agriculture

Article

ResViT-Rice: A Deep Learning Model Combining Residual Module and Transformer Encoder for Accurate Detection of Rice Diseases

Yujia Zhang, Luteng Zhong, Yu Ding *, Hongfeng Yu and Zhaoyu Zhai *

College of Artificial Intelligence, Nanjing Agricultural University, Nanjing 210095, China; 9203011528@stu.njau.edu.cn (Y.Z.); 9203011126@stu.njau.edu.cn (L.Z.); hongfengyu@njau.edu.cn (H.Y.)

* Correspondence: 2022009@njau.edu.cn (Y.D.); zhaoyu.zhai@njau.edu.cn (Z.Z.)

Abstract: Rice is a staple food for over half of the global population, but it faces significant yield losses: up to 52% due to leaf blast disease and brown spot diseases, respectively. This study aimed at proposing a hybrid architecture, namely ResViT-Rice, by taking advantage of both CNN and transformer for accurate detection of leaf blast and brown spot diseases. We employed ResNet as the backbone network to establish a detection model and introduced the encoder component from the transformer architecture. The convolutional block attention module was also integrated to ResViT-Rice to further enhance the feature-extraction ability. We processed 1648 training and 104 testing images for two diseases and the healthy class. To verify the effectiveness of the proposed ResViT-Rice, we conducted comparative evaluation with popular deep learning models. The experimental result suggested that ResViT-Rice achieved promising results in the rice disease-detection task, with the highest accuracy reaching 0.9904. The corresponding precision, recall, and F1-score were all over 0.96, with an AUC of up to 0.9987, and the corresponding loss rate was 0.0042. In conclusion, the proposed ResViT-Rice can better extract features of different rice diseases, thereby providing a more accurate and robust classification output.

Keywords: leaf blast disease; brown spot disease; hybrid architecture; transformer encoder; convolutional neural network

check for updates

Citation: Zhang, Y.; Zhong, L.; Ding, Y.; Yu, H.; Zhai, Z. ResViT-Rice: A Deep Learning Model Combining Residual Module and Transformer Encoder for Accurate Detection of Rice Diseases. *Agriculture* **2023**, *13*, 1264. https://doi.org/10.3390/agriculture13061264

Academic Editors: Zhou Zhang and Wen-Hao Su

Received: 22 May 2023
Revised: 15 June 2023
Accepted: 17 June 2023
Published: 19 June 2023

1. Introduction

As one of the world's most important food crops, rice is of great significance to global food security and agricultural sustainability. Rice is a major food source in countries around the world, and it also supports the livelihoods and economic development of millions of people. Especially in many developing countries, rice as a major food crop not only provides the food and energy needed for human life but also plays a positive role in promoting employment and economic growth in rural areas.

However, rice diseases pose a serious threat to rice yield and quality, thereby affecting the economic benefits and food security of farmers. Various types of rice diseases, such as blast disease, sheath blight, brown spot, and bacterial leaf blight, have a significant impact on rice yield and quality. Leaf blast (*Magnaporthe oryzae*) is a serious fungal disease in rice caused by *Pyricaria oryzae Cavara*. It can infect any aboveground tissue of the rice plant at any stage of its growth, causing lesions on leaves, leaf collars, stems, nodes, neck nodes, and panicles [1]. Leaf blast has a negative impact on the physical properties of rice. Rice grains infected by this disease dry out 10% [2] more than normal rice grains, and the thickness of the rice grains decreases by 10%. The impact of leaf blast on rice yield is enormous [3], as studies show that leaf blast can reduce rice yield by an average of 35% [4].

Brown spot (*Bipolaris oryzae*) is also a fungal disease, and it infects coleoptiles, leaves, leaf sheaths, panicle branches, glumes, and spikelets. Brown spot, caused by *Bipolaris zeicola*, is a major fatal disease in rice that can cause qualitative and quantitative crop damage [5–7]. Research indicates that bacterial brown spot in rice can reduce rice yields by up to 52% [8]. Due to the widespread presence and serious threat of rice diseases, the detection of rice diseases is particularly important. Early and accurate detection of rice diseases can help farmers take timely prevention and control measures and reduce the impact of diseases on rice yield and quality. Therefore, the development of efficient and accurate rice disease-detection methods is urgently needed.

To date, there have been many studies on the detection of rice diseases involving various methods applied in rice disease detection. Among them, many scholars have utilized deep learning techniques [9,10]. Wang et al. [11] proposed the ADSNN-BO model, which is a method based on attention neural networks and Bayesian optimization for rice disease detection and classification, using the MobileNet structure and enhanced attention mechanism. This model can effectively learn feature information and achieve an accuracy of 94.65%. Daniya et al. [12] proposed the RideSpider water-wave algorithm based on deep recurrent neural network and achieved a maximum accuracy of 90.5% in detecting rice plant diseases. These studies extracted more distinctive features such as texture and color for feature extraction. In addition to algorithmic recognition research, there has also been research on rice disease detection. These methods typically use the YOLO object-detection algorithm to achieve automation. Kim et al. [13] proposed a system for predicting and automatically detecting the infection rate of rice bakanae disease (RBD) through drone images, using the YOLOv3 and RestNETV2 101 models for detecting infected bundles and classifying infected panicles, with average accuracies of 90.49% and 80.36%, respectively. Haque et al. [14] achieved 90% accuracy in rice disease detection using the YOLOv5 deep learning method. It is worth noting that several significant studies have begun to explore the use of deep learning techniques for the detection of diseases in other crops, such as the fusarium head blight in wheat [15], the Alternaria leaf blotch disease in apple trees [16], as well as broader grain-crop phenotyping [17]. These studies demonstrated the potential and feasibility of utilizing deep learning for the detection of diseases in rice [18]. Despite the many applied rice disease-detection methods, the accuracy of these methods still needs further improvement. Some methods also need to improve their applicability, especially in dealing with complex rice diseases, and adaptability in different environments. Finally, some detection methods such as fluorescence quantitative PCR and digital PCR are limited by equipment and technical requirements, making it difficult to deploy them in the areas with limited sources. Future research needs to address these issues to better support the prevention and control of rice diseases.

In order to develop a more accurate and efficient method for detecting rice diseases, this study proposes a deep learning model to improve the accuracy and efficiency of rice disease detection. The model utilizes image processing and machine learning techniques for training and optimization using similar images of different categories of rice diseases, with stronger representation and learning capabilities. Specifically, this study proposes a novel hybrid model called ResViT-Rice, which combines the convolutional neural network and transformer architecture. The model is specifically designed for detecting rice diseases. The contribution of this article can be summarized by the following three main points:

- The incorporation of the ResNet [19] model as the backbone network of our structure enabled effective extraction of image features. The employment of residual blocks paved the way for efficient information transfer, mitigating the gradient vanishing issue and thereby enhancing the stability during the training phase, all the while reducing the overall parameters;
- We incorporated the transformer architecture into our model, aiming to leverage its powerful self-attention mechanism, which demonstrated exceptional performance in image-processing tasks. Our approach adopted a hybrid structure that combined CNNs and transformer encoder. The CNN component provided spatial inductive bias and accelerated network convergence, thereby enhancing the stability of the training process;

- The convolutional block attention module (CBAM) attention mechanism was integrated into the ResViT-Rice block, allowing the model to adjust adaptively to the significance of different regions within the input feature map. This was especially beneficial for rice disease-detection tasks where disease localization within the image might be random. By deploying attention mechanisms, we ensured that the model prioritized disease-afflicted areas, thereby boosting the model's accuracy.

To evaluate the performance of the model, we conducted comparative experiments with traditional rice disease-detection methods and mainstream CNN models. Meanwhile, to further underscore the superiority of our ResViT block, we also carried out ablation experiments. These investigations served to underline the integral role that the ResViT block played in the overall performance of the model, thereby solidifying its place in our future efforts in the field of rice disease detection. The workflow of this study is shown in Figure 1. The results showed that ResViT-Rice obtained better accuracy in complex disease situations and can provide strong support for early warning and precise control of rice diseases. Therefore, ResViT-Rice is expected to be widely used in future rice disease detection.

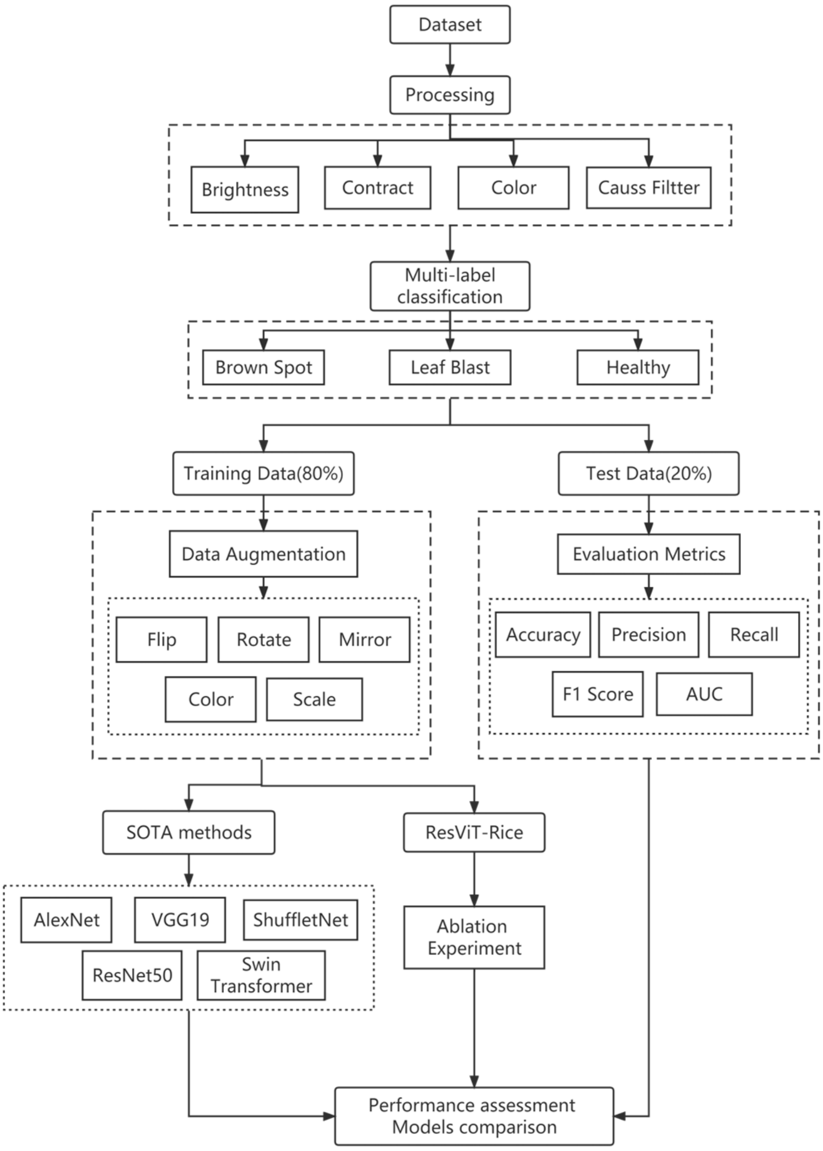

Figure 1. Workflow of this study.

2. Materials and Methods

2.1. Data Source

The data used in this study were obtained from Kaggle (https://www.kaggle.com/datasets/tiffanyjade/rice-disease-image-dataset, accessed on 1 March 2023), a well-known open data science community that provides a large number of public datasets for researchers and data scientists. We obtained a dataset from Kaggle that includes images of two types of rice diseases, brown spot and leaf blast, as well as a category of healthy rice leaves. Each category consists of 516 images. This dataset covers different types of rice diseases and has a rich sample size, providing ample data resources for training and validating our deep learning model in this study.

2.2. Data Preprocessing

Data preprocessing plays a crucial role in deep learning, as it can greatly improve the performance and robustness of the model. In this study, we performed data preprocessing on the rice disease images obtained from Kaggle. The preprocessing included adjusting the contrast, brightness, and color of the images as well as applying Gaussian filtering to remove noise. Adjusting the contrast, brightness, and color can enhance the details in the images, while Gaussian filtering is a commonly used image-filtering method that can effectively reduce noise and smooth the image. These processes helped improve the image quality and make subsequent feature extraction and model training more accurate.

To ensure the stability of the model and the convergence of the training, we performed the normalization operation over the original dataset. Normalization can scale the pixel values of the images to a fixed range (e.g., 0 to 1), while the size of original images was resized to 224 × 224 to better fit the input requirements of the deep learning model. Meanwhile, the original dataset was split into training and testing sets in an 8:2 ratio. In addition, to increase the diversity and richness of the training set, we also performed data augmentation on the images in the dataset, including random rotation, translation, scaling, and flipping. Data augmentation can increase the size of the training set and reduce the risk of overfitting. Through normalization and data augmentation, we can enhance the model's ability to process and detect rice disease images. The final dataset size is shown in Figure 2, with each category processed from the original 516 images into 1648 training images and 104 test images.

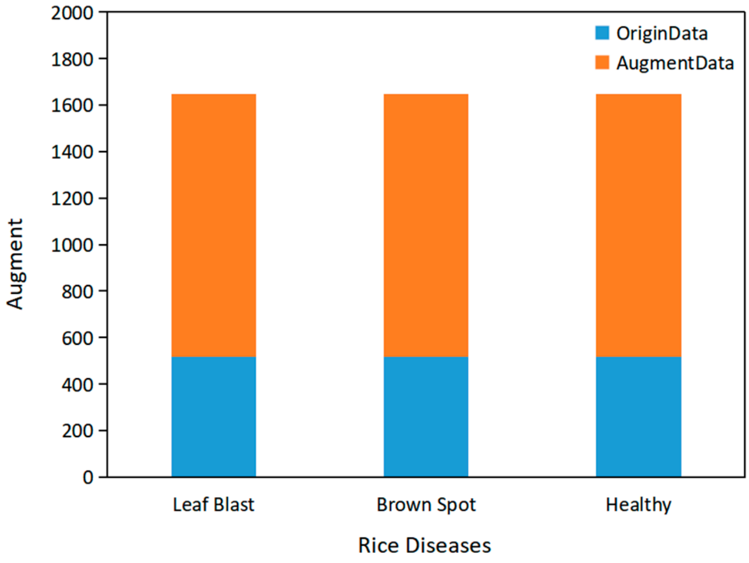

Figure 2. The amount of data after pre-processing.

2.3. Proposal of ResViT-Rice

In this study, ResViT-Rice was proposed based on the architecture of deep convolutional neural networks and vision transformers. A newly designed module, ResViT-Rice block, was added to ResNet [19] to introduce the self-attention mechanism and global view on the basis of CNN to improve the performance of the model. Figure 3 shows the proposed model architecture, which is mainly composed of ordinary convolution, bottleneck, ResViT-Rice block, adaptive pooling, and fully connected layers.

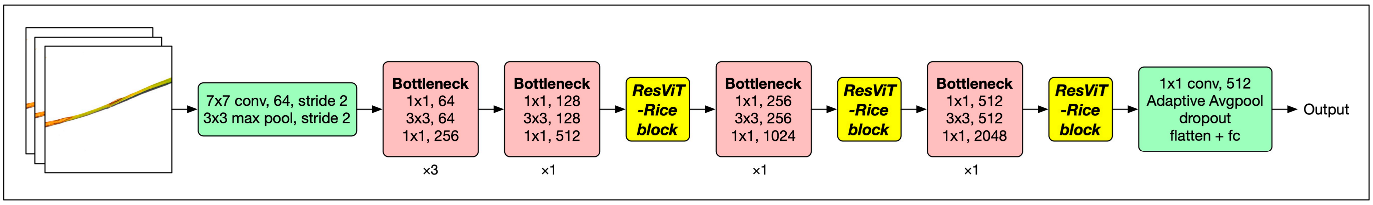

Figure 3. Architecture of ResViT-Rice model.

For the general ResNet [19] architecture, it used residual connections to alleviate the gradient-vanishing problem and make the network more easily trained. As the core structure of ResNet, the bottleneck is mainly composed of three convolutional layers, which were the 1 × 1 convolutional layer, 3 × 3 convolutional layer, and 1 × 1 convolutional layer, respectively. The output channels of these convolutional layers were 1/4 of the original input channel, 1/4 of the original input channel, and 4 times the original input channel, respectively. This setting allowed the bottleneck module to reduce computational complexity, increase network depth, and improve feature-extraction ability. The main idea of the bottleneck module was to introduce a bottleneck structure that mapped the input features to a low-dimensional space through a low-dimensional bottleneck layer and then mapped the features back to the original dimension through a high-dimensional expansion layer. This structure can reduce the number of parameters and computational complexity. It can also improve the feature-extraction ability, thus achieving better performance. Given the aforementioned reasons, coupled with its impressive capabilities in the visual domain, we chose ResNet as our backbone network.

The most novel contribution, the ResViT-Rice block in ResViT-Rice, is shown in Figure 4. It mainly consists of four components. First is the input channel transformation, which reduces the dimension of the input channel using a 1 × 1 convolutional layer. This can reduce computational complexity and the number of parameters and keep the feature dimensionality consistent when inputting into the transformer encoder each time. Next are the transformers with global views. Assuming that the feature map after the input channel transformation is (H, W, D), the feature map can be unfolded on the surface of H and W axis to obtain a word vector of (HW, D). After adding the positional encoding to this word vector, it was sent to the encoder component. The formula of positional encoding can be written in Equations (1) and (2) as follows:

$$\mathrm{PE}(pos, 2i) = \sin\left(\frac{pos}{10000^{2i/d_{model}}}\right) \tag{1}$$

$$\mathrm{PE}(pos, 2i+1) = \cos\left(\frac{pos}{10000^{2i/d_{model}}}\right) \tag{2}$$

where pos represents the position in the input sequence, i represents the dimension index in the PE vector, and d_{model} represents the embedding dimension in the transformer model.

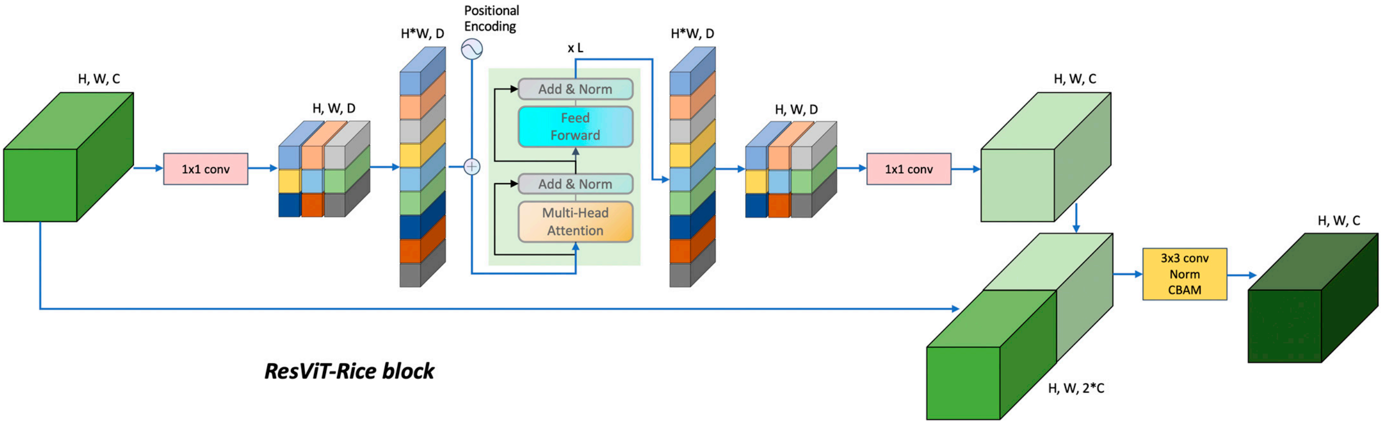

Figure 4. Architecture of ResViT-Rice block module.

The PE vector has a dimension of d_{model}, so each position has a d_{model}-dimensional PE vector. This formula uses sine and cosine functions, which have different periods in different dimensions. Therefore, for different dimensions, the values in the PE vector vary according to different periods, providing unique encoding for different positions. By adding the PE vector to the word-embedding vector, the transformer model can capture the positional information in the input sequence. Additionally, the formula for the multi-head attention mechanism in transformer encoder is given in Equation (3):

$$\mathrm{Multi-Head}(Q,K,V) = \mathrm{Concat}(\mathrm{head}_1,\ldots,\mathrm{head}_h)W^O \tag{3}$$

where Q, K, and V represent the query, key, and value vectors, respectively. h represents the number of heads, Concat refers to concatenating the heads together, and W^O is the weight matrix for the output. The calculation for each head follows Equation (4):

$$\mathrm{head_i} = \mathrm{Attention}\left(QW_i^Q, KW_i^K, VW_i^V\right) \tag{4}$$

where W_i^Q, W_i^K, and W_i^V are the weight matrices used to perform linear transformations on the query, key, and value vectors, respectively. Attention refers to the attention function used in the calculation. The attention function is calculated by Equation (5):

$$\mathrm{Attention}(Q,K,V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{5}$$

where d_k represents the dimensionality of the query or key vectors.

After obtaining the output, it can be transformed to the original feature map size (H, W, D) based on the processing method used for the original input. Finally, the fusion and output part first use a 1 × 1 convolution to transform the feature channel number to the original size and then perform residual concatenation with the input feature map along the channel direction. Then, a 3 × 3 convolutional kernel is used for feature fusion and dimension reduction, and after batch normalization, the CBAM attention mechanism is used again for global feature fusion and extraction. The CBAM attention module used here includes a channel attention module and a spatial attention module. The channel attention module adjusts the feature representation of different channels by learning channel weights, thereby enhancing the model's attention to different channel features. The spatial attention module adjusts the feature representation of different spatial positions by learning spatial weights, thereby enhancing the model's attention to different spatial positions. The CBAM attention module can adaptively adjust the weights of channel and spatial features to better fuse and extract features. Finally, in Table 1, specific values for H, W, C, and D are provided.

The serial number corresponds to the three ResViT-Rice blocks in Figure 3, respectively (the three blocks marked in bright yellow from left to right).

Table 1. Pre-defined parameters of transformer encoder.

The Serial Number of ResViT-Rice Block	1	2	3
H	28	14	7
W	28	14	7
C	512	1024	2048
D	64	128	256
dim of feedforward	128	256	512
number of layers	2	4	3

As illustrated in Table 2, we aimed at reducing the training cost of the ResViT-Rice model. To this end, we cataloged in the table the quantity of parameters contained within each layer of the model while simultaneously tracing the evolution of image feature maps at every stage. Through this comprehensive presentation of data, we can effectively demonstrate the resource consumption involved during the training process of our model.

Table 2. ResViT-Rice architectural dimensions and params.

Layer Name	Output Size	Parameters
Input image	224 × 224 × 3	0
Conv 1	112 × 112 × 64	9408
BatchNorm	112 × 112 × 64	128
ReLu+MaxPool	112 × 112 × 64	0
Bottleneck 1-1	56 × 56 × 256	75,008
Bottleneck 1-2	56 × 56 × 256	70,400
Bottleneck 1-2	56 × 56 × 256	70,400
Bottleneck 2	28 × 28 × 512	379,392
ResViT-Rice block 1	28 × 28 × 512	4,967,906
Bottleneck 3	14 × 14 × 1024	1,512,448
ResViT-Rice block 2	14 × 14 × 1024	19,955,810
Bottleneck 4	7 × 7 × 2048	6,039,552
ResViT-Rice block3	7 × 7 × 2048	79,192,674
Conv 2	7 × 7 × 512	1,048,576
AdaptiveAvgPool+Dropout	1 × 1 × 512	0
Linear	2	1026
Total	/	113,322,728

2.4. Other Mainstream Models

In this study, we also compared ResViT-Rice with other mainstream models, including AlexNet, ResNet50, VGG19, ShuffleNet, and Swin transformer. ResNet50 uses a special type of cross-layer connection [19], and this design allowed for smoother information flow and avoiding the problem of gradient vanishing, making the training process more stable. VGG19 uses small (3 × 3) convolutional kernels and a large number of convolutional layers to extract richer features [20] and employs max pooling layers to reduce the size of feature maps, followed by three fully connected layers for classification. ShuffleNet employs group convolution and channel shuffling to achieve efficient feature extraction and computation [21]. AlexNet is a deep convolutional neural network model proposed by Alex Krizhevsky in 2012 [22]. The Swin transformer is a hierarchical transformer model that achieves high efficiency and better performance through layered attention mechanisms [23].

These five network models all use the Adam optimizer and cosine annealing, with a learning rate of 0.0001 and a batch size of 32. The Adam optimizer can adaptively adjust the learning rate for each parameter, and the cosine annealing algorithm can dynamically adjust the learning rate to improve the model's generalization ability. The specific parameters of the cosine annealing algorithm are T_0 = 10 and T_mul = 2.

2.5. Model Evaluation

In this manuscript, we evaluated the deep learning models using various evaluation metrics, including accuracy, precision, recall, F1-score, AUC, confusion matrix, and ROC curve. Among them, accuracy is the most commonly used evaluation metric for classification models; precision and recall are used to measure the prediction accuracy and coverage of the model; and F1-score is a harmonic mean that takes both precision and recall into consideration. The confusion matrix can be used to visualize the classification model's prediction results, providing more detailed performance analysis. The ROC curve is used to evaluate the performance of binary classification models, where the larger the area under the curve (AUC value), the better the model's performance. Table 3 shows the formulas and explanations of the various evaluation metrics used in this study, which can comprehensively evaluate the model's performance and provide strong support for model selection and improvement.

Table 3. Formulas of evaluation metrics.

Metric	Equation
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$
Precision	$\frac{TP}{TP + FP}$
Recall	$\frac{TP}{TP + FN}$
F1-score	$\frac{precision \times recall}{precision + recall}$

In addition, to demonstrate the generalization capability of our model, namely its excellent performance under changes in environmental conditions and plant diseases, we conducted a set of experiments on a dataset [24] with a complex background and a wider range of rice disease categories. The corresponding results were added to the Supplementary Materials.

2.6. Ablation Experiments

Ablation experiments are a vital approach in machine learning, and they are used to understand the contribution of individual components to the overall performance of a model. By systematically ablating parts of the model, the effect on performance can be observed, providing a way to evaluate the importance of these components.

In this research, we conducted three groups of comparison experiments. We applied the ResViT-Rice model to two detection tasks, namely leaf blast and brown spot. For each task, we performed ablation experiments on different elements of the model: block3, a combination of block2 and block3, and the convolutional block attention module (CBAM), respectively. These ablation experiments allowed us to measure the contribution of each individual block and the CBAM attention mechanism to the overall performance of the ResViT-Rice model.

3. Result

3.1. Experimental Setup

In this study, the dataset was divided into training and validation sets at an 8:2 ratio. The experiments were conducted on a computer equipped with an NVIDIA RTX 3090 GPU and 11th generation Intel Core i7 CPU, providing sufficient computing power to support deep learning model training and evaluation. We used PyTorch 2.0 with CUDA 11.7 as the deep learning framework and Python 3.9.16 as the programming language. To assist in implementing deep learning models and model evaluation, we also utilized several commonly used Python libraries, such as scikit-learn 1.2.2, numpy 1.23.5, and pandas 1.5.3.

3.2. Results of Data Preprocessing

In order to enhance the diversity of the dataset and improve the generalization ability of the model, we first preprocessed the data. As shown in Figure 5a–c, from top to bottom, comparisons of brown spot, leaf blast, and healthy images before and after data preprocessing and augmentation are presented. It can be observed that after image preprocessing, the characteristics of each category became more distinct, while the semantic information was preserved well, which is beneficial for the model to learn and classify effectively.

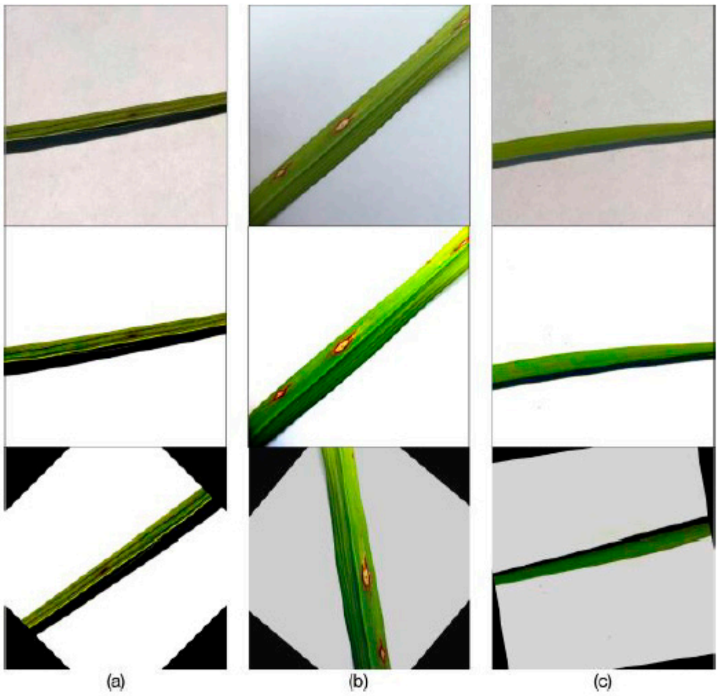

Figure 5. The sample images of the rice disease dataset before and after data preprocessing and augmentation: (**a**) brown spot, (**b**) leaf blast, and (**c**) healthy.

3.3. Results of ResViT-Rice

The results suggested that ResViT-Rice achieved excellent performance. As shown in Table 4, it achieved a recognition accuracy of 99.04% and 96.63% for the two types of rice diseases, respectively, which were the highest among all the models evaluated, with an average accuracy of 97.84%. Compared with AlexNet, ResNet50, VGG19, ShuffleNet, and Swin transformer, the ResViT-Rice model outperformed the best-performing ResNet50 model by 4.81% in terms of accuracy. In addition, when evaluated with precision, recall, F1-score, and AUC value, the ResViT-Rice model consistently outperformed the other four classic CNN models by at least 4% across all evaluation metrics.

From another perspective, Figure 6a,b depict the confusion matrix of the five classic models and ResViT-Rice on the brown spot and leaf blast tasks, respectively. By comparing the confusion matrix, it was evident that the ResViT-Rice model achieved the best classification performance, as indicated by the darkest colors on the main diagonal. As shown in Figure 7, the ROC curves further illustrate the significant differences in performance among the models. In Figure 7b, the ResViT-Rice curve almost entirely overlaps with the top-left corner, indicating the largest area under the curve (AUC). The AUC values of ResViT-Rice for both rice diseases reached 0.99, demonstrating its strong generalization ability and superior performance in disease detection.

Table 4. Evaluation Results of AlexNet, ResNet50, ShuffleNet, VGG19, Swin transformer, and ResViT-Rice. Bold fonts indicate the best performance in each category.

	Model	Accuracy	Precision	Recall	F1 Score	AUC
Leaf Blast	AlexNet [22]	0.9087	0.9087	0.9087	0.9087	0.9735
	ResNet50 [19]	0.9567	0.9578	0.9567	0.9567	0.9850
	VGG19 [20]	0.8221	0.8236	0.8221	0.8219	0.9056
	ShuffleNet [21]	0.7067	0.7072	0.7067	0.7066	0.8108
	Swin Transformer [23]	0.9326	0.9352	0.9326	0.9325	0.9766
	ResViT-Rice	**0.9904**	**0.9904**	**0.9904**	**0.9904**	**0.9987**
Brown Spot	AlexNet [22]	0.8990	0.9020	0.8990	0.8988	0.9449
	ResNet50 [19]	0.9038	0.9044	0.9038	0.9038	0.9583
	VGG19 [20]	0.9038	0.9040	0.9038	0.9038	0.9517
	ShuffleNet [21]	0.7788	0.7826	0.7788	0.7781	0.8760
	Swin Transformer [23]	0.8557	0.8623	0.8557	0.8557	0.9495
	ResViT-Rice	**0.9663**	**0.9674**	**0.9663**	**0.9663**	**0.9873**
Healthy	AlexNet [22]	0.9294	0.9330	0.9294	0.9302	0.9705
	ResNet50 [19]	0.9358	0.9362	0.9358	0.9360	0.9863
	VGG19 [20]	0.8974	0.9015	0.8974	0.8985	0.9436
	ShuffleNet [21]	0.7500	0.7500	0.7500	0.7500	0.8347
	Swin Transformer [23]	0.8846	0.8963	0.8846	0.8867	0.9636
	ResViT-Rice	**0.9839**	**0.9840**	**0.9839**	**0.9839**	**0.9962**

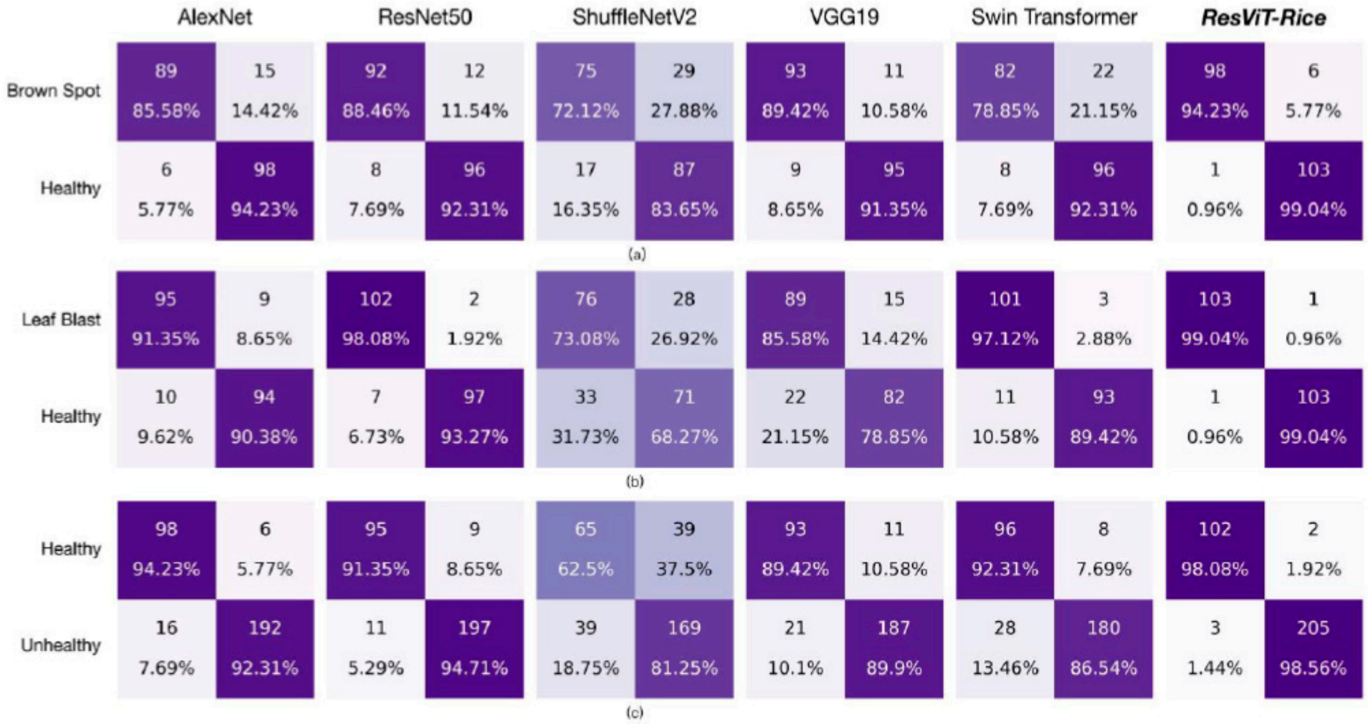

Figure 6. Confusion matrix of AlexNet, ResNet50, ShuffleNet, VGG19, Swin transformer, and ResViT-Rice: (**a**) brown spot, (**b**) leaf blast, and (**c**) healthy.

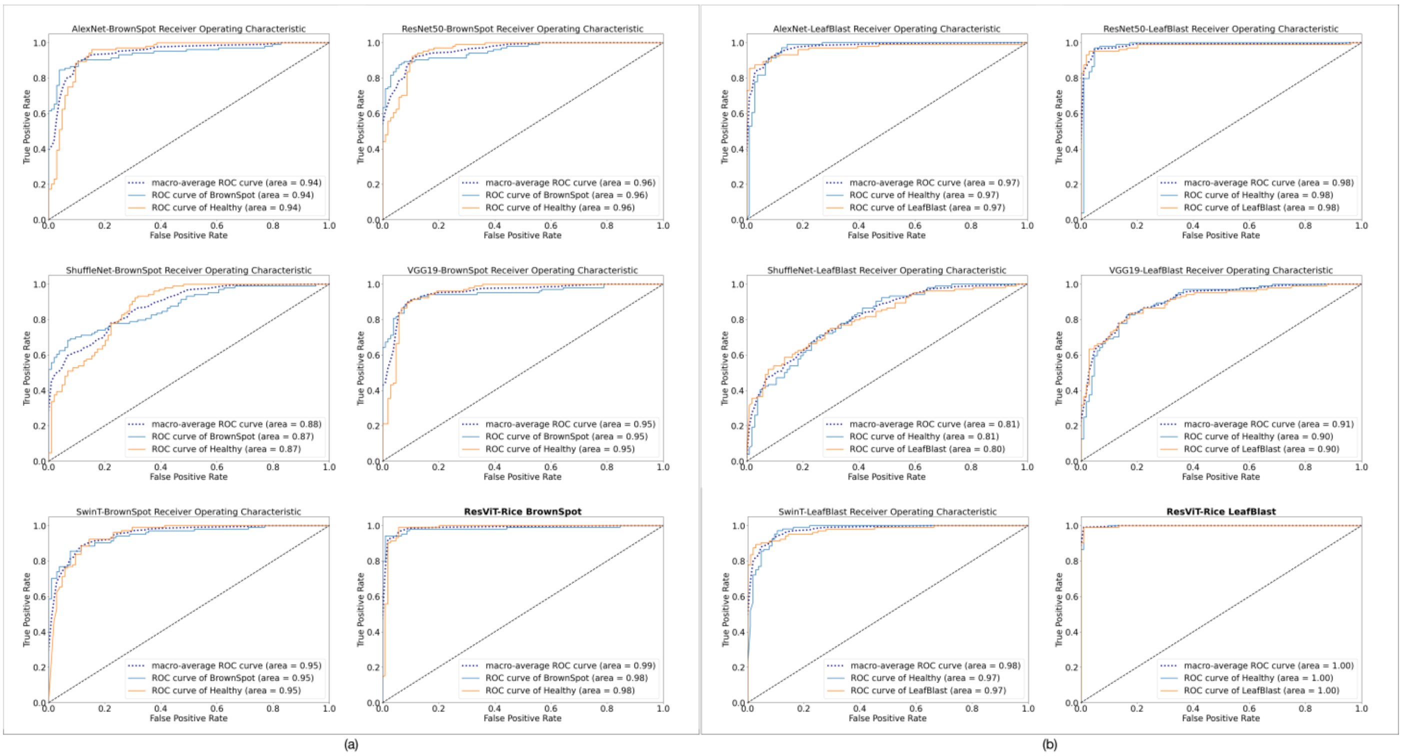

Figure 7. ROC curves of AlexNet, ResNet50, ShuffleNet, VGG19, Swin transformer, and ResViT-Rice: (**a**) brown spot and (**b**) leaf blast.

As shown in Table 4, ShuffleNet performed the worst, with an accuracy of less than 0.8 in the rice disease-detection task. This may be due to the fact that lightweight networks were limited to extracted features. ResNet50 had the best classification performance among the all models, achieving an accuracy of over 0.9. The use of residual modules allowed ResNet50 to better extract features, which greatly improved network performance, so residual neural networks often performed well in various classification tasks. However, there was still a considerable gap between ResNet50 and ResViT-Rice.

The Swin transformer demonstrated an acceptable performance, achieving 0.9326 accuracy in the leaf-blast-classification task. However, its performance in the brown-spot-classification task was mediocre, indicating both the potential and limitations of the Swin transformer. In this classification task, the performance of AlexNet ranked second to ResNet50, with an average accuracy above 0.9038 in both rice disease-detection tasks. VGG19, as the most complex and parameter-heavy model among these networks, had accuracy of 0.8221 and 0.9038 in the two classification tasks, respectively. However, its overall accuracy was less than 0.9.

3.4. Results of SOTA Models

Table 5 summarizes the latest research on rice disease classification and the corresponding accuracy. Various methods were used to complete the rice disease classification task. For instance, Wang et al. [11] applied both the attention mechanism and Bayesian optimization to a depth-wise separable neural network. Kim et al. [13] and Haque et al. [14] adopted the YOLO serious model to classify different classes of rice diseases. It is also noted that various backbones were used, including ResNet, GoogLeNet, VGG, ShuffleNet, and so on. Although the dataset used in each study varied in sample sizes, the disease categories in these remained almost the same. Judging from the evaluation metrics (accuracy, precision, recall, etc.), it can be concluded that ResViT-Rice achieved the optimal performance among all.

Table 5. Summary of different latest studies on rice disease classification with the corresponding accuracies.

Reference	Method	Dataset Used	Accuracy	Precision	Recall	F1-Score
[11]	Neural network with Bayesian optimization	500 images	0.9465	0.9260	0.8740	0.8960
[12]	RideSpider water wave (RSW)	NA	0.9050	NA	0.7900	NA
[13]	YOLOv3	28,365 images	0.9049	NA	NA	NA
[14]	YOLOv5	1500 images	0.9000	0.9000	0.6700	0.7600
[25]	VGG-16	NA	0.9246	NA	NA	NA
[26]	Inception-ResNet-V2	984 images	0.9268	0.9370	0.9260	0.9280
[27]	Residual neural network	120 images	0.9583	0.9400	0.9400	0.9400
[28]	Improved ShuffleNet V2	1608 images	0.9440	0.9680	0.9670	0.9680
Our Work (ResViT-Rice)	**Hybrid architecture of CNN and transformer**	**1548 images**	**0.9910**	**0.9900**	**0.9900**	**0.9900**

Note: Bold values indicate the optimal performance.

3.5. Feature-Visualization Process

Deep learning models have shown remarkable performance in tasks such as image classification and object detection. However, due to their black-box nature, explaining the decision-making process behind their predictions remains a challenge. Visualizing the feature maps of neural networks has been widely used to gain further insights into the model. Grad-CAM is a gradient-based explainability technique that generates heatmaps and highlights the areas (pixels) of a given image that the model focuses on, thereby aiding in understanding the model's decision-making process and inference basis. To better understand the differences between feature maps extracted from different rice leaf disease images and evaluate the model's attention regions, this experiment used Grad-CAM for visualization. The experimental results, shown in Figure 8, help further elucidate the model's decision-making process and feature-extraction process.

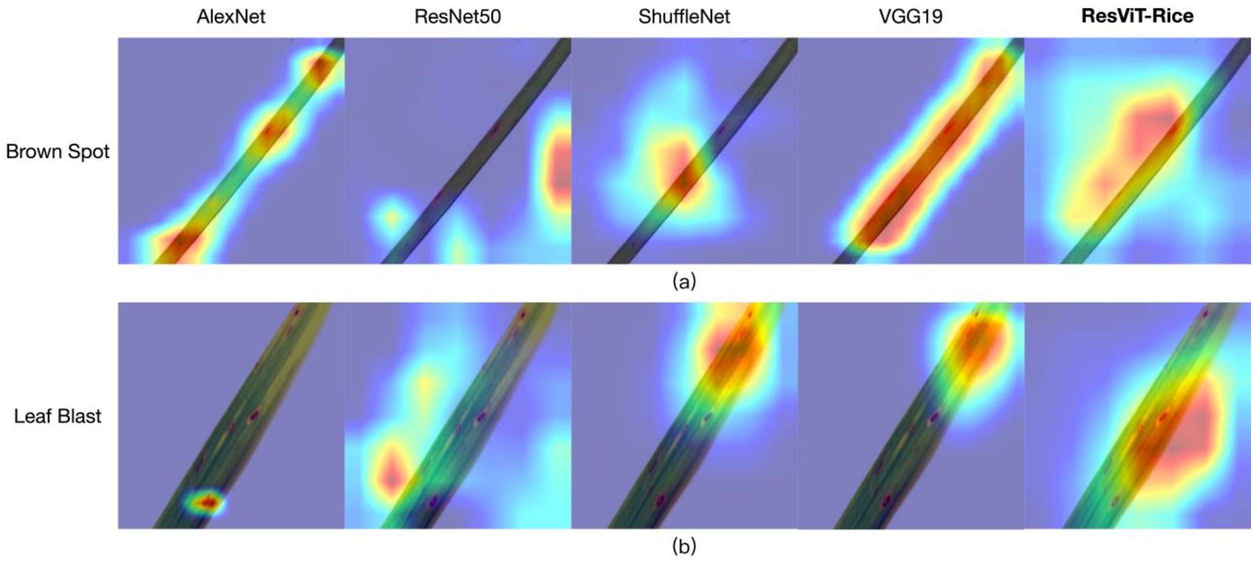

Figure 8. Grad-CAM heatmap of AlexNet, ResNet50, ShuffleNet, VGG19, and ResViT-Rice: (**a**) brown spot and (**b**) leaf blast.

From the results of ResViT-Rice shown in Figure 8a, it can be observed that the model exhibited deeper colors and textures in the vicinity of the disease spots, indicating that the model can accurately capture the features of these disease spots. In comparison, AlexNet, ShuffleNet, and VGG also highlighted the colors and textures in the regions where the disease spots were located to varying degrees. However, AlexNet additionally highlighted an area that should not be focused on, mistaking it for a disease spot, while ShuffleNet failed to cover all the disease spots comprehensively. VGG19 almost focused on the entire leaf, failing to better highlight the prominent features. Although ResNet50 achieved the highest accuracy compared with other mainstream models, it can be observed from both Figure 8a,b that the features were not related to the disease spots. For the leaf-blast-diseased leaves shown in Figure 8b, ResViT-Rice also generated correct colors and textures in the most densely concentrated disease spots, obtaining the best features in comparison.

3.6. Results of Ablation Experiments

The results of the ablation experiments for the ResViT-Rice model are presented in Table 6. These outcomes offer a detailed insight into the relative significance of each module/block in the model.

Table 6. Results of the ResViT-Rice model ablation experiment. Bold fonts indicate the best performance in each category.

	Module to be Ablated	Accuracy	Precision	Recall	F1 Score	AUC
Leaf Blast	Block 3	0.8605	0.8704	0.8605	0.8596	0.9581
	Block 2 and 3	0.8509	0.8512	0.8509	0.8509	0.9378
	CBAM	0.8221	0.8235	0.8221	0.8219	0.9000
	None (ResViT-Rice)	**0.9904**	**0.9904**	**0.9904**	**0.9904**	**0.9987**
Brown Spot	Block 3	0.8990	0.9035	0.8990	0.8987	0.9596
	Block 2 and 3	0.8557	0.8605	0.8557	0.8552	0.9498
	CBAM	0.8653	0.8675	0.8653	0.8651	0.9258
	None (ResViT-Rice)	**0.9663**	**0.9674**	**0.9663**	**0.9663**	**0.9873**

For the leaf-blast-detection task, the removal of block 3 resulted in an accuracy, precision, recall, and F1-score of 0.8605, 0.8704, 0.8605, and 0.8596, respectively, with an AUC of 0.9581. This reflected a decrease in performance compared to the intact ResViT-Rice model, thereby indicating the importance of block 3 to the model's functioning. When both blocks 2 and 3 were ablated, the metrics showed a further slight decline. However, the most considerable reduction in model performance was observed when the CBAM was removed, with a drop in all metrics, signifying the CBAM's crucial role in the model's effectiveness.

Similarly, the ablation experiments for the brown-spot-detection task revealed the value of each module. The removal of block 3 led to a decrease in all performance metrics, indicating its importance. However, unlike in the leaf blast task, the ablation of blocks 2 and 3 in the brown spot task resulted in more considerable performance degradation, signifying the potentially greater role of the ResViT block in this task. The removal of the CBAM also led to lower performance, underscoring its essential role across tasks.

In both tasks, the highest performance across all metrics was achieved with the full ResViT-Rice model, which suggests that each component of the model contributes significantly to its overall performance. The results from these ablation experiments demonstrate the model's robustness and the importance of its individual blocks as well as the CBAM's attention mechanism.

4. Discussion

4.1. Advantages of ResViT-Rice

In this paper, a hybrid CNN and transformer model called ResViT-Rice is proposed for rice disease detection. The ablation studies reinforce the fact that all parts of the ResViT-Rice model played essential roles in its superior performance in detecting rice diseases. They underline the significance of using a comprehensive and intact model to ensure the most accurate detection for rice diseases. The experimental results of comparing with mainstream models such as ResNet50 and VGG19 demonstrated that our model performed remarkably well in the rice disease detection, achieving a highest accuracy of 99%, which was about 5% higher than other mainstream models, indicating its precise identification of rice diseases.

The ResViT-Rice model achieved its excellent performance for three main reasons:

- ResViT-Rice employed the outstanding ResNet model as the backbone network to extract image features. The use of residual structures allowed for a smoother information transfer, avoiding the problem of gradient vanishing and making the network more stable during training while also reducing the number of parameters [29,30];
- The transformer architecture was introduced to the proposed model. Although many excellent transformer-based works have emerged in the visual domain, such as the Swin transformer [31], there is still a gap in terms of model parameters and inference speed compared to lightweight models based on CNN [32]. Most importantly, training a transformer architecture on images is difficult, as it requires more training data, epochs, and regularization, and the transformer architecture is sensitive to data augmentation [33]. However, we are unwilling to abandon the powerful performance of the self-attention mechanism of the transformer architecture when applied to images. Therefore, we adopted a hybrid architecture of CNN and transformer. CNN can provide spatial inductive bias and accelerate network convergence, making the network training process more stable;
- The CBAM was adopted in the ResViT-Rice block, which can adaptively adjust the importance of different regions in the input feature map [34], thereby increasing the model's attention to important areas. In rice disease-detection tasks, the location of some diseases in the image may be random, so by introducing attention mechanisms, we can make the model pay more attention to disease areas, thereby improving the model's accuracy [35].

4.2. Limitations of Other Mainstream Models

In our study, we found that although other mainstream models such as AlexNet and VGG19 performed well in rice disease-detection tasks, numerous scholars have also employed these two networks to detect rice diseases. However, their accuracy rates largely plateaued around 92% [25,36,37]. On the other hand, several scholars utilized ResNet or networks with residual structures for rice disease detection, generally achieving higher accuracies, primarily in the vicinity of 95% [26,27,38,39]. These scholars' work attested to the superior performance of the residual structure, which is one reason we chose ResNet as our backbone network. Despite these models' commendable performances [39], they still fell short when compared to our proposed ResViT-Rice. This disparity can be ascribed to a couple of critical factors. Firstly, the differences among rice diseases were subtle, particularly between brown spot and leaf blast, which are very similar. Although some neural networks such as ResNet50 and VGG19 have deep layers, the feature-extraction ability of these models is insufficient [40], particularly without the help of the attention mechanism. Judging from Figure 8, it was noted that the decision-making mechanism of each model varied since each model focused on different contributing areas (pixels). The proposed ResViT-Rice can detect most of the symptoms, thereby generating more accurate results. Secondly, the power of the traditional convolution kernel is limited [41]. This motivated us to take advantage of the transformer encoder architecture, where the multi-head attention was integrated. The multi-head attention mechanism enabled the model to capture various representations from subspaces at different positions.

4.3. Limitation of Our Work

Although our research achieved certain success, there are still limitations that need to be further addressed. For instance, deploying well-trained models in edge devices is trending. However, this requires establishing a disease-detection model with fewer trainable parameters and low complexity. Under such a circumstance, using popular lightweight models such as MobileNet, ShuffleNet, and EfficientNet as the backbone would become a preferable approach [42,43]. We examined one of the lightweight neural networks, ShuffleNet, in the comparative evaluation, and the experimental result showed that ShuffleNet achieved the worst performance. Therefore, it is important to achieve a balance between the model size and its performance. Another lightweight approach would be the use of model-compression techniques [44,45]. For example, knowledge distilling, transfer learning, deep compression, network slimming, etc., are well-known methods. On the one hand, guiding "small" neural networks by "large" ones can usually accelerate the training process and avoid the overfitting issue. On the other hand, removing the parameters that have minor contributions would not affect the overall performance. Considering the model size of ResViT-Rice, in the future, we will attempt to narrow down the model size and deploy it to an embedded device for practical applications in the field. In addition, it is worth mentioning that due to the limitations of our dataset, we were unable to recognize and detect different stages of disease progression. In the future, we will collect more sample images to further our understanding and detection of disease-development stages.

5. Conclusions

The aim of this study was to develop an accurate and efficient method for identifying rice diseases. Our work can be summarized into three highlights:

- By integrating the residual module with the encoder from the transformer architecture and introducing attention mechanisms, we proposed an improved deep learning model, ResViT-Rice, and compared it with other mainstream models. The results showed that our model performed the best in all evaluation metrics, with an accuracy of up to 99%;
- Based on the results of our ablation experiments, we had made the significant finding that our ResViT block was essentially an attention mechanism module. It was not only compatible with the ResNet50 network model but can also be combined with various other network models. This suggests that our ResViT block has wide applicability and can be broadly applied in various scenarios. By incorporating the ResViT block into other network models, we can further enhance their performance, thereby boosting their expressiveness. Therefore, our research is not limited to ResNet50 and can be extended to other network structures, providing more competitive solutions for different fields and tasks;
- Finally, our method is not only limited to rice disease detection but is also applicable to the detection and identification of diseases in other crops. In the future, this method is expected to be widely used in agriculture, contributing to the improvement of agricultural production and economic benefits.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/agriculture13061264/s1, Figure S1: The sample images of the rice disease dataset before and after data preprocessing and augmentation: (a) Brown Spot, (b) Leaf Blast, (c) Bacterial Blight, and (d) Tungro; Figure S2. The number of data after pre-processing; Table S1: Additional experimental Evaluation Results of AlexNet, ResNet50, ShuffleNet, VGG19, Swin-Transformer, and ResViT-Rice.

Author Contributions: Data curation, Y.Z.; formal analysis, Y.Z. and Y.D.; funding acquisition, Z.Z.; methodology, Y.Z.; project administration, Z.Z.; resources, H.Y.; software, Y.Z.; supervision, Z.Z.; validation, Y.Z. and L.Z.; visualization, Y.D.; writing—original draft, Y.Z. and Z.Z.; writing—review and editing, H.Y. and Z.Z. All authors have read and agreed to the published version of the manuscript.

Funding: The research leading to the presented results has been undertaken within the Startup Foundation of New Professor at Nanjing Agricultural University (Grant No. 106-804005).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The data used in this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Asibi, A.E.; Chai, Q.; Coulter, J.A. Rice Blast: A Disease with Implications for Global Food Security. *Agronomy* **2019**, *9*, 451. [CrossRef]
2. Candole, B.L.; Siebenmorgen, T.J.; Lee, F.N.; Cartwright, R.D. Effect of Rice Blast and Sheath Blight on Physical Properties of Selected Rice Cultivars. *Cereal Chem. J.* **2000**, *77*, 535–540. [CrossRef]
3. Ng, L.C.; Sariah, M.; Sariam, O.; Radziah, O.; Zainal Abidin, M.A. Bio-efficacy of microbial-fortified rice straw compost on rice blast disease severity, growth and yield of aerobic rice. *Australas. Plant Pathol.* **2012**, *41*, 541–549. [CrossRef]
4. Chukwu, S.C.; Rafii, M.Y.; Ramlee, S.I.; Ismail, S.I.; Hasan, M.M.; Oladosu, Y.A.; Magaji, U.G.; Akos, I.; Olalekan, K.K. Bacterial leaf blight resistance in rice: A review of conventional breeding to molecular approach. *Mol. Biol. Rep.* **2019**, *46*, 1519–1532. [CrossRef]
5. Chhabra, R.; Sharma, R.; Hunjan, M.S.; Sharma, V.K.; Sharma, P.; Chauhan, S.K. Microstructural and metabolic variations induced by Bipolaris oryzae inciting brown spot disease of rice. *Cereal Res. Commun.* **2023**. [CrossRef]
6. Aslam, H.M.U.; Naveed, K.; Hussain, S.I.; Shakeel, Q.; Ashraf, W.; Anwaar, H.A.; Raza, M.M.; Sarfraz, S.; Tariq, I. First Report of Brown Leaf Spot of Rice Caused by Bipolaris zeicola in Pakistan. *Plant Dis.* **2021**, *105*, 212. [CrossRef]
7. Nur Ain Izzati, M.Z.; Madihah, M.Z.A.; Nor Azizah, K.; Najihah, A.; Muskhazli, M. First Report of Bipolaris cactivora Causing Brown Leaf Spot in Rice in Malaysia. *Plant Dis.* **2019**, *103*, 1021. [CrossRef]
8. Barnwal, M.K.; Kotasthane, A.S.; Magculia, N.; Mukherjee, P.K.; Savary, S.; Sharma, A.K.; Singh, H.B.; Singh, U.; Sparks, A.H.; Variar, M.; et al. A review on crop losses, epidemiology and disease management of rice brown spot to identify research priorities and knowledge gaps. *Eur. J. Plant Pathol.* **2013**, *136*, 443–457. [CrossRef]
9. Mamat, N.; Othman, M.F.; Abdulghafor, R.; Alwan, A.A.; Gulzar, Y. Enhancing Image Annotation Technique of Fruit Classification Using a Deep Learning Approach. *Sustainability* **2023**, *15*, 901. [CrossRef]
10. Gulzar, Y. Fruit Image Classification Model Based on MobileNetV2 with Deep Transfer Learning Technique. *Sustainability* **2023**, *15*, 1906.
11. Wang, Y.; Wang, H.; Peng, Z. Rice diseases detection and classification using attention based neural network and bayesian optimization. *Expert Syst. Appl.* **2021**, *178*, 114770. [CrossRef]
12. Daniya, T.; Vigneshwari, S. Deep Neural Network for Disease Detection in Rice Plant Using the Texture and Deep Features. *Comput. J.* **2021**, *65*, 1812–1825. [CrossRef]
13. Kim, D.; Jeong, S.; Kim, B.; Kim, S.-j.; Kim, H.; Jeong, S.; Yun, G.-y.; Kim, K.-Y.; Park, K. Automated Detection of Rice Bakanae Disease via Drone Imagery. *Sensors* **2023**, *23*, 32. [CrossRef]
14. Haque, M.E.; Rahman, A.; Junaeid, I.; Hoque, S.U.; Paul, M. Rice Leaf Disease Classification and Detection Using YOLOv5. *arXiv* **2022**, arXiv:2209.01579.
15. Gao, Y.; Wang, H.; Li, M.; Su, W.-H. Automatic Tandem Dual BlendMask Networks for Severity Assessment of Wheat Fusarium Head Blight. *Agriculture* **2022**, *12*, 1493. [CrossRef]
16. Liu, B.-Y.; Fan, K.-J.; Su, W.-H.; Peng, Y. Two-Stage Convolutional Neural Networks for Diagnosing the Severity of Alternaria Leaf Blotch Disease of the Apple Tree. *Remote Sens.* **2022**, *14*, 2519. [CrossRef]
17. Wang, Y.-H.; Su, W.-H. Convolutional Neural Networks in Computer Vision for Grain Crop Phenotyping: A Review. *Agronomy* **2022**, *12*, 2659. [CrossRef]
18. Su, W.-H.; Zhang, J.; Yang, C.; Page, R.; Szinyei, T.; Hirsch, C.D.; Steffenson, B.J. Automatic Evaluation of Wheat Resistance to Fusarium Head Blight Using Dual Mask-RCNN Deep Learning Frameworks in Computer Vision. *Remote Sens.* **2021**, *13*, 26. [CrossRef]
19. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
20. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
21. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6848–6856.
22. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2012**, *60*, 84–90. [CrossRef]
23. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, BC, Canada, 10–17 October 2021; pp. 9992–10002.
24. Sethy, P.K.; Barpanda, N.K.; Rath, A.K.; Behera, S.K. Deep feature based rice leaf disease identification using support vector machine. *Comput. Electron. Agric.* **2020**, *175*, 105527. [CrossRef]
25. Ghosal, S.; Sarkar, K. Rice Leaf Diseases Classification Using CNN With Transfer Learning. In Proceedings of the 2020 IEEE Calcutta Conference (CALCON), Salt Lake City, UT, USA, 28–29 February 2020; pp. 230–236.
26. Islam, M.A.; Shuvo, M.N.R.; Shamsojjaman, M.; Hasan, S.; Shahadat, M.A.; Khatun, T. An Automated Convolutional Neural Network Based Approach for Paddy Leaf Disease Detection. *Int. J. Adv. Comput. Sci. Appl.* **2021**, *12*. [CrossRef]
27. Patidar, S.; Pandey, A.; Shirish, B.A.; Sriram, A. Rice Plant Disease Detection and Classification Using Deep Residual Learning. In Proceedings of the International Conference on Machine Learning, Vienna, Austria, 12–18 July 2020.

28. Zhou, Y.; Fu, C.; Zhai, Y.; Li, J.; Jin, Z.; Xu, Y. Identification of Rice Leaf Disease Using Improved ShuffleNet V2. *Comput. Mater. Contin.* **2023**, *75*, 4501–4517. [CrossRef]
29. He, F.X.; Liu, T.L.; Tao, D.C. Why ResNet Works? Residuals Generalize. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *31*, 5349–5362. [CrossRef] [PubMed]
30. Allen-Zhu, Z.; Li, Y. What can ResNet learn efficiently, going beyond kernels? In Proceedings of the 33rd International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019; Curran Associates Inc.: Red Hook, NY, USA, 2019; p. 809.
31. Liu, Z.; Hu, H.; Lin, Y.; Yao, Z.; Xie, Z.; Wei, Y.; Ning, J.; Cao, Y.; Zhang, Z.; Dong, L.; et al. Swin Transformer V2: Scaling Up Capacity and Resolution. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 11999–12009.
32. Touvron, H.; Cord, M.; Sablayrolles, A.; Synnaeve, G.; J'egou, H.e. Going deeper with Image Transformers. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, BC, Canada, 10–17 October 2021; pp. 32–42.
33. Graham, B.; El-Nouby, A.; Touvron, H.; Stock, P.; Joulin, A.; Jégou, H.; Douze, M. LeViT: A Vision Transformer in ConvNet's Clothing for Faster Inference. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, BC, Canada, 10–17 October 2021; pp. 12239–12249.
34. Yang, X. An Overview of the Attention Mechanisms in Computer Vision. *J. Phys. Conf. Ser.* **2020**, *1693*, 012173. [CrossRef]
35. Peng, J.; Wang, Y.; Jiang, P.; Zhang, R.; Chen, H. RiceDRA-Net: Precise Identification of Rice Leaf Diseases with Complex Backgrounds Using a Res-Attention Mechanism. *Appl. Sci.* **2023**, *13*, 4928. [CrossRef]
36. Yakkundimath, R.; Saunshi, G.; Anami, B.; Palaiah, S. Classification of Rice Diseases using Convolutional Neural Network Models. *J. Inst. Eng. (India) Ser. B* **2022**, *103*, 1047–1059. [CrossRef]
37. Prasetyo, H.D.; Triatmoko, H.; Nurdiansyah; Isnainiyah, I.N. The Implementation of CNN on Website-based Rice Plant Disease Detection. In Proceedings of the 2020 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS), Jakarta, Indonesia, 19–20 November 2020; pp. 75–80.
38. Ajra, H.; Nahar, M.K.; Sarkar, L.; Islam, M.S. Disease Detection of Plant Leaf using Image Processing and CNN with Preventive Measures. In Proceedings of the 2020 Emerging Technology in Computing, Communication and Electronics (ETCCE), Dhaka, Bangladesh, 21–22 December 2020; pp. 1–6.
39. Acharya, A.; Muvvala, A.; Gawali, S.; Dhopavkar, R.; Kadam, R.; Harsola, A. Plant Disease detection for paddy crop using Ensemble of CNNs. In Proceedings of the 2020 IEEE International Conference for Innovation in Technology (INOCON), Bangaluru, India, 6–8 November 2020; pp. 1–6.
40. Zhao, X.; Huang, P.; Shu, X. Wavelet-Attention CNN for image classification. *Multimed. Syst.* **2022**, *28*, 915–924. [CrossRef]
41. Wu, H.; Xiao, B.; Codella, N.; Liu, M.; Dai, X.; Yuan, L.; Zhang, L. CvT: Introducing Convolutions to Vision Transformers. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, BC, Canada, 10–17 October 2021; pp. 22–31.
42. Kumar, P.R.; Kiran, R.; Singh, U.P.; Rathore, Y.; Janghel, R.R. Rice Leaf Disease Detection using Mobile Net and Inception V.3. In Proceedings of the 2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT), Indore, India, 23–24 April 2022; pp. 282–286.
43. Masykur, F.; Adi, K.; Nurhayati, O.D. Classification of Paddy Leaf Disease Using MobileNet Model. In Proceedings of the 2022 IEEE 8th International Conference on Computing, Engineering and Design (ICCED), Virtual, 28–29 July 2022; pp. 1–4.
44. Chavan, A.; Shen, Z.; Liu, Z.; Liu, Z.; Cheng, K.T.; Xing, E. Vision Transformer Slimming: Multi-Dimension Searching in Continuous Optimization Space. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 4921–4931.
45. He, J.; Ding, Y.; Zhang, M.; Li, D. Towards efficient network compression via Few-Shot Slimming. *Neural Netw.* **2022**, *147*, 113–125. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.