**外 文 翻 译**

**毕业设计题目：基于深度学习的植物病害检测模型设计与移动应用系统开发实现**

**原文1：Plant Disease Identification Using a**

**Novel Convolutional Neural Network**

**译文1：利用新型卷积神经网络的植物病害识别**

**原文2： Recent advances in plant disease severity**

**assessment using convolutional neural**

**networks**

**译文2：****卷积神经网络在植物病害严重程度评估中的**

**最新进展**

原文1

**Plant Disease Identification Using a Novel Convolutional Neural Network**

SK MAHMUDUL HASSAN AND ARNAB KUMAR MAJI , (Member, IEEE)

Department of Information Technology, School of Technology, North-Eastern Hill University (NEHU), Shillong 793022, India

Corresponding author: Arnab Kumar Maji (arnab.maji@gmail.com)

**Abstract**：The timely identification of plant diseases prevents the negative impact on crops. Convolutional neural network, particularly deep learning is used widely in machine vision and pattern recognition task. Researchers proposed different deep learning models in the identification of diseases in plants. However, the deep learning models require a large number of parameters, and hence the required training time is more and also difficult to implement on small devices. In this paper, we have proposed a novel deep learning model based on the inception layer and residual connection. Depthwise separable convolution is used to reduce the number of parameters. The proposed model has been trained and tested on three different plant diseases datasets. The performance accuracy obtained on plantvillage dataset is 99.39%, on the rice disease dataset is 99.66%, and on the cassava dataset is 76.59%. With fewer number of parameters, the proposed model achieves higher accuracy in comparison with the state-of-art deep learning models.

**Index Terms：**Plant disease, machine learning, deep learning, depthwise convolution, pointwise convolution.

**I. INTRODUCTION**

Diseases in crops caused mainly by bacteria and fungi negatively impact the production and quality of crops . Timely and early identification of disease symptoms is a major challenge in protecting crops. Visual identification of diseases in a large farm by experts and agronomists is the primary approach in developing countries which is time-consuming and costly. Automatic identification of diseases using smart devices is a promising approach to identifying and reducing overall costs . In recent times, deep learning particularly convolution neural network (CNN) gained much attention in agricultural field such as plant detection , fruit detection , disease identification –​, weeds detection , pest recognition etc. The reason behind the CNN-based model’s popularity is the automatic extraction of appropriate features from the data set. Several popular deep learning-based models such as AlexNet , GoogleNet , VGGNet , ResNet , DenseNet have been developed for identification of plant diseases.

Real-time applications and the identification of diseases using deep learning architectures are gaining attention in the current scenario. The number of parameters used and the computation cost in the deep learning models depend on the depth and the number of filters used in the model. The deep learning models usually require a large number of parameters. So the computational cost of the deep learning models is also high. Therefore there will be difficulties to implement on small devices having small resources . In recent times, researchers have implemented deep learning architectures using high-powered devices having GPUs and servers. Uses of sophisticated devices having GPU’s are not feasible in the agricultural field due to its high cost. Therefore there is a dearth need for applications having fewer parameters, less power consumption, and computation .

In prospect of the above consideration, we have designed a novel lightweight deep learning model to identify the diseases in the plant. This paper proposes a novel CNN architecture based on Inception and ResNet with fewer parameters for determining plant diseases. The Inception architecture extracts better features using multiple convolutions of different filter sizes. To tackle the vanishing gradient problem, we have used residual connections. Instead of standard convolution, we have used depth-wise separable convolution, which reduces the parameter size and the computational complexity without affecting performances. The model has trained on three different datasets, and the performances are evaluated. The main contribution of the paper is summarized as follows:

A new CNN architecture is proposed using Inception and Residual connection, which extracts better features and produces higher performance results.

In this paper, the standard convolution is replaced with depth-wise separable convolution, which reduces the parameter number by a large margin without affecting performances.

The proposed architecture uses fewer parameters than other deep learning architecture and is also faster than the standard deep learning models.

To check the robustness, the model’s performance is evaluated on three different plant disease data sets. We have considered three different conditioned images. In the PlantVillage dataset, the images were captured on uniform background and laboratory setup conditions. The images in the rice disease dataset were captured in real-time field conditions, and in the cassava plant dataset, the images are field conditioned, and images contain multiple leaves. We have compared the performance of the proposed model with other state-of-art deep learning models on three different datasets. The result shows that our proposed model outperforms other deep learning models.

The rest of the paper is organized as follows: Section 2 provides the existing literature on the identification of plant diseases using deep learning models. Materials and methods are discussed in Section 3. Section 4 presents the experimental results and discussions. Finally, the paper concludes with Section 5.

**II. Related Work**

This section presents a discussion and review of recent work on identifying plant diseases based on deep learning models. Mohanty et al. used AlexNet and GoogleNet to identify 26 diseases of 14 different plant species. To train the models they have used both transfer learning and learning from the scratch method. They achieved a maximum accuracy of 99.34% using GoogleNet. Five different pre-trained deep learning models, such as VGG, AlexNet, AlexNetOWTBn, Overfeat, and GoogleNet used by Ferentinos to identify 58 plant leaf diseases. Geetharamani and Pandian used nine-layer deep CNN to identify the diseases in plants and achieved an accuracy rate of 96.46%. Inspired from AlexNet and GoogLeNet architecture, Liu et al. designed a model by replacing the fully connected layer of AlexNet with the inception layer to identify four different apple leaf diseases and recorded an accuracy rate of 97.62%.

Four different pre-trained deep learning architecture namely VGG16, VGG19, ResNet, and InceptionV3 used by Ahmad et al. to identify different tomato leaf diseases. They fine-tuned the network parameter to get the optimal result. They achieved the highest performance accuracy using InceptionV3 is 99.60% and 93.70% on laboratory and field images respectively. Rangarajan and Purushothaman used a pre-trained VGG16 model to identify different eggplant diseases. VGG16 was used for the extraction of features. For classification, they used multiclass SVM. To check the robustness of the model performances, they have used three different color model images, namely RGB, YCbCr and HSV. Using RGB images, they recorded maximum accuracy of 99.4%.

Too et al. used several pre-trained deep learning models and fine-tuned the model parameter in the identification and classification of diseases in the plant. They have achieved maximum testing accuracy of 99.75% using DenseNet architecture. Deep feature-based and SVM classifier used by Sethy et al. to identify rice leaf diseases. They have used 11 deep CNN models to extract the features and SVM for classification purposes. Deep features from the ResNet50 model with SVM perform better with an f1-score of 98.38%. Six different pre-trained deep learning architecture used by Rangarajan Aravind and Raja to identify 10 different diseases of four varieties of plants. Among the architectures, VGG16 gives the highest performance accuracy of 90% on the test dataset.

Ramacharan et al. used transfer learning on InceptionV3 to identify three diseases and two pest damages on cassava plants. The dataset used consists of leaves with multiple numbers of leaves in a single image. The accuracy obtained in the dataset having single leaf images is higher than the images with multiple leaves. Later on, Ramacharan et al. used mobile devices to identify six different diseases in the cassava plant. They have used MobileNet deep learning architecture to train the network. They have used both image and video files of diseased leaf images to evaluate the performance. They achieved an accuracy of 80.6% and 70.4% on the image file and video files respectively. Oyewola et al. shown that deep learning with residual connection outperforms plain convolution neural network in the identification of different diseases in cassava plants.

Deep residual neural network with 50 layers along with batch normalization and ReLU activation after each convolution used by Picon et al. to identify 3 different wheat diseases. They recorded an accuracy rate of 96% on real-time field images captured in Germany. Durmus et al. used SqueezeNet architecture to identify different tomato leaf diseases. The structure of SqueezeNet is similar to AlexNet architecture, whose size is 227.6MB, while the size of SequenzeNet is 2.9MB. Modified Cifar10 quick CNN model used by Hu et al. and replaced the standard convolution with depthwise separable convolution to identify four different tea leaf diseases. They achieved an improved accuracy rate of 92.5%. Atila et al. used EfficientNet architecture to identify diseases in plants and shown that EfficientNet outperforms the standard CNN model such as AlexNet, VGG, etc., with an accuracy rate of 99.97% using EfficientNetB4 on PlantVillage dataset. The EfficientNet model takes less time to train the network as the network uses fewer parameters compared to other deep learning models.

Pre-trained VGG model with inception layer termed as INC-VGGN used by Chen et al. to identify the different corn and rice diseases. They replaced the fully connected layers of VGGNet and added two inception layers. The average testing accuracy obtained in rice diseases is 92% and 80.38% in corn diseases. Shallow CNN from the pre-trained VGG16 model used by Li et al. to identify different diseases in corn, apple, and grape. They have used only the first four-layer of VGG16 architecture a global pooling layer. Shallow CNN was used to extract the features and PCA to reduce the dimensions. They obtained an f1-score of 94% using SVM and RF respectively. Self-attention CNN was used by Zeng and Li to identify different crop diseases. Attention network is effective in extracting the image features from the critical region. Table 1 summarizes the related works along with the deep learning models and performance results.

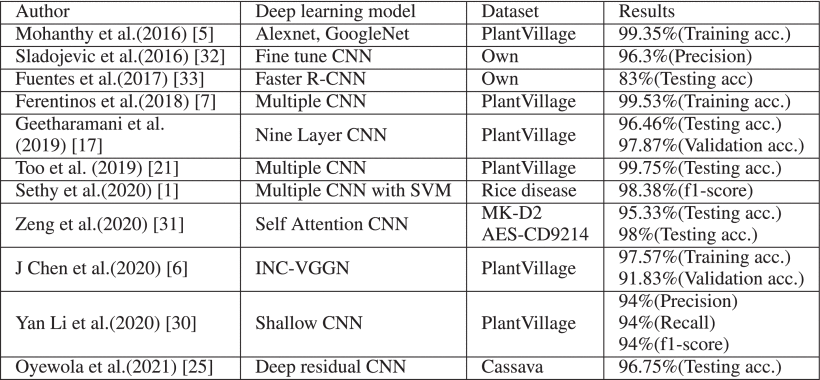


TABLE 1 Summary of Related Works on Identification of Plant Diseases

**III. Materials and Methods**

**A. Convolution Neural Network**

Convolutional Neural Network (CNN) is a type of Neural network that is very effective in several computer vision tasks such as pattern recognition and classification. CNN has an advantage that it can learn and extracts the features automatically from the training images whereas in traditional approach there is a need of manual extraction of feature from the images. CNN consists of different layers: Convolutional layer, pooling layer, and fully connected layer. The convolutional layer is the primary and most crucial component of CNN which extracts the feature from the input image. Convolution layers consist of a small array of numbers called kernels applied over the input which produces an output called feature maps. Different convolutional kernels are used to extract different types of features. Several convolutional layers are there, which depend on the size of the input images.

After the convolutional layer, pooling is performed which is responsible for reducing the dimension of the convolutional feature map. The pooling layer performs down sampling operations by reducing the dimension of the feature map, which ultimately helps in reducing the required computational complexity to process the data. Different types of pooling operations are there, such as max-pooling, min-pooling, average-pooling. The output feature maps of the convolution or pooling layer are transformed into a one-dimensional vector in which every input is connected to every output by weight. This layer is also called a dense layer. There can be one or more fully connected layers, and the final fully connected layer has the same number output as the number of classes.

**B. Residual Network**

The convolutional neural network may achieve high performances in classification tasks. With the rise in network depth, the performance accuracy gets saturated and degrades rapidly. To address this issue in 2015, He et al. introduced a deep residual learning network known as ResNet. In deep learning, using the residual connection in the network, we can train a large network and parallely solve the vanishing gradient problem, which usually occurs due to the increase in network depth. Figure 1 shows the basic block diagram of the ResNet model. ResNet introduced a skip connection known as “identity mapping” which combines previous layer’s output with the forthcoming layer. To perform identity mapping, it doesn’t generate any parameter. Therefore ResNet has the ability to train a deeper network with lower complexity than other networks such as VGG.

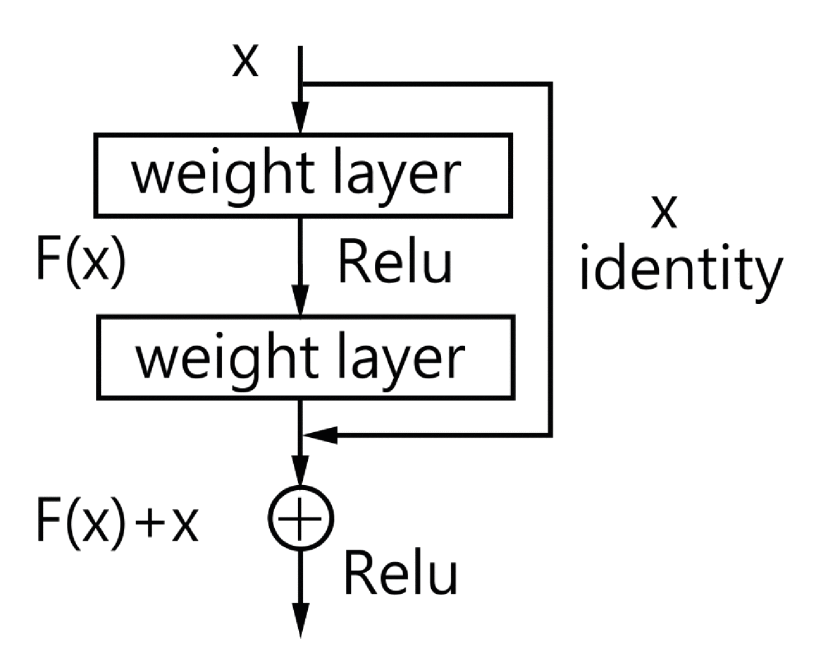
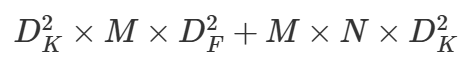


FIGURE 1.Basic block diagram of residual network.

**C. Depthwise Separable Convolution**

Depthwise separable convolution introduced by Chollet in the Xception model. Later on, Howard et al. used depthwise separable convolution in MobileNet architecture. Depthwise separable convolution factorizes standard convolution into depthwise convolution and 1×1 pointwise convolution. In standard convolution, it takes one step to filter the input images and combine these values. Figure 2 shows the block diagram of depthwise separable convolution. The Depthwise layer performs filtering operation, and the pointwise layer combines the output of the depthwise layer. The computation cost of depthwise separable convolution is computed as



whereas the computation cost in standard convolution is –



where DF is the dimension of the input image of square size, DK is the dimension of the kernel, M is the number of channels and N is the number of kernel/filters.

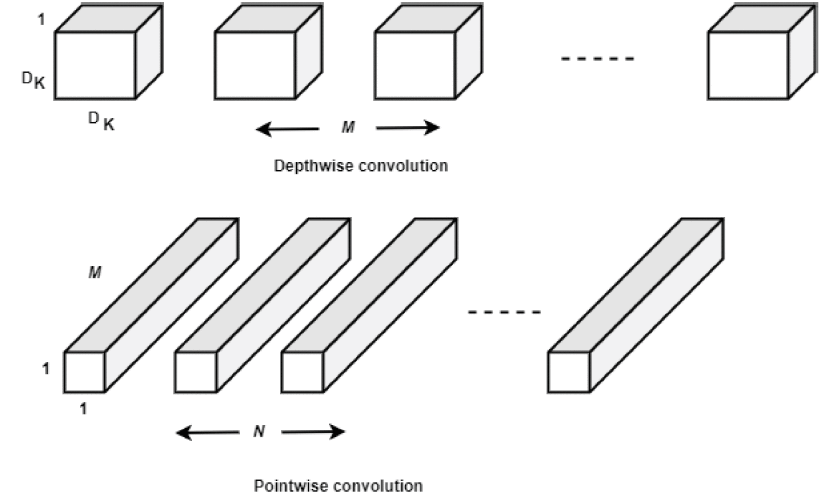


FIGURE 2.Depthwise separable convolution.

**D. Proposed Novel CNN Approach for Identification of Plant Diseases**

In this paper, we have proposed a novel light weight CNN based on Inception and Residual connections with fewer parameters than the InceptionV3, ResNet50 as well as other deep learning models. The Inception in deep learning architecture was introduced by Szegedy et al. in 2015 and named as GoogleNet (Inception-v1). Later on, additional factorization is added in the inception network and referred it as Inception-v3. Inception architecture is different from typical convolutional architecture and performs multiple convolutions and pooling operations simultaneously to extract the better feature. It concatenates the outputs of all the convolutions of different filter sizes. Inception v3 consist of numbers of Inception A, reduction A, Inception B and Reduction B blocks as shown in Figure 3(a) and Figure 4(a). The inception A block consist of a convolution layer with kernel size 1×1 , 1×1 convolution followed by 3×3 convolutions, 1×1 convolution followed by 5×5 convolutions, 3×3 max-pooling layer followed by 1×1 convolution. Inception B block consists of convolution layer with kernel size 1×1 , 1×1 convolution followed by 7×7 convolutions, 1×1 convolution followed by two 7×7 convolutions, 3×3 max-pooling layer followed by 1×1 convolution. 1×1 convolutions were used to reduce the computation of the model. The reduction A block uses a 3×3 max-pooling layer, 3×3 convolution layer, and 1×1 convolution layer followed by a 3×3 convolution layer. The reduction B block of inceptionv3 uses a 3×3 max-pooling layer, 3×3 convolution layer followed by 1×1 convolution layer, and 7×7 convolution layer followed by 3×3 convolution layer. Figure 3(a) and Figure 4(a) represent the block diagram of inception-A and inception-B block of inception-V3 architecture.

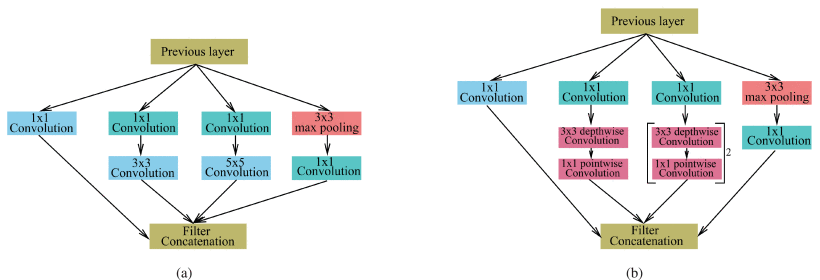


FIGURE 3.Inception architecture of (a) Original inception-A block (b) Modified inception-A block.

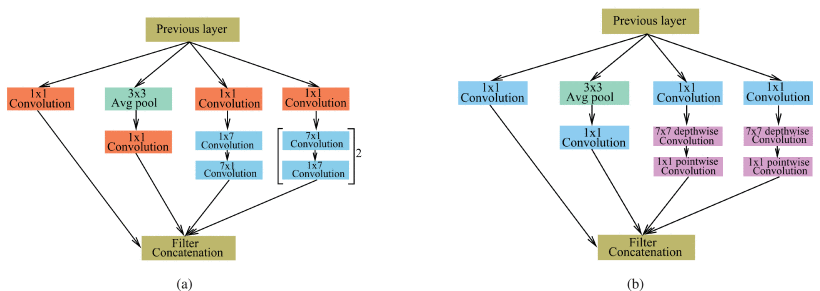


FIGURE 4.Inception architecture of (a) Original inception-B block (b) Modified inception-B block.

In our work, we have proposed a novel CNN model which uses the inception architecture with normal convolution as well as depthwise separable convolutions. In the proposed model, 3×3 convolution of Inception A block is replaced by 3×3 depthwise separable convolution (3×3 depthwise convolution and 1×1 pointwise convolution), and 5×5 convol- ution is replaced with two 3×3 depthwise separable convolutions as shown in Figure 3(b). The 7×7 convolution used in Inception B is replaced with 7×7 depthwise separable convolution (7×7 depthwise convolution and 1×1 pointwise convolution) as shown in Figure 4(b). The 3×3 convolution layer of reduction A replaced with a 1×1 convolution layer followed by 3×3 depthwise separable convolution. In reduction B block, we have replaced the 3×3 and 7×7 convolution with a 1×1 convolution layer and a 3×3 depthwise separable convolution layer. Table 2 and Table 3 shows the parameter comparison between the original inception-A block and the modified inception-A block with depthwise separable convolution. From Table 2 and Table 3 it is seen that the parameter used in the modified inception-A block is much less as compared to the original inception-A blocks.

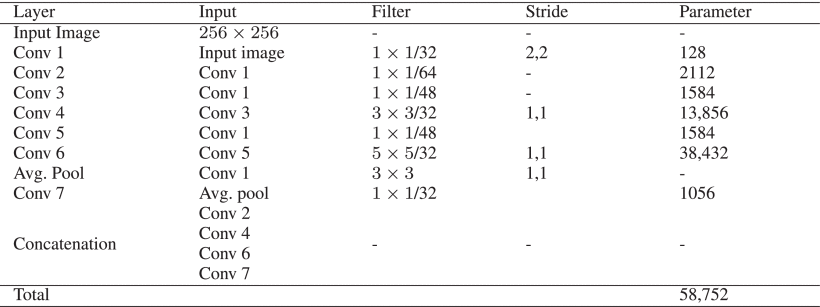


TABLE 2 Parameter Required in Original Inception-A Block

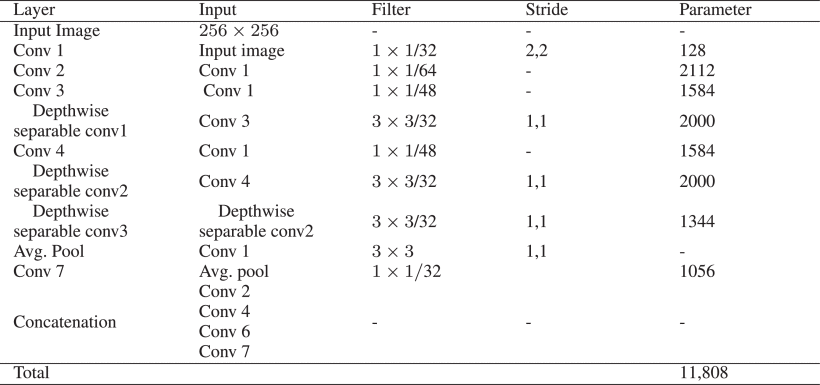


TABLE 3 Parameters Required in Modified Inception-A Block With Depth-Wise Separable Convolution

Figure 5 shows the proposed CNN architecture used to identify the diseases in plants. The proposed implemented model consists of a convolution layer, batch normalization and activation layer, inception blocks with depthwise separable convolution layer, pooling layer, and fully connected layer. In this architecture, we have replaced the standard convolution with depthwise separable convolution. We have used one standard convolution, three depthwise separable convolutions, two max-pooling, and one global average pooling operation, three modified inception A block with residual connection followed by modified reduction A block, three modified inception B block with residual connection followed by modified reduction B block. After each convolution layer, we have used Batch Normalization and activation. The activation function used is ReLu. Batch Normalization and activation function improves the performance and speeds up the process. After global average pooling, we have used dropout, which reduces the chances of overfitting the model. The required parameter in our proposed model is 428,100 while the number of parameter used in standard inceptionv V3 model is 23,851,784 which is much higher than the proposed model. It is observed that the proposed model uses 70% less parameter as compared to the inception V3 architecture.

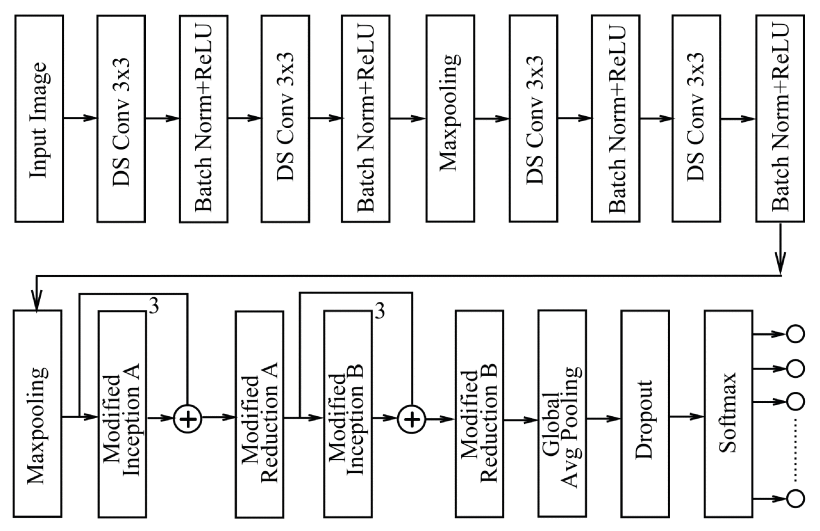


FIGURE 5.Proposed CNN approach in identification of plant diseases.

**IV. Results and Discussion**

**A. Dataset**

In this paper, we consider three different plant disease datasets to evaluate the performances of the model. We have used the Rice plant dataset , cassava plant dataset , and Plant village dataset which is the large and available plant disease dataset. From the Plant village dataset specifically, we have used the corn, potato, and tomato plant diseases. It is seen that the images in the PlantVillage dataset are captured in a uniform background, and the intensity is also relatively uniform. The rice plant dataset consists of 5932 images which are divided into four categories, namely 1584 bacterial blight, 1440 blast, 1600 brown spot, and 1308 tungro images. The cassava plant dataset consists of 5,656 images includes five different categories as 316 healthy cassava leaf images and four sets of diseased leaves as 466 cassava bacteria blight, 1,443 cassava brown streak, 773 cassava green mite, and 2,658 cassava mosaic diseased images. The images in the cassava dataset are captured in the complex background, and multiple leaves are also present in a single image. The datasets are divided randomly into training and testing set in the ratio of 80:20. The dimensions of the images are resized into 256×256 pixels. Figure 6 shows the sample images of different datasets. Table 4, Table 5, Table 6 shows the dataset details along with the disease common name, disease scientific name, Type of disease along with the number of images in each class.

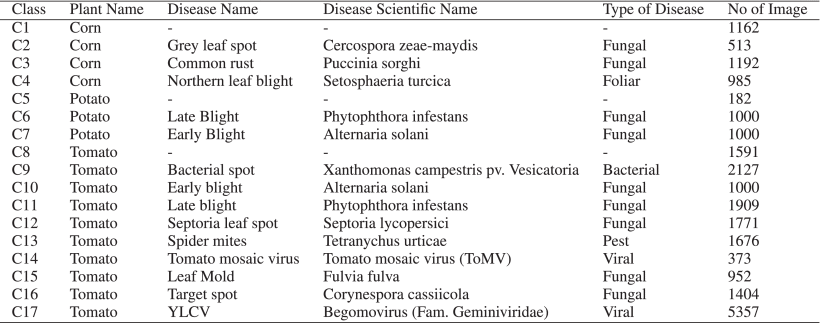


TABLE 4 Data Description of Plantvillage Dataset (Corn, Potato, Tomato)

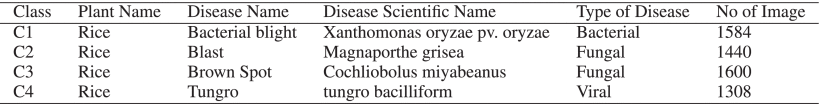


TABLE 5 Data Description of Rice Disease Data

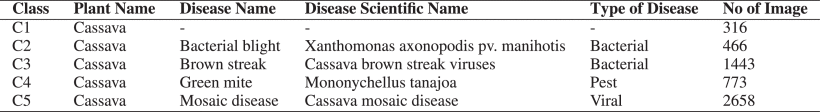


TABLE 6 Data Description of Cassava Dataset

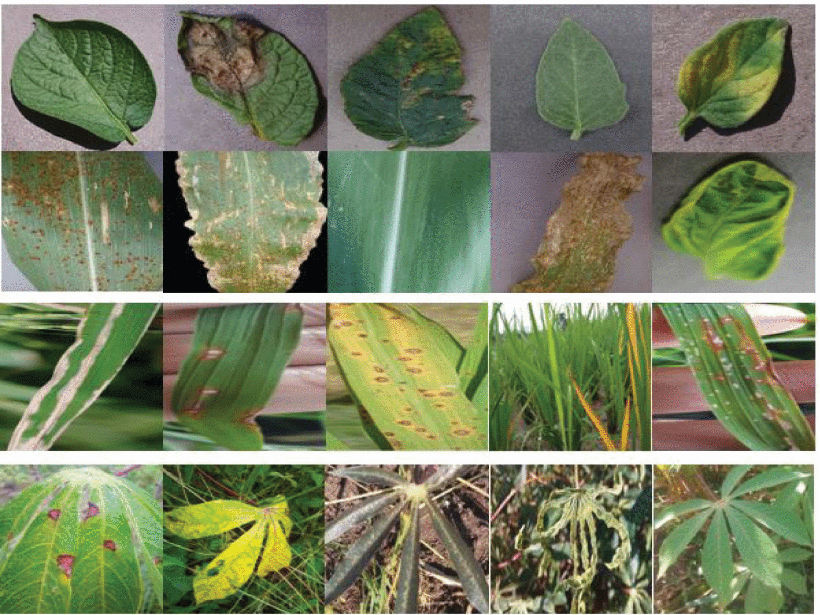
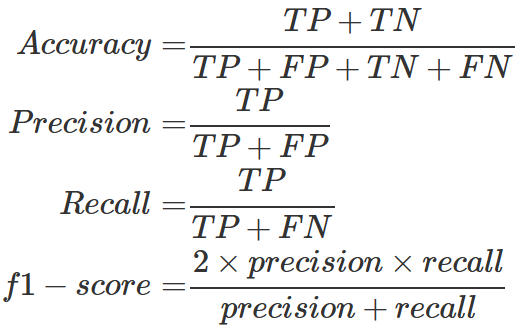


FIGURE 6.Sample images of Plantvillage dataset, rice and cassava plant dataset.

**B. Experimental Result**

To evaluate the performance of the model, we consider different performance statistics such as number of parameter, accuracy, precision, recall, f1-score and expressed as follows-



where TP = true positive, TN = true negative, FP = false positive and FN = false negative.

Table 7 shows the performances of the implemented model on different datasets. From Table 7, it is seen that after 50 epochs of training, the model achieved the highest training accuracy and loss of 99.81% and 0.0015, validation accuracy and loss of 99.39% and 0.0549 respectively in PlantVillage dataset (potato, corn, tomato). In the rice dataset, the model achieves the highest training accuracy of 99.94%, training loss of 0.0030, validation accuracy of 99.66%, and validation loss of 0.0041. To evaluate the robustness of our proposed model, we have considered the Cassava plant disease dataset in which the images were captured in real-time and in field conditions. Moreover, the images have complex backgrounds and having multiple leaves present on a single image. Compared with the PlantVillage and Rice plant datasets, the Cassava plant dataset gives less training and validation accuracy. The obtained training and validation accuracies are 98.17% and 76.59%, respectively. The cassava dataset’s performance accuracy drops compared to the plantvillage and rice dataset due to data imbalance, and the images in the dataset contain images with complex backgrounds.

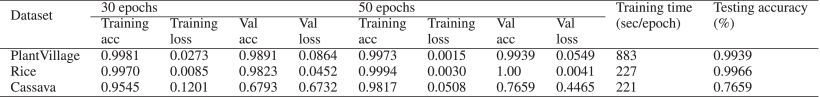


TABLE 7 Summary of Deep Learning Based Implemented Methods

After splitting the dataset into 80%-20% training and validation set we train the model upto 50 epoch and evaluate the performances on training and validation set. Figure 7–​9 shows the accuracy and loss for training and validation of the proposed model on PlantVillage, rice and cassava dataset respectively. From Figure 7–​9 it is seen that the proposed model gives satisfactory performances with fewer number of parameters.

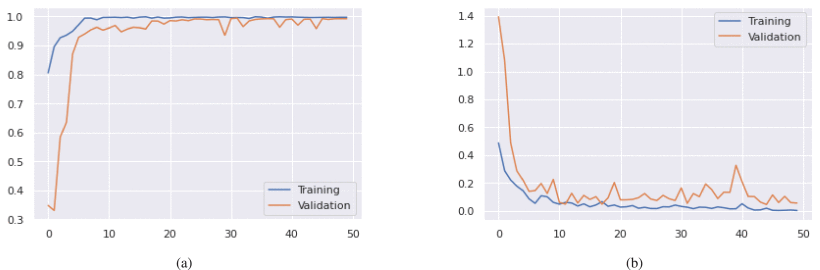


FIGURE 7.(a) Training & validation accuracy (b) Training & validation loss on PlantVillage dataset.

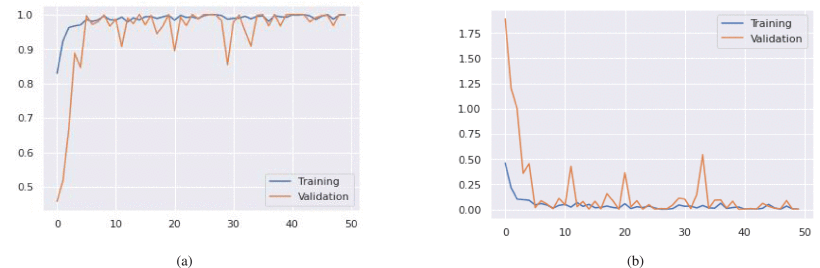


FIGURE 8.(a) Training & validation accuracy (b) Training & validation loss on rice disease dataset.

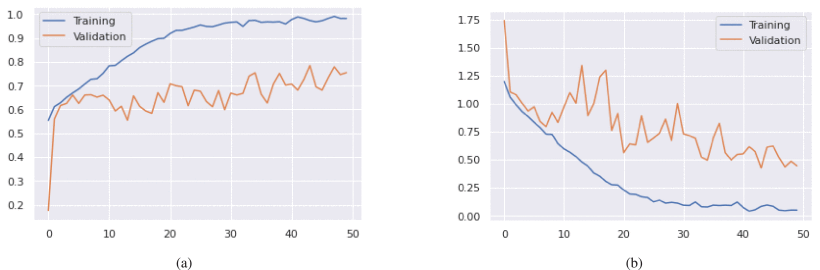


FIGURE 9.(a) Training & validation accuracy (b) Training & validation loss on cassava plant dataset.

We have also calculated the confusion matrix to evaluate the performance of the model. Table 8 shows the performances of the implemented model in terms of testing accuracy, precision, recall, and f1-score on three plant datasets. From Table 8, it is seen that the Rice plant dataset gives the highest testing accuracy than PlantVillage and Cassava dataset. Figure 10 gives the performance metrics of the implemented model on different datasets. From Figure 10, it is seen that the average prediction accuracy, precision, recall, and F1-score is more than 99% on the rice and plantvillage dataset.

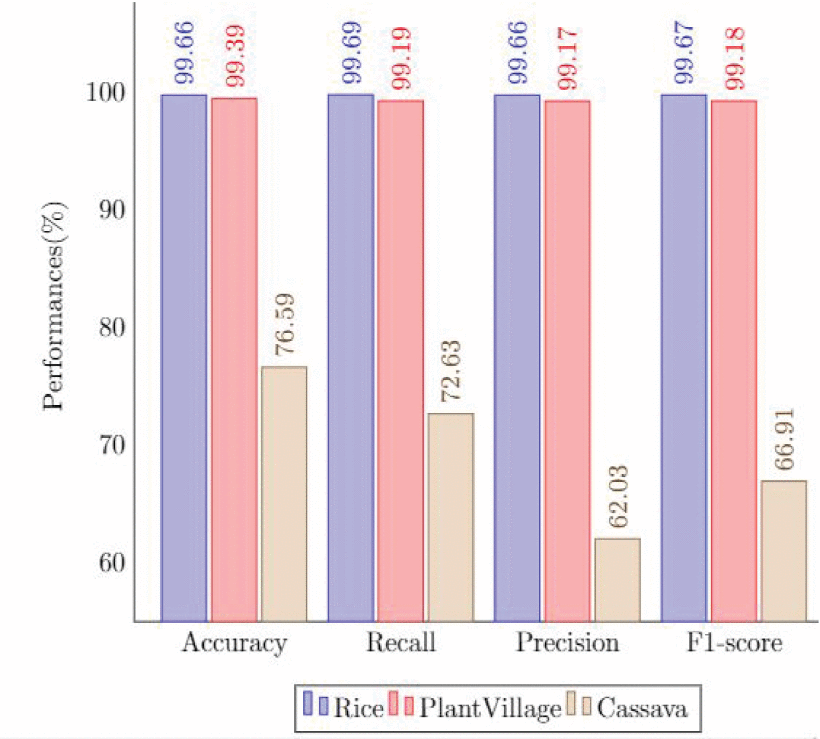


FIGURE 10.Performance metric of proposed model on different dataset.

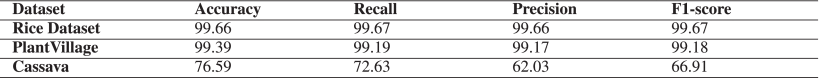


TABLE 8 Performance Metric of the Proposed Model on Testing Images

**C. Comparison of Performances And Robustness of the Model**

To validate the stability of our proposed CNN model, we use k-fold cross-validation (5 fold) to evaluate the performances on the disease datasets. The disease datasets are divided randomly into five equal parts. Each part of the dataset is considered as the test set, and the remaining part of the dataset is considered as the training set. Based on the splitting of the dataset, we get the performance of the model on a different set of training and testing images. Table 9 shows the performances of the proposed model on each fold. From Table 9, it is seen that there is not much performance deviation in each dataset. In the PlantVillage dataset, the accuracy ranges from 99.29% to 99.37%, in the Rice dataset, it ranges from 99.33% to 99.66% and in the Cassava dataset, the accuracy ranges from 76.42% to 76.58%. The result shows that our proposed CNN method has good stability in the identification of diseases and is consistent in each data split part.

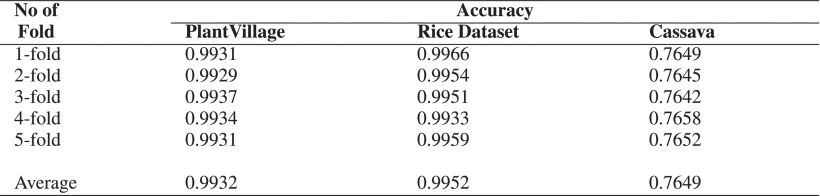


TABLE 9 Result of Proposed CNN Based on k-Fold Cross Validation

**D. Performance Comparison With Pre-Trained Network**

We have evaluated the performances of several pre-trained deep learning models such as VGG16, VGG19, InceptionV3, ResNet50, and DenseNet201 on three different datasets. The performance is compared with our proposed CNN model in terms of performance accuracy and training time. The performance comparison with other pre-trained deep learning models on three different datasets is shown in Table 10. The accuracy defines the correctly identified classes of the images from the test image set. From the performance results, it is seen that our proposed model gives satisfactory performances on three different datasets (99.39% on PlantVillage, 99.66% on Rice, 76.59% on imbalance cassava) with fewer parameters (428,100). Table 10 shows the time required to train the models per epoch on three different datasets. From Table 10, it is seen that the required training time is much less in comparison with the other pre-trained networks in all three datasets as the number of layer used in our proposed model is less. The performance accuracies of the pre-trained deep learning models are less as the models uses the pre-trained weight where the models are trained on Imagenet Dataset. Due to the uses of pre-trained weight, these models did not achieve the optimal results. Whereas, the proposed model has an advantages of inception layer which has ability to extract better features as well as residual connection that removes vanishing gradient problem. Moreover, the uses of batch normalization and dropout increases the performance of model. Although the proposed model uses depthwise separable convolution which reduces the model parameter, also the number of layers used in this model is less. So the required time to train the model is much less as compared to pre-trained models. DenseNet201 requires more training time than the other deep learning models as the number of layers are more.

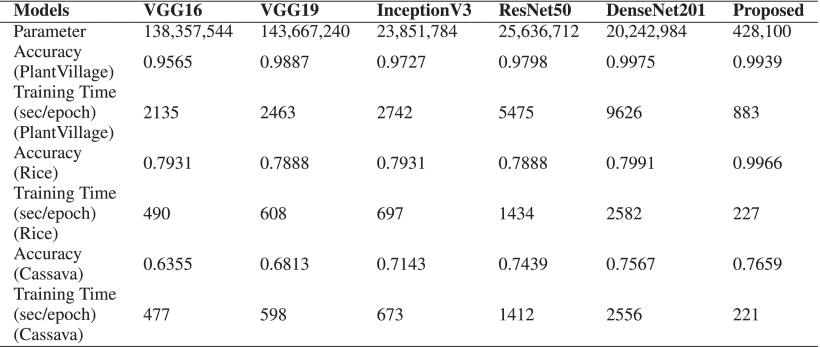


TABLE 10 Performance Comparison With Pre-Trained Network

**E. Comparison With some Existing Literature**

To verify the performances of our proposed CNN model, we compare the performances of our proposed model with other deep learning models from the literature. Table 11 shows the performance comparison with some deep learning models. From Table 11, it is seen that the proposed CNN model performs better. From Table 11, we can also see that the proposed model requires much fewer parameter than other deep learning models. As the model uses fewer parameters, it requires less time to train the model compared to other deep learning models. From Table 11 it is also seen that the models were implemented on PlantVillage dataset where the images are captured in controlled environment. In this work, we have used three different dataset where rice disease images were captured in field condition and contains background image. The cassava dataset consist of field images along with more than one leaf in single image. Our proposed CNN model has advantages in terms of performance accuracy, parameter size, and training time.

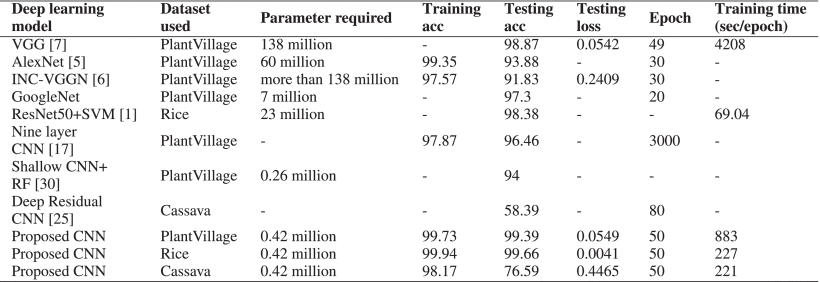


TABLE 11 Performance Comparison With Different Deep Learning Models

**V. Conclusion**

Deep learning is a recent and advanced technique in the field of image pattern recognition and much effective in the identification of diseases in plants. In this paper, we have proposed a novel CNN model based on the inception and residual connection that can effectively classify the diseases in plants. In addition, we have used depthwise separable convolution in inception architecture which reduces the computation cost by reducing the number of parameters by a margin of 70%. Therefore, training the network requires much less time as compared to the standard CNN. The experimental result shows that the proposed model achieves higher performance accuracy. To evaluate the robustness of the model, we have used three different plant datasets. The testing accuracies of the proposed model is 99.39%,99.66% and 76.59% on Plantvillage, Rice, and Cassava dataset, respectively. In author recorded an accuracy rate of 52.87% and 46.26% using plain convolution neural network and deep residual neural network respectively on imbalance dataset. In athor achieved an accuracy rate of 80.6% on balance dataset. In comparison with, our proposed model achieves a higher accuracy rate of 76.59% on the imbalance cassava dataset. For the rice dataset Sethy et al. achieved an accuracy rate of 98.25%, and our proposed model achieved higher accuracy of 99.66% on the same dataset. In the future, we carry forward and try to investigate the performance of the proposed model on the different agricultural fields such as weed detection, pest identification, etc. Another future work includes the identification of plant diseases with different plant disease datasets with a different variety of images, different geographical regions. Using clustering based unsupervised technique in identification of diseases also an important aspects.

译文1

作者：Sk Mahmudul Hassan,Arnab Kumar Maji

出处：IEEE Access, 2022, 10: 5390-5401.

**利用新型卷积神经网络的植物病害识别**

**摘要**：及时识别植物病害对于防止作物受损至关重要。卷积神经网络（CNN），尤其是深度学习方法，广泛应用于机器视觉和模式识别任务中。研究人员提出了多种深度学习模型用于植物病害识别。然而，这些模型通常需要大量参数，导致训练时间较长，并且难以在小型设备上实现。在本文中，我们提出了一种基于Inception层和残差连接的新型深度学习模型。通过引入深度可分离卷积以减少参数数量。该模型在三个不同的植物病害数据集上进行了训练和测试。在PlantVillage数据集上，模型达到了99.39%的准确率；在水稻病害数据集上达到了99.66%；在木薯病害数据集上达到了76.59%。与现有的深度学习模型相比，该模型在减少参数数量的同时实现了更高的准确率。

**关键词：**植物病害，机器学习，深度学习，深度卷积，逐点卷积

1. **引言**

作物病害主要由细菌和真菌引起，对作物的产量和质量产生负面影响。及时、早期识别病害症状是保护作物的重大挑战。在发展中国家，农学家和专家通过目视方式识别大面积农场的病害是主要方法，但此方法耗时且成本高昂。利用智能设备进行自动病害识别是一种有前景的方法，可以降低整体成本并提高识别效率。

近年来，深度学习，尤其是卷积神经网络（CNN）在农业领域备受关注，包括植物检测、水果检测、病害识别、杂草识别和害虫分类等应用。CNN模型之所以受到青睐，是因为其能够自动从数据集中提取适当的特征。多种流行的深度学习模型如AlexNet、GoogleNet、VGGNet、ResNet和DenseNet已被开发用于植物病害识别。

在当前背景下，基于深度学习架构的实时应用及病害识别正逐渐受到关注。深度学习模型的参数数量和计算成本取决于模型的深度及使用的滤波器数量。由于深度学习模型通常需要大量参数，因此计算成本较高，这给在小型设备上部署带来了困难。近年来，研究人员在高性能GPU和服务器上实现了深度学习架构。然而，在农业领域，使用GPU等昂贵设备并不可行。因此，迫切需要开发具有更少参数、更低功耗和计算量的应用程序。

针对上述问题，我们设计了一种新型轻量级深度学习模型用于植物病害识别。本文提出了一种基于Inception和ResNet的CNN架构，该模型参数较少但能够有效识别植物病害。Inception架构通过不同滤波器尺寸的多次卷积提取更优特征。为了应对梯度消失问题，我们引入了残差连接。此外，我们用深度可分离卷积替代了标准卷积，从而在不影响性能的前提下降低了参数数量和计算复杂度。模型在三个不同数据集上进行训练和测试，验证了其性能。主要贡献：

提出一种结合Inception和残差连接的新型CNN架构，能够提取更优特征并提供更高的性能。

本文用深度可分离卷积取代标准卷积，显著减少参数数量，同时保持模型性能。

该模型的参数数量少于其他深度学习架构，执行速度更快。

为验证模型的健壮性，我们在三个不同的植物病害数据集上评估了其性能。这些数据集包括在实验室和田间条件下采集的图像。结果表明，我们提出的模型优于现有的深度学习模型。

本文其余部分组织如下：第2节介绍现有的植物病害识别相关文献。第3节讨论材料与方法。第4节展示实验结果与讨论。最后，第5节为结论部分。

**II．相关研究**

本节对基于深度学习模型的植物病害识别的近期研究进行了讨论和综述。Mohanty等人使用AlexNet和GoogleNet识别了14种不同植物物种的26种疾病。在训练模型时，他们采用了迁移学习和从零开始学习两种方法。通过GoogleNet，他们获得了99.34%的最高准确率。Ferentinos使用五种不同的预训练深度学习模型（VGG、AlexNet、AlexNetOWTBn、Overfeat和GoogleNet）识别了58种植物叶片病害。Geetharamani和Pandian使用了九层深度卷积神经网络（CNN）对植物病害进行识别，达到了96.46%的准确率。受AlexNet和GoogleNet架构启发，Liu等人通过将AlexNet的全连接层替换为Inception层设计了一个模型，识别了四种不同的苹果叶片病害，记录的准确率为97.62%。

Ahmad等人使用了四种不同的预训练深度学习架构（VGG16、VGG19、ResNet和InceptionV3）识别不同番茄叶片病害。他们通过微调网络参数获得了最优结果。其中InceptionV3在实验室和田间图像上的最高准确率分别达到了99.60%和93.70%。Rangarajan和Purushothaman使用预训练的VGG16模型识别不同的茄子病害。在特征提取过程中使用VGG16，而在分类过程中采用了多类支持向量机（SVM）。为验证模型的健壮性，他们使用了三种不同的颜色模型图像，即RGB、YCbCr和HSV。在RGB图像中，他们记录的最高准确率为99.4%。

Too等人使用了多种预训练深度学习模型，并在植物病害识别和分类中对模型参数进行了微调。通过DenseNet架构，他们获得了99.75%的最高测试准确率。Sethy等人使用基于深度特征的SVM分类器识别了水稻叶片病害。他们使用了11个深度CNN模型提取特征，并使用SVM进行分类。ResNet50模型的深度特征与SVM结合后表现最佳，F1分数为98.38%。Rangarajan Aravind和Raja使用了六种不同的预训练深度学习架构，识别了四种植物的10种不同病害。在所有架构中，VGG16在测试数据集上的最高准确率达到了90%。

Ramacharan等人通过InceptionV3迁移学习识别了木薯植物的三种疾病和两种害虫损害。所使用的数据集包含了具有多片叶子的单个图像。在包含单片叶子图像的数据集中，获得的准确率高于包含多片叶子图像的数据集。随后，Ramacharan等人使用移动设备识别木薯植物的六种不同病害。他们采用MobileNet深度学习架构进行网络训练，并使用病害叶片图像和视频文件评估模型性能。其在图像和视频文件上的准确率分别为80.6%和70.4%。Oyewola等人表明，在识别木薯植物不同病害方面，带有残差连接的深度学习优于普通的卷积神经网络（CNN）。

Picon等人使用带有50层的深度残差神经网络（ResNet），在每个卷积层后加入批量归一化和ReLU激活函数，识别了三种不同的小麦病害。在德国采集的实时田间图像中，他们记录的准确率为96%。Durmus等人使用SqueezeNet架构识别不同的番茄叶片病害。SqueezeNet的结构类似于AlexNet，但其大小为2.9MB，而AlexNet的大小为227.6MB。Hu等人通过修改Cifar10快速CNN模型，将标准卷积替换为深度可分卷积，识别了四种不同的茶叶病害，准确率提升至92.5%。Atila等人使用EfficientNet架构识别植物病害，结果表明EfficientNet的性能优于AlexNet、VGG等标准CNN模型。在PlantVillage数据集上，EfficientNetB4的准确率达到了99.97%。EfficientNet由于参数量较少，训练时间比其他深度学习模型更短。

Chen等人使用带有Inception层的预训练VGG模型（称为INC-VGGN）识别了不同的玉米和水稻病害。他们替换了VGGNet的全连接层，并添加了两个Inception层。在水稻病害上的平均测试准确率为92%，玉米病害上的准确率为80.38%。Li等人使用预训练VGG16模型的浅层CNN识别玉米、苹果和葡萄的不同病害。他们仅使用VGG16架构的前四层和一个全局池化层。浅层CNN用于特征提取，并通过主成分分析（PCA）减少维度。SVM和随机森林（RF）分别获得了94%的F1分数。Zeng和Li使用自注意力CNN识别不同的农作物病害。注意力网络在提取关键区域图像特征方面表现出色。

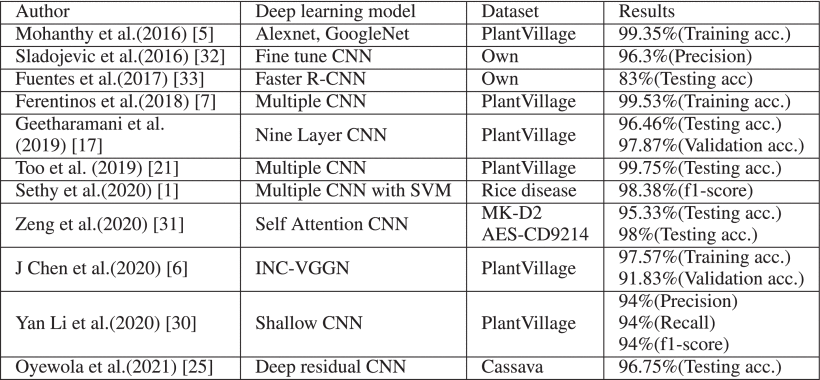


表1植物病害鉴定相关研究总结

**III．资料与方法**

A.卷积神经网络

卷积神经网络是一类在计算机视觉任务（如模式识别和分类）中非常有效的神经网络。卷积神经网络的优势在于能够自动从训练图像中学习和提取特征，而传统方法需要手动提取特征。卷积神经网络包含不同的层次结构：卷积层、池化层和全连接层。卷积层是卷积神经网络的核心组成部分，从输入图像中提取特征。卷积层由一组称为核的数字数组组成，应用于输入数据后生成称为特征图的输出。不同的卷积核用于提取不同类型的特征。卷积层的数量取决于输入图像的大小。

卷积层后进行池化操作，负责减少卷积特征图的维度。池化层通过下采样减少特征图的尺寸，从而降低处理数据所需的计算复杂度。池化操作有多种类型，包括最大池化、最小池化和平均池化。卷积层或池化层的输出特征图最终转换为一维向量，每个输入与输出之间由权重相连。这个层称为密集层，可以包含一个或多个全连接层。最终的全连接层的输出数量等于类别的数量。

B.残差网络(ResNet)

卷积神经网络在分类任务中可能达到较高的性能。然而，随着网络深度的增加，模型的准确率趋于饱和，并迅速下降。为了解决这个问题，He等人在2015年提出了一种深度残差学习网络ResNet。

在深度学习中，通过在网络中使用残差连接，可以训练更深层次的网络，并解决由于网络深度增加而导致的梯度消失问题。ResNet引入了一种称为“恒等映射”的跳跃连接，将前一层的输出与后续层的输出相结合。恒等映射无需生成额外参数，因此ResNet能够训练比VGG等其他网络更深且复杂度更低的模型。

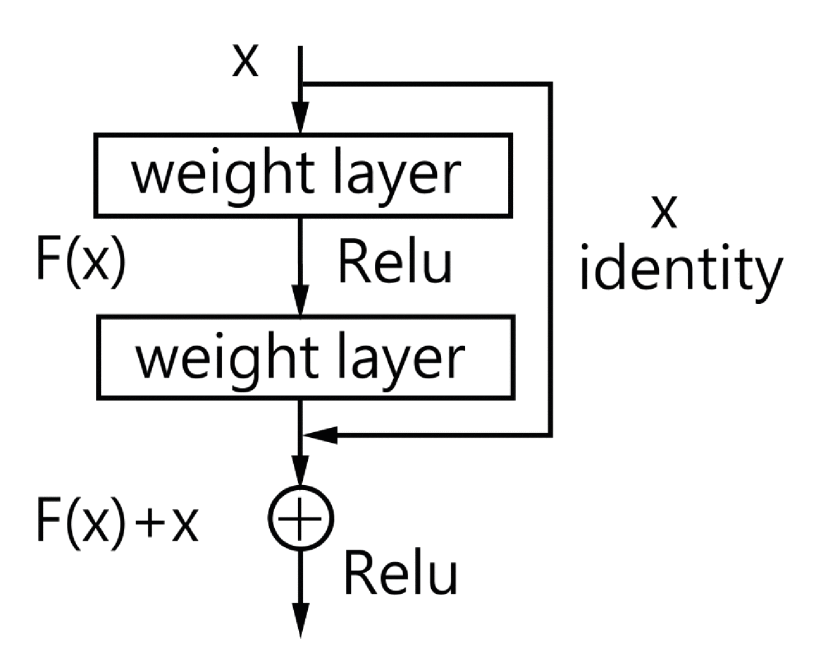
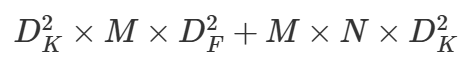


图1.残差网络的基本框图

C.深度可分离卷积

Chollet在Xception模型中引入了深度可分离卷积。后来，Howard等人在MobileNet架构中使用了深度可分离卷积。深度可分离卷积将标准卷积分解为深度卷积和1×1逐点卷积。在标准卷积中，过滤输入图像并组合这些值需要一步。图2显示了深度可分离卷积的框图。深度层执行过滤操作，逐点层组合深度层的输出。深度可分离卷积的计算成本计算为：



而标准卷积的计算成本是：



其中DF是方形输入图像的维度，DK是内核的维度，M是通道数，N是内核/过滤器的数量。

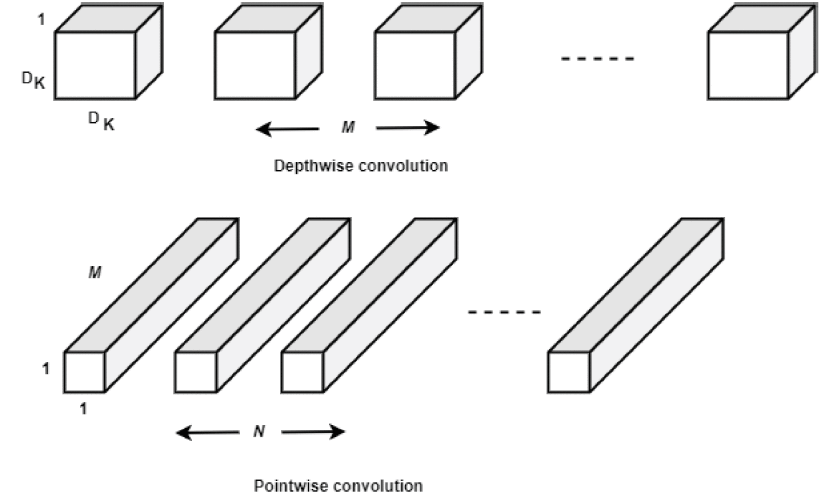


图 2.深度可分离卷积

D.植物病害识别的新型卷积神经网络方法

本文提出了一种基于Inception和残差连接的新型轻量级卷积神经网络，其参数数量远少于InceptionV3、ResNet50及其他深度学习模型。Inception架构由Szegedy等人于2015年提出，被命名为GoogleNet（Inception-v1）。随后，该网络通过增加因式分解进一步优化，被称为Inception-v3。Inception架构与典型的卷积神经网络不同，它通过同时执行多个卷积和池化操作以提取更优的特征，并将不同卷积核尺寸的输出进行拼接。Inception v3架构由多个Inception A块、Reduction A块、Inception B块和Reduction B块组成，其结构如图3(a)和图4(a)所示。Inception A块 包含一个1×1卷积层，后接一个1×1卷积层和一个3×3卷积层，或者一个1×1卷积层和一个5×5卷积层。此外，还包含一个3×3最大池化层，后接一个1×1卷积层。Inception B块包含一个1×1卷积层，后接一个1×1卷积层和一个7×7卷积层，或者一个1×1卷积层和两个连续的7×7卷积层。此外，还包含一个3×3最大池化层，后接一个1×1卷积层。Reduction A块包括一个3×3最大池化层、一个3×3卷积层和一个1×1卷积层，后接一个3×3卷积层。Reduction B块包括一个3×3最大池化层、一个3×3卷积层，后接一个1×1卷积层和一个7×7卷积层，后接一个3×3卷积层。图3(a)和图4(a)展示了Inception-v3架构中Inception A块和B块的结构图。

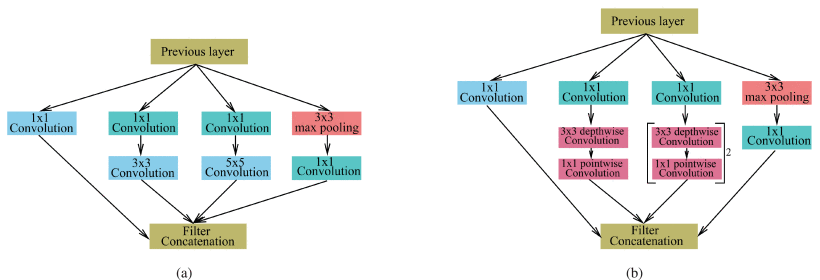


图 3.Inception架构(a)原始inception-A块(b)修改后的inception-A块

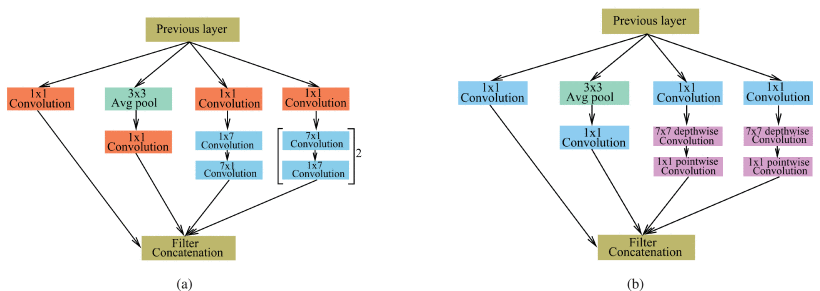


图 4.Inception架构(a)原始inception-B块(b)修改后的inception-B块

在本文工作中，提出了一种新型卷积神经网络模型，该模型结合了Inception架构的标准卷积和深度可分离卷积（Depthwise Separable Convolution）。Inception A块中的3×3卷积层被替换为3×3深度可分离卷积（即3×3深度卷积和1×1点卷积）。5×5卷积层被两个3×3深度可分离卷积所取代，如图3(b)所示。Inception B块中的7×7卷积层被7×7深度可分离卷积（7×7深度卷积和1×1点卷积）取代，如图4(b)所示。Reduction A块中的3×3卷积层被1×1卷积层和3×3深度可分离卷积取代。Reduction B块中的3×3和7×7卷积层被1×1卷积层和3×3深度可分离卷积所取代。表2和表3分别展示了原始Inception A块与经过深度可分离卷积修改后的Inception A块所需的参数数量对比。从表中可以看出，修改后的Inception A块所需参数量显著减少。

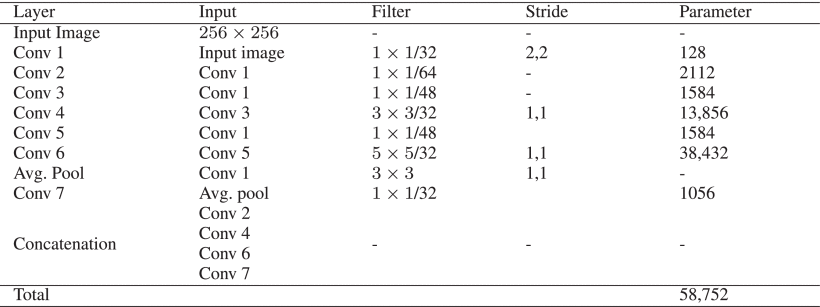


表 2 原始Inception-A块中所需的参数

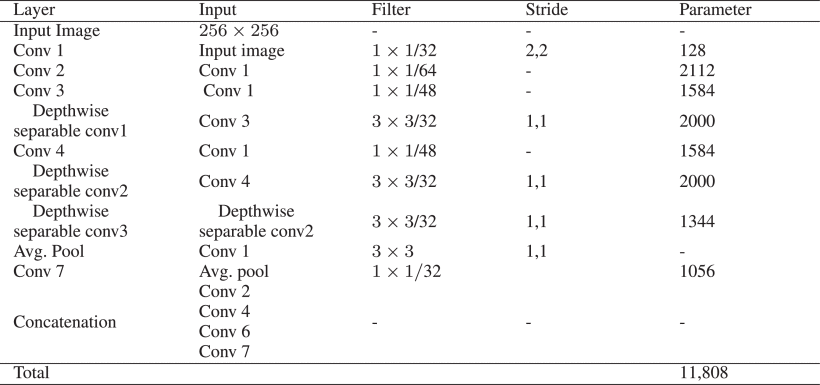


表 3 改进的Inception-A块中具有深度可分离卷积所需的参数

图5展示了所提出的卷积神经网络模型架构，该模型用于识别植物病害。模型主要包含以下部分：卷积层、批量归一化（Batch Normalization）和激活层、带有深度可分离卷积的Inception块、池化层和全连接层。

在该架构中，标准卷积被深度可分离卷积取代。模型采用了一个标准卷积层、三个深度可分离卷积层、两个最大池化层和一个全局平均池化层。此外，模型包括三个带有残差连接的改进型Inception A块，后接改进型Reduction A块，以及三个带有残差连接的改进型Inception B块，后接改进型Reduction B块。每个卷积层后均添加批量归一化和ReLU激活函数，以提升模型性能并加速训练过程。在全局平均池化后，添加了Dropout层，以减少模型过拟合的风险。该模型的总参数数量为428,100，而标准Inception V3模型的参数数量高达23,851,784。结果表明，所提出的模型相比Inception V3减少了约70%的参数量，同时在准确率方面保持了较高水平。

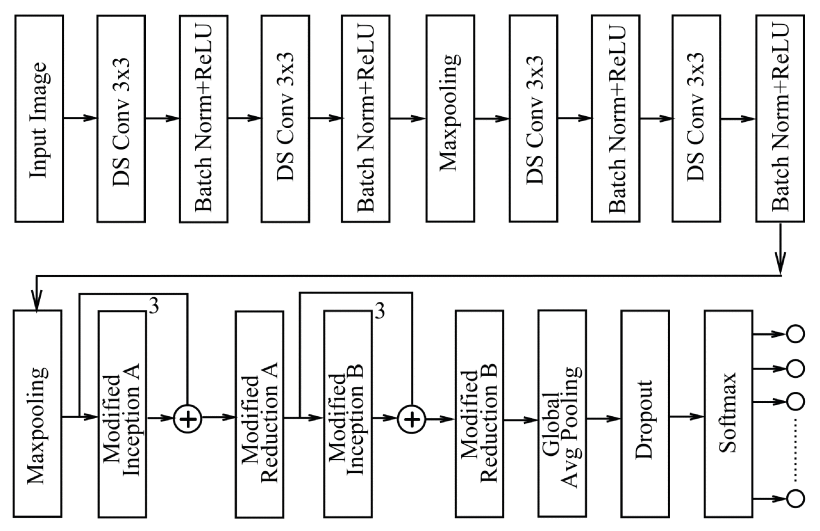


图 5. 提出的卷积神经网络方法在植物疾病识别中的应用

**IV. 结果与讨论**

**A. 数据集**

本文使用三个不同的植物病害数据集对模型性能进行评估，包括水稻病害数据集、木薯病害数据集和PlantVillage数据集。PlantVillage数据集是一个大型且可用的植物病害数据集，本文使用了其中的玉米、马铃薯和番茄病害图像。PlantVillage数据集的图像在统一背景和实验室设置条件下拍摄，而水稻病害数据集的图像在田间实时采集，木薯数据集的图像在田间条件下拍摄，并包含多片叶子。水稻病害数据集包含5932张图像，分为四类：1584张细菌性条斑病图像，1440张稻瘟病图像，1600张褐斑病图像，以及1308张黄矮病图像。木薯病害数据集包含5656张图像，包括五个类别：316张健康木薯叶片图像，466张木薯细菌性枯萎病图像，1443张木薯褐条病图像，773张木薯绿螨病图像，以及2658张木薯花叶病图像。木薯病害数据集的图像包含复杂背景，且单张图像中存在多片叶子。数据集按照80:20的比例随机划分为训练集和测试集，图像尺寸调整为256×256像素。图6展示了不同数据集的示例图像，表4、表5和表6详细描述了数据集的组成、病害类别以及每类图像数量。

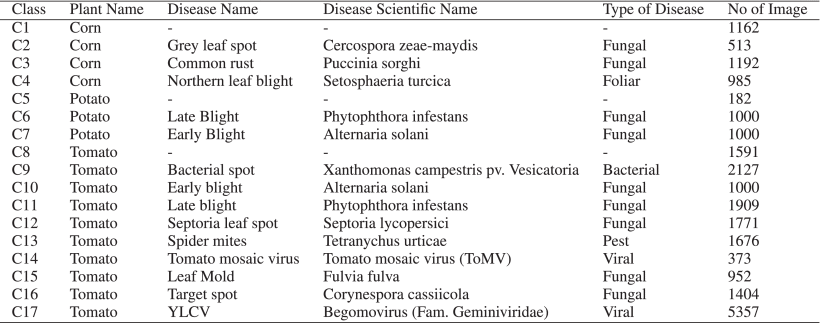


表4 Plantvillage数据集（玉米、马铃薯、番茄）数据描述

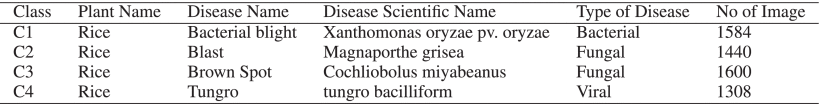


表5 水稻病害数据描述

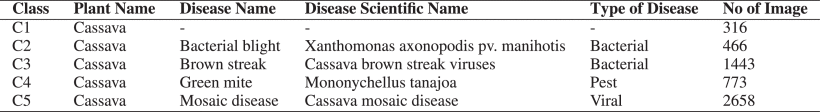


表6 木薯数据集的数据描述

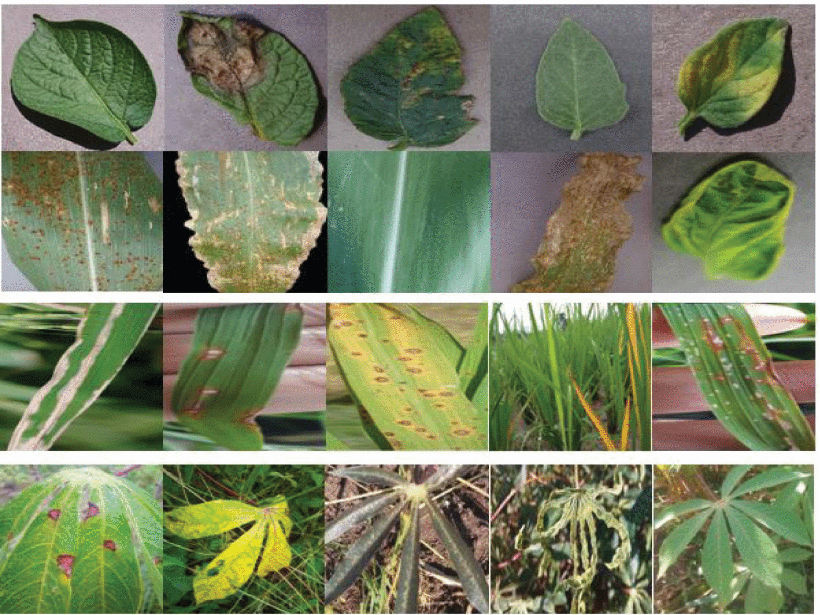
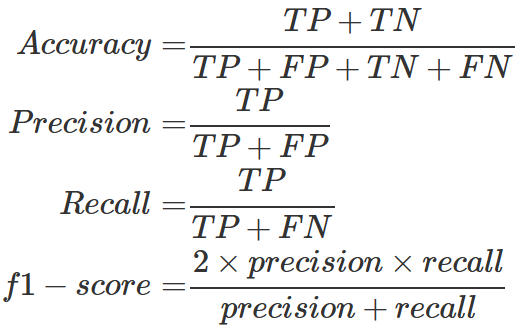


图 6.Plantvillage数据集、水稻和木薯植物数据集的样本图像

**B. 实验结果**

为了评估模型的性能，本文使用了多项性能指标，包括参数数量、准确率、精确率、召回率和F1分数，其计算公式如下：



其中，TP为真阳性，TN为真阴性，FP为假阳性，FN为假阴性。

表7展示了所提出模型在不同数据集上的性能表现。经过50轮训练后，该模型在PlantVillage数据集上的最高训练准确率和损失分别达到99.81%和0.0015，验证准确率和损失分别为99.39%和0.0549。在水稻病害数据集上，模型取得了99.94%的最高训练准确率和0.0030的训练损失，验证准确率和损失分别为99.66%和0.0041。为了验证模型的健壮性，本文还使用了木薯病害数据集，该数据集的图像在真实田间条件下拍摄，包含复杂背景，并且单张图像中可能包含多片叶子。与PlantVillage和水稻数据集相比，由于数据不平衡和图像复杂性，木薯病害数据集的训练准确率和验证准确率分别为98.17%和76.59%。与PlantVillage和水稻病害数据集相比，木薯数据集的性能较低，这可能是由于数据不平衡以及图像背景复杂造成的。图7至图9展示了模型在PlantVillage、水稻和木薯病害数据集上的训练和验证准确率以及损失情况。

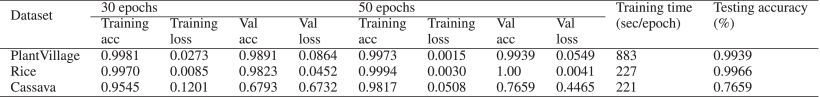


表 7 基于深度学习的实施方法总结

将数据集分成80%-20%的训练集和验证集后，我们对模型进行最多50次训练，并评估训练集和验证集上的性能。图7-9分别显示了所提模型在PlantVillage、水稻和木薯数据集上的训练和验证的准确率和损失。从图7-9可以看出，所提模型以较少的参数提供了令人满意的性能。

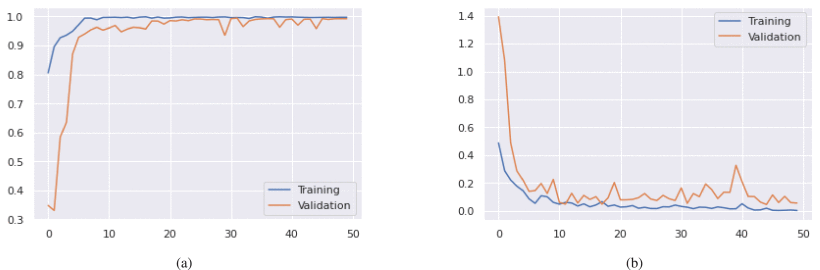


图 7. (a)PlantVillage数据集上的训练和验证准确率(b)训练和验证损失

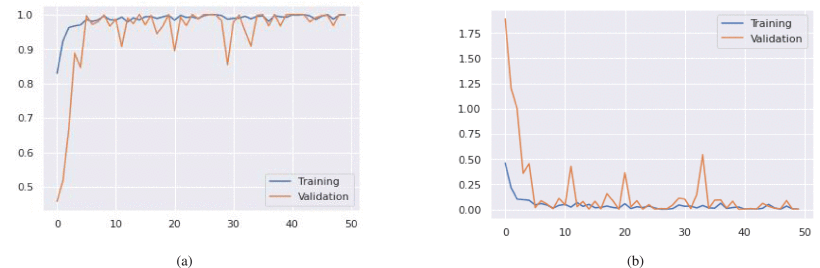


图 8. (a)水稻病害数据集的训练和验证准确率(b)训练和验证损失

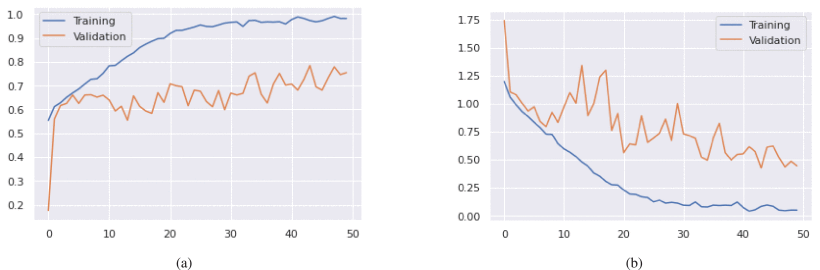


图 9. (a)训练和验证准确度(b)木薯植物数据集的训练和验证损失

我们还计算了混淆矩阵来评估模型的性能。表8显示了实施模型在三种植物数据集上的测试准确度、精确度、召回率和f1分数方面的性能。从表8可以看出，水稻植物数据集的测试准确度高于PlantVillage和木薯数据集。图10给出了实施模型在不同数据集上的性能指标。从图10可以看出，水稻和plantvillage数据集的平均预测准确度、精确度、召回率和F1分数超过99%。

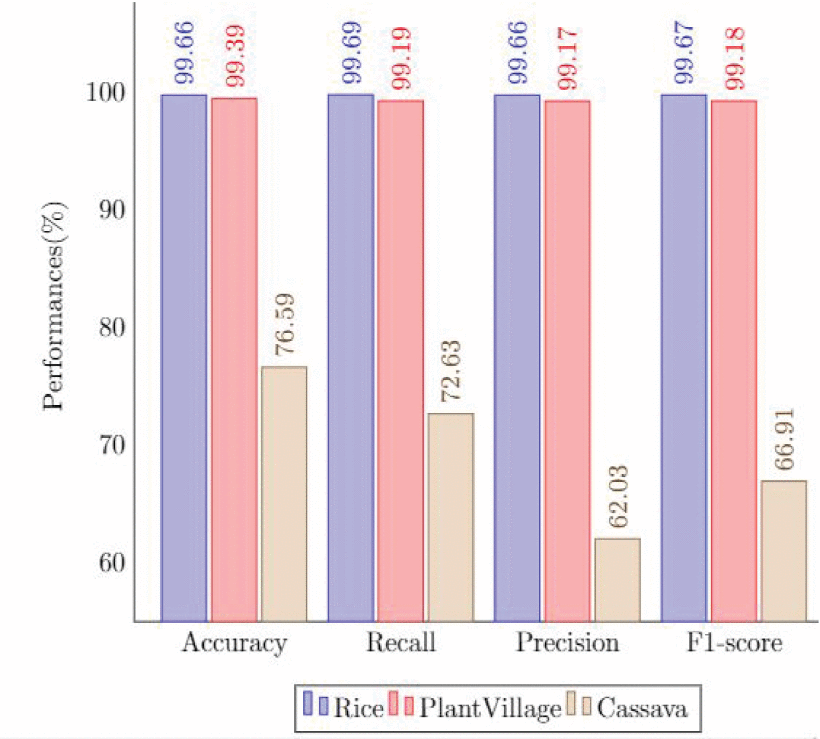


图 10. 所提出的模型在不同数据集上的性能指标

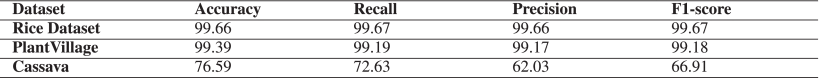


表 8 所提模型在测试图像上的性能指标

**C. 模型性能与健壮性比较**

为了验证所提出模型的稳定性，本文采用了5折交叉验证方法对病害数据集进行评估。数据集被随机分为五个相等部分，每次选择其中一部分作为测试集，其余部分作为训练集。基于不同的训练-测试划分，可以获得模型在不同数据集上的表现。表9展示了所提出模型在交叉验证过程中的性能表现。结果表明，每个数据集不同折之间的性能差异较小。在PlantVillage数据集上，模型准确率在99.29%到99.37%之间浮动；在水稻病害数据集上，准确率在99.33%到99.66%之间；而在木薯病害数据集上，准确率范围为76.42%到76.58%。结果表明，本文提出的CNN模型在植物病害识别任务中具有良好的稳定性和一致性。

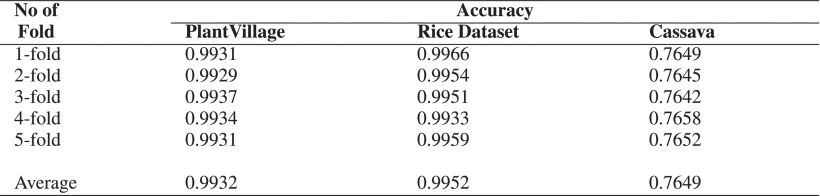


表9 基于k倍交叉验证的CNN结果

**D.与预训练网络的性能比较**

本文还对比了多种预训练深度学习模型，包括VGG16、VGG19、InceptionV3、ResNet50和DenseNet201，并将这些模型的性能与本文提出的CNN模型进行对比。性能比较指标包括模型在不同数据集上的准确率以及每轮训练所需的时间。表10展示了所提出模型与其他预训练模型在不同数据集上的性能对比。结果显示，本文模型在PlantVillage数据集上的测试准确率为99.39%，在水稻病害数据集上的准确率为99.66%，在木薯病害数据集上的准确率为76.59%。此外，本文模型的参数数量远少于InceptionV3模型。每轮训练的时间显著减少，说明本文模型在小型设备或计算资源受限的环境中更具实际应用价值。

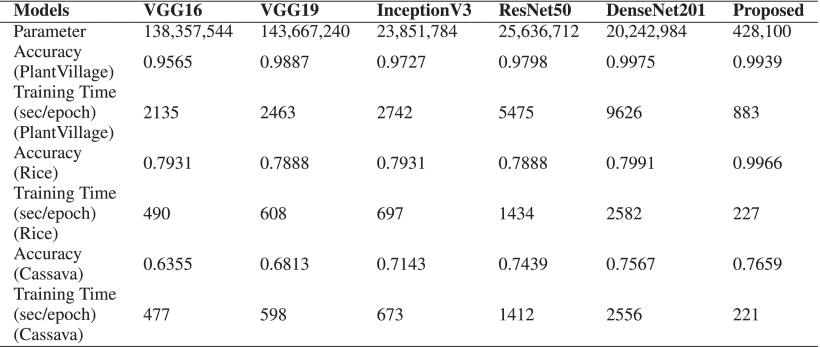


表10 与预训练网络的性能比较

**E.与现有文献的性能比较**

为了进一步验证所提出模型的有效性，本文将模型的性能与现有文献中的深度学习模型进行了对比。表11展示了本文模型与其他模型的性能比较。结果表明，本文模型在准确率、参数数量和训练时间等方面均优于现有模型。表11的结果还显示，本文模型所需参数数量远少于其他深度学习模型，因此在训练过程中耗时更少。与其他模型主要在PlantVillage数据集（实验室环境下）进行测试不同，本文模型在真实田间环境下的数据集（如水稻和木薯病害数据集）上表现出较强的健壮性和泛化能力。综上所述，所提出的CNN模型在植物病害识别任务中展现出更高的准确率、更少的参数量和更快的训练速度，同时在资源受限的设备上更具部署优势。

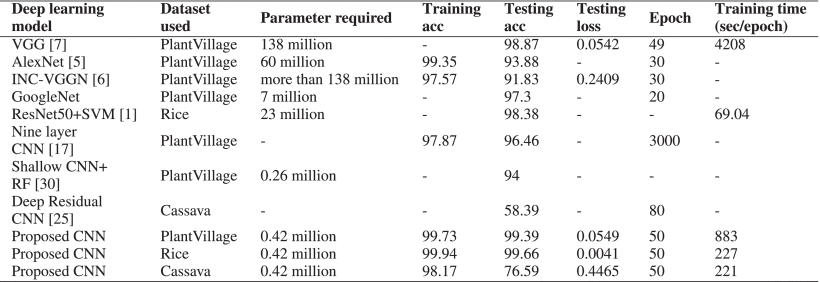


表 11 不同深度学习模型的性能比较

**V. 结论**

深度学习是一种在图像模式识别领域中具有前沿性的先进技术，在植物病害识别方面表现出极高的有效性。本文提出了一种基于Inception和残差连接的新型卷积神经网络（CNN）模型，该模型能够有效地对植物病害进行分类。此外，在Inception架构中引入了深度可分离卷积，从而减少了约70%的参数量，显著降低了计算成本。因此，与标准CNN模型相比，训练该网络所需的时间大大减少。实验结果表明，所提出的模型在不同植物病害数据集上均取得了较高的识别准确率。为了验证模型的健壮性，本文采用了三个不同的植物病害数据集进行测试。该模型在PlantVillage数据集上的测试准确率为99.39%，在水稻病害数据集上的准确率为99.66%，在木薯病害数据集上的准确率为76.59%。在不平衡数据集上，其他研究中使用普通卷积神经网络和深度残差神经网络所取得的准确率分别为52.87%和46.26%，而在平衡数据集上，其最高准确率可达80.6%。相比之下，本文模型在不平衡木薯病害数据集上取得了76.59%的准确率，表现优于现有模型。在水稻病害数据集上，Sethy等人的模型取得了98.25%的准确率，而本文模型在相同数据集上达到了更高的99.66%的准确率。未来的研究方向包括将该模型应用于农业领域的其他任务，如杂草检测、害虫识别等。此外，还将探索在不同种类的植物病害数据集上进行模型验证，这些数据集将涵盖不同的植物种类、图像类型以及来自不同地理区域的图像数据。未来的工作还将尝试使用基于聚类的无监督技术进行植物病害识别，以进一步提升模型的泛化能力和实用性。

原文2

**Recent advances in plant disease severity assessment using convolutional neural networks**

Tingting Shi, Yongmin Liu, Xinying Zheng, Kui Hu, Hao Huang, Hanlin Liu & Hongxu Huang

**Abstract**：In modern agricultural production, the severity of diseases is an important factor that directly affects the yield and quality of plants. In order to effectively monitor and control the entire production process of plants, not only the type of disease, but also the severity of the disease must be clarified. In recent years, deep learning for plant disease species identification has been widely used. In particular, the application of convolutional neural network (CNN) to plant disease images has made breakthrough progress. However, there are relatively few studies on disease severity assessment. The group first traced the prevailing views of existing disease researchers to provide criteria for grading the severity of plant diseases. Then, depending on the network architecture, this study outlined 16 studies on CNN-based plant disease severity assessment in terms of classical CNN frameworks, improved CNN architectures and CNN-based segmentation networks, and provided a detailed comparative analysis of the advantages and disadvantages of each. Common methods for acquiring datasets and performance evaluation metrics for CNN models were investigated. Finally, this study discussed the major challenges faced by CNN-based plant disease severity assessment methods in practical applications, and provided feasible research ideas and possible solutions to address these challenges.

**I. Introduction**

Plant diseases caused by various organisms that damage plant growth, such as pests, bacteria or fungi, are a major cause of agricultural losses. Reliable and accurate methods for assessing disease severity are essential for effective disease control and minimizing yield loss. There are several ways to assess the severity of plant diseases. The traditional method of determining disease severity is by Visual Assessment, which is highly unreliable due to the similarity of diseases and the diversity of characteristics that are susceptible to external factors and subjective individual differences. Visual Assessment usually needs to be carried out by experienced specialists, which is not efficient, and many farmers do not have access to specialists, making accurate and timely disease severity identification very difficult. In addition, hyperspectral imaging has been used to measure the severity of plant diseases, but this technique requires sophisticated equipment such as sensors and a certain level of expertise, making it costly and inefficient.

In recent years, with the rapid development of computer imaging technology and the continuous improvement of the hardware performance of related electronic devices, computer vision and artificial intelligence (AI) have been widely used in the field of agricultural diagnosis, such as plant species classification, leaf disease identification and plant disease severity estimation. Deep learning has now made significant breakthroughs in the field of computer vision, and CNN has shown excellent performance in plant disease detection applications. Compared to traditional methods, CNN is able to automatically and directly extract features from the input image, eliminating the need for complex image pre-processing and enabling end-to-end detection methods. To date, satisfactory results have been achieved in identifying plant disease species using CNN, but little research has been done in the area of disease severity assessment. This study focuses on the application of CNN for plant disease severity assessment, and systematically reviews the related research to provide reference ideas for further research work.

The remainder of this review is organized as follows: the second part provides an overview of the concepts related to the Visual Assessment of plant disease severity. The third part reviews the history of the development of CNN. The fourth part deals with the specific application of CNN on plant disease severity, illustrating the differences between single-task and multi-task systems. And we focus on the basic working principles of CNN-based plant disease severity assessment methods from three aspects: classical CNN framework, improved CNN architecture, and CNN-based semantic segmentation network, and analyze the advantages and disadvantages of each method. The fifth part summarizes the relevant public datasets and presents CNN performance evaluation metrics. The sixth section discusses the major challenges that CNN-based plant disease severity assessment may face in practical applications, and provides feasible research ideas and possible solutions to these challenges.

**II. visual assessments**

**Definition of plant disease severity**

Plant disease severity, defined as the ratio of plant units with visible disease symptoms to the total plant unit (e.g. leaves), is an important quantitative indicator for many diseases. Timely and accurate assessment of disease severity is critical in crop production because disease severity directly affects crop yield and is often used as a predictor to estimate crop loss with excellent accuracy. For example, severity indicators can be used as decision thresholds or disease forecasts to help growers rationalize disease control, such as deciding on the dose and type of pesticide and the time of day to spray.

**Visual assessment methods for plant disease severity**

Accurate measurement and evaluation of disease severity is critical to agricultural production. Because it ensures a correct analysis of treatment effects, an accurate understanding of the correlation between yield loss and disease severity, and a reasonable assessment of plant growth stages. Inaccurate or unreliable disease assessment can lead to erroneous conclusions, resulting in the wrong disease management actions, which can further exacerbate losses. Assessment of disease severity is typically done using a variety of scales, including nominal (descriptive) scales, ordinal rating scales, interval (category) scales, and ratio scales. The following is an overview of these scales for visual assessment of disease severity, both qualitative and quantitative.

Qualitative scales.

(1) Descriptive scale: This is one of the simplest and most subjective criteria in the disease severity grading scales. The disease is divided into several categories with descriptive terms such as mild, moderate, and severe. Due to the subjectivity and lack of quantitative definitions, the value of this scale is very limited, except for ratings in a specific condition.

(2) Qualitative ordinal scale: This is still the descriptive disease scale, but provides more variety in the categories of disease severity levels than the descriptive scale. For example, Xu et al. assigned a scale of 0–5 to describe the severity of symptoms of zucchini yellow mosaic virus and watermelon mosaic virus to indicate increasing disease severity. This scale is widely used for certain diseases, especially for assessing viral diseases with symptoms that are not easily quantifiable.

Quantitative scales.

(1) Quantitative ordinal scale: This scale consists of numbers in known categories, usually the percentage of symptomatic areas. It can be further divided into two types: equal interval and unequal interval. However, equal interval rating scales may give a higher average severity, especially if the actual severity is at the lower end of a category, because the interval is so wide that it is difficult to show differences, leading to an inaccurate rating. Some disease rating scales have unequal intervals. The Horsfall-Barratt scale (H–B scale) is a widely used unequal interval scale. It was developed by Horsfall and Barratt, which effectively alleviates the problem of equal intervals. For example, Bock et al. used the scale to estimate the severity of citrus ulcer disease. Forbes et al. used the H-B scale to estimate the severity of potato late blight in the field, etc.

(2) Ratio scale: This scale is widely used for visual assessment of severity. The grader measures the percentage of symptomatic organs, defined as 0% to 100%, and rates the severity accordingly. Therefore, the ratio scale places greater demands on the rater to identify and measure the actual disease more accurately.

Although plant disease severity can be assessed by several different methods, both qualitative and quantitative assessment methods tend to result in assessments that are inconsistent with reality due to factors such as the subjectivity of individual raters, the tendency to overestimate disease severity when it is low, and the bias of raters toward 5% whole number intervals. To improve the accuracy of rater estimates, the Standard Area Map (SAD) has long been used as a tool to help estimate plant disease severity. Professional training of raters can also be effective in improving the accuracy of the assessment.

**III. History of CNN development**

Deep learning began with the introduction of threshold logic in 1943 and is essentially a process of building computer models that closely resemble human neural networks16. CNN is a subset of deep learning that first appeared in the 1980s. In the beginning, the concept of receptive field was developed and later introduced into CNN research. Later, with the introduction of the BackPropagation (BP) algorithm and the training of multi-layer perceptron, researchers tried to automatically extract features instead of manually designing features. LeCun et al. proposed a CNN architecture called “LeNet-5” using BP networks, which outperformed all other techniques on a standard handwritten digit recognition task at the time. Research on deep neural network models was put on hold due to a number of problems encountered with traditional BP neural networks, such as local optima, overfitting, and gradient disappearance with increasing number of network layers, and the accompanying proposal of some shallow machine models at that time19. Until about 2006, Hinton et al. found that artificial neural networks with multiple hidden layers have excellent feature learning capabilities. Glorot et al. mitigated the problem of disappearing gradients during training with a normalization method. Attention shifted back to deep learning. In 2012, AlexNet won the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC), and since then, DL has attracted the attention of more and more researchers, and AlexNet is considered a major breakthrough in the field of deep learning. Next, the CNN architecture continues to evolve, and many algorithms with excellent performance emerge. The main classical CNN networks are LeNet, AlexNet, VGG, GoogLeNet, Resnet, DenseNet, and so on. The evolutionary sequence order from LeNet to DenseNet is shown in Fig. 1.

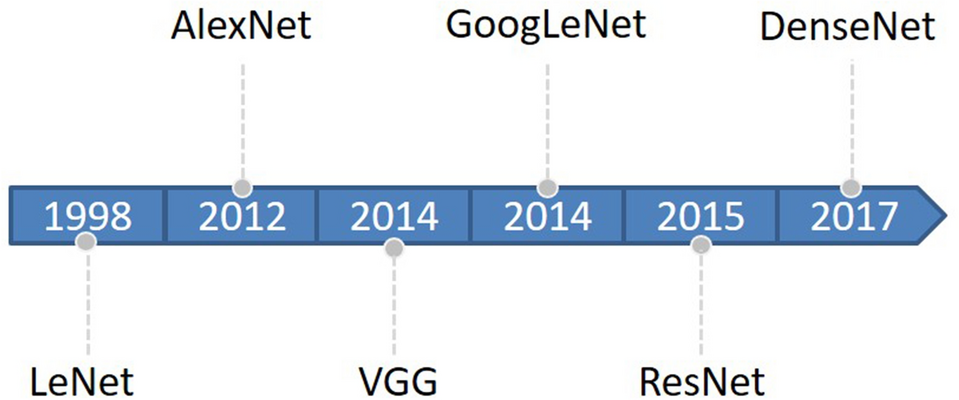


Figure 1 Evolution timeline of CNNs from LeNet to DenseNet

As CNN evolves, new CNN models are constantly emerging that implement different features. For example, lightweight networks: SqueezeNet, MobileNet, ShuffleNet, Xception, EfficientNet. Target detection networks: R-CNN, Fast R-CNN, Faster R-CNN, YOLO, SSD. Segmentation networks: FCN, SegNet, U-Net, PSPNet, DeepLab, Mask RCNN, etc., they show excellent performance and great research value.

**IV. CNN-based plant disease severity assessment method**

CNN has been used with great success to assess the severity of plant diseases. Automatic estimation of plant disease severity based on CNN was first proposed by Wang et al.44 in 2017. They used different CNN models to classify apple black rot images with four severity levels and achieved an overall accuracy of 90.4% on the test set, suggesting that CNN is a promising new technique for fully automated plant disease severity classification. Liang et al.3 proposed PD2SE-Net to implement a multitask system for disease severity estimation, plant species identification, and plant disease classification with overall accuracies of 0.91%, 0.99%, and 0.98%, respectively. Su et al.45 combined ResNet-101 network and semantic segmentation to rapidly predict the severity of Fusarium head blight (FHB) in wheat with a prediction accuracy of 77.19%.

**Single-task versus multi-task systems**

Deep learning tends to focus on optimizing for specific metrics. In other words, a model or a set of models is often trained to perform the single target task, and such systems are known as single-task systems46. On the other hand, there is the concept of multi-task learning (MTL), where multiple tasks can be learned simultaneously if they are linked together47. Experimental studies have shown that learning features from multiple related tasks simultaneously is more beneficial than learning them independently in terms of prediction performance. MTL can reduce the risk of overfitting in each task by learning tasks in parallel and thus using more features from different tasks, leading to better generalization of the model.

Studies using CNN for plant disease detection include single-task systems that individually identify plant disease species or estimate disease severity. For example, Prabhakar et al. used ResNet101 to assess the severity of leaf blight in tomato. Zeng et al. trained six different CNN models to classify the severity of citrus yellow shoot. There are also multitasking systems that perform both tasks simultaneously. For example, José G.M. Esgario et al. used CNN to implement a classification of coffee leaf disease species and severity grading. Fenu et al. considered five pre-trained CNN architectures as feature extractors for the classification of three diseases and six severity levels, whose experimental results show that the trained model is robust in automatically extracting disease leaf identification features using a multi-task learning model.

**Application of CNN in plant disease severity assessment**

To clarify the specific implementation process of CNN for plant disease severity assessment, 16 high-quality articles that fit the research topic were selected for this study. First, a search was conducted on the Web of Science platform, one of the world's largest and most comprehensive scientific information resources54. According to55, the process of collecting research sets requires the definition of search terms, so the keywords of “Convolutional neural network” (Topic) and “plant disease severity” (Topic) were entered into Web of Science, and as of 2022, 57 articles were retrieved with the year of publication shown in Fig. 2. Among the 57 papers, 16 papers were selected for specific analysis based on the research object (plant disease) and research method (CNN). On this basis, the most recent research in 2022 is analyzed separately. According to the different CNN network architectures used in these 16 articles, they are further divided into three categories: classical CNN framework, improved CNN architecture, and CNN-based segmentation network. The flowchart of the CNN-based method for plant disease severity assessment method is shown in Fig. 3.

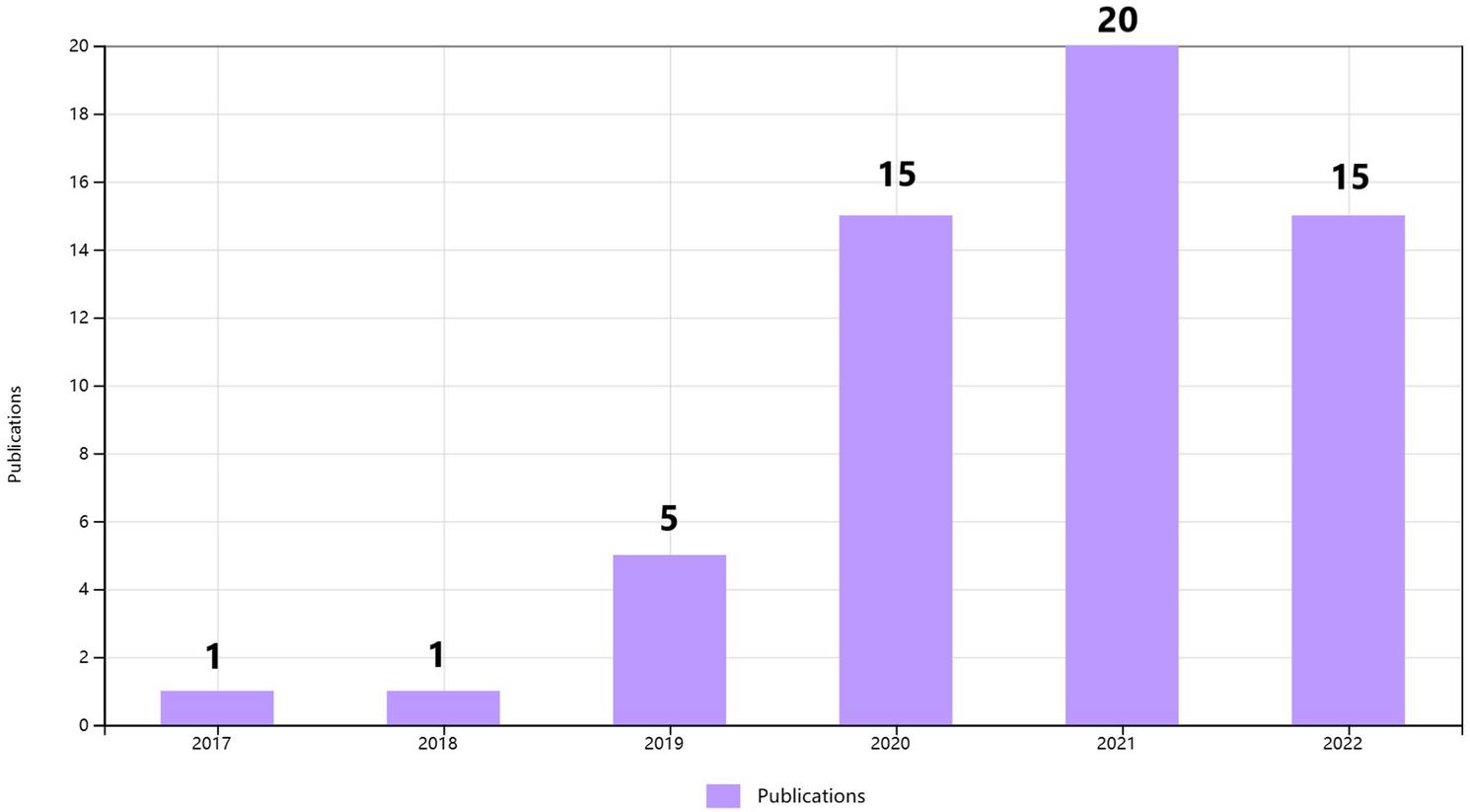


Figure 2 The distribution graph of the publication years of 57 articles based on the keywords of “convolutional neural network” and “plant disease severity”

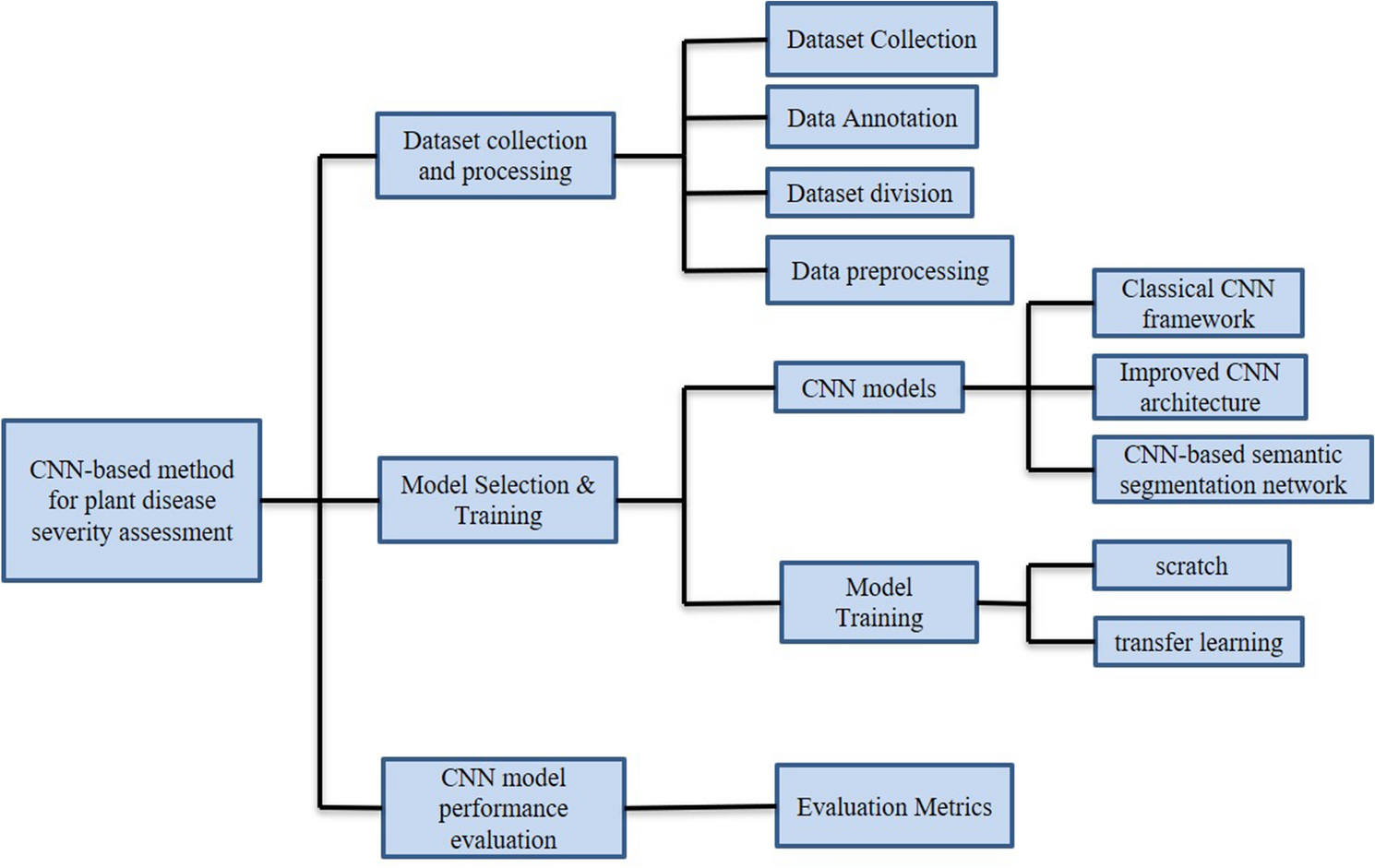


Figure 3 The flowchart of the CNN-based plant disease severity assessment method

**Classical CNN framework**

10 of the 16 articles are based on the classical CNN framework for implementing severity grading. The 10 studies differ in specific CNN frameworks and research topics, but are similar in that they are CNN-based approaches to assessing plant disease severity. Therefore, they have similarities in the specific implementation process. The process of implementing plant disease severity assessment based on the classical CNN framework can be divided into the following three main steps.

The first step is to collect and process datasets, and is described in four aspects.

(1)Dataset characteristics. Of the 10 studies, 6 studies used self-made datasets and 4 studies used the images from PlantVillage images. The self-made datasets can be further divided into two types. One is for images taken under controlled conditions. For example, in46, The photos were taken from the abaxial (lower) side of the leaves under partially controlled conditions and placed on a white background. The other is for images taken under natural conditions with a complex background. In contrast, the background of the images in PlantVillage is uniform and homogeneous. Making your own dataset is a time-consuming and expensive process, but it is more in line with what happens in a real environment. A large number of studies have demonstrated that when models trained on controlled images are used to predict images collected from real-world environments, their accuracy is significantly reduced56,57,58. If a public dataset does not meet the needs of a particular study, self-made datasets must be produced.

(2)Dataset annotation. One of the necessary conditions for assessing severity is that the records are labeled with different severity levels. Of the 10 articles, 3 were labeled according to the descriptive scale, 1 according to the qualitative ordinal scale, 4 according to the quantitative ordinal scale, 1 article did not indicate the labeling method in the article. For example, in46, a quantitative ordinal scale was used. Severity was classified into five levels according to the proportion of diseased leaves: healthy (< 0.1%), very low (0.1–5%), low (5.1–10%), high (10.1%-15%), and very high (> 15%).

(3)Dataset division. The dataset is usually divided into three parts: training dataset, validation dataset and test dataset. The training set is used to train the model, the validation set is used to tune the hyperparameters, and the test set is used to evaluate the model performance59. The 10 studies all basically used 70% to 85% of the dataset for training. Mohanty et al.60 tried five different separation ratios to partition the dataset, and the experimental results showed that using 80% of the dataset for training and 20% for validation was ideal for their data.

(4)Data preprocessing. Typically, two preprocessing operations are performed before the images are fed into the CNN. One is to resize the images to match the input layer requirements. For example, The image size in PlantVillage is 256 × 256, and the AlexNet input layer requires a size of 227 × 227, then the original photos need to be resized. This processing is reflected in all 10 studies. Second, the images are normalized to help the model converge faster, significantly improving the efficiency of end-to-end training61.

The second step is the model selection and training phase, which is described below in two aspects.

(1)CNN framework selection. The CNN frameworks used in the 10 studies include AlexNet, VGG, GoogLeNet, ResNet, DenseNet, MobileNet, Inception, Faster R-CNN, YOLO, EfficientNet, SqueezeNet, Xception, etc. The vast majority of these studies have used multiple CNN frameworks in comparative experiments to determine which model is better at detecting the severity of a particular plant disease under the same training conditions52.

(2)Training methods. There are two ways to train CNN, one is to start from scratch and the other is transfer learning. Transfer learning refers to adapting a pre-trained network on a large set of images, such as ImageNet (1.2 million images in 1000 classes), to a different task, which is implemented by the underlying CNN learning non-specific features62. There are two approaches to transfer learning: feature extraction and fine-tuning. Feature extraction is the process of keeping the weights of a pre-trained model unchanged and then using them to train a new classifier on the target dataset. Fine-tuning involves initializing the model using the weights from a pre-trained model, and then training some or all of the weights on the target dataset63. Brahimi et al.64 used three approaches of feature extraction, fine-tuning, and training from scratch to train six CNN models. And the results suggested that the fine-tuning models had the highest accuracy and the feature extraction models had the shortest training time. 8 of the 10 studies used transfer learning and only one was trained from scratch. In44, the two methods of training models were compared and the results showed that transfer learning alleviated the problem of insufficient training data.

**The final step is to evaluate the performance of the CNN models.**

The performance of CNN models is obtained by using the test set on the trained model. It is critical that the test set is independent of the training and validation sets, otherwise the evaluation results may be highly biased. Mohanty et al.60 trained a model to identify 14 crops and 26 diseases with an overall accuracy of 99.35% on a test set, where there was no clear separation between the validation and test set. When they tested the model on a set of images taken under different conditions than the training images, the model’s accuracy dropped dramatically to 31%. It is worth noting that only 4 of the 10 studies explicitly distinguish three types of datasets. Sibiya et al.65 explicitly separated the training set, validation set, and test set. The experimental results showed that the proposed model was neither overfitting nor underfitting, as the model achieved the accuracy of 95.63% on the validation set and a high accuracy of 89% on the test set.

Quantitative assessment of model performance is achieved through evaluation metrics. Evaluation metrics typically include accuracy, precision, recall, mean average precision (mAP), and F1 score based on precision and recall. With the development of deep learning, the performance of CNN models on different datasets has been improved, and various evaluation metrics have been increased. A consistent performance comparison of CNNs from different studies is difficult to achieve because most CNN-based studies for plant disease severity assessment apply to specific datasets, many of which are not yet publicly available and do not provide all the parameters needed to reproduce experiments.

**Improved CNN architecture**

2 of the 16 articles are based on an improved CNN architecture for severity assessment. Comparing the classical CNN framework with the improved CNN architecture, the similarity is that the implementation process is basically the same, and the difference is that the latter uses an improved network based on the classical CNN with the aim of designing a higher performance and more practical system for plant disease diagnosis.

In3, a network, PD2SE-Net, was proposed to design a more excellent and practical plant disease diagnosis system. PD2SE-Net introduced the ResNet50 network as the base model and integrated the building blocks of ShuffleNet-V230. The PD2SE-Net architecture is shown in Fig. 4. There are two key components of the PD2SE-Net that make it so effective. One is the introduction of a residual structure to construct the parameter sharing layers, which allows the model more information to update per batch. Inspired by ShaResNet66, ResNet50 was used to build the basic framework and integrated with parameter sharing to reduce the redundant information in the network. The other is the introduction of shuffle units. The ShuffleNet-V2 units were used to extract the feature maps of different plant species and diseases with low computational complexity. Finally, PD2SE-Net achieved plant species recognition, disease classification, and severity estimation with overall accuracies of 0.99, 0.98, and 0.91, respectively.

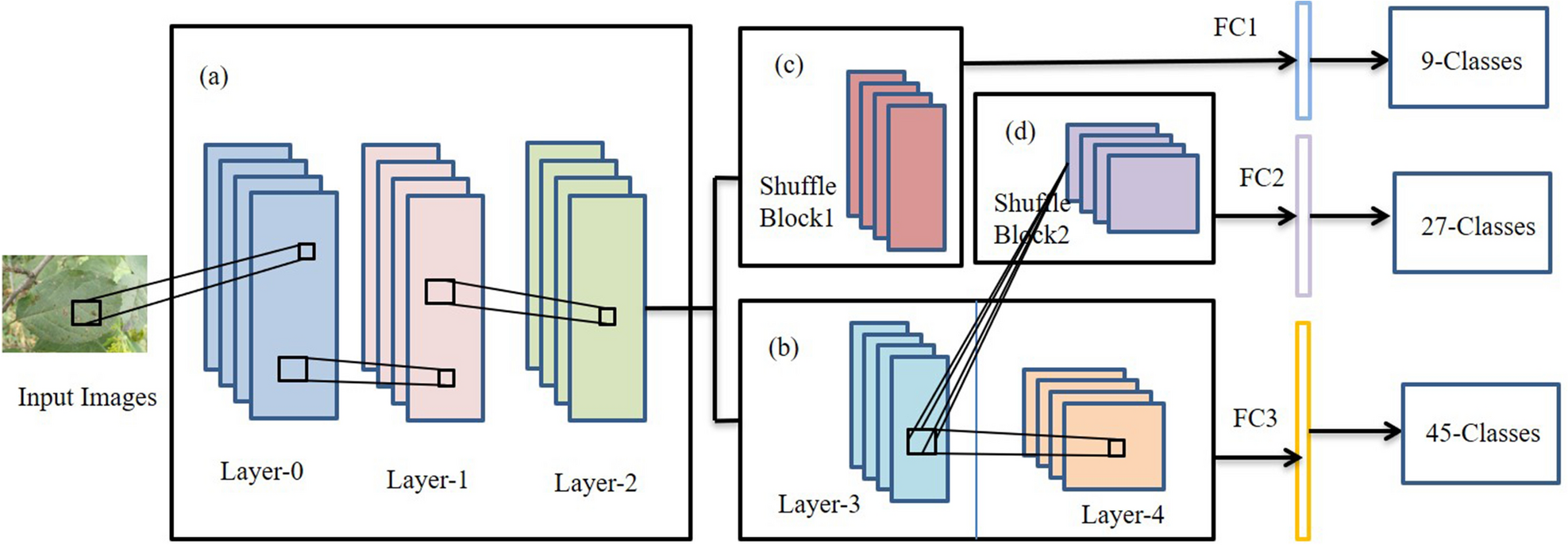


Figure 4 The architecture of PD2SE-Net. Five parts: (a) parameter sharing layer; (b) the third layer is the parameter sharing layer between the fourth layer and shuffling block 2, and the fourth layer is the high-dimensional feature extractor for severity estimation; (c) a feature extractor for plant species recognition; (d) feature extractor for plant disease diagnosis; (e) fully connected layers3.

Xiang et al. proposed a lightweight network, L-CSMS, based on residual networks, channel shuffle operation and multi-size module for plant disease severity assessment. Multiple convolution kernels of different sizes were used in the multiscale convolution module to extract different receptive fields in order to obtain robust features and spatial relationships from feature maps68. Channel shuffle operation was introduced to enable information communication between different channel groups and to improve accuracy. The channel shuffle operation and the multi-size convolution module were integrated into the building block as a stacked topology, as is shown in Fig. 5. L-CSMS used the residual learning approach of ResNet to build a deep network by stacking modules of the same topology. To validate the performance of the L-CSMS model, Xiang et al.67 conducted comparative experiments between the L-CSMS model and ResNet, DenseNet, Inception-V4, PD2SE-Net, ShuffleNet, and MobileNet. The results showed that L-CSMS achieved a competitive advantage with fewer parameters, FLOPs, and comparatively good accuracy.

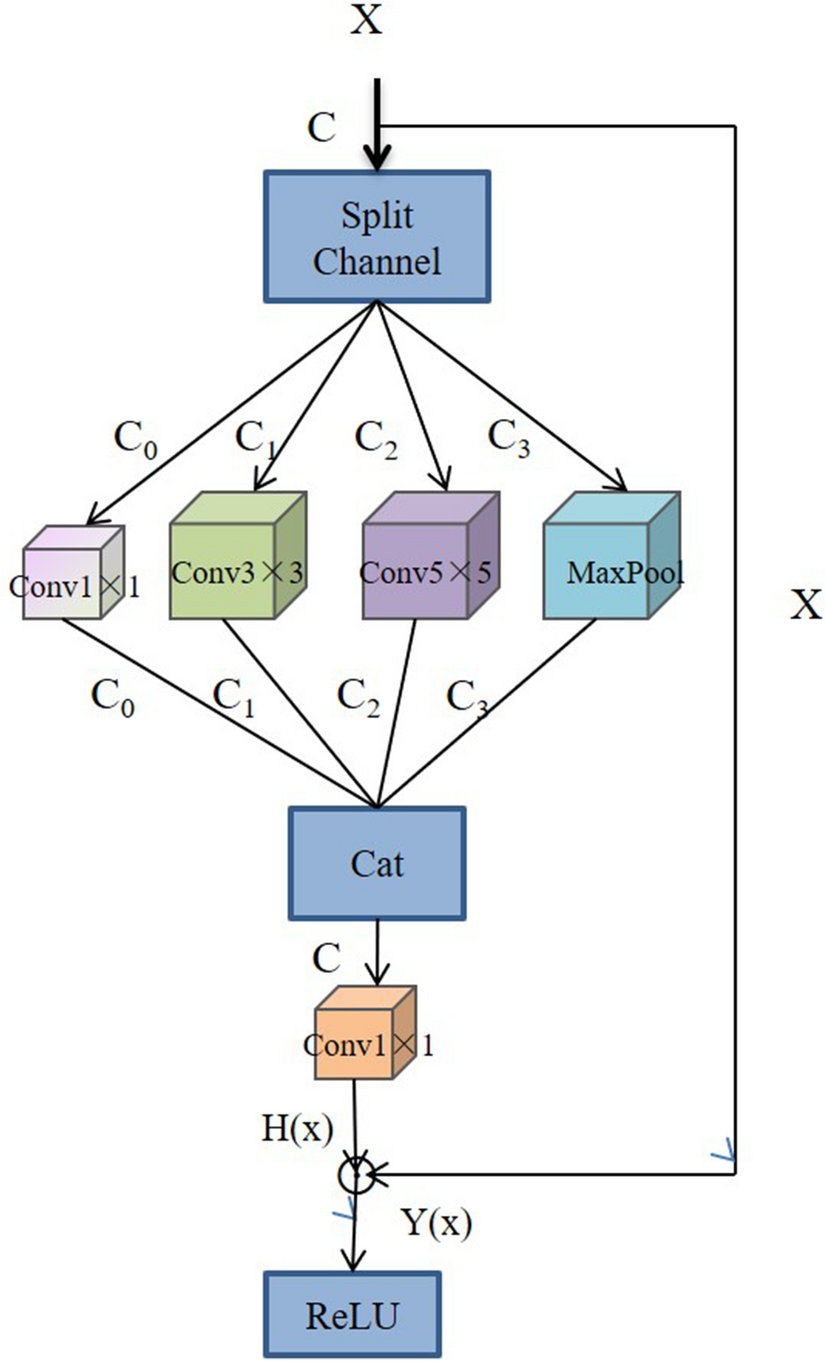


Figure 5 The building block with channel shuffle operation and multi-size convolution module67

The ways to improve the CNN architecture are different, but the improvements are basically aimed at the same goal, which is to design a more accurate and practical plant disease severity assessment system with better generalization performance. Both studies focused on the residual structure in the ResNet and the channel shuffle.

**CNN-based semantic segmentation network**

Image semantic segmentation has received increasing attention from computer vision and DL researchers, and research work on semantic segmentation using DL techniques continues to evolve. In particular, CNN has far outperformed other methods in terms of accuracy and efficiency69. CNN-based segmentation provides not only category information, but also additional information about the spatial location of those categories. The task of semantic segmentation is to label each pixel as a kind of closed objects or a category of regions70. CNN-based segmentation theory has been applied to plant disease severity estimation and other related research in agriculture. The main goal of semantic segmentation applied to plant disease severity estimation is to assign appropriate labels to each pixel in order to obtain the percentage of diseased areas required for disease severity estimation.

Typically, the architecture of semantic segmentation is divided into two parts: the encoder network and the decoder network. The encoder is typically based on CNN networks to generate low-resolution image representations or feature maps that are mapped to pixel-level images and then perform prediction and segmentation. The differences between the different semantic segmentation models are often reflected in the decoder networks70. The first successful application of deep learning to semantic segmentation was achieved by a fully convolutional network (FCN) constructed by Long et al.38. After that, a number of variants of semantic segmentation emerged, such as U-Net, SegNet, DeepLab, and so on.

4 of the 16 articles used the CNN-based semantic segmentation network for plant disease severity assessment. Chen et al.71 proposed a BLSNet for estimating the severity of rice bacterial leaf streak (BLS). BLSNet was based on U-Net with the addition of an attention mechanism and multi-scale extraction to improve the accuracy of lesion segmentation. Compared with DeepLabv3+ and U-Net, the experimental results suggested that BLSNet was more suitable for adapting to scale changes of images, and the prediction time of BLSNet was slightly longer than U-Net, but shorter than DeepLabV3+ . Gao et al.72 proposed a SegNet-based network to segment potato late blight (PLB) lesions for quantification of the PLB severity. Goncalves et al.73 conducted comparative experiments on six semantic segmentation networks (U-net, SegNet, PSPNet, FPN, and 2 variants of DeepLabv3+) applied to three types of plant disease severity estimation (coffee leaf miner, soybean rust, and wheat tan spot).

Although CNNs have provided good results in assessing the severity of plant diseases, the CNN-based semantic segmentation network also has its advantages. The achievement of CNN models for plant disease severity assessment is to directly establish a relationship between severity and samples that is applicable to certain plant diseases, but may not be appropriate for others. For other diseases, the model needs to be retrained. The CNN-based semantic segmentation network is a good solution to this problem by obtaining the percentage of diseased leaf area to reflect the severity through pixel-level segmentation. Previous studies have demonstrated the feasibility of a CNN-based semantic segmentation network for plant disease severity assessment.

**The research from 2022**

In the research of 2022, the methods of using CNN to evaluate the severity of plant diseases can be roughly divided into two categories. One is based on segmentation, and the other is based on improving the CNN, specifically adding the Attention Mechanism. In the segmentation evaluation method, the commonly used segmentation networks include DeepLabV3+ , U-Net, PSPNet and Mask R-CNN. For example, Zhang et al.74 used the three-stage method to classify “Huangguan” pears. In the first stage, Mask R-CNN was used to segment “Huangguan” pears from complex backgrounds, and in the second stage, DeepLabV3+ , U-Net and PSPNet were used to segment the “Huangguan” pear spot, and the ratio of the spot area to the pixel area of the “Huangguan” pear was calculated, which was divided into three levels. In the third stage, ResNet-50, VGG-16 and MobileNetV3 were used to obtain the grade of “Huangguan” pear. Liu et al.75 also used the idea of stage segmentation. Apple leaves were first segmented from the complex background using the deep learning algorithm, then the disease area of the segmented leaves was identified, and the ratio of the disease area to the leaf area was calculated to evaluate the severity of the disease. Instance segmentation can effectively separate the target from the complex background, which is conducive to dealing with the real environment. In the other method, Attention Mechanism has attracted people's attention. Yin et al.76 improved the DCNN based on the addition of multi-scale and attention mechanism, and realized the classification of maize small leaf spot. Liu et al.77 introduced a multi-scale convolution kernel and coordinate attention mechanism in SqueezeNext78 to estimate disease severity, which was 3.02% higher than the original SqueezeNext model.

**V. Datasets and evaluation metrics**

**Plant disease severity datasets**

The correct construction and rational use of plant disease severity datasets is a prerequisite and basis for severity assessment work. Unlike ImageNet, PlantVillage, and COCO in computer vision, there are no large unified datasets for plant disease severity. Plant disease severity datasets can be collected by taking one's own photographs and annotating the images, or by using public datasets and then annotating the images and citing other people's annotated images. With the development and popularity of electronic devices, image collection is typically done through cameras and smartphones. PlantVillage is a common public dataset used in plant disease severity, and common image annotation software is LabelMe, LabelImg, etc. This section provides links to datasets and annotation software used in the 16 studies, as shown in Table 1.

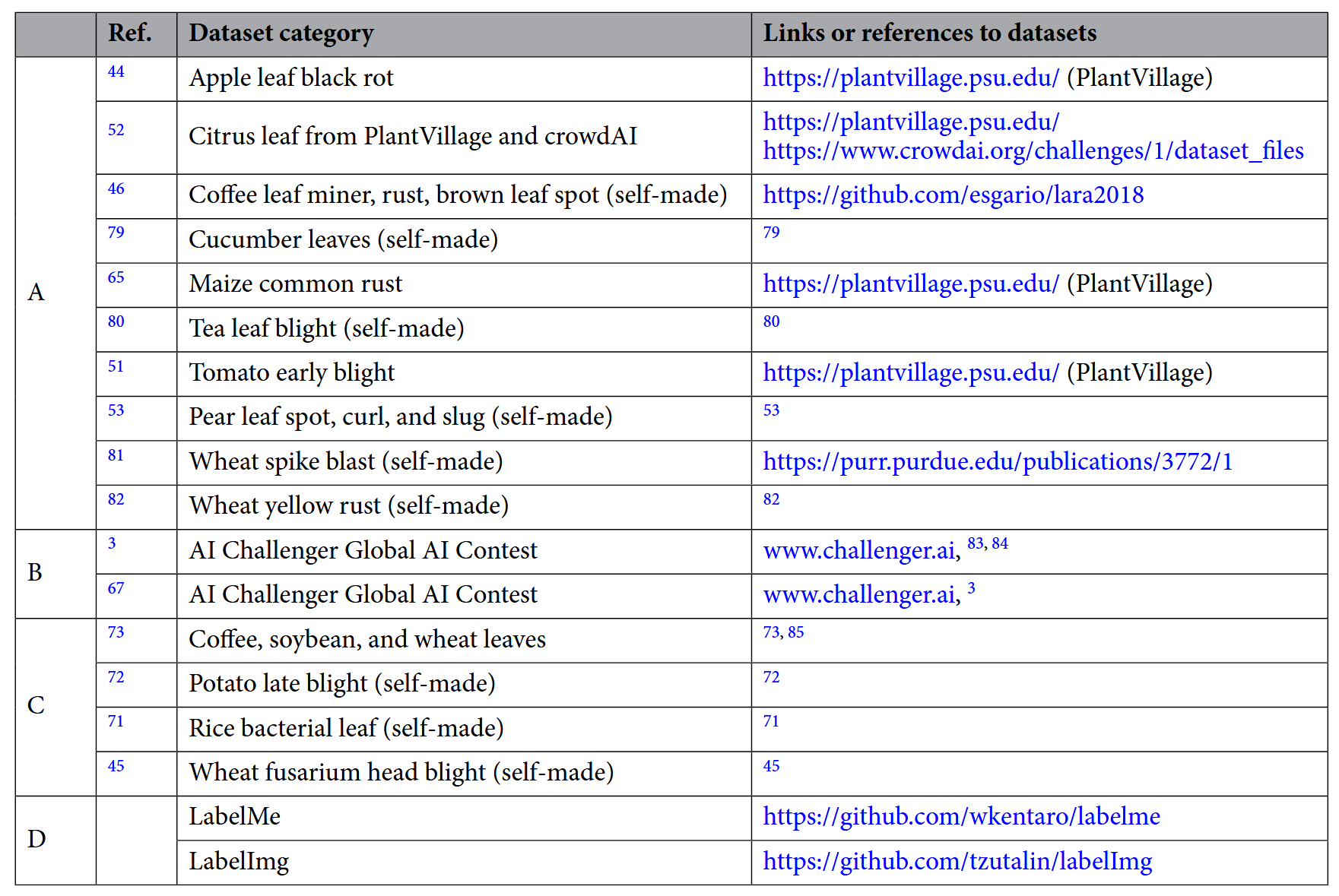
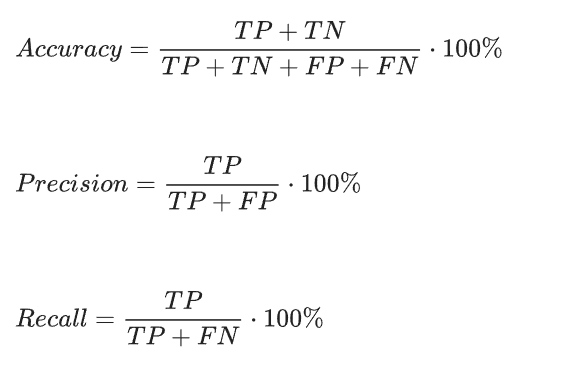


Table 1 Datasets and labeling software in the 16 studies, A for Classical CNN framework, B for Improved CNN architecture, C for CNN-based semantic segmentation network, and D for Labeling Software

**Evaluation metrics**

The common evaluation metrics mentioned in the previous model performance evaluation include accuracy, precision, recall, mean average precision (mAP), and F1 score based on precision and recall. Their specific definitions of them are described separately below.

Accuracy, Precision and Recall are expressed by the following equations:



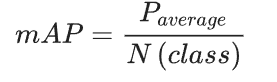
In Eqs. (1) and (2), the true positive (TP), with a predicted value of 1 and an actual value of 1, indicates the number of correctly identified lesions. The false positive (FP), with a predicted value of 1 and an actual value of 0, indicates the number of misidentified lesions. The false negative (FN), with a predicted value of 0 and an actual value of 1, indicates the number of lesions not identified. The true negative (TN), with a predicted value of 0 and an actual value of 0, indicates the number of correctly identified non-lesions.

First, it is necessary to calculate the average precision for each category in the data set for mAP.

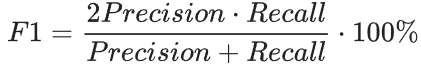


In the equation above, N is the number of all classes and j is the specific class in the dataset.

The average precision for each category is defined as follows:



The F1 score takes into account both the accuracy and recall of the model, and the equation is:



**VI. Challenges and future outlook**

One of the most important and time-consuming parts of traditional machine learning (ML) methods is the manual feature extraction, while CNN can learn features automatically. Hedjazi MA et al. addressed the task of visual identification of leaves in images by pre-training a CNN model. The experimental results showed that the pre-trained CNN model outperformed the classical machine learning methods using local binary patterns (LBPs). Bhujel A et al.87 designed and tested a semantic segmentation model based on deep learning to detect and measure gray mold on strawberry plants. The results showed that the Unet model outperformed the traditional XGBoost, K-means, and image processing technologies in detecting and quantifying gray mold. Compared with traditional image processing methods and machine learning, plant disease severity assessment has broad application prospects and great development potential, either through the classical CNN framework, improved CNN architecture, or CNN-based semantic segmentation network. Although the technology of plant disease severity assessment is developing rapidly and has gradually moved from academic research to agricultural applications, there is still a certain gap from mature applications in real natural environments, and many problems need to be solved.

**Dataset issues**

Dataset problems can be divided into two main aspects: dataset insufficiency and dataset imbalance.

(1)Dataset Insufficiency. Adequate datasets are necessary and fundamental for training the network. However, collecting and constructing datasets is an extremely time-consuming, labor-intensive, and costly process. Although there are a number of publicly available datasets for plant diseases, such as PlantVillage, ImageNet, and some publicly available self-made datasets. However, severity annotated datasets are really needed for plant disease severity research. Severity annotation of images is a more tedious process. There are two problems to face in the annotation process, one is the efficiency problem and the other is the accuracy problem. To address the time-consuming and complex annotation process that occurs in manual annotation, a possible solution is to automate the annotation with advanced software, and this automated annotation algorithm is urgently needed. In addition, semi-supervised training and auxiliary labeling methods can be used to increase the speed of agricultural sample processing and help alleviate the workload problem of manual semantic labeling. For the accuracy problem, errors are inevitable whether the annotation is done by manual visual assessment or by software, which is a challenge for future research1.

(2)Dataset imbalance. The imbalance problem can have a serious impact on the performance of the model, for example, the misclassification rate becomes higher which has been demonstrated in the experiments of many studies45, 46, 71. This problem can be well mitigated by data augmentation and weighted loss functions81.

**Complex background issues**

The dataset can be divided into two types based on image backgrounds: images with uniform backgrounds taken under controlled conditions and images with complex backgrounds taken in natural environments. CNN models are more generalized by using images taken in a natural environment for training compared to a uniform background. At the same time, complex backgrounds in images can cause other negative problems. For example, in realistic environments, ground stains resemble disease symptoms, leading to classification errors in the mode. Reflections from natural lighting can lead to misclassification of shaded healthy areas or failure to detect disease area. In addition, it is more common for multiple diseases to occur simultaneously in real-world environments. In46, the researchers mentioned that the presence of multiple diseases on a single sheet leaf can significantly change the characteristics of the symptoms, especially when the symptoms overlap, making the system more prone to misclassification. Many studies have shown that when disease symptoms are similar, their error separation rate increases significantly. Due to the problems caused by the complex background, the application of the theoretical results of CNN-based plant disease severity assessment to the actual agricultural production process faces serious obstacles. To solve some of the problems caused by the complex environment, the images can be pre-processed, but this increases the complexity of the whole detection process. For the problem of simultaneous identification and assessment of multiple diseases, researchers in proposed to alleviate this problem by training a similarity-based architecture that classifies symptoms that are not similar to any disease in the dataset into new classes, such as other classes. This idea has not yet been realized, and further solutions need to be brainstormed.

**Practicality issues**

In order to apply theoretical research to practical situations, various solutions have been proposed. As we all know, DCNN is an effective autonomous feature extraction model. Some researches combine deep learning and machine learning methods to build hybrid models. Usually, CNN is used as the feature extraction part and machine learning method is used as the classifier. Saberi Anari et al.88 used improved CNN for feature extraction. And multiple Support Vector Machine (SVM) model was used to improve the speed of feature recognition and processing. Kaur et al.89 used EfficientNet- B7 for feature extraction. After migration learning, they used logical regression technique to sample the collected features. Finally, the proposed variance technique was used to remove irrelevant features from the feature extraction vector. And the classification algorithm was used to classify the resulting features, and the most discriminative features are identified with the highest constant accuracy of 98.7%. By eliminating irrelevant features, the parameters of the model are greatly reduced. Vasanthan et al.89 adopted AlexNet and VGG-19 for feature extraction, and selected the best subset of features by correlation coefficient, and fed them to K-nearest neighbor, SVM, Pulse Neutron Neutron (PNN), Fuzzy Logic, Artificial Neural Network (ANN) and other classifiers. The experimental results showed that the average accuracy of this method was more than 96%.

A server with supercomputing power is needed to ensure that the plant disease severity model built by CNN in the lab is widely used. Cloud computing is essentially a shared pool of computing resources. Cloud computing gathers many computing resources and realizes automatic management through software. Not limited by time and space, anyone who uses the Internet can use the huge computing resources and data centers on the network90. PaaS cloud is a concrete implementation of cloud computing. PaaS providers provide many infrastructure and other IT services, and users can access them anywhere through web browsers. The ability to pay for use allows organizations to eliminate the capital expenditures traditionally used for local hardware and software. Lanjewar et al.91 deployed the CNN model used to evaluate tea diseases in the PaaS cloud, and the smartphone can access the hyperlink of the deployed model. The image of the tea can be captured by the smartphone camera and uploaded to the cloud. The cloud system automatically predicts the disease and displays it on the mobile display. Lanjewar M G et al.92 used the PaaS cloud platform to deploy the CNN model for Curcuma longa detection. While cloud computing brings convenience to us, it inherits the security problems shared by computers and the Internet. In particular, privacy issues, resource theft, attack, and computer viruses. These potential security problems are serious and deserve our attention.

To deploy CNN models on the cloud computing platform, the smaller the size of the model, the better. However, whether the smaller the model can achieve the same evaluation effect is a question worth discussing. Increasing the model size to a certain extent shows better feature extraction effect, such as the comparison between AlexNet and DCNN models such as VGG and GoogLeNet. However, as the model becomes deeper and larger, the degradation problem occurs. The residual structure of ResNet effectively mitigates this problem. More and more lightweight networks have been proposed. Their efficiency may not be the best, but it is worth trading a small amount of effectiveness for a large amount of efficiency. Liu et al.77 improved SquezeNext and performed comparative experiments with ReseNet-50, Xception, and MobileNet-V2. The experimental results showed that the accuracy of the proposed method was slightly better than that of Xconcept, while the model size was only 2.83 MB, which was only 3.45% of Xconcept. Model structure is a key factor to balance model size and performance.

Although CNN has shown excellent performance and great potential in assessing the severity of plant diseases, CNN also has its own limitations, such as translation invariance, pooling layer leading to information loss, and inability to obtain global features well. As a possible contribution and future work, new techniques that have become quite popular recently, such as vision transformers. The main feature of vision transformers93, 94 is the self-attention mechanism, which can capture the global information well. As far as I know, no research has applied it to severity estimation.

At present, some problems still have not found appropriate solutions, which indicates that the current research on automatic assessment of plant disease severity is far from mature and perfect practical application, which requires more scholars to continue to struggle to study the unsolved problems. The review article of our group on “Recent Advances in Plant Disease Severity Assessment Using Convolutional Neural Networks” provides some references for related types of research work. And more importantly, it can provide new ideas for the subsequent research work.

译文2

作者：Tingting Shi, Yongmin Liu, Xinying Zheng, Kui Hu, Hao Huang, Hanlin Liu,Hongxu Huang

出处：Scientific Reports, 2023, 13(1): 2336.

**卷积神经网络在植物病害严重程度评估中的最新进展**

**摘要**：在现代农业生产中，病害严重程度是直接影响植物产量和品质的重要因素。为了对植物的整个生产过程进行有效的监测和控制，不仅要明确病害的种类，还要明确病害的严重程度。近年来，深度学习在植物病害种类识别方面的应用越来越广泛，尤其是卷积神经网络（CNN）在植物病害图像中的应用取得了突破性进展，但对病害严重程度评估的研究相对较少。该研究首先追溯了现有病害研究者的主流观点，为植物病害严重程度的分级提供标准；然后根据网络架构，从经典CNN框架、改进的CNN架构和基于CNN的分割网络三个方面概括了16篇基于CNN的植物病害严重程度评估的研究，并对各自的优缺点进行了详细的比较分析，并研究了数据集获取的常用方法和CNN模型的性能评价指标。最后，本研究讨论了基于CNN的植物病害严重程度评估方法在实际应用中面临的主要挑战，并提出了可行的研究思路和可能的解决方案。

**I. 引言**

植物病害是造成农业损失的主要原因，这些病害是由各种损害植物生长的生物引起的，例如害虫、细菌或真菌。可靠而准确的病害严重程度评估方法对于有效控制病害和最大限度减少产量损失至关重要。评估植物病害严重程度的方法有很多种。传统的确定病害严重程度的方法是目测评估，但由于病害的相似性和特征的多样性，容易受到外界因素和主观个体差异的影响，这种方法非常不可靠。目测评估通常需要由经验丰富的专家进行，效率不高，而许多农民无法接触到专家，因此很难准确及时地识别病害严重程度。此外，高光谱成像也被用于测量植物病害的严重程度，但这种技术需要传感器等精密设备和一定的专业知识，成本高昂且效率低下。

近年来，随着计算机图像技术的快速发展和相关电子设备硬件性能的不断提升，计算机视觉与人工智能在植物种类分类、叶片病害识别、植物病害严重程度评估等农业诊断领域得到了广泛的应用。深度学习目前在计算机视觉领域取得了重大突破，而CNN在植物病害检测应用中更是表现出了优异的性能。与传统方法相比，CNN能够自动直接从输入图像中提取特征，无需复杂的图像预处理，实现端到端的检测方法。目前，利用CNN进行植物病害种类识别已经取得了令人满意的效果，但在病害严重程度评估方面的研究较少。本研究重点关注CNN在植物病害严重程度评估中的应用，并系统地综述相关研究，为进一步的研究工作提供参考思路。

本综述其余部分安排如下：第二部分概述了与植物病害严重程度视觉评估相关的概念。第三部分回顾了CNN的发展历史。第四部分讨论CNN在植物病害严重程度上的具体应用，说明单任务和多任务系统之间的差异。并从经典CNN框架、改进的CNN架构和基于CNN的语义分割网络三个方面重点介绍基于CNN的植物病害严重程度评估方法的基本工作原理，并分析每种方法的优缺点。第五部分总结了相关的公开数据集并提出了CNN性能评估指标。第六部分讨论了基于CNN的植物病害严重程度评估在实际应用中可能面临的主要挑战，并针对这些挑战提供了可行的研究思路和可能的解决方案。

**II. 视觉评估**

**植物病害严重程度的定义**

植物病害严重程度定义为可见病害症状的植物单位与植物总单位（例如叶子）的比例，是许多病害的重要定量指标。及时准确地评估病害严重程度对作物生产至关重要，因为病害严重程度直接影响作物产量，并且通常用作预测指标，以极高的准确度估计作物损失。例如，严重程度指标可用作决策阈值或病害预测，以帮助种植者合理化病害控制，例如决定农药的剂量和类型以及喷洒时间。

**植物病害严重程度的视觉评估方法**

准确测量和评估病害严重程度对农业生产至关重要。因为它可以确保正确分析治疗效果、准确了解产量损失与病害严重程度之间的相关性以及合理评估植物生长阶段。不准确或不可靠的病害评估会导致错误的结论，从而导致错误的病害管理措施，从而进一步加剧损失。病害严重程度的评估通常使用各种尺度，包括名义（描述性）尺度、序数评级尺度、区间（类别）尺度和比率尺度。以下概述了这些用于直观评估病害严重程度的尺度，包括定性和定量。

定性尺度。

（1）描述性量表：这是疾病严重程度分级量表中最简单、最主观的标准之一。疾病被分为几类，并用轻度、中度、重度等描述性术语来表示。由于主观性和缺乏量化定义，该量表的价值非常有限，除非在特定情况下进行评级。

（2）定性序数标度：这仍然是描述性病害标度，但病害严重程度的类别比描述性标度更加多样化。例如，徐等人使用0-5的标度来描述西葫芦黄花叶病毒和西瓜花叶病毒的症状严重程度，以表示病害严重程度的增加。该标度被广泛用于某些病害，尤其是用于评估症状不易量化的病毒性疾病。

定量尺度。

（1）定量序数标度：这种标度由已知类别中的数字组成，通常是有症状区域所占的百分比。它可以进一步分为两种类型：等间隔和不等间隔。然而，等间隔评定量度可能会给出更高的平均严重程度，特别是当实际严重程度处于某一类别的低端时，因为间隔太宽，难以显示差异，导致评级不准确。有些病害评定量度的间隔是不等的。Horsfall-Barratt标度（H-B标度）是一种广泛使用的不等间隔标度。它是由Horsfall和Barratt开发的，有效地缓解了等间隔的问题。例如，Bock等人用该标度估计柑橘溃疡病的严重程度。Forbes等人用HB标度估计田间马铃薯晚疫病的严重程度等。

（2）比例量表：该量表广泛用于严重程度的直观评估。评分者测量症状器官的百分比，定义为0%至100%，并据此评估严重程度。因此，比例量表对评分者的要求更高，需要更准确地识别和测量实际疾病。

虽然植物病害严重程度可以通过多种不同的方法进行评估，但定性和定量评估方法往往会导致评估结果与现实不符，这是由于评估人员的主观性、在病害严重程度较低时倾向于高估以及评估人员倾向于使用5%的整数间隔等因素造成的。为了提高评估人员估计的准确性，标准面积图(SAD)长期以来一直被用作帮助评估植物病害严重程度的工具。对评估人员进行专业培训也可以有效提高评估的准确性。

**CNN的发展历史**

深度学习始于1943年阈值逻辑的提出，本质上是建立与人类神经网络极为相似的计算机模型的过程。CNN是深度学习的一个子集，出现于20世纪80年代，最初发展了感受野的概念，随后被引入到CNN研究中，后来随着BP算法的引入和多层感知器的训练，研究者们尝试自动提取特征，而不是手工设计特征。LeCun et al.提出了一种使用BP网络的CNN架构“LeNet-5”，在当时标准的手写数字识别任务上，其表现优于所有其他技术。由于传统BP神经网络随着网络层数的增加，出现了局部最优、过拟合、梯度消失等问题，以及当时一些浅层机器模型的提出，深度神经网络模型的研究一度被搁置。直到2006年左右，Hinton等人发现具有多个隐藏层的人工神经网络具有出色的特征学习能力，Glorot等人通过一种归一化方法缓解了训练过程中梯度消失的问题。人们的注意力又转移到深度学习上。2012年，AlexNet赢得了ImageNet大规模视觉识别挑战赛（ILSVRC），自此DL受到了越来越多研究者的关注，AlexNet被认为是深度学习领域的重大突破。接下来，CNN架构不断发展，涌现出许多性能优异的算法。主要的经典CNN网络有LeNet、AlexNet、VGG、GoogLeNet、Resnet、DenseNet等，从LeNet到DenseNet的演进序列顺序如图1所示。

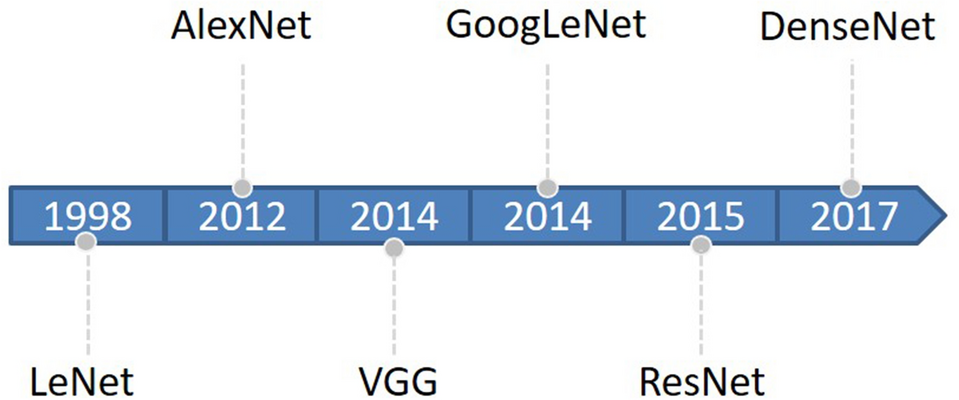


图 1 CNN从LeNet到DenseNet的演进时间线

随着CNN的发展，新的CNN模型不断涌现，实现不同的功能。例如轻量级网络：SqueezeNet，MobileNet，ShuffleNet，Xception，EfficientNet。目标检测网络：R-CNN，Fast R-CNN，Faster R-CNN，YOLO，SSD。分割网络：FCN，SegNet，U-Net，PSPNet，DeepLab，Mask RCNN等，它们表现出优异的性能和巨大的研究价值。

**IV.** **基于CNN的植物病害严重程度评估方法**

CNN在评估植物病害严重程度方面已取得巨大成功。基于CNN的植物病害严重程度自动评估首次由Wang等人于2017年提出。他们使用不同的CNN模型对四个严重程度等级的苹果黑腐病图像进行分类，在测试集上取得了90.4%的总体准确率，这表明CNN是一种很有前途的全自动植物病害严重程度分类新技术。Liang等人提出了PD2SE-Net，实现了一个用于病害严重程度评估、植物物种识别和植物病害分类的多任务系统，总体准确率分别为0.91%、0.99%和0.98%。Su等人结合ResNet-101网络和语义分割，快速预测小麦赤霉病(FHB)的严重程度，预测准确率为77.19%。

**单任务与多任务系统**

深度学习倾向于针对特定指标进行优化。换句话说，一个模型或一组模型通常被训练来执行单个目标任务，这样的系统被称为单任务系统。另一方面，还有多任务学习（MTL）的概念，如果将多个任务链接在一起，则可以同时学习。实验研究表明，在预测性能方面，同时从多个相关任务中学习特征比单独学习它们更有益。MTL可以通过并行学习任务并使用来自不同任务的更多特征来降低每个任务中过度拟合的风险，从而实现更好的模型泛化。

利用CNN进行植物病害检测的研究包括单任务系统，可单独识别植物病害种类或估计病害严重程度。例如，Prabhakar等人使用ResNet101评估番茄叶枯病的严重程度。Zeng等人训练了六种不同的CNN模型来对柑橘黄芽病的严重程度进行分类。也有多任务系统同时执行这两项任务。例如，José GM Esgario等人使用CNN实现了咖啡叶病种类的分类和严重程度分级。Fenu等人考虑了五种预训练的CNN架构作为特征提取器，对三种疾病和六种严重程度进行分类，其实验结果表明，训练后的模型在使用多任务学习模型自动提取病害叶片识别特征方面具有很强的健壮性。

**CNN在植物病害严重程度评估中的应用**

为了明确CNN用于植物病害严重程度评估的具体实现过程，本研究选取了16篇符合研究主题的高质量文章。首先，在世界最大、最全面的科学信息资源之一Web of Science平台上进行搜索。收集研究集的过程需要定义搜索词，因此在Web of Science中输入关键词“卷积神经网络”（主题）和“植物病害严重程度”（主题），截至2022年，共检索到57篇文章，出版年份如图2所示。在57篇论文中，根据研究对象（植物病害）和研究方法（CNN）选取16篇论文进行具体分析。在此基础上，对2022年的最新研究进行单独分析。根据这16篇文章采用的CNN网络架构的不同，进一步分为三类：经典CNN框架、改进的CNN架构和基于CNN的分割网络。基于CNN的植物病害严重程度评估方法流程图如图3所示。

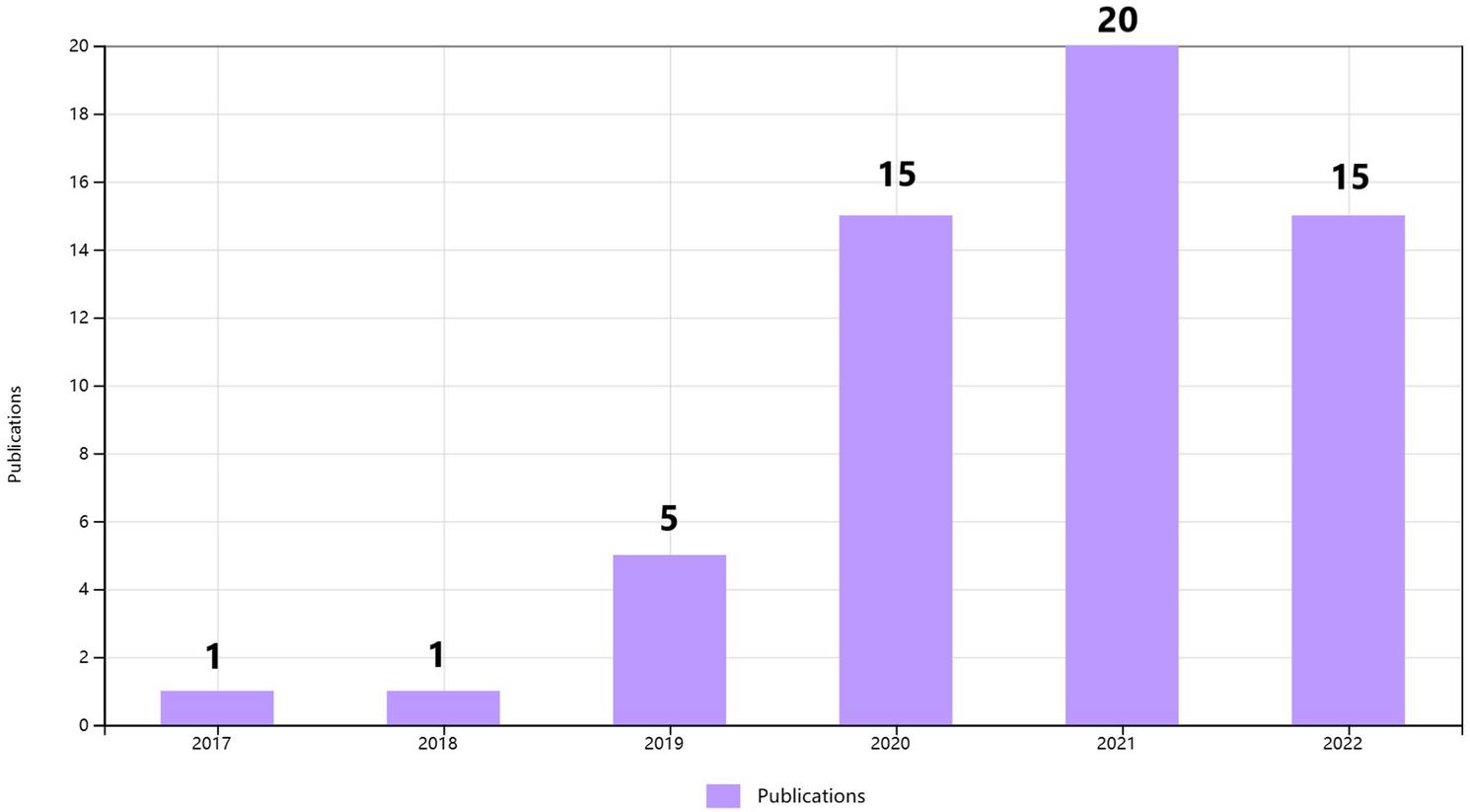


图 2以“卷积神经网络”和“植物病害严重程度”为关键词的57篇文章发表年份分布图

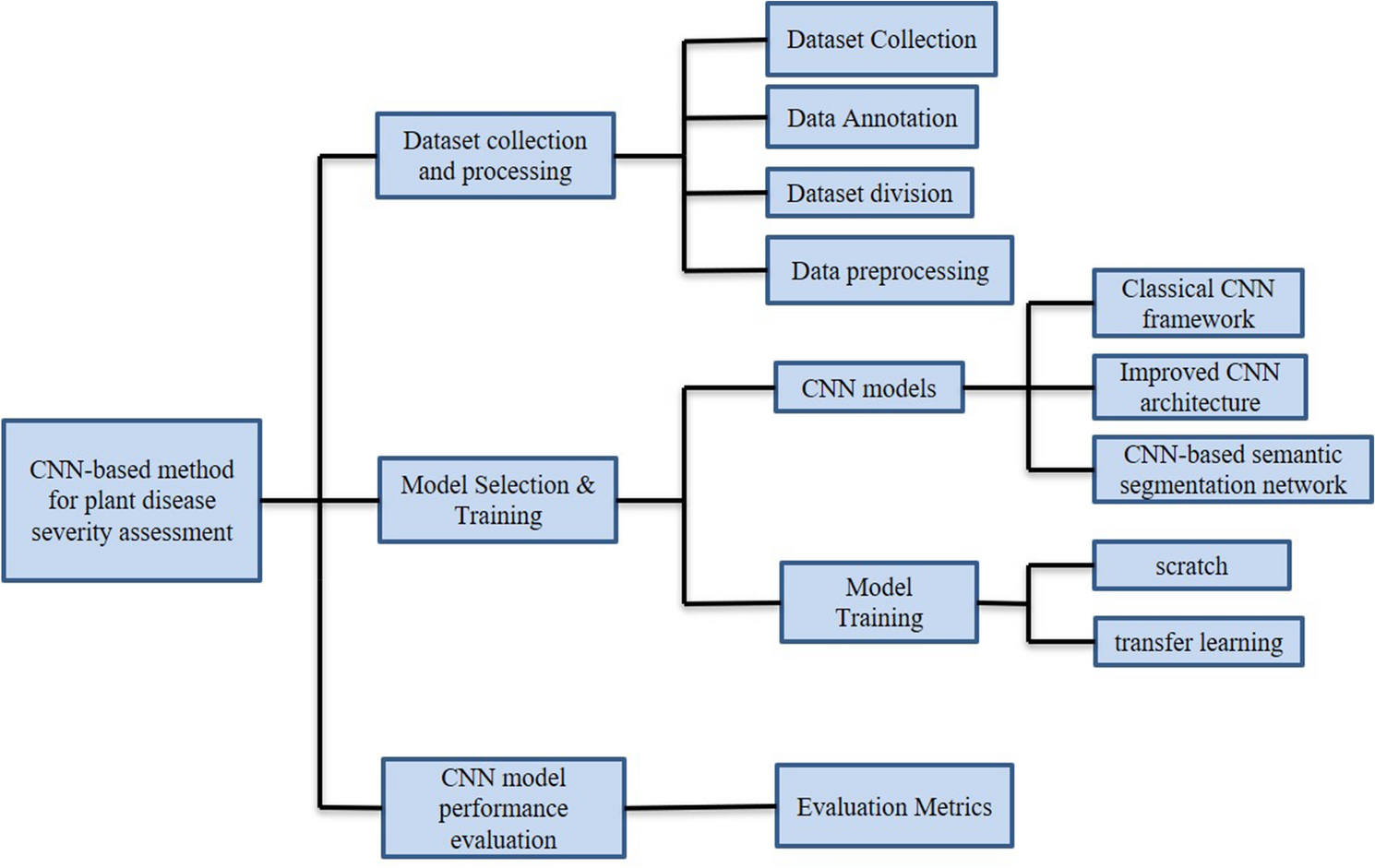


图 3 基于CNN的植物病害严重程度评估方法流程图

**经典CNN框架**

16篇文章中有10篇是基于经典CNN框架实现病害严重程度分级的，这10篇研究在具体的CNN框架和研究主题上有所不同，但都是基于CNN实现植物病害严重程度评估，因此在具体实现过程上有相似之处。基于经典CNN框架实现植物病害严重程度评估的过程可以分为以下三个主要步骤。

第一步是收集和处理数据集，从四个方面进行描述。

（1）数据集特点。10篇研究中，6篇使用自制数据集，4篇使用PlantVillage图片。自制数据集又可分为两类，一类是在受控条件下拍摄的图像，照片是在部分受控条件下从叶片背面拍摄的，背景为白色；另一类是在自然条件下拍摄的，背景复杂，而PlantVillage中的图像背景均匀、均质。自制数据集耗时长、成本高，但更符合真实环境。大量研究表明，用受控图像训练的模型预测真实环境图像时，准确率会显著降低。如果公开的数据集不能满足某项研究的需求，就必须自制数据集。

(2)数据集标注。评估严重程度的必要条件之一是记录被标注不同的严重程度等级。10篇文章中，3篇按描述性尺度标注，1篇按定性序数尺度标注，4篇按定量序数尺度标注，1篇文章未在文章中注明标注方法。例如，在46篇文章中，采用了定量序数尺度。根据患病叶片的比例将严重程度分为五个等级：健康（<0.1%）、非常低（0.1–5%）、低（5.1–10%）、高（10.1%–15%）和非常高（>15%）。

（3）数据集划分。数据集通常分为三部分：训练数据集、验证数据集和测试数据集。训练集用于训练模型，验证集用于调整超参数，测试集用于评估模型性能。这10项研究基本上都使用了70%到85%的数据集进行训练。Mohanty等人尝试了五种不同的分离比例来划分数据集，实验结果表明，对他们的数据来说，使用80%的数据集进行训练，20%的数据集进行验证是理想的。

（4）数据预处理。通常，在将图像输入CNN之前会进行两项预处理操作。一是调整图像大小以匹配输入层要求。例如，PlantVillage中的图像大小为256×256，而AlexNet输入层要求大小为227×227，则需要调整原始照片的大小。此处理在所有10项研究中均有体现。其次，对图像进行归一化，以帮助模型更快地收敛，显著提高端到端训练的效率。

第二步是模型选择和训练阶段，下面从两个方面进行描述。

（1）CNN框架选择。10项研究中使用的CNN框架包括AlexNet、VGG、GoogLeNet、ResNet、DenseNet、MobileNet、Inception、Faster R-CNN、YOLO、EfficientNet、SqueezeNet、Xception等。这些研究中绝大多数都使用了多个CNN框架进行对比实验，以确定在相同训练条件下哪种模型更能检测出某种植物疾病的严重程度。

（2）训练方法。训练CNN有两种方法，一种是从头开始，另一种是迁移学习。迁移学习是指将一个在大量图像集（如ImageNet（1000个类别，120万张图像））上训练好的网络适应不同的任务，这是通过底层CNN学习非特定特征来实现的。迁移学习有两种方法：特征提取和微调。特征提取是保持预训练模型的权重不变，然后使用它们在目标数据集上训练新的分类器的过程。微调涉及使用预训练模型中的权重初始化模型，然后在目标数据集上训练部分或全部权重。Brahimi等人使用了特征提取、微调和从头训练三种方法来训练六个CNN模型。结果表明，微调模型的准确率最高，特征提取模型的训练时间最短。10篇研究中有8篇采用了迁移学习，只有一篇是从头开始训练的。在44篇论文中，对两种训练模型的方法进行了比较，结果表明迁移学习缓解了训练数据不足的问题。

**最后一步是评估CNN模型的性能**

CNN模型的性能是通过对训练好的模型使用测试集来获得的。测试集必须独立于训练集和验证集，否则评估结果可能会有很大偏差。Mohanty等人训练了一个模型来识别14种作物和26种疾病，在测试集上的总体准确率为99.35%，其中验证集和测试集之间没有明显的区分。当他们在一组与训练图像不同条件下拍摄的图像上测试该模型时，模型的准确率急剧下降到31%。值得注意的是，10项研究中只有4项明确区分了三类数据集。Sibiya等人明确区分了训练集、验证集和测试集。实验结果表明，所提出的模型既没有过度拟合也没有欠拟合，因为该模型在验证集上的准确率为95.63%，在测试集上的准确率高达89%。

模型性能的定量评估是通过评估指标实现的。评估指标通常包括准确率、精确率、召回率、平均精确率(mAP)以及基于精确率和召回率的F1分数。随着深度学习的发展，CNN模型在不同数据集上的性能得到了提高，各种评估指标也得到了增加。很难对来自不同研究的CNN进行一致的性能比较，因为大多数基于CNN的植物病害严重程度评估研究都适用于特定数据集，其中许多数据集尚未公开，并且没有提供重现实验所需的所有参数。

**改进的CNN架构**

16篇文章中有2篇是基于改进的CNN架构进行严重程度评估的。将经典的CNN框架与改进的CNN架构进行比较，相同之处在于实现过程基本相同，不同之处在于后者采用基于经典CNN的改进网络，目的是设计出性能更高、更实用的植物病害诊断系统。

在3中，提出了一个网络PD2SE-Net，以设计一个更优秀、更实用的植物病害诊断系统。PD2SE-Net引入了ResNet50网络作为基础模型，并集成了ShuffleNet-V230的构建块。PD2SE-Net架构如图4所示。PD2SE-Net有两个关键组件使其如此有效。一是引入残差结构来构建参数共享层，这使得模型每批可以更新更多信息。受ShaResNet66的启发，ResNet50被用来构建基本框架，并与参数共享相结合，以减少网络中的冗余信息。另一个是引入了shuffle单元。ShuffleNet-V2单元用于以较低的计算复杂度提取不同植物种类和疾病的特征图。最后，PD2SE-Net实现了植物种类识别、疾病分类和严重程度估计，总体准确率分别为0.99、0.98和0.91。

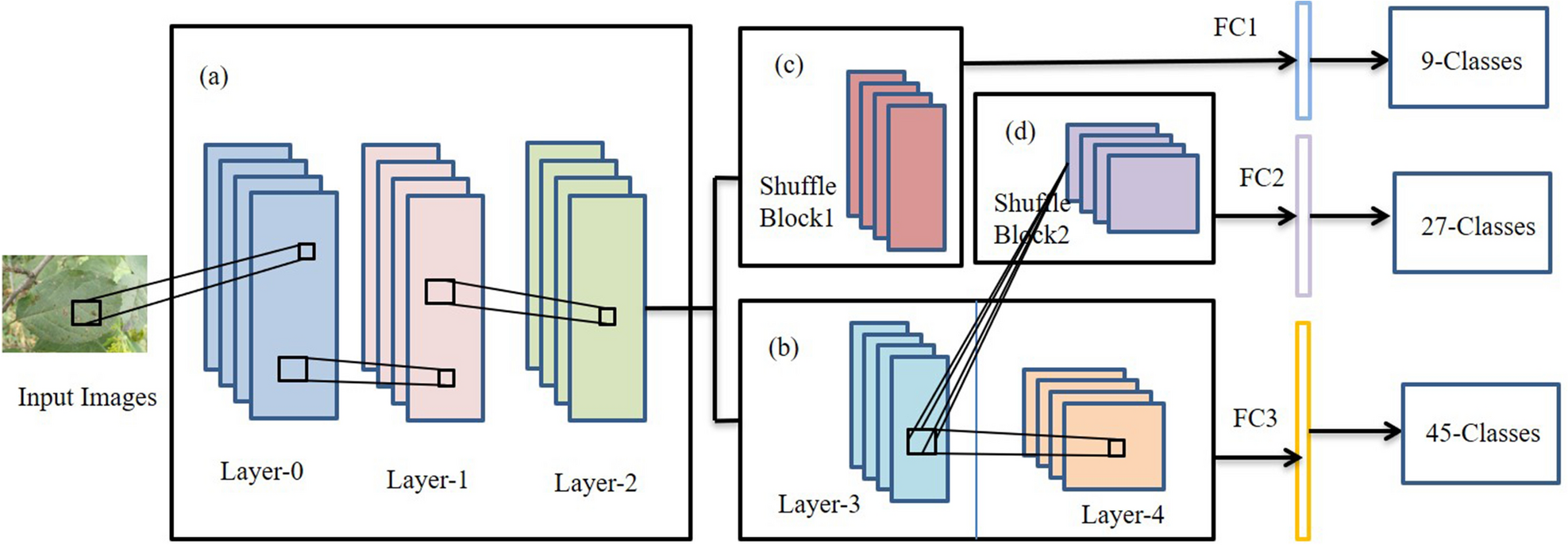


图 4 PD2SE-Net的架构。分为五部分：（a）参数共享层；（b）第三层是第四层与混洗块2之间的参数共享层，第四层是用于严重程度估计的高维特征提取器；（c）用于植物物种识别的特征提取器；（d）用于植物疾病诊断的特征提取器；（e）全连接层3

Xiang等提出了一种基于残差网络、通道重排操作和多尺寸模块的轻量级网络L-CSMS，用于植物病害严重程度评估。多尺度卷积模块使用多个不同大小的卷积核来提取不同的感受野，从而从特征图中获得鲁棒的特征和空间关系。引入通道重排操作以实现不同通道组之间的信息通信并提高精度。通道重排操作和多尺寸卷积模块以堆叠拓扑的形式集成到构建块中，如图5所示。L-CSMS使用ResNet的残差学习方法，通过堆叠相同拓扑的模块来构建深度网络。为了验证L-CSMS模型的性能，Xiang等将L-CSMS模型与ResNet、DenseNet、Inception-V4、PD2SE-Net、ShuffleNet和MobileNet进行了对比实验。结果表明，L-CSMS以更少的参数、FLOP和较好的精度取得了竞争优势。

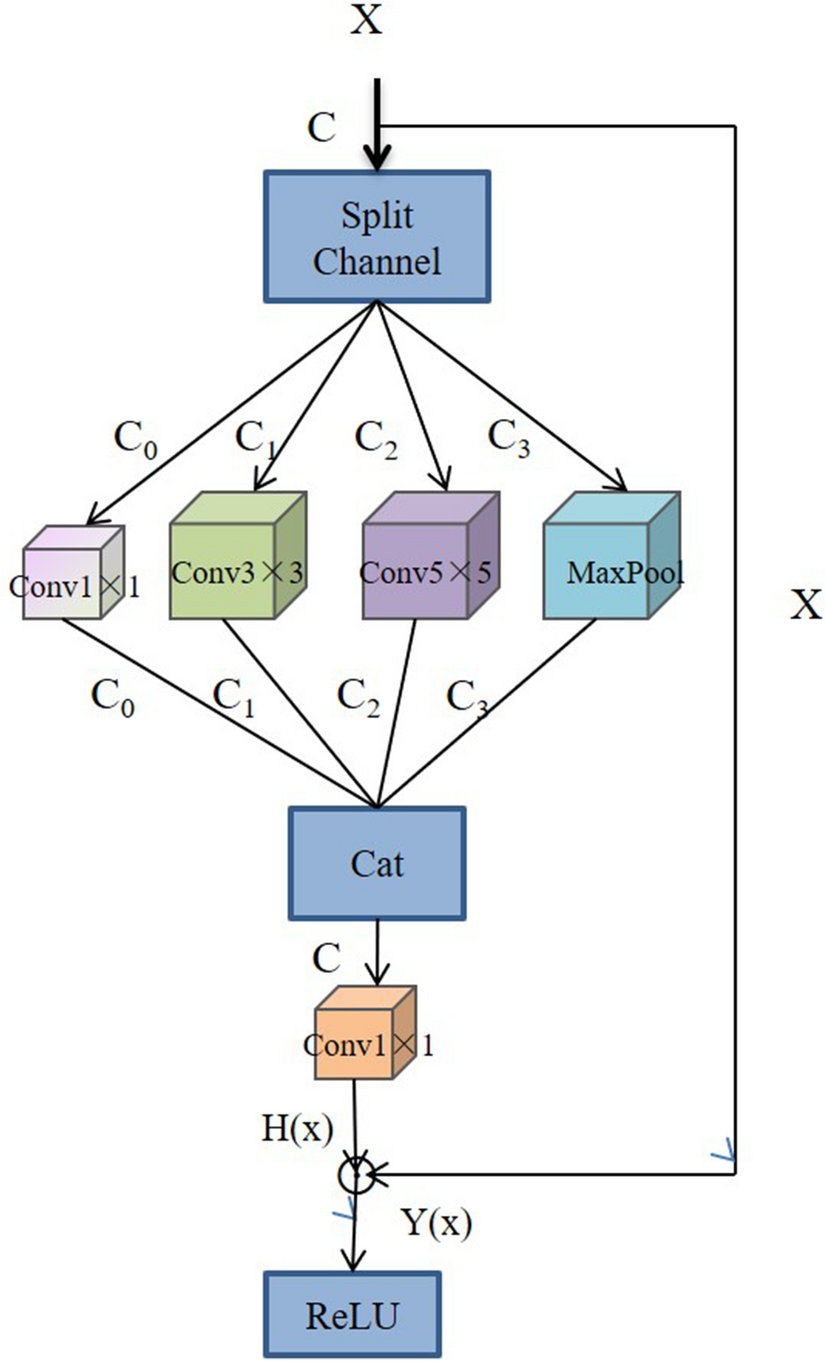


图 5 具有通道混洗操作和多尺寸卷积模块的构建块

改进CNN架构的方式各不相同，但改进的目的基本一致，即设计出更准确、更实用、泛化性能更好的植物病害严重程度评估系统。两篇研究均侧重于ResNet中的残差结构和通道改组。

**基于CNN的语义分割网络**

图像语义分割越来越受到计算机视觉和深度学习研究人员的关注，使用深度学习技术进行语义分割的研究工作也在不断发展。特别是，CNN在准确性和效率方面远远超过了其他方法。基于CNN的分割不仅提供类别信息，还提供关于这些类别的空间位置的附加信息。语义分割的任务是将每个像素标记为一种封闭对象或一类区域。基于CNN的分割理论已经应用于植物病害严重程度估计和农业中的其他相关研究。语义分割应用于植物病害严重程度估计的主要目标是给每个像素分配合适的标签，以获得病害严重程度估计所需的患病区域百分比。

通常，语义分割的架构分为两部分：编码器网络和解码器网络。编码器通常基于CNN网络，生成低分辨率图像表示或特征图，映射到像素级图像，然后进行预测和分割。不同的语义分割模型之间的差异通常体现在解码器网络上。深度学习在语义分割中的第一个成功应用是由Long等人构建的全卷积网络（FCN）。此后，出现了许多语义分割的变体，例如U-Net、SegNet、DeepLab等。

16篇文章中有4篇使用基于CNN的语义分割网络进行植物病害严重程度评估。陈等提出了一种用于评估水稻细菌性叶纹病（BLS）严重程度的BLSNet。BLSNet在U-Net的基础上加入了注意力机制和多尺度提取以提高病斑分割的准确性。与DeepLabv3+和U-Net相比，实验结果表明BLSNet更能适应图像的尺度变化，BLSNet的预测时间略长于U-Net，但短于DeepLabV3+。高等提出了一种基于SegNet的网络来分割马铃薯晚疫病（PLB）病斑，以量化PLB的严重程度。Goncalves等人对六种语义分割网络（U-net、SegNet、PSPNet、FPN和DeepLabv3+的2个变体）应用于三类植物病害严重程度估计（咖啡潜叶虫病、大豆锈病和小麦褐斑病）进行了比较实验。

虽然CNN在植物病害严重程度评估方面已经取得了不错的效果，但是基于CNN的语义分割网络也有其优势。CNN模型在植物病害严重程度评估方面的成就在于直接建立了严重程度与样本之间的关系，这种关系对某些植物病害适用，但对另一些病害可能并不适用，对于其他病害，模型需要重新训练。基于CNN的语义分割网络通过像素级分割得到病害叶面积百分比来反映严重程度，很好地解决了这一问题。前期研究已经证明了基于CNN的语义分割网络在植物病害严重程度评估方面的可行性。

**2022年的研究**

在2022年的研究中，利用CNN评估植物病害严重程度的方法大致可以分为两类，一类是基于分割的，另一类是基于改进CNN，具体来说是增加了Attention机制。在分割评估方法中，常用的分割网络有DeepLabV3+、U-Net、PSPNet和Mask R-CNN等。例如，张等采用三阶段的方法对黄冠梨进行分级，第一阶段利用Mask R-CNN从复杂背景中分割出黄冠梨，第二阶段利用DeepLabV3+、U-Net和PSPNet对黄冠梨斑点进行分割，并计算斑点面积与黄冠梨像素面积的比值，分为三个等级；第三阶段采用ResNet-50、VGG-16和MobileNetV3得到黄冠梨的等级。刘等也借鉴了阶段分割的思想，首先利用深度学习算法将苹果叶片从复杂背景中分割出来，然后对分割后的叶片进行病害面积识别，计算病害面积与叶片面积的比值来评估病害严重程度。实例分割可以有效地将目标从复杂背景中分离出来，有利于处理真实环境。在另一种方法中，注意力机制（Attention Mechanism）引起了人们的关注，尹等在加入多尺度和注意力机制的基础上对DCNN进行了改进，实现了玉米小叶斑病的分类。刘等在SqueezeNext78中引入了多尺度卷积核和坐标注意力机制来估计病害严重程度，比原SqueezeNext模型提高了3.02%。

**V. 数据集和评估指标**

**植物病害严重程度数据集**

正确构建和合理使用植物病害严重程度数据集是开展严重程度评估工作的前提和基础。与计算机视觉领域的ImageNet、PlantVillage、COCO等不同，植物病害严重程度数据集尚无统一的大型数据集。植物病害严重程度数据集可以通过自己拍照并标注图片，也可以利用公开的数据集然后标注图片并引用他人标注的图片来收集。随着电子设备的发展和普及，图像采集一般通过摄像头和智能手机进行。PlantVillage是植物病害严重程度领域常用的公开数据集，常用图像标注软件有LabelMe、LabelImg等。本节提供16篇研究中使用的数据集及标注软件链接，如表1所示。

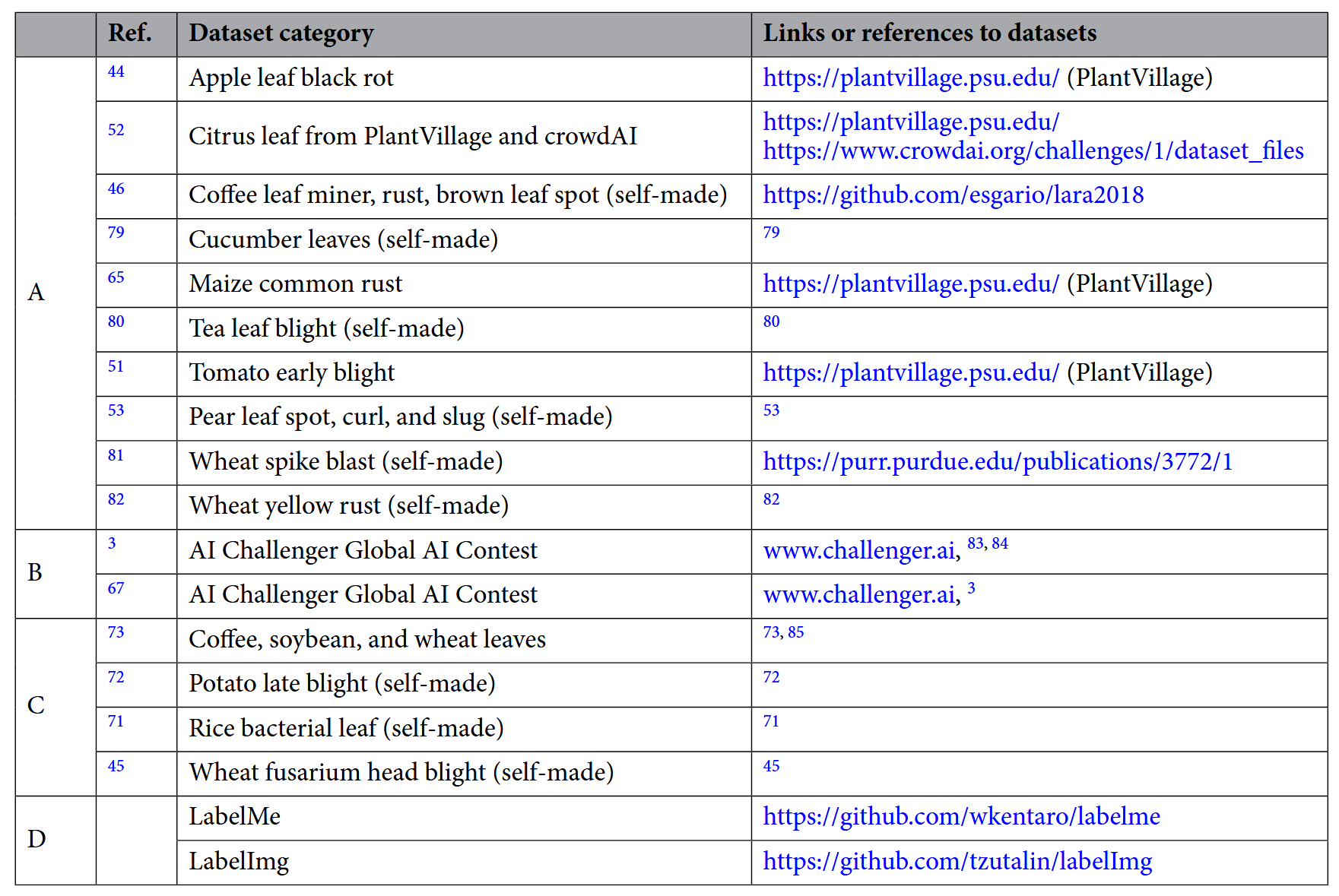
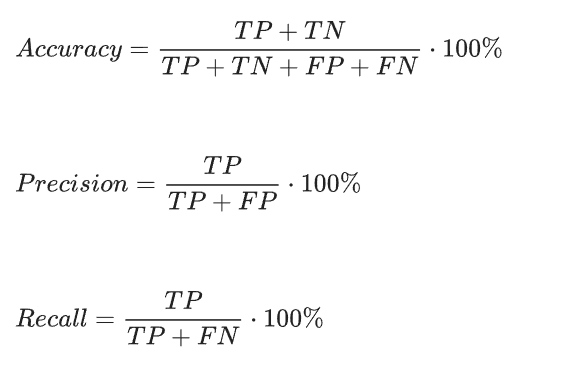


表 1 在这16项研究中使用的数据集和标注软件：A代表经典CNN框架，B代表改进型CNN架构，C代表基于CNN的语义分割网络，D代表标注软件

**评估指标**

前面模型性能评估中提到的常见评估指标有准确率、精确率、召回率、平均精确率（mAP），以及基于精确率和召回率的F1分数，下面分别介绍它们的具体定义。

准确率、精确率和召回率由以下公式表示：



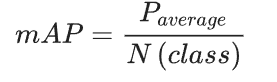
在公式（1）和公式（2）中，真实阳性（TP）表示正确识别的病变数量，预测值为1，实际值为1。假阳性（FP）表示错误识别的病变数量，预测值为1，实际值为0。假阴性（FN）表示未识别的病变数量，预测值为0，实际值为1。真实阴性（TN）表示正确识别的非病变数量，预测值为0，实际值为0。

首先需要计算数据集中每个类别的平均精度，得到mAP。

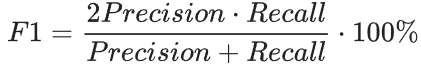


在上式中，N是所有类别的数量，j是数据集中的具体类别。

每个类别的平均精度定义如下：



F1分数同时考虑了模型的准确率和召回率，其公式为：



**VI. 挑战与未来展望**

传统机器学习(ML)方法中最重要的和最耗时的部分之一是手动特征提取，而CNN可以自动学习特征。Hedjazi MA等人通过预训练CNN模型解决了图像中叶子的视觉识别任务。实验结果表明，预训练的CNN模型优于使用局部二值模式(LBP)的传统机器学习方法。Bhujel A等人设计并测试了基于深度学习的语义分割模型，以检测和测量草莓植株上的灰霉病。结果表明，Unet模型在检测和量化灰霉病方面优于传统的XGBoost、K-means和图像处理技术。与传统图像处理方法和机器学习相比，无论是通过经典的CNN框架、改进的CNN架构还是基于CNN的语义分割网络，植物病害严重程度评估都具有广阔的应用前景和巨大的发展潜力。植物病害严重程度评估技术虽然发展迅速，并逐渐从学术研究走向农业应用，但距离现实自然环境中的成熟应用还有一定差距，还有很多问题需要解决。

**数据集问题**

数据集问题主要可以分为两个方面：数据集不足和数据集不平衡。

（1）数据集不足。充足的数据集是训练网络的必要和基础。然而，收集和构建数据集是一个极其耗时、费力且成本高昂的过程。虽然有许多关于植物病害的公开数据集，如PlantVillage、ImageNet，以及一些公开的自制数据集。然而，植物病害严重程度研究确实需要严重程度注释的数据集。图像的严重程度注释是一个更繁琐的过程。在注释过程中要面对两个问题，一个是效率问题，另一个是准确性问题。为了解决手工注释中出现的耗时和复杂的注释过程，一个可能的解决方案是使用先进的软件实现注释的自动化，这种自动注释算法是迫切需要的。此外，半监督训练和辅助标记方法可用于提高农业样品处理的速度，并有助于减轻手工语义标记的工作量问题。对于准确性问题，无论是通过人工视觉评估还是通过软件进行注释，错误都是不可避免的，这是未来研究的挑战。

(2)数据集不平衡。不平衡问题会对模型的性能产生严重影响，例如，误分类率会变得更高，这在许多研究的实验中已经得到证实。数据增强和加权损失函数81可以很好地缓解这一问题。

**复杂的背景问题**

根据图像背景，数据集可分为两类：在受控条件下拍摄的具有均匀背景的图像和在自然环境中拍摄的具有复杂背景的图像。与均匀背景相比，使用在自然环境中拍摄的图像进行训练可以使CNN模型更具泛化能力。同时，图像中的复杂背景还会带来其他负面问题。例如，在现实环境中，地面污渍与疾病症状相似，从而导致模型出现分类错误。自然光的反射会导致阴影健康区域被错误分类或无法检测到疾病区域。此外，在现实环境中，多种疾病同时发生的情况更为常见。研究人员提到，一片叶子上同时存在多种疾病会显著改变症状的特征，尤其是当症状重叠时，使系统更容易出现错误分类。众多研究表明，当病害症状相似时，它们的错误分离率会显著增加。由于复杂背景带来的问题，基于CNN的植物病害严重程度评估的理论结果应用到实际农业生产过程中面临严重的障碍。为了解决一些复杂环境带来的问题，可以对图像进行预处理，但这增加了整个检测过程的复杂性。针对多种病害同时识别评估的问题，研究人员提出通过训练一种基于相似性的架构来缓解这一问题，该架构将数据集中与任何病害都不相似的症状归为新的类，例如其他类。这个想法还没有实现，需要集思广益寻找进一步的解决方案。

**实用性问题**

为了将理论研究应用到实际情况中，人们提出了各种各样的解决方案。众所周知，DCNN是一种有效的自主特征提取模型。一些研究将深度学习和机器学习方法结合起来建立混合模型。通常，CNN用作特征提取部分，机器学习方法用作分类器。Saberi Anari等人使用改进的CNN进行特征提取。并使用多个支持向量机（SVM）模型来提高特征识别和处理的速度。Kaur等人使用EfficientNet-B7进行特征提取。经过迁移学习后，他们使用逻辑回归技术对收集到的特征进行采样。最后，使用提出的方差技术从特征提取向量中去除不相关的特征。并使用分类算法对得到的特征进行分类，以98.7％的最高恒定准确率识别出最具判别力的特征。通过消除不相关的特征，模型的参数大大减少。Vasanthan等采用AlexNet和VGG-19进行特征提取，通过相关系数选取最佳特征子集，并输入到K近邻、SVM、脉冲中子网(PNN)、模糊逻辑、人工神经网络(ANN)等分类器中，实验结果表明该方法的平均准确率在96%以上。

需要一台具有超强计算能力的服务器来确保实验室中CNN构建的植物病害严重程度模型得到广泛应用。云计算本质上是一个共享的计算资源池。云计算将众多计算资源聚集起来，通过软件实现自动化管理。不受时间和空间的限制，任何使用互联网的人都可以使用网络上庞大的计算资源和数据中心。PaaS云是云计算的具体实现。PaaS提供商提供许多基础设施和其他IT服务，用户可以通过Web浏览器在任何地方访问它们。按使用付费的能力使组织可以消除传统上用于本地硬件和软件的资本支出。Lanjewar等在PaaS云中部署了用于评估茶叶病害的CNN模型，智能手机可以访问已部署模型的超链接。茶叶的图像可以通过智能手机摄像头拍摄并上传到云端。云系统自动预测疾病并将其显示在手机显示屏上。Lanjewar MG等使用PaaS云平台部署了用于姜黄检测的CNN模型。云计算在给我们带来便利的同时，也继承了计算机和互联网所共有的安全问题，特别是隐私问题、资源窃取、攻击、计算机病毒等，这些潜在的安全问题十分严重，值得我们重视。

在云计算平台上部署CNN模型，模型尺寸越小越好，但模型越小是否能达到同样的评测效果是一个值得讨论的问题。增加模型尺寸到一定程度，特征提取效果会更好，比如AlexNet与VGG、GoogLeNet等DCNN模型的对比，但随着模型越来越深越来越大，会出现退化问题，ResNet的残差结构可以有效缓解这一问题。越来越多的轻量级网络被提出，它们的效率可能不是最好的，但值得用少量的有效性换取大量的效率。刘等对SquezeNext进行了改进，并与ReseNet-50、Xception、MobileNet-V2进行了对比实验，实验结果表明，所提方法的准确率略优于Xconcept，而模型大小只有2.83MB，仅为Xconcept的3.45%。模型结构是平衡模型大小和性能的关键因素。

尽管CNN在评估植物疾病严重程度方面表现出色且潜力巨大，但CNN也有其自身的局限性，例如平移不变性、池化层导致信息丢失以及无法很好地获取全局特征。作为可能的贡献和未来的工作，最近变得相当流行的新技术，例如视觉转换器。视觉转换器的主要特点是自注意力机制，可以很好地捕捉全局信息。据我所知，还没有研究将其应用于严重程度估计。

目前，有些问题仍然没有找到合适的解决方案，这表明目前植物病害严重程度自动评估的研究还远远没有达到成熟和完善的实际应用，需要更多的学者继续努力研究尚未解决的问题。本课题组关于“利用卷积神经网络进行植物病害严重程度评估的最新进展”的综述文章为相关类型的研究工作提供了一些参考，更重要的是可以为后续的研究工作提供新的思路。