Computer Fraud and Security
ISSN (online): 1873-7056

Residual Network with Multi-scale Information Fusion for Apple Fungal Infection Classification

Yufeng Long[1], Xiaofei Wang[1,*]

[1] School of Electronic Engineering, Heilongjiang University, Harbin, China
Corresponding Author's Email: nk_wxf@hlju.edu.cn

Abstract: Black rot, Cedar rust, and Scab are three common apple diseases caused by fungal infections. The infected areas not only impair the function of the current parts but also spread to healthy regions of the plant, necessitating early detection in production. Addressing the challenge of subtle lesion features, we have designed a deep learning model that integrates multi-scale information fusion by combining wavelet transform with a residual network, and introducing a multi-channel Collaborative Attention mechanism. In comparisons with similar models, our proposed model achieved the best results, with a classification accuracy of 98.1% for various lesions. It demonstrated excellent stability in data detection imaging under various conditions. This model provides a new perspective for apple lesion detection and may potentially be analogously applied to other types of detection in the future.

Keywords: Apple, Defect classification, Wavelet transform, computer vision, Multiscale convolution

1. Introduction

Apple is one of the significant economic crops globally, with a worldwide production of 95.83 million tonnes in 2024. China's apple production in 2024 reached 49.6017 million tonnes, accounting for 51.8% of the global total. The yield and quality of apples directly relate to agricultural economic stability and food security. However, the widespread occurrence of apple leaf diseases, particularly scab, black rot, and cedar rust, poses a severe threat to the apple industry. These diseases caused by fungal infections[1,2,3], which can lead to the degradation of leaf function and the impediment of photosynthesis. Worse still, damaged areas may serve as entry points for other pathogens to infect healthy parts[4], further spreading to fruits, causing fruit deformation, rot, and even the death of entire trees, which can result in annual yield losses of up to 20%-50%. In the field of crop disease control, detecting and classifying plant diseases is a crucial task. Traditional disease detection primarily relies on manual visual inspection or biochemical analysis in laboratories. However, manual methods are inefficient, have high misjudgment rates, and are difficult to apply across large-scale orchards. While laboratory testing offers higher accuracy, it is time-consuming and costly, making it unsuitable for real-time field monitoring. With climate change and the increase in pathogen resistance, the frequency of disease outbreaks and the speed of their spread have risen significantly, necessitating a disease detection technology that is efficient, precise, and scalable.

In research on similar fruits, deep learning is a highly utilized method. Ramachandran et al.[5] developed a deep learning model cascaded by two CNN models, achieving a classification accuracy of 91.61% for banana Wilt's Disease. Sijan Karki et al.[6] used ResNet-50 to detect strawberry angular leaf spot, anthracnose, gray mold, and powdery mildew, with an accuracy of 94.1%. Sharifah et al.[7] cascaded two CNN networks for image segmentation and classification to detect and classify disease severity on orange surfaces, achieving an accuracy of 94.37%. Existing research has demonstrated the effectiveness of deep learning in detecting fruit diseases. For single-class classification tasks of fungal infections in apples, Alexander et al.[8] combined two CNN models, MobileNetV2 and EfficientNetV2L, with multispectral imaging, achieving classification accuracies of 93.84% and 94.22%, respectively. Deepak[9] combined CNN with Random Forest to classify the black rot disease, achieving an average classification accuracy of 95%. Rishabh et al.[10] combined Dual CNN with LSTM to classify and predict the severity of Black rot lesions, achieving an overall accuracy of 99.2%. Dounia et al.[11] established multiple CNN models (efficientnet_v2, inception_resnet_v2, inceptionV3, mobilenet_v2, VGG-16), achieving a prediction accuracy of up to 97.3% for apple and grape varieties and Black rot infections

Single-class classification tasks generally yield better results because samples of a single disease often have significant feature differences from healthy samples, whereas classifying multiple diseases poses greater challenges. Diksha et al. used the VGG network to detect multiple diseases in apples, achieving an average accuracy of 85%. Zhong et al.[12] used a DenseNet-121 model with an improved focal loss function to classify six diseases in apples, achieving an accuracy of 93.71%. Devanshi et al. used various deep learning models to detect apple scab, cedar rust, and black rot, with the VGG network performing best, achieving a detection accuracy

of 92.5%. Prakhar et al.[13] integrated deep learning models including ResNet-121, achieving an average accuracy of 96.25% for detecting multiple apple diseases.

Based on the analysis of current research trends in this area, this study proposes a deep learning model that integrates multi-scale information fusion. This model combines wavelet transform convolution with a residual network and introduces a multi-channel collaborative attention mechanism. Its purpose is to classify and predict various apple leaf diseases caused by fungal infections. The classification results will be compared with those obtained using classic network models employed in similar studies

2. Materials and Methods

2.1 Materials

This research project focuses on accurately classifying and detecting three common fungal diseases affe cting apple leaves: Scab, Cedar Rust, and Black Rot. To establish a comprehensive and representative dataset, w e have included sample images of these three diseases, along with healthy leaf images as controls, totaling 1,400 images. Each category within the dataset encompasses diverse lighting conditions, complex backgrounds, and v arious shooting angles. This design not only enhances the diversity of the dataset but also ensures its high versati lity and wide applicability in practical scenarios, catering to disease identification needs under different lighting, environmental, and shooting conditions. Currently, these data samples are publicly accessible, which is availabl e at https://www.kaggle.com/datasets/sadmansakibmahi/plant-disease-expert.

As shown in Figure 1, during the input stage of training, we conducted a series of preprocessing operations on the images. For the color information, we made certain adjustments to the images by adding a 20% random fluctuation to their brightness, contrast, and saturation based on their original values before inputting them into the network for training. For the geometric information of the images, we applied flipping and rotation transformations with a 50% probability. The purpose of these operations is to artificially increase the diversity of the training data, helping the model to learn and generalize better, and thereby improving its performance on unseen data.

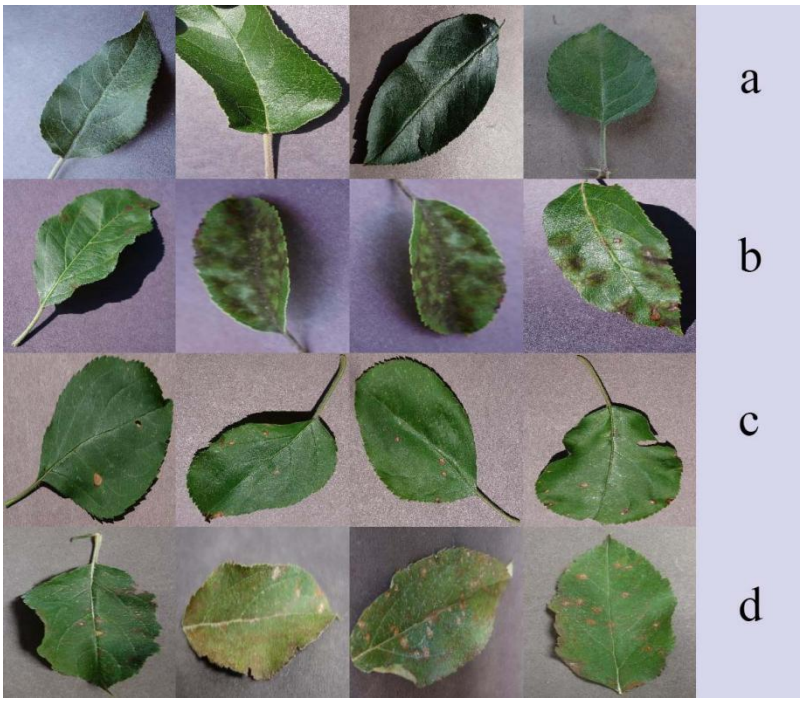

Figure 1. RGB images of samples: healthy(a), Scab(b), Black Rot(c), Cedar Rust(d)

2.2 Multiscale information fusion

For various diseases affecting apple leaves, we propose a model based on wavelet transform information fusion and channel-wise coordinated attention. This model integrates the aforementioned modules with residual

blocks from ResNet to ultimately achieve precise classification tasks. The proposed network architecture is illustrated in Figure 2.

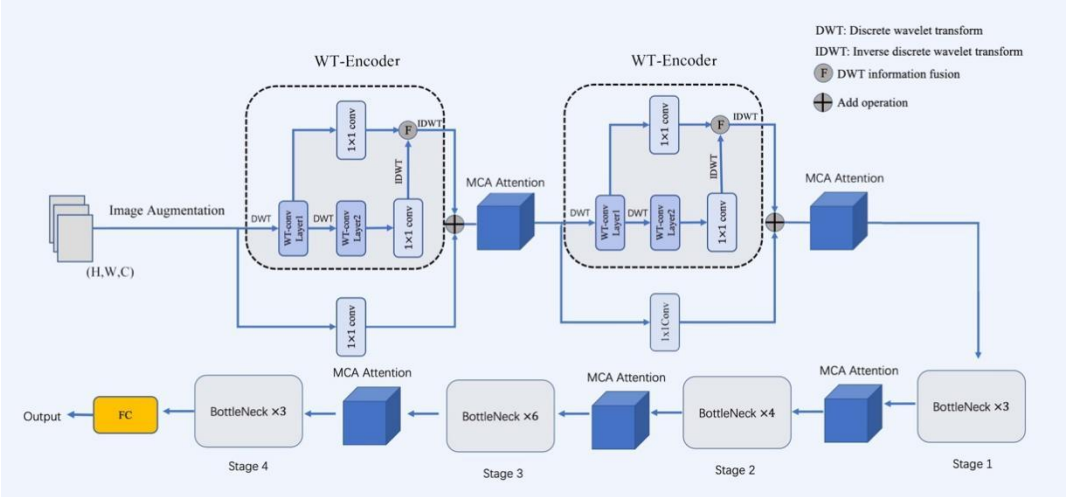

Figure 2. Structure of proposed model

Initially, image augmentation is applied to the input to enhance the model's generalization capability. Subsequently, two WT-Encoders and the main convolutional layers elevate the number of feature map channels to 16 and 64, respectively, during the feature fusion stage, yielding frequency-domain-enhanced feature maps. Following this, there are four stages, each comprising a varying number of BottleNecks. The structure of a BottleNeck is shown in Figure 2(b). Within each stage, changes in the number of convolutional channels occur in the first BottleNeck. For a feature map with an input channel count of C_1, the number of channels is first reduced to $C_1/4$ in the first 1×1 convolutional layer. After feature extraction in the 3×3 convolutional layer, another 1×1 convolutional layer is used to increase the number of feature map channels to C_2, where C_2 is typically an integer multiple of C_1. At the output of each WT-Encoder and Stage, we apply MCA Attention to integrate feature information from different dimensions of the feature maps. The final extracted features undergo global average pooling and are then input into a fully connected layer to output the final classification probabilities. Convolution parameters of modules can be seen in table 1.

Table 1. Convolution parameters of modules

Module	Input channels	Output channels	Kernel size
WT-Encoder 1	3	16	1*1
WT-Encoder 2	16	64	1*1
Stage 1	64	256	1*1，3*3，1*1
Stage 2	256	512	1*1，3*3，1*1
Stage 3	512	1024	1*1，3*3，1*1
Stage 4	1024	2048	1*1，3*3，1*1

2.3 Residential Blocks

ResNet (Residual Network)[14] is a deep neural network architecture whose core idea is the introduction of Residual Learning. Residual Learning mainly consists of two crucial components. The first is the residual block, a module composed of a set of convolutional layers and tensor operations aimed at increasing network depth while reducing convolutional computation costs. There are two classic forms of residual blocks: the basic residual block, whose internal structure is shown in Figure 3(a). The basic residual block contains two 3x3 convolutional layers, each followed by a ReLU activation function. Due to its simpler structure, the basic residual block is suitable for building shallow neural networks.

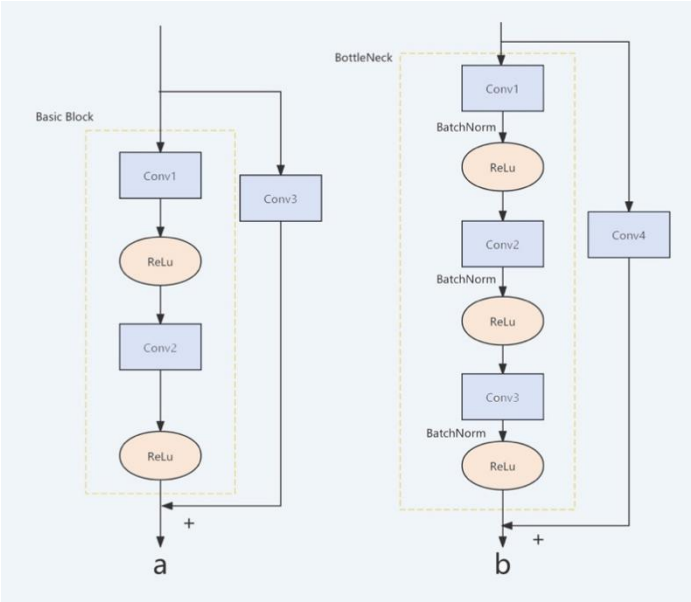

Figure 3. Structure of Basic Block (a) and BottleNeck Block (b)

The other type of residual block is the Bottleneck block, which has a more complex internal structure. The Bottleneck block contains two 1×1 convolutional layers and one 3×3 convolutional layer. The first 1×1 convolutional layer reduces the number of channels in the input feature map, aiming to decrease the computational load during feature extraction. This is followed by a 3×3 convolutional layer responsible for extracting the input's feature information. The last 1×1 convolutional layer increases the number of channels in the feature map. And within the Bottleneck, after each convolution, batch normalization and the ReLU activation function are applied to the convolution's output. Their purposes are to accelerate network training and introduce nonlinearity, respectively. The formula for batch normalization is shown in Equation 2, and the formula for the ReLU function is shown in Equation 3. Another important improvement in Residual Learning is skip connections, where the input of the module is directly added to its final output. This significantly mitigates the issue of vanishing gradients in deep learning. If there is a mismatch in the number of channels between the input and output of a residual block, a 1×1 convolutional layer is applied to the input x to match the number of channels.

$$\mu = \frac{1}{m}\sum_{i=1}^{m} x_i \, , \sigma = \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu)^2 \tag{1}$$

$$\hat{x} = \frac{x-\mu}{\sqrt{\sigma+\varepsilon}} \tag{2}$$

$$ReLu(x) = \begin{cases} x \; x > 0 \\ 0 \; x \leq 0 \end{cases} \tag{3}$$

2.4 Wavelet Multiscale Convolution

In deep learning, traditional Convolutional Neural Networks (CNNs) have achieved great success in image feature extraction through a series of convolutional, pooling, and nonlinear activation operations. However, traditional convolutions have some deficiencies in multi-scale feature extraction, especially in preserving high-frequency features and over-extracting low-frequency features. Additionally, due to the limitation of convolutional kernel sizes, CNNs often face the issue of having too small a receptive field. Therefore, Finder et al. proposed a method using wavelet transform to enhance the receptive field and details[15]. Building upon this, we have made improvements in the utilization of channel numbers and frequency domain details, introducing a multi-scale convolutional module based on wavelet transform, known as the WT-Encoder.

Figure 4 illustrates the internal structure of the two-layer wavelet transform encoder used in this study. It accepts input feature maps with an arbitrary number of channels and first performs a standard 1×1 convolution operation to increase the feature dimensions and integrate multi-channel information. Then, each channel undergoes two layers of wavelet transforms, generating a total of 8 corresponding frequency images. Layer 0 only contains the input feature map, while layers 1 and 2 respectively contain the component images obtained by decomposing the low-frequency components from the previous layer. The low-frequency components represent the smooth, overall structure and brightness information of the image, while the high-frequency components

capture edges, textures, and fine details, encoded in three directions: horizontal, vertical, and diagonal. Starting from the second layer, 1×1 convolution operations are applied to the frequency component images. After convolution, the transformed feature maps are inverse-transformed, and the result is added to the low-frequency components from the previous layer. The final inverse-transform output is combined with the 1×1 convolution result of Layer 0 and passed to the next layer. During this process, specific frequency components can be selected for convolution encoding based on the characteristics of the feature map, while redundant components can be discarded. Each layer's 1×1 convolution outputs the same number of channels to ensure the accurate fusion of information in the final Figure 4.

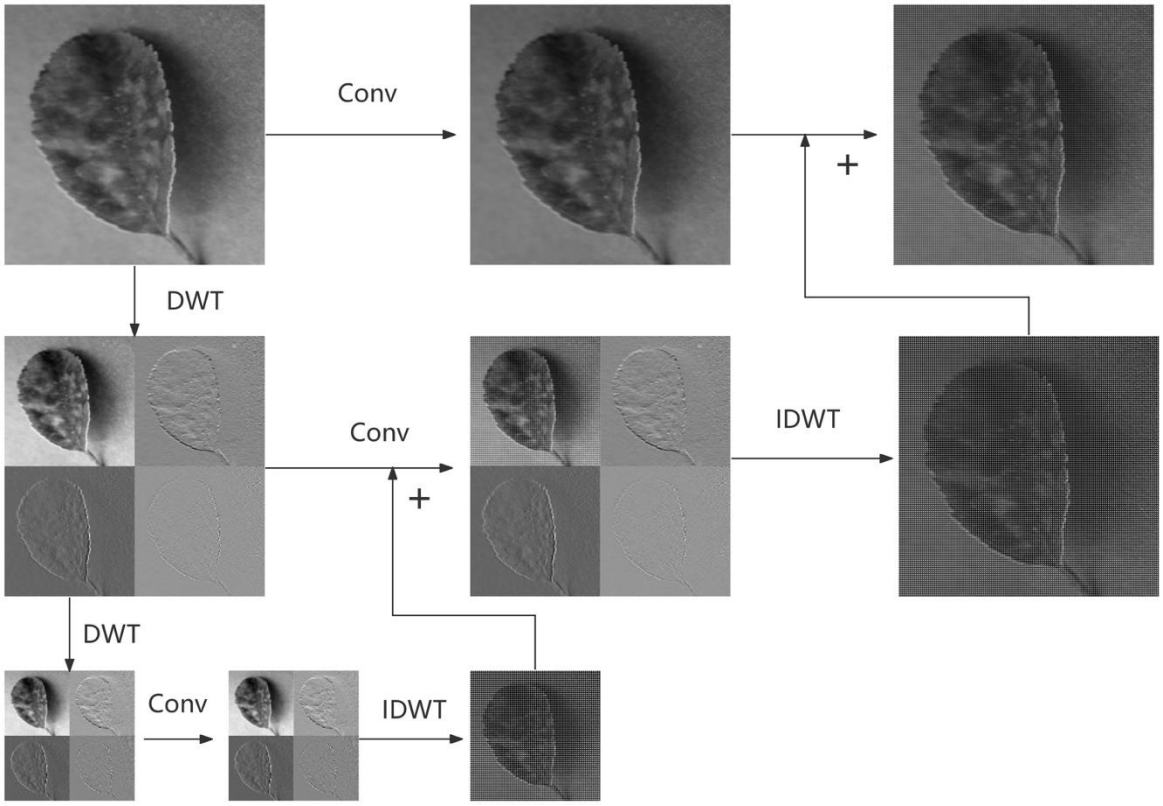

Figure 4. Structure of WT-conv Encoder

2.5 Multichannel Collaborative Attention

In the field of deep learning, especially when dealing with complex data structures, the attention mechanism of models plays a crucial role. Traditional attention modules, while capable of capturing key information, often struggle when processing multimodal data or complex feature interactions. To more efficiently integrate multi-source information and enhance the model's ability to perceive details, we introduce Multi-Collaborative-Attention (MCA)[16]. This mechanism, by constructing multiple attention heads that work collaboratively, not only independently focuses on different subsets of features but also achieves deeper feature fusion through cross-head information interaction. This design enables the model to more comprehensively capture complex relationships in data, thereby demonstrating significant performance improvements in tasks such as multimodal learning, image recognition, and natural language processing. In this paper, we will delve into the architectural design of Multi-Collaborative-Attention and its application effects in specific tasks in detail.

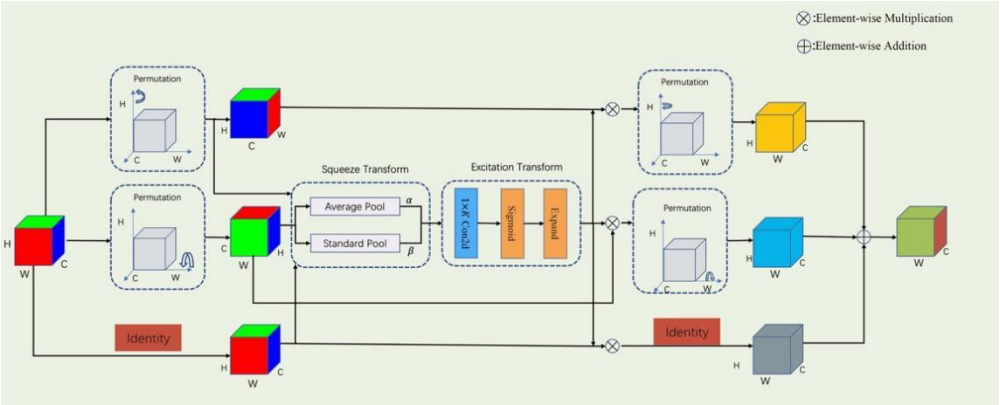

Figure 5. Structure of MCA attention

The MCA module is an advanced multi-dimensional collaborative attention mechanism that comprehensively captures feature dependencies through three meticulously designed parallel branches, with each branch dedicated to processing the height, width, and channel dimensions of the feature map, respectively. The internal structure of the MCA module is illustrated in Figure 5. Starting with the top width branch, the feature map is first rotated 90 degrees counterclockwise along the height axis, and then processed through squeeze transformation and excitation transformation to capture long-range dependencies in the spatial dimension height-wise, generating attention weights for the width dimension. Following that is the middle height branch, where the feature map is rotated 90 degrees counterclockwise along the width axis before undergoing squeeze transformation and excitation transformation to capture long-range dependencies in the spatial dimension width-wise, resulting in attention weights for the height dimension. In the bottom channel weight branch, the feature map undergoes identity mapping to produce an identical feature map, which is then processed through squeeze transformation and excitation transformation to capture interactions between channels and generate attention weights for the channel dimension. Since the MCA module can simultaneously capture feature dependencies in both channel and spatial dimensions, and its spatial branches further enhance the representation capability of these features, by integrating the outputs of the three branches, the MCA module is able to generate more comprehensive and accurate attention weights. Process of Squeeze transform can be seen in Figure 6.

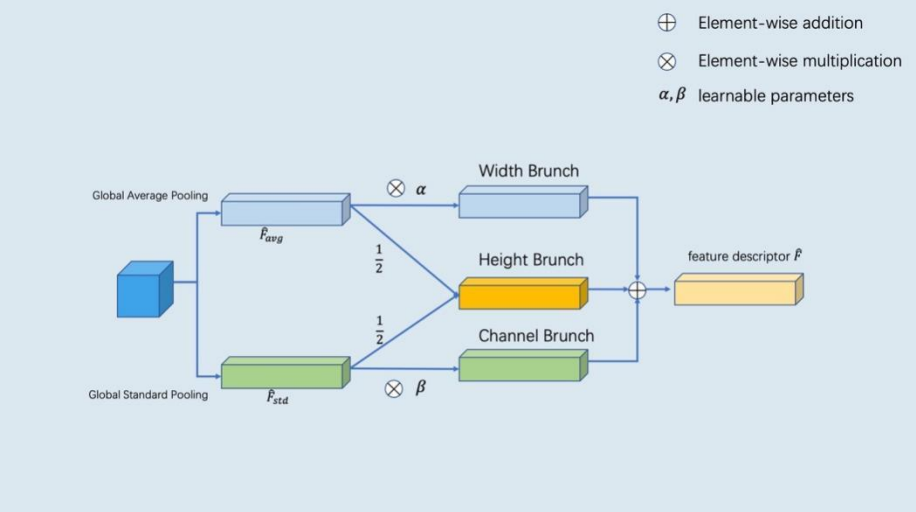

Figure 6. Process of Squeeze transform

3. Results and discussion

The computer used for data processing features an Intel(R) Xeon(R) Platinum 8481C CPU and an NVIDIA GeForce RTX 4090 GPU, running in the operation system of Ubuntu 22.04. The software environment is configured with Python version 3.9.1, and the primary toolkit for model development relies on Pytorch version 2.4.1, along with pytorch-wavelets version 1.3.0[17]

3.1 Training and Evaluation Metrics

In the validation phase of the proposed model's prediction performance, we conducted comparative testing of the model presented in this study against classical models of the same type used in similar research. Eventually, we established and compared models utilized in similar studies, with the final experimentally validated models including Proposed model, DenseNet[10,18], AlexNet[19] and the ResNet [6,20,21]. In the pre-loaded dataset, the batch size for training was set to 16, the number of epochs was set to 200, the Adam optimizer was adopted, and the learning rate was set to 0.0001.

During the testing process, corresponding confusion matrices were generated. To evaluate the prediction performance of the proposed model compared to similar models, we employed various classification task evaluation metrics, including accuracy, precision, recall, and F1 score. The definitions of these metrics are shown in Table 2

Table 2. Evaluation metrics of the classifier based on the confusion matrix.

Criterion	Description
$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$	The proportion of samples which are correctly classified.
$\text{Precision} = \frac{TP}{TP + FP} \times 100\%$	The proportion of true positives among all predicted positives
$\text{Recall} = \frac{TP}{TP + FN} \times 100\%$	The proportion of true positives among all positives
$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\%$	The balance measure between precision and recall
TP: True positive (Correctly predicted as positive)	FP: False positive (Incorrectly predicted as positive)
TN: True negative (correctly predicted as negative)	FN: False negative (Incorrectly predicted as negative)

3.2 Experiment Results

To ensure randomness in the training and testing processes, we divided the dataset according to a 60% and 40% probability ratio. Subsequently, we conducted multiple sampling iterations to partition the dataset into training and testing sets using each model. We validated the performance of the trained models on the testing sets and obtained the average of the validation results. The types, quantities of various samples, and classification outcomes are presented in Table 3. Furthermore, based on the number of classified samples, we calculated the confusion matrices for each model in the classification task, as shown in Figure 7.

Table 3 . Result of classification on different category

Model	Category	Healthy	Black rot	Cedar rust	Scab	Total
Proposed model	Healthy	119	0	0	2	121
	Black rot	4	109	0	1	114
	Cedar rust	0	0	131	0	131
	Scab	3	0	0	119	122
DenseNet	Healthy	91	1	2	15	109
	Black rot	0	123	1	3	127
	Cedar rust	1	0	121	1	123
	Scab	1	0	2	126	129
AlexNet	Healthy	100	1	8	6	115
	Black rot	3	122	1	1	127
	Cedar rust	0	0	121	3	124
	Scab	1	1	0	120	122
ResNet-50	Healthy	114	3	6	1	124
	Black rot	2	123	0	0	125
	Cedar rust	0	0	123	0	123

Scab	3	3	0	110	116

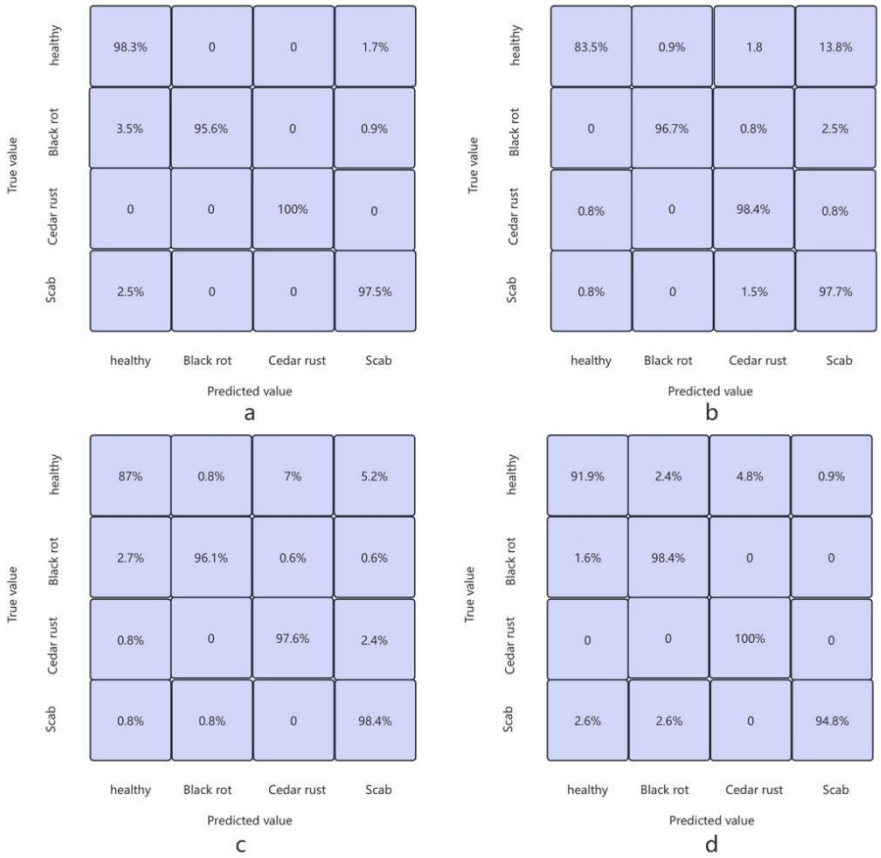

Figure 7. Confusion matrices of models: proposed model(a), DenseNet (b), AlexNet (c), ResNet(d)

From the prediction results of various samples in the test set, it can be observed that in the classification task targeting Black rot lesion samples, ResNet achieved the best classification performance, correctly classifying 98.4% of the Black rot lesion samples. Our proposed model performed slightly inferior to ResNet on this type of sample, correctly classifying 95.6% of the lesion samples, with 4 out of 114 Black rot samples being misclassified as healthy and 1 being misclassified as Scab-infected. For Cedar rust samples, all models exhibited good performance, with our proposed model and ResNet achieving 100% correct classification of Cedar rust samples. DenseNet and AlexNet followed closely, with classification accuracies of 98.4% and 97.6%, respectively. For Scab-type samples, AlexNet achieved the best classification performance, correctly classifying 98.4% of the samples. Our proposed model had a classification accuracy of 97.5%, while ResNet performed the worst, correctly classifying only 94.8% of Scab lesion samples. For healthy samples, our model significantly improved accuracy compared to other models, correctly classifying 98.3% of healthy samples, with only 2 healthy samples being misclassified as Scab by our model. In contrast, among the other models, ResNet had the best classification effect, but only 91.9% of the samples were correctly classified. DenseNet performed the worst, with 15 healthy samples being misclassified as Scab. Most of the mispredicted healthy samples were identified as either Black rot or Scab. Based on these classification results, we calculated the average accuracy, precision, recall, and F1 score from multiple experimental trials for further analysis. The evaluation metrics are shown in Table 4. Among similar models, ResNet exhibits the best classification performance, with a Recall score of 96.4%, indicating that 96.4% of lesion samples can be accurately predicted by ResNet. Our proposed model achieves an average recall rate of 97.9%, which means it precisely predicts the highest number of lesion samples.

Table 4. Evaluation metrics of models

	Accuracy	Precision	Recall	F1
Proposed model	98.1%	98.0%	97.9%	97.5%
DenseNet	94.0%	94.8%	94.1%	94.8%
AlexNet	94.9%	95.0%	94.7%	94.4%
ResNet	96.0%	96.3%	96.4%	96.2%

3.3 Discussion

As can be seen from Table 4, our proposed model outperforms existing similar models across various indicators. The high F1 score further indicates that its performance is superior to classical models in its category, especially when compared to the benchmark model ResNet-50. It is evident that our proposed model achieves excellent results in the classification and prediction of various lesion samples. Figure 7 and Table 3 show that similar models perform well in predicting various diseases. However, they misclassify a considerable number of healthy samples as diseased samples, leading to a significant drop in classification accuracy. Upon analysis, most of the misclassified healthy samples are categorized as Black rot or Scab, and similarly, most misclassified lesion samples are also among these three categories. After analysis, it was found that these three types of samples share significant similarities, and the characteristics of early lesion samples are not pronounced. Traditional classification models may fail to correctly learn the features of various lesions, resulting in large classification errors. Various similar samples are shown in Figure 8

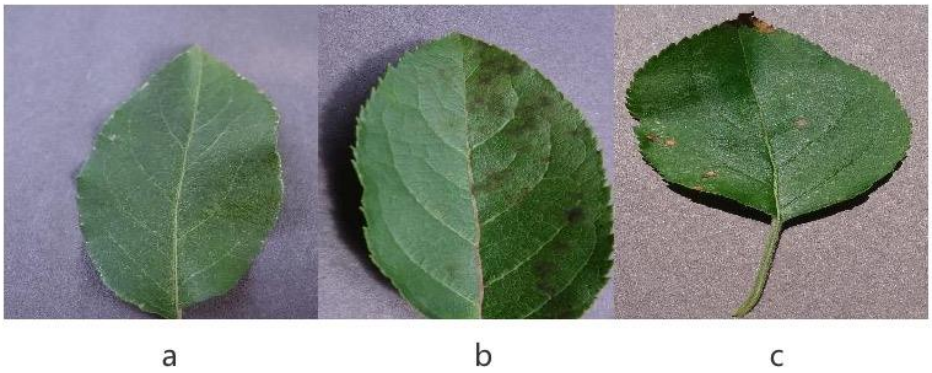

Figure 8. Similar samples of healthy(a), Scab(b), Black rot(c)

Compared to ReseNet, which performs best among similar models, our main improvements lie in the introduction of multi-scale convolution based on wavelet transform and channel collaborative attention. The role of the WT-Encoder is to maximize the utilization of image information in frequency domain feature maps of various sizes through a U-shaped wavelet transform, combining global image information with local feature information. This enables the model to better capture detailed features of lesion samples and improve classification accuracy. The feature map results of the WT-Encoder convolution with double-layer wavelet transform are shown in Figure 4. Following that is the MCA attention module. The purpose of introducing MCA is to capture complex interactions between features through multiple collaboratively working attention heads, thereby enhancing the model's ability to focus on key information. The channel processing results of the MCA module on the feature map are shown in Figure 9. The results demonstrate that the improved network significantly outperforms similar networks.

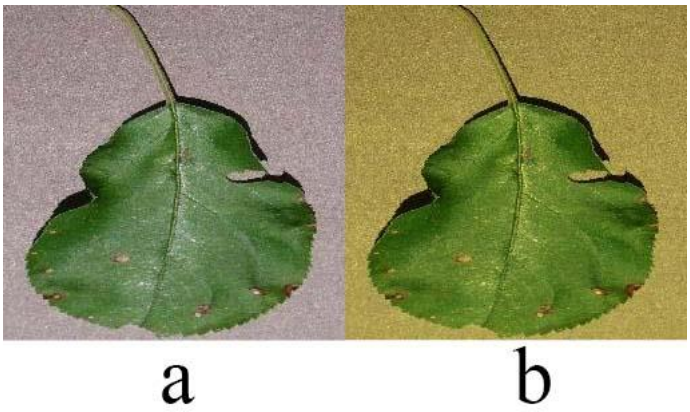

Figure 9. Input(a) and output(b) of MCA layer

4. Conclusion

The fungal infection diseases, including Black rot, Cedar rust, and Scab, that apple leaves are exposed to pose a severe threat to apple production, making the development of efficient lesion detection algorithms particularly urgent. Traditional deep learning algorithms have limitations in detecting various lesion samples. To address this issue, we innovatively designed a multi-scale residual network. This network integrates wavelet transform convolution and MCA attention modules to enhance the classification and prediction accuracy of multiple fungal infection diseases in apple leaves. Specifically, we mainly overcame two major challenges:

(1) Achieving a multi-classification task by accurately predicting multiple lesion samples simultaneously

(2) Significantly improving the classification and prediction accuracy for samples with indistinct features.

To comprehensively verify the effectiveness of the proposed method, we conducted rigorous experimental validation using multiple evaluation indicators and compared its performance with commonly used models in similar studies. Experimental results show that our proposed multi-scale residual network is a feasible and efficient method that can accurately distinguish between healthy samples and various infected lesion samples. This achievement has profound significance for the assessment and prevention of apple leaf diseases in practical production.

Acknowledgments
The authors acknowledge the Heilongjiang Provincial Natural Science Foundation of China (PL2024F026)

Data sharing agreement
The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Declaration of Conflicting Interests
The authors declared no potential conflicts of interest with respect to the research, author-ship, and publication of this article.

References

[1] Crespo, M., Moral, J., Michailides, T. J., & Trouillas, F. P. (2017). First report of black rot on apple fru it caused by Diplodia seriata in California. Plant Disease. https://doi.org/10.1094/PDIS-07-17-1023-PD N

[2] Kumar, S., Kumar, R., & Gupta, M. (2022). Analysis of Apple Plant Leaf Diseases Detection and Class ification: A Review. In 2022 Seventh International Conference on Parallel, Distributed and Grid Comp uting (PDGC) (pp. 361-365). Solan, Himachal Pradesh, India

[3] Okoro, C. A., El-Hasan, A., & Voegele, R. T. (2024). Integrating biological control agents for enhanced management of apple scab (Venturia inaequalis): Insights, risks, challenges, and prospects. Agrochemi cals, 3(2), 118-146. https://doi.org/10.3390/agrochemicals3020010

[4] Mahanti, N. K., Pandiselvam, R., Kothakota, A., Ishwarya, P. S., Chakraborty, S. K., Kumar, M., & Co zzolino, D. (2022). Emerging non-destructive imaging techniques for fruit damage detection: Image pro cessing and analysis. Trends in Food Science & Technology, 120, 418-438. https://doi.org/10.1016/j.tif s.2021.10.00678

[5] Sangeetha, R., Logeshwaran, J., Rocher, J., & Lloret, J. (2023). An improved agro deep learning model for detection of Panama wilts disease in banana leaves. AgriEngineering, 5(2), 660-679. https://doi.org/ 10.3390/agriengineering5020042

[6] Karki, S., Basak, J. K., Tamrakar, N., Deb, N. C., Paudel, B., Kook, J. H., Kang, M. Y., Kang, D. Y., & Kim, H. T. (2024). Strawberry disease detection using transfer learning of deep convolutional neural n etworks. Scientia Horticulturae, 332, 113241. https://doi.org/10.1016/j.scienta.2024.113241

[7] Syed-Ab-Rahman, S. F., Hesamian, M. H., & Prasad, M. (2022). Citrus disease detection and classifica tion using end-to-end anchor-based deep learning model. Applied Intelligence, 52, 927–938. https://doi. org/10.1007/s10489-021-02452-w

[8] Bleasdale, A. J., & Whyatt, J. D. (n.d.). Classifying early apple scab infections in multispectral imagery with convolutional neural networks. SSRN. https://ssrn.com/abstract=4693776 https://doi.org/10.2139/

ssrn.4693776

[9] Banerjee, D. (2024). Unified framework for apple scab disease recognition: Utilizing CNN and random forest models. In 2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI), Coimbatore, India, pp. 1220-1225. https://doi.org/10.1109/ICoICI62503.2024.10695973

[10] Sharma, R., Kukreja, V., Sood, P., & Bhattacharjee, A. (2023). Classifying the severity of apple black rot disease with deep learning: A dual CNN and LSTM approach. In 2023 3rd International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS), Kalady, Ernakulam, India, pp. 173-177. https://doi.org/10.1109/ACCESS57397.2023.10199549

[11] Bourzig, D. K. D., Abed, M., & Merah, M. (2024). Black rot disease classification of apples and grapes using convolutional neural network and transfer learning. In 2024 8th International Conference on Image and Signal Processing and their Applications (ISPA), Biskra, Algeria, pp. 1-6. https://doi.org/10.1109/ISPA59904.2024.10536790

[12] Zhong, Y., & Zhao, M. (2020). Research on deep learning in apple leaf disease recognition. Computers and Electronics in Agriculture, 168, 105146. https://doi.org/10.1016/j.compag.2019.105146

[13] Bansal, P., Kumar, R., & Kumar, S. (2021). Disease detection in apple leaves using deep convolutional neural network. Agriculture, 11(7), 617. https://doi.org/10.3390/agriculture11070617

[14] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. arXiv preprint arXiv:1512.03385. https://arxiv.org/abs/1512.03385

[15] Finder, S. E., Amoyal, R., Treister, E., & Freifeld, O. (2024). Wavelet convolutions for large receptive fields. In Proceedings of the European Conference on Computer Vision (ECCV). https://doi.org/10.48550/arXiv.2407.05848

[16] Yang, Y., Zhang, Y., Cheng, Z., Song, Z., & Tang, C. (2023). MCA: Multidimensional collaborative attention in deep convolutional neural networks for image recognition. Engineering Applications of Artificial Intelligence, 126(Part C), 107079. https://www.sciencedirect.com/science/article/abs/pii/S0952197623012630

[17] Cotter, F. (2020). Uses of complex wavelets in deep convolutional neural networks [Doctoral dissertation, University of Cambridge]. Apollo - University of Cambridge Repository. https://www.repository.cam.ac.uk/handle/1810/306661

[18] Dong, Y.-Y., Huang, Y.-S., Xu, B.-L., Li, B.-C., & Guo, B. (2022). Bruise detection and classification in jujube using thermal imaging and DenseNet. Journal of Food Process Engineering, 45(1), e13981. https://onlinelibrary.wiley.com/doi/full/10.1002/fpe2.13981

[19] Qiu, J., et al. (2024). Research on image recognition of tomato leaf diseases based on improved AlexNet model. Heliyon, 10(13), e33555. https://doi.org/10.1016/j.heliyon.2024.e33555

[20] Desanamukula, V. S., Teja, T. D., & Rajitha, P. (2024). An in-depth exploration of ResNet-50 and transfer learning in plant disease diagnosis. In 2024 International Conference on Inventive Computation Technologies (ICICT), Lalitpur, Nepal, pp. 614-621. https://doi.org/10.1109/ICICT60155.2024.10544802

[21] Kunduracioglu, I. (2024). Utilizing ResNet architectures for identification of tomato diseases. Journal of Intelligent Decision Making and Information Science, 1, 104–119. https://doi.org/10.59543/jidmis.v1i.11949