FCA-ResNet: An Improved Model with Enhanced Multi-Scale Feature Fusion and Coordinate Attention for Wheat Leaf Disease Classification

Hongyan Zang[1,2], Annie Anak Joseph[1,*], Shourong Zhang[3], Haiyan Fu[2], Lili Huang[2], Zhen Huang[2]

[1]Faculty of Engineering, Universiti Malaysia Sarawak, Sarawak, Malaysia

[2]College of Computer and Information Engineering, Qilu Institute of Technology, Jinan, China

[3]Shandong Vanform High Energy Physics Technology Co. Ltd., Jinan, China

Received 18 September 2018; received in revised form 03 December 2024; accepted 04 December 2024

DOI: https://doi.org/10.46604/ijeti.2024.14304

Abstract

Rapid and accurate identification of leaf disease is essential in intelligent agriculture. Current methods often struggle with balancing precision and speed. This research introduces the fusion coordinate attention and residual network (FCA-ResNet) model to improve classification accuracy while maintaining a lightweight structure for both healthy wheat leaves and five common wheat leaf diseases. FCA-ResNet incorporates a coordinate attention (CA) mechanism along with a multi-branch Inception module. The model consists of an Inception-based multi-branch structure and CA mechanism fusion module, which optimizes feature focus and weight allocation. Additionally, a multi-scale fusion module utilizes both channel and spatial attention mechanisms to effectively integrate shallow and deep features, improving the detection accuracy of small lesions. The multi-branch structure is designed to replace traditional multi-layer convolution, resulting in a lightweight model. The model achieves an average accuracy of 91.6% on custom datasets, demonstrating its effectiveness in plant disease detection for agriculture.

Keywords: convolutional neural networks, wheat leaf disease classification, coordinate attention, multi-scale feature fusion

1. Introduction

In agricultural production, the classification and diagnosis of leaf diseases are essential to ensuring the health of crops, especially wheat, an important food crop that provides food for more than 40% of the world's population. Accurate diagnosis of wheat leaf diseases is significant to wheat quality and agricultural economic development. With computer vision and deep learning development, image-based disease classification methods have gradually become a research hotspot. More scholars use deep convolutional neural networks (CNN) for smart agriculture [1-2]. A series of customized improved CNN models [3-8] have been applied to plant disease classification and achieved reliable results. High accuracy coexists with high model complexity, such as VGG-16, ResNet [9], DenseNet [10], and InceptionNet [11] trained on ImageNet. However, it is still challenging to deploy such models on mobile devices with limited computational resources, such as Internet of things devices [12].

The rapid detection and identification of crop diseases are expected to be implemented on mobile devices for more convenient applications. However, current lightweight CNN models cannot achieve exact classification results. For example, a disease detection model for apples based on a single shot multi-box detector (SSD) has been developed [13]. However, when deployed on mobile devices, the detection accuracy is only 83.12%. Kumar et al. [14] developed a CNN based on the DeepLens

* Corresponding author. E-mail address: jannie@unimas.my

English language proofreader: Chih-Wen Teng

classification and detection model on the mobile side. The model has 9,914 training parameters. Users can take pictures and upload images through their mobile phones and use cloud services and the trained model for disease diagnosis. Results can be obtained within less than 1 second, with an average inference time of 0.349 seconds. Aziz Hosen Foysal et al. [15] integrated the model into an Android application, which can quickly process images and give diagnostic results. The entire application is approximately 142 MB and can be used offline. It takes an average of 10 seconds to process an image.

Therefore, developing a model with higher accuracy and better suited for mobile deployment remains critical. An important reason why wheat diseases are challenging to identify is that the lesion size, shape, and color of different diseases are different, or the lesion of the same disease will show noticeable differences under the influence of environmental factors, varieties, and other variables. These problems pose a significant challenge to the classification of diseases. When dealing with field images featuring complex backgrounds, traditional methods or pre-trained models for feature extraction may no longer suffice to meet actual needs.

To improve the classification accuracy and keep the model lightweight, a shortened model combining coordinate attention (CA) mechanism and multi-branch Inception module is designed to classify wheat leaf diseases. The proposed model deals with disease spots of varied sizes and shapes through a multi-branch network structure and enhances the model's ability to perceive prominent features with the help of a CA mechanism. The main contributions of this study are summarized as follows:

(1) The Inception coordinate attention (INCA) module integrates the CA mechanism and multi-branch difference hierarchical stacking to emphasize the extraction and application of multi-scale deep features.
(2) The multi-scale feature fusion enhanced (MFFE) module explores the multi-scale fusion of different local spatial features.
(3) Fusion coordinate attention and residual network (FCA-ResNet) wheat leaf disease image classification model that replaces multi-layer convolution with a multi-branch structure is developed to achieve a lightweight architecture.

The remainder of this paper is structured as follows: Section 2 provides an overview of related work, outlining the current research landscape in deep learning for identifying and classifying wheat leaf diseases. Section 3 introduces the FCA-ResNet model, outlining its structure and characteristics. Section 4 details the experimental process, covering the experimental validation of the proposed method and a comparative analysis of the results. Section 5 presents the conclusion and outlines the future work of this paper.

2. Related Work

In intelligent agriculture, extensive research has been conducted, and numerous public datasets have been established. Notably, publicly available wheat disease datasets such as LWDCD2020 [16], WFD2020 [17], Yellow-Rust-19 [18], and CGIAR [7] play a crucial role in confirming the model's performance, as depicted in Table 1. Field data is instrumental in enhancing the model's practical utility. Researchers employ both self-built and large datasets to advance practical classification tasks further, aiming to optimize the model's application in real-world scenarios. It is worth noting that 70% of the studies utilize custom datasets.

To supplement the public PlantVillage dataset of single-leaf images under controlled conditions, field-based wheat disease images (FWDI) [19] have been developed based on images collected in the field. It has 561 images of wheat powdery mildew, 808 of leaf rust, 1,015 of stripe rust, and 259 of healthy wheat leaves. These images come from different capture conditions and disease development stages and have complex environmental backgrounds. A multi-scene wheat disease dataset MSWDD2022 [20] has images of wheat stripe rust, wheat powdery mildew, wheat yellow dwarf, and wheat scab. Nigam et al. [21] used CNN to detect wheat yellow rust, where a dataset with 2,000 images was collected using a digital camera. The proposed deep learning model achieves the highest test accuracy of 97.3%. However, it only includes two types: healthy wheat leaves and yellow rust.

Classifying crop diseases poses significant challenges due to large variations within the same class and slight differences between different classes. For instance, early leaf lesions are not familiar, the characteristics of early-stage diseases are not clearly defined, and the difference between some diseases' early and late stages is minimal. Classification models often use MFFE and attention mechanisms to address these challenges and improve accuracy. MFFE combines features from different depths to reduce detail loss caused by downsampling. It also prevents disease features from being lost during image extraction. On the other hand, the attention mechanism identifies important information and suppresses irrelevant details. These techniques have gained increased attention for their potential to optimize the role of the attention mechanism.

Table 1 Related works on wheat leaf disease classification

Ref.	Category	Details
Haider et al. [8]	Methodology	RBNet, residual block, and self-attention mechanisms
	Dataset	Two public datasets on Kaggle [22-23]
	The number of images	4,086
	Accuracy	98.60% (maximum)
	Disease	5 types: brown rust, healthy, septoria, stripe rust, and yellow rust
Xu et al. [24]	Methodology	RFE-CNN [25], residual channel attention block (RCAB), feedback block (FB), elliptic metric learning (EML), and CNN
	Dataset	Custom dataset, CGIAR dataset, LWDCD2020 [16], and plant pathology
	The number of images	7,239 and 12,000
	Accuracy	Custom dataset: 99.95% (maximum); Others: 99.50% (average)
	Disease	5 types: healthy and four kinds of diseases: powdery mildew, septoria, blight, and rust
Mao et al. [20]	Methodology	DAE-Mask, diversification-augmented features, and edge features
	Dataset	MSWDD2022, PlantDoc
	The number of images	111,180
	Accuracy	MSWDD2022: 96.02% (mAP); PlantDoc:57.68% (mAP)
	Disease	4 types: wheat stripe rust, wheat powdery mildew, wheat yellow dwarf, and wheat scab
This study	Methodology	Custom CNN model
	Dataset	LWDCD2020 [16]
	The number of images	12,000
	Accuracy	97.88% (maximum)
	Disease	10 types: Karnal bunt, black chaff, crown and root rot, fusarium head blight, healthy wheat, leaf rust, powdery mildew, tan spot, wheat loose smut, and wheat streak mosaic
Genaev et al. [17]	Methodology	EfficientNet-B0
	Dataset	WFD2020
	The number of images	2,414
	Accuracy	94.2%
	Disease	5 types: leaf rust, stem rust, yellow rust, powdery mildew, and septoria

Xu et al. [25] propose an integrated deep learning algorithm that combines the residual channel attention module, feedback module, EML, and CNN. The algorithm achieves a maximum accuracy of 99.95% on the self-built dataset and an average accuracy of 99.50% on open-source databases, including CGIAR, Plant Diseases, LWDCD2020, and Plant Pathology. Mao et al. [20] propose the DAE-Mask method for intelligent detection of wheat diseases based on diversified enhancement features and edge features. After extracting the primary features of DenseNet, a backbone feature extraction network combining a Feature Pyramid Network (FPN) and attention mechanism is used to extract diversified enhanced features. On the self-built dataset MSWDD2022 and the public dataset PlantDoc, the mean average precision (mAP) is 96.02% and 57.68%, respectively. A multi-scale cross-level attention learning network is proposed for hyperspectral image classification to fully exploit pixels' global and local multi-scale features.

Haider et al. [8] propose three new residual blocks and self-attention mechanisms, evaluating the proposed method on publicly available crop disease images, including cotton, wheat, and Eurostat datasets. The technique achieves maximum accuracies of 98.60%, 93.90%, and 83.1%, respectively. Feng et al. [26] improved the MobileNetV2 model by adding Coord Attention to the bottleneck and proposed a new lightweight wheat seed identification method (IRB-5-CA Net). The self-built dataset's average accuracy, average recall, and F1 score are 99.5%, 99.6%, and 99.6%, respectively. Goyal et al. [16] propose a ResNet variant for classifying ten wheat disease categories. The method achieved a test accuracy of 97.88%, which is 15.92% higher than that of ResNet-50.

To better deploy neural network models on mobile devices, researchers have carried out research from various aspects and obtained remarkable results. Zhu et al. [27] use asymmetric and dilated (AD) convolution instead of standard convolution to construct a lightweight model using asymmetric and dilated convolution (LAD-Net) inception, reducing model parameters and enhancing the ability to extract multi-scale features of disease spots of varied sizes. The LAD-Net model is constructed by low-rank convolutional block attention module (CBAM) and LAD-Inception modules, and the fully connected layer is replaced by global average pooling to reduce the parameters further. Experimental results show that LAD-Net can achieve 98.58% recognition performance with only 1.25 MB size. Yu et al. [28] propose a lightweight ResNet (LW-ResNet) model for apple leaf disease recognition based on ResNet-18. They construct a multi-scale feature extraction layer and realize model compression. The model parameter memory is 2.32 MB, 94% less than ResNet-18.

Bao et al. [29] designed a SimpleNet, a lightweight CNN model that automatically identifies wheat ear diseases on mobile terminals. The model has only 2.13 M parameters and achieves 94.1% recognition accuracy on the test data set. Genaev et al. [17] use the EfficientNet-B0 neural network model to identify wheat leaf rust, stem rust, yellow rust, powdery mildew, and septoria. From the CGIAR dataset with a total of 2,414 RGB images taken, the EfficientNet-B0 neural network model has an average test accuracy of 94.2% for wheat disease classification. Xiao et al. [30] propose a lightweight model based on an improved residual network and attention mechanism, which combines Deep variant residual network (VRNet) and squeeze-and-excitation (SE) module with attention mechanism and obtained high accuracy on NewData and SelfData. It is verified that the new module effectively recognizes leaf diseases on mobile devices. Meanwhile, the size of the model is 119.8 MB.

The analysis of existing research shows that in the study of various computer vision models, high accuracy and lightweight are still the two most important indicators pursued by deep learning in practical applications. The balance between the two has always been essential to research content. While traditional deep convolutional networks typically aim to enrich the feature space by increasing network depth and the number of filters, this often leads to a significant increase in the scale of network parameters. In the classification of wheat leaf diseases from images, a common challenge arises from variations in the size of disease spots for the same type of disease and the high similarity of disease spots across different kinds of diseases.

3. Proposed Methodology

Recognition accuracy must be enhanced by extracting features from multiple scales to address the above questions. In this context, a lightweight multi-feature fusion network called FCA-ResNet is proposed to capture richer multi-scale features while mitigating parameter increases. The overall architecture of FCA-ResNet is shown in Fig. 1. It consists of two core components: INCA and MFFE. First, the MFFE is proposed to extract global and local information from leaf disease spots while considering multi-scale information. The MFFE module fuses shallow and deep features through CNN and attention mechanisms. Secondly, INCA is proposed to obtain more contextual semantic details and consider the feature extraction accuracy of different lesion sizes. Attention modules such as CA are embedded into the multi-branch convolution block to focus on multi-scale information and lesion features at the same time. In addition, the overall network structure is shortened to reduce the depth of the network. Finally, the extracted deep features are linearized into one-dimensional vectors to complete the classification.

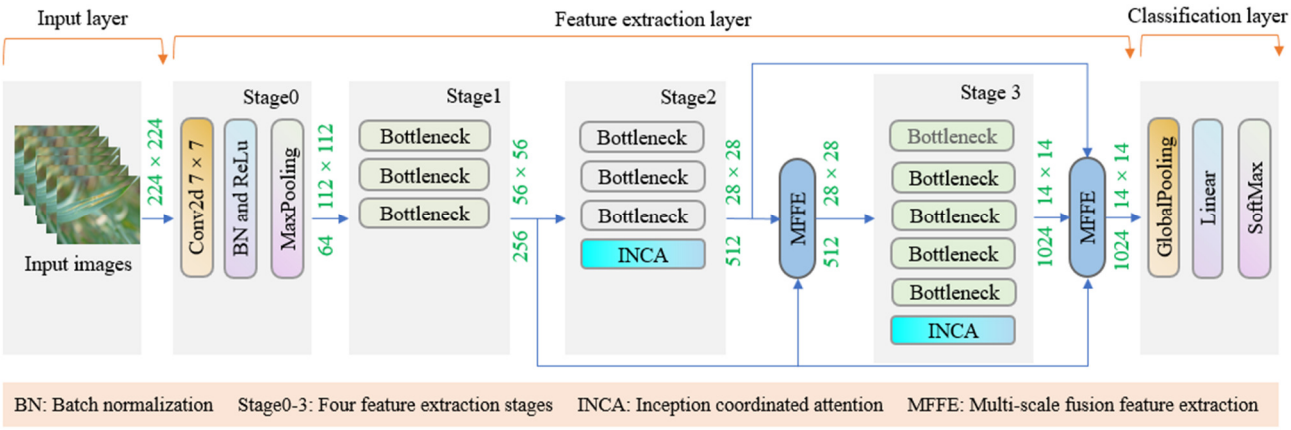


Fig. 1 Flowchart of FCA-ResNet

(1) Inception coordinate attention (INCA) module

The widespread distribution of wheat disease images requires a multi-scale feature extraction method to capture long-range dependencies. Combining the Inception module with the CA mechanism can enhance the model's ability to recognize disease features, improving classification accuracy. The Inception structure involves simultaneous convolution operations at multiple scales, enabling the extraction of diverse features at different scales. CA, a lightweight channel attention mechanism, compensates for the spatial feature limitations of SE attention and the long-range dependency shortcomings of CBAM. The overall structure of the INCA module is shown in Fig. 2. Although the INCA module has the same number of module parameters as other attention mechanisms, due to its multi-scale and CA combination, it can achieve higher accuracy when the model is shortened.

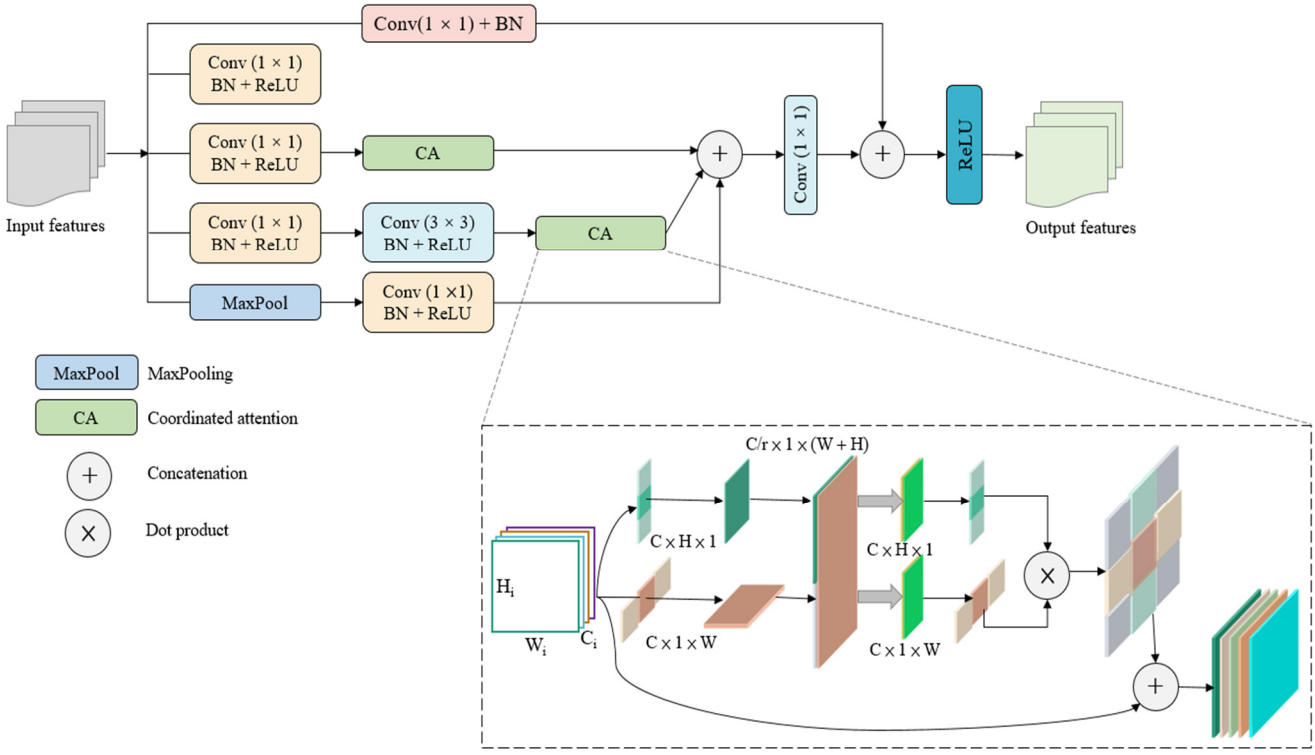


Fig. 2 Structure diagram of INCA module

(2) Multi-scale feature fusion enhanced (MFFE) module

The MFFE module combines shallow feature mAPs using the spatial attention mechanism and subsampling and deep feature mAPs through the attention mechanism. The module fuses features from different layers using various fusion strategies. An example of the MFFE module structure involves fusing three layers, as depicted in Fig. 3.

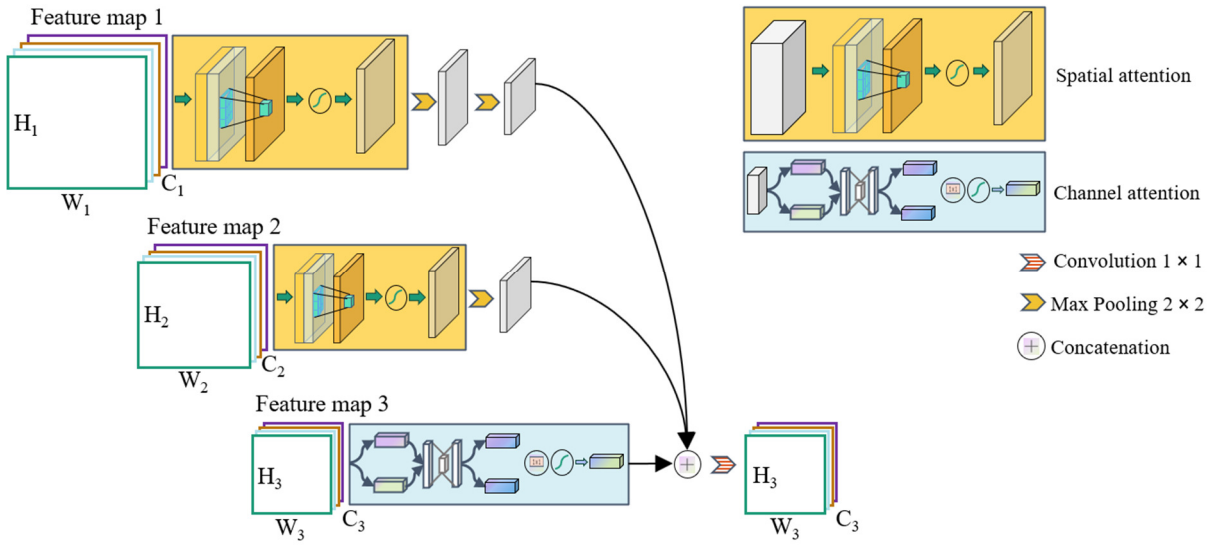


Fig. 3 Structure diagram of the MFFE module

4. Experiment Results and Analysis

The following section outlines the experimental results. It begins with providing information about the dataset, unique network configurations, and related evaluation metrics. It then delves into conducting ablation experiments and comparative studies to analyze the performance of the proposed FCA-ResNet classification algorithm.

4.1. Construction of dataset and experimental setup

(1) Construction of the dataset

The wheat leaf disease data in this paper were collected from five public sources: LWDCD2020 [16], wheat leaf dataset [29], yellow rust [18], dataset-of-DAE-Mask [20], as well as some local data from Shandong wheat fields. The selected leaves include leaf blotch (290), powdery mildew (264), leaf rust (310), stripe rust (346), yellow dwarf (323), and healthy wheat leaves (323), totaling 1,856 image data. The image format used is JPEG. Examples of the samples are shown in Fig. 4. All 1,856 samples are randomly divided into a training set, validation set, and test set according to the ratio of 7:2:1. The specific division is shown in Table 2.

Fig. 4 Six samples of wheat leaf disease dataset [14, 16, 18]

Table 2 Statistics of the dataset

Index	Disease classes	Training set	Validation set	Testing set
0	Healthy wheat leaves	226	64	33
1	Leaf blotch	133	37	20
2	Leaf rust	217	62	31
3	Powdery mildew	184	53	27
4	Stripe rust	242	69	35
5	Yellow dwarf	226	64	33
Total		1,128	349	179

(2) Hardware and software

The model proposed in this project is implemented using the PyTorch 2.1.0 framework. The hardware platform consists of an Intel® Xeon® Silver 4214R CPU @ 2.40 GHz processor and an NVIDIA GeForce RTX 3080 Ti (12 GB) GPU workstation with CUDA 12.1. The programming language used is Python 3.10 (Ubuntu 22.04), as it provides built-in libraries and functions for the full exploitation of deep learning models with complete support for GPU computing. Before implementing the deep learning model, libraries such as NumPy, OpenCV, Multiprocessing, scikit-image, SciPy, os, Matplotlib, and math were installed.

(3) Performance metrics

This task involves multi-classification and primarily assesses the model's classification performance for various diseases. The evaluation criteria include accuracy, recognition accuracy, specificity, recall, F1 score, and precision-recall curve. The definition of these measures is as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100\% \tag{1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\% \tag{2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\% \tag{3}$$

$$\text{F1 score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\% \tag{4}$$

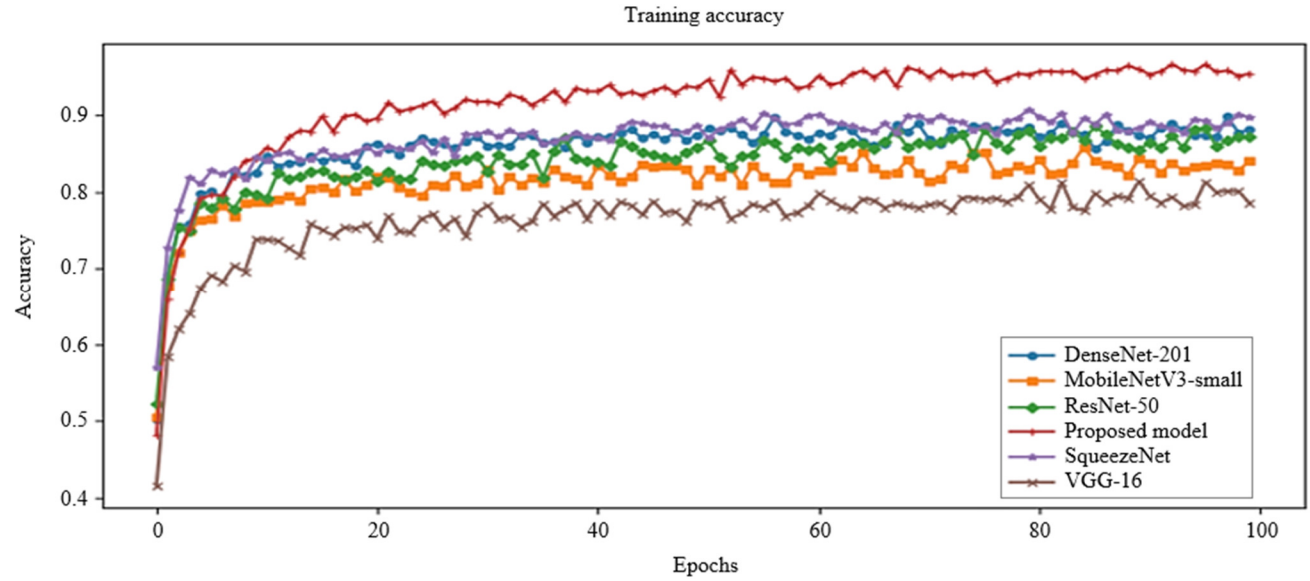
$$\text{mAP} = \frac{\sum_{j}^{N_j} \text{AP}(j)}{N_{class}} = \frac{\sum_{j=1}^{N(class)} \text{Precision}(j) \times \text{Recall}(j) \times 100\%}{N(calss)} \tag{5}$$

where true positive (TP) indicates the number of samples predicted to be positive, false negative (FN) represents the number of samples that are positive but predicted to be negative, false positive (FP) (FP) represents the number of samples that are negative but predicted to be positive, and true negative (TN) indicates the number of negative samples predicted. Additionally, the evaluation considers the computational complexity and parameters, given the focus on the model's performance during future deployment. Moreover, the mAP metric, which is the average of average precision over all categories, is also considered for evaluating the performance of multiple classifiers. AP(j) represents the area under the precision-recall curve (average precision) for the j-th class.

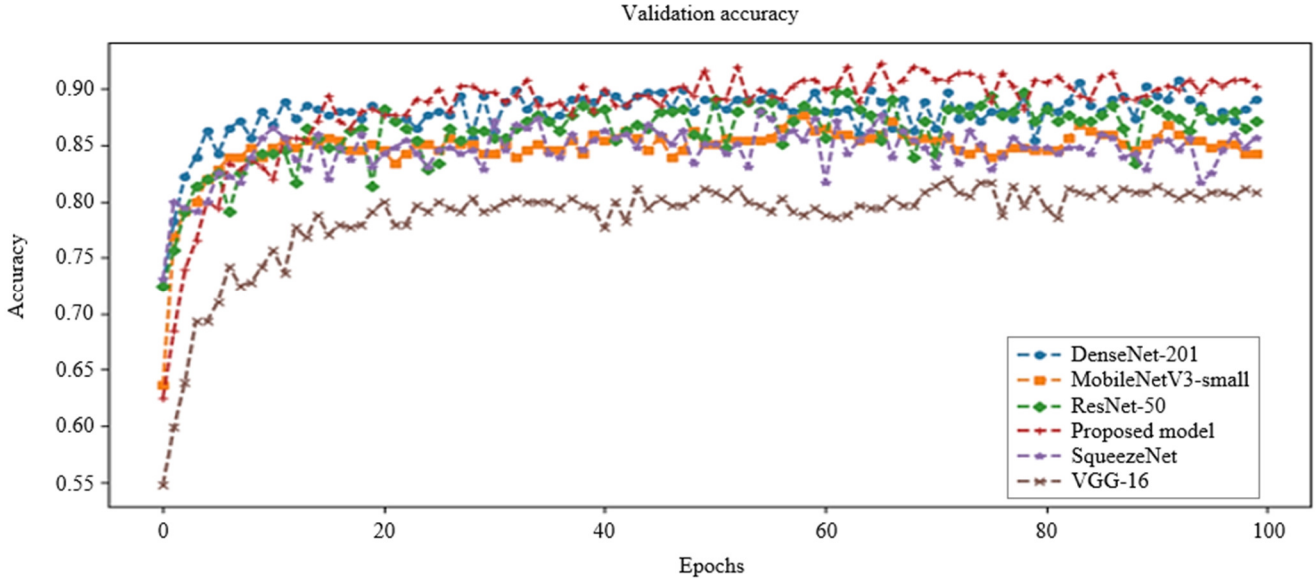
4.2. Comparison with other methods

To effectively demonstrate the FCA-ResNet model's ability to capture feature information from wheat leaf disease images, pre-trained models of five classical models were used for comparison. The pre-trained model is frozen and fine-tuned to classify six different disease states. To ensure fairness, all images were standardized to a size of 224 × 224. The models shared the same operating environment and hyperparameters. K-fold cross-validation is commonly used to select the most appropriate hyperparameter but increases computational complexity. In this paper, 5-fold cross-validation is chosen to avoid excessive computational complexity.

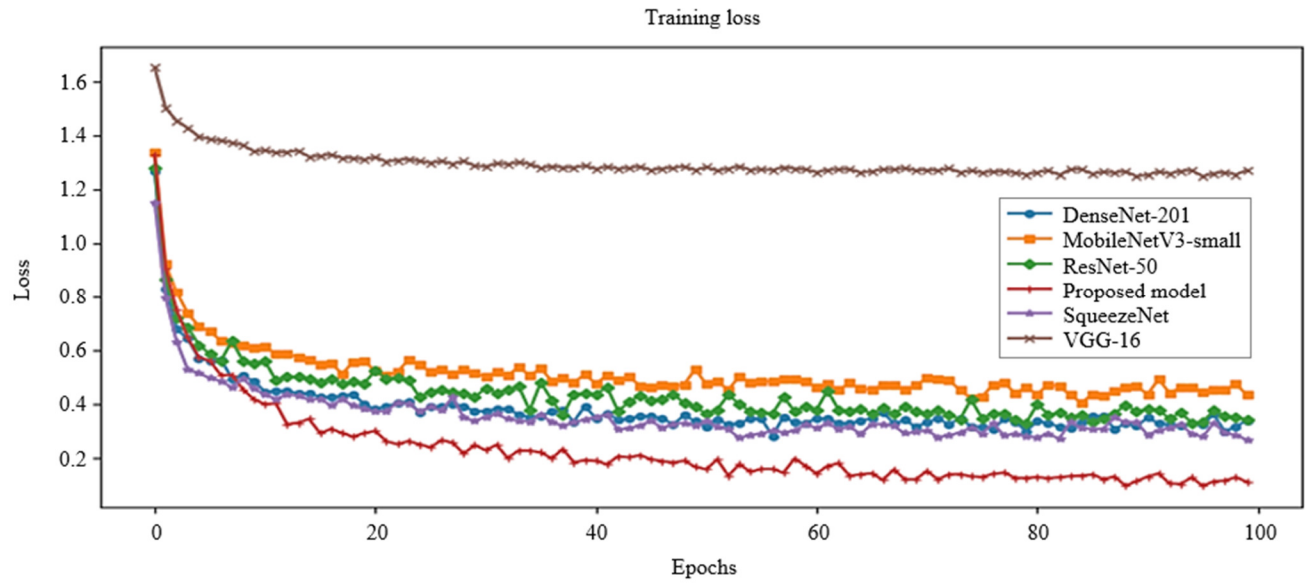
The training data is partitioned into five equal parts, with one part for testing and the other four for training in each run. The best hyperparameter, determined by the highest test accuracy on the verification set, is selected for the model. The batch size is 16, the epoch is 100, the initial learning rate is 0.002, and the weight decay is 0.001, with SoftMax as the output layer and classification cross-entropy as the chosen loss function. To present the performance of each model more visually, the accuracy and loss of the different models on the training and validation sets are shown in Fig. 5.



(a) Training accuracy of different models

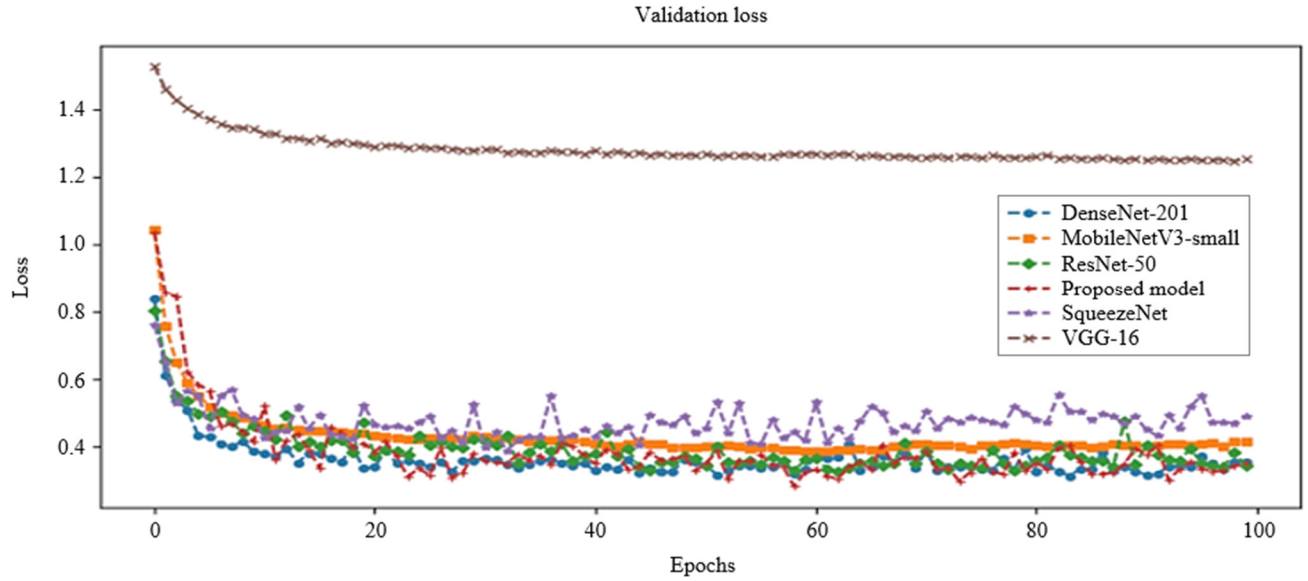


(b) Validation accuracy of different models



(c) Training loss of different models

Fig. 5 Comparison of accuracy and loss for training and validation sets



(d) Validation loss of different models

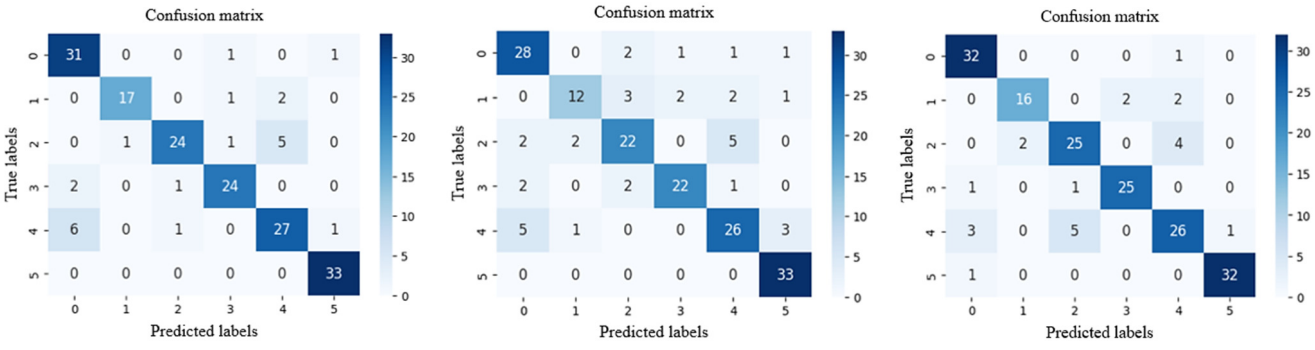
Fig. 5 Comparison of accuracy and loss for training and validation sets (continued)

The accuracy, recall, precision, specificity, F1 score, floating-point operations (FLOPs), and total training parameters for each model were documented on the test set. Table 3 presents the test results for various deep-learning models. The average accuracy achieved is 91.6%, approximately 4.5% higher than the highest accuracy attained by the classical model. In terms of FLOPs and total training parameters, despite having more parameters and requiring more computations than lightweight models such as MobileNet and SqueezeNet, the overall computational load is reduced by 8.1%, and the parameter count is decreased by 57.9% as compared to the baseline model ResNet-50.

Table 3 Performance comparison of different models on the test set

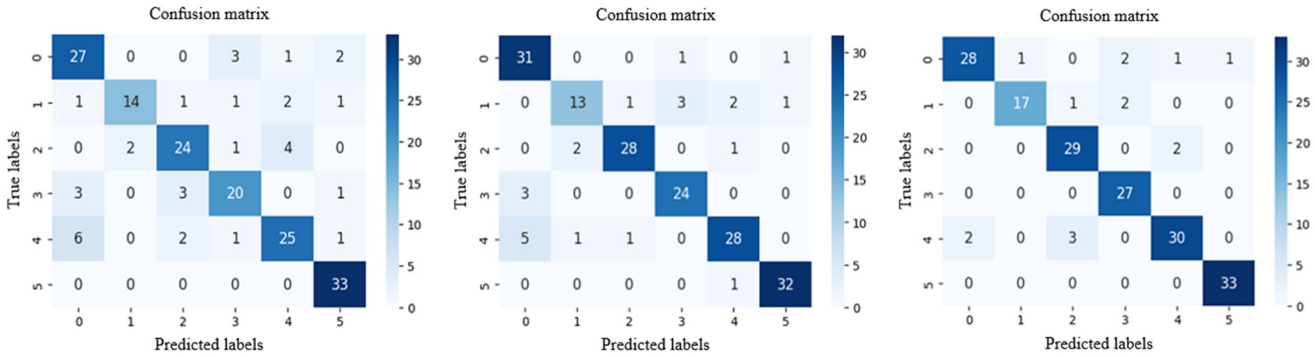
Model	Accuracy	Recall	Precision	Specificity	F1 score	FLOPs	Parameters
ResNet-50	87.1%	86.9%	87.4%	97.4%	87.1%	4,133.74 M	25.56 M
MobileNetV3-small	79.9%	78.6%	80.1%	95.9%	79.0%	62.48 M	2.54 M
DenseNet-201	87.1%	87.1%	88.1%	97.4%	87.3%	4,390.38 M	20.01 M
VGG-16	80.0%	79.1%	80.4%	95.9%	79.4%	15,470.31 M	138.36 M
SqueezeNet	87.1%	85.9%	86.9%	97.4%	86.1%	349.32 M	1.24 M
Proposed model	91.6%	91.5%	91.8%	98.3%	91.5%	3,798.14 M	10.76 M

The confusion matrix compares the accurate and inaccurate classification results. Each row corresponds to the actual class, each column corresponds to the predicted class, and the main diagonal represents the number of correct classifications. Fig. 6 displays the confusion matrices of the six different models on the test set. Although FCA-ResNet is not the best in class 0 (healthy wheat leaves), it performs better than other models on the remaining five diseases.



(a) DesNet-201 (b) MobilenetV3-small (c) ResNet-50

Fig. 6 Confusion matrices of different models on the test set



(d) VGG-16 (e) SqueezeNet (f) FCA-ResNet

Fig. 6 Confusion matrices of different models on the test set (continued)

4.3. Ablation experiment

(1) Performance analysis of different modules

Ablation experiments were used to test each network block and evaluate the effectiveness of each module. As shown in Table 4, the impact of each part on the model is assessed on the data set built in this article. Here, ResNet-50 is the baseline network, and the numbers in parentheses indicate the number of blocks in each stage, respectively, starting from the baseline model and going up to the proposed model. The inception block (IN), CA, INCA, and MFFE were gradually added to construct different network configurations for ablation analysis. Short_ResNet-50 is a shortened ResNet-50 for cutting the stage 4 feature extraction.

Table 4 Performance analysis of different modules

Name	Models	Accuracy	Recall	F1 score	FLOPs	Parameters	Time (ms)	FPS
B	ResNet-50	87.1%	86.9%	87.1%	4,133.74 M	25.56 M	7.32	136.61
B1	ResNet-50 + IN	87.1%	87.5%	87.1%	6,003.58 M	42.25 M	11.4	87.72
B2	ResNet-50 + CA	81.0%	80.9%	79.9%	3,482.3 M	20.22 M	9.71	102.99
B3	ResNet-50 + INCA	91.6%	90.8%	91.0%	4,540.0 M	29.28 M	14.06	71.12
B4	ResNet-50 + MFFE	89.4%	88.9%	89.2%	3,680.90 M	19.99 M	8.21	121.8
B5	ResNet-50 + CA + MFFE	88.9%	88.1%	88.4%	3,688.8 M	20.52 M	10.12	98.81
B6	ResNet-50 + INCA + MFFE	87.7%	87.3%	87.9%	4,746.47 M	29.58 M	12.92	77.4
B7	FCA-ResNet	91.6%	91.5%	91.5%	3,798.14 M	10.76 M	10.56	94.7

The performance of models B1 and B indicates that the IN module does not improve performance. However, the accuracy of model B3 with the INCA module enhanced by 10.6% compared to B2. This demonstrates that the INCA module has strong feature extraction capabilities for classifying wheat leaf diseases. Therefore, the effectiveness of multi-scale feature extraction, which enhances the coordination attention mechanism, has been confirmed. The number of parameters increased by approximately 44.1% in the transition from B5 to B6, while the accuracy decreased by 1.2%. The only difference is that B5 uses CA while B6 uses INCA. However, when model B utilizes the superposition of bottleneck blocks, small lesions may not receive enough attention compared to more extensive lesions and are easily ignored due to the excessive enhancement of the receptive field.

Models B4 and B5 are based on the baseline models B and B2, with the addition of the MFFE. The MFFE combines two shallow features, and the shallow feature mAP has a small receptive field, making it more suitable for extracting features of small leaf disease spots. The accuracy of B4 and B5 is 89.4% and 88.9%, respectively, 2.3% and 7.9% higher than that of the baseline models B and B2, with minimal changes in the model's calculations and parameters. In the comparison of B3 and B6, it's evident that the complexity of the model degrades its performance. To further demonstrate the superior performance of the MFFE module in improving accuracy and recall rate, model B7 is proposed. B7 achieved 91.6% accuracy, 91.5% recall, 91.5%

F1 score, and had 10.76 M parameters. The number of parameters in B7 is 63.25% lower than that of B3, 46.17% lower than that of B4, and 63.62% lower than that of B6. However, the accuracy of B7 is not reduced compared to B3; it is increased by 2.2% and 3.9% compared with B4 and B6, respectively.

The experimental results presented above indicate that the proposed model offers significant advantages in terms of accuracy and efficiency. However, a notable drawback is its extended inference time. The average inference time for all samples in the test set is used as the measure. Specifically, B7 has an average inference time of 10.56 ms per picture, equating to 94.7 frames per second (fps). This represents a 3.24 ms per picture difference compared to the baseline model.

(2) Performance comparison of INCA module

Comparison and ablation experiments were conducted to assess the effectiveness of the INCA module and determine the optimal number of permutation bottlenecks. Three attention mechanisms (CBAM, CA, and multi-scale convolutional attention (MSCA)) are utilized to construct three fused attention modules: INCA, INMSCA, and INCBAM. These modules are embedded in ResNet-50, and the experimental comparison results are presented in Table 5.

Table 5 Performance comparison of different attention mechanism fusion modules

Base model	*m*	Module	Test accuracy	mAP	Parameters
ResNet-50	1	INCA	91.6%	83.1%	29.28 M
		INMSCA	88.3%	77.5%	29.31 M
		INCBAM	85.5%	72.5%	29.31 M
ResNet-50	2	INCA	87.7%	76.5%	23.42 M
		INMSCA	87.2%	76.1%	23.45 M
		INCBAM	83.8%	67.0%	23.45 M
Short_ResNet-50	1	INCA	92.2%	83.9%	10.46 M
		INMSCA	87.7%	76.5%	10.46 M
		INCBAM	87.7%	76.3%	10.46 M
Short_ResNet-50	2	INCA	79.9%	63.5%	9.06 M
		INMSCA	87.7%	75.9%	9.07 M
		INCBAM	89.3%	78.9%	9.07 M

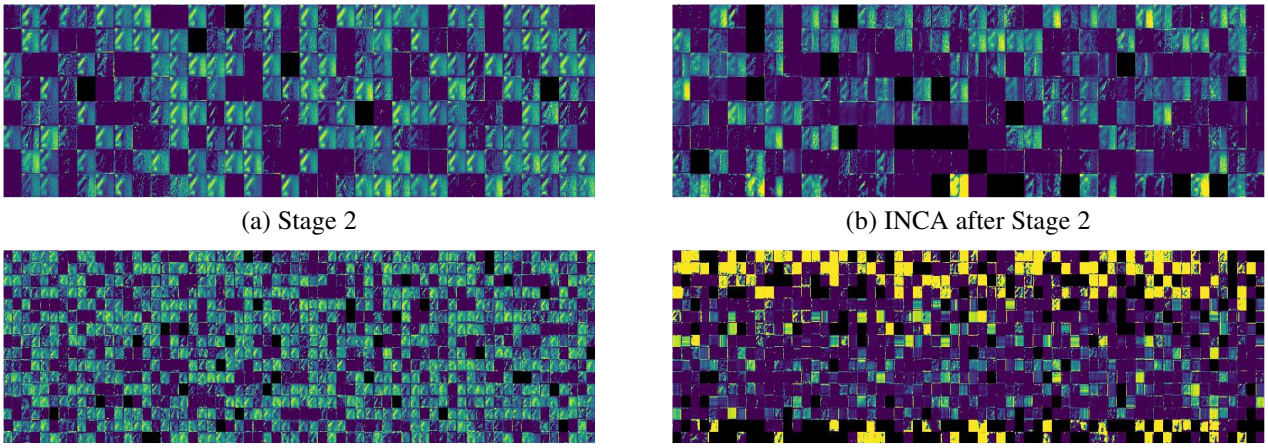
The comparison is based on three different attention modules. Among the three modules, INCA performed the best. When using INCA with a single permutation block (m = 1), the test accuracy reaches 91.6%, and the mAP reaches 83.1%. This is 3.3% and 6.1% higher than the accuracy achieved by INMSCA and INCBAM, respectively. The mAP is 5.6% and 10.6% higher than that achieved by INMSCA and INCBAM, respectively. With two permutation blocks (m = 2), the test accuracy of the INCA model is 87.7%, and the mAP is 76.5%. This is 0.5% and 3.9% higher than the accuracy achieved by INMSCA and INCBAM, respectively. The mAP is 0.4% and 9.5% higher than that achieved by INMSCA and INCBAM, respectively. When the depth of the model is reduced, the accuracy and mAP of both cases are improved correspondingly, with accuracy up to 92.2% and mAP up to 83.9%.

As m increases, the accuracy and mAP decrease in all three cases. However, the model with the INCA module still performs the best. The INCA module can extract more critical information due to increased CA in one branch. As the model is shortened, the number of parameters is reduced while the accuracy and mAP are kept or improved. The fused attention mechanism module's feature extraction ability is comparable to the performance of the superposition of multiple bottleneck blocks. In the experiment, the INCA module replaced the last bottleneck block in Stages 2, 3, and 4. There are 11 different embedding schemes, as detailed in Table 6. Among them, Schemes 3 and 9 stood out. When INCA replaced the last bottleneck block of Stage 3 or when Stages 2, 3, and 4 were replaced by INCA simultaneously, the model achieved 91.6% accuracy on the test set. However, when the model is modified such that only Stage 3 utilized INCA, Scheme 10's performance decreased by 24% compared to Scheme 3 in terms of accuracy and precision. On the other hand, replacing Stage 2 and Stage 3 modules improved accuracy to 89.4%. So, the proposed model utilized Schemes 9 and 11.

Table 6 Comparison of different schemes for the INCA module

Scheme	Stage 1	Stage 2	Stage 3	Stage 4	Parameters	Accuracy	mAP
1	√	--	--	--	25.60 M	88.8%	78.4%
2	--	√	--	--	25.73 M	88.8%	78.3%
3	--	--	√	--	26.27 M	91.6%	83.9%
4	--	--	--	√	28.39 M	91.1%	83.1%
5	√	√	--	--	25.79 M	84.9%	71.8%
6	--	√	√	--	26.48 M	89.4%	79.7%
7	--	--	√	√	29.10 M	89.9%	80.8%
8	√	√	√		22.03 M	82.1%	66.5%
9	--	√	√	√	29.28 M	91.6%	83.1%
10	--	--	√	--	10.27 M	67.6%	45.2%
11	--	√	√	--	10.46 M	92.2%	83.9%

Fig. 7 shows the output characteristics with and without the INCA module added at the Stage 2 and Stage 3 locations. Comparing the two groups reveals that the INCA module's MFFE and attention mechanism improve the extraction ability of lesion features.



(a) Stage 2 (b) INCA after Stage 2

(c) Stage 3 (d) INCA after Stage 3

Fig. 7 Comparison of feature mAPs before and after adding INCA

(3) Performance comparison of MFFE multi-scale feature extraction and fusion module

Comparative experiments using different fusion strategies were conducted to test the performance of the proposed MFFE module. The results of the two proposed feature extraction fusion modules can be found in Table 7. For example, FF12 denotes the feature fusion of Stage 1 and Stage 2, with other identifiers following a similar naming convention.S+C indicates spatial attention for shallow features and channel attention for deep features in feature fusion, while C+C means that channel attention is used for shallow and deep features. The experimental results show that the fusion of shallow traits is more effective in the feature extraction of leaf lesions. Notably, the fusion of FF12 and FF123 performs well under both fusion strategies. The S+C strategy has fewer parameters each way, and only the FF123 group had lower accuracy and mAP than the C+C strategy. Therefore, the S+C strategy is adopted in the proposed model.

Table 7 Performance comparison of different fusion strategies for MFFE

Model	S+C			C+C		
	Parameters	Test accuracy	mAP	Parameters	Test accuracy	mAP
ResNet+FF13+FF24	31.46 M	91.1%	82.4%	32.81 M	88.8%	78.5%
ResNet+FF12+FF34	30.58 M	88.3%	77.0%	32.94 M	88.3%	78.7%
ResNet+FF123	26.74 M	91.1%	83.4%	27.56 M	91.6%	83.4%
ResNet+FF234	30.28 M	89.9%	80.8%	33.58 M	90.5%	82.0%
ResNet+FF1234	30.28 M	85.5%	74.0%	34.12 M	84.9%	72.6%
ResNet+FF12	25.85 M	92.7%	85.6%	25.99 M	88.8%	78.4%
ResNet+FF13	26.73 M	90.5%	81.1%	27.01 M	87.7%	76.6%

4.4. Testing model generalization performance

The proposed model is tested on the DAE-Mask dataset and the wheat leaf disease dataset. Table 8 calculated the model's average time to make predictions on a single sample and five evaluation indicators on the test set. The average prediction time on the DAE-Mask dataset is 8.72 ms/pic; on the wheat leaf disease dataset, it is 8.46 ms/pic. The model demonstrated good generalization ability, with all five evaluation indicators scoring above 96%.

Table 8 Test effects of FCA-ResNet on other datasets

Dataset	Categories of diseases	Average inference time (ms/pic)	Test accuracy	Recall	Precision	Specificity	F1 score
DAE-Mask	0: Wheat yellow dwarf; 1: Wheat powdery mildew; 2: Wheat stripe rust	8.72	99.7%	99.7%	99.7%	99.8%	99.1%
Wheat leaf disease	0: Brown_rust; 1: Healthy; 2: Septoria; 3: Yellow_rust	8.46	97.6%	97.1%	96.0%	99.2%	96.0%

FCA-ResNet is introduced to meet the requirements of high precision and lightweight applications in natural environments. This model addresses the limitations of larger models like ResNet-50, which have numerous parameters and high complexity in natural settings. The proposed model enhances classification accuracy and generalization capability and quickly identifies diseased wheat and healthy wheat leaves against complex backgrounds. The average accuracy, recall rate, and F1 score are 91.6%, 91.5%, and 91.5%, where accuracy is 5.2% higher than ResNet-50. Moreover, FCA-ResNet uses only 40% of the parametric memory of the original ResNet-50 model, with 8% lower FLOPs. Although FCA-ResNet does not demonstrate recognition and training speed advantages, it can process 94.7 images per second. The robust and generalizable model suits small mobile longitudinal equipment with limited resources. It has the potential to positively impact the prevention and control of wheat leaf diseases in agricultural production.

5. Conclusion

This study proposed a wheat leaf disease classification model named FCA-ResNet, which outperforms models such as ResNet-50, InceptionV3, and DenseNet, providing a new approach to research on numerous wheat leaf diseases. Experiments showed that the proposed method achieves an accuracy rate of 91.6% on a six-class wheat leaf dataset. The main findings are summarized as:

(1) The proposed model adopts two multi-scale feature fusion techniques: an INCA multi-scale attention module for extracting local features of wheat leaf diseases and an MFFE fusion module to enhance recognition accuracy when the disease spot scales vary significantly within the same disease type.

(2) Reducing the layer structure of the backbone network ResNet-50 can reduce the model size while ensuring that accuracy does not decrease. The proposed model has an average inference time reduced to 10.56 ms per picture from 12.92 ms per picture, while the accuracy increases from 87.5% to 91.6%.

(3) Generalization testing of the DAE-Mask and the wheat leaf disease datasets showed average prediction times of 8.72 ms/pic and 8.46 ms/pic, respectively, and all five evaluation indicators scored above 96%.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Y. Liu, G. Gao, and Z. Zhang, "Crop Disease Recognition Based on Modified Light-Weight CNN with Attention Mechanism," IEEE Access, vol. 10, pp. 112066-112075, 2022.

[2] S. Gupta and A.K. Tripathi, "Fruit and Vegetable Disease Detection and Classification: Recent Trends, Challenges, and Future Opportunities," Engineering Applications of Artificial Intelligence, vol. 133, part C, article no. 108260, 2024.

[3] S. Huang, G. Zhou, M. He, A. Chen, W. Zhang, and Y. Hu, "Detection of Peach Disease Image Based on Asymptotic Non-Local Means and PCNN-IPELM," IEEE Access, vol. 8, pp. 136421-136433, 2020.

[4] E. Hirani, V. Magotra, J. Jain, and P. Bide, "Plant Disease Detection Using Deep Learning," 6th International Conference for Convergence in Technology, pp. 1-4, 2021.

[5] S. M. Hassan and A. K. Maji, "Plant Disease Identification Using a Novel Convolutional Neural Network," IEEE Access, vol. 10, pp. 5390-5401, 2022.

[6] P. Bedi and P. Gole, "Plant Disease Detection Using Hybrid Model Based on Convolutional Autoencoder and Convolutional Neural Network," Artificial Intelligence in Agriculture, vol. 5, pp. 90-101, 2021.

[7] V. Kukreja, R. Sharma, V. Sharma, and A. Verma, "Crop Vigil: Automated Wheat Bunt Disease Multi-Classification with a CNN-RNN Hybrid Model and Attention Block," 14th International Conference on Computing Communication and Networking Technologies, pp. 1-6, 2023.

[8] I. Haider, M. A. Khan, M. Nazir, A. Hamza, O. Alqahtani, M. T. H. Alouane, et al., "Crops Leaf Disease Recognition from Digital and RS Imaging Using Fusion of Multi Self-Attention RBNet Deep Architectures and Modified Dragonfly Optimization," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 17, pp. 7260-7277, 2024.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778, 2016.

[10] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700-4708, 2017.

[11] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, et al., "Going Deeper with Convolutions," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-9, 2015.

[12] R. Maurya, S. Mahapatra, and L. Rajput, "A Lightweight Meta-Ensemble Approach for Plant Disease Detection Suitable for IoT-Based Environments," IEEE Access, vol. 12, pp. 28096-28108, 2024.

[13] H. Sun, H. Xu, B. Liu, D. He, J. He, H. Zhang, et al., "MEAN-SSD: A Novel Real-Time Detector for Apple Leaf Diseases Using Improved Light-Weight Convolutional Neural Networks," Computers and Electronics in Agriculture, vol. 189, article no. 106379, 2021.

[14] P. Kumar, S. Raghavendran, K. Silambarasan, K. S. Kannan, and N. Krishnan, "Mobile Application Using DCDM and Cloud-Based Automatic Plant Disease Detection," Environmental Monitoring and Assessment, vol. 195, no. 1, article no. 44, 2023.

[15] M. Aziz Hosen Foysal, F. Ahmed, and M. Zahurul Haque, "Multi-Class Plant Leaf Disease Detection: A CNN-Based Approach with Mobile App Integration," International Journal of Computer Applications, vol. 186, no. 41, pp. 62-68, 2024.

[16] L. Goyal, C. M. Sharma, A. Singh, and P. K. Singh, "Leaf and Spike Wheat Disease Detection & Classification Using an Improved Deep Convolutional Architecture," Informatics in Medicine Unlocked, vol. 25, article no. 100642, 2021.

[17] M. A. Genaev, E. S. Skolotneva, E. I. Gultyaeva, E. A. Orlova, N. P. Bechtold, and D. A. Afonnikov, "Image-Based Wheat Fungi Diseases Identification by Deep Learning," Plants, vol. 10, no. 8, article no. 1500, 2021.

[18] T. Hayit, H. Erbay, F. Varçın, F. Hayit, and N. Akci, "Determination of the Severity Level of Yellow Rust Disease in Wheat by Using Convolutional Neural Networks," Journal of Plant Pathology, vol. 103, no. 3, pp. 923-934, 2021.

[19] J. Jiang, H. Liu, C. Zhao, H. He, J. Ma, T. Cheng, et al., "Evaluation of Diverse Convolutional Neural Networks and Training Strategies for Wheat Leaf Disease Identification with Field-Acquired Photographs," Remote Sensing, vol. 14, no. 14, article no. 3446, 2022.

[20] R. Mao, Y. Zhang, Z. Wang, X. Hao, T. Zhu, S. Gao, et al., "Dae-Mask: A Novel Deep-Learning-Based Automatic Detection Model for In-Field Wheat Diseases," Precision Agriculture, vol. 25, no. 2, pp. 785-810, 2024.

[21] S. Nigam, R. Jain, S. Marwaha, A. Arora, V. K. Singh, A. K. Singh, et al., "Automating Yellow Rust Disease Identification in Wheat Using Artificial Intelligence," Indian Journal of Agricultural Sciences, vol. 91, no. 9, pp. 1391-1395, 2021.

[22] O. Getch, "Wheat Leaf Dataset," https://www.kaggle.com/datasets/olyadgetch/wheat-leaf-dataset, 2021.

[23] S. Dunk, "Wheat Disease Detection," https://www.kaggle.com/datasets/sinadunk23/behzad-safari-jalal, 2020.
[24] L. Xu, B. Cao, F. Zhao, S. Ning, P. Xu, W. Zhang, et al., "Wheat Leaf Disease Identification Based on Deep Learning Algorithms," Physiological and Molecular Plant Pathology, vol. 123, article no. 101940, 2023.
[25] F. Xu, G. Zhang, C. Song, H. Wang, and S. Mei, "Multiscale and Cross-Level Attention Learning for Hyperspectral Image Classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 61, article no. 5501615, 2023.
[26] Y. Feng, C. Liu, J. Han, Q. Lu, and X. Xing, "IRB-5-CA Net: A Lightweight, Deep Learning-Based Approach to Wheat Seed Identification," IEEE Access, vol. 11, pp. 119553-119559, 2023.
[27] X. Zhu, J. Li, R. Jia, B. Liu, Z. Yao, and A. Yuan, "LAD-Net: A Novel Light Weight Model for Early Apple Leaf Pests and Diseases Classification," IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 20, no. 2, pp. 1156-1169, 2023.
[28] H. Yu, X. Cheng, Z. Li, Q. Cai, and C. Bi, "Disease Recognition of Apple Leaf Using Lightweight Multi-Scale Network with ECANet," Computer Modeling in Engineering & Sciences, vol. 132, no. 3, pp. 711-738, 2022.
[29] W. Bao, X. Yang, D. Liang, G. Hu, and X. Yang, "Lightweight Convolutional Neural Network Model for Field Wheat Ear Disease Identification," Computers and Electronics in Agriculture, vol. 189, article no. 106367, 2021.
[30] Z. Xiao, Y. Shi, G. Zhu, J. Xiong, and J. Wu, "Leaf Disease Detection Based on Lightweight Deep Residual Network and Attention Mechanism," IEEE Access, vol. 11, pp. 48248-48258, 2023.