



## OPEN ACCESS

## EDITED BY

Muhammad Fazal Ijaz,  
Melbourne Institute of Technology,  
Australia

## REVIEWED BY

Parvathaneni Naga Srinivasu,  
Prasad V. Potluri Siddhartha Institute of  
Technology, India  
Debaleena Datta,  
University of Engineering and Management,  
India  
Jana Shafi,  
Prince Sattam Bin Abdulaziz University,  
Saudi Arabia

## \*CORRESPONDENCE

Xin Zhang

✉ xzhang1@gzu.edu.cn

Xiaoyulong Chen

✉ chenxiaoyulong@sina.cn

RECEIVED 17 May 2023

ACCEPTED 22 August 2023

PUBLISHED 11 September 2023

## CITATION

Cai J, Pan R, Lin J, Liu J, Zhang L, Wen X,  
Chen X and Zhang X (2023) Improved  
EfficientNet for corn disease identification.  
*Front. Plant Sci.* 14:1224385.  
doi: 10.3389/fpls.2023.1224385

## COPYRIGHT

© 2023 Cai, Pan, Lin, Liu, Zhang, Wen, Chen  
and Zhang. 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.

# Improved EfficientNet for corn disease identification

Jitong Cai<sup>1,2</sup>, Renyong Pan<sup>1,2</sup>, Jianwu Lin<sup>1,2</sup>, Jiaming Liu<sup>1,2</sup>,  
Licai Zhang<sup>1,2</sup>, Xingtian Wen<sup>1,2</sup>, Xiaoyulong Chen<sup>2,3,4\*</sup>  
and Xin Zhang<sup>1,2\*</sup>

<sup>1</sup>College of Big Data and Information Engineering, Guizhou University, Guiyang, China, <sup>2</sup>Guizhou-Europe Environmental Biotechnology and Agricultural Informatics Oversea Innovation Center in Guizhou University, Guizhou Provincial Science and Technology Department, Guiyang, China,

<sup>3</sup>International Jointed Institute of Plant Microbial Ecology and Resource Management in Guizhou University, Ministry of Agriculture, China Association of Agricultural Science Societies, Guiyang, China,

<sup>4</sup>College of Life Sciences, Guizhou University, Guiyang, China

**Introduction:** Corn is one of the world's essential crops, and the presence of corn diseases significantly affects both the yield and quality of corn. Accurate identification of corn diseases in real time is crucial to increasing crop yield and improving farmers' income. However, in real-world environments, the complexity of the background, irregularity of the disease region, large intraclass variation, and small interclass variation make it difficult for most convolutional neural network models to achieve disease recognition under such conditions. Additionally, the low accuracy of existing lightweight models forces farmers to compromise between accuracy and real-time.

**Methods:** To address these challenges, we propose FCA-EfficientNet. Building upon EfficientNet, the fully-convolution-based coordinate attention module allows the network to acquire spatial information through convolutional structures. This enhances the network's ability to focus on disease regions while mitigating interference from complex backgrounds. Furthermore, the adaptive fusion module is employed to fuse image information from different scales, reducing interference from the background in disease recognition. Finally, through multiple experiments, we have determined the network structure that achieves optimal performance.

**Results:** Compared to other widely used deep learning models, this proposed model exhibits outstanding performance in terms of accuracy, precision, recall, and F1 score. Furthermore, the model has a parameter count of 3.44M and Flops of 339.74M, which is lower than most lightweight network models. We designed and implemented a corn disease recognition application and deployed the model on an Android device with an average recognition speed of 92.88ms, which meets the user's needs.

**Discussion:** Overall, our model can accurately identify corn diseases in realistic environments, contributing to timely and effective disease prevention and control.

## KEYWORDS

Convolutional Neural Network, corn leaf disease, real scene, lightweight model, fully-convolution-based coordinate attention

## 1 Introduction

Corn is a globally important crop with high nutritional value and significant economic importance for farmers (Rosas-Castor et al., 2014). Corn is affected by a variety of unfavorable factors during the planting process, among which disease is the main disaster that affects corn yield, leading to a significant reduction in the yield of diseased corn and causing economic losses to farmers. Therefore, timely detection of corn diseases is crucial. However, traditional methods of diagnosing corn diseases, which require diagnosticians to physically enter the field and rely on preliminary diagnosis or quantitative analysis to determine the type of disease, are both time-consuming and labor-intensive. Agricultural experts and technicians cannot reach the field in time, while most farmers lack the necessary knowledge to accurately diagnose diseases, resulting in time, cost, and inefficiency in problem solving. In the field of plant disease recognition, existing research suggests two main approaches: traditional machine learning and deep learning (DL).

Traditional image processing methods has improved the efficiency of corn disease prevention and control to some extent (Guruprasad, 2020; O'Mahony et al., 2020). Kai et al. (2011) segmented the lesions based on the texture characteristics of corn diseases using the YCbCr color space technique, extracted the lesion texture features using the co-occurrence matrix spatial gray-level layer, and classified corn diseases using a BP neural network, achieving an accuracy rate of 98%. Aravind et al. (2018) processed the image to obtain a feature bag and texture features based on statistical histograms, and used a multi-class support vector machine to classify the diseases based on the obtained features, achieving an average best accuracy rate of 83.7%.

Recognizing crop diseases using machine learning algorithms requires features to be designed manually, which is time-consuming and laborious. In contrast, DL models are able to autonomously learn information about image target features, and are therefore widely used in the field of image recognition. Praveen et al. (2022) using the FastAI technology with ResNet-32 to diagnose ductal carcinoma, The experimental results show that 93.60% recognition accuracy was achieved on IDC dataset. Srinivasu et al. (2022) proposed an RNN model integrating GRU and LSTM with auxiliary memory components and designed one, for predicting type 2 diabetes, and the experimental results showed that the model achieved a correct recognition rate of 81.8%. For plant disease recognition, Nigam et al. (2023) used fine-tuned EfficientNet-B4 model for recognizing wheat disease and the experimental results showed that 99.35% recognition accuracy was obtained on their collected dataset. Al-Gaashani et al. (2023) based on the ResNet50 architecture, combined with kernel attention mechanism proposed SANET for recognizing rice disease, and the experimental results showed that 98.71% recognition accuracy was obtained on the publicly available rice disease dataset.

In the above study, the images of crop disease datasets usually have a simple background collected in the laboratory, while in the real environment, it is difficult to distinguish the differences between disease features with complex backgrounds to achieve the accuracy of specific disease identification in the previous

model. The use of max pooling can highlight the most significant features, help filter some noise, and reduce the interference from complex backgrounds. Sun et al. (2020) proposed a novel max pooling method for CNN trained by noisy samples, and the classification accuracy of CNN with this method for noisy images is much higher than that of traditional pooling methods. Lin et al. (2022b) proposed CAMFFNet for tobacco disease identification using multiple feature fusion module, which introduced additional max pooling branches for feature extraction at different locations of the structure. The results showed that the recognition accuracy of the test set was 89.71%. At the same time, for complex background images, the attention mechanism can make the model focus on the disease region in a targeted manner, improve the ability of the model to learn disease features, and reduce the influence of the model by noise to improve the accuracy of the model. Hu et al. (2023) proposed a Class-Attention-based Lesion Proposal Convolutional Neural Network to locate disease objects of complex backgrounds, and the experimental results showed that the recognition accuracy was 92.56% on a self-constructed field strawberry disease dataset. Stephen et al. (2023) designed a self-attention-based ResNet architecture for rice disease classification, and the experimental results showed that the accuracy of recognition of four types of rice diseases was 98.54%.

The above studies have well demonstrated the superiority of CNN for plant disease identification. However, the above models are heavy-weight and require large computational resources, which are not suitable for efficient deployment on mobile devices for timely disease identification. Therefore, there is a need to design lightweight networks that can be deployed on mobile devices. Vishnoi et al. (2022) proposed a new CNN model using only fewer convolutional layers to reduce the computational burden, and the experimental results showed a 98% classification accuracy for four classes of apple diseases. Lin et al. (2022a) proposed a lightweight CNN model called GrapeNet for recognizing specific grape diseases at different symptom stages, and the experimental results showed that the recognition accuracy for seven classes of grape diseases was 97.85%. Chen et al. (2022) proposed a lightweight CNN model, DFCANet, for recognizing corn diseases in real environments, and the experimental results showed that the classification accuracy for six classes of corn diseases reached 98.47%.

Identification of corn leaf diseases in field environments faces several difficulties, such as complex background disturbances, variability and irregularity of lesion areas. In addition, traditional CNN models with a large number of parameters in crop identification tasks require more computational resources, cannot be deployed on mobile devices and are difficult to scale widely.

To solve the above problems, we focus on corn diseases and propose a lightweight CNN model with high recognition accuracy, FCA-EfficientNet. on top of EfficientNet, we use adaptive fusion (AF) module for shallow feature extraction, fully-convolution-based coordinate attention (FCA) module to focus on disease areas in complex backgrounds, better normalization methods and activation functions to enhance the recognition of the network speed, and removing the network redundant structure of the corn disease dataset to make the network easier to deploy. This study explores

the accurate recognition of corn disease images and provides new insights for plant disease recognition research. The main contributions of this study are as follows:

- Since the field environment often has complex background noise, in order to reduce the interference of background on the network recognition results, we use pooling layer to reduce the background interference, and through the AF module, fuse the features of pooling layer and convolutional layer, so that the network can better extract shallow information and improve the recognition accuracy.
- To better focus the network on the lesion area while reducing the interference of background noise, we propose the FCA module, which uses a convolutional structure to extract spatial information, allowing the network to better extract disease areas and reduce the recognition difficulty.
- In order to make the network have better generalization ability, we use more advanced normalization method with activation function to improve the robustness of the network.
- To make the model easy to deploy due to performance limitations of mobile devices, we removed the redundant structure of the model to make the model have faster inference speed and smaller number of parameters.

The paper generally consists of several sections. The “Introduction” section provides a brief overview of the research field, motivation, and objectives. The “Materials and Methods” section describes the source of the dataset and the methods used for data processing. The “Proposed method” section provides an

explanation of the technical details adopted in this paper. The next section presents the “Experimental results and analysis”. The “Conclusions” section provides conclusions and outlines future research directions. Finally, the last section includes the references cited in this paper.

## 2 Materials and methods

### 2.1 Image acquisition

The corn disease data for this paper comes from four different sources. We obtained images of three types of corn diseases images from the CD&S (Ahmad et al., 2021) dataset, namely northern leaf blight (NLB), gray leaf spot (GLS), and northern leaf spot (NLS) respectively. We obtained images of corn rust leaf (RL) under real conditions from the PlantDoc (Singh et al., 2020) dataset. Added to that, we got images of corn leaf infected by fall armyworm on a public website (<https://github.com/FXD96/Corn-Diseases>). Finally, we collected images of healthy corn leaves (LH) and a few images of other diseases through web crawler technology. Figure 1 shows a sample corn disease dataset in a real environment.

### 2.2 Image preprocessing

In this paper, we collected 3258 images, including 432 images of Fall Armyworm, 613 images of Gray Leaf Spot, 688 images of Leaf Blight, 537 images of Leaf Healthy, 551 images of Northern Leaf Spot, 437 images of Rust Leaf. After that, the image data of each category are assigned to the training set, validation set and test set in

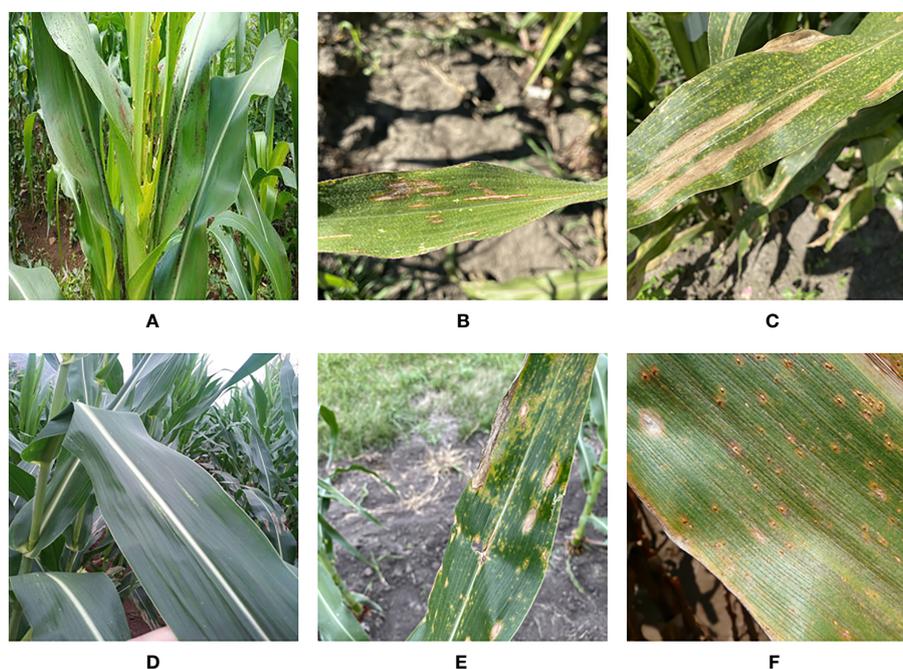


FIGURE 1

Example of corn leaves: (A) Fall Armyworm. (B) Gray Leaf Spot. (C) Leaf Blight. (D) Leaf Healthy. (E) Northern Leaf Spot. (F) Rust Leaf.

the ratio of 8:1:1 respectively. Table 1 shows the data distribution of corn leaf disease images.

In order to obtain better generalization and robustness and avoid overfitting problems, DL models usually need a lot of data as support (Lin et al., 2023). Therefore, three different online data augmentation methods are used in this paper, namely Cutout (DeVries and Taylor, 2017), Grid Mask (Chen et al., 2020) and Random Erasing (Zhong et al., 2020). Figure 2 illustrates the online data enhancement method used in this paper.

### 3 Proposed method

EfficientNet (Tan and Le, 2019) is a network series published by Google in May 2019, designed to improve model accuracy while keeping the model parameter count relatively low. Taking into account the deployment of the network on mobile devices, we chose the EfficientNet-B0 architecture and improved on it. It consists of nine stages, which comprise convolutional layers, MBConv1 layers, MBConv6 layers, pooling layers, and a fully connected layer. The overall architecture of EfficientNet-B0 is shown in Figure 3.

To identify corn disease images more accurately, we propose a new model based on the EfficientNet-B0 architecture, named FCA-EfficientNet. Table 2 provides a detailed layer-by-layer description of FCA-EfficientNet. The overall architecture of FCA-EfficientNet is shown in Figure 4.

#### 3.1 Fully-convolution-based coordinate attention

In order to avoid the interference of complex background, we replace the SE module in the MBConv structure with the FCA module, which enables the network to better focus on the lesion area. Figure 5 shows the overall structure of the FCA-MBConv module.

FCA module based on Coordinate Attention (CA) (Hou et al., 2021). While CA employs global average pooling to aggregate horizontal and spatial features, this approach fails to capture critical disease information in the feature map for fine-grained recognition tasks, such as corn disease recognition, leading to recognition accuracy bottlenecks. To overcome this limitation, we

remove the global average pooling layer and employ a convolutional kernel with a size of horizontally, and a convolutional kernel with a size of vertically, as shown in Figure 6. This modification enables the attention module to automatically capture salient disease information in different spatial directions and aggregate it, thereby strengthening the ability to capture dependencies between remote disease feature information while preserving precise location information of disease features. The structure of the FCA module is shown in Figure 7.

Assuming the input feature map is  $x_c$ , where  $c$  denotes the number of feature map channel, a new feature map  $\hat{x}_c$  can be obtained by two convolution operations for feature extraction.

$$\hat{x}_{tc} = DWConv_{tc}^{k \times k}(Conv_{tc}^{1 \times 1}(x_c)) \quad (1)$$

Convolution of  $\hat{x}$  in the horizontal and vertical directions separately. This results in two output feature maps.

$$Z_{tc}^h(h) = Conv^{1 \times W}(\hat{x}_{tc}) \quad (2)$$

$$Z_{tc}^w(w) = Conv^{H \times 1}(\hat{x}_{tc}) \quad (3)$$

The two output feature maps are concatenated along the channel dimension and passed through a  $1 \times 1$  convolutional layer  $Conv^{1 \times 1}$  followed by a non-linear activation function  $\delta$  to obtain the output feature map  $f$ .

$$f = \delta(Conv^{1 \times 1}(concatz^h;)(z^w)) \quad (4)$$

$f$  is decomposed into two outputs along the spatial dimension and transformed into tensors with the same number of channels as the input by two  $1 \times 1$  convolutions ( $g$  and  $h$ ). The two outputs can be represented by  $g^h$  and  $g^w$ :

$$g^h = Sigmoid(Conv_h^{1 \times 1}(f^h)) \quad (5)$$

$$g^w = Sigmoid(Conv_w^{1 \times 1}(f^w)) \quad (6)$$

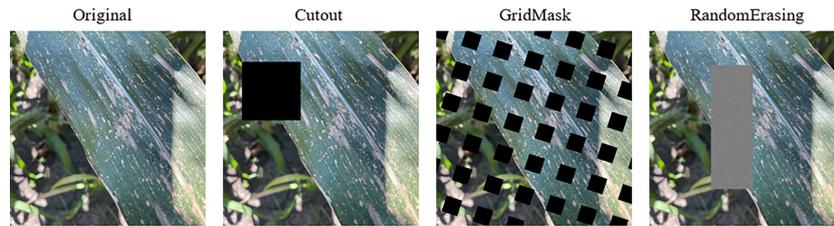
the output expression of the FCA module is:

$$y_{FCA}(i, j) = \hat{x}_{tc}(i, j) \times g^h(i) \times g^w(j) \quad (7)$$

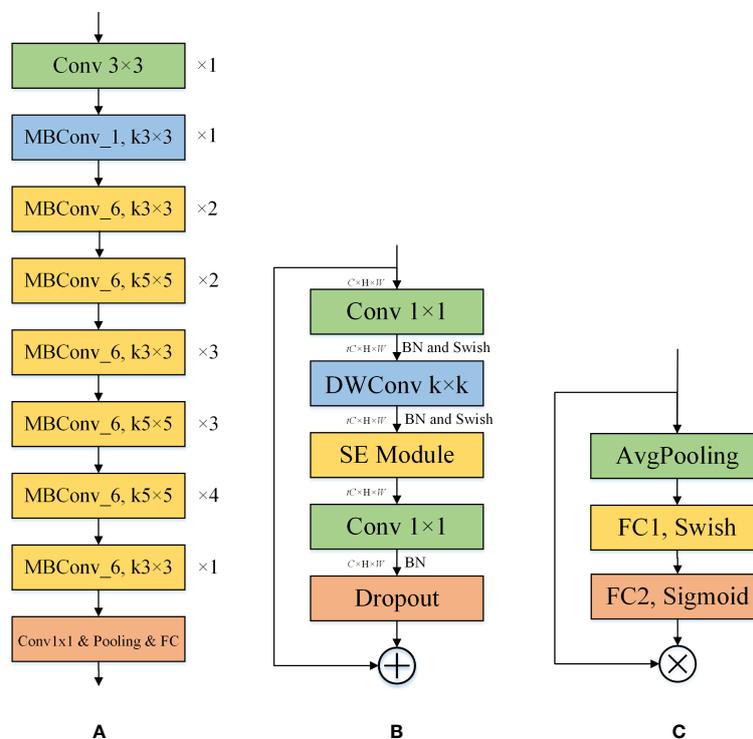
Finally, the feature dimensions are reduced and residual joins are performed through the Conv layer, and the output of the FCA-MBConv module is:

TABLE 1 Distribution of image data by category.

| Disease classes          | Train | Valid | Test |
|--------------------------|-------|-------|------|
| Fall Armyworm (FA)       | 346   | 43    | 43   |
| Gray Leaf Spot (GLS)     | 491   | 61    | 61   |
| Leaf Blight (LB)         | 550   | 69    | 69   |
| Leaf Healthy (LH)        | 429   | 54    | 54   |
| Northern Leaf Spot (NLS) | 441   | 55    | 55   |
| Rust Leaf (RL)           | 351   | 43    | 43   |
| Total                    | 2608  | 325   | 325  |



**FIGURE 2** Classical online data augmentation. The Cutout method randomly cuts out part of the sample and fills it with 0 pixel values. Grid Mask generates a mask with the same resolution as the original image, and then multiplies that mask with the original image. Random Erasing randomly selects a region and then overwrites it with a random value.



**FIGURE 3** The individual structures included in EfficientNet-B0. (A) The structure of the EfficientNet-B0. xn indicates that the layer is repeated n times, and knxn indicates that the size of the convolution kernel is n. (B) The structure of the MBConv\_t. k denotes the size of the convolution kernel of the DWConv layer. t indicates that the input channel is expanded by a factor of t. (C) The structure of the SE. SE can perform channel feature enhancement on the input feature map to improve the performance of the model by learning the relationship between different channels.

$$MBConv_{FC_A}(x_c) = x_c + Dropout(Conv_c^{1 \times 1} y_{FCA}(\hat{x}_{tc})) \quad (8)$$

### 3.2 Adaptive fusion

In real-world crop disease images, the presence of complex background noise often makes it challenging to extract shallow features using convolutional layers alone, which can significantly affect the recognition accuracy of the model. To address this issue,

we introduced a max-pooling layer to assist in shallow feature extraction. Unlike the conventional feature fusion approaches that involve simple addition or concatenation, we adopted two learnable weighting coefficients to evaluate the feature maps derived from the two methods. This approach enables the model to dynamically allocate weights to the pooled and convolved feature maps based on the complexity of the background information, resulting in effective extraction of shallow convolutional layer features while avoiding interference from complex background noise. As a result, the effectiveness and robustness of the model is significantly improved. The structure of the AF module is shown in Figure 8.

TABLE 2 Detailed configuration information for each layer of FCA-EfficientNet.

| Stage $i$ | Operator $\hat{F}_i$   | Resolution $\hat{H}_i \times \hat{W}_i$ | #Channels $\hat{C}_i$ | #Layers $\hat{L}_i$ | Dropout factor |
|-----------|------------------------|---|-----------------------|---------------------|----------------|
| 1         | Conv3×3                | 224×224                                 | 32                    | 1                   | 0              |
| 2         | AF & FCA-MBConv1, k3×3 | 112×112                                 | 16                    | 1                   | 0.0125         |
| 3         | FCA-MBConv6, k3×3      | 112×112                                 | 24                    | 2                   | 0.025          |
| 4         | FCA-MBConv6, k5×5      | 56×56                                   | 40                    | 2                   | 0.05           |
| 5         | FCA-MBConv6, k3×3      | 28×28                                   | 80                    | 2                   | 0.075          |
| 6         | FCA-MBConv6, k5×5      | 14×14                                   | 112                   | 3                   | 0.1            |
| 7         | FCA-MBConv6, k5×5      | 14×14                                   | 192                   | 2                   | 0.1375         |
| 8         | FCA-MBConv6, k3×3      | 7×7                                     | 320                   | 1                   | 0.1625         |
| 9         | Conv1×1 & Pooling & FC | 7×7                                     | 1280                  | 1                   | 0              |

Assuming the input image is  $x$ , Two output  $y_m, y_f$  can be obtained by max pooling and backbone. We use the learnable weight parameters  $w_1, w_2$  to weight the two and obtain the final output feature map  $y_{AF}$ :

$$y_m = \text{Conv}^{1 \times 1}(\text{Maxpool}^{2 \times 2}(x)) \quad (9)$$

$$y_f = \text{MBConv}_{\text{FCA}}(\text{Conv}_{3 \times 3}(x)) \quad (10)$$

$$y_{AF} = w_1 \times y_m + w_2 \times y_f \quad (11)$$

### 3.3 Layer normalization and gaussian error linear unit

In CNN models, normalization techniques are used to adjust the distribution of the input data to better fit the training process of the neural network. The formula for normalization is as follows.

$$f(x) = \frac{x - E[x]}{\sqrt{\text{Var}[x] + \delta}} * \gamma + \beta \quad (12)$$

where  $x$  denotes the input feature map and  $f(x)$  the output feature map after normalization.

The use of batch normalization (BN) (Ioffe and Szegedy, 2015) in EfficientNet allows the model to accelerate convergence during training, but the inconsistency of BN between training and testing may adversely affect the model performance. Therefore, we adopt layer normalization (LN) (Ba et al., 2016) instead of BN, which has the same performance during training and testing and is easier to implement. LN normalizes the output of each convolutional layer to a normalized value equal to the mean and variance of the inputs, which reduces the memory consumption during the training of the model and improves the generalization ability of the model. As shown in Figure 9. Unlike BNs that are normalized in the batch direction, LN pairs are normalized in the channel direction. This makes LN independent of batch size.

Also to speed up the convergence of the model, we used gaussian error linear unit (GELU) (Hendrycks and Gimpel, 2016)

to replace Swish (Ramachandran et al., 2017). Their formulas are as follows.

$$\text{Swish}(x) = x \cdot \text{sigmoid}(x) = \frac{x}{1 + e^{-x}} \quad (13)$$

$$\text{GELU}(x) = x \times P(X \leq x) = x \times \phi(x), x \sim N(0, 1) \quad (14)$$

The plots of Swish and GELU functions and their derivatives are shown in Figure 10. Compared to Swish, GELU tends to converge faster, which makes it more suitable for model training.

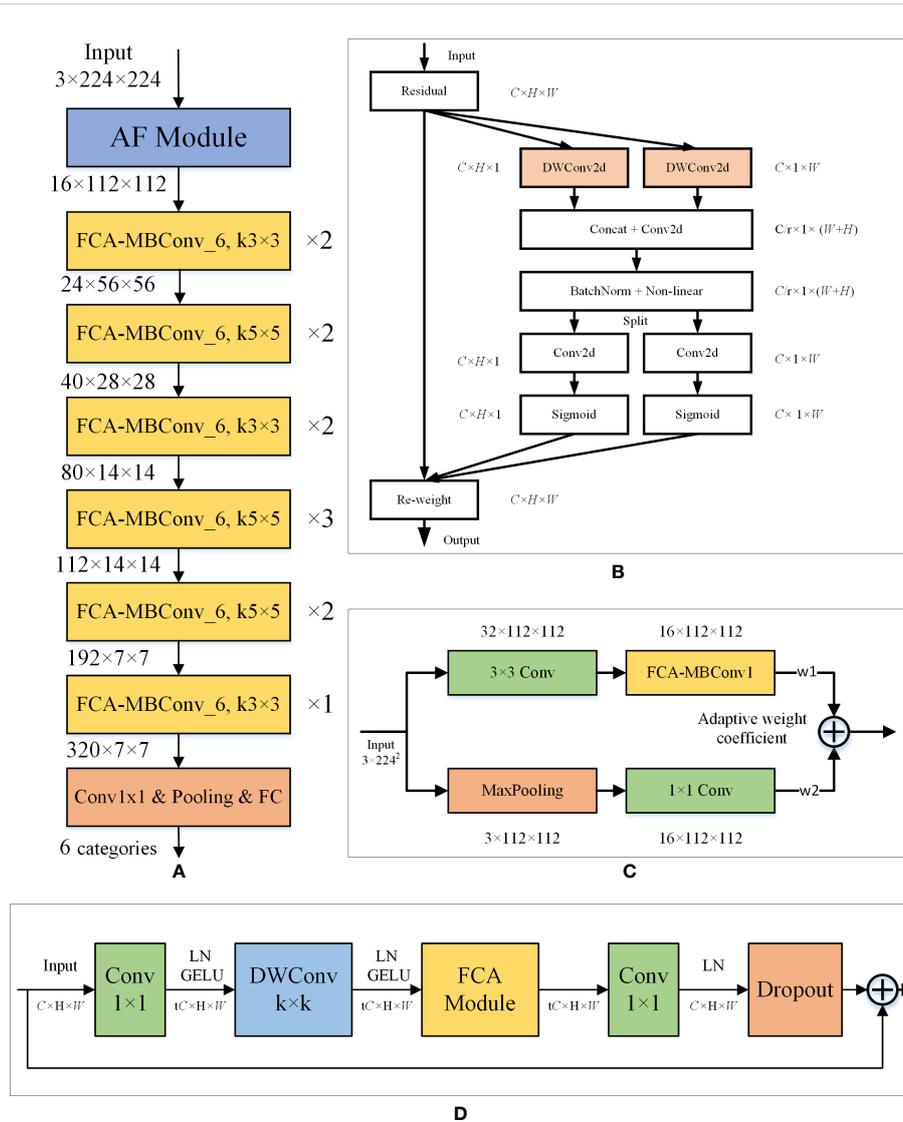
### 3.4 Optimal architecture

Multiple MBConv layers are included in EfficientNet, and by simply repeating the MBConv layers, the model can be made to learn complex features and achieve better performance. However, EfficientNet was trained by a large dataset like ImageNet with a large number of samples and categories, compared to the corn disease dataset we used with fewer samples and categories. As a result, there are some redundant structures in the original network, and although the effect of these redundant structures on the accuracy can be reduced by the residual structures, these structures cause the model's to perform unnecessary computations, which affects the recognition speed of the model. Therefore, we remove these redundant structures to reduce the number of parameters of the model and increase the speed of computation without affecting the accuracy of model recognition, which helps in deployment on mobile devices.

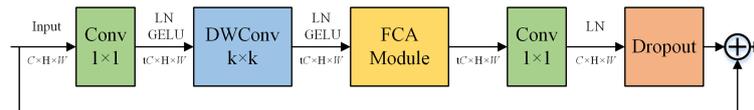
## 4 Experimental results and analysis

### 4.1 Evaluation indexes

Accuracy (Acc), Precision (P), Recall (R), F1 score (F1) are utilized as the evaluation metrics in this study, they are defined as follows:



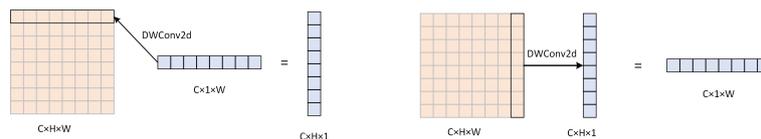
**FIGURE 4** Layered architecture diagram of the FCA-EfficientNet. **(A)** The structure of the FCA-EfficientNet.  $xn$  indicates that the layer is repeated  $n$  times, and  $k \times n \times n$  indicates that the size of the convolution kernel is  $n$ . **(B)** The structure of the FCA module. Attentional information about spatial orientation is extracted through a convolutional model, allowing the network to better focus on the disease region and reduce interference from complex backgrounds. **(C)** The structure of the AF module. The AF module uses maximum pooling and  $1 \times 1$  convolution, which reduces the interference of complex backgrounds and allows the model to better extract shallow features. **(D)** The structure of the FCA-MBCConv<sub>t</sub>.  $k$  denotes the size of the convolution kernel of the DWConv layer.  $t$  indicates that the input channel is expanded by a factor of  $t$ .



**FIGURE 5** The structure of the FCA-MBCConv.  $k$  denotes the size of the convolution kernel of the DWConv layer.  $t$  indicates that the input channel is expanded by a factor of  $t$ .

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (15) \quad R = \frac{TP}{TP + FN} \quad (17)$$

$$P = \frac{TP}{TP + FP} \quad (16) \quad F1 = \frac{P \times R}{P + R} \times 2 \quad (18)$$



**FIGURE 6**  
The computational procedure of DWConv2d, using a DW convolution of  $1 \times W$ , can obtain attention in the width direction of size  $H \times 1$  by focusing on the feature information in the width direction of the picture. Similarly, using DW convolution of  $H \times 1$ , attention in the height direction of size  $1 \times W$  can be obtained.

Where TP, TN, FP, and FN are the abbreviations used to represent the count of true positive, true negative, false positive, and false negative samples, respectively. Furthermore, the parameter count of the model is also considered to evaluate its performance.

### 4.2 Experimental configuration and hyperparameter setting

This study utilized the language of python-3.7.13 and pytorch-1.12.1 deep learning frameworks for experimentation. All experiments were conducted on a Windows 10 operating system, with an Intel(R) Xeon(R) W-1390 CPU and an NVIDIA RTX3090 GPU. The hyperparameters for model training were set to the configurations listed in Table 3 for all models.

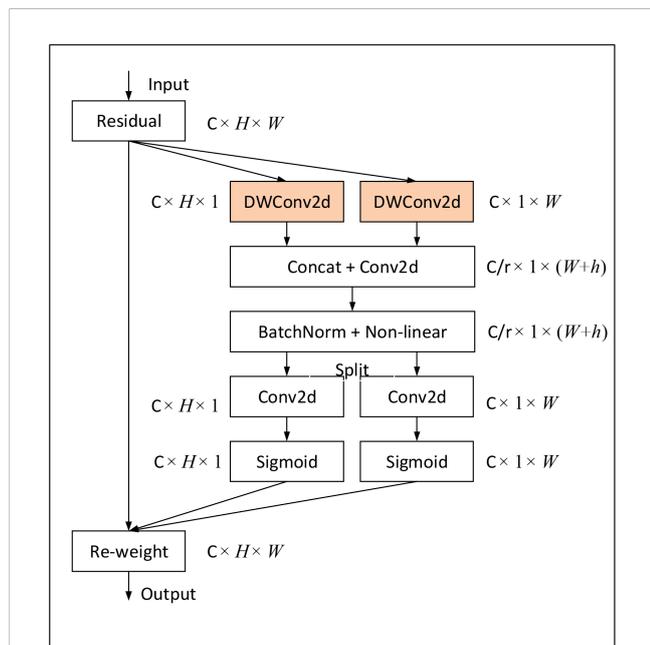
### 4.3 Performance of FCA-EfficientNet

The performance of the model on the training and validation sets is shown in Figure 11. To assess the recognition performance of the model, we conducted a separate analysis of its recognition capabilities on various categories of disease images in the test set. By employing a comprehensive set of evaluation metrics, we performed a quantitative analysis and the experimental results are presented in Table 4. As shown in the table, our model achieved an accuracy of up to 100% in identifying various types of corn diseases, with none falling below 95%. This demonstrates the model’s precise recognition of corn diseases. Notably, the model’s performance in identifying Northern Leaf Spot images was exceptional, with all evaluation metrics reaching the highest levels. Conversely, the model’s recognition performance on Rust Leaf images was comparatively lower, possibly due to the similarity to images of other categories. Additionally, the average precision, recall, and F1 scores of 98.63%, 98.76%, and 98.68%, respectively, demonstrate the model’s superior overall performance, making it highly effective in accomplishing the corn disease classification task.

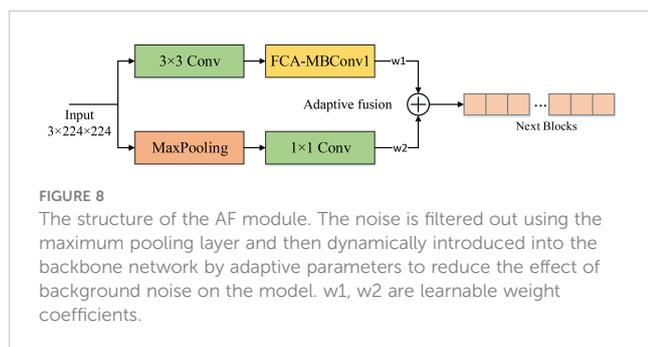
In Figure 12, a confusion matrix is presented to visualize the recognition performance of the designed model on corn disease images. The matrix displays the true and predicted classes of the data along its horizontal and vertical axes, respectively, with the diagonal entries indicating the number of correct predictions. The confusion matrix suggests that the proposed model is capable of effectively identifying multiple classes of corn disease images, as the majority of images in the provided dataset are classified correctly. The misclassifications observed may be attributed to small inter-class differences between certain categories and background interference. Notably, it achieves a high degree of accuracy in correctly classifying all disease images of the GLF, LH, and NLS categories, while maintaining a low rate of false positives for images of other categories. These results underscore the model’s high precision and reliability for identifying corn diseases.

### 4.4 Comparison of different nodes for adaptive fusion

In order to determine the optimal fusion location, we conducted experiments on three different nodes in the shallow network: after the first convolutional layer, after the first FCA-MBCConv module, and after the second FCA-MBCConv module. Figure 13 shows different fusion nodes. The fusion nodes were selected based on



**FIGURE 7**  
The structure of the FCA module. We use DWConv 2d to replace the average pooling layer in CA, which is able to capture critical disease information through convolutional operations while being less susceptible to interference from complex backgrounds.



their potential to improve the model's performance. The results, as presented in Table 5, indicate that the addition of AF at Node1, Node2, and Node3 led to improvements in various evaluation metrics, while maintaining a similar parameter count. Notably, applying AF to the shallow layers of the network resulted in better recognition results compared to the deep layers. Specifically, the performance improvement was most significant when applying AF to Node2, resulting in a 0.92% increase in recognition accuracy and a 0.91% increase in F1-score. These findings suggest that incorporating AF into specific locations of the network can effectively enhance its performance.

#### 4.5 Comparison of different attention mechanisms

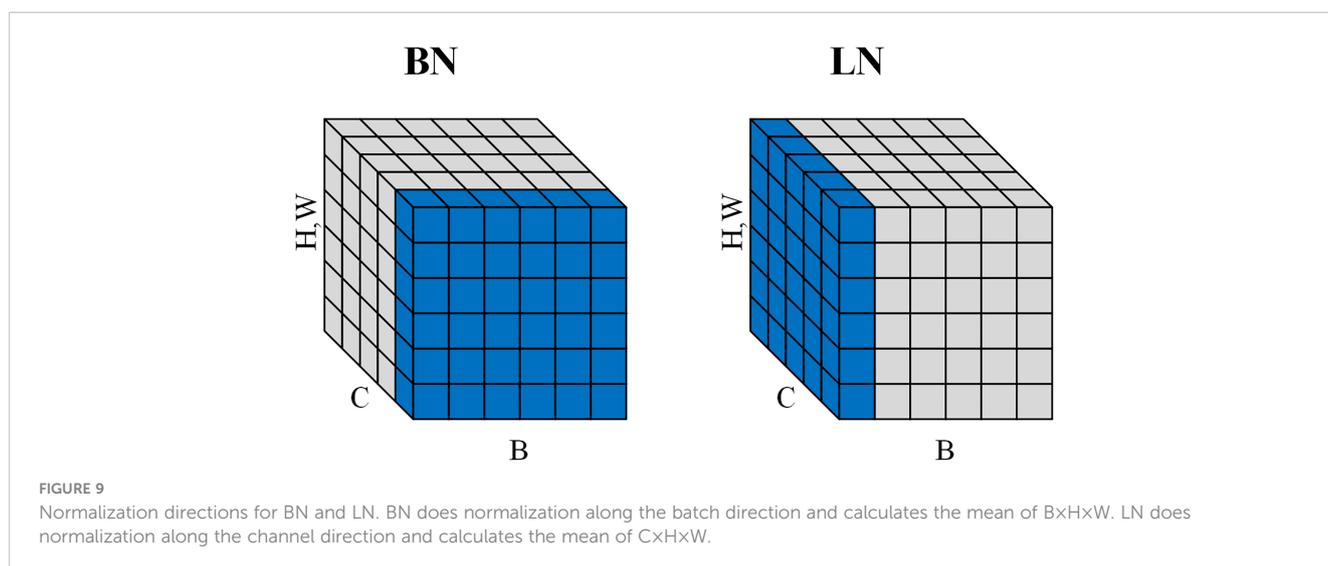
Different attention mechanisms have a significant impact on the performance of CNN models, and adding suitable attention mechanisms can optimize recognition results. This study compared and analyzed the effects of three attention mechanism modules, namely SE, CA, and our proposed FCA on model performance. As shown in the Table 6, adding the CA module did not significantly improve the overall recognition performance of the model, and the F1 score even decreased. However, after adding the FCA module, the model's performance was significantly

improved, with an increase in recognition accuracy of 0.93%. This result verifies that adding the FCA module can effectively promote the model's focus on disease information, leading to better recognition results.

To analyze the learning ability of models with different attention mechanisms on disease information, we used Grad-Cam (Selvaraju et al., 2017) to generate class activation maps for visualization, as shown in Figure 14. The first column shows the original disease image, and the other columns show the visualization results of different models. The results in the first row of the figure demonstrate that the use of the FCA module enables the model to more comprehensively focus on the primary diseased regions. Moreover, the comparative results presented in the second row of the figure reveal significant disparities between the models. Specifically, the model utilizing the SE module inadequately attends to the entire diseased area, while the model leveraging the CA module primarily focuses on background information. Conversely, the FCA module precisely localizes the distinctive diseased feature region, underscoring its ability to promote the model's attention to disease information. The results indicate that the FCA module can promote the model's attention to disease information and is more suitable for identifying corn diseases.

#### 4.6 Comparison of different normalization methods and activation function

In order to investigate the impact of different normalization methods and activation functions on the model, we conducted comparative analysis experiments by combining BN, LN, SiLU, and GELU. As shown in Table 7, replacing the BN with the LN reduces the complexity of the model, as evidenced by the reduction in Flops, while the parameter counts remain relatively unchanged. In addition, the combination of LN and GELU yields the highest recognition accuracy and precision, and sub-optimal recall and F1-score. The consistency of LN's performance in training and testing,



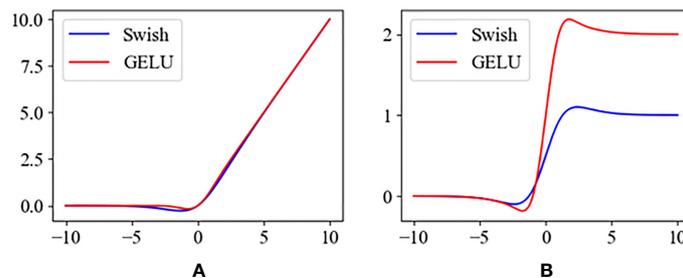


FIGURE 10

(A) The activation functions of the Swish and the GELU. (B) The derivatives of Swish and GELU.

and the approximately 3% reduction in Flops, make it more suitable for model deployment.

## 4.7 Optimal architecture

We removing some redundant layers in Stages 5, 6, and 7, the performance of all networks on the test set is shown in the following Table 8. From the table, it can be observed that removing only one redundant layer does not significantly decrease the performance of the model (rows 6, 9, and 11 in the table). The model can still maintain good performance even after removing all redundant

layers (row 1 in the table). However, removing all redundant layers in shallow stages while keeping the same number of layers in Stage 7 may lead to overfitting and a significant decrease in model performance (row 3 in the table). With consideration of the accuracy, parameter quantity, and computational complexity of the models, we finally selected 2 layers in Stage 5, 3 layers in Stage 6, and 2 layers in Stage 7 (row 4 in the table).

## 4.8 Ablation experiment

This study conducted ablation experiments to demonstrate the efficacy of the proposed model in improving model performance. Table 9 presents the results of the ablation experiments on the test set. The introduction of the AF module, FCA module, LN and GELU led to accuracy improvements of 0.9%, 0.9%, and 0.3%, respectively. Moreover, we have identified the optimal neural network architecture that achieves a 0.3% improvement in accuracy on the test set while reducing the model's parameter count. The improved model reduced the parameters by 0.58M compared to the original EfficientNet model and achieved a 2.4% increase in accuracy. These findings highlight the ability of the designed model to effectively learn disease feature information in corn disease images, thereby achieving superior recognition performance with lower parameter consumption following model optimization.

TABLE 3 Hyperparameter configuration for model training.

| Parameter           | Value         |
|---------------------|---------------|
| Optimizer           | Adam          |
| Loss Function       | Cross-Entropy |
| Training Epochs     | 200           |
| Warmup Epochs       | 20            |
| Learning Rate       | 0.001         |
| ExponentialLR Gamma | 0.99          |
| Batch Size          | 32            |

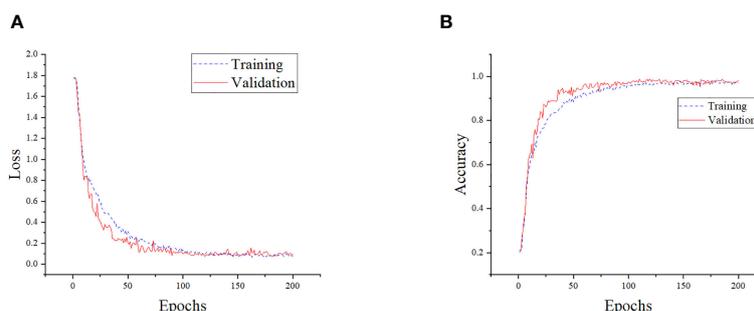
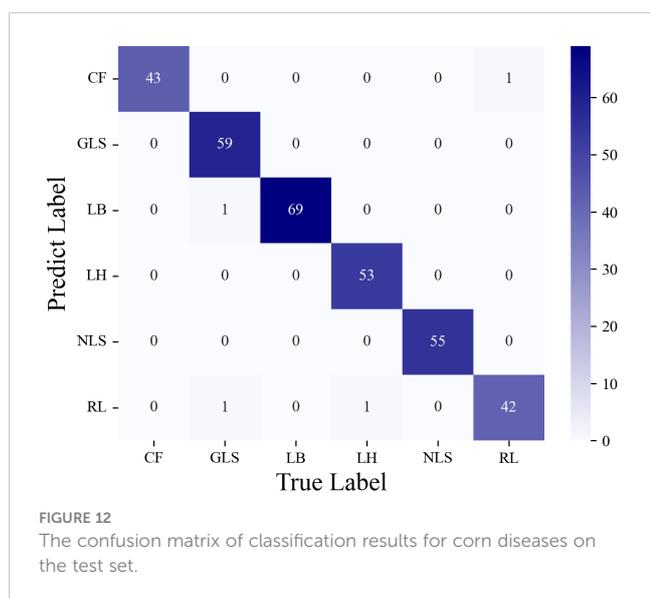


FIGURE 11

Training details of the model. (A) Training loss and validation loss of the model. (B) Training accuracy and validation accuracy of the model.

TABLE 4 Evaluation results of FCA-EfficientNet on corn images of different categories.

| Categories         | Precision | Recall | F1     | TP   | FP  | TN  | FN    |
|--------------------|-----------|--------|--------|------|-----|-----|-------|
| Fall Armyworm      | 0.9773    | 1.0    | 0.9885 | 43.0 | 1.0 | 0.0 | 281.0 |
| Gray Leaf Spot     | 1.0       | 0.9672 | 0.9833 | 59.0 | 0.0 | 2.0 | 264.0 |
| Leaf Blight        | 0.9857    | 1.0    | 0.9928 | 69.0 | 1.0 | 0.0 | 255.0 |
| Leaf Healthy       | 1.0       | 0.9815 | 0.9907 | 53.0 | 0.0 | 1.0 | 271.0 |
| Northern Leaf Spot | 1.0       | 1.0    | 1.0    | 55.0 | 0.0 | 0.0 | 270.0 |
| Rust Leaf          | 0.9545    | 0.9767 | 0.9655 | 42.0 | 2.0 | 1.0 | 280.0 |



### 4.9 Comparison with different models

CNN has shown superior performance in the field of image classification, and CNN-based classification models have diversity for different domains. To demonstrate the classification performance of the designed model, this study conducted experiments on corn disease recognition using existing classic CNN models under the same training conditions. The experiments used different models with varying parameter sizes, including VGG16, ResNet50, Shufflenetv2, and MobileNetv2, and the accuracy change curves of each model on the validation set during 200 epochs of training are shown in Figure 15. It is not difficult to observe from the validation curve graph that, compared with existing classic models, the proposed model can achieve a higher accuracy at a faster rate and with less oscillation, demonstrating better accuracy and stability.

Table 10 shows the recognition performance of different models on the test set, and compares and analyzes the advantages of the proposed model using metrics such as accuracy, precision, recall,

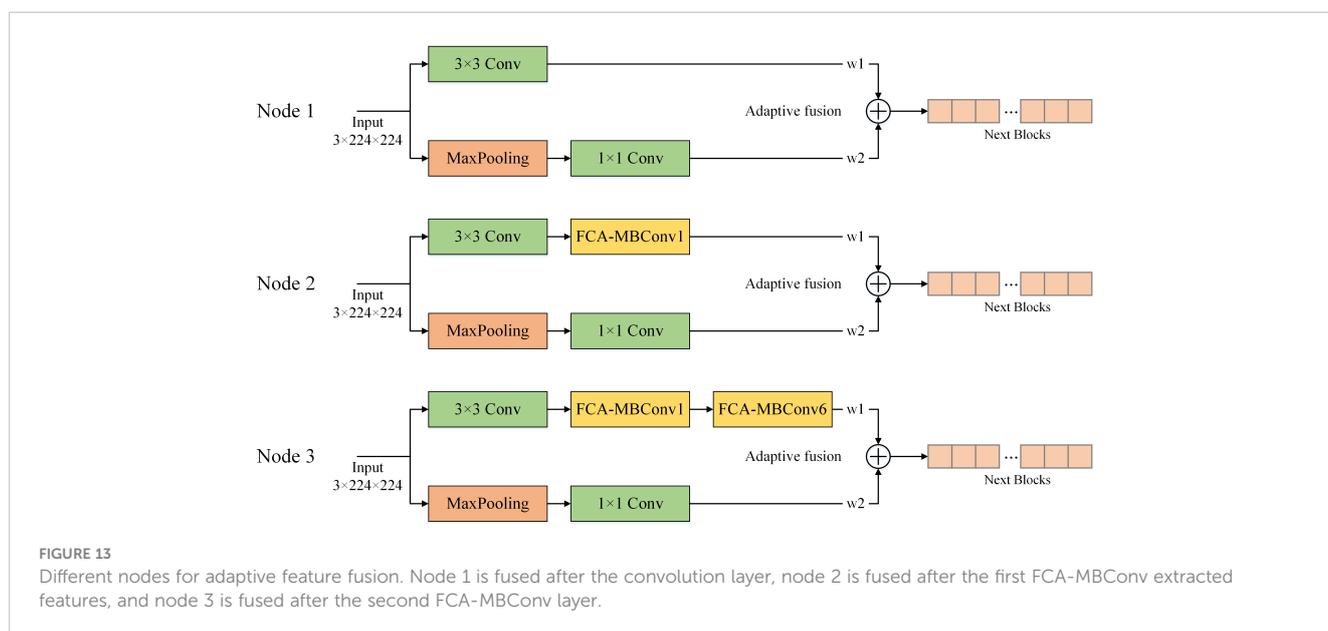


TABLE 5 The AF for the different nodes.

| Model              | Accuracy | Precision | Recall | F1     | Param  |
|--------------------|----------|-----------|--------|--------|--------|
| EfficientNet (FCA) | 0.9631   | 0.9650    | 0.9603 | 0.9619 | 4.02 M |
| AF on Node1        | 0.9692   | 0.9709    | 0.9670 | 0.9682 | 4.02 M |
| AF on Node2        | 0.9723   | 0.9726    | 0.9698 | 0.9710 | 4.02 M |
| AF on Node3        | 0.9631   | 0.9611    | 0.9593 | 0.9592 | 4.02 M |

TABLE 6 The effects of different attention mechanisms.

| Attention Mechanism     | Accuracy | Precision | Recall | F1     | Param  |
|-------------------------|----------|-----------|--------|--------|--------|
| AF-EfficientNetB0 (SE)  | 0.9723   | 0.9726    | 0.9698 | 0.9710 | 4.02 M |
| AF-EfficientNetB0 (CA)  | 0.9723   | 0.9728    | 0.9701 | 0.9711 | 4.79 M |
| AF-EfficientNetB0 (FCA) | 0.9816   | 0.9848    | 0.9837 | 0.9841 | 5.00 M |

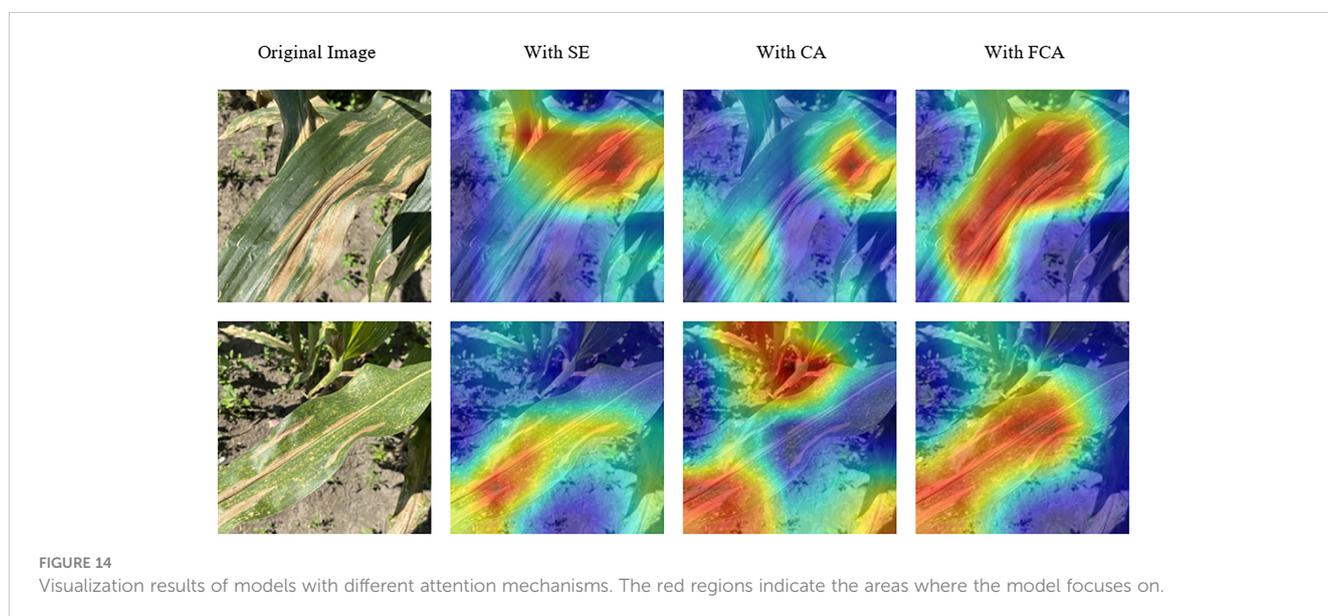


TABLE 7 The effects of different combinations.

| Method     | Accuracy | Precision | Recall | F1     | Param  | Flops   |
|------------|----------|-----------|--------|--------|--------|---------|
| BN & Swish | 0.9816   | 0.9848    | 0.9837 | 0.9841 | 5.00 M | 420.23M |
| BN & GELU  | 0.9815   | 0.9822    | 0.9802 | 0.9806 | 5.00M  | 420.23M |
| LN & Swish | 0.9785   | 0.9782    | 0.9771 | 0.9775 | 5.05M  | 407.68M |
| LN & GELU  | 0.9846   | 0.9857    | 0.9818 | 0.9831 | 5.05M  | 407.68M |

F1-score, parameter count, Flops and inference time. Among the existing classic CNN models, VGG16, ResNet50, and DenseNet169 all have an accuracy of around 95%, while lightweight models such as MobileNet and ShuffleNet have lower accuracy at around 94%. In

contrast, our model achieves an accuracy of 98.77% with only 3.44M parameters. Moreover, our model has precision, recall, and F1-score all greater than 98% with the lowest number of parameters, indicating strong generalization ability and significant advantages

TABLE 8 The performance of different layers on the test set.

| Row | Layers (Stage 5, 6, 7) | Accuracy      | Precision     | Recall        | F1            | Param        | Flops          |
|-----|------------------------|---------------|---------------|---------------|---------------|--------------|----------------|
| 1   | 2, 2, 2                | 0.9815        | 0.9826        | 0.9796        | 0.9806        | 3.17M        | 305.1M         |
| 2   | 2, 2, 3                | 0.9846        | 0.9854        | 0.9842        | 0.9846        | 3.90M        | 330.62M        |
| 3   | 2, 2, 4                | 0.9754        | 0.9763        | 0.9738        | 0.9748        | 4.64M        | 356.15M        |
| 4   | <b>2, 3, 2</b>         | <b>0.9877</b> | <b>0.9863</b> | <b>0.9876</b> | <b>0.9868</b> | <b>3.44M</b> | <b>339.74M</b> |
| 5   | 2, 3, 3                | 0.9846        | 0.9840        | 0.9853        | 0.9844        | 4.18M        | 365.26M        |
| 6   | 2, 3, 4                | 0.9815        | 0.9808        | 0.9818        | 0.9811        | 4.91M        | 390.79M        |
| 7   | 3, 2, 2                | 0.9846        | 0.9846        | 0.9845        | 0.9845        | 3.31M        | 321.99M        |
| 8   | 3, 2, 3                | 0.9785        | 0.9808        | 0.9776        | 0.9790        | 4.04M        | 347.52M        |
| 9   | 3, 2, 4                | 0.9846        | 0.9870        | 0.9849        | 0.9857        | 4.78M        | 373.04M        |
| 10  | 3, 3, 2                | 0.9846        | 0.9833        | 0.9844        | 0.9838        | 3.58M        | 356.64M        |
| 11  | 3, 3, 3                | 0.9815        | 0.9805        | 0.9806        | 0.9801        | 4.31M        | 382.16M        |
| 12  | Origin (3,3,4)         | 0.9846        | 0.9857        | 0.9818        | 0.9831        | 5.05M        | 407.68M        |

Bold indicates the optimal structure used in this paper.

TABLE 9 Ablation experiments on the test set.

| Method              | Accuracy | Precision | Recall | F1     | Param | Flops   |
|---------------------|----------|-----------|--------|--------|-------|---------|
| EfficientNetB0      | 0.9631   | 0.9650    | 0.9603 | 0.9619 | 4.02M | 398.03M |
| +AF                 | 0.9723   | 0.9726    | 0.9698 | 0.9710 | 4.02M | 399.38M |
| + FCA               | 0.9816   | 0.9848    | 0.9837 | 0.9841 | 5.00M | 420.23M |
| +LN & GELU          | 0.9846   | 0.9857    | 0.9818 | 0.9831 | 5.05M | 407.68M |
| Architecture Search | 0.9877   | 0.9863    | 0.9876 | 0.9868 | 3.44M | 339.74M |

over other models, making it more suitable for the recognition task of corn disease images.

<sup>1</sup>Model's inference time tested on NVIDIA RTX3090

Meanwhile, in order to further verified the robustness of the model, we changed the contrast and brightness of the test set, images to test the recognition accuracy of the model under different lighting

conditions. We used four different magnifications (0.5, 0.67, 1.5, 2) to change the contrast or brightness of the image, respectively. The images before and after enhancement are shown in Figure 16.

Table 11 lists the recognition accuracies of the different networks on the augmented test set, and it can be seen that the recognition accuracies of the networks all produce a large drop due to the substantial changes made to the images, but the recognition accuracies of our proposed network drop to a lesser extent compared to the other networks, which suggests that the model's is more robust under different lighting conditions.

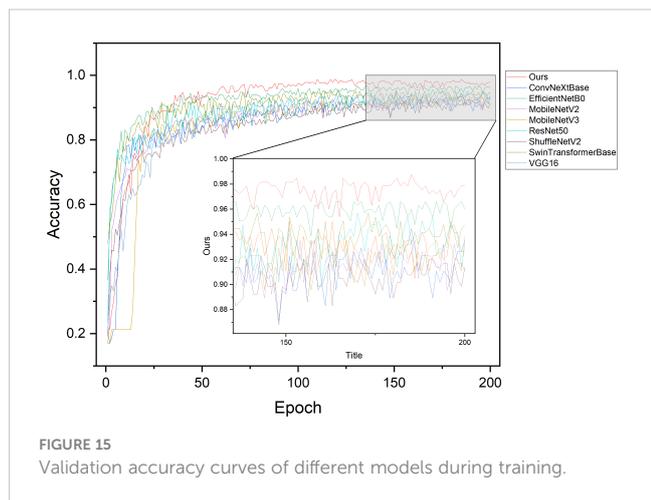


FIGURE 15 Validation accuracy curves of different models during training.

### 4.10 Comparison with relevant studies

To further evaluated the performance of the model, we compared it with some related studies that have been conducted. Table 12 lists some of the studies on corn disease recognition, including the techniques used, data sources, disease categories and numbers, and the recognition accuracy of corn images.

In these studies, CNNs showed significant performance strength, while conventional image processing techniques performed poorly in terms of recognition accuracy for multiple

TABLE 10 Performance and parameter comparison of different models on the test set.

| Model                | Accuracy      | Precision     | Recall        | F1            | Param        | Flops          | Inference Time <sup>1</sup> |
|----------------------|---------------|---------------|---------------|---------------|--------------|----------------|-----------------------------|
| ConvNeXt-Base        | 0.9446        | 0.9467        | 0.9433        | 0.9446        | 87.52M       | 15.35G         | 12.096ms                    |
| DenseNet             | 0.9508        | 0.9539        | 0.9456        | 0.9475        | 6.96M        | 2.90G          | 14.316ms                    |
| EfficientNet-B0      | 0.9631        | 0.9650        | 0.9603        | 0.9619        | 4.02M        | 398.03M        | 9.490ms                     |
| MobileNetV2          | 0.9477        | 0.9545        | 0.9441        | 0.9467        | 2.23M        | 318.96M        | 6.847ms                     |
| MobileNetV3          | 0.9262        | 0.9235        | 0.9227        | 0.9213        | 1.52M        | 58.79M         | 4.672ms                     |
| ResNet50             | 0.9538        | 0.9581        | 0.9501        | 0.9531        | 23.52M       | 4.12G          | 6.422ms                     |
| ShuffleNetV2         | 0.9200        | 0.9191        | 0.9185        | 0.9178        | 1.26M        | 149.58M        | 5.675ms                     |
| SwinTransformer-Base | 0.9569        | 0.9553        | 0.9550        | 0.9545        | 86.69M       | 15.17G         | 16.2103ms                   |
| VGG16                | 0.9508        | 0.9498        | 0.9487        | 0.9489        | 134.29M      | 15.5G          | 11.9590ms                   |
| <b>Ours</b>          | <b>0.9877</b> | <b>0.9863</b> | <b>0.9876</b> | <b>0.9868</b> | <b>3.44M</b> | <b>339.74M</b> | <b>7.218ms</b>              |

Bold indicates the final performance of the network.



FIGURE 16 Sample image after enhancement of the test set. Where f denotes the enhancement factor.

TABLE 11 Comparison of the performance of different models under different light conditions.

| Model                | Accuracy      | Precision     | Recall        | F1            |
|----------------------|---------------|---------------|---------------|---------------|
| ConvNeXt-Base        | 0.6614        | 0.7005        | 0.6769        | 0.6724        |
| DenseNet             | 0.6275        | 0.6958        | 0.6484        | 0.6345        |
| EfficientNet-B0      | 0.6704        | 0.6866        | 0.6709        | 0.6662        |
| MobileNetV2          | 0.5673        | 0.5888        | 0.5945        | 0.5739        |
| MobileNetV3          | 0.6451        | 0.6823        | 0.6560        | 0.6498        |
| ResNet50             | 0.9538        | 0.9581        | 0.9501        | 0.9531        |
| ShuffleNetV2         | 0.5215        | 0.6043        | 0.5320        | 0.5296        |
| SwinTransformer-Base | 0.6924        | 0.7229        | 0.6936        | 0.6868        |
| VGG16                | 0.6519        | 0.6847        | 0.6580        | 0.6444        |
| <b>Ours</b>          | <b>0.7389</b> | <b>0.7397</b> | <b>0.7489</b> | <b>0.7393</b> |

Bold indicates the final performance of the network.

TABLE 12 Performance comparison of relevant studies.

| Techniques  | Data sources   | Categories and quantities     | Acc (%) |
|---|----------------|-------------------------------|---------|
| SVM (Aravind et al., 2018)                            | PlantVillage   | 4 categories and 2000 images  | 83.7    |
| SVM (Kusumo et al., 2018)                             | PlantVillage   | 4 categories and 3852 images  | 87.2    |
| Improved LeNet<br>(Ahila Priyadharshini et al., 2019) | PlantVillage   | 4 categories and 3852 images  | 97.89   |
| DISE-NET (Yin et al., 2022)                           | Self-collected | 5 categories and 1268 images  | 97.12   |
| DMS-Robust Alexnet (Lv et al., 2020)                  | Self-collected | 7 categories and 12227 images | 98.62   |
| CNN (Mishra et al., 2020)                             | Self-collected | 3 categories and 4382 images  | 88.46   |
| SKPSNet-50 (Zeng et al., 2022a)                       | Self-collected | 6 categories and 1,452 images | 92.9    |
| LDSNet (Zeng et al., 2022b)                           | Self-collected | 6 categories and 3363 images  | 95.4    |
| DFCANet (Chen et al., 2022)                           | Self-collected | 6 categories and 3271 images  | 98.47   |
| Ours  | Self-collected | 6 categories and 3258 images  | 98.76   |

types of diseases (Aravind et al., 2018; Kusumo et al., 2018). Meanwhile, CNNs have shown better recognition performance in datasets with simple backgrounds (Ahila Priyadharshini et al., 2019; Lv et al., 2020; Yin et al., 2022), but the robustness of CNNs decreases when they are applied to corn disease identification in real environments (Zeng et al., 2022a). Moreover, our image dataset is similar to the studies of Chen et al. (2022) and Zeng et al. (2022b), but our accuracy is higher and the Flops is smaller. This indicates that FCA-EfficientNet has excellent computational efficiency and potential for practical implementation. Therefore, FCA-EfficientNet has certain performance advantages in the field of corn disease image recognition.

#### 4.11 Deployed on Android

We deploy the model on a mobile device to verify its performance. The mobile device used MediaTek Dimensity 1100, 8GB RAM, and Android 12. We conducted 5 consecutive experiments on this device, each experiment recognizing 1000 images. The average recognition speed for each image on the current device was: 88.34ms, 87.99ms, 91.87ms, 93.16ms, 93.02ms. The average recognition time for the 5 experiments was 92.88ms. Based on these current experimental results, the model can meet daily recognition needs. Figure 17 shows the detailed operation of the system.

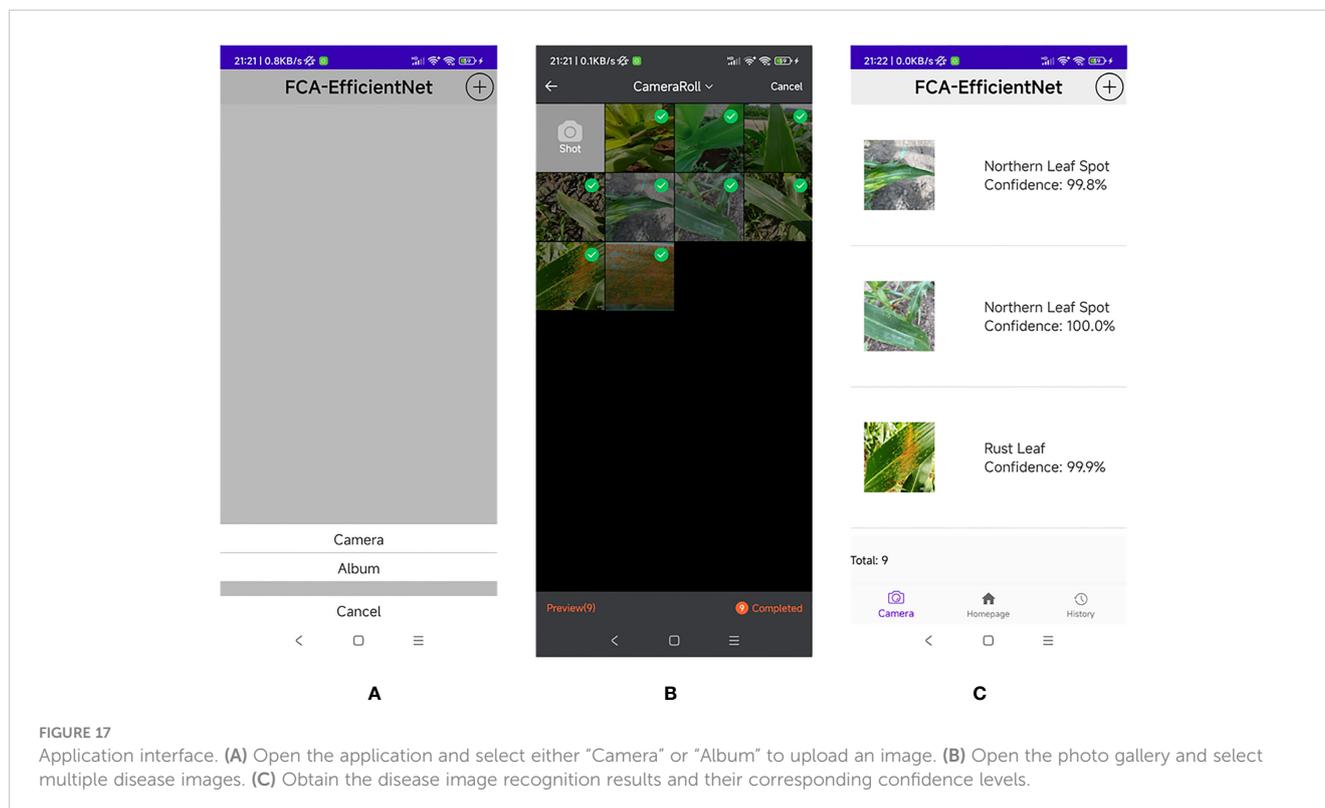


FIGURE 17

Application interface. (A) Open the application and select either "Camera" or "Album" to upload an image. (B) Open the photo gallery and select multiple disease images. (C) Obtain the disease image recognition results and their corresponding confidence levels.

## 5 Conclusions

In this study, we propose an improved EfficientNet model, FCA-EfficientNet, to achieve accurate identification of multiple corn diseases. To recognize corn diseases in the field environment, we designed the AF module for filtering background noise and the FCA module that enables the network to be able to focus on diseased areas. In addition, we used LN and GELU to improve the generalization ability of the network. Finally, we removed redundant network structures to make the model more lightweight. In the corn disease dataset, the classification accuracy of FCA-EfficientNet for six corn images is 98.78%, and all the evaluation metrics in the comparison experiments are better than the classical CNN model with good generalization ability. The number of parameters of FCA-EfficientNet is 3.44M, which meets the conditions for deployment on mobile devices. The average recognition speed of each image on the test device is 92.88ms, which has good recognition speed and meets the recognition needs of growers. However, at the present time, the model can only quickly identify a small number of common corn diseases and cannot cover the full range of corn diseases. In the future, we plan to use transfer learning on large-scale plant disease datasets to enhance the model performance, as well as autonomously collect images of more types of corn diseases to further optimize the utility of the model and provide technical support to relevant growers.

## Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: The datasets for this manuscript are not publicly available because this research has been carried out as part of a project funded by the Government of China. Requests to access these datasets should be directed to xzhang1@gzu.edu.cn.

## References

- Ahila Priyadarshini, R., Arivazhagan, S., Arun, M., and Mirnalini, A. (2019). Maize leaf disease classification using deep convolutional neural networks. *Neural Computing Appl.* 31, 8887–8895. doi: 10.1007/s00521-019-04228-3
- Ahmad, A., Saraswat, D., Gamal, A. E., and Johal, G. (2021). Cd&s dataset: Handheld imagery dataset acquired under field conditions for corn disease identification and severity estimation. arXiv preprint arXiv:2110.12084.
- Al-Gaashani, M. S., Samee, N. A., Alnashwan, R., Khayyat, M., and Muthanna, M. S. A. (2023). Using a resnet50 with a kernel attention mechanism for rice disease diagnosis. *Life* 13, 1277. doi: 10.3390/life13061277
- Aravind, K., Raja, P., Mukesh, K., Anirudh, R., Ashwin, R., and Szczepanski, C. (2018). "Disease classification in maize crop using bag of features and multiclass support vector machine," in *2018 2nd International Conference on Inventive Systems and Control (ICISC)*. (IEEE), 1191–1196.
- Ba, J. L., Kiros, J. R., and Hinton, G. E. (2016). Layer normalization. arXiv preprint arXiv:1607.06450.
- Chen, Y., Chen, X., Lin, J., Pan, R., Cao, T., Cai, J., et al. (2022). Dfcnet: A novel lightweight convolutional neural network model for corn disease identification. *Agriculture* 12, 2047. doi: 10.3390/agriculture12122047
- Chen, P., Liu, S., Zhao, H., and Jia, J. (2020). Gridmask data augmentation. arXiv preprint arXiv:2001.04086.
- DeVries, T., and Taylor, G. W. (2017). Improved regularization of convolutional neural networks with cutout. arXiv preprint arXiv:1708.04552.
- Guruprasad, P. (2020). "Overview of different thresholding methods in image processing," in *TEQIP Sponsored 3rd National Conference on ETACC*.
- Hendrycks, D., and Gimpel, K. (2016). Gaussian error linear units (gelus). arXiv preprint arXiv:1606.08415.
- Hou, Q., Zhou, D., and Feng, J. (2021). "Coordinate attention for efficient mobile network design," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 13713–13722.
- Hu, X., Wang, R., Du, J., Hu, Y., Jiao, L., and Xu, T. (2023). Class-attention-based lesion proposal convolutional neural network for strawberry diseases identification. *Front. Plant Sci.* 14, 1091600. doi: 10.3389/fpls.2023.1091600
- Ioffe, S., and Szegedy, C. (2015). "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*. (PMLR), 448–456.
- Kai, S., Zhikun, L., Hang, S., and Chunhong, G. (2011). "A research of maize disease image recognition of corn based on bp networks," in *2011 third international conference on measuring technology and mechatronics automation (IEEE)*, 1, 246–249.
- Kusumo, B. S., Heryana, A., Mahendra, O., and Pardede, H. F. (2018). "Machine learning-based for automatic detection of corn-plant diseases using image processing," in *2018 International conference on computer, control, informatics and its applications (IC3INA)*. (IEEE), 93–97.

## Author contributions

XC and XZ designed the experiments. JC, RP, JWJ, JML, LZ, and XW performed the experiments. JC, RP, and JWJ analyzed data. JC drafted the manuscript. XC and XZ conducted visualization and proofreading of the manuscript. All authors contributed to the article and approved the submitted version.

## Funding

This study was supported by the National Key Research and Development Program of China (2021YFE0107700), the National Natural Science Foundation of China (32360705; 31960555), Guizhou Provincial Science and Technology Program (2019-1410; HZJD[2022]001), Outstanding Young Scientist Program of Guizhou Province (KY2021-026), and Guizhou University Cultivation Project (2019-04). In addition, the study received support by the Program for Introducing Talents to Chinese Universities (111 Program, D20023).

## 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.

- Lin, J., Chen, X., Pan, R., Cao, T., Cai, J., Chen, Y., et al. (2022a). Grapenet: A lightweight convolutional neural network model for identification of grape leaf diseases. *Agriculture* 12, 887. doi: 10.3390/agriculture12060887
- Lin, J., Chen, Y., Pan, R., Cao, T., Cai, J., Yu, D., et al. (2022b). Camffnet: A novel convolutional neural network model for tobacco disease image recognition. *Comput. Electron. Agric.* 202, 107390. doi: 10.1016/j.compag.2022.107390
- Lin, J., Yu, D., Pan, R., Cai, J., Liu, J., Zhang, L., et al. (2023). Improved YOLOX-Tiny network for detection of tobacco brown spot disease. *Front. Plant Sci.* 14.
- Lv, M., Zhou, G., He, M., Chen, A., Zhang, W., and Hu, Y. (2020). Maize leaf disease identification based on feature enhancement and dms-robust alexnet. *IEEE Access* 8, 57952–57966. doi: 10.1109/ACCESS.2020.2982443
- Mishra, S., Sachan, R., and Rajpal, D. (2020). Deep convolutional neural network based detection system for real-time corn plant disease recognition. *Proc. Comput. Sci.* 167, 2003–2010. doi: 10.1016/j.procs.2020.03.236
- Nigam, S., Jain, R., Marwaha, S., Arora, A., Haque, M. A., Dheeraj, A., et al. (2023). Deep transfer learning model for disease identification in wheat crop. *Ecol. Inf.* 75, 102068. doi: 10.1016/j.ecoinf.2023.102068
- O'Mahony, N., Campbell, S., Carvalho, A., Harapanahalli, S., Hernandez, G. V., Krpalkova, L., et al. (2020). "Deep learning vs. traditional computer vision," in *Advances in Computer Vision: Proceedings of the 2019 Computer Vision Conference (CVC)*, (Springer) 1, 128–144.
- Praveen, S. P., Srinivasu, P. N., Shafi, J., Wozniak, M., and Ijaz, M. F. (2022). Resnet-32 and fastai for diagnoses of ductal carcinoma from 2d tissue slides. *Sci. Rep.* 12, 20804. doi: 10.1038/s41598-022-25089-2
- Ramachandran, P., Zoph, B., and Le, Q. V. (2017). Searching for activation functions. arXiv preprint arXiv:1710.05941.
- Rosas-Castor, J. M., Guzmán-Mar, J. L., Hernández-Ramírez, A., Garza-González, M. T., and Hinojosa-Reyes, L. (2014). Arsenic accumulation in maize crop (*zea mays*): a review. *Sci. Total Environ.* 488, 176–187. doi: 10.1016/j.scitotenv.2014.04.075
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 618–626.
- Singh, D., Jain, N., Jain, P., Kayal, P., Kumawat, S., and Batra, N. (2020). "Plantdoc: A dataset for visual plant disease detection," in *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*. 249–253.
- Srinivasu, P. N., Shafi, J., Krishna, T. B., Sujatha, C. N., Praveen, S. P., and Ijaz, M. F. (2022). Using recurrent neural networks for predicting type-2 diabetes from genomic and tabular data. *Diagnostics* 12, 3067. doi: 10.3390/diagnostics12123067
- Stephen, A., Punitha, A., and Chandrasekar, A. (2023). Designing self attention-based resnet architecture for rice leaf disease classification. *Neural Computing Appl.* 35, 6737–6751. doi: 10.1007/s00521-022-07793-2
- Sun, S., Hu, B., Yu, Z., and Song, X. (2020). A stochastic max pooling strategy for convolutional neural network trained by noisy samples. *Int. J. Comput. Commun. Control* 15 (1). doi: 10.15837/ijccc.2020.1.3712
- Tan, M., and Le, Q. (2019). "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*. (PMLR), 6105–6114.
- Vishnoi, V. K., Kumar, K., Kumar, B., Mohan, S., and Khan, A. A. (2022). Detection of apple plant diseases using leaf images through convolutional neural network. *IEEE Access* 11, 6594–6609. doi: 10.1109/ACCESS.2022.3232917
- Yin, C., Zeng, T., Zhang, H., Fu, W., Wang, L., and Yao, S. (2022). Maize small leaf spot classification based on improved deep convolutional neural networks with a multi-scale attention mechanism. *Agronomy* 12, 906. doi: 10.3390/agronomy12040906
- Zeng, W., Li, H., Hu, G., and Liang, D. (2022a). Identification of maize leaf diseases by using the skpsnet-50 convolutional neural network model. *Sustain. Computing: Inf. Syst.* 35, 100695. doi: 10.1016/j.eswa.2020.114514
- Zeng, W., Li, H., Hu, G., and Liang, D. (2022b). Lightweight dense-scale network (ldsnet) for corn leaf disease identification. *Comput. Electron. Agric.* 197, 106943. doi: 10.1016/j.compag.2022.106943
- Zhong, Z., Zheng, L., Kang, G., Li, S., and Yang, Y. (2020). "Random erasing data augmentation," in *Proceedings of the AAAI conference on artificial intelligence*, 34 (07), 13001–13008.