



OPEN

## A classification method for soybean leaf diseases based on an improved ConvNeXt model

Qinghai Wu<sup>1,2,7</sup>, Xiao Ma<sup>1,2,7</sup>, Haifeng Liu<sup>3</sup>, Cunguang Bi<sup>4</sup>, Helong Yu<sup>4</sup>, Meijing Liang<sup>5</sup>, Jicheng Zhang<sup>6</sup>, Qi Li<sup>1</sup>, You Tang<sup>1,3</sup>✉ & Guanshi Ye<sup>1</sup>✉

Deep learning technologies have enabled the development of a variety of deep learning models that can be used to detect plant leaf diseases. However, their use in the identification of soybean leaf diseases is currently limited and mostly based on machine learning methods. In this investigation an enhanced deep learning network model was developed to recognize soybean leaf diseases more accurately. The improved network model consists of three parts: feature extraction, attention calculation, and classification. The dataset used was first diversified through data augmentation operations such as random masking to enhance network robustness. An attention module was then used to generate feature maps at various depths. This increased the network's focus on discriminative features, reduced background noise, and enabled the use of the LeakyReLU activation function in the attention module to prevent situations in which neurons fail to learn when the input is negative. Finally, the extracted features were then integrated using a fully connected layer, and the predicted disease category inferred to improve the classification accuracy of soybean leaf diseases. The average recognition accuracy of the improved network model for soybean leaf diseases was 85.42% both higher than the six deep learning comparison models (ConvNeXt (66.41%), ResNet50 (72.22%), Swin Transformer (77.00%), MobileNetV3 (67.27%), ShuffleNetV2 (59.89%), and SqueezeNet (72.92%)), thus proving the effectiveness of the improved method. The model proposed in this paper was also tested on the grapevine leaf dataset, and the performance ability of the improved network model remained due to other common network models, and overall the proposed network model was very effective in leaf disease identification.

Soybean is one of the most widely cultivated legumes, as it can be processed into various foods and is an important global feed crop<sup>1</sup>. Disease is one of the main factors affecting soybean yields, with infected soybeans generally reducing yields by 10–30%, or more than 50% in severe cases. In particular, soybean leaf diseases can alter leaf color and induce leaf loss which when severe can lower their disease resistance, thus decreasing crop yield and quality. Common and serious leaf diseases include soybean leaf spots and soybean rust. According to treatment plans, appropriate actions should be taken at the earliest possible stages of disease occurrence<sup>2</sup>. It is thus essential that we detect such diseases in a timely and accurate manner. Traditional disease identification requires manual intervention, and due to the complexity and diversity of leaf diseases, a great deal of experience is required to accurately identify the disease type, which makes it a time-consuming and labor-intensive process<sup>3</sup>. Numerous algorithms have been used to identify plant leaf diseases owing to the development of machine learning and neural networks<sup>4</sup>. Currently, there are two primary approaches for plant leaf disease classification tasks: leaf disease recognition based on traditional machine learning algorithms and disease recognition based on deep learning.

Feature extraction and classification based on conventional machine learning algorithms is one of the traditional methods used to categorize plant leaf diseases. The benefits of this approach include a quick recognition speed and minimal hardware requirements. Panigrahi et al.<sup>5</sup> compared and analyzed traditional machine learning algorithms such as naïve Bayes (NB), decision tree (DT), K-nearest neighbor (KNN), support vector

<sup>1</sup>Electrical and Information Engineering College, Jilin Agricultural Science and Technology University, Jilin 132101, Jilin, China. <sup>2</sup>School of Information and Control Engineering, Jilin Institute of Chemical Technology, Jilin 132022, China. <sup>3</sup>College of Agriculture, Yanbian University, Yanji 133002, Jilin, China. <sup>4</sup>College of Information Technology, Jilin Agricultural University, Changchun 132101, Jilin, China. <sup>5</sup>Department of Crop and Soil Sciences, Washington State University, Pullman, WA 99164, USA. <sup>6</sup>College of Electronic and Information, Northeast Agricultural University, Harbin, Heilong 150030, Jiang, China. <sup>7</sup>These authors contributed equally: Qinghai Wu and Xiao Ma. ✉email: tangyou@jlnku.edu.cn; yeguanshi@163.com

machine (SVM), and random forest (RF) for maize disease detection. Random forest had the highest accuracy when compared to the other algorithms. Rahaman et al.<sup>6</sup> proposed a k-nearest neighbor (KNN) classifier to detect and classify plant leaf diseases using texture features extracted from disease images, which was successful in detecting and identifying the selected diseases with 96.76% accuracy. The general steps of image classification using machine-learning algorithms are image preprocessing, feature extraction, and classifier training. Among these, feature extraction is the most crucial step as it directly impacts on the accuracy of the classification result. Traditional feature extraction relies mainly on manual design, which can make feature extraction challenging<sup>7</sup>.

With the development of neural networks and deep learning<sup>8</sup> researchers have shifted their focus away from automated feature extraction using neural networks. Lin et al.<sup>9</sup> proposed the GrapeNet lightweight model to identify the different stages of grape leaf disease, which addresses the problem of diverse disease morphologies and leads to increased recognition accuracy and reduced parameter quantity. Bansal et al.<sup>10</sup> combined three pre-trained models to identify apple tree leaves and achieved satisfactory classification results. Haque et al.<sup>11–14</sup> conducted experiments on common maize diseases, using GoogLeNet, Inception V3, and designing a new CNN network to classify and assess the severity of common maize diseases, respectively, and the proposed methods achieved good results in the identification of maize leaf diseases, and achieved certain results in the lightweighting of the model.

To date, however, studies on soybean leaf disease recognition have been limited<sup>15–17</sup>, with the majority focusing on the images of soybean leaf disease captured against a single background, and mostly using machine learning methods. In this study, soybean leaf images with complex background information were used as the training, validation, and testing samples, and an improved ConvNeXt model-based soybean leaf disease image classification algorithm was proposed that addresses the issue of complex background interference typically present in real collections. The main findings of this study can be summarized as follows: (1) by performing image augmentation operations, such as random masking, on the dataset before training the network, the dataset was enriched to improve model robustness; (2) attention mechanisms were included at various depths to enable the ConvNeXt network to focus on discriminative features and reduce background interference; (3) the LeakyReLU activation function was used in the attention module to avoid situations in which the neurons did not learn when the input was negative; and (4) the improved network model was compared to existing models under the same conditions to validate its disease classification ability.

## Materials and method

### Data augmentation

The experiment was conducted to investigate soybean leaf spot, soybean rust, and healthy leaves. Soybean leaf spot disease initially appears as watery spots on the leaves, then turns brown or black with a yellow-green halo around the edges. Soybean rust is caused by the phakopsora pachyrhiz and usually spreads upward from the lower leaves, with small yellow spots appearing on the leaves in the early stages of the disease, and then the spots expand slightly. Pictures displaying each class of disease are shown in Table 1.

Due to the variations in weather conditions and leaf-covered lesions identified when collecting disease images, the generalization ability and robustness of the model were poor during soybean disease classification.

Name	Leaf Spot	Rust	Healthy
Samples			
			
			

**Table 1.** Samples of soybean leaf disease.

To improve the classification accuracy of the model, as well as its generalization ability and robustness, a regularization method for image data augmentation is typically used<sup>18</sup>. To simulate the different angles, background leaf occlusions, and weather conditions encountered during image acquisition, this study used data augmentation methods such as rotation, Gaussian blur, adding random noise, adding random occlusion at random positions, and brightness adjustment. This prevented overfitting and improved the robustness and generalization of the model. The “Smart Agriculture” platform dataset from Jilin Agricultural Science and Technology College (Collected from soybean plantations at Jilin College of Agricultural Science and Technology, Jilin Province, China) was used as the experimental objects. The original dataset was split into training, validation, and testing sets using a ratio of 6:2:2 to prevent information leakage. A total of 1296 images were used before image augmentation, and 11655 images were used after data augmentation. Detailed dataset information is presented in Table 2, and typical data augmentation effects are shown in Fig. 1.

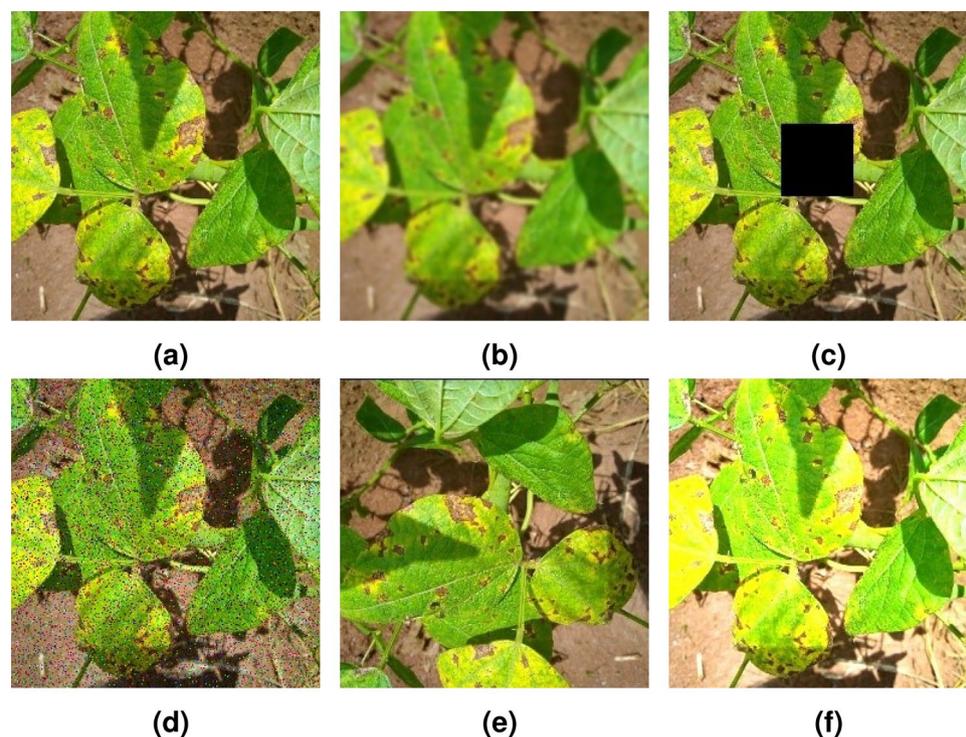
### Attention mechanism

The human visual system serves as the inspiration for the attention mechanism as attention can quickly be focused on the key elements of a scene, allocating more neural network computational resources to critical tasks when processing complex information<sup>19</sup>. By using backpropagation to guide the attention module, important features can be identified using parameter updates, enabling efficient and accurate task completion. Attention mechanisms have been widely used in a variety of fields<sup>20–23</sup>.

Common attention mechanisms include Squeeze and excitation (SE-Net), efficient channel attention (ECA-Net), and the convolutional block attention modules (CBAM). SE-Net<sup>24</sup> explicitly models the interdependencies between feature channels; it automatically acquires the importance of each channel through learning. ECA-Net<sup>25</sup> proposes a non-diminished local cross-channel interaction strategy and an adaptive method for selecting one-dimensional convolution kernel size. CBAM<sup>26</sup> is a combined spatial and channel attention mechanism module, which can well obtain the attention information between space and channel, it includes two parts: spatial attention module (SAM) and channel attention module (CAM), which are used to collect the attention information

Name	Original images	After data enhancement
Leaf spot	432	3,888
Rust	436	3,924
Healthy	428	3,843
Total	1,296	11,655

**Table 2.** Details of the soybean leaf disease dataset.



**Figure 1.** Data enhancement effect. (a) Original images, (b) Gaussian blur, (c) Random occlusion, (d) Add noise, (e) Rotate, (f) Brightness adjustment.

from both space and channel respectively. It is this attentional mechanism that is used in this study. CBAM's overall structure of the CBAM is shown in Fig. 2. The computation formula of the channel attention module (CAM) is shown in Eq. 1:

$$CAM(F) = \sigma(W_1(W_0(F_{avg}^c)) + W_1(W_0(F_{max}^c))) \tag{1}$$

where  $\sigma$  describes the sigmoid activation function,  $F_{avg}$  and  $F_{max}$  describe the output results of global average pooling and global maximum pooling, respectively, and  $W_0$  and  $W_1$  describe two different neural network operations in two layers.

The computation formula of the spatial attention module (SAM) is shown in Eq. 2:

$$SAM(F) = \sigma(f^{7 \times 7}(F_{avg}^s; F_{max}^s)) \tag{2}$$

where  $f^{7 \times 7}$  describes the convolution operation with a kernel size of  $7 \times 7$ , and  $[\ ]$  describes the channel concatenation operation.

The overall CBAM process can be described by Eqs. 3 and 4:

$$F' = CAM(F) \otimes F \tag{3}$$

$$F'' = SAM(F') \otimes F' \tag{4}$$

where the input feature is  $F(F \in \{R^{C \times H \times W}\})$ , the output of the channel attention module is  $CAM(CAM \in \{R^{C \times 1 \times 1}\})$ , the output of the spatial attention module is  $SAM(SAM \in \{R^{1 \times H \times W}\})$ , and the output results of the channel and spatial attention are denoted as  $F'$  and  $F''$  respectively. This study uses an attention module that combines channel and spatial attention.

### Improved ConvNeXt architecture

The ConvNeXt<sup>27</sup> model used in this experiment was based on ResNet50<sup>28</sup> and was improved by integrating the ideas of the Swin Transformer<sup>29</sup>. The improved ConvNeXt (CBAM-ConvNeXt) structure proposed in this study (Fig. 3) consists of ConvNeXt Block modules for feature extraction (Fig. 4), downsampling modules, and attention modules to eliminate complex background interference. Typical attention mechanisms include SENet, ECA-Net, selective kernel networks (SK-Net)<sup>30</sup> and CBAM. In this experiment, a comparative analysis was conducted using the SENet, ECANet, SKNet, and CBAM attention modules. CBAM was selected as the attention module. CBAM is an attention mechanism that combines channel and spatial attention. It maps the extracted intermediate features to the channel and spatial dimensions for attention analysis. The obtained attention scores are multiplied by the input of intermediate feature maps to obtain feature maps with added attention, which are then processed by the next convolution operation.

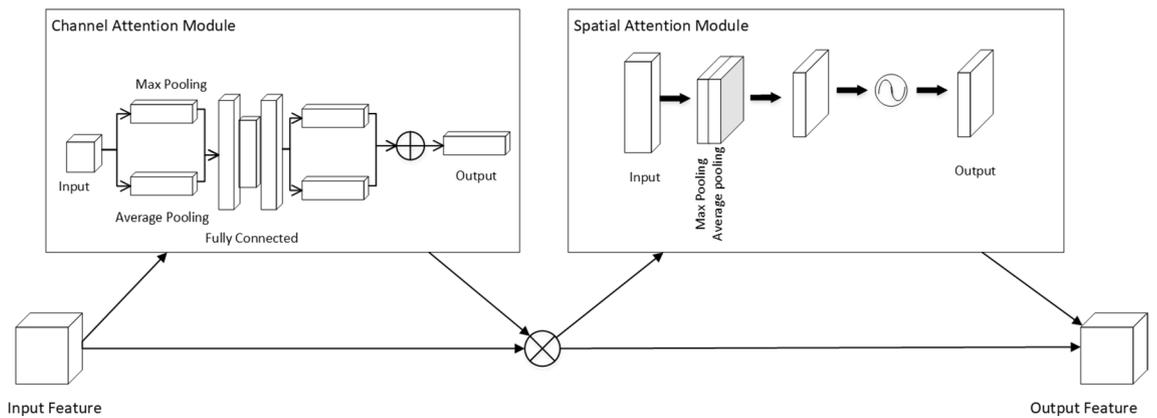


Figure 2. CBAM monolithic construction.

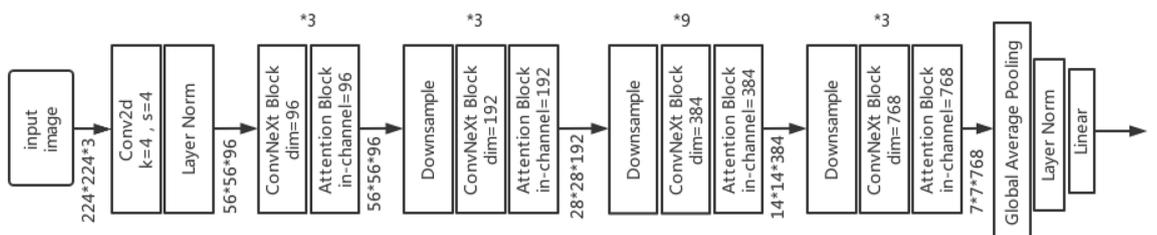
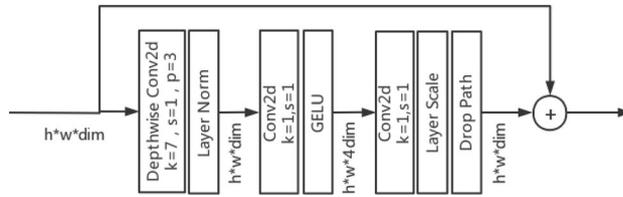


Figure 3. Overall structure of the CBAM-ConvNeXt.



**Figure 4.** ConvNeXt Block structure.

The LeakyReLU<sup>31</sup> activation function was used in the CBAM attention module to overcome the problem in which neurons do not learn when the input is negative by correcting its nonlinear unit. The LeakyReLU activation function formula is shown in Eq. 5:

$$LeakyReLU = \begin{cases} x, & x > 0 \\ ax, & x \leq 0 \end{cases} \quad (5)$$

The improved ConvNeXt model first extracted shallow features from 224×224 three-channel colored soybean leaf disease images using a 4×4 convolutional operation with a stride of four. A 56×56 feature map with 96 channels was created using a normalization layer. The formulae for calculating the width and height of the feature map obtained after the convolutional operation are shown in Eqs. 6 and 7:

$$H = (h - k + 2p) / s + 1 \quad (6)$$

$$W = (w - k + 2p) / s + 1 \quad (7)$$

where H and W are the height and width of the feature map after convolution, respectively, h and w are the height and width of the input feature map, respectively, k is the size of the convolution kernel, p is the padding size, and s is the stride of the convolution. Next, four ConvNeXt blocks, four attention blocks, and three downsampling modules were used to extract features, add attention scores, and downsample feature maps, enabling the network to focus on disease features and reduce its attention on the complex backgrounds, thereby reducing interference.

The softmax function was used as the output of the model to calculate the predicted class of the target. The definition of softmax is given in Eq 8:

$$Y(P) = P(y = p|x, \theta_p) = \frac{e^{x^T \cdot \theta_p}}{\sum_{p=1}^c e^{x^T \cdot \theta_p}} \quad (8)$$

### Evaluation metrics

The cross-entropy loss function was used as the standard for network optimization, and the Adam optimizer was used to optimize the model parameters. The expression for the cross-entropy loss function is given by Eq. 9:

$$\min_{\{W, b; \theta\}} \varepsilon(W, b; \theta) = \frac{1}{N} \sum_{n=1}^N \left[ - \sum_{p=1}^N \gamma(y_n = p) \cdot \log P(y = p|x, \theta_p) \right] + \lambda R(\theta) \quad (9)$$

where C are the number of categories, N are the number of samples,  $\gamma$  are the Dirichlet function, the parameters  $\theta = (\theta_1, \theta_2, \dots, \theta_k)$ ,  $R(\bullet)$  are the regularization constraint term, and  $\lambda$  are the regularization factor.

This experiment used accuracy, precision, recall, and  $f_1$  - score to evaluate model performance. Accuracy describes the overall accuracy of the model's predictions; however, in cases of imbalanced datasets, it may not be a good metric for evaluating model performance. Precision describes the accuracy of positive samples predicted by the model, whereas recall describes the probability of correctly predicting positive samples among all positive samples. The accuracy, precision, recall, and  $f_1$  - score for the binary classifications were defined using Eqs. 10–13:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

$$Precision = \frac{TP}{TP + FP} \quad (11)$$

$$Recall = \frac{TP}{TP + FN} \quad (12)$$

$$f_1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (13)$$

True positive (TP) describes the number of positive samples predicted correctly, false positive (FP) describes the number of negative samples predicted as positive, true negative (TN) describes the number of negative samples predicted correctly, and false negative (FN) describes the number of positive samples predicted as negative.

A confusion matrix was adopted to display the classification results of the model, where each column describes the predicted label category, and each row describes the true label category of the data. The more concentrated the data are on the diagonal, the better the classification performance of the model.

## Results and discussion

Python 3.7 was used as the programming language, and the PaddlePaddle 2.3.2 deep learning framework, was adopted for this experiment. The training process used an accelerated 4-core CPU and Tesla V100 GPU. The network was trained using cross-entropy loss combined with an Adam optimizer<sup>32</sup>, which can adaptively adjust the learning rate based on the training parameters. The network was trained for 100 iterations with a batch size of 64 and learning rate of 0.000001.

To verify the performance of the improved CBAM-ConvNeXt model, two comparative experiments were conducted: a comparison of the effects of training the improved model with the augmented dataset versus the original dataset, and a performance comparison between the improved model and common traditional classification models. In order to validate the classification ability of the network model proposed in this study on other datasets, this paper also conducts a comparison between the proposed model and other network models on the grape leaf dataset.

### Data augmentation experiment results

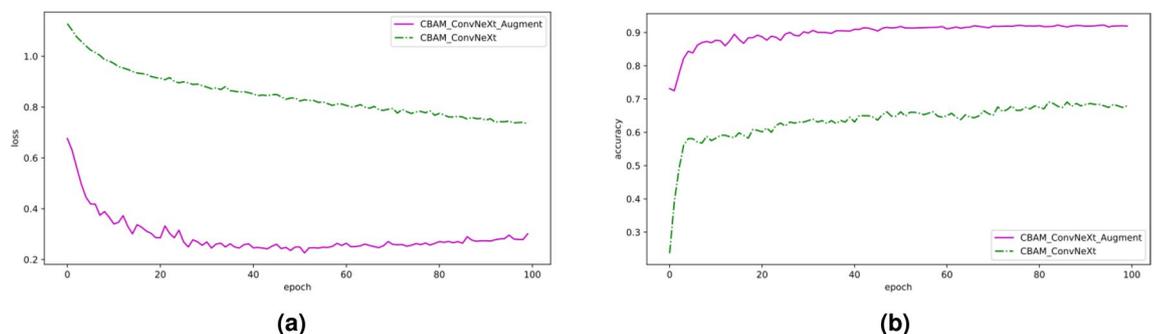
To verify the improvements in model performance that occur when using the data augmentation method proposed in this paper, the original and augmented datasets were separately fed into the CBAM-ConvNeXt model for training. A comparison of the loss value and accuracy of the validation set during the model training process is shown in Fig. 5. The confusion matrix of the classification performance on the test set is presented in Fig. 6, where larger values and darker colors on the diagonal of the confusion matrix indicate better model classification performance.

The model trained using the augmented dataset was improved in terms of loss values and classification accuracy, both for convergence speed and stable final values, when compared with the model trained using the original dataset (Fig. 9). A comparison of the experimental results before and after data augmentation is presented in Table 3. The data show that the model trained using the augmented dataset had improved accuracy, precision, recall, and f1-score on the test set when compared to the model trained using the original dataset. The improvement of the model's ability to classify diseases depends on the fact that the data-enhanced training set effectively avoids the generation of overfitting situations during the training process by adding interference information, which improves the robustness of the model. In summary, the data augmentation method used in this study could improve the generalizability and robustness of the network model during the data pre-processing stage, which has a positive effect on the classification of soybean leaf diseases.

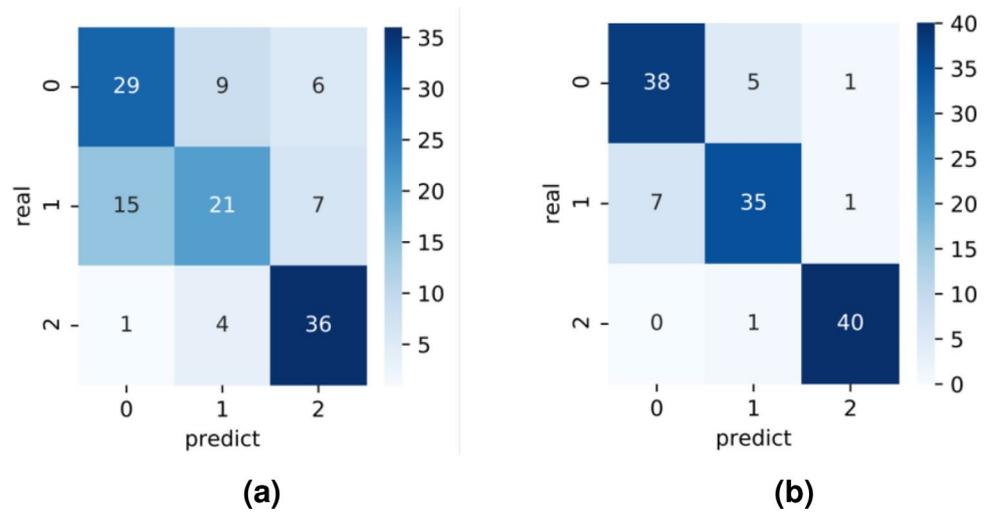
### Model improvement experimental results

To verify the effectiveness of the improved model for soybean leaf disease classification, the enhanced dataset was input into the original ConvNeXt, ResNet50, and swine transformer models for training. Comparison curves of the validation loss and accuracy were obtained (Fig. 7). Compared to the other three models, the improved network model showed higher accuracy and loss values and faster convergence due to fuller access to spatial and inter-channel feature information during training. In addition, several common classification models were added as controls. The experimental results for each model are listed in Table 4.

From Table 4, it can be concluded that the proposed model is effective in improving the classification accuracy of the model compared to traditional deep learning classification models in complex contexts with guaranteed



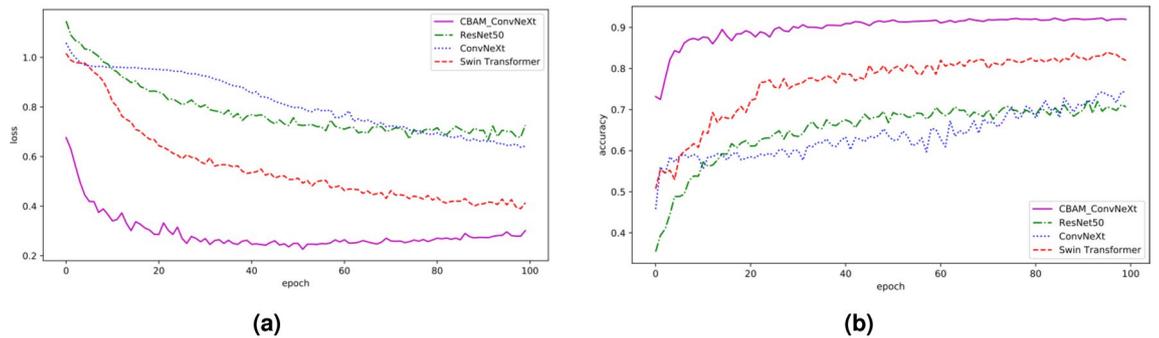
**Figure 5.** Loss value and accuracy curve before and after data enhancement. **(a)** Loss curves before and after data augmentation, **(b)** Accuracy curves before and after data augmentation.



**Figure 6.** Confusion matrix comparison before and after data enhancement. (a) Original dataset, (b) After data augmentation.

Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Original dataset	67.19	66.56	67.52	66.57
After data augmentation	88.28	88.35	88.44	88.37

**Table 3.** Comparison of experimental effects before and after data enhancement.



**Figure 7.** Loss values and accuracy curves for each model. (a) Loss curves for each model, (b) Accuracy curves for each model.

Models	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Total params	Training time (s)
ResNet50	72.22	72.32	72.26	72.15	23,567,299	7,820.79
ConvNeXt	66.41	65.97	66.52	66.14	27,822,435	9,290.13
Swin transformer	77.00	76.71	77.12	76.48	27,521,661	8,772.59
CBAM-ConvNeXt	85.42	85.53	85.46	85.87	28,162,455	8,707.11
MobileNetV3	67.27	67.12	67.33	67.19	4,230,275	6,682.45
ShuffleNetV2	59.89	59.07	59.97	59.23	1,272,859	6,369.60
SqueezeNet	72.92	72.70	72.98	72.73	736,963	6,761.99

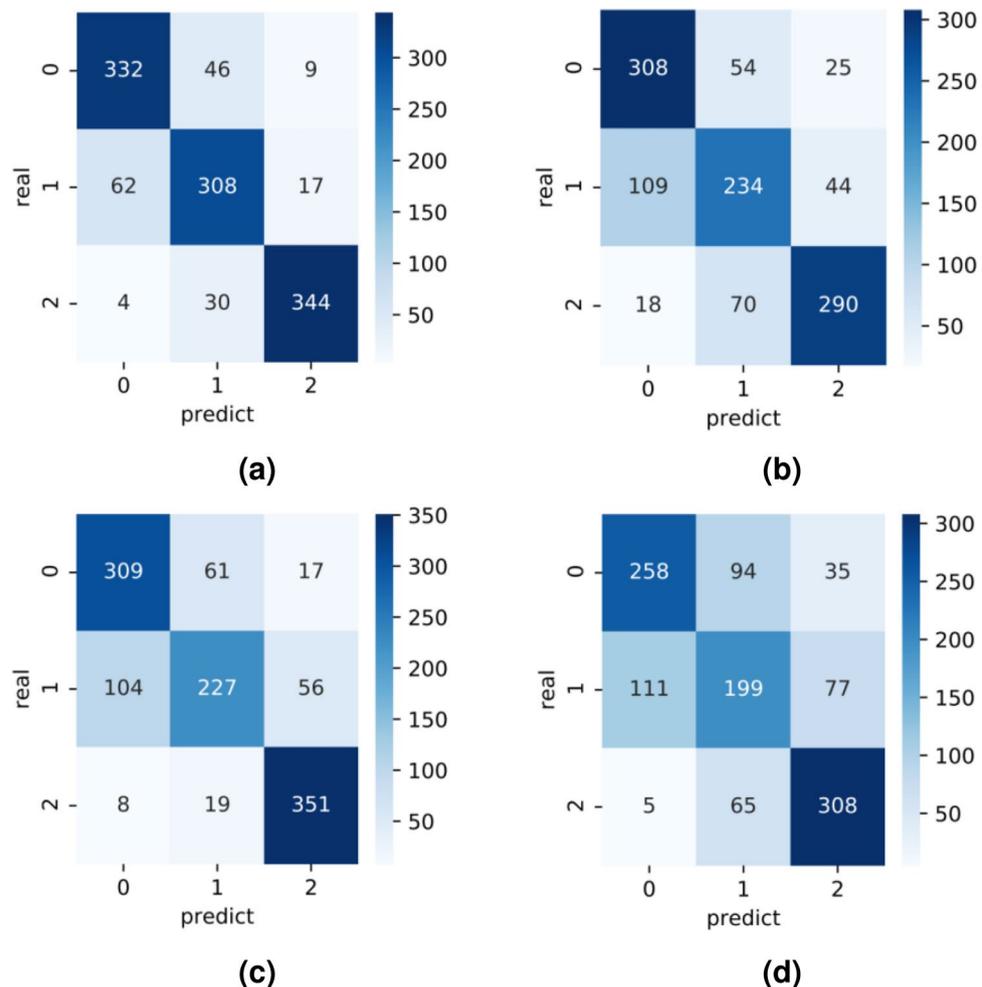
**Table 4.** Comparison of the experimental results of each model.

parameters and training speed, but lacks in the number of parameters and training speed compared to light-weight network models. The models were then compared using a confusion matrix (Fig. 8). The results showed that misclassifications mostly occurred between classes 0 and 1 because of the high similarity in their lesion features. However, compared to the other three models, the improved ConvNeXt model proposed in this study was superior in identifying and characterizing lesions and their features due to its excellent feature extraction ability to discriminate complex disturbances. A heat map of the output of the last attention module shows that the attention mechanism was effective in the improved network (Fig. 9)<sup>33</sup>.

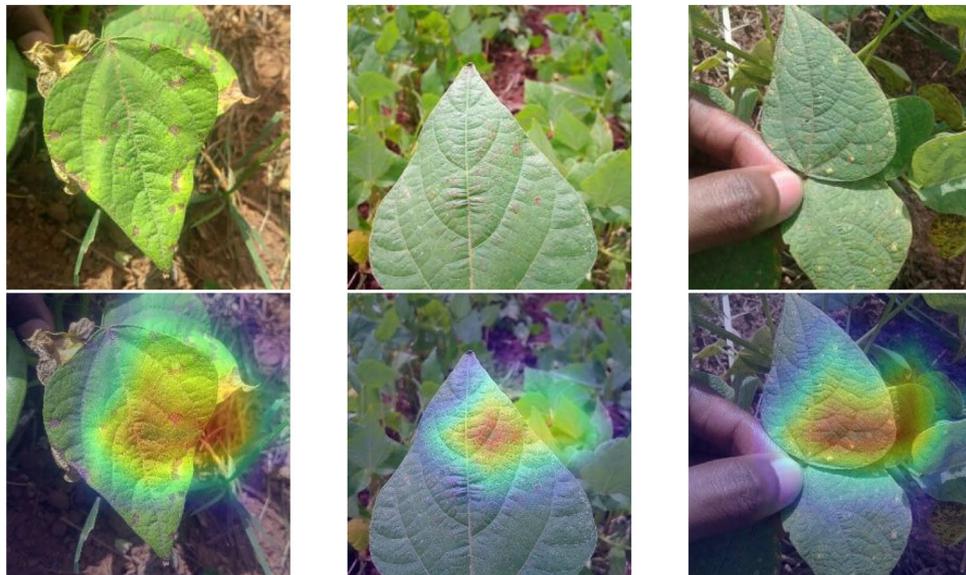
In addition, to demonstrate the performance of the proposed models on other datasets, we also conducted comparative experiments on the grape leaf dataset in the PlantVillage public dataset, and the experimental results for each model are shown in Table 5.

## Conclusion

To address the problem of complex backgrounds in soybean leaf images and unsatisfactory recognition accuracy, a soybean leaf disease classification method based on ConvNeXt and an attention module has been proposed. The proposed method optimizes soybean leaf disease classification by obtaining attention feature maps at different depths in the network. The major conclusions were as follows: By incorporating four CBAM attention modules, the CBAM-ConvNeXt network described in this study was an improvement when compared with the original ConvNeXt model as it enhanced the attention of the network to the interchannel and spatial positions of the feature maps. The issue of neuronal inactivity when the input was negative was addressed by implementing the LeakyReLU activation function in the CBAM attention module. A soybean leaf disease dataset redundancy experiment showed that the CBAM-ConvNeXt model outperformed the other network models in terms of generalization ability and robustness. The average recognition accuracy of the CBAM-ConvNeXt model on the augmented test set was 85.42%, which was higher than that of the other network models under the same experimental conditions. This indicates that the proposed model improvement method has a favorable impact on soybean leaf disease classification.



**Figure 8.** Confusion matrix contrast. (a) CBAM-ConvNeXt, (b) ResNet50, (c) Swin Transformer, (d) ConvNeXt.



**Figure 9.** Heat map of attention output. The red and blue regions highlight the most and the least discriminative regions, respectively.

Models	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	training time (s)
ResNet50	87.60	87.50	87.60	87.52	15,495.57
ConvNeXt	85.14	85.13	85.14	84.99	19,618.43
Swin transformer	93.87	93.85	93.87	93.86	21,336.28
CBAM-ConvNeXt	94.52	94.46	94.52	94.47	17,135.20
MobileNetV3	83.56	83.66	83.56	83.59	14,989.98
ShuffleNetV2	78.56	78.79	78.56	78.60	14,905.33
SqueezeNet	87.07	86.87	87.07	86.91	14,070.32

**Table 5.** Experimental results of the model on the grape leaf dataset.

The model exhibits the lowest performance in detecting leaf diseases that have a high similarity in their essential features. In the future, a combination of small-sample learning methods should be used to improve model performance when using a smaller training data set. In addition, the number of parameters in the model was found to be a gap compared to the lightweight network model in the experiments, which is not conducive to the deployment of the model to mobile, so lightweighting of the model is also an important research direction.

### Data availability

As the Jilin College of Agricultural Science and Technology “Smart Agriculture” dataset is not yet fully established, the datasets analysed in the current study are not publicly available, but are available on request from the corresponding author, You Tang, upon reasonable request.

### Code availability

The main code can be accessed at the PaddlePaddle platform, the access link is: <https://aistudio.baidu.com/projectdetail/5363906?contributionType=1>.

Received: 15 May 2023; Accepted: 1 November 2023

Published online: 06 November 2023

### References

- Jianing, G. *et al.* Crispr/cas9 applications for improvement of soybeans, current scenarios, and future perspectives. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca* **50**, 12678–12678 (2022).
- Huang, W. *et al.* New optimized spectral indices for identifying and monitoring winter wheat diseases. *IEEE J. Sel. Top. Appl. Earth Obs Remote Sens.* **7**, 2516–2524 (2014).
- Dhakal, A. & Shakya, S. Image-based plant disease detection with deep learning. *Int. J. Comput. Trends Technol.* **61**, 26–29 (2018).
- Ngugi, L. C., Abelwahab, M. & Abo-Zahhad, M. Recent advances in image processing techniques for automated leaf pest and disease recognition—a review. *Inf. Process. Agric.* **8**, 27–51 (2021).

5. Panigrahi, K. P., Das, H., Sahoo, A. K. & Moharana, S. C. Maize leaf disease detection and classification using machine learning algorithms. In *Progress in Computing, Analytics and Networking: Proceedings of ICCAN 2019*, 659–669 (Springer, 2020).
6. Hossain, E., Hossain, M. F. & Rahaman, M. A. A color and texture based approach for the detection and classification of plant leaf disease using knn classifier. In *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, 1–6 (IEEE, 2019).
7. Wagle, S. A. et al. Comparison of plant leaf classification using modified alexnet and support vector machine. *Traitement du Signal* **38** (2021).
8. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
9. Lin, J. et al. Grapenet: A lightweight convolutional neural network model for identification of grape leaf diseases. *Agriculture* **12**, 887 (2022).
10. Bansal, P., Kumar, R. & Kumar, S. Disease detection in apple leaves using deep convolutional neural network. *Agriculture* **11**, 617 (2021).
11. Haque, M. A. et al. Image-based identification of maydis leaf blight disease of maize (zea mays) using deep learning. *Indian J. Agric. Sci.* **91**, 1632–7 (2021).
12. Haque, M. A. et al. Deep learning-based approach for identification of diseases of maize crop. *Sci. Rep.* **12**, 6334 (2022).
13. Haque, M. A. et al. A lightweight convolutional neural network for recognition of severity stages of maydis leaf blight disease of maize. *Front. Plant Sci.* **13**, 1077568 (2022).
14. Haque, M. A., Marwaha, S., Deb, C. K., Nigam, S. & Arora, A. Recognition of diseases of maize crop using deep learning models. *Neural Comput. Appl.* **35**, 7407–7421 (2023).
15. Karlekar, A. & Seal, A. Soynet: Soybean leaf diseases classification. *Comput. Electron. Agric.* **172**, 105342 (2020).
16. Elfatimi, E., Eryigit, R. & Elfatimi, L. Beans leaf diseases classification using mobilenet models. *IEEE Access* **10**, 9471–9482 (2022).
17. Yu, M., Ma, X., Guan, H., Liu, M. & Zhang, T. A recognition method of soybean leaf diseases based on an improved deep learning model. *Front. Plant Sci.* **13**, 878834 (2022).
18. Fujita, E., Kawasaki, Y., Uga, H., Kagiwada, S. & Iyatomi, H. Basic investigation on a robust and practical plant diagnostic system. In *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 989–992 (IEEE, 2016).
19. Xu, Z. J., Lleras, A. & Buetti, S. Predicting how surface texture and shape combine in the human visual system to direct attention. *Sci. Rep.* **11**, 6170 (2021).
20. Guo, W., Feng, Q., Li, X., Yang, S. & Yang, J. Grape leaf disease detection based on attention mechanisms. *Int. J. Agric. Biol. Eng.* **15**, 205–212 (2022).
21. Wang, Y., Tao, J. & Gao, H. Corn disease recognition based on attention mechanism network. *Axioms* **11**, 480 (2022).
22. Zhang, M., Su, H. & Wen, J. Classification of flower image based on attention mechanism and multi-loss attention network. *Comput. Commun.* **179**, 307–317 (2021).
23. Qian, Z., Mu, J., Tian, F., Gao, Z. & Zhang, J. Facial expression recognition based on strong attention mechanism and residual network. *Multimed. Tools Appl.* **82**, 14287–14306 (2023).
24. Hu, J., Shen, L. & Sun, G. Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7132–7141 (2018).
25. Wang, Q. et al. Eca-net: Efficient channel attention for deep convolutional neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11534–11542 (2020).
26. Woo, S., Park, J., Lee, J.-Y. & Kweon, I. S. Cbam: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 3–19 (2018).
27. Liu, Z. et al. A convnet for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11976–11986 (2022).
28. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778 (2016).
29. Liu, Z. et al. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 10012–10022 (2021).
30. Li, X., Wang, W., Hu, X. & Yang, J. Selective kernel networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 510–519 (2019).
31. Xu, B., Wang, N., Chen, T. & Li, M. Empirical evaluation of rectified activations in convolutional network. arXiv preprint [arXiv:1505.00853](https://arxiv.org/abs/1505.00853) (2015).
32. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014).
33. Selvaraju, R. R. et al. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: *Proceedings of the IEEE International Conference on Computer Vision*, 618–626 (2017).

## Acknowledgements

This study was supported by the Science and Technology Development Plan Project of the Jilin Province (YDZJ202201ZYTS692) and the project of the Doctoral Initial Scientific Research Fund Supported by Jilin Agricultural Science and Technology University (No.2023706).

## Author contributions

Q.W. and X.M. wrote the main manuscript text and the main experiments, H.L., C.B. and H.Y. prepared Tables 1 and 2, M.L., J.Z. and Q.L. prepared Figures 1, 2, 3, 4, 5, 6, 7 and 8, Y.T. and G.Y. prepared the translation and edited for this paper.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to Y.T. or G.Y.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023