



# Effects of Image Dataset Configuration on the Accuracy of Rice Disease Recognition Based on Convolution Neural Network

Huiru Zhou, Jie Deng, Dingzhou Cai, Xuan Lv and Bo Ming Wu\*

College of Plant Protection, China Agricultural University, Beijing, China

## OPEN ACCESS

### Edited by:

Wanneng Yang,  
Huazhong Agricultural University,  
China

### Reviewed by:

Yang Lu,  
Heilongjiang Bayi Agricultural  
University, China  
Gongfa Li,  
Wuhan University of Science  
and Technology, China  
Liang Wanjie,  
Jiangsu Academy of Agricultural  
Sciences (JAAS), China

### \*Correspondence:

Bo Ming Wu  
bmwu@cau.edu.cn

### Specialty section:

This article was submitted to  
Technical Advances in Plant Science,  
a section of the journal  
Frontiers in Plant Science

Received: 01 April 2022

Accepted: 10 May 2022

Published: 5 July 2022

### Citation:

Zhou H, Deng J, Cai D, Lv X and  
Wu BM (2022) Effects of Image  
Dataset Configuration on  
the Accuracy of Rice Disease  
Recognition Based on Convolution  
Neural Network.  
Front. Plant Sci. 13:910878.  
doi: 10.3389/fpls.2022.910878

In recent years, the convolution neural network has been the most widely used deep learning algorithm in the field of plant disease diagnosis and has performed well in classification. However, in practice, there are still some specific issues that have not been paid adequate attention to. For instance, the same pathogen may cause similar or different symptoms when infecting plant leaves, while the same pathogen may cause similar or disparate symptoms on different parts of the plant. Therefore, questions come up naturally: should the images showing different symptoms of the same disease be in one class or two separate classes in the image database? Also, how will the different classification methods affect the results of image recognition? In this study, taking rice leaf blast and neck blast caused by *Magnaporthe oryzae*, and rice sheath blight caused by *Rhizoctonia solani* as examples, three experiments were designed to explore how database configuration affects recognition accuracy in recognizing different symptoms of the same disease on the same plant part, similar symptoms of the same disease on different parts, and different symptoms on different parts. The results suggested that when the symptoms of the same disease were the same or similar, no matter whether they were on the same plant part or not, training combined classes of these images can get better performance than training them separately. When the difference between symptoms was obvious, the classification was relatively easy, and both separate training and combined training could achieve relatively high recognition accuracy. The results also, to a certain extent, indicated that the greater the number of images in the training data set, the higher the average classification accuracy.

**Keywords:** deep learning, convolutional neural network, rice diseases, image recognition, crop disease dataset, model fitting

## INTRODUCTION

Rice production is facing many threats, especially many diseases caused by fungi, bacteria, and environmental factors (Zhang et al., 2018). Timely and accurate diagnosis of rice diseases is critical to the management of these diseases. Traditionally, disease diagnosis was mainly done by experienced personnel based on visible symptoms and laboratory identification (Sethy et al., 2020). However, experienced personnel are in short supply at grass-roots plant protection stations

in China and many other developing countries. Besides, the identification of crop diseases using laboratory technology is often laborious and time-consuming (Feng et al., 2020). Therefore, efforts have been made to develop alternative techniques, including image recognition based on machine learning for its timely feedback and low cost (Coulbaly et al., 2019; Abade et al., 2021; Bari et al., 2021).

Early automatic diagnoses of crop diseases were mainly done *via* image recognition based on traditional machine learning (Li et al., 2020). Many traditional machine learning algorithms, including self-organizing maps (Phadikar and Sil, 2008), back propagation neural network (Xiao et al., 2018), Naive Bayes (Islam et al., 2018), K-means clustering (Ghyar and Birajdar, 2017), and support vector machine (Yao et al., 2009), have been applied to the recognition of rice disease images. These algorithms achieved classification accuracy ranging from 92 to 97.2% in these studies, but the small training dataset and the huge feature extraction engineering have been two huge obstacles to the practical application of traditional machine learning algorithms in the field of rice diseases recognition (DeChant et al., 2017; Lu J. et al., 2017).

Deep learning, with the advantages of automatic feature extraction and efficient processing of big data, triggered a boom of research on image recognition these years (Min et al., 2017). Among many deep learning algorithms, the convolutional neural network (CNN) is most widely used in the field of computer vision (Voulodimos et al., 2018). The CNN automatically learns the features of the image through convolution and pooling operations, mimicking the processes of image recognition by the cerebral perception cortex (Yamins and DiCarlo, 2016), which suggested that CNN could perform like the human visual nerves in some way (Cadieu et al., 2014).

Recently, many researchers all over the world have also paid attention to apply deep learning, especially CNN, in the diagnosis of rice diseases. Some researchers trained existing CNN models with rice disease images (Ghosal and Sarkar, 2020; Deng et al., 2021; Krishnamoorthy et al., 2021), some built their own CNN models (Lu Y. et al., 2017), and some modified the classical CNN models such as DenseNet by adding inception module (Chen et al., 2020). Lightweight models, such as simple CNN in which model parameters were greatly reduced without precision loss, have also been developed for application with mobile devices (Rahman et al., 2020). As CNN is excellent in extracting features, Liang et al. (2019) also used a traditional SVM classifier for subsequent image classification based on image features extracted by CNN from images of rice leaf blast and achieved a significantly better classification accuracy by combining SVM with CNN than by combining SVM with two traditional feature extraction methods, namely, LBPH and Haar-WT.

The existing research results suggested that deep learning-based image recognition has become more and more mature and achieved high performance in the recognition of rice diseases, both in accuracy and efficiency. Therefore, instead of building new models or improving algorithms, more attention has been paid to solve specific and practical issues in training existing models by some researchers recently. For example, Mohanty et al. (2016) found that the image type used in model training and

the image allocation ratio between the training set and test set would have effects on the diagnosis accuracy of the resulted model. Picon et al. (2019) proved that training a model for multi-crops performed slightly better than developing specific models for individual crops. Lee et al. (2020) proved that if a model was trained with datasets containing plant diseases that were not associated with a specific crop, the model would be more suitable for a wider range of uses, especially for images obtained in different fields and images from unseen crops.

Similarly, automatic diagnosis of rice diseases has encountered some practical problems because of the high complexity of rice disease symptoms under field conditions. For example, similar or different symptoms can develop at different stages, under different weather conditions, or on different plant parts. Previous studies on the diagnosis of rice diseases concentrated on the recognition of typical symptoms of different rice diseases, but rarely addressed how the images of different symptoms caused by the same disease should be tagged in the construction of the training dataset. Should they be divided into different classes or combined into a single class? How will the different data configurations affect the accuracy of models? This has become an urgent problem to be solved before the automatic disease diagnosis can really be applied to field conditions.

Therefore, taking rice blast and rice sheath blight as examples in this study, experiments were conducted to explore how the split or merged disease classes in the configuration of training databases affect the recognition accuracy of the model. The specific objectives of this study were as follows:

- (1) To select an appropriate model from 5 common CNN models for the subsequent investigation;
- (2) To evaluate the effects of three training data configuration methods on the performance of CNN models during the training and test processes;
- (3) To identify where the misclassifications lie *via* constructing a normalized confusion matrix for each method; and
- (4) To explore the possible causes for misclassification by visualizing the recognition process.

## MATERIALS AND METHODS

### Choosing Crop Diseases and Construction of Datasets

#### Collection of Disease Images

Images of healthy rice leaves (HRL), and rice leaves or sheaths with symptoms of the three common diseases, rice blast (RB), rice brown spots (RBS), and rice sheath blight (RSB) were collected mainly from experimental rice fields in Panjin and Dandong cities of Liaoning Province, China. In addition, images were also collected from the greenhouse on the campus of China Agricultural University (CAU), from CAU experimental fields in Haidian District, Beijing, and from commercial fields in Wuyuan County of Jiangxi Province and in Lu'an City of Anhui Province. These images were photographed using smartphones or cameras following three rules: (1) avoid overexposure caused by direct sunlight; (2) ensure that the targeted lesion was in the center

of the picture; and (3) avoid different disease symptoms in a single picture.

The rice leaves, necks, heads, and whole plants with no visible symptoms were photographed and regarded as healthy rice plants. For rice leaf blast, images of chronic (RLBC), and acute (RLBA) leaf lesions were collected in this study at the early growth stage of rice because of their importance and prevalence under field conditions, while the other two less common symptom types, namely, white spot and brown spot, were not included in this study. According to Kato (2001), the chronic type leaf symptoms were defined as spindle-shaped leaf lesions with a yellow outside halo, a brown inner ring, and a gray white center (**Figure 1A**), while acute type symptoms were defined as the leaf lesions that are nearly round or oval in shape, which often become irregular, and look like water stains with a layer of dark green mold on the surface (**Figure 1B**). Besides, images of rice neck blast (RNB), the most economically important symptom of rice blast, were also collected at the late growth stages of rice in this study. According to Kumar et al. (1992), neck blast was defined as the symptoms that appeared around the neck of rice panicles as light brown spots at the initial stage and then gradually expand up and down, leading to a white gray color of the whole rice ear, and sometimes the death of whole ear (**Figure 1C**).

Since rice brown spot caused by *Bipolaris oryzae* has a similar shape and yellow halo to those of rice blast leaf lesions, images of rice leaves with brown spots were collected and used to test the recognition accuracy of the outcome models. According to Quintana et al. (2017), the infected leaves with sesame-like oval dark brown spots surrounded by yellow halos were considered as typical symptoms of rice brown spot (**Figure 1D**).

Another important disease, rice sheath blight caused by *Rhizoctonia solani*, which can cause similar symptoms on leaves (RSBL) and sheaths (RSBS), was also included in this study to illustrate how the classification of similar symptoms on different plant parts caused by the same pathogen would affect the accuracy of recognition. According to Lee and Rush (1983), the typical symptoms of this disease are cloud-shaped lesions on the leaf sheaths and leaves, with brown to dark brown edges and grayish green to grayish white middle parts (**Figures 1E,F**).

## Preprocessing of Images

As CNN requires squared input images, in order to avoid image deformation caused by the forced compression of non-squared images during input, the automatic clipping method was used to cut each image into a square, with the side length equal to the length of the short side of the original image and using the original image center as the clipping center. The clipped images were then compressed to  $500 \times 500$  pixels. Subsequently, normalization was applied on each image by dividing all pixel values with 255 to accelerate the convergence of models during the subsequent training procedure.

As the number of acquired images in some classes was inadequate for model training and validation, more images in these classes were generated to meet the requirement by image augmentation (**Table 1**). The methods used in augmentation included flip, translocation, rotation, and zoom (Francois, 2018).

## Experimental Scheme

Three experiments were designed to investigate the effects of dataset configuration on rice disease images recognition. In each experiment, two symptoms of one disease were selected for training and testing together with the other three diseases. In experiment 1, training datasets with separate and combined classes of RLBC and RLBA were compared. In experiment 2, training datasets with separate and combined classes of RLBC and RNB were compared. In experiment 3, training datasets with separate and combined classes of RSBL and RSBS were compared.

In each experiment, a method using two separated classes and two methods with one combined classes were compared. Considering that the imbalance of data may affect the training results, two methods were used in the construction of the combined class, directly combining all the images of two classes into one class, and randomly selecting half images from each class and combining them into one class.

## Construction of Datasets

Images of each class were randomly numbered after preprocessing, with a unique ID for each image. For example, the first image of RSB was named "RSB (0)." For each class, the first 500 or 1,000 images were used in training and validation datasets as required, and the images 1,001–1,099 were used to build test sets.

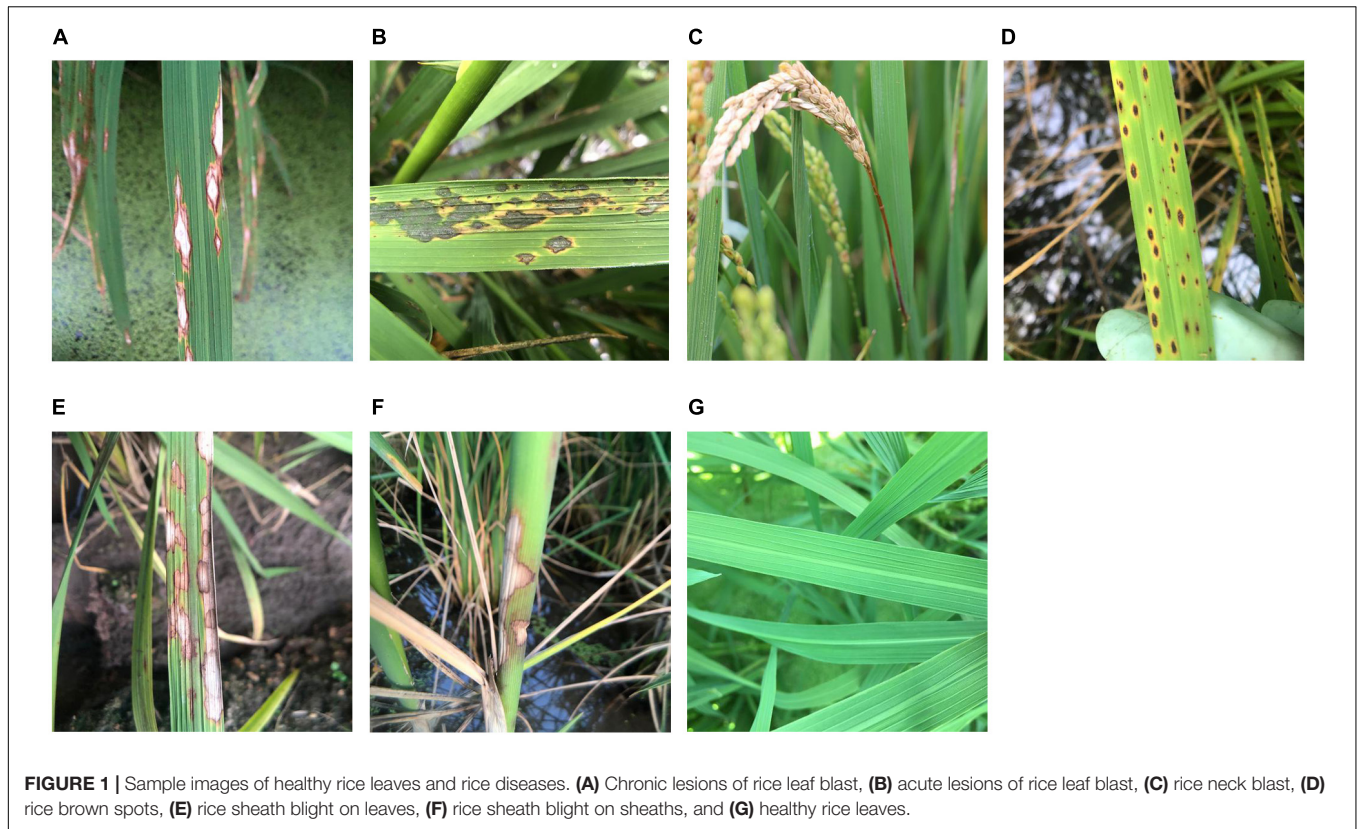
There were three independent datasets for each experiment. In experiment 1, 1,000 images of each class were divided into training set and validation set according to a ratio of 8:2 for method A. In method B, 1,000 images of RLBC and RLBA were directly merged into one class, with twice as many images as the other classes. In method C, 500 images were randomly taken from RLBC and RLBA, respectively, to form a combined class. The same ratio of 8:2 was used dividing image data into constructing training and validation sets in both methods B and C. In addition to the 1,000 images, other 100 images of each class were randomly selected to form a 500-image test set. These 500 images were used to test all three methods A, B, and C, but classes of RLBC and RLBA would be merged into one class for testing methods B and C. In the same way, training, validation, and test datasets were constructed in experiment 2 and experiment 3 (**Table 2**).

## Hardware and Software

Keras/Tensorflow backend framework based on Anaconda3 platform was used in this study (version: keras 2.2.4, tensorflow 1.15.0), and the training and validation processes were coded using Python 3.7 programming language. The computer was equipped with 32 g memory module and GTX 1080Ti graphics card. The computer operation system was the 64-bit Windows 10 professional edition. The programs were all run on a single graphic processing unit (GPU) because the training speed on GPU is much faster than that on the central processing unit (CPU).

## Training Parameter Setting

Instead of starting from scratch, transfer learning was applied in all model training experiments to saving



**TABLE 1 |** The number of images within each disease class obtained in this study.

	Rice blast			Rice sheath blight		Rice brown spots (RBS)	Healthy rice leaves (HRL)	Total
	Leaf		Neck	Leaf	Sheath			
	Acute (RLBA)	Chronic (RLBC)	(RNB)	(RSBL)	(RSBS)			
Initial number	1,146	1,186	599	1,193	598	1,143	1,146	7,011
Number after augmentation	1,146	1,186	1,198	1,193	1,196	1,143	1,146	8,208

**TABLE 2 |** The number of images in each disease class in training experiments using different methods.

Disease classes	Experiment 1			Experiment 2			Experiment 3		
	Method A	Method B <sup>b</sup>	Method C	Method J	Method K	Method L	Method X	Method Y	Method Z
HRL <sup>a</sup>	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
RBS	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
RLBA	1,000	[1,000]	[500]	/	/	/	/	/	/
RLBC	1,000	[1,000]	[500]	1,000	[1,000]	[500]	1,000	1,000	1,000
RNB	/	/	/	1,000	[1,000]	[500]	/	/	/
RSBL	1,000	1,000	1,000	1,000	1,000	1,000	1,000	[1,000]	[500]
RSBS	/	/	/	/	/	/	1,000	[1,000]	[500]

<sup>a</sup>HRL, healthy rice leaves; RBS, rice brown spot; RLBA, rice leaf blast-acute lesions; RLBC, rice leaf blast-chronic lesions; RNB, rice neck blast; RSBL, rice sheath blight on leaves; RSBS, rice sheath blight on sheaths.

<sup>b</sup>The images from the two classes within the braces were combined into one single class for training.

training time by carrying the weights from the training on ImageNet dataset (Russakovsky et al., 2015). The learning rate was set as 0.001, and the training was run

for 50 epochs with a momentum of 0.9, an optimization function of stochastic gradient descent (SGD), and a mini-batch size of 32.

## Model Selection

Different algorithms have their own purposes or specific application scenarios when designing or modifying. For example, a multi-stream residual network (MResLSTM) was designed for dynamic hand movement recognition (Yang et al., 2021), and a modified YOLO v3 algorithm was applied to detect helmet wearing by construction personnel (Huang et al., 2021). At present, however, there is no widely used model for the diagnosis of rice diseases, so we conducted a preliminary experiment to select from five representative CNN models for subsequent experiments on the construction of datasets.

The VGG series (Simonyan and Zisserman, 2014), first developed by the VGG group of Oxford University, were CNN models with stacked  $3 \times 3$  convolution kernels for extracting complex features with a manageable number of parameters. Considering the moderate size of our disease data, VGG16 (16 layers) was selected as the representative of this model series. Compared with the VGG series, some CNN models used more network layers to extract higher dimension features and took different approaches to handle the gradient dispersion problem associated with deeper networks (Gao et al., 2019). Inception v3 was chosen as a candidate model in this study for its deep depths and its inception module, which uses convolution kernels of different sizes in the same layer to realize feature fusion of different scales and batch normalization to speed up the learning rate (Szegedy et al., 2015). ResNet50 (50 layers) was included as a representative of ResNet series, in which a residual module was introduced for a shortcut connection in the network allowing the original input information to be directly transmitted to the later layer (He et al., 2015). In addition, MobileNet v2 (Howard et al., 2017) and NASNetMobile (Zoph et al., 2018), two representatives of the current lightweight models in the application scenarios of mobile terminals or embedded devices, were also selected for their relatively excellent performance and small number of parameters (Wang et al., 2020).

A pre-experiment was conducted to compare the performances of the five CNN models in recognition of the three rice leaf diseases and healthy rice leaves (**Figure 1G**). The 1,000 training images from each of the five classes, namely, RLBC, RLBA, RBS, RSBL, and HRL, were divided into a training dataset and a validation dataset according to the ratio of 8:2. The models were trained for 50 epochs using the transfer learning method, and the initial weights of five models were all set as the shared weights from training on ImageNet as described in the “Training parameter setting” section. The size of models, speed of training (in seconds per epoch), the highest validation accuracy, the final validation accuracy, the average validation accuracy, and standard deviation of validation accuracy were used to evaluate the models.

## Experiments and Statistical Analysis

Subsequently, 3 experiments were done using the best model selected from the pre-experiment. Due to the random input order of mini-batches, the results of training could vary at each run. To estimate this variation and assess the reliability of the results, each of the three experiments was repeated three times. The final

validation accuracy, final validation loss, test accuracy, and test loss were analyzed using the GLM procedure in SAS (version 9.4, SAS Institute Inc., Cary, NC, United States) to determine whether the effects of the training dataset configuration were statistically significant.

Over the 50 epochs of the training processes, the average validation accuracy and average validation loss of three repeated experiments were calculated every four epochs. As the performance of each method fluctuated over epochs in the training process, to better express the whole trend during the process, regression was performed to fit a negative exponential decay model to the average validation accuracy and an exponential decay model to average validation loss over the training processes for each method using the non-linear regression procedure in SAS (Version 9.4, SAS Institute Inc., Cary, NC, United States).

For validation accuracy, the following model was used:

$$A = A_{max} - (A_{max} - A_0) e^{(-r_a \cdot x)}$$

where  $A$  was the validation accuracy and  $x$  was the epoch number in training, while  $A_{max}$ ,  $A_0$ , and  $r_a$  were parameters to be estimated in model fitting.  $A_{max}$  reflects the highest validation accuracy that the method can reach,  $A_0$  reflects the initial validation accuracy, and  $r_a$  can reflect the increase rate of  $A$  or improvement rate of validation accuracy over epochs.

For validation loss, the following model was used in regression:

$$L = L_{min} + e^{(-r_l \cdot x + b)}$$

where  $L$  was the validation loss and  $x$  was the epoch number in training, while  $L_{min}$ ,  $r_l$ , and  $b$  were parameters to be estimated during model fitting.  $L_{min}$  represents the lowest validation loss rate obtained by this method after unlimited epochs,  $L_{min} + e^{(-r_l + b)}$  reflects the initial validation loss at epoch #1, and  $r_l$  is related to the decline rate of validation loss.

After model fitting, the parameters were compared between different methods using Student's  $t$ -test (Steel and Torrie, 1980) to characterize the disparity of the three methods in the training process.

## Normalized Confusion Matrix

Confusion matrix, which was widely used in the evaluation of classification accuracy in many areas, was constructed for comparison of different training dataset configuration methods based on test results. As the image numbers of the classes to be tested in this study varied among different training dataset configuration methods, to better reflect their difference in classification accuracy, the normalized confusion matrix was used. For any classification with  $c$  classes, the confusion matrix consisted of  $c$  rows  $\times$   $c$  columns, and the element in the  $i$ -th row and  $j$ -th column was calculated by dividing the number of images that belonged to the  $i$  class and were classified into the  $j$  class with the total number of images in the row.

## Heatmap

To understand which parts of the input image, such as the lesion edge, the center, or other areas, had contributed more to the

automatic classification by the models, for each representative image with a high frequency of misclassification in recognition, a heatmap of class activation was generated using the GRAD-CAM algorithm (Selvaraju et al., 2020), in which the pixels that contributed heavily to the final classification will be presented as yellow to red colors and those that contributed less will be presented in green to purple colors. The heatmap generated this way serves as a tool for visualization of the feature extraction process of deep neural network.

## RESULTS

### Performance of the Five Models in Pre-experiments

The results from the pre-experiment demonstrated that the five CNN models performed differently in the classification of these images (Table 3). VGG16 and Inception v3 all achieved a validation accuracy higher than 99%, but ResNet50 had the highest average validation accuracy and a smaller standard deviation among these models, suggesting that its convergence speed was the fastest and its performance was the most stable. Considering ResNet50's excellent performance in training, including good speed (38 s/epoch), the highest final validation accuracy, the highest average validation accuracy, and the smallest standard deviation, it was selected for the subsequent training experiments.

### Experiment 1: Different Symptoms of the Same Disease on the Same Part

The results from experiment 1 revealed that the training curves of validation accuracy and validation loss using method A differed from those using methods B and C (Figures 2A,B). Method A consistently had lower validation accuracy and higher validation loss than methods B and C did during the whole training process over 50 epochs (Figures 2A,B). Differences also existed between method B and method C in the early epochs of the training process, but the difference gradually decreased to an ignorable level with the increase of training epochs. Regardless of methods used in training dataset configuration, the trends of validation accuracy over training epochs could be fitted well to the negative

exponential decay model (Table 4A) and those of validation loss fitted well to the exponential decay model (Table 4B). The  $t$ -test indicated that the highest accuracy ( $A_{max}$ ) obtained using method A was lower than those using the other two methods, while the lowest validation loss ( $L_{min}$ ) using method A was significantly greater than those using methods B and C (Tables 4A,B). The growth rate  $r_a$  of validation accuracy and the decline rate  $r_l$  of validation loss were significantly faster for method B than for the other two methods.

The ANOVA and multiple mean comparison revealed that on both validation and test datasets, the validation accuracy and test accuracy obtained using method A were significantly lower than those obtained using methods B and C, and the validation loss and test loss obtained using method A were significantly greater than those obtained using methods B and C (Table 5).

The confusion matrix of test results using method A revealed that the class with the lowest accuracy was RLBC, and the main classification errors came from the misclassification of RLBC images into RLBA by the model (Figure 3A). When combining the two classes into one for training, the test accuracy ranged from 96 to 99% in every class with little variation among classes (Figures 3B,C). To understand why the misclassifications occurred, the original images of these misclassified RLBC images were visually examined again. It was found that although the leaf lesions in these images were nearly spindle shaped, the edges and corners were not obvious enough. When there were many lesions on leaves, they connected into pieces that were more like water stains, and the surfaces of some lesions were even gray green, which were typical symptoms of RLBA at the early stage of developing into RLBC (Kumar et al., 1992).

### Experiment 2: Different Symptoms of the Same Disease at Different Parts

The validation accuracy obtained using method K was highest among the three methods at the beginning of the training processes, and the lowest accuracy was gained using method L, but the accuracy increase rates  $r_a$  were higher for methods L (0.6564) and J (0.5377) than for method K (0.2800), and as a result, the three methods differed very less in accuracy after 20 training epochs (Figure 2C and Table 4A), and the maximum accuracy gained after 50 epochs varied from 0.9935 to 0.9940,

**TABLE 3** | Performance of five CNN models in the classification of rice disease images.

CNN models	Seconds/epoch	Highest validation accuracy <sup>a</sup>	Final validation accuracy <sup>b</sup>	Average validation accuracy <sup>c</sup>	Standard deviation of validation accuracy <sup>d</sup>
VGG16	36	0.9900	0.9890	0.8758	0.2473
Inception v3	60	0.9980	0.9920	0.9803	0.4020
<b>ResNet50</b>	<b>38</b>	<b>0.9940</b>	<b>0.9920</b>	<b>0.9851</b>	<b>0.0107</b>
MobileNet v2	37	0.9830	0.9520	0.7138	0.2199
NASNetMobile	57	0.9870	0.9870	0.9693	0.0352

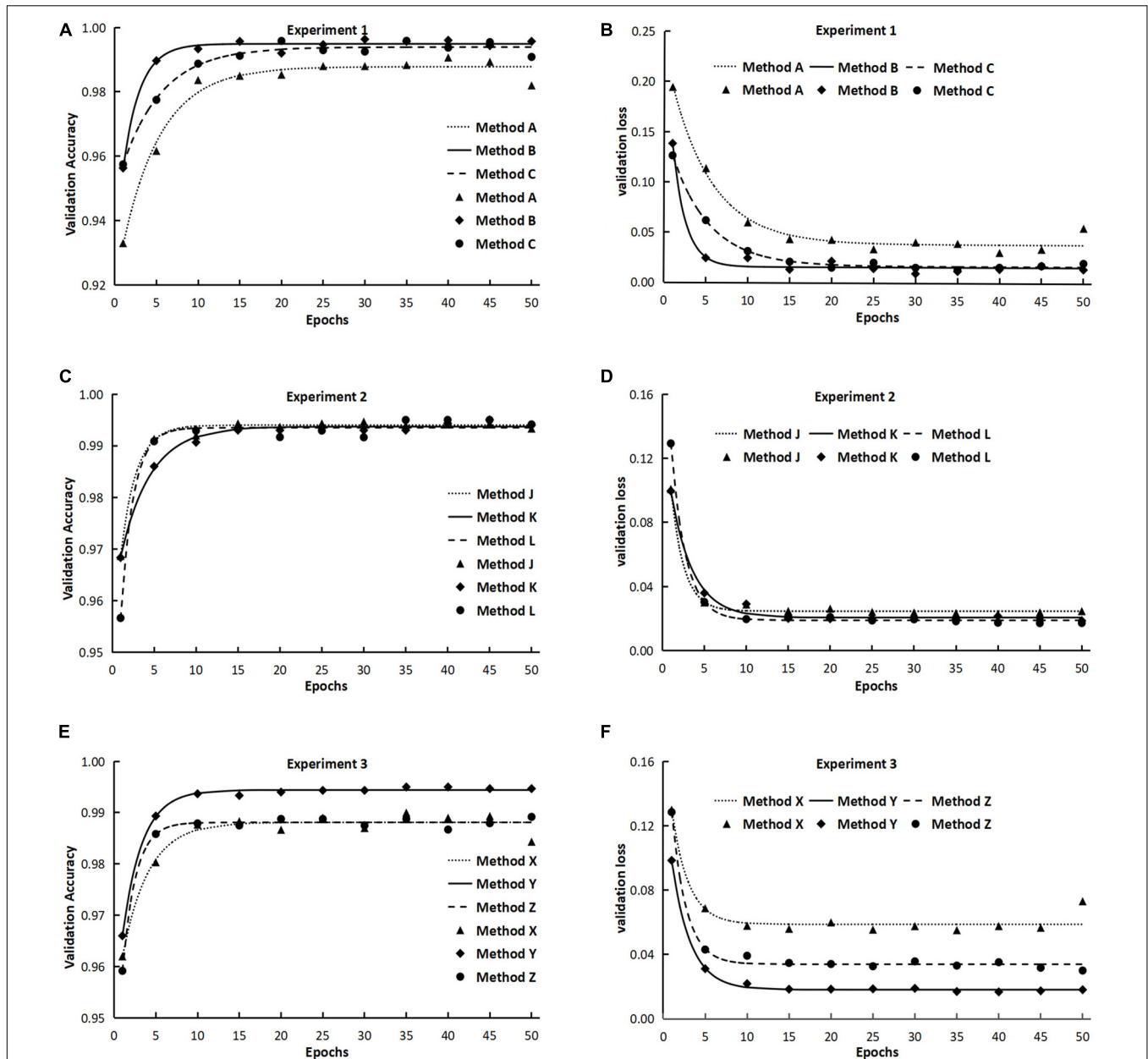
<sup>a</sup>The highest validation accuracy achieved within the first 50 epochs.

<sup>b</sup>The final validation accuracy after 50 epochs.

<sup>c</sup>The average validation accuracy over the first 50 epochs.

<sup>d</sup>The standard deviation of validation accuracy over the first 50 epochs, which reflects the convergence rate of five models.

The training results of the selected model ResNet50 were shown in bold.



**FIGURE 2 |** The validation accuracy and validation loss during the training processes in experiments 1, 2, and 3. (The points in the figures were means from three repeated runs, and the lines represented the fitted models of validation accuracy and validation loss.) Method A: Training with two separate classes, namely, acute type of rice leaf blast (RLBA) and chronic type of rice leaf blast (RLBC). Method B: Combining RLBA and RLBC as one class for training and the total number of images in the combined class was two times as those in the other three classes. Method C: Combining RLBA and RLBC as one class for training and the total number of images in the combined class was equal to those in the other three classes. Method J: Training with two separate classes of RLBC and rice neck blast (RNB). Method K: Combining RLBC and RNB as one class for training and the total number of images in the combined class was two times as those in the other three classes. Method L: Combining RLBC and RNB as one class for training and the total number of images in the combined class was equal to those in the other three classes. Method X: Training with two separate classes of rice sheath blight on leaves (RSBL) and rice sheath blight on sheath (RSBS). Method Y: Combining RSBL and RSBS as one class for training and the total number of images in the combined class was two times as those in the other three classes. Method Z: Combining RSBL and RSBS as one class for training and the total number of images in the combined class was equal to those in the other three classes.

showing no significant difference among the three methods (Table 4A). On the contrary, the validation loss using method L was the highest among the three methods early in the training, but it declined quickly as the training progressed and ended the

training with a loss value that was very close to the other two methods (Figure 2D and Table 5).

Interestingly, although the accuracy using methods J, K, and L differed slightly (insignificantly) on validation data, the test

**TABLE 4A** | Parameters and determinant coefficients of models [ $A = A_{max} - (A_{max} - A_0) e^{-r_a \cdot x}$ ] fitted to the validation accuracy over training epochs in 3 experiments using different methods.

Experiment-method	$A_0$	$A_{max}$	$r_a$	$R^2$
1-method A	0.9191 ± 0.0047 <sup>b</sup>	0.9878 ± 0.0011 <sup>b</sup>	0.2154 ± 0.0289 <sup>b</sup>	0.977
1-method B	0.9319 ± 0.0050 <sup>b</sup>	0.9949 ± 0.0005 <sup>a</sup>	0.4896 ± 0.0667 <sup>a</sup>	0.985
1-method C	0.9489 ± 0.0028 <sup>a</sup>	0.9939 ± 0.0007 <sup>a</sup>	0.2077 ± 0.0254 <sup>b</sup>	0.985
2-method J	0.9507 ± 0.0035 <sup>b</sup>	0.9940 ± 0.0003 <sup>a</sup>	0.5377 ± 0.0718 <sup>a</sup>	0.991
2-method K	0.9603 ± 0.0014 <sup>a</sup>	0.9937 ± 0.0003 <sup>a</sup>	0.2800 ± 0.0230 <sup>b</sup>	0.992
2-method L	0.9223 ± 0.0099 <sup>c</sup>	0.9935 ± 0.0005 <sup>a</sup>	0.6564 ± 0.1332 <sup>a</sup>	0.992
3-method X	0.9520 ± 0.0037 <sup>a</sup>	0.9881 ± 0.0006 <sup>b</sup>	0.3206 ± 0.0629 <sup>b</sup>	0.960
3-method Y	0.9510 ± 0.0015 <sup>a</sup>	0.9944 ± 0.0002 <sup>a</sup>	0.4244 ± 0.0260 <sup>b</sup>	0.997
3-method Z	0.9340 ± 0.0053 <sup>b</sup>	0.9881 ± 0.0003 <sup>b</sup>	0.6274 ± 0.0921 <sup>a</sup>	0.992

$R^2$ : Degree of coincidence between test data and fitting function. The closer the value of  $R^2$  is to 1, the higher the degree of coincidence is. Within each experiment, the fitted value followed by different letters in the same column differed significantly at confidence level  $p = 0.05$ .

**TABLE 4B** | Parameters and determinant coefficients of models [ $L = L_{min} + e^{-(r_l \cdot x + b)}$ ] fitted to the validation loss over training epochs in 3 experiments using different methods.

Experiment-method	$L_{min}$	$r_l$	$b$	$R^2$
1-method A	0.0382 ± 0.0030 <sup>a</sup>	-0.2020 ± 0.0242 <sup>b</sup>	-1.6449 ± 0.0591 <sup>a</sup>	0.980
1-method B	0.0156 ± 0.0016 <sup>b</sup>	-0.6331 ± 0.1274 <sup>a</sup>	-1.4649 ± 0.1325 <sup>a</sup>	0.985
1-method C	0.0165 ± 0.0009 <sup>b</sup>	-0.2226 ± 0.0126 <sup>b</sup>	-1.9856 ± 0.0281 <sup>b</sup>	0.991
2-method J	0.0246 ± 0.0006 <sup>a</sup>	-0.6433 ± 0.0788 <sup>a</sup>	-1.9319 ± 0.0816 <sup>b</sup>	0.996
2-method K	0.0206 ± 0.0005 <sup>b</sup>	-0.3797 ± 0.0360 <sup>b</sup>	-2.1610 ± 0.0499 <sup>c</sup>	0.991
2-method L	0.0189 ± 0.0005 <sup>b</sup>	-0.5632 ± 0.0353 <sup>a</sup>	-1.6410 ± 0.0382 <sup>a</sup>	0.991
3-method X	0.0587 ± 0.0019 <sup>a</sup>	-0.5010 ± 0.1536 <sup>a</sup>	-2.1412 ± 0.1754 <sup>ab</sup>	0.948
3-method Y	0.0181 ± 0.0004 <sup>c</sup>	-0.4453 ± 0.0216 <sup>a</sup>	-2.0777 ± 0.0266 <sup>b</sup>	0.998
3-method Z	0.0339 ± 0.0009 <sup>b</sup>	-0.5686 ± 0.0682 <sup>a</sup>	-1.7911 ± 0.0734 <sup>a</sup>	0.994

Within each experiment, the fitted value followed by different letters in the same column differed significantly at confidence level  $p = 0.05$ .

accuracy obtained using method L was significantly lower, and the test loss was significantly greater than those obtained using method J and method K (Table 5).

The confusion matrix for method J in experiment 2 revealed that the model can distinguish RNB from other classes well, and the accuracy of RLBC was the lowest among the five classes, with majority of misclassification errors between RLBC and RSBL, but its recognition accuracy of RNB was relatively

high (Figure 3D). When a combined class of RNB with RLBC was used in methods K and L, the accuracy of the combined RNB/RLBC class was between those of the two separate classes (Figures 3E,F). It was also noted that considerable errors existed in misclassifying RSBL into RLBC or combined class of RLBC with RNB regardless of the methods used (Figures 3D–F). This revealed that the identification of different plant parts is an indispensable part of classification by CNN models, and this identification could help to distinguish diseases on different plant parts, but similar symptoms on the same plant parts could not use this information and therefore become a more difficult task. It was also very interesting to note that method L had lower accuracy on combined RLBC/RNB class than method K. This might have been because method K had been trained with more images of the combined class than method L.

### Experiment 3: Similar Symptoms of the Same Disease at Different Parts

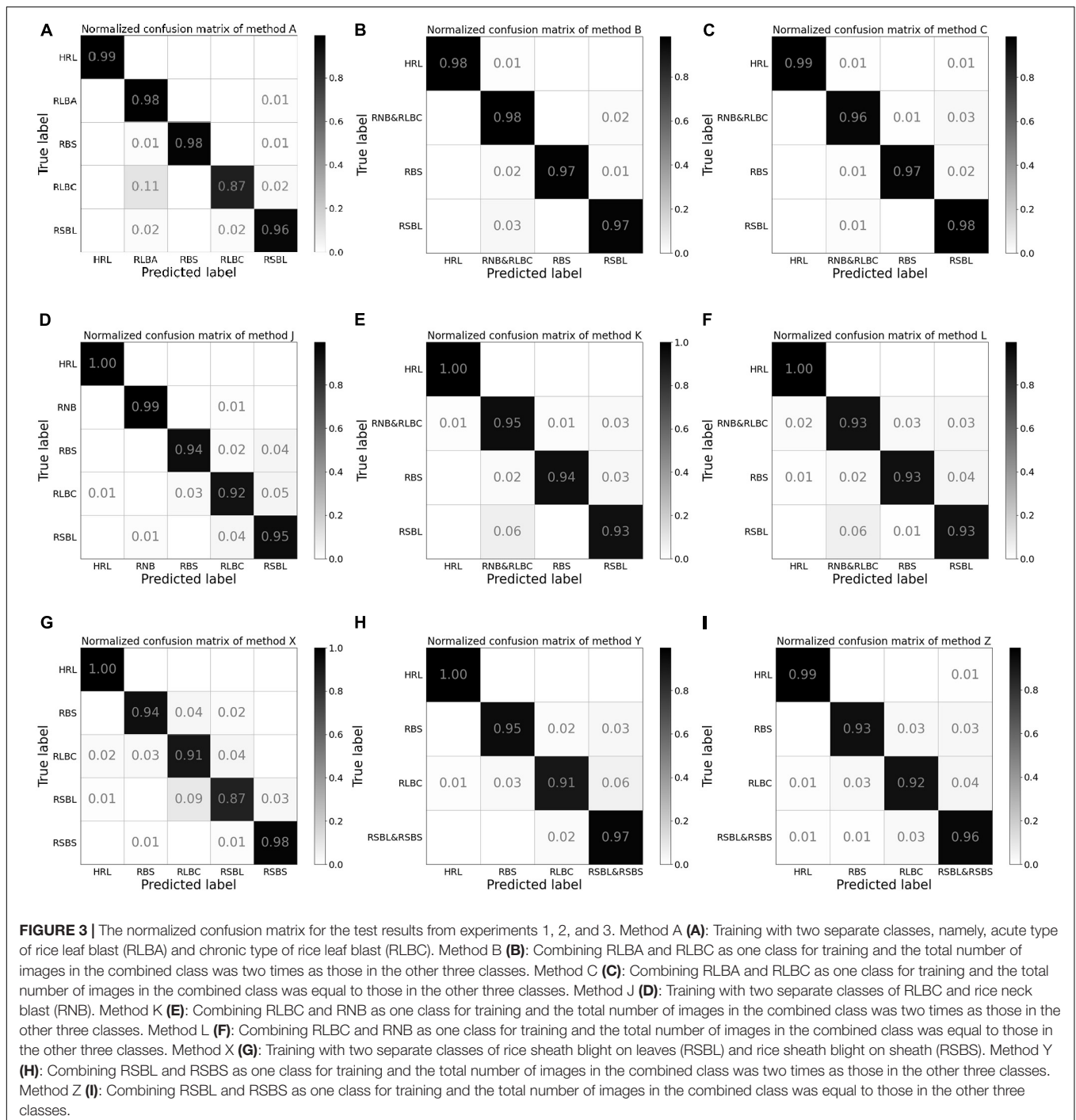
The initial validation accuracy of method Y was the highest among the three methods, and with the increase in training epochs, its validation accuracy remained highest all way to the end (Figure 2E). The results from *t*-test on model parameters  $A_{max}$  and  $r_a$  revealed that the highest validation accuracy  $A_{max}$  from method Y was significantly higher than those from the other two methods, but no significant difference in  $r_a$  was detected

**TABLE 5** | The accuracy and loss obtained with validation and test datasets in three experiments using different methods.

Experiment-method	Val_acc	Val_loss	Test_acc	Test_loss
1-Method A	0.9831 <sup>b</sup>	0.0530 <sup>a</sup>	0.9573 <sup>b</sup>	0.1680 <sup>a</sup>
1-Method B	0.9928 <sup>a</sup>	0.0211 <sup>b</sup>	0.9767 <sup>a</sup>	0.0685 <sup>b</sup>
1-Method C	0.9912 <sup>a</sup>	0.0254 <sup>b</sup>	0.9727 <sup>a</sup>	0.0957 <sup>b</sup>
2-Method J	0.9928 <sup>a</sup>	0.0282 <sup>a</sup>	0.9593 <sup>a</sup>	0.1223 <sup>c</sup>
2-Method K	0.9918 <sup>a</sup>	0.0251 <sup>a</sup>	0.9547 <sup>a</sup>	0.1586 <sup>b</sup>
2-Method L	0.9919 <sup>a</sup>	0.0236 <sup>a</sup>	0.9420 <sup>b</sup>	0.1884 <sup>a</sup>
3-Method X	0.9867 <sup>b</sup>	0.0610 <sup>a</sup>	0.9400 <sup>a</sup>	0.2174 <sup>a</sup>
3-Method Y	0.9927 <sup>a</sup>	0.0225 <sup>c</sup>	0.9593 <sup>a</sup>	0.1060 <sup>a</sup>
3-Method Z	0.9871 <sup>b</sup>	0.0376 <sup>b</sup>	0.9527 <sup>a</sup>	0.1263 <sup>a</sup>

Each of the values in the table was the mean from 3 repeated runs. Within each experiment, the means followed by different letters in the same column differed significantly at confidence level  $p = 0.05$ .





between method Y and method Z (Table 4A). The validation loss curves obtained with three methods displayed trends reverse to validation accuracy, in that no significant difference in the decline rate  $r_l$  of validation loss was detected among three methods (Figure 2F), but method Y had the lowest validation loss among the three methods, and method X had the highest validation loss.

The test results showed that the average test accuracy of method Y was much higher than that of method X and slightly

higher than that of method Z (Table 5), although the ANOVA detected no significant difference between the three methods (see Supplementary Material). The confusion matrix of the test results illustrated that the model trained with method X misclassified 4% RLBC images as RSBL and 9% RSBL images as RLBC (Figure 3G). When the model was trained using method Y, with a combined class of RSBL&RSBS, its accuracy was greatly improved that it misclassified 6% of RLBC images into the combined class, but only misclassified 2% RSBL&RSBS images

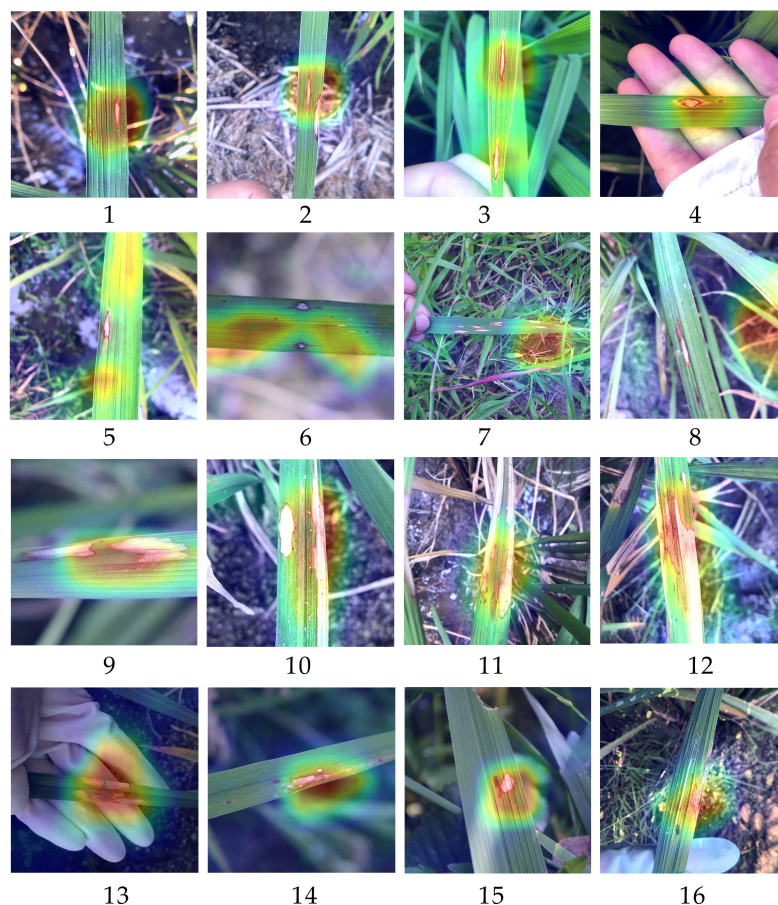
into RLBC (**Figure 3H**). Similar results were gained with method Z (**Figure 3I**). Once again, classifying between RLBC and RSBL was a difficult task, and the accuracy for the combined class was higher for method Y than for method Z, where more combined class images were used in training for method Y than method Z.

To further explore the reasons why differentiating RLBC and RSBL was difficult and easy to be misclassified for the outcome models, heatmaps of RLBC samples correctly classified, RLBC images misclassified as RSBL, RSBL samples correctly classified, and RSBL images misclassified as RLBC were compared (**Figure 4**). For those correctly classified RLBC, the areas with hot color were concentrated around the disease lesions, suggesting an excellent feature extraction by the model (**Figures 4.1–4.4**). However, it was observed that in most of the misclassified RLBC samples, the hot loci were not well overlapped with the disease lesions, suggesting the model didn't extract important lesion features for decision-making, interfered either by other leaf damages (**Figures 4.5,4.6**) or field background (**Figures 4.7,4.8**). The existence of RLBC that directly led to the test results of the three methods of experiment 3 was not significantly

different. Unlike RLBC, for all RSBL images, regardless of whether correctly classified or misclassified, the classification areas were mainly concentrated on the disease lesion area (**Figures 4.9–4.16**). It can be seen that compared with the typical symptoms of RSBL, in most of the misclassified samples, lesions were relatively small, had gray center areas, and were surrounded by brown halos, which was, to a certain degree, similar to the atypical RLBC, except for the subtle difference in lesion shapes (**Figures 4.15,4.16**). This may be one of the reasons why more rice sheath blight images were identified as RLBC than RBS.

## DISCUSSION AND CONCLUSION

In this study, we explored some specific problems encountered in dataset configuration for automatic recognition of rice diseases. The results from this study demonstrated that whether a combined class or several separated classes should be used depend on the similarity of these classes. For example, our results from experiment 1 demonstrated that using a combined class



**FIGURE 4 |** Heatmaps generated based on the classification by models trained with methods X for some rice leaf blast samples and rice sheath blight samples. **(1–4)** The samples of chronic type of rice leaf blast (RLBC) that were correctly recognized as RLBC; **(5–8)** The samples of RLBC that were mistakenly recognized as RSB. **(9–12)** The samples of rice sheath blight (RSB) that were correctly recognized as RSB. **(13–16)** The samples of RSB that were mistakenly recognized as RLBC. The red part has the highest contribution to the final prediction results. On the contrary, the purple part has the lowest contribution to the final prediction results of an image).

for RLBA and RLBC, two very similar symptoms on rice leaves, could achieve better recognition performance (higher accuracy and lower loss) than using two separate classes. A possible explanation might be that similar lesions sometimes are difficult to differentiate even for human experts because acute lesions often gradually develop into chronic lesions in the later stage (Kumar et al., 1992). This was also supported by the high misclassification rate between these two classes by method A using separate classes, but relatively low misclassification rates between any of these two classes and other class by method A. Similarly, for RSBL and RSBS in experiment 3, using a combined class in the training dataset could achieve a better performance than using two separate classes. A possible explanation is that using two separate classes of RSBL and RSBS will require the model to differentiate the similar cloud-shaped lesions on leaves and on sheaths and therefore will increase the possibility for the model to make mistakes in recognition of the background plant parts. However, using a combined class and using separate classes for RLBC and RNB had no significant impact on the performance of resulted models in experiment 2 where two symptoms were on different plant parts. A possible explanation for this was that with information from areas surrounding lesions, it is relatively easy for CNN model to differentiate two different symptoms, and thus, using a combined class or two separated classes didn't have any significant impact on the final recognition as illustrated among methods J, K, and L in experiment 2 of this study.

The results from this study illustrated that a large number of images were required for training to achieve a high and repeatable recognition accuracy. As revealed in experiment 2, method L, in which the model was trained with half as many images of RLBC/RNB class as in methods J and K, although gained very high validation accuracy, performed significantly worse than methods J and K when tested with unseen images. So, for deep learning model, how many images are required to achieve best recognition effect? Is the more the number of images, the better the result will be? So far, few experts have explored this issue, and the number of images used in the existing literature varied from dozens to thousands. Rangarajan et al. (2018) discussed the influence of different number of images on the accuracy of the model, but the total number of images was small, and a scientific validation process has not been established yet. More in-depth studies are needed to answer this question in the future.

Through this study, we further prove the excellent ability of CNN in feature extraction. Based on the results of three experiments, it can be seen that the main features affecting the decision-making of rice disease classification models came from the disease lesion area, then from the area of plant organs, and finally from the image background. This is also consistent with the logic of human beings when classifying crop diseases. It can be seen from the heatmaps that for most samples, whether by correctly classified or by misclassified, the main feature areas that affect the model decision-making were still concentrated on the lesion area, and the areas were covered with red or yellow. The nearby areas of rice organs were also yellow, while the less important background areas were covered with blue or purple. The results of experiment 3 showed that when the disease

lesions of RSB were similar, even if they existed on different organs, there would be confusion between RSBL and RSBS to some extent, indicating that the main distinguishing features still came from the lesion. At the same time, the results of experiment 2 showed that when the symptoms and organs were different, the model could extract more favorable information except the features of disease lesions, and that is why it can well distinguish the two classes when separately training RLBC and RNB. Does this mean that the image background is not important? Studies have shown that although the targets on the simple indoor background image and the complex field image were the same, the models trained by the two image sets could not be universal (Ferentinos, 2018). From the heatmaps of RLBC images misclassified as RSBL, it can be seen that although the error rate was low, the main factor causing the wrong model decision was the feature extraction of the field background. This also showed that the recognition of the background played an auxiliary role for the model. Therefore, it is very important to collect disease images under different conditions to improve the generalization ability of the models.

The results of three experiments showed that if the data configuration scheme was correct, the overall accuracy could be effectively improved. In experiment 1, combining the two similar leaf symptoms of rice and training, the validation accuracy was improved from 0.9831 to 0.9928, and the test accuracy was improved from 0.9573 to 0.9767, which was statistically significant at the confidence level of 0.05. Similarly, in experiment 3, the average accuracy of two symptoms of rice sheath blight was improved to 0.9700 from 0.9250 by using a combined class for similar symptoms on different plant parts. However, the results of experiment 2 revealed that for disparate symptoms on different plant parts, training with one combined class or two separate classes makes no difference, and the amount of data is a key factor affecting the overall accuracy. The average test accuracy of method L with a smaller data set was significantly lower (at a confidence level of 0.05) than that of the other two methods with larger datasets.

This study proposed a database configuration scheme among different symptoms of the same rice disease. Similar problems are often encountered in the diagnosis of other crop diseases (Barbedo, 2016). If our goal is to achieve a high overall classification accuracy, the findings from this study provide a reference. However, if the purpose is to differentiate multiple similar disease symptoms on different plant parts or at different stages, even if the symptoms are similar, they should be separately trained. Hopefully, the findings from this study can inspire researchers to put more efforts in automatic crop disease identification and think about the problems of disease identification from more different perspectives.

## DATA AVAILABILITY STATEMENT

The original contributions presented in this study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

HZ led the whole experimental process and wrote the manuscript. JD provided suggestions about experimental design and guided the operation of the algorithms. DC participated in the construction of the dataset and collated data. XL participated in the image collection process. BW supervised the project and wrote the manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

This research was funded by the National Natural Science Foundation of China (No. 31471727) and the National Key Research and Develop Program of China (#2016YFD0300702).

## REFERENCES

- Abade, A., Ferreira, P. A., and Vidal, F. D. B. (2021). Plant diseases recognition on images using convolutional neural networks: a systematic review. *Comput. Electron. Agric.* 185:106125. doi: 10.1016/j.compag.2021.106125
- Barbedo, J. G. A. (2016). A review on the main challenges in automatic plant disease identification based on visible range images. *Biosyst. Eng.* 144, 52–60. doi: 10.1016/j.biosystemseng.2016.01.017
- Bari, B. S., Islam, M. N., Rashid, M., Hasan, M. J., Razman, M. A. M., Musa, R. M., et al. (2021). A real-time approach of diagnosing rice leaf disease using deep learning-based faster R-CNN framework. *PeerJ Comput. Sci.* 7:e432. doi: 10.7717/peerj-cs.432
- Cadiou, C. F., Hong, H., Yamins, D. L. K., Pinto, N., Ardila, D., Solomon, E. A., et al. (2014). Deep neural networks rival the representation of primate IT cortex for core visual object recognition. *PLoS Comput. Biol.* 10:e1003963. doi: 10.1371/journal.pcbi.1003963
- Chen, J., Zhang, D., Nanekaran, Y. A., and Li, D. (2020). Detection of rice plant diseases based on deep transfer learning. *J. Sci. Food Agric.* 100, 3246–3256. doi: 10.1002/jsfa.10365
- Coulibaly, S., Kamsu-Foguem, B., Kamissoko, D., and Traore, D. (2019). Deep neural networks with transfer learning in millet crop images. *Comput. Ind.* 108, 115–120. doi: 10.1016/j.compind.2019.02.003
- DeChant, C., Wiesner-Hanks, T., Chen, S., Stewart, E. L., Yosinski, J., Gore, M. A., et al. (2017). Automated identification of northern leaf blight-infected maize plants from field imagery using deep learning. *Phytopathology* 107, 1426–1432. doi: 10.1094/PHYTO-11-16-0417-R
- Deng, R., Tao, M., Xing, H., Yang, X., Liu, C., Liao, K., et al. (2021). Automatic diagnosis of rice diseases using deep learning. *Front. Plant Sci.* 12:701038. doi: 10.3389/fpls.2021.701038
- Feng, L., Wu, B., Zhu, S., Wang, J., Su, Z., Liu, F., et al. (2020). Investigation on data fusion of multisource spectral data for rice leaf diseases identification using machine learning methods. *Front. Plant Sci.* 11:577063. doi: 10.3389/fpls.2020.577063
- Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Comput. Electron. Agric.* 145, 311–318. doi: 10.1016/j.compag.2018.01.009
- Francois, C. (2018). *Deep Learning with Python*. New Jersey: Manning Publications.
- Gao, Q., Liu, J., Ju, Z., and Zhang, X. (2019). Dual-Hand detection for human-robot interaction by a parallel network based on hand detection and body pose estimation. *IEEE Trans. Ind. Electron.* 66, 9663–9672. doi: 10.1109/TIE.2019.2898624
- Ghosal, S., and Sarkar, K. (2020). “Rice leaf diseases classification using CNN with transfer learning,” in *Proceedings of 2020 IEEE Calcutta Conference*. (Piscataway: IEEE), 230–235.
- Ghyar, B. S., and Birajdar, G. K. (2017). “Computer vision based approach to detect rice leaf diseases using texture and color descriptors,” in *Proceedings of*

## ACKNOWLEDGMENTS

We would like to thank Donggang Demonstration Farm, Panjin Saline Alkali Land Utilization, and Research Institute for providing experimental fields in collecting rice disease images. We would also like to thank Huayuan Yao, Kai Yuan, Meina Hu and Hui Liu for their help in the process of collecting images.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.910878/full#supplementary-material>

- the International Conference on Inventive Computing and Informatics(ICIC)*. (Piscataway: IEEE), 1074–1078.
- He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep residual learning for image recognition. *arXiv [preprint]*. doi: 10.48550/arXiv.1512.03385
- Howard, A. G., Menglong, Z., Chen, B., Kalenichenko, D., Weijun, W., Weyand, T., et al. (2017). MobileNets: efficient convolutional neural networks for mobile vision applications. *arXiv [preprint]*. doi: 10.48550/arXiv.1704.04861
- Huang, L., Fu, Q., He, M., Jiang, D., and Hao, Z. (2021). Detection algorithm of safety helmet wearing based on deep learning. *Concurr. Comput. Pract. Exp.* 33:e6234. doi: 10.1002/cpe.6234
- Islam, T., Sah, M., Baral, S., and RoyChoudhury, R. (2018). “A Faster Technique on Rice Disease Detection using Image Processing of Affected Area in Agro-Field,” in *Proceedings of the 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)* (Coimbatore, India: IEEE), 62–66. doi: 10.1109/ICICCT.2018.8473322
- Kato, H. (2001). Rice blast disease - an introduction. *Pestic. Outlook* 12, 23–25. doi: 10.1039/b100803j
- Krishnamoorthy, N., Prasad, L. V. N., Kumar, C. S. P., Subedi, B., Abraha, H. B., and Sathishkumar, V. E. (2021). Rice leaf diseases prediction using deep neural networks with transfer learning. *Environ. Res.* 198:111275. doi: 10.1016/j.envres.2021.111275
- Kumar, J., Chaube, H. S., Singh, U. S., and Mukhopadhyay, A. N. (1992). *Plant Diseases of International Importance*. New Jersey: Prentice Hall, Inc.
- Lee, F. N., and Rush, M. C. (1983). Rice sheath blight: a major rice disease. *Plant Dis.* 67, 829–832. doi: 10.1094/PD-67-829
- Lee, S. H., Goeau, H., Bonnet, P., and Joly, A. (2020). New perspectives on plant disease characterization based on deep learning. *Comput. Electron. Agric.* 170:105220. doi: 10.1016/j.compag.2020.105220
- Li, Y., Nie, J., and Chao, X. (2020). Do we really need deep CNN for plant diseases identification? *Comput. Electron. Agric.* 178:105803. doi: 10.1016/j.compag.2020.105803
- Liang, W. J., Zhang, H., Zhang, G. F., and Cao, H. X. (2019). Rice blast disease recognition using a deep convolutional neural network. *Sci. Rep.* 9:2869. doi: 10.1038/s41598-019-38966-0
- Lu, J., Hu, J., Zhao, G., Mei, F., and Zhang, C. (2017). An in-field automatic wheat disease diagnosis system. *Comput. Electron. Agric.* 142, 369–379. doi: 10.1016/j.compag.2017.09.012
- Lu, Y., Yi, S. J., Zeng, N. Y., Liu, Y. R., and Zhang, Y. (2017). Identification of rice diseases using deep convolutional neural networks. *Neurocomputing* 267, 378–384. doi: 10.1016/j.neucom.2017.06.023
- Min, S., Lee, B., and Yoon, S. (2017). Deep learning in bioinformatics. *Brief. Bioinform.* 18, 851–869. doi: 10.1093/bib/bbw068
- Mohanty, S. P., Hughes, D. P., and Salathe, M. (2016). Using deep learning for Image-Based plant disease detection. *Front. Plant Sci.* 7:1419. doi: 10.3389/fpls.2016.01419
- Phadikar, S., and Sil, J. (2008). “Rice Disease Identification using Pattern Recognition Techniques,” in *Proceedings of the 2008 11th International*

- Conference on Computer and Information Technology*. (Khulna, Bangladesh: IEEE), 420–423.
- Picon, A., Seitz, M., Alvarez-Gila, A., Mohnke, P., Ortiz-Barredo, A., and Echazarra, J. (2019). Crop conditional Convolutional Neural Networks for massive multi-crop plant disease classification over cell phone acquired images taken on real field conditions. *Comput. Electron. Agric.* 167:105093. doi: 10.1016/j.compag.2019.105093
- Quintana, L., Gutierrez, S., Arriola, M., Morinigo, K., and Ortiz, A. (2017). Rice brown spot *Bipolaris oryzae* (Breda de Haan) Shoemaker in Paraguay. *Trop. Plant Res.* 4, 419–420. doi: 10.22271/tp.2017.v4.i3.055
- Rahman, C. R., Arko, P. S., Ali, M. E., Khan, M. A. I., Apon, S. H., Nowrin, F., et al. (2020). Identification and recognition of rice diseases and pests using convolutional neural networks. *Biosyst. Eng.* 194, 112–120. doi: 10.1016/j.biosystemseng.2020.03.020
- Rangarajan, A. K., Purushothaman, R., and Ramesh, A. (2018). Tomato crop disease classification using pre-trained deep learning algorithm. *Procedia Comput. Sci.* 133, 1040–1047. doi: 10.1016/j.procs.2018.07.070
- Russakovsky, O., Deng, J., and Su, H. (2015). ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* 115, 211–252. doi: 10.1007/s11263-015-0816-y
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2020). Grad-CAM: visual Explanations from Deep Networks via Gradient-Based Localization. *Int. J. Comput. Vis.* 128, 336–359. doi: 10.1007/s11263-019-01228-7
- Sethy, P. K., Barpanda, N. K., Rath, A. K., and Behera, S. K. (2020). Deep feature based rice leaf disease identification using support vector machine. *Comput. Electron. Agric.* 175:105527. doi: 10.1016/j.compag.2020.105527
- Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for Large-Scale image recognition. *arXiv* [preprint]. doi: 10.48550/arXiv.1409.1556
- Steel, R. G. D., and Torrie, J. H. (1980). *Principles and Procedures of Statistics*. New York: McGraw-Hill Book Company.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2015). Rethinking the inception architecture for computer vision. *arXiv* [preprint]. doi: 10.48550/arXiv.1512.00567
- Voulodimos, A., Doulamis, N., Doulamis, A., and Protopapadakis, E. (2018). Deep learning for computer vision: a brief review. *Comput. Intel. Neurosci.* 2018:7068349. doi: 10.1155/2018/7068349
- Wang, W., Hu, Y., Zou, T., Liu, H., Wang, J., and Wang, X. (2020). A new image classification approach via improved MobileNet models with local receptive field expansion in shallow layers. *Comput. Intel. Neurosci.* 2020:8817849. doi: 10.1155/2020/8817849
- Xiao, M., Ma, Y., Feng, Z., Deng, Z., Hou, S., Shu, L., et al. (2018). Rice blast recognition based on principal component analysis and neural network. *Comput. Electron. Agric.* 154, 482–490. doi: 10.1016/j.compag.2018.08.028
- Yamins, D. L. K., and DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nat. Neurosci.* 19, 356–365. doi: 10.1038/nn.4244
- Yang, Z., Jiang, D., Sun, Y., Tao, B., Tong, X., Jiang, G., et al. (2021). Dynamic gesture recognition using surface EMG signals based on Multi-Stream residual network. *Front. Bioeng. Biotechnol.* 9:779353. doi: 10.3389/fbioe.2021.779353
- Yao, Q., Guan, Z., Zhou, Y., Tang, J., Hu, Y., and Yang, B. (2009). “Application of support vector machine for detecting rice diseases using shape and color texture features,” in *Proceedings of the 2009 International Conference on Engineering Computation*. (Hong Kong, China: IEEE), 79–83. doi: 10.1109/ICEC.2009.73
- Zhang, J., Yan, L., and Hou, J. (2018). “Recognition of rice leaf diseases based on salient characteristics,” in *Proceedings of the 2018 13th World Congress on Intelligent Control and Automation (WCICA)*. (Changsha, China: IEEE), 801–806.
- Zoph, B., Vasudevan, V., Shlens, J., and Le, Q. V. (2018). “Learning transferable architectures for scalable image recognition,” in *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. (Salt Lake City, USA: IEEE), 8697–8710. doi: 10.1109/CVPR.2018.00907

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher’s Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhou, Deng, Cai, Lv and Wu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.