Supplementary Materials A

In order to demonstrate the generalization ability of our model, we have carried out a series of experiments using a dataset with a complex background and a wider variety of rice disease categories. This dataset is challenging due to its variability and diversity, encompassing more environmental conditions and disease types.

A.1: Dataset Description

The dataset used in this study was obtained from Mendeley Data(https://data.mendeley.com/datasets/fwcj7stb8r/), a renowned open data platform that offers a wide array of public datasets for researchers and data scientists. We procured a comprehensive dataset that includes images of four types of rice diseases: Leaf Blast, Brown Spot, Bacterial Blight, and Tungro, with each category consisting of 1440, 1600, 1584, and 1308 images respectively. This extensive dataset, characterized by its diverse and complex backgrounds and environments, is invaluable in assessing the robustness and adaptability of our deep learning model. The variety and complexity inherent in this dataset ensure a thorough validation of our model's capability to perform under intricate and varied conditions.

The data preprocessing method for this study adheres to the same standards outlined in the main text, so we will not elaborate on it here. Figure S1 provides a display of the dataset samples. We have adjusted the random division of data to an 8:2 ratio, with 80% serving as training data and 20% as validation data, and the training set data has been augmented to four times its original size. Figure S2 depicts the quantity and distribution of this dataset.

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.

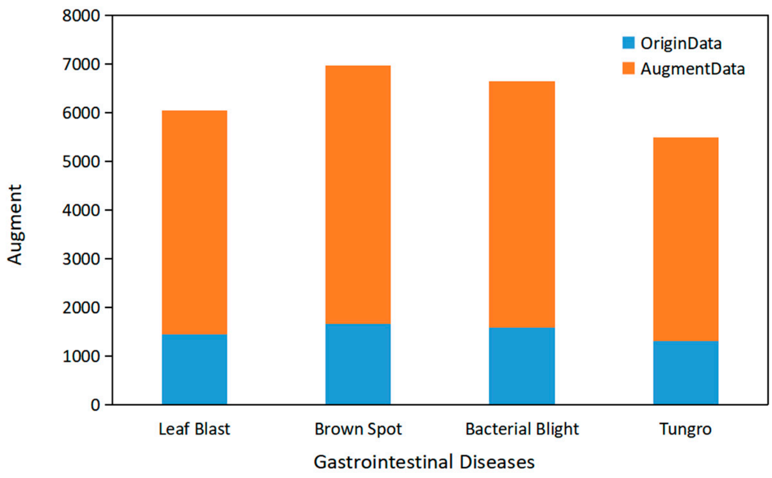

Figure S2. The number of data after pre-processing.

A.2: Model Performance

As shown in Table S1, the proposed model, ResViT-Rice, outperforms all other models, demonstrating the highest accuracy, precision, recall, F1 score, and AUC. Specifically, ResViT-Rice achieved an accuracy of 0.9966, a precision of 0.9973, a recall of 0.9966, an F1 score of 0.9966, and an AUC of 0.9993. These results indicate a very high level of model performance.

While AlexNet, ResNet50, VGG19, ShuffleNet, and Swin Transformer yielded commendable results, it is clear that ResViT-Rice delivered superior performance, demonstrating its strength and reliability.

Of note is the resilience and adaptability of ResViT-Rice. Despite the complexities presented by the various backgrounds of the images and potential imbalances in the dataset distribution, our model maintained a high level of performance. This is a testament to the model's robustness and ability to effectively learn and classify even when presented with challenging conditions.

In conclusion, ResViT-Rice is a powerful model, capable of delivering top-tier results even in the face of complex and challenging data conditions. The model's high degree of accuracy, precision, recall, and F1 score solidifies its potential for effective use in diverse and intricate data environments.

Table S1. Additional experimental Evaluation Results of AlexNet, ResNet50, ShuffleNet, VGG19, Swin-Transformer and ResViT-Rice.

Model	Accuracy	Precision	Recall	F1 Score	AUC
AlexNet [10]	0.9758	0.9762	0.9758	0.9758	0.9912
ResNet50 [7]	0.9803	0.9826	0.9803	0.9797	0.9963
VGG19 [8]	0.9764	0.9764	0.9761	0.9723	0.9921
ShuffleNet [9]	0.9539	0.9576	0.9523	0.9539	0.9856
Swin Transformer [11]	0.9820	0.9823	0.9820	0.9817	0.9975
ResViT-Rice	**0.9966**	**0.9973**	**0.9966**	**0.9966**	**0.9993**

A.3: Conclusion

We hope that these experiments and their results provide a clear demonstration of our model's robustness and generalization capability across different environmental conditions and rice disease types. We believe that these findings attest to the model's potential as a powerful tool in the diagnosis and management of rice diseases.