Recent Advances in Plant Disease Severity Assessment Using Convolutional Neural Networks

Yong-min LIU
Central South University of Forestry and Technology
Ting-ting SHI
Central South University of Forestry and Technology
Xin-ying ZHENG (lym37212004@126.com)
Business School of Hunan Normal University
Kui HU
Central South University of Forestry and Technology
Hao HUANG
Central South University of Forestry and Technology
Han-lin LIU
Central South University of Forestry and Technology
Hong-xu HUANG
Central South University of Forestry and Technology

Article

Keywords: Convolutional neural network, plant diseases, severity assessment, precision agriculture

Posted Date: August 16th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1904357/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Scientific Reports on February 9th, 2023. See the published version at https://doi.org/10.1038/s41598-023-29230-7.

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 whole production process of crops, it is necessary to clarify not only the types of diseases, but also the severity of diseases. In recent years, the use of Deep Learning (DL) for plant disease species identification has been widely applied. In particular, the application of Convolutional Neural Networks (CNNs) for plant disease images has made breakthrough progress. However, little research has been done on disease severity assessment. This group first traces the mainstream views of existing disease research scholars and accordingly gives various scales for visual assessment of plant disease severity grading. According to the difference of network architecture, then, this study outlines the 16 researches on CNN-based plant disease severity assessment from three aspects of classical CNN framework, improved CNN architecture and CNN-based semantic segmentation network, and the advantages and disadvantages of each research method are compared and analyzed in detail. The common methods of datasets acquisition and the performance evaluation index of CNN models are examined in depth. Finally, this study discusses the major challenges in the practical application of CNN-based plant disease severity assessment methods, and provides feasible research ideas and possible solutions to address these challenges.

1 Introduction

Plant diseases caused by various organisms that damage plant growth, such as pests, bacteria or fungi, are one of the major causes of agricultural losses. Reliable and accurate disease severity assessment methods are essential for effective disease control and minimizing yield losses [1]. There are many methods for assessing the severity of plant diseases. Traditional method is to determine the severity by visual recognition, which affects recognition accuracy severely due to the similarity of diseases and the diversity of features susceptible to external factors, as well as subjective individual differences. Visual recognition usually has to be done by experienced specialists, which is not efficient and many farmers do not have access to specialists, making accurate and timely identification of disease severity very difficult. In addition, hyperspectral imaging has been used to measure the severity of plant diseases, but this technique requires sophisticated instruments such as sensors, as well as a certain level of expertise, so the cost of this method is high and the efficiency is low [2].

In recent years, with the rapid development of computer image technology and the continuous improvement of 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 [3]. Now deep learning has made significant breakthroughs in the field of computer vision, especially CNN has shown excellent performance in the application of plant disease detection. Compared to traditional methods, CNN is able to automatically and directly extract features from the input image, avoiding complex pre-processing of the image, thus enabling end-to-end detection methods [4]. Currently, satisfying results have been achieved in plant disease species identification using CNNs, but little research has been conducted 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 the next research work.

The remainder of this survey 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 CNN development. The fourth part addresses 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 analyzes 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.

2 Visual Assessments

2.1 Definition of plant disease severity

The severity of plant disease, defined as the proportion of plant units exhibiting visible disease symptoms to the entire plant unit (e.g. leaves), is a key quantitative indicator of many diseases [5].Timely and accurate assessment of plant disease severity is critical in crop production, because plant disease severity directly affects crop yield and is commonly used as a predictor for estimating crop losses with excellent effect [6]. Thus, severity indicator can be used as decision thresholds or for disease forecasting to help farmers rationalize disease control, such as deciding on the dose and variety of pesticide and the time of day to spray.

2.2 Visual assessment methods for plant disease severity

Accurate measurement and correct assessment of disease severity is essential for 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 rating of crop growth stages [7]. Inaccurate or unreliable disease assessment can lead to erroneous conclusions, resulting in the wrong measures being taken in the disease management process, which in turn can further exacerbate losses. Assessment of disease severity is usually done by various scales, including nominal (descriptive) scales, ordinal rating scale, interval (category) scales, and ratio scales [1]. The following will outline these scales for visual assessment of disease severity, both qualitatively and quantitatively.

Qualitative scales. (1) Descriptive Scale: This is one of the simplest and most subjective criteria in the disease severity grading scales. The disease is classified into several categories with descriptive terms such as "slight", "moderate" or "severe". Due to the subjectivity and the lack of quantitative definitions, the value of this scale is very limited, except for ratings in a certain condition. (2) Qualitative Ordinal Scale:

This is still the descriptive disease scale, which provides more diversity in the categories of disease severity levels than the Descriptive Scale. For example, Xu et al. [8] assigned a scale of 0–5 for zucchini yellow mosaic virus and watermelon mosaic virus severity symptom descriptions to indicate the increasing severity of the disease. This scale is widely used for specific diseases, especially for rating viral diseases with symptoms that are not easily quantifiable [1, 5].

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: equally interval and unequally interval. However, for equally interval rating scales there may be a higher average severity, especially when the actual disease severity is at the lower end of a category, as the interval is so wide that it is difficult to show differences, leading to inaccurate assessment [9]. Some disease assessment scales have unequal intervals. Horsfall-Barratt scale (H-B scale) is a widely used scale with unequal intervals, as shown in Table 1. It was developed by Horsfall and Barratt (1945), which is effective in alleviating the problem of equal intervals [10]. For example, Bock et al. [11] used the scale for citrus ulcer disease severity estimation; Forbes et al. [12] used the H-B scale to assess potato late blight severity in the field, etc. Interval scales can be analyzed directly, but they are usually converted to disease indices (DSI), providing a continuous variable for parametric statistical analysis.

Table 1
Horsfall and Barratt interval scale [10]

H-B category	Disease severity range	Midpoint	Interval size
1	0	0	0
2	0^+-3	2.34	3
3	3^+-6	4.69	3
4	6^+-12	9.38	6
5	12^+-25	18.75	13
6	25^+-50	37.50	25
7	50^+-75	62.50	25
8	75^+-87	81.25	13
9	87^+-94	90.62	6
10	94^+-97	95.31	3
11	97^+-100	97.66	3
12	100	100	0

(2) Ratio Scales: This scale is widely used for visual estimation of severity. The rater measures the percentage of disease symptomatic organs, which is defined from 0–100%, and rates the severity accordingly. Therefore, Ratio Scales places greater demands on the rater, who needs to more accurately identify and measure the actual disease.

Although plant disease severity can be assessed by several different methods, both qualitative and quantitative assessment methods tend to cause the assessment to be 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 towards 5% whole number intervals [13]. To improve the accuracy of raters' estimates, the standard area map (SAD) has long been used as a tool to help estimate plant disease severity [14, 15]. Professional training of raters can also be effective in improving the accuracy of the assessment.

3 Cnn Development History And Principles

The Deep Learning (DL) began with the introduction of threshold logic in 1943 and is essentially a process of building computer models that closely resemble human neural networks [16]. CNN is a subcategory of DL, that first appeared in the 1980s [17]. At the beginning, the concept of receptive field was come up and later introduced into CNN research work [18]. Subsequently, with the introduction of the BackPropagation (BP) algorithm and the training of multi-layer perceptron, researchers tried to automatically extract features instead of manual design of features [19]. LeCun et al. proposed a CNN architecture called "LeNet-5" using BP networks, which showed better performance than all other techniques in a standard handwritten digit recognition task at that time [20]. Research on deep neural network models was put on hold due to a series of problems encountered with traditional BP neural networks such as local optima, overfitting and gradient disappearance as the number of network layers increases, as well as the accompanying proposal of some shallow machine models at that time[19]. Until around 2006, Hinton et al. [21] found that artificial neural networks with multiple hidden layers have excellent feature learning capabilities. Glorot et al. [22] alleviated the problem of gradient disappearance during training by a normalization method. Attention was shifting back to deep learning. In 2012, AlexNet [23] 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, CNN architecture continues to evolve and many algorithms with excellent performance emerge. The main classical CNN networks are LeNet, AlexNet, VGG [24], GoogLeNet [25], Resnet [26], DenseNet [27] and so on. The temporal development order from LeNet to DenseNet is shown in Fig. 1.

With the development of CNN, new CNN models are constantly emerging and implement different functions. Such as lightweight networks: SqueezeNet[28], MobileNet [29], ShuffleNet [30], Xception[31], EfficientNet [32]; target detection networks: R-CNN [33], Fast R-CNN [34], Faster R-CNN [35], YOLO [36], SSD [37]; semantic segmentation networks: FCN [38], SegNet [39], U-Net [40], PSPNet [41], DeepLab [42], Mask RCNN [43], etc., they display excellent performance and great research value.

4 Cnn-based Plant Disease Severity Assessment Method

CNN has been used to assess the severity of plant diseases with significant success. Automatic estimation of plant disease severity based on CNN was first proposed by Wang et al. in 2017. They used diverse CNN models to classify apple black rot images with four severity stages, achieving an overall accuracy of 90.4% on the test set, which suggests that CNN is a promising new technique in fully automated plant disease severity classification [44]. Liang et al. [3] proposed PD^{2}SE-Net to implement a multi-task 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 for rapid prediction of the severity of Fusarium head blight (FHB) in wheat with a prediction accuracy of 77.19%.

4.1 Single-task versus multi-task systems

In DL, the focus is usually on optimization for specific metrics. In other words, a model or a set of models is often trained to perform the single target task [46], and such systems are known as single-task systems. On the other hand, there is the concept of multi-task learning (MTL), where multiple tasks can be learned simultaneously when they are associated to each other [47]. Experimental studies have revealed 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, which leads to better generalization of the model [48–50].

Among the studies using CNN for plant disease detection, there are 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 early blight of tomato leaves [51]; Zeng et al. trained six different CNN models for severity rating of citrus yellow shoot [52]. There are also multitasking systems that perform both tasks simultaneously. For instance, José G.M. Esgario et al. used CNN to implement classification of coffee leaf disease species and grading of severity [46]. Fenu et al. considered five pre-trained CNN architectures as feature extractors for the classification of three diseases and six levels of severity, whose experimental results show that the trained model is robust in automatically extracting disease leaf identification features using a multi-task learning model [53].

4.2 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. The selection process and basis are as follows: firstly, a search was conducted on the Web of Science platform which is one of the world's largest comprehensive academic information resource platforms [54]. According to [55], the process of collecting research sets requires the definition of search terms, therefore the keywords of “Convolutional neural network" (Topic) and "plant disease severity" (Topic) were entered in the Web of Science, and 44 articles were retrieved with the year of publication shown in Fig. 2. From the graph of publication years, we can observe that the application of CNN in plant disease severity assessment has gradually become a research hotspot in the past two years, but the number of related articles is not much, and thus there is much room for research and development. These 44 articles were screened and 16 articles were finally identified. A synthesis description of these 16 articles is shown in Appendix A. According to the different CNN network architectures used in these 16 articles, they are further subdivided into three categories: classical CNN framework, improved CNN architecture, and CNN-based semantic segmentation network. The flowchart of CNN-based method for plant disease severity assessment is shown in Fig. 3.

4.2.1 Classical CNN framework

A total of 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 subjects, and are similar in that they are CNN-based approach to assess the severity of plant diseases. 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 major steps.

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

(1) Dataset characteristics. Among the 10 studies, 6 studies used self-made datasets and 4 studies used the images from PlantVillage. The self-made datasets can be further divided into two types. One is for images taken under controlled conditions. For example, in[46], 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 by controlled images are used to predict images collected from real-world environments, their accuracies are significantly reduced [56–58]. When a public dataset does not meet the needs of a particular study, self-made datasets must be produced.

(2) Dataset annotation. For the assessment of severity, one of the necessary conditions is to have datasets labeled with different severity levels. Among the 10 articles, 3 were labeled according to Descriptive Scale, 1 according to Qualitative Ordinal Scale, 4 according to Quantitative Ordinal Scale, 1 article did not indicate the labeling method in the article. For example, in [46], Quantitative Ordinal Scale was used. The severity level 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 for the training of the model, the validation set is used to tune the hyperparameters, and the evaluation of the model performance is done by the test set [59]. The 10

studies all basically used 70–85% of the dataset for training. Mohanty et al. 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 [60].

(4) Data preprocessing. Two pre-processing operations are usually performed before the images are fed into CNN. One is to resize the images to match the requirements of the input layer. 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 of the 10 studies. Second, the images are normalized to help the model converge faster and significantly improve the efficiency of end-to-end training [61].

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

(1) CNN framework selection. The CNN frameworks used in the 10 studies are AlexNet, VGG, GoogLeNet, ResNet, DenseNet, MobileNet, Inception, Faster R-CNN, YOLO, EfficientNet, SqueezeNet, Xception, etc. The vast majority of these studies used multiple CNN frameworks for comparative experiments, with the aim of finding out which model is better suited to detect the severity of a particular plant disease under the same training conditions[52].

(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 that a pre-trained network on a large set of images, such as ImageNet (1.2 million images in 1000 classes), is adapted to another task, which is implemented by the underlying network of CNN learning non-specific features [62]. There are two approaches to transfer learning: feature extraction and fine-tuning. Feature extraction is to keep the weights of a pre-trained model unchanged and then using them to train a new classifier on the target dataset. Fine-tuning is to initialize the model using the weights from a pre-trained model and then train some or all of the weights on the target dataset [63]. Brahimi et al. used three approaches of feature extraction, fine-tuning and training from scratch to train six CNN models. And the results suggested that fine-tuning models had the highest accuracy and feature extraction models had the shortest training time [64]. 8 of the 10 studies used transfer learning and only one trained from scratch. In [44], the two methods of training models are compared and the results showed that transfer learning alleviated the problem of insufficient training data.

The final step is to evaluate the performance of 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 might be biased high. Mohanty et al. trained a model to identify 14 crops and 26 diseases with overall accuracy of 99.35% on 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 conditions different from the training images, the accuracy of the model decreased dramatically to 31% [60]. It is worth noting that only 4 of the 10 studies explicitly delineate three types of datasets. In [65], Sibiya et al. explicitly divided training set, validation set, and test set. 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-test accuracy of 89%.

Quantitative assessment of model performance is achieved by 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 DL, the performance of CNN models on different datasets has been improved, and various evaluation metrics have been increased. A uniform performance comparison of CNNs from different studies is difficult to achieve since 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 parameters needed to reproduce experiments.

The specific statistics of the dataset type, data annotation methods, dataset division and CNN models in the 10 articles are shown in Table 2.

Table 2
The specific statistics in the 10 articles

References	Dataset type	Data annotation methods	Dataset division	CNN models
[51]	PlantVillage	Descriptive Scale	70: 30	AlexNet VGG16 VGG19 GoogLeNet ResNet50 and ResNet101
[46]	Homemade (in the background of the whiteboard)	Quantitative Ordinal Scale	70: 15:15	AlexNet VGG16 GoogLeNet ResNet50 and MobileNetV2
[44]	PlantVillage	Descriptive Scale	80: 20	VGG16 VGG19 ResNet50 and Inception-V3
[66]	Homemade	Not specified	80: 20	NAS-FPN-ResNet101 VGG16 YOLO V3 Faster R-CNN and Faster R-CNN + FPN
[53]	Homemade	Quantitative Ordinal Scale	70: 20:10	VGG-16 VGG-19 ResNet50 InceptionV3 MobileNetV2 and EfficientNetB0
[52]	PlantVillage	Quantitative Ordinal Scale	80: 20	AlexNet VGG-13 ResNet34 DenseNet169 SqueezeNet1-1 and Inception V3
[67]	Homemade	Ratio Scales	80: 20	ResNet101
[68]	Homemade	Qualitative Ordinal Scale	85: 15	Faster R-CNN
[69]	Homemade	Quantitative Ordinal Scale	70: 10:20	ResNet Xception and MobileNet
[65]	PlantVillage	Descriptive Scale	13: 2:1	VGG-16

4.2.2 Improved CNN architecture

2 of the 16 articles are based on improved CNN architecture for severity grading. Compared classical CNN framework with 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.

In [3], A network, PD^{2}SE-Net, was proposed to design a more excellent and practical diagnosis system for plant diseases. PD^{2}SE-Net introduced the ResNet50 network as the basic model and integrated the building blocks of ShuffleNet-V2 [30]. The architecture of PD^{2}SE-Net is shown in Fig. 4. There are two key components of the PD^{2}SE-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 updates per batch. Inspired by ShaResNet [70], 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, PD^{2}SE-Net achieved plant species recognition, disease classification and severity estimation with overall accuracies of 0.99, 0.98 and 0.91, respectively.

In [71], Xiang et al. proposed a lightweight network, L-CSMS, based on residual networks, channel shuffle operation and multi-size module for plant disease severity recognition. 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 maps [72]. Channel shuffle operation was introduced to enable information communication between different channel groups and to improve accuracy. The channel shuffle operation and the multiple-size convolution module were incorporated 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. conducted comparison experiments between the L-CSMS model and ResNet, DenseNet, Inception-V4, PD^{2}SE-Net, ShuffleNet, and MobileNet. The results showed that L-CSMS achieved a competitive advantage with fewer parameters, FLOPs and comparatively good accuracy.

The ways to improve the CNN architecture vary, but the improvements are basically aimed at the same goal, which is to design a more accurate and practical system for plant disease severity assessment with better generalization performance. Compared with the classical CNN framework, there are relatively few studies based on the improved CNN architecture. From the available research results, this technique deserves focused attention and further in-depth study.

4.2.3 CNN-based semantic segmentation network

Image semantic segmentation has received increasing attention for computer vision and DL researchers, and research work on semantic segmentation using DL techniques continues to develop. In particular, CNNs have greatly surpassed other methods in terms of accuracy and efficiency [73]. Semantic segmentation provides not only category information but also additional information about the spatial location of these categories. The task of semantic segmentation is to label each pixel as a kind of closed objects or a category of regions [74]. Semantic segmentation theory has been applied to plant disease severity estimation and other related research work in agricultural. 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.

Usually, the architecture of semantic segmentation is divided into two parts: the encoder network and the decoder network. The encoder is usually based on CNN networks for generating low-resolution image representations or feature maps which are mapped to pixel-level images and then performs prediction and segmentation. The differences of the different semantic segmentation models are often reflected among the decoder networks [74]. The first successful application of DL techniques to semantic segmentation was achieved by a fully convolutional network (FCN) constructed by Long et al.. Afterwards, a series of variant forms 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 to achieve plant disease severity assessment. In [75], Chen et al. proposed a BLSNet for the severity estimation of rice bacterial leaf streak (BLS). BLSNet is 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 suggests that BLSNet is better suited for the adaptation of scale changes of images, and the prediction time of BLSNet is slightly longer than U-Net but shorter than DeepLabV3+. In [76], Gao et al. proposed a SegNet-based network to segment potato late blight (PLB) lesions to achieve quantification of the PLB severity. Goncalves et al. [77] conducted comparative experiments on six semantic segmentation networks (U-net, SegNet, PSPNet, FPN and 2 variants of DeepLabv3+), applied for three kinds of plant disease severity estimation (coffee leaf miner, soybean rust and wheat tan spot).

Although CNNs have yielded good results in the assessment of plant disease severity, the CNN-based semantic segmentation network also has its advantages. The achievement of CNN models to detect plant disease severity is by directly establishing a relationship between severity and samples, which is applicable to specific plant diseases, but may not be suitable 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 reflecting the severity by pixel-level segmentation. Existing studies have shown the feasibility of CNN-based semantic segmentation network for plant disease severity assessment.

5 Datasets And Evaluation Metrics

4.1 Plant disease severity datasets

The correct construction and rational use of plant disease severity datasets is a prerequisite and basis for severity assessment work. In contrast to ImageNet, PlantVillage and COCO in computer vision, there is no large unified datasets available for plant disease severity. Plant disease severity datasets can be collected by taking your own photos 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 usually acquired through cameras and smartphones. PlantVillage is a common public dataset used in plant disease severity, and common software for labeling images is LabelMe, LabelImg, etc. This section provides links to datasets and annotation software used in the research work of plant disease severity in conjunction with existing studies, as shown in Table 3.

Table 3
Datasets and labeling software for plant disease severity

Species	Link
PlantVillage dataset	https://plantvillage.psu.edu/
Coffee leaf severity dataset[46]	https://github.com/esgario/lara2018
AI Challenger Global AI Contest dataset	www.challenger.ai
Wheat Spike Blast Severity dataset[67]	https://purr.purdue.edu/publications/3772/1
LabelMe Software	https://github.com/wkentaro/labelme
LabelImg Software	https://github.com/tzutalin/labelImg

4.2 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. The specific definitions of them are described separately below.

Accuracy, precision and recall are expressed by the following equations:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \bullet 100\%$$

1

$$Precision = \frac{TP}{TP + FP} \bullet 100\%$$

2

$$Recall = \frac{TP}{TP + FN} \bullet 100\%$$

3

In Eq. (1) and Eq. (2), the true positive (*TP*) with the prediction value of 1 and the actual value of 1, indicates the number of correctly identified lesions. The false positive (*FP*) with the prediction value of 1 and the actual value of 0, indicates the number of misidentified lesions. The false negative (*FN*) with the prediction value of 0 and the actual value of 1, indicates the number of unidentified lesions. The true negative (*TN*) with the prediction value of 0 and the actual value of 0, indicates the number of correctly identified non-lesions.

It is first necessary to calculate the average precision for each category in the dataset for mAP.

$$P_{average} = \sum_{j=1}^{N(class)} Precision\,(j) \bullet Recall\,(j) \bullet 100\%$$

4

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

The average precision for each category is defined as follows:

$$mAP = \frac{P_{average}}{N\,(class)}$$

5

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

$$F1 = \frac{2Precision \bullet Recall}{Precision + Recall} \bullet 100\%$$

6

6 Challenges And Future Outlook

Compared with traditional image processing methods and machine learning, the assessment of plant disease severity has broad application prospects and great potential for development, through either 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 gulf away from mature applications in real natural environments, and many problems need to be solved.

6.1 Dataset issues

Dataset issues can be divided into two main aspects: the insufficiency of dataset and the imbalance of the various classes of dataset.

(1) The insufficiency of dataset. Adequate datasets are necessary and fundamental to train the network. However, the collection and construct of 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, for the research work on plant disease severity, datasets with severity annotation are really needed. Severity annotation of

images is a more arduous 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 research [1].

(2) The imbalance of the various classes of dataset. The imbalance problem can have a serious impact on the performance of the model, such as the misclassification rate becomes higher which has been demonstrated in the experiments of many studies [45, 46, 75]. This problem can be well alleviated by data augmentation and weighted loss functions [67].

6.2 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 [56–58]. At the same time, complex background in images can bring other negative problems. For example, in realistic environments, soil patches are similar to disease symptoms, leading to classification errors in the model [76]. Reflections due to natural illumination might cause shaded healthy areas to be misclassified, or disease areas to be undetected [68]. In addition, it is more common to see multiple diseases occurring simultaneously in real-world environments. In [46], the researchers mentioned that the presence of multiple diseases on a single 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 [45, 51, 60]. The problems caused by the complex background make it possible to apply the theoretical findings of CNN-based plant disease severity assessment to the actual agricultural production process with serious obstacles. To solve some problems caused by the complex environment, the images can be pre-processed, however which in turn increases the complexity of the whole detection process. For the problem of simultaneous multi-disease identification and assessment, in [46], researchers 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 is yet to be realized and more solutions need to be brainstormed.

6.3 Practicality issues

The application of CNN to the real-world production of plant disease severity assessment requires not only a certain standard of accuracy but also a high enough detection efficiency. The study of lightweight CNN models for plant disease severity assessment has opened our eyes to a new research idea. For example, proposed a cloud-based fruit classification framework based on 5G networks to apply deep learning in practice.

Currently, some problems still have not found appropriate solutions which indicates that the current researches 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.

Declarations

Acknowledgement:

We deeply acknowledge The National Natural Science Foundation of China, Natural Science Foundation of Hunan Province China, and Hunan Provincial Education Science "13th Five-Year Plan" Foundation.

Conflicts of Interest:

The authors declare that they have no conflicts of interest to report regarding the present study.

Data Availability

All data generated or analysed during this study are included in this published article and its supplementary information files.

Funding Statement:

This research is funded by The National Natural Science Foundation of China, No. 31870532 Natural Science Foundation of Hunan Province China, No. 2021JJ31163, Hunan Provincial Education Science "13th Five-Year Plan" Fund, No. XJK20BGD048.

References

1. Bock, C.H., et al., *Plant Disease Severity Estimated Visually, by Digital Photography and Image Analysis, and by Hyperspectral Imaging*. Critical Reviews in Plant Sciences, 2010. **29**(2): p. 59–107.
2. Mahlein, A.K., et al., *Development of spectral indices for detecting and identifying plant diseases*. Remote Sensing of Environment, 2013. **128**: p. 21–30.
3. Liang, Q., et al., *(PDSE)-S-2-Net: Computer-assisted plant disease diagnosis and severity estimation network*. Computers and Electronics in Agriculture, 2019. **157**: p. 518–529.
4. Liu, L., et al., *PestNet: An End-to-End Deep Learning Approach for Large-Scale Multi-Class Pest Detection and Classification*. Ieee Access, 2019. **7**: p. 45301–45312.

5. Madden, L.V., G. Hughes, and F. Van Den Bosch, *The study of plant disease epidemics.* 2007.
6. Cooke, B., *Disease assessment and yield loss*, in *The epidemiology of plant diseases*. 2006, Springer. p. 43–80.
7. Chiang, K.S., et al., *Effects of rater bias and assessment method on disease severity estimation with regard to hypothesis testing*. Plant Pathology, 2016. **65**(4): p. 523–535.
8. Xu, Y., et al., *Inheritance of resistance to zucchini yellow mosaic virus and watermelon mosaic virus in watermelon*. The Journal of heredity, 2004. **95**(6): p. 498–502.
9. Horsfall, J. and E. Cowling, *Pathometry: the measurement of plant disease*. Vol. 2. 1978: Academic Press New York.
10. Barratt, R. and J.J.P. Horsfall, An improved grading system for measuring plant disease. 1945. **35**: p. 655.
11. Bock, C.H., et al., *The Horsfall-Barratt scale and severity estimates of citrus canker*. European Journal of Plant Pathology, 2009. **125**(1): p. 23–38.
12. Forbes, G. and J.J.P.P. Korva, *The effect of using a Horsfall-Barratt scale on precision and accuracy of visual estimation of potato late blight severity in the field*. 1994. **43**(4): p. 675–682.
13. Bock, C.H., et al., *From visual estimates to fully automated sensor-based measurements of plant disease severity: status and challenges for improving accuracy*. Phytopathology Research, 2020. **2**(1).
14. Del Ponte, E.M., et al., *Standard Area Diagrams for Aiding Severity Estimation: Scientometrics, Pathosystems, and Methodological Trends in the Last 25 Years*. Phytopathology, 2017. **107**(10): p. 1161–1174.
15. Cobb, N.A.J.A.G.o.N.S.W., *Contributions to an economic knowledge of the Australian rusts (Uredineae).* 1892. **3**: p. 44–48.
16. McCulloch, W.S. and W.J.T.b.o.m.b. Pitts, A logical calculus of the ideas immanent in nervous activity. 1943. **5**(4): p. 115–133.
17. Fukushima, K., *Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position*. Biological cybernetics, 1980. **36**(4): p. 193–202.
18. Hubel, D.H. and T.N. Wiesel, *Receptive fields, binocular interaction and functional architecture in the cat's visual cortex*. The Journal of physiology, 1962. **160**: p. 106–54.
19. Wang, W., et al., *Development of convolutional neural network and its application in image classification: a survey*. Optical Engineering, 2019. **58**(4).
20. LeCun, Y., et al., Gradient-based learning applied to document recognition. 1998. **86**(11): p. 2278–2324.
21. Hinton, G.E., S. Osindero, and Y.-W. Teh, *A fast learning algorithm for deep belief nets*. Neural computation, 2006. **18**(7): p. 1527–54.
22. Glorot, X. and Y. Bengio. *Understanding the difficulty of training deep feedforward neural networks*. in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. 2010.

JMLR Workshop and Conference Proceedings.
23. Krizhevsky, A., I. Sutskever, and G.E.J.A.i.n.i.p.s. Hinton, Imagenet classification with deep convolutional neural networks. 2012. **25**.
24. Simonyan, K. and A.J.a.p.a. Zisserman, *Very deep convolutional networks for large-scale image recognition.* 2014.
25. Szegedy, C., et al. *Going deeper with convolutions*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
26. He, K., et al. *Deep residual learning for image recognition*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
27. Huang, G., et al. *Densely connected convolutional networks*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
28. Iandola, F.N., et al., *SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size.* 2016.
29. Howard, A.G., et al., *Mobilenets: Efficient convolutional neural networks for mobile vision applications.* 2017.
30. Ma, N., et al. *Shufflenet v2: Practical guidelines for efficient cnn architecture design*. in *Proceedings of the European conference on computer vision (ECCV)*. 2018.
31. Chollet, F. *Xception: Deep learning with depthwise separable convolutions*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
32. Tan, M. and Q. Le. *Efficientnet: Rethinking model scaling for convolutional neural networks*. in *International conference on machine learning*. 2019. PMLR.
33. Girshick, R., et al. *Rich feature hierarchies for accurate object detection and semantic segmentation*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014.
34. Girshick, R. *Fast r-cnn*. in *Proceedings of the IEEE international conference on computer vision*. 2015.
35. Ren, S., et al., *Faster r-cnn: Towards real-time object detection with region proposal networks.* 2015. **28**.
36. Redmon, J., et al. *You only look once: Unified, real-time object detection*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
37. Liu, W., et al. *Ssd: Single shot multibox detector*. in *European conference on computer vision*. 2016. Springer.
38. Long, J., E. Shelhamer, and T. Darrell. *Fully convolutional networks for semantic segmentation*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
39. Badrinarayanan, V., et al., *Segnet: A deep convolutional encoder-decoder architecture* for image segmentation. 2017. **39**(12): p. 2481–2495.
40. Ronneberger, O., P. Fischer, and T. Brox. *U-net: Convolutional networks for biomedical image segmentation*. in *International Conference on Medical image computing and computer-assisted intervention*. 2015. Springer.

41. Zhao, H., et al. *Pyramid scene parsing network*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
42. Chen, L.-C., et al., *Semantic image segmentation with deep convolutional nets and fully connected crfs.* 2014.
43. He, K., et al. *Mask r-cnn*. in *Proceedings of the IEEE international conference on computer vision*. 2017.
44. Wang, G., et al., *Automatic image-based plant disease severity estimation using deep learning.* 2017. **2017**.
45. Su, W.-H., et al., *Automatic Evaluation of Wheat Resistance to Fusarium Head Blight Using Dual Mask-RCNN Deep Learning Frameworks in Computer Vision*. Remote Sensing, 2021. **13**(1).
46. Esgario, J.G.M., R.A. Krohling, and J.A. Ventura, *Deep learning for classification and severity estimation of coffee leaf biotic stress*. Computers and Electronics in Agriculture, 2020. **169**.
47. Caruana, R.J.M.l., Multitask learning. 1997. **28**(1): p. 41–75.
48. Evgeniou, T. and M. Pontil. *Regularized multi–task learning*. in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. 2004.
49. Ruder, S.J.a.p.a., *An overview of multi-task learning in deep neural networks.* 2017.
50. Zhang, Y., Q.J.I.T.o.K. Yang, and D. Engineering, *A survey on multi-task learning.* 2021.
51. Prabhakar, M., et al., Deep learning based assessment of disease severity for early blight in tomato crop. 2020. **79**(39): p. 28773–28784.
52. Zeng, Q., et al., *Gans-based data augmentation for citrus disease severity detection using deep learning*. 2020. **8**: p. 172882–172891.
53. Fenu, G. and F.M.J.C. Malloci, *Using Multioutput Learning to Diagnose Plant Disease and Stress Severity.* 2021. **2021**.
54. Mongeon, P. and A.J.S. Paul-Hus, The journal coverage of Web of Science and Scopus: a comparative analysis. 2016. **106**(1): p. 213–228.
55. Afzal, W., R. Torkar, and R. Feldt. *A Systematic Mapping Study on Non-Functional Search-based Software Testing*. in *Proceedings of the Twentieth International Conference on Software Engineering & Knowledge Engineering (SEKE'2008), San Francisco, CA, USA, July 1–3, 2008*. 2008.
56. Goncharov, P., et al. *Disease Detection on the Plant Leaves by Deep Learning*. 2018.
57. Thapa, R., et al., *The Plant Pathology 2020 challenge dataset to classify foliar disease of apples.* 2020.
58. Zeng, W., et al. *High-order residual convolutional neural network for robust crop disease recognition*. in *Proceedings of the 2nd international conference on computer science and application engineering*. 2018.
59. Goodfellow, I., Y. Bengio, and A. Courville, *Deep learning*. 2016: MIT press.
60. Mohanty, S.P., D.P. Hughes, and M.J.F.i.p.s. Salathé, *Using deep learning for image-based plant disease detection.* 2016. **7**: p. 1419.

61. Ioffe, S. and C. Szegedy. *Batch normalization: Accelerating deep network training by reducing internal covariate shift*. in *International conference on machine learning*. 2015. PMLR.
62. Zeiler, M.D. and R. Fergus. *Visualizing and understanding convolutional networks*. in *European conference on computer vision*. 2014. Springer.
63. Chollet, F., *Deep learning with Python*. 2021: Simon and Schuster.
64. Brahimi, M., et al., *Deep learning for plant diseases: detection and saliency map visualisation*, in *Human and machine learning*. 2018, Springer. p. 93–117.
65. Sibiya, M. and M.J.P. Sumbwanyambe, Automatic fuzzy logic-based maize common rust disease severity predictions with thresholding and deep learning. 2021. **10**(2): p. 131.
66. Hu, G., et al., *Detection and severity analysis of tea leaf blight based on deep learning*. 2021. **90**: p. 107023.
67. Fernández-Campos, M., et al., Wheat spike blast image classification using deep convolutional neural networks. 2021. **12**: p. 1054.
68. Ozguven, M.M.J.F.E.B., *DEEP LEARNING ALGORITHMS FOR AUTOMATIC DETECTION AND CLASSIFICATION OF MILDEW DISEASE IN CUCUMBER*. 2020. **29**(8): p. 7081–7087.
69. Hayit, T., et al., *Determination of the severity level of yellow rust disease in wheat by using convolutional neural networks*. 2021. **103**(3): p. 923–934.
70. Boulch, A.J.a.p.a., Sharesnet: reducing residual network parameter number by sharing weights. 2017.
71. Xiang, S., et al., *L-CSMS: novel lightweight network for plant disease severity recognition*. 2021. **128**(2): p. 557–569.
72. Szegedy, C., et al. *Inception-v4, inception-resnet and the impact of residual connections on learning*. in *Thirty-first AAAI conference on artificial intelligence*. 2017.
73. Garcia-Garcia, A., et al., *A review on deep learning techniques applied to semantic segmentation.* 2017.
74. Hariharan, B., et al. *Simultaneous detection and segmentation*. in *European conference on computer vision*. 2014. Springer.
75. Chen, S., et al., *An approach for rice bacterial leaf streak disease segmentation and disease severity estimation*. 2021. **11**(5): p. 420.
76. Gao, J., et al., *Automatic late blight lesion recognition and severity quantification based on field imagery of diverse potato genotypes by deep learning*. 2021. **214**: p. 106723.
77. Goncalves, J.P., et al., *Deep learning architectures for semantic segmentation and automatic estimation of severity of foliar symptoms caused by diseases or pests*. 2021. **210**: p. 129–142.

Figures

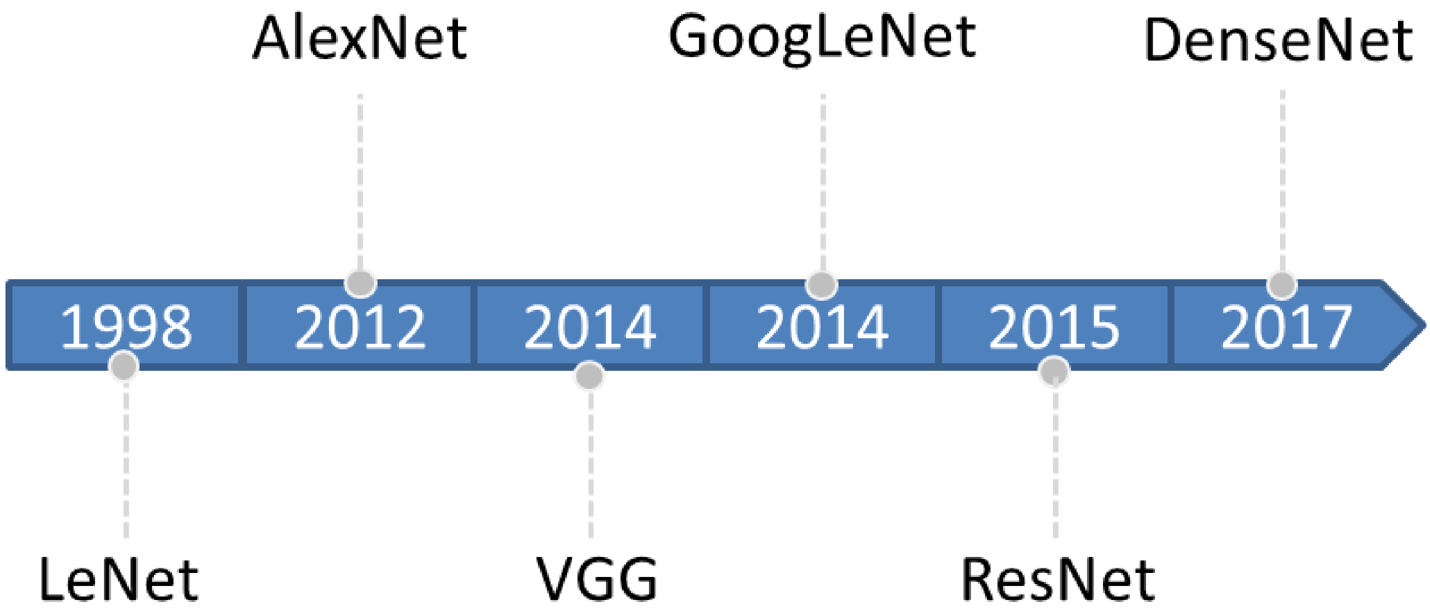

Figure 1

Evolution timeline of CNNs from LeNet to DenseNet

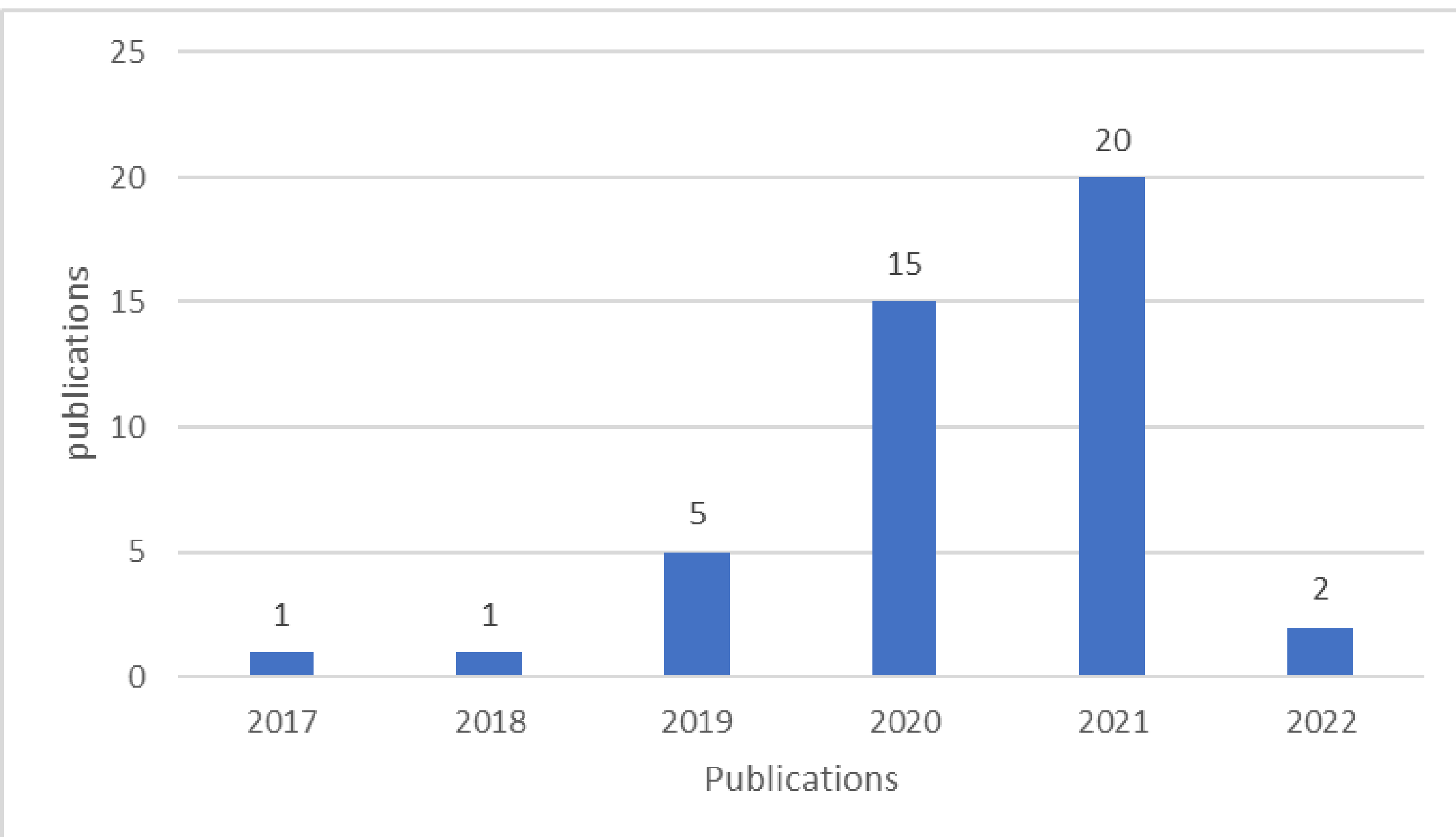

Figure 2

The distribution graph of publishing years of 44 articles based on the keywords of "convolutional neural network" and "plant disease severity" (from Web of Science)

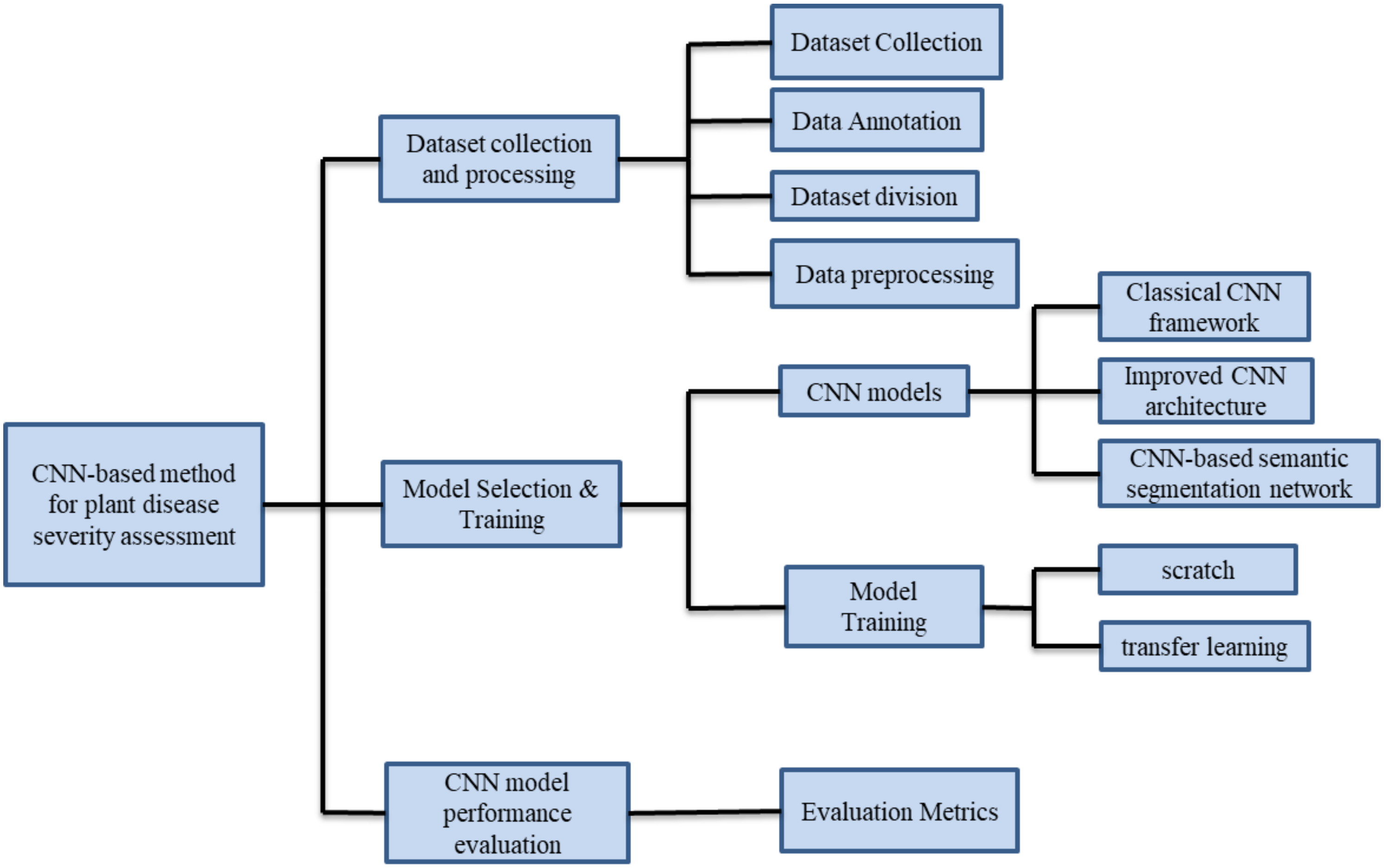

Figure 3

The flowchart of CNN-based plant disease severity assessment method

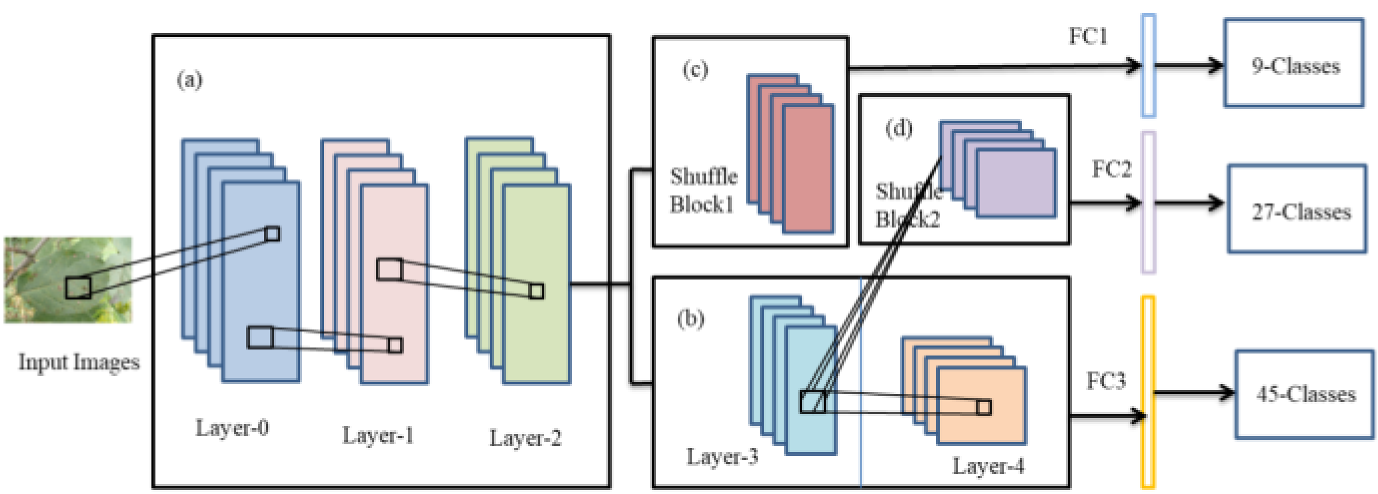

Figure 4

The architecture of PD^{2}SE-Net. Five parts: (a) parameter sharing layer; (b) The third layer is the parameter sharing layer between the fourth layer and the 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 layers. [3]

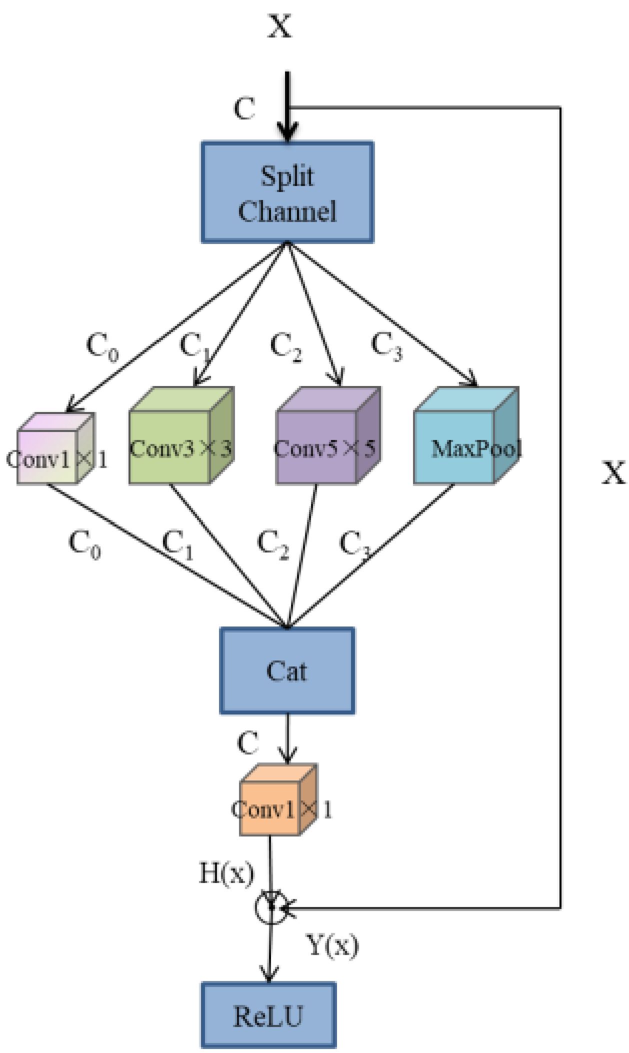

Figure 5

The building block with channel shuffle operation and multi size convolution module[71]

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AppendixA.docx