



Rice disease identification method based on improved CNN-BiGRU

Yang Lu^{a,*}, Xiaoxiao Wu^a, Pengfei Liu^a, Hang Li^{b,c}, Wanting Liu^a

^a College of Information and Electrical Engineering, Heilongjiang Bayi Agricultural University, Daqing, Heilongjiang 163319, China

^b Artificial Intelligence Energy Research Institute, Northeast Petroleum University, Daqing, Heilongjiang 163318, China

^c Sanya Offshore Oil and Gas Research Institute, Northeast Petroleum University, Sanya, Hainan 572025, China

ARTICLE INFO

Article history:

Received 4 February 2023

Received in revised form 8 August 2023

Accepted 20 August 2023

Available online 25 August 2023

Keywords:

Deep learning

CNN-BiGRU

Rice disease

Feature relationship

ABSTRACT

In the field of precision agriculture, diagnosing rice diseases from images remains challenging due to high error rates, multiple influencing factors, and unstable conditions. While machine learning and convolutional neural networks have shown promising results in identifying rice diseases, they were limited in their ability to explain the relationships among disease features. In this study, we proposed an improved rice disease classification method that combines a convolutional neural network (CNN) with a bidirectional gated recurrent unit (BiGRU). Specifically, we introduced a residual mechanism into the Inception module, expanded the module's depth, and integrated an improved Convolutional Block Attention Module (CBAM). We trained and tested the improved CNN and BiGRU, concatenated the outputs of the CNN and BiGRU modules, and passed them to the classification layer for recognition. Our experiments demonstrate that this approach achieves an accuracy of 98.21% in identifying four types of rice diseases, providing a reliable method for rice disease recognition research.

© 2023 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co., Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Rice is a major crop in Asia (Khush, 1997). With the intensification of rice cultivation, its yield and economic benefits have increased continuously. However, this has also led to the frequent occurrence of rice diseases, which pose a threat to food security (He et al., 2010). In China, blast disease, bacterial leaf blight, sheath blight, and rice stripe disease are the major types of diseases that are prone to occur (Zhu et al., 2004; Sun et al., 1998; Peng et al., 2015). These diseases can occur throughout the entire growth cycle of rice, and 10%–15% of rice yield losses are caused by rice diseases (Peng et al., 2009). Timely and effective monitoring of rice diseases is particularly important for stabilizing yield and saving costs. Traditional crop disease monitoring methods involve destructive sampling, laboratory analysis, and specialized instruments, which are destructive and lagging. Computer vision is a scientific method for obtaining information about the investigated crop without contact. Rice diseases mainly manifest on the stems and leaves of rice, and mobile devices such as cameras can record and obtain information about plant phenotypes and their changes, providing an objective means for the quantification of rice diseases compared to visual methods. In the field of image recognition research, machine learning methods such as random forests, support vector machines, and KNN have had a profound impact. (Rumpf et al., 2010) used support vector

machines and hyperspectral technology to identify early sugar beet diseases, with an accuracy rate of 97% for distinguishing healthy sugar beet from diseased leaves. Wang et al. (2019) extracted color, shape, and texture features of fungal diseases such as powdery mildew, rust, and leaf spots on wheat leaves and used support vector machines to classify and identify them. (Majumdar et al., 2015) used fuzzy c-means clustering algorithm to extract 25 disease features of wheat leaves and classified them using an artificial neural network to determine whether wheat was infected. Traditional machine learning relies on manually generated specific features, which results in low recognition efficiency. Remote sensing methods have achieved certain research results in crop stress monitoring, but lack spectral specificity of diseases, which may encounter uncertainty issues in monitoring. With the continuous development of big data and deep learning research, plant phenotype research based on deep learning methods provides a new means for crop disease recognition and has achieved significant results in crop monitoring and management.

(Sethy et al., 2020) evaluated the performance of 13 rice disease identification models. Statistical analysis showed that the ResNet50 and SVM classification models had higher recognition accuracy than the other models. (Waheed et al., 2020) proposed an optimized dense convolutional neural network called DenseNet, which achieved an accuracy of 98.06% and had fewer parameters and computation time than similar CNNs. (Da Costa et al., 2020) studied external defects in tomatoes and compared them with transfer learning training, showing that the ResNet18 model was the best. In peanut variety identification,

* Corresponding author.

E-mail address: luyanga@sina.com (Y. Lu).

(Yang et al., 2021) improved the VGG16 and fine-tuned it to achieve an accuracy of 96.7%, which was 8.9% higher than the original VGG16. (Ferentinos, 2018) used multiple deep learning models to detect plant leaf diseases and achieved the highest accuracy of 99.53% on the test set using the VGG network. (Barbedo, 2018) used transfer learning on GoogleNet to detect 56 diseases in 12 types of plants. (Atole and Park, 2018) used AlexNet to differentiate between three categories of rice plants - healthy, diseased, and snail-infected - based on 227 images. (Too et al., 2019) evaluated the performance of deep convolutional neural networks in plant disease classification using fine-tuning, and the DenseNet network achieved a testing accuracy of 99.75% on the plant village dataset. (Jiang et al., 2020) used convolutional neural networks (CNNs) to extract image features of rice leaf diseases. They then applied the SVM method to classify and predict specific diseases, obtaining an accuracy of 96% using 10-fold cross-validation to determine the optimal SVM parameters. (Chen et al., 2020) used transfer learning to identify different diseases in rice and corn, and classified plant disease images using the pre-trained VGGNet network. (Rahman et al., 2020) used the architectures of VGG16 and InceptionV3 and fine-tuned them to detect and recognize rice pests and diseases. The experimental results showed that VGG16 and InceptionV3 had recognition accuracies of 97.12% and 96.37%, respectively. (Woo et al., 2018) proposed a spatial and channel-wise feedforward convolutional neural network attention module. It is a lightweight model that can be integrated into most CNNs. In the lightweight CNN model proposed by (Bao et al., 2021), which integrates CBAM, the recognition accuracy of wheat spikes in the field reached 94.1%.

These work have demonstrated the reliability of computer vision techniques in diagnosing crop diseases. As researchers tackle old issues and discover new challenges, they are exploring ways to combine different methods to address complex problems. However, current CNNs used in visual applications have limitations, as they only emphasize local features and overlook the relationships between these extracted features. For instance, the distribution patterns and internal weight differences of features are crucial in capturing long-term feature dependencies, which is necessary for identifying rice disease symptoms. This information cannot be fully captured by local features alone, thus requiring special attention. Recently, RNNs such as LSTM (Graves and Graves, 2012) and GRU (Chung et al., 2014) have gained attention for their excellent performance in handling sequence features in language and signal processing tasks. The expression of disease features can be viewed as numerical matrix information, which can be flattened into nonlinear sequence data for effective processing by RNNs. This allows research to focus on information that CNNs may overlook. Specific processing methods need to be developed for downstream tasks in computer vision, and for modeling rice diseases, wider networks are better able to detect finer-grained disease features. This study improves the model by considering disease features as semantic information describing the type of disease and transforming it into sequence data, which is then trained and summarized using RNNs. This method combines CNNs and RNNs to diagnose rice diseases, attempting to learn the relationships between features and address current research limitations.

This paper makes the following contributions:

- (1) We propose an improved CNN-BiGRU model for diagnosing rice diseases using images.
- (2) To accurately capture the features of rice diseases, we integrate an improved Convolutional Block Attention Module (CBAM), which increases the model's focus on regions of interest in the image, thereby reducing the negative impact of noise on feature capture.
- (3) The high-level features extracted by the CNN are flattened and treated as sequential data, and BiGRU is introduced to further investigate the dependency between the preceding and following information of rice disease characteristics.
- (4) The model achieves a recognition accuracy of 98.21% on four rice disease samples, making it a reliable automated detection method.

2. Methods

2.1. The structure of improved CNN-BiGRU

The proposed method in this paper combines CNN and RNN for diagnosing rice diseases. In order to enable the model to learn contextual features between upper and lower layers of features, we used BiGRU to learn relevant information from both forward and backward directions. To reduce the computational cost incurred by flattening the entire image, we employed an improved CNN module as an image feature encoder. The CNN module works by aggregating information and shrinking data normalization. The output of the improved CNN is then transformed from matrix to vector form, and fed into the RNN for training. The output data is then transmitted back for further refinement. The output from both the CNN and RNN modules were concatenated to form the final decision data, which is then passed to the classification layer for classification. The overall architecture is illustrated in Fig. 1. The specific method will be described in detail below.

2.2. The structure of improved CNN

To improve the performance of Convolutional Neural Networks (CNNs), increasing the network's width and depth is a direct method, but this may lead to higher computational costs and network degradation. Fixed convolutional kernel sizes can also result in the loss of disease feature information. The Inception module includes a wide network layer and several different sizes of convolutional kernels. By using multi-scale convolutional calculations, it extracts multi-scale features that effectively describe the image. To accurately extract rice disease information, we established a CNN model based on the InceptionV1 (Szegedy et al., 2015) module, with an increased depth of the module. Two 3×3 convolutional kernels are equivalent to a 5×5 , with a larger receptive field for detecting larger disease points in rice diseases while reducing computational complexity. We replaced a branch and added a 3×3 convolutional kernel to extract small point information, capturing local disease information and avoiding the loss of small target information. A 1×1 convolutional kernel can be used for upsampling and downsampling, and can flexibly adjust the output feature size. As the network depth increases, we adopt a residual mechanism (He et al., 2016) to address network degradation issues by using skip connections to combine shallow and deep layer information. The residual structure $H(x) = F(x) + x$ allows for small weight updates when $F(x)$ is close to zero, and concatenating processed data during optimization can effectively improve efficiency. We improve the Conv shortcut by using convolutional operations to calculate $H(x) = F(x) + \rho(x)$, where $\rho(x)$ denotes the convolutional operation. As can be seen in part 2 of Fig. 2, we combine the small field of view information from the output of the 3×3 convolutional kernel calculation with the large field of view information from the output of the 3×3 convolutional kernel in part 1. This approach not only alleviates gradient problems, but also allows for richer feature combinations. Finally, we merge the outputs of each branch to obtain multi-scale features that effectively describe the disease image.

The training time for the Inception module is lengthy. To improve its performance, we incorporated four enhanced CNN modules in our experiment as detailed in Section 3.4. These modules produced varying numbers of feature maps. Specifically, in the enhanced CNN, module 1 produced 128 feature maps, module 2 produced 256 feature maps, module 3 produced 512 feature maps, and module 4 produced 1024 feature maps. For more information, please refer to Table 1.

2.3. Improved CBAM

In order to capture rice stress features more accurately, the convolutional block attention module was introduced and improved,

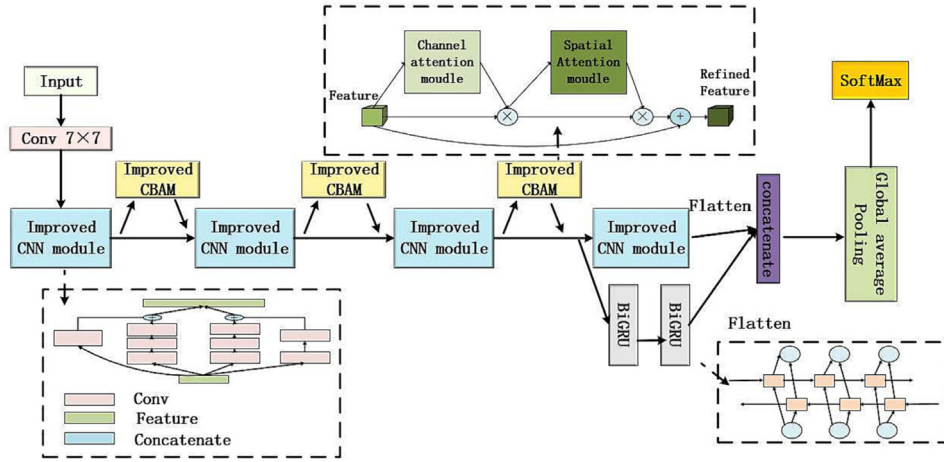


Fig. 1. Structure of Improved CNN-BIGRU.

which used the feature Map $F \in R_{c \times h \times w}$, as input to infer the 1-dimensional channel attention Map $M_c F \in R_{c \times 1 \times 1}$ and the 2-dimensional spatial attention Map $M_s F \in R_{1 \times h \times w}$. The feature map f received by the channel attention module, after maximum pooling and average pooling, obtains two different feature descriptions, F_{cAvg} and F_{cMax} , and then transferred to the multilayer perceptron (mlp) layer. The original channel attention module is shown in Fig. 3(a). The dimension reduction operation of the initial MLP layer will lose some feature information. In order to retain the features that are meaningful for diagnosis, this paper replaces it with three one-dimensional convolution filters to generate features. F_{cAvg} and F_{cMax} sums the elements and outputs the eigenvector to generate the channel note figure $M_c F \in R_{c \times h \times w}$, the improved channel attention module is shown in Fig. 3(b). The channel note diagram can be calculated according to Eq. (1).

$$M_c F = \rho(C1(C1(C1(Avgpool(F)))) + C1(C1(C1(Maxpool(F)))) \quad (1)$$

where ρ denotes the sigmoid function and C1 represents a one-dimensional convolutional layer.

Spatial attention is a complement to the Channel Attention module, in order to calculate spatial attention, max pooling and average pooling operations are applied along the channel attention axis, connecting them together to form a valid feature descriptor. Finally, the spatial attention is generated by convolutional $M_s F \in R_{h \times w}$ layers, which can be calculated by Eq. (2).

$$M_s F = \rho(f(7 \times 7)[maxpool, avgpool]) \quad (2)$$

where ρ denotes the sigmoid function, $f(7 \times 7)$ represents the convolutional kernel size 7×7 . The integration mode of improved CBAM in convolutional neural network is shown in Fig. 3(c).

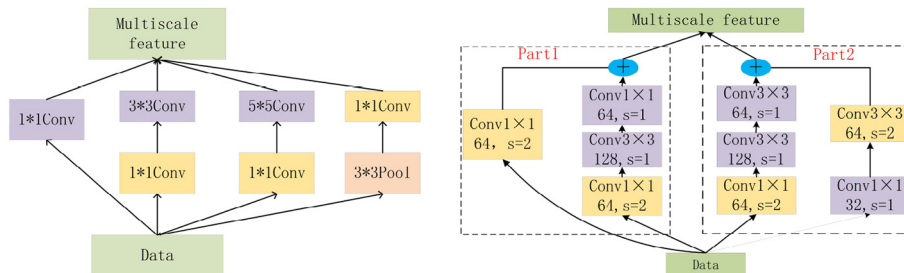


Fig. 2. The internal structure of improved CNN module.

2.4. Bidirectional gated recurrent neural network

The internal structure of GRU is shown in Fig. 4.

GRU obtains the gate state by the previous hidden state h_{t-1} and the current input x_t , the reset gate determines how the new input information is combined with the previous memory, which is expressed as Eq. (3):

$$r_t = \sigma(\omega_r \cdot [h_{t-1}, x_t]) \quad (3)$$

The update gate determines the amount of previous memory saved to the current time step, which is expressed as Eq. (4):

$$u_t = \sigma(\omega_u \cdot [h_{t-1}, x_t]) \quad (4)$$

where σ denotes the Sigmoid function, through which data can be transformed into values in the range of 0–1 to act as gating signals to control the amount of information flowing through the valve in all dimensions. ω_r and ω_u are the weights of the reset gate and the update gate, respectively. After obtaining the gated signal, the current input x_t and reset data are stitched together, and then activated by the tanh function to obtain the candidate hidden state of the moment h'_t at time t , which is expressed as an Eq. (5):

$$h'_t = \tanh(\omega \cdot [r_t \otimes h_{t-1}, x_t]) \quad (5)$$

The size of the $\omega \cdot [r_t \otimes h_{t-1}]$ determines the degree of combination of the current input x_t and historical information, the larger it is, the more information needs to be retained, the higher the degree of integration, conversely, the less information is retained.

The current hidden state h_t is expressed as Eq. (6):

$$h_t = (1 - u_t)h_{t-1} + u_t \otimes h'_t \quad (6)$$

Table 1
Structural parameters of improved CNN.

Layer	Kernel size	Output shape
Input		(None,224,224,3)
Conv	$7 \times 7, s = 2$	(None,112,112,64)
Improved Block1	$S = 2$	(None,56,56,128)
Improved Block2	$S = 2$	(None,28,28,256)
MaxPool	$3 \times 3, s = 2$	(None,14,14,512)
Improved Block3	$S = 2$	(None,7,71,024)
MaxPool	$3 \times 3, s = 1$	(None,14,14,512)
Improved Block4	$S = 2$	(None,7,71,024)
Conv	$1 \times 1, s = 1$	(None,7,71,024)
Global average pool		(None,1024)
softmax		(None,4)

The size of the u_t is positively correlated with the information to be remembered in interval $[0,1]$, $(1-u_t)h_{t-1}$ indicating selective forgetting the hidden state at the previous moment, $u_t \otimes h'_t$ indicating selective memory of the hidden state, h_t indicating forget certain types of the previous step information in h_{t-1} , and add the current new memory information in h'_t , thus forming the final memory within the unit.

Hidden states use reset gates to control how much information has been forgotten in the past, and update gates control how much new information needs to be received from historical states in the current state. This design addresses the gradient decay problem in recurrent neural networks and better captures dependency between

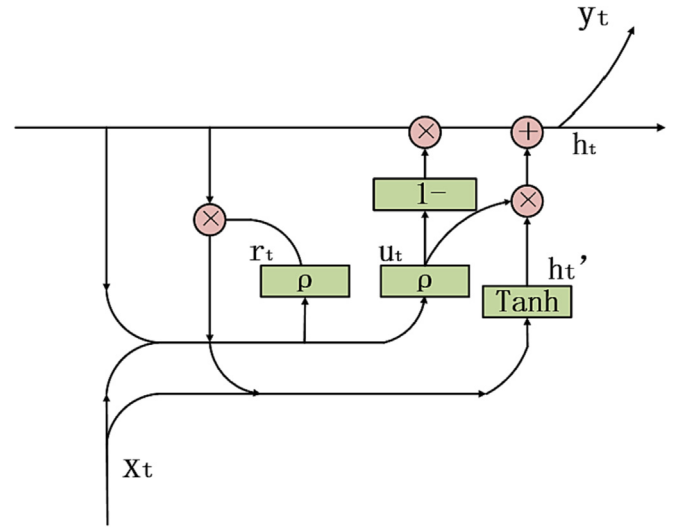
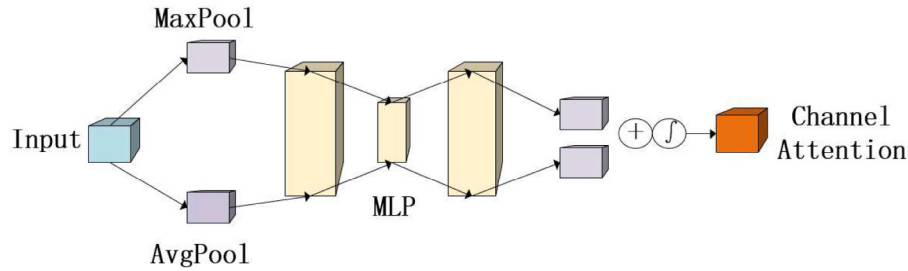
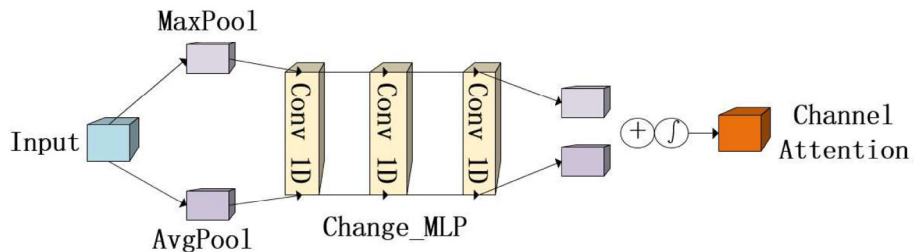


Fig. 4. The internal structure of the GRU unit.

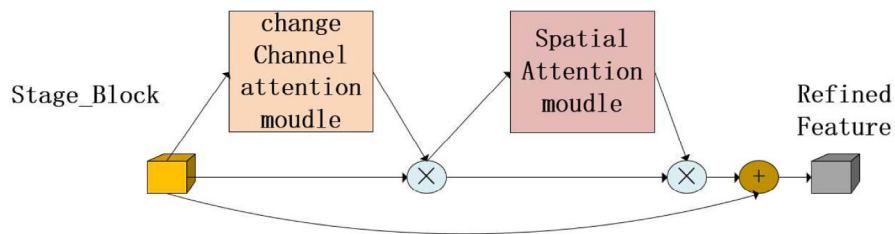
features in sequence data. The output information of the neural units of the last layer of the entire RNN network model is the most abundant, and we stitch it together as a secondary feature with the output feature of CNN.



(a).Original channel attention module.



(b).Improved channel attention module.



(c).CBAM is integrated with convolutional module.

Fig. 3. (a) is the original channel attention module, (b) is the improved channel attention module, and (c) is the integration of improved CBAM in convolutional neural network.

BiGRU consists of two sets of GRU that input coercive features from forward and backward, respectively, providing contextual global features for the neural network, with a structure shown in Fig. 5, where GRU1 represents the backward, GRU2 represents forward input.

2.5. Evaluation indicators

Using these indicators, they are *Accuracy*, *Precision*, *Recall*, and *F1* value for evaluation, and the formulas are equations such as Eqs. (7)–(10), respectively as shown.

$$\text{Accuracy} = \frac{TN + TP}{TP + TN + FP + FN} \quad (7)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (8)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (9)$$

$$F1 = \frac{2 * \text{Recall} * \text{Precision}}{\text{Precision} + \text{Recall}} \quad (10)$$

where *TP* denotes the number of true positive samples, *TN* denotes the number of true negative samples, *FP* denotes the number of false positive samples, and *FN* denotes the number of false negative samples.

3. Experiment

3.1. Experimental environment

This experiment was conducted on the Windows operating system (64-bit), GPU is NVIDIA RTX3050, CPU is Intel (R) Core (TM) i511400 (H) with 16GB of memory and all code is python, deep learning platform Tensorflow2.4.0.

3.2. Experimental data

Four plant diseases, namely, Rice blast, Sheath blight, Brownspot, and Leaf blight were collected. Experimental datasets were collected by laboratory personnel from Heilongjiang Bayi Agricultural University paddy fields (45°46'46"55" N, 124°19'125"12E) during the planting periods of June–August 2020 and 2021. The diversity of data sources can resist over fitting and strengthen the generalization ability of the model, this research combined public data with self-made data (Yang et al., 2023), proved that data from different sources is feasible and improved the generalization ability of model. The diversity of data obtained in the field can be enlarged by different devices, the images were captured using iPhone12, Redmi K30Pro, and Huawei P30Pro devices, with resolutions of 2532 × 1170, 2400 × 1080, and 2340 × 1080, respectively. The data was captured in well-lit environments, and the devices were placed approximately 20–30 cm away from the samples to ensure image clarity. After sorting and screening, a total of 2414 images were included in the dataset. Fig. 6. displays some example images.

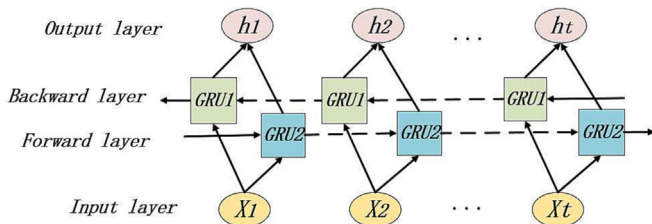


Fig. 5. The structure of the BiGRU.

3.3. Data preprocessing and data enhancement

To ensure standardization, we converted the collected rice images, which had varying resolutions and sizes, into 224 × 224 pixels for ease of training. These images were in RGB color format and had significant noise, making it difficult to extract image features. To overcome this challenge, we converted the images into the HSV color space, which describes colors based on hue, saturation, and value. The HSV color space is better suited for feature extraction, and thus we used this format for the images. To prevent overfitting during model training, we enhanced and expanded the dataset. This involved rotating some images, adjusting their width and height, randomly scaling their amplitude, and flipping some images horizontally. We first divided the dataset into training, validation, and testing sets in a ratio of 6:2:2 and then conduct augmentation. Table 2 shows the dataset size before and after expansion and Fig. 7 shows the distribution of the dataset.

3.4. Experimental parameters and results

The model training is divided into two stages. In the first stage, to find the optimal weight and scale of the improved CNN, we superimposed the proposed improved CNN module to carry out the test, and adopted the 10 cross-validation method to divide the data into ten parts, and gradually carried out, recording the results of each test, as shown in Table 3. The accuracy of the model increases with the depth, and the model degrades after it is stacked to the fifth module. If the residual structure is used to improve, the increase in depth will cause a large investment in computing resources, and the training speed of the wide network is slow due to it contains different branches. Therefore, according to the current data, the CNN module can be improved by adding four modules to obtain the optimal performance.

The second stage is to find the best parameters and the best data source acquisition location of BiGRU on the basis of the first stage. Transferring the original data to BiGRU will increase a lot of calculation costs. As mentioned above, improving the output of CNN module as a carrier of sensitive information removes some irrelevant information, and gradually discards low-sensitive information in the process of hidden layer processing, but it does not mean that these information is useless for decision-making. We introduced BiGRU to conduct separate independent tests, and gradually processed the output of the improved CNN modules 1 to 4 into one-dimensional vector form, and transferred it to BiGRU with different parameter settings for experiments. We spliced the output data and recorded the experimental results. The specific results are shown in Table 4.

It can be seen in table that the model achieves the highest accuracy when the network has 2 layers and 80 nodes. Compared to the improved CNN alone, the accuracy has increased by 4.09%, which confirms our hypothesis. Models that can successfully handle sequence data can be applied in the field of image recognition and can achieve good results. Compared to increasing the scale by stacking models, the cost of BiGRU is relatively small and can still achieve significant improvements. From the data, it is evident that as the number of nodes or layers increases, the model's accuracy improves, but after reaching a certain scale, it becomes challenging to improve performance. Next, we will concatenate and conduct experiments on the data of the improved CNN with improved CBAM and BiGRU. We propose an improved CNN-BiGRU model for rice disease image diagnosis, and the model parameters are set as shown in Table 5.

The proposed improved CNN-BiGRU model was trained, as shown in Fig. 8, which displays the accuracy and loss curves of the proposed method. The model was well-trained, with fast convergence and low loss rate. Our proposed approach is to improve CNN by combining it with BiGRU. The improved CNN module output was used for BiGRU computation, effectively capturing feature regularity information. The output features were concatenated with the CNN module output to

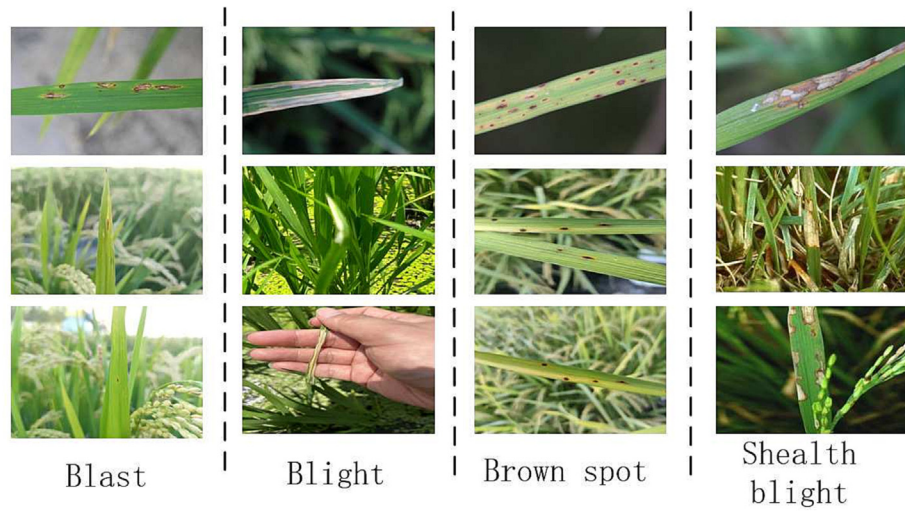


Fig. 6. Disease images data.

Table 2

The amount of rice disease data before and after data enhancement.

Disease type	Oringin	Expansion
Blast	783	1500
Blight	553	1500
Brown spot	632	1500
Shealth blight	446	1500
Total	2414	6000

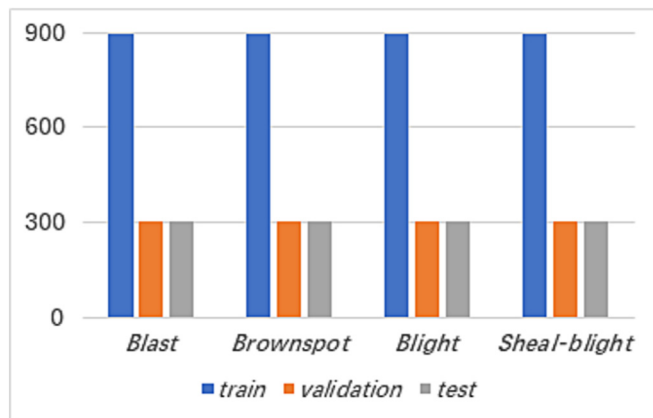


Fig. 7. Distribution of training set, validation set and test set.

overcome the limitations of CNN, and were transmitted to the classification layer for final diagnosis. The accuracy of the proposed model reached 97.42% after feature concatenation. We also compared our results with those of classical models that have achieved good results in rice disease diagnosis-related work, as shown in Fig. 9, which displays the performance of deep learning models in the rice dataset classification.

Different models were trained in the rice disease dataset, and the experimental results were compared. The 0, 1, 2, and 3 in the figure represent the disease categories. The experimental results show that the proposed method has more dense internal data points, so the similarity is higher and the best classification effect is achieved. Experiments show that in the experiment of rice disease data, the accuracy of the improved CNN-BiGRU fusion model proposed in this paper is higher than that of other classical deep learning models,

Table 3

Initial network scores.

Number of modules	Single CNN accuracy	With Improved CBAM accuracy
1	60.02%	60.56%
2	84.25%	85.31%
3	89.61%	90.36%
4	92.12%	95.26%
5	85.31%	89.21%
6	84.56%	91.23%

Table 4

Initial network scores.

Layers	Nodes	First moudle	Second moudle	Third moudle	Fourth moudle
1 layer	20	52.21%	55.12%	62.14%	66.11%
	40	62.14%	71.21%	85.22%	88.51%
	80	66.31%	75.21%	89.32%	91.67%
2 layers	20	58.32%	60.11%	70.01%	79.62%
	40	63.22%	65.54%	73.12%	84.52%
	80	75.22%	88.16%	96.21%	95.22%
3 layers	20	71.06%	80.64%	91.64%	92.01%
	40	70.21%	73.06%	92.61%	94.35%
	80	69.42%	74.32%	93.11%	95.31%

Table 5

Model parameters.

Parameters	Value
learning rate	0.001
Optimizer	Adam
Epoch	18
Batch size	20
Loss function	categorical crossentropy

which verifies the effectiveness of the improved method proposed in this study.

3.5. Ablation experiment

3.5.1. Ablation experiment of improved CBAM

We conducted ablation experiments on different CNN modules integrated with the improved CBAM attention mechanism to enhance the model's focus on disease spots. In the initial CBAM, the MLP layer

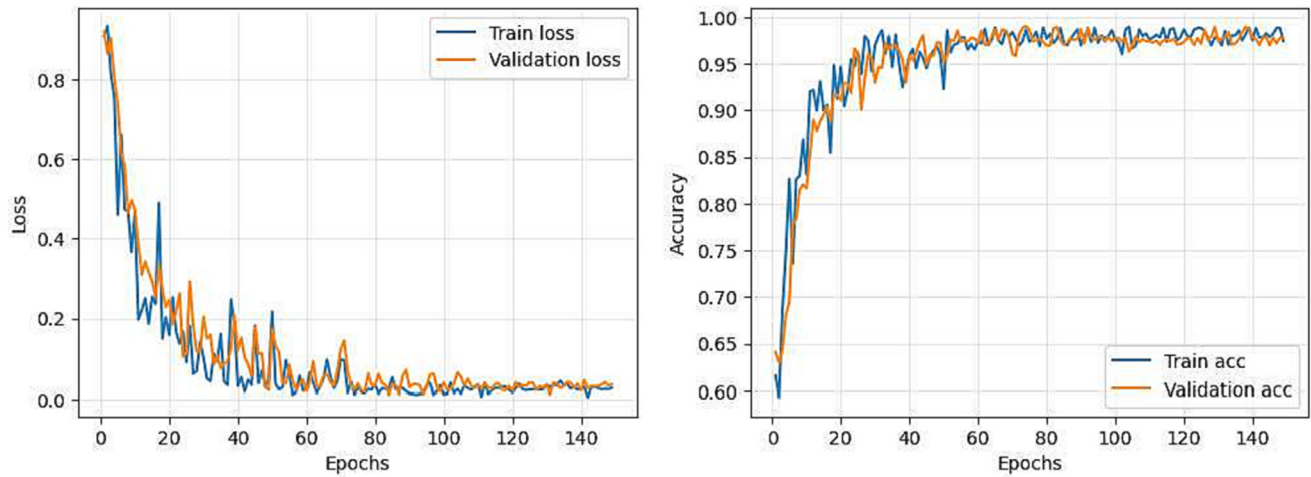


Fig. 8. Accuracy and loss rate curves of improved CNN-BiGRU.

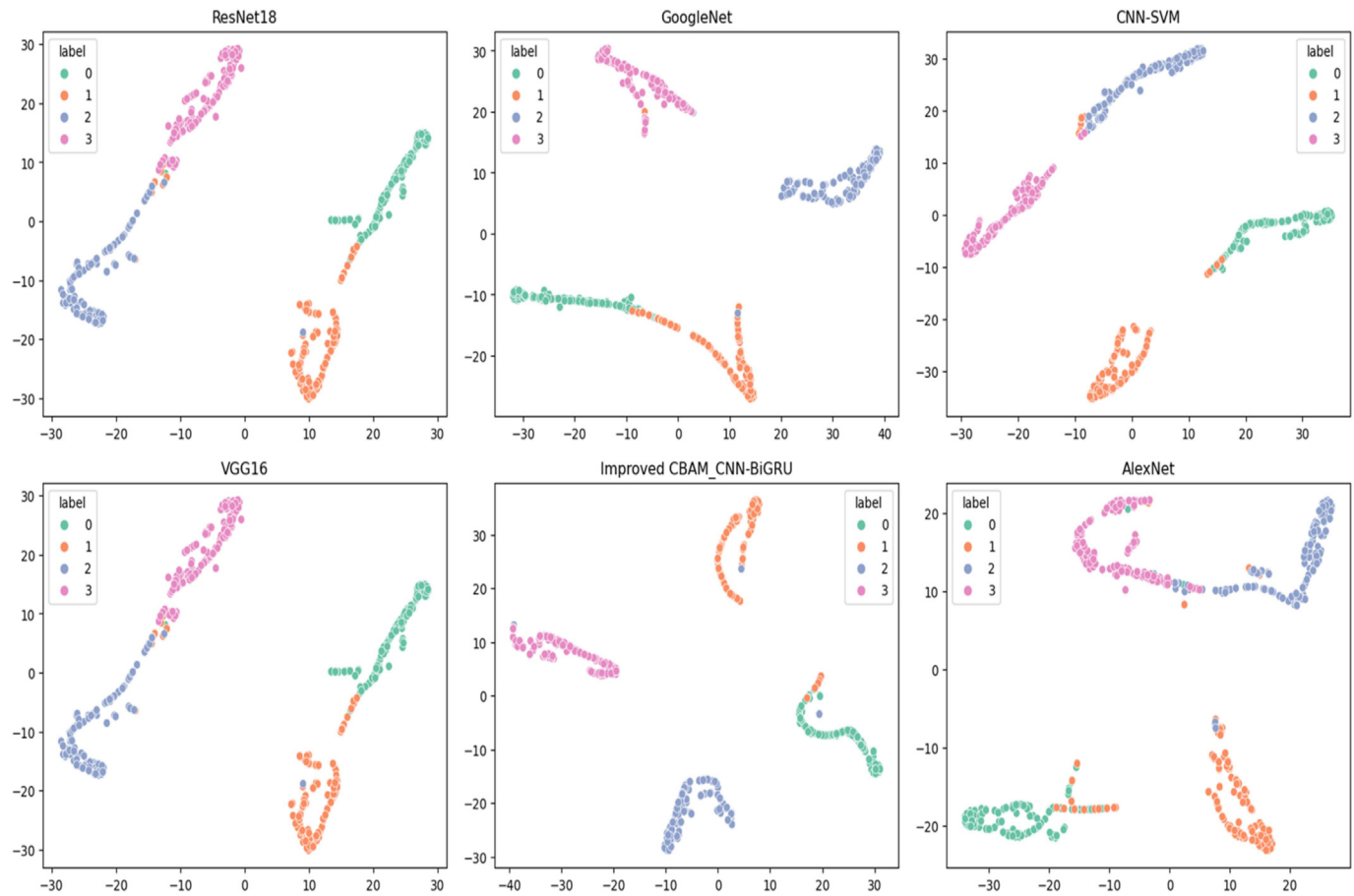


Fig. 9. Deep learning model performance in rice dataset.

extracts information through dimensionality reduction and downsampling, which may lead to the loss of some information, particularly for features with scattered spot layouts. We compared this improvement against the original method to study which aspects of attention mechanism bring performance improvements to the model. Table 6 provides the experimental data.

The data in the Table 6 suggests that the attention mechanism's impact on the shallow layers is subtle. Due to the processing of shallow features is not deep enough. As processing continues, higher-level features are gradually extracted, and the model strengthens its focus on this part.

Table 6
The ablation experimental data of CBAM in Improved CNN.

	First module	Second module	Third module	Accuracy
Improved CNN				92.12%
Improved CNN with CBAM	✓			92.23%
	✓	✓		92.79%
	✓	✓	✓	93.11%
Improved CNN with improved CBAM	✓			92.26%
	✓	✓		93.21%
	✓	✓	✓	93.96%

Table 7
Analysis of Recurrent neural network combined with improved CNN.

	GRU	BiGRU	LSTM	BiLSTM	Accuracy
Improved CNN	✓				95.11% \uparrow 2% 97.42% \uparrow 4.31%
		✓			94.96% \uparrow 2.84% 96.56% \uparrow 3.45%
			✓		92.56% \uparrow 2.33% 95.31% \uparrow 5.08%
				✓	91.26% \uparrow 1.03% 94.23% \uparrow 4% 93.36% \uparrow 2.13%
VGG16	✓				96.23% \uparrow 5% 92.96% \uparrow 1.73%
		✓			95.32% \uparrow 4.09% 94.63% \uparrow 1.32%
			✓		96.79% \uparrow 3.48% 93.89% \uparrow 0.58%
AlexNet	✓				96.23% \uparrow 2.52%
		✓			
			✓		
ResNet18	✓				
		✓			
			✓		
				✓	

In addition, improvements have weakened the impact of dimensionality reduction on the data, and this has been achieved through the use of one-dimensional convolution. This approach has effectively enhanced the impact of CBAM on CNN, resulting in improved performance.

3.5.2. Ablation experiment of BiGRU

As CNN lacks the ability to quantify the inter-relationship between features during the process of extracting local features, we introduced BiGRU to analyze this information. And also analyzed the impact of BiGRU on the experiment and conducted ablation experiments. We investigated the support provided by the recurrent neural network in analyzing the features extracted by CNN, while maintaining consistency in node numbers. Table 7 presents the experimental data.

From the data in the Table 7, it can be seen that in the case where the performance of the model has room for improvement, the recurrent neural network can effectively improve the performance of the model by analyzing the interdependence between features. Research on other deep learning models has proven the reliability of this method. GRU and LSTM have structural differences and perform differently in different tasks. Obviously, BiGRU is worth adopting in this task.

4. Discussion

4.1. Confusion matrix analysis

The performance of the improved CNN-BiGRU model in the test set is analyzed. The confusion matrix is shown in Fig. 10.

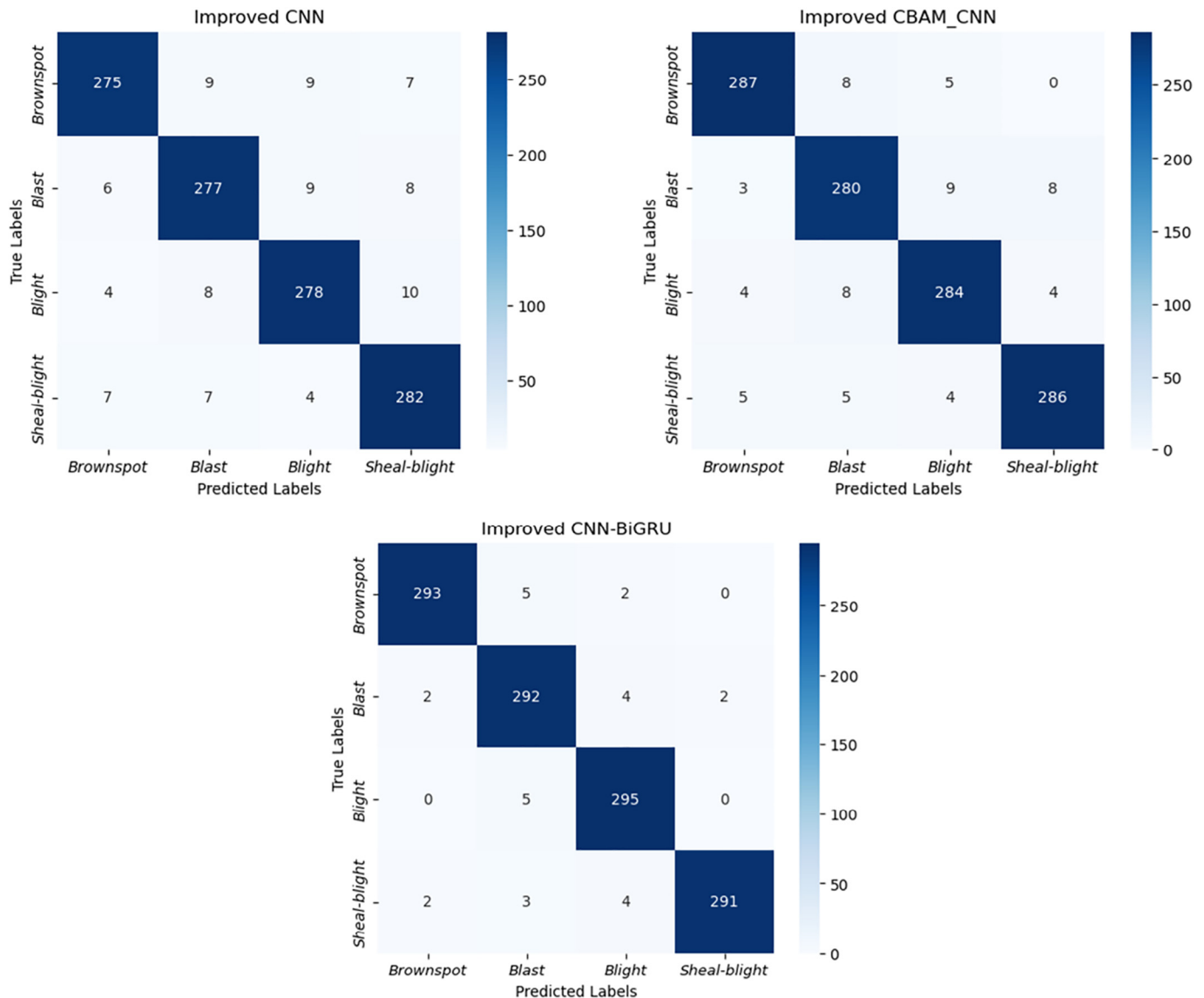


Fig. 10. The confusion matrix of the proposed method.

Table 8
Analysis of the number of model nodes.

	α	β	Accuracy
Improved CNN	0.75		93.15%↓2.11%
	1		95.26%
	1.25		95.29%↓0.03%
	1.5		95.36%↓0.1%
Improved CNN-BiGRU	1	0.75	96.12%↓2.09%
		1	97.26%
		1.5	98.21%↓0.95%
		2	98.20%↓0.94%

300 samples each of rice diseases: Brownspot, Rice blast, Blight and Sheal blight participated in test. Our improved CNN showed reduced misclassification rates for each disease 25, 23, 22, and 18 respectively, which further decreased to 13, 20, 16, and 14 after integrating the CBAM module. Notably, the CBAM module significantly improved recognition of sesame spot disease. By combining the BiGRU module to extract advanced features, misclassification rates were further reduced to 7, 8, 5, and 9 respectively. The results demonstrate the effectiveness of using RNNs to process features of rice diseases, thereby enhancing the performance of the CNNs.

4.2. Parameter analysis

The number of nodes in this paper is obtained through human exploration and experience, and there is still a certain degree of subjectivity. We will analyze the number of nodes in the model. We set α and β to represent the magnification factors of CNN and BiGRU respectively, and discuss the influence of the number of nodes on the model performance through scaling. In addition, we also adjust the size of the input data to explore the impact of this move on the model.

According to Table 8, the model's accuracy changes as the number of nodes changes. Increasing the number of nodes improves accuracy but comes at the cost of increased computation and training time. Effective methods must balance cost and performance. The model achieves the highest accuracy of 98.21% when α is maintained at 1 and β is maintained at 1.5.

Different input sizes yield features of varying granularity. Generally, larger inputs are believed to enhance model performance, but they come with inevitable computation cost increases. This cost can be significant, especially with large amounts of data. Therefore, it is necessary to explore an input size suitable for a particular task, such as the rice disease diagnosis model. Table 9 shows that size of inputs have significant impacts on model performance, but larger sizes inevitably come with huge costs. It is necessary to consider an input size that is suitable for the current task while being practical and cost-effective. In this paper, the effect of 224×224 is relatively ideal compared to others.

5. Conclusion

This paper proposed an improved CNN for extracting features of rice diseases by combining Inception and ResNet theories. The network addresses the high misidentification rate by incorporating information about the rice disease images and their surrounding context. The improved CNN also includes CBAM module for more precise feature extraction. Additionally, BiGRU was introduced to recognize the

relationships between stress image features, deepening the model's understanding of the images and the structural characteristics of rice diseases. The effectiveness of the proposed model is demonstrated through experiments, which show higher accuracy and lower cost compared to other models. However, the current dataset still has limitations, with only four types of rice disease samples. To address this, future work will focus on collecting more comprehensive real-field samples and improving the model structure to increase its generalization ability in complex scenarios.

Conflicts of Interest Statement

The authors declare that there is no conflict of interests.

CRediT authorship contribution statement

Yang Lu: Writing – original draft, Investigation, Software. **Xiaoxiao Wu:** Conceptualization, Supervision, Funding acquisition. **Pengfei Liu:** Methodology, Writing – review & editing, Project administration. **Hang Li:** Formal analysis, Visualization. **Wanting Liu:** Validation.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants U21A2019, 61873058 and 61933007, the Hainan Province Science and Technology Special Fund under Grant ZDYF2022-SHFZ105, Heilongjiang Natural Science Foundation of China under Grant LH2020F042 and the Scientific Research Starting Foundation for Post Doctor from Heilongjiang under Grant LBH-Q17134.

References

- Atole, R.R., Park, D., 2018. A multiclass deep convolutional neural network classifier for detection of common rice plant anomalies. *Int. J. Adv. Comput. Sci. Appl.* 9 (1). <https://doi.org/10.14569/IJACSA.2018.090109>.
- Bao, W., Yang, X., Liang, D., Hu, G., Yang, X., 2021. Lightweight convolutional neural network model for field wheat ear disease identification. *Comput. Electron. Agric.* 189, 106367. <https://doi.org/10.1016/j.compag.2021.106367>.
- Barbedo, J., 2018. Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. *Comput. Electron. Agric.* 153, 46–53. <https://doi.org/10.1016/j.compag.2018.08.013>.
- Chen, J., Chen, J., Zhang, D., Sun, Y., Nanekharan, Y.A., 2020. Using deep transfer learning for image-based plant disease identification. *Comput. Electron. Agric.* 173, 105393. <https://doi.org/10.1016/j.compag.2020.105393>.
- Chung, J., Gulcehre, C., Cho, K., Bengio, Y., 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*. <https://arxiv.org/pdf/1412.3555>.
- Da Costa, A.Z., Figueroa, H.E., Fracarolli, J.A., 2020. Computer vision based detection of external defects on tomatoes using deep learning. *Biosyst. Eng.* 190, 131–144. <https://doi.org/10.1016/j.biosystemseng.2019.12.003>.
- Ferentinos, K.P., 2018. Deep learning models for plant disease detection and diagnosis. *Comput. Electron. Agric.* 145, 311–318. <https://doi.org/10.1016/j.compag.2018.01.009>.
- Graves, A., Graves, A., 2012. Long short-term memory. *Supervised Sequence Labelling with Recurrent Neural Networks*, 37–45. <https://doi.org/10.1162/neco.1997.9.8.1735>.
- He, W., Huang, D., Liu, C., Cen, Z., Zhang, Y., Ma, Z., Li, R., 2010. Genetic analysis of resistance to bacterial leaf streak in common wild rice. *J. Plant Pathol.* 40 (180–185). <https://doi.org/10.13926/j.cnki.apps.2010.02.012>.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. *Conference on Computer Vision and Pattern Recognition*, pp. 770–778. <https://arxiv.org/abs/1512.03385>.
- Jiang, F., Lu, Y., Chen, Y., Cai, D., Li, G., 2020. Image recognition of four rice leaf diseases based on deep learning and support vector machine. *Comput. Electron. Agric.* 179, 105824. <https://doi.org/10.1016/j.compag.2020.105824>.
- Khush, G.S., 1997. Origin, dispersal, cultivation and variation of rice. *Plant Mol. Biol.* 35, 25–34. <https://doi.org/10.1023/A:1005810616885>.
- Majumdar, D., Kole, D.K., Chakraborty, A., Majumder, D.D., 2015. An integrated digital image analysis system for detection, recognition and diagnosis of disease in wheat leaves. *Proceedings of the Third International Symposium on Women in Computing and Informatics*, pp. 400–405. <https://doi.org/10.1145/2791405.2791474>.
- Peng, S., Tang, Q., Zou, Y., 2009. Current status and challenges of rice production in China. *Plant Prod. Sci.* 12 (1), 3–8. <https://doi.org/10.1626/pps.12.3>.
- Peng, W., Li, K., Zeng, Y., 2015. Research progress on microbial control of rice diseases. *J. Jiangxi Agric. Univ.* 37 (4), 625–631. <https://doi.org/10.13836/j.jjau.2015096>.
- Rahman, C.R., Arko, P.S., Ali, M.E., Khan, M.A.I., Apon, S.H., Nowrin, F., Wasif, A., 2020. Identification and recognition of rice diseases and pests using convolutional neural

Table 9
Experimental data with different input sizes.

Inputsize	Accuracy	Time
150 × 150	96.95%	2h10min
224 × 224	98.21%	3h45min
299 × 299	98.23%	7h3min

- networks. *Biosyst. Eng.* 194, 112–120. <https://doi.org/10.1016/j.biosystemseng.2020.03.020>.
- Rumpf, T., Mahlein, A.-K., Steiner, U., Oerke, E.-C., Dehne, H.-W., Plümer, L., 2010. Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. *Comput. Electron. Agric.* 74 (1), 91–99. <https://doi.org/10.1016/j.compag.2010.06.009>.
- Sethy, P.K., Barpanda, N.K., Rath, A.K., Behera, S.K., 2020. Deep feature based rice leaf disease identification using support vector machine. *Comput. Electron. Agric.* 175, 105527. <https://doi.org/10.1016/j.compag.2020.105527>.
- Sun, G., Du, X., Rongxiang, T., Sun, S., 1998. Strategies for controlling rice blast and prospects for research in the 21st century. *J. Plant Pathol.*, 289–292. <https://doi.org/10.13926/j.cnki.apps.1998.04.001>.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. *Conference on Computer Vision and Pattern Recognition*, pp. 1–9. <https://arxiv.org/abs/1409.4842>.
- Too, E.C., Yujian, L., Njuki, S., Yingchun, L., 2019. A comparative study of fine-tuning deep learning models for plant disease identification. *Comput. Electron. Agric.* 161, 272–279. <https://doi.org/10.1016/j.compag.2018.03.032>.
- Waheed, A., Goyal, M., Gupta, D., Khanna, A., Hassanien, A.E., Pandey, H.M., 2020. An optimized dense convolutional neural network model for disease recognition and classification in corn leaf. *Comput. Electron. Agric.* 175, 105456. <https://doi.org/10.1016/j.compag.2020.105456>.
- Wang, H., Chen, D., Li, C., Tian, N., Zhang, J., Xu, J.-R., Wang, C., 2019. Stage-specific functional relationships between tub1 and tub2 beta-tubulins in the wheat scab fungus *Fusarium graminearum*. *Fungal Genet. Biol.* 132, 103251. <https://doi.org/10.1016/j.fgb.2019.103251>.
- Woo, S., Park, J., Lee, J.-Y., Kweon, I.S., 2018. Cbam: convolutional block attention module. *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 3–19. <https://arxiv.org/abs/1807.06521>.
- Yang, H., Ni, J., Gao, J., Han, Z., Luan, T., 2021. A novel method for peanut variety identification and classification by improved vgg16. *Sci. Rep.* 11 (1), 15756. <https://www.nature.com/articles/s41598-021-95240-y>.
- Yang, L., Yu, X., Zhang, S., Long, H., Zhang, H., Xu, S., Liao, Y., 2023. Googlenet based on residual network and attention mechanism identification of rice leaf diseases. *Comput. Electron. Agric.* 204, 107543. <https://doi.org/10.1016/j.compag.2022.107543>.
- Zhu, Y., Leung, H., Chen, H., Wang, Y., Tang, K., Zhao, X., Zhou, J., Tu, J., Li, Y., He, X., Zhou, J., Sun, Y., Twng, W., 2004. Sustainable control of rice diseases by using disease resistance gene diversity. *China Agric. Sci.* 37 (6), 832–839.