



OPEN ACCESS

EDITED BY

Uzair Aslam Bhatti,
Hainan University, China

REVIEWED BY

Mir Muhammad Nizamani,
Guizhou University, China
Sibghat Ullah Bazai,
BUIITEMS, Pakistan
Qiong Chen,
Tsinghua University, China
Ahmad Hasnain,
Nanjing Normal University, China

*CORRESPONDENCE

Mingxin Hou
houmx@gdou.edu.cn

SPECIALTY SECTION

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

RECEIVED 03 November 2022

ACCEPTED 21 November 2022

PUBLISHED 22 December 2022

CITATION

Liu M, Liang H and Hou M (2022)
Research on cassava disease
classification using the multi-scale
fusion model based on EfficientNet
and attention mechanism.
Front. Plant Sci. 13:1088531.
doi: 10.3389/fpls.2022.1088531

COPYRIGHT

© 2022 Liu, Liang and Hou. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

Research on cassava disease classification using the multi-scale fusion model based on EfficientNet and attention mechanism

Mingxin Liu¹, Haofeng Liang¹ and Mingxin Hou^{2*}

¹School of Electronic and Information, Guangdong Ocean University, Zhanjiang, China, ²School of Mechanical and Power Engineering, Guangdong Ocean University, Zhanjiang, China

Cassava disease is one of the leading causes to the serious decline of cassava yield. Because it is difficult to identify the characteristics of cassava disease, if not professional cassava growers, it will be prone to misjudgment. In order to strengthen the judgment of cassava diseases, the identification characteristics of cassava diseases such as different color of cassava leaf disease spots, abnormal leaf shape and disease spot area were studied. In this paper, deep convolutional neural network was used to classify cassava leaf diseases, and image classification technology was used to recognize and classify cassava leaf diseases. A lightweight module Multi-scale fusion model (MSFM) based on attention mechanism was proposed to extract disease features of cassava leaves to enhance the classification of disease features. The resulting feature map contained key disease identification information. The study used 22,000 cassava disease leaf images as a data set, including four different cassava leaf disease categories and healthy cassava leaves. The experimental results show that the cassava leaf disease classification model based on multi-scale fusion Convolutional Neural Network (CNN) improves EfficientNet compared with the original model, with the average recognition rate increased by nearly 4% and the average recognition rate up to 88.1%. It provides theoretical support and practical tools for the recognition and early diagnosis of plant disease leaves.

KEYWORDS

deep learning, classification, EfficientNet, multi-scale feature fusion, attention mechanism

1 Introduction

With the current climate posing a threat to human health, vegetation and biodiversity (Bhatti et al., 2022a), and the outbreak of the novel coronavirus pneumonia, major cities across the country have suspended production in order to effectively prevent the spread of the epidemic (Bhatti et al., 2022b). The importance of food is self-evident. In recent years, the planting area has continued to expand, and it is also a key food security crop for smallholder farmers because it can withstand harsh conditions. However, with the increase in cassava planting areas, the disease problem is becoming increasingly prominent. Cassava disease can generally be diagnosed according to the shape, color, and leaf shape characteristics of the disease spots on cassava leaves. Under the influence of environmental factors, cassava disease is more likely to occur, which affects the yield and quality of cassava. According to the different characteristics, cassava disease can be mainly classified into Cassava Bacterial Blight (CBB), Cassava Brown Streak Disease (CBSD), Cassava Green Mottle (CGM), Cassava Mosaic Disease (CMD), etc., which can lead to reduced and/or diseased cassava output. Due to the small number of professional plant personnel of cassava and the lack of professional knowledge of general plant personnel of cassava, the symptoms of cassava leaf disease are not typically studied with good understanding, which can lead to inaccurate and incorrect diagnoses of cassava disease. At the same time, the artificial diagnosis and treatment of cassava leaf diseases not only wastes a lot of manpower and material resources but also results in omission and error from subjective judgment due to the relatively similar characteristics of each leaf disease. The diversity of solutions for cassava disease often prevents effective treatment of cassava, so it is very important to correctly identify the disease. Therefore, as an auxiliary means, computer technology can be applied to help planting personnel determine whether there is cassava disease and of which type, and then the best treatment can be given, to avoid yield decline.

There have been many research achievements in judging plant diseases through traditional machine learning methods. Since they are all based on artificial designs of features, they are inefficient and have a large workload (Pujari et al., 2016). In addition, people tend to rely on experience when selecting features, which is highly subjective and not only consumes manpower but also has a large amount of uncertainty. Combining machine learning with deep learning can solve this problem well. Jamil (Jamil et al., 2021) used artificial neural network (ANN) and support vector machine (SVM) to solve the problem of land classification. When the accuracy of ANN was 82.60% and the accuracy of SVM was 73.66%, they combined the two models and weighted them, and finally the average accuracy reached 86.18%. A CNN can automatically extract image features, greatly reducing the workload, while providing a good research direction for plant disease classification.

Meanwhile, K-nearest Neighbor (KNN) classifier (Bazai et al., 2017) and other algorithms are also studied in deep learning in data analysis scenarios, Bhatti (Bhatti et al., 2021a) proposed a method for edge detection of color images by using Clifford algebra and its subalgebra quaternion in image processing, which improved object detection and classification as well as extraction of other features. Bhatti (Bhatti et al., 2021b) also proposed a spatial spectrum HIS classification algorithm – local similarity Projection Gabor Filtering (LSPGF), which uses the reduced-dimension Convolutional Neural Network based on local similarity projection (LSP) and two-dimensional Gabor filtering algorithm. The performance of the proposed method is compared with other algorithms in the public Host Integration Server (HIS) database, and the overall accuracy is better than all datasets. Based on the data information, Bhatti (Aamir et al., 2021; Bhatti et al., 2022c) uses regression analysis algorithm and path analysis algorithm to extract the relationship between variables and get the relationship between algorithms.

With the proposals of AlexNet (Krizhevsky et al., 2012) and Visual Geometry Group Network (VGG) (Simonyan and Zisserman, 2014), the number of network parameters is greatly reduced and the network is more suitable for complex samples under multiple training times, paving the way for deep learning to be applied in future computer vision. In the same year that VGG was proposed, Google proposed GoogleNet (Szegedy et al., 2015). This network adopts the Inception modular structure, by which convolutional kernels of different sizes are used to capture features of feature maps and expand their receptive fields, and then splice the results into channels. Finally, the accuracy of the network is improved by fusing multiple features. In the following years, the proposal of Resnet (He et al., 2016) residual network provided new ideas for the CNN. With the advent of EfficientNet (Tan and Le, 2019), the model has become more capable at capturing features, and its application in computer vision is developing day by day, especially in plant disease recognition.

Hewitt (Hewitt and Mahmoud, 2018) only used a simple shape feature set to judge and recognize relevant plant leaves, in which the feature set included shape features of original leaves and signal features extracted from different convolution models for recognition and obtained good recognition results. In the face of wheat disease leaf identification and differentiation, (Van Hieu and Hien, 2020) used a variety of classification algorithms to compare the prediction accuracy of various neural networks, among which GoogleNet proved to have the highest accuracy of 98%, more suitable for wheat disease detection. As for the detection and recognition of tea diseases, Lee (2018) used Faster Region Convolutional Neural Network (FR-CNN) and candidate objects proposed by Region Proposal Network (RPN) to detect, identify, and distinguish three kinds of tea diseases, with recognition accuracies up to 63%, 81%, and 64%, respectively. Two major crop damage modes in maize production were evaluated, and three commonly used object

detectors were evaluated. It was concluded that YOLOv2 had better performance and was more suitable for the assessment of maize growth damage (Turkoglu and Hanbay, 2019). The adversarial robustness of the model (Zoran et al., 2020) was significantly improved by adding an attention mechanism, and the robustness was effectively improved by changing the model expansion steps. On the lightweight model, Wang (Wang et al., 2022) proposed an Individualized activity space modeler (IASM) mechanism to improve the accuracy and efficiency of the model, and achieved the classification accuracy of 92.57% on the self-made data set by using Ghostnet and Weighted Boxes Fusion (WBF) structures. In the classification of banana diseases, Narayanan (Narayanan et al., 2022) combined the mixed algorithm of CNN and Fuzzy Support Vector Machine (FSVM) to classify banana diseases, CNN to detect, and FSVM algorithm to strengthen the classification, and finally achieved good results. In terms of attention mechanism, Zhu (Zhu et al., 2021) added an attentional mechanism module combining Convolutional Block Attention Module (CBAM) and ECA-Net module to the model, which improved the accuracy of the model by 3.4%. Zakzouk, S. (Zakzouk et al., 2021) used AlexNet to classify new rice diseases with an accuracy of 99.71%. The accuracy of the results indicated the feasibility of the automatic rice disease classification system. Tang (Tang et al., 2020) proposed a new two-stage Convolutional Neural Network image classification network. InnerMove, a new image data enhancement method, was used to enhance images and increase the number of training samples, so as to improve the generalization ability of the deep CNN model for image classification tasks. There are also many innovative neural network methods on algorithm models that can provide ideas for classification. At present, the 3D Convolutional Neural Network method (Hameed et al., 2022a) is innovatively used to extract the feature information for the data set. This method can solve the problem with the data better than the pixel-based support vector machine classifier. When it comes to the impact of food security on local and global economies, Mazhar (Hameed et al., 2022b) applied the sequential model in deep learning to classify the outer layer air particles through the analysis and characteristics of objects and fusion. Compared with the existing deep learning method of surface landscape, the accuracy rate reaches 98%.

In view of the above mentioned contents and problems, in order to improve the efficiency of cassava disease classification and recognition, this paper uses the deep convolutional neural network method according to the characteristics of crop disease images in real scenes, takes cassava leaf disease images as the research object, and designs a cassava disease classification model based on multi-scale fusion CNN. A multi-scale fusion module is proposed to extract multi-scale information features of images. Focal loss was adopted to reduce the emphasis on most categories caused by data imbalance, and to solve the problem of low classification accuracy for categories with few

samples. CBAM (Woo et al., 2018) module was introduced to obtain key information such as texture and color of cassava leaves, and the result of precise positioning of specific features was achieved.

We enhance and amplify the existing data images and add these images to the existing data set to form new mixed data for training the model. The effectiveness of the proposed model was verified by designing several comparison experiments and comparing them with network models such as Resnet and VGG.

The rest of this article is organized as follows. The Materials and Methods section introduces our proposed cassava disease classification model based on multi-scale fusion Convolutional Neural Network. See the results section for experimental results. Finally, in the conclusion part of the summary of this article.

To summarize, the main contributions of this study are as follows:

- (i) a model is developed to recognize cassava disease based on Convolutional Neural Network deep learning.
- (ii) the accuracy of cassava disease classification model is evaluated using images taken from nature and artificial enhanced images.
- (iii) a lightweight module based on attention mechanism to enhance the classification accuracy of the cassava disease model.

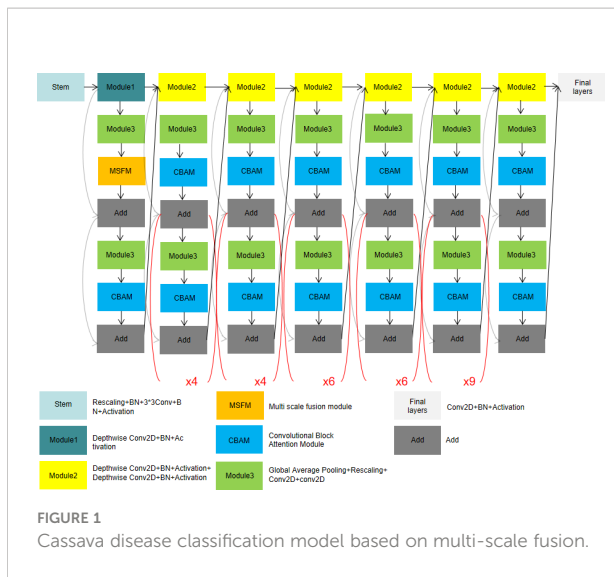
2 Method and materials

2.1 Cassava disease classification model based on multi-scale fusion Convolutional Neural Network

In this paper, a classification model of cassava disease was proposed based on multi-scale fusion (shown in Figure 1). The proposed model adopts Efficientnet-B6 as the backbone network, by which a multi-scale fusion module is designed to improve shallow feature extraction. Through CBAM, the channel and spatial weights of subsequent modules are recalibrated and the classification capacity of cassava leaf disease is enhanced.

2.1.1 EfficientNet

We generally expand the network scale by increasing network depth D , receptive field W , and resolution R . Compared with AlexNet, VGG network convolution kernels were all replaced with smaller 3×3 convolution kernels (including a few 1×1 convolution kernels), which achieved better training results through deeper network structure. However, with the gradual deepening of network depth, network training becomes more difficult due to the emergence of Vanishing Gradient, over-



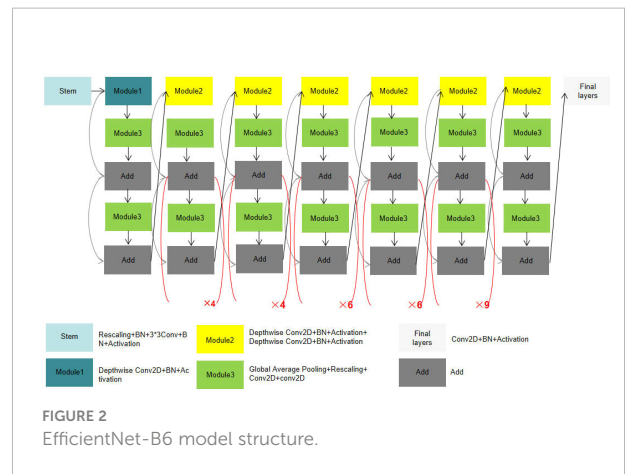
fitting, and other problems. Even if the problem of Vanishing Gradient is solved, the low precision return will lead to high calculation cost and low efficiency of increasing network depth blindly. For another example, MobileNet (Howard et al., 2017) can adjust the number of feature data channels by reducing the amount of computation. However, like deepening the network structure, low precision return will also be found when the width of the network structure reaches a certain level. In the final approach, the neural network can capture finer patterns by using higher-resolution input images. It has developed from 224×224 pixels to 229×229 pixels, or even 512×512 pixels. However, accuracy problems are inevitable as the parameter becomes larger. Before EfficientNet, network improvements were generally realized by changing only one of the following variables of network depth, receptive field, and resolution size. However, EfficientNet can obtain better training results by increasing the depth, receptive field, and image resolution through one adjustment. Compared with the aforementioned model, EfficientNet can get a better result because it is capable of adjusting the proportions in three dimensions (shown in Figure 2).

2.1.2 Proposed model

2.1.2.1 Multi-scale fusion module

The low-level feature map has a small stride, a large size, and a small receptive field to detect the details of small targets. The high-level feature map has a larger stride, smaller size, larger receptive field, and rich semantic information. The model extracts detailed features such as color and texture from the low-level network and extracts the blade shape feature from the high-level network.

The key features in the map can be selectively enhanced and the features can be accurately located by redistributing both channel and spatial weights through the Attention Mechanism.



Compared with the previous EfficientNet model, which did not include the MSFM module, the new model adds modules based on the attention mechanism to allocate computing resources to more important tasks. The operations of different pooling layers in channel and space were added to enhance important features and reduce the proportion of unnecessary features. At the same time, the sensitivity field of the low-level feature map augmented by expansion convolution with different expansion rates can not only extract the details of color and texture, but also fully obtain the context information of the image. In the end, the feature map output obtained by different expansion rates is fused to improve the classification effect.

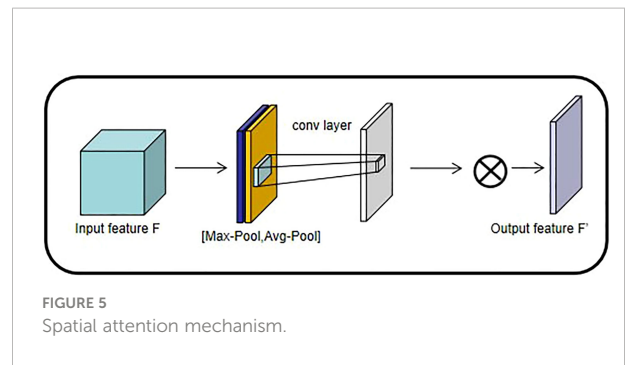
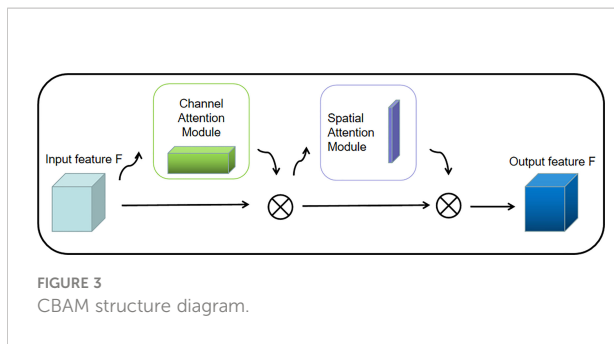
In the result part, we can see from the comparison experiment that our model has higher accuracy compared with other models, and in the heat map, we can see that it has more accurate judgment on key points.

2.1.2.1.1 Convolutional Block Attention Module

The attention mechanism is a relatively efficient data processing method developed in machine learning in recent years, and is widely used in various types of machine learning tasks such as image recognition and natural language processing. When people observe things outside, they usually focus on what they think is important. The attention mechanism focuses on local information that allows the network to achieve better results. Therefore, in this paper, Convolutional Block Attention Module is added before each regularization of the pre-training network and improves the features of the selected maps to increase the accuracy of the model (shown in Figure 3).

2.1.2.1.1.1 Channel Attention Mechanism (CAM)

SENet (Hu J. et al., 2018), as the champion network of the 2017 ImageNet classification Contest, is essentially a model based on a channel attention mechanism, which gives rewards and punishments of different weights according to the importance of each feature channel. In this paper, the channel attention mechanism adopts avg-pool and max-pool for fusion.



After convolution and activation of the Relu function, the results of the two pooling layers are added together. Finally, it is outputted by the Sigmoid function (shown in Figure 4). The size of the input feature map is $H \times W \times C$. Firstly, the global maximum pooling layer and average pooling layer are carried out, respectively, to obtain two-channel weight matrices of $1 \times 1 \times C$.

The two matrix results are fed into a two-layer multilayer perceptron (MLP), respectively, and the MLP can share parameters.

After adding the two feature vectors, the weight coefficients are obtained by the Sigmoid activation function again.

The weight coefficient is multiplied by the original input feature to obtain the final output feature.

2.1.2.1.1.2 Spatial Attention Mechanism (SAM)

Different from the weight of each feature plane of the channel attention allocation, the spatial attention model is to find the most important part of the network for processing (shown in Figure 5).

The input characteristic graph is $H \times W \times C$. The Max-Pool and Avg-Pool of one dimension are spliced and dimensionally reduced to generate two one-dimensional feature maps.

The weight parameters are generated by the Sigmoid activation function, and then the final output feature is obtained by multiplying the original input feature.

2.1.2.1.2 Multi-scale fusion module

First, the MSFM redistributes the channel and spatial information through the Convolutional Block Attention Module to enhance the characteristics of small target information. Then Feature Extraction is carried out using

dilated convolution of different dilatation rates, and the context information of feature images is fully extracted by expanding different receptive fields. The convolution kernel size of dilated convolution (Yu and Koltun, 2015) is the same as that of ordinary convolution, and the number of parameters in the neural network remains unchanged. The difference lies in that the dilated convolution has a larger receptive field. A 3×3 convolution kernel with an expansion rate of two has the same receptive field as a 5×5 convolution kernel. However, the number of parameters is only nine, much less than the 25 parameters of the 5×5 convolution kernel. The size of the convolution kernel after expansion:

$$k_d = k + (k - 1) \times (r - 1) \tag{1}$$

where

k_d = Size of the expanded convolution kernel

k = Size of original convolution

r = Dilatation rate

The calculation of the receptive field of dilated convolution is as follows:

$$r_f = [(k+1) \times (r-1)+k] \times [(k+1) \times (r-1)+k] \tag{2}$$

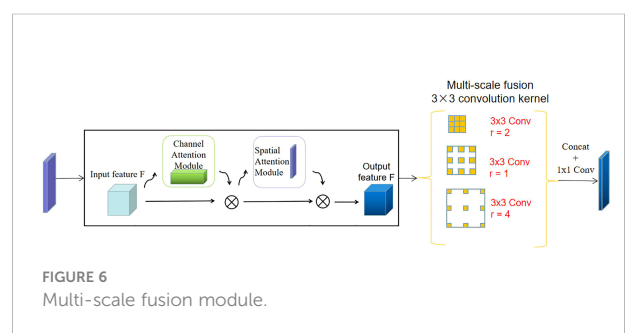
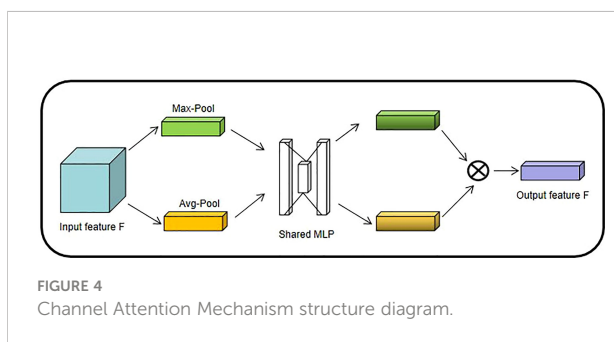
where

r_f = Receptive field

k = Size of original convolution

r = Dilatation rate

The multi-scale fusion module designed in this paper is shown in Figure 6. First, feature maps are learned by the Convolutional Block Attention Module, and weight calibration



is carried out for channel and space to strengthen the weight reward of important features of feature maps. The invalid features are punished and the weight is reduced to highlight the important information of the detected image. In this paper, a 3×3 convolution kernel is used for feature extraction, and the expansion convolution with expansion rates of 1, 2, and 4 are used for feature extraction, respectively. The features of different scales are extracted from each layer, and the new features are obtained after dimensionality reduction by fusion.

2.2 Loss function

2.2.1 Cross-entropy loss function

In recent years, transfer learning has been widely used in machine learning, which presents satisfactory application results in deep learning. For multi-classification tasks, the cross-entropy loss function (Hu K. et al., 2018) is generally used.

The most commonly used cross-entropy loss function is:

$$CE(p_t) = -a_t \log(p_t) \quad (3)$$

where

$CE = Loss$

$P_t = Predictive\ value$

$a_t = Added\ parameters\ that\ represent\ weights\ for\ different\ categories$

Cross entropy loss function under multiple classifications:

$$CE = -\sum_{j=1}^N a_j \log(p_j) \quad (4)$$

(b) Focal loss

To a certain extent, traditional methods can solve the problem of fewer categories and unbalanced image distribution, but when there are many easily classified samples, the samples will still dominate the training process, so some difficult-to-classify samples have little chance of gaining the attention of the model. The focus function (Lin et al., 2017) treats the difficult-to-classify samples and the easy-to-classify samples differently, focusing on the difficult-to-classify samples and reducing the weight of the easy-to-classify samples. Therefore, the focus function is adopted as the loss function in this study.

Focal loss adjusts the weights of the difficult-to-classify and easy-to-classify samples in the formula:

$$FL(p_t) = -a_t(1-p_t)^\gamma \log(p_t) \quad (5)$$

γ is a constant, and the magnitude of γ determines the weight of small and difficult samples.

When $\gamma < 0$, the focusing parameter can be adjusted. When the value of γ is larger, the loss of the sample that is easy to classify is small, and the focus of the model is on the sample that is difficult to classify. This is because when γ is larger, the loss of

small samples and difficult samples will be larger, so that they can obtain greater weight.

When $\gamma > 1$, the training loss of large and simple samples can be reduced, while the loss of small and difficult samples will not be reduced much.

When $\gamma = 1$, the equation degenerates into the cross-entropy loss function mentioned above.

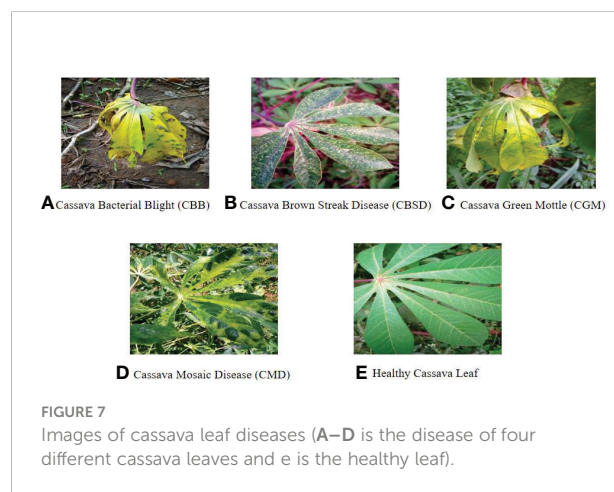
The focus loss function of multi-classification:

$$FL = -\sum_{j=1}^N (1-p_j)^\gamma \log(1-p_j) \quad (6)$$

2.3 Input dataset

2.3.1 Data sources and features

The data used in this paper are a cassava leaf dataset manually photographed in Uganda and annotated by experts from the National Crop Resources Institute in collaboration with the Artificial Intelligence (AI) Laboratory at Makerere University, Kampala. The data set contains five kinds of cassava leaf images, and disease images are cassava white leaf blight, brown streak disease, green mottling disease, Mosaic disease and healthy cassava leaf images. It can better reflect the characteristics and symptoms of healthy cassava leaves and diseased cassava leaves in natural environment, and it also represents the real and low diagnostic format that farmers need in real life. The dataset includes images taken under field conditions (some of which are shown in Figure 7). And images enhanced by data. There are 22,000 of them. Each image has a pixel size of 800×600. The standard input size of neural networks such as ResNet, EfficientNet and Resnet-50 is 224×224 pixels. The whole data set was randomly divided into a training set (90%) and a test set (10%). Therefore, 19,800 images were used for model training, and the remaining 2,200 images were used to test the performance of the model.



2.3.2 Data augmentation

To prevent network overfitting, OpenCV data enhancement was used to expand the data set appropriately. The data set was enhanced by increasing the brightness, decreasing the brightness, and reversing the image, and the data of various cassava leaves amounted to 4000 pieces. The data enhancement methods used are as follows (shown in Figures 8–12):

Rotation: Rotate the image by 180°.

Brightness reduction: The enhancement factor is 0.7, which means the brightness becomes 70% of the original image.

Horizontal flip: Flip the input image horizontally.

2.4 Computer hardware

The proposed method was tested for training and test configuration for neural network models (shown in Table 1). This result is achieved in Table 2, where PyTorch, as a popular learning framework today, is capable of powerful GPU acceleration and includes deep neural networks. Meanwhile, GPU RTX2080Ti has 11 GB of memory, which can better train the model by adjusting batch size.

2.5 Experimental hyperparameter setting

The cassava leaf data set was divided into a training set, a validation set, and a test set with the ratio of 8:1:1. The training set was trained as an epoch 150 times, 1e-4 was selected as the



FIGURE 8
Cassava Bacterial Blight. (A) original (B) Rotation by 180° (C) Brightness reduction (D) Horizontal flipping.



FIGURE 9
Cassava Brown Streak Disease. (A) original (B) Rotation by 180° (C) Brightness reduction (D) Horizontal flipping.



FIGURE 10
Cassava Green Mottle. (A) original (B) Rotation by 180° (C) Brightness reduction (D) Horizontal flipping.

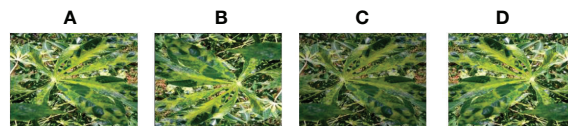


FIGURE 11
Cassava Mosaic Disease. (A) original (B) Rotation by 180° (C) Brightness reduction (D) Horizontal flipping.



FIGURE 12
Healthy cassava leaf. (A) original (B) Rotation by 180° (C) Brightness reduction (D) Horizontal flipping.

TABLE 1 Experimental configuration.

Experimental environment	Model and version
Deep learning framework	Pytorch
Programming language	Python3.7
GPU	NVIDIA GeForce RTX 2080 Ti
The hardware environment	Intel(R) Xeon(R) Silver 4110 CPU @ 2.10 GHz 2.10 GHz

initial value of the learning rate in the form of a small amount, and the batch size was set to 16. Batch size not only affects the efficiency of the training model but also affects the accuracy. To find a group balance between efficiency and memory capacity, the batch size is used to calculate the batch size.

2.6 Model evaluation criteria

Accuracy (%) is used as an evaluation index for multi-classification problems in the laboratory. The accuracy of the

TABLE 2 Classification accuracy of the test set in different models.

Model	Accuracy of the test set/%
VGG-16	70.5
Resnet-18	83.2
Resnet-34	83.9
Resnet-50	85.2
Inception v2	84.1
Inception v3	85.3
MobileNet v3	84.1
ShuffleNet	83.5
EfficientNet-b3	85.5
EfficientNet-b6+	86.5
Our model	88.1

experimental model classification can be obtained by removing the number of labels in the test set according to the evaluation criteria. The calculation formula of indicators is as follows:

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (7)$$

TP=Positive sample prediction is the number of positive classes.

FP=Negative sample prediction is the number of negative classes.

3 Result

3.1 Validation and comparison of proposed Convolutional Neural Network

In order to verify that the improved model in this paper has better image recognition ability compared with the traditional model, this paper uses multiple groups of comparative experiments. The experimental results are shown in Table 2. According to the evaluation index results, as shown in the figure above, the improved model is compared with the model with better performance recently. Among them, the accuracy rate of the model proposed in this paper is 88.1% in the test set separated from the data set, both of which are better than the previous models. To verify that the improved model in this paper has better image recognition ability compared with the traditional model, multi-group comparative experiments were carried out (shown in Table 2).

3.2 Ablation experiments

The confusion matrix (Song et al., 2015) is also one of the evaluation indicators of the classification model (shown in Figure 13). The confusion matrix parameters are converted by classification report. The parameters are given as follows:

(i)The percentage of the total that all predictions are correct, as in Equation 1:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

(ii)The proportion of those correctly predicted to be positive that are actually positive, as in Equation 2:

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

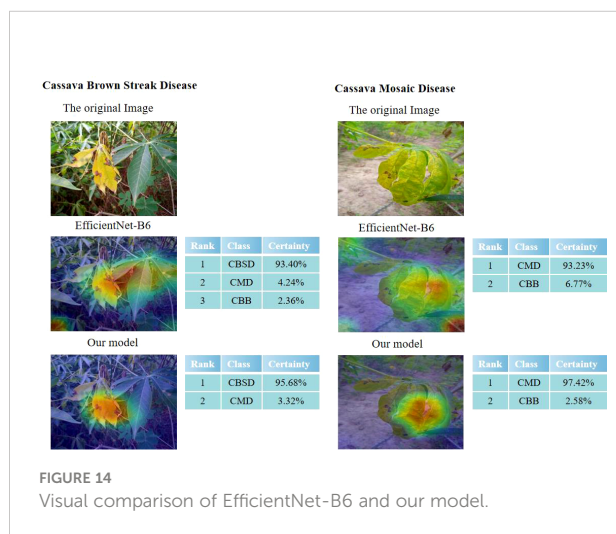
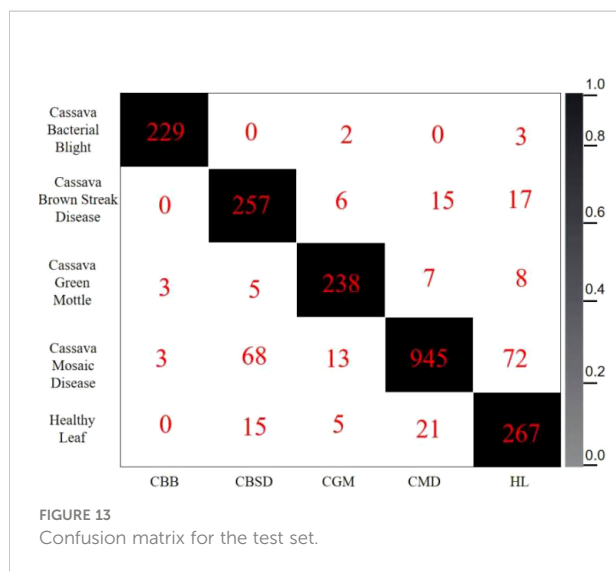
(i)The proportion of correct positive precisions to total positive precisions, as in Equation 3:

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

The column labels of the confusion matrix represent the predicted cassava leaf disease type, and the sum of the corresponding row values represents the sum of the samples of this type. The diagonal line indicates the number of correctly predicted labels. Each value on the diagonal line indicates the number of correctly predicted labels. The value at the intersection of the columns represents the value of the corresponding tag predicted. If it is not on the diagonal, it can be seen as the number of wrongly predicted tags. The darker diagonal suggests the better model. In the classification results, the judgment accuracy is high, and most of the results of the test set are concentrated on the diagonal of the confusion matrix. The identification accuracy of all kinds of blades is greater than 90%. However, the identification accuracy of Cassava Mosaic Disease in the test set is lower than that of other diseases, and it is easy to misjudge it as other cassava leaf diseases. By observing the confusion matrix, it can be found that the pictures of Cassava Mosaic Disease can be easily identified as Cassava Bacterial Blight and Cassava Green Mottle, because the symptoms of these three diseases are relatively similar, so classification errors are prone to occur. The obfuscation matrix of the improved model presented in this paper has better performance and a higher average recognition rate.

3.3 Visual output comparative analysis

Class Activation Mapping (CAM) was used to visualize each trained model (Zhou et al., 2016) to better compare the expression process of image features between an improved network and a traditional network. Feng (Feng et al., 2022) used the Grad-CAM thermal map of interpretative analysis, the feature extraction effect of the model can be better expressed. In visualization, the thermal map and the original image are superimposed (shown in Figure 14). This is a visual output of the original image and Efficientnet-B6 and our model, respectively. The darker the color is, the larger the value and the more feasible it will be to serve as the judgment. Compared with the model in this paper, our model has a stronger feature extraction ability and better effects in the face of Cassava leaf disease, which not only extracts different colors of cassava leaf disease but also better captures features in the context.



In addition, it can achieve more accurate extraction of key information. It can also be seen from the comparative experiment that the original model has some judgment errors in the color discrimination of leaves and the discrimination ability of background information, while the improved one has a better grasp of sample information and a better capture effect for judgment features.

4 Conclusions

Cassava, as one of the three varieties of potato, is an important food. Compared with the diagnosis of cassava disease by human, the identification of cassava disease by

computer and deep learning method not only has low cost and higher accuracy than manual diagnosis, but also greatly reduces the efficiency. The multi-scale cassava leaf classification model proposed in this paper can better ensure the safety and efficiency of cassava food production and judge the cassava disease type more accurately. Compared with the diagnosis of cassava leaf disease using the human eye, the identification of cassava leaf disease by computer and deep learning method is characterized by lower cost, higher accuracy, and greatly increased efficiency. In this paper, a multi-scale fusion module was proposed, and the Focal Loss function and CBAM module were introduced. An optimization network model of a multi-scale fusion network based on EfficientNet and attention mechanism was proposed. The model was used to train the cassava leaf disease data set and compare with the EfficientNet, ResNet50, and VGG16 networks. The experimental results show that the improved network proposed in this paper has higher precision and better generalization ability. The problem of uneven data was solved by changing the loss function, the distinguishing ability of cassava leaf diseases was improved through an attention mechanism, and the recognition ability of the model was enhanced by multi-layer fusion. According to the pricing standard of the model, the model proposed in this paper can be used for image recognition of Cassava leaf disease.

Since our model adds the lightweight module MSFM based on EfficientNet, our model can be installed on mobile devices, such as microprocessors. Due to the large-scale application of 5th Generation Mobile Communication Technology (5G) (Johannes et al., 2017), there is efficient transmission. According to the improvement of the hardware configuration of mobile terminal equipment, the image to be detected can be uploaded to the cloud server for processing, and then the recognition and classification results can be returned to the terminal. For some cassava planting technicians, when they have doubts about cassava disease judgment, the mobile terminal deployed with the model can detect and classify cassava disease in real time, which is equivalent to having a valuable consulting tool. In the future, a cassava disease detection system can be developed based on the classification results of cassava diseases, which can judge the disease categories and provide corresponding management methods. This can greatly improve the planting efficiency of cassava, improve the production efficiency of cassava, achieve scientific and technological progress of agriculture, and promote agriculture into the era of intelligence. Although the model in this study achieved a good success rate in a limited number of cassava diseases, cassava diseases are not limited to these diseases. To improve this, more images can be collected in different cassava planting areas and field conditions, and the model can be more effective in identifying cassava diseases based on field conditions by amplifying the dataset.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: <https://www.kaggle.com/competitions/cassava-leaf-disease-classification/data>.

Author contributions

ML: Modeling and experimental design of cassava leaf disease; HL: Experimental design and data analysis and arrangement; MH: AI model selection and debugging. All authors contributed to the article and approved the submitted version.

Funding

This work was partly supported by the National Natural Science Foundation of China (62171143, 6187465) and the Innovation and Entrepreneurship Team Induced Navigation Plan Project of Zhanjiang (2020LHJH003).

References

- Amir, M., Li, Z., Bazai, S., Wagan, R. A., Bhatti, U. A., Nizamani, M. M., et al. (2021). Spatiotemporal change of air-quality patterns in hubei province—a pre-to post-COVID-19 analysis using path analysis and regression. *Atmosphere* 12 (10), 1338. doi: 10.3390/atmos12101338
- Bazai, S. U., Jang-Jaccard, J., and Wang, R. (2017). “Anonymizing k-NN classification on MapReduce,” in *International conference on mobile networks and management* (Cham: Springer), 364–377.
- Bhatti, U. A., Ming-Quan, Z., Qing-Song, H., Ali, S., Hussain, A., Yuhuan, Y., et al. (2021a). Advanced color edge detection using Clifford algebra in satellite images. *IEEE Photonics J.* 13 (2), 1–20. doi: 10.1109/JPHOT.2021.3059703
- Bhatti, U. A., Yu, Z., Chanussot, J., Zeeshan, Z., Yuan, L., Luo, W., et al. (2021b). Local similarity-based spatial-spectral fusion hyperspectral image classification with deep CNN and gabor filtering. *IEEE Trans. Geosci. Remote Sens.* 60, 1–15. doi: 10.1109/TGRS.2021.3090410
- Bhatti, U. A., Nizamani, M. M., and Mengxing, H. (2022a). Climate change threatens pakistan’s snow leopards. *Science* 377 (6606), 585–586. doi: 10.1126/science.add9065
- Bhatti, U. A., Zeeshan, Z., Nizamani, M. M., Bazai, S., Yu, Z., and Yuan, L. (2022b). Assessing the change of ambient air quality patterns in jiangsu province of China pre-to post-COVID-19. *Chemosphere* 288, 132569. doi: 10.1016/j.chemosphere.2021.132569
- Bhatti, U. A., Wu, G., Bazai, S. U., Nawaz, S. A., Baryalai, M., Bhatti, M. A., et al. (2022c). A pre-to post-COVID-19 change of air quality patterns in anhui province using path analysis and regression. *Polish J. Environ. Stud.* 31 (5), 4029–4042. doi: 10.15244/pjoes/148065
- Feng, S., Liu, Q., Patel, A., Bazai, S. U., Jin, C. K., Kim, J. S., et al. (2022). Automated pneumothorax triaging in chest X-rays in the new Zealand population using deep-learning algorithms. *J. Med. Imaging Radiat. Oncol.* 66 (8), 1035–1043. doi: 10.1111/1754-9485.13393
- Hameed, M., Yang, F., Bazai, S. U., Ghafoor, M. I., Alshehri, A., Khan, I., et al. (2022a). Convolutional autoencoder-based deep learning approach for aerosol emission detection using LiDAR dataset. *J. Sensors* 2022, 17. doi: 10.1155/2022/3690312
- Hameed, M., Yang, F., Bazai, S. U., Ghafoor, M. I., Alshehri, A., Khan, I., et al. (2022b). Urbanization detection using LiDAR-based remote sensing images of azad Kashmir using novel 3D CNNs. *J. Sensors* 2022, 9. doi: 10.1155/2022/6430120
- Hewitt, C., and Mahmoud, M. (2018). Shape-only features for plant leaf identification. *arXiv preprint arXiv* 1811, 08398. doi: 10.48550/arXiv.1811.08398
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778. doi: 10.48550/arXiv.1512.03385
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., et al. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv* 1704, 04861. doi: 10.48550/arXiv.1704.04861
- Hu, J., Shen, L., and Sun, G. (2018). “Squeeze-and-excitation networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 7132–7141. doi: 10.1109/CVPR.2018.00745
- Hu, K., Zhang, Z., Niu, X., Zhang, Y., Cao, C., Xiao, F., et al. (2018). Retinal vessel segmentation of color fundus images using multiscale convolutional neural network with an improved cross-entropy loss function. *Neurocomputing* 309, 179–191. doi: 10.1016/j.neucom.2018.05.011
- Jamil, A., ali Hameed, A., and Bazai, S. U. (2021). Land cover classification using machine learning approaches from high resolution images. *J. Appl. Emerging Sci.* 11 (1), 108. doi: 10.36785/jaes.111501
- Johannes, A., Picon, A., Alvarez-Gila, A., Echazarra, J., Rodriguez-Vaamonde, S., Navajas, A. D., et al. (2017). Automatic plant disease diagnosis using mobile capture devices, applied on a wheat use case. *Comput. Electron. Agric.* 138, 200–209. doi: 10.1016/j.compag.2017.04.013
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Commun. ACM.* 60 (6). doi: 10.1145/3065386
- Lin, T. Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). “Focal loss for dense object detection,” in *Proceedings of the IEEE international conference on computer vision*. 2980–2988. doi: 10.48550/arXiv.1708.02002
- Narayanan, K. L., Krishnan, R. S., Robinson, Y. H., Julie, E. G., Vimal, S., Saravanan, V., et al. (2022). Banana plant disease classification using hybrid convolutional neural network. *Comput. Intell. Neurosci.* 2022, 13. doi: 10.1155/2022/9153699
- Pujari, D., Yakkundimath, R., and Byadgi, A. S. (2016). SVM and ANN based classification of plant diseases using feature reduction technique. *IJIMAI* 3 (7), 6–14. doi: 10.9781/ijimai.2016.371
- Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv* 1409, 1556. doi: 10.48550/arXiv.1708.02002
- Song, Y., Wang, X., Lei, L., and Xing, Y. (2015). Evaluating dynamic reliability of sensors based on evidence theory and confusion matrix. *Control Decision* 30 (6), 1111–1115. doi: 10.1007/s10489-018-1188-0

Acknowledgments

The authors would like to thank the reviewers and editors for their help in improving this paper.

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.

- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1–9. doi: 10.1109/CVPR.2015.7298594
- Tang, C., Zhu, Q., Wu, W., Huang, W., Hong, C., and Niu, X. (2020). PLANET: improved convolutional neural networks with image enhancement for image classification. *Math. Problems Eng.* 2020, 1–10. doi: 10.1155/2020/1245924
- Tan, M., and Le, Q. (2019). "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning* (PMLR), 6105–6114. doi: 10.48550/arXiv.1905.11946
- Turkoglu, M., and Hanbay, D. (2019). Recognition of plant leaves: An approach with hybrid features produced by dividing leaf images into two and four parts. *Appl. Mathematics Comput.* 352, 1–14. doi: 10.1016/j.amc.2019.01.054
- Van Hieu, N., and Hien, N. L. H. (2020). Automatic plant image identification of Vietnamese species using deep learning models. *arXiv preprint arXiv* 2005.02832. doi: 10.14445/22315381/IJETT-V68I4P2055
- Wang, H., Shang, S., Wang, D., He, X., Feng, K., and Zhu, H. (2022). Plant disease detection and classification method based on the optimized lightweight YOLOv5 model. *Agriculture* 12 (7), 931. doi: 10.3390/agriculture12070931
- Woo, S., Park, J., Lee, J. Y., and Kweon, I. S. (2018). "CBAM: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*. 3–19. doi: 10.1007/978-3-030-01234-2_1
- Yu, F., and Koltun, V. (2015). Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv* 1511.07122. doi: 10.48550/arXiv.1511.07122
- Zakzouk, S., Ehab, M., Atef, S., Yousri, R., Tawfik, R. M., and Darweesh, M. S. (2021). "Rice leaf diseases detector based on AlexNet," in *2021 9th international Japan-Africa conference on electronics, communications, and computations (JAC-ECC)* (IEEE), 170–174. doi: 10.1109/JAC-ECC54461.2021.9691435
- Zhang, S. (2016). *Research on plant leaf images identification algorithm based on deep learning* (Beijing, China: Beijing Forestry University).
- Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., and Torralba, A. (2016). "Learning deep features for discriminative localization," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2921–2929. doi: 10.1109/cvpr.2016.319
- Zhu, L., Geng, X., Li, Z., and Liu, C. (2021). Improving yolov5 with attention mechanism for detecting boulders from planetary images. *Remote Sens.* 13 (18), 3776. doi: 10.3390/rs13183776
- Zoran, D., Chrzanowski, M., Huang, P. S., Goyal, S., Mott, A., and Kohli, P. (2020). "Towards robust image classification using sequential attention models," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 9483–9492. doi: 10.48550/arXiv.1912.02184