus disease recognition based on weighted scalable vocabulary tree. Precision agriculture, 2014; 15(3): 321–330.

[8] Kussul N, Lavreniuk M, Skakun S, Shelestov A. Deep learning classification of land cover and crop types using remote sensing data. IEEE Geoscience and Remote Sensing Letters, 2017; 14(5): 778–782.

[9] Arulkumaran K, Deisenroth M P, Brundage M, Bharath A A. A brief survey of deep reinforcement learning. IEEE Signal Processing Magazine, 2017; 34(6): 26–38.

[10] McCool C, Perez T, Upcroft B. Mixtures of lightweight deep convolutional

neural networks: Applied to agricultural robotics. IEEE Robotics and Automation Letters, 2017; 2(3): 1344–1351.

[11] Dechant C, Wiesner–Hanks T, Chen S, Stewart E L, Yosinski J, Gore M A, et al. Automated identification of northern leaf blight-infected maize plants from field imagery using deep learning. Phytopathology, 2017; 6(107): 1426–1432.

[12] Mohammed B, Kamel B, Abdelouahab M. Deep learning for tomato diseases: Classification and symptoms visualization. Applied Artificial Intelligence, 2017; 31(4): 299–315.

[13] Nachmani E, Marciano E, Lugosch L, Warren J G, David B. Deep learning methods for improved decoding of linear codes. IEEE Journal of Selected Topics in Signal Processing, 2018; 12(1): 119–131.

[14] Saleem M, Potgieter J, Arif K M. Plant disease detection and classification by deep learning. Plants, 2019; 8(11): 468–510.

[15] Geetharamani G, Arun P J. Identification of plant leaf diseases using a nine-layer deep convolutional neural network. Computers & Electrical Engineering, 2019; 76: 323–338.

[16] Pan S J, Yang Q. A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 2010; 22(10): 1345–1359.

[17] Abbas A, Abdelsamea M M, Gaber M. 4S-DT: Self supervised super sample decomposition for transfer learning with application to COVID-19 detection. IEEE Transactions on Neural Networks and Learning Systems, 2021; 32(7): 2798–2808.

[18] Liu Y, Jing L, Jian Y, Ng M K. Learning transferred weights from Co-occurrence data for heterogeneous transfer learning. IEEE Transactions on Neural Networks and Learning Systems, 2016; 27(11): 2187–2200.

[19] Zhuang F, Qi Z, Duan K, Xi B D, Zhu Y C, Zhu H S, et al. A comprehensive survey on transfer learning. Proceedings of the IEEE, 2020; 109(1): 43–76.

[20] Coulibaly S, Kamsu F B, Kamissoko D, Traore D. Deep neural networks with transfer learning in millet crop images. Computers in industry, 2019; 108(2019): 115–120.

[21] Long M S, Ouyang C J, Liu H, Fu Q. Image recognition of *Camellia oleifera* diseases based on convolutional neural network and transfer learning. Transactions of the CSAE, 2018; 34(18): 194–201. (in Chinese)

[22] Arnal B. Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. Computers and Electronics in Agriculture, 2018; 153: 46–53.

[23] Lumini A, Nanni L. Deep learning and transfer learning features for plankton classification. Ecological Informatics, 2019; 51: 33–43.

[24] Chen J, Chen J, Zhang D, Sun Y, Nanehkaran Y. Using deep transfer learning for image-based plant disease identification. Computers and Electronics in Agriculture, 2020; 173: 105393. doi: 10.1016/j.compag.2020.105393.

[25] Zhu R, Li S, Wang P, Xu M L, Yu S. Energy-efficient deep reinforced traffic grooming in elastic optical networks for cloud-fog computing. IEEE Internet of Things Journal, 2021; 8(15): 12410–12421.

[26] Lin F, Chen J, Ding G, Jiao Y, Sun J C, Wang H C. Spectrum prediction based on GAN and deep transfer learning: A cross-band data augmentation framework. China Communications, 2021; 8(1): 18–32.

[27] Abidin A Z, Deng B, Dsouza A M, Nagarajan M B, Coan P, Wismülleret A. Deep transfer learning for characterizing chondrocyte patterns in phase contrast X-Ray computed tomography images of the human patellar cartilage. Computers in Biology and Medicine, 2018; 95: 24–33.

[28] Zhang X, Wang Z, Liu D, Lin Q, Ling Q. Deep adversarial data augmentation for extremely low data regimes. IEEE Transactions on Circuits and Systems for Video Technology, 2020; 3(1): 15–28.

[29] Frid-Adar M, Diamant I, Klang E, Amitai M, Goldberger J, Greenspan J. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. Neurocomputing, 2018; 321(10): 321–331.

[30] Rangarajan A K, Purushothaman R, Ramesh A. Tomato crop disease classification using pre-trained deep learning algorithm. Procedia Computer Science, 2018; 133: 1040–1047.

[31] Chen Z, Le Z, Cao Z, Jing J. Distilling the knowledge from handcrafted features for human activity recognition. IEEE Transactions on Industrial Informatics, 2018; 14(10): 4334–4342.

[32] Liang Z, Tao M, Wang L, Su J, Yang X. Automatic modulation recognition based on adaptive attention mechanism and ResNeXt WSL model. IEEE Communications Letters, 2021; 25(9): 2953–2957.

[33] Hui J, Du M, Ye X, Qin Q, Sui J. Effective building extraction from high-resolution remote sensing images with multitask driven deep neural network. IEEE Geoscience and Remote Sensing Letters, 2019; 16(5): 786–790.