International Journal of Computer Applications (0975 - 8887)
Volume 187 - No.76, January 2026

Advancing Rice Leaf Disease Detection using Vision Transformer on Real Datasets from Bangladesh

Apurbo Deb Nath, Mohammad Shoaib Rahman, Md. Shahrear Ahmed Shuvon, Boby Rani Das, Nayebul Jannath Chowdhury, Md. Jalal Uddin Chowdhury and Sadia Afrin Rimi
DeepNet Research and Development Lab, Sylhet 3100, Bangladesh

ABSTRACT

Rice leaf diseases pose a significant threat to our food security as they can reduce crop yields, cause plant deaths or even complete destruction in some cases, and result in shortfalls for both farmers and global agricultural production. Typically, farmers and other agricultural experts identify these diseases merely through visual examination, which leads to laboriousness, undesirable subjectivity, and faulty diagnosis. The main aim of this study is to provide farmers with accurate visual information so that they can protect their crops on time. Results: To accurately model a disease identification, we first propose our RiceLeafBD dataset. This dataset has been the subject of several studies, but our approach applies a model to it for the first time. We employed a proposed framework, achieving a superior accuracy of 92.75%. Notably, when assessing the performance on the tungro virus class, the model demonstrated exceptional precision, recall, and F1-score values of 100%, 98%, and 99%, respectively. The proposed framework does better than current convolutional neural network (CNN) and hybrid CNN-transfer learning models, according to the results of experiments. It has the highest accuracy and the least amount of model complexity that has been seen so far.

Keywords

Rice Leaf, Disease, RiceLeafBD Dataset, Proposed Framework, Vision Transformer

1. INTRODUCTION

Agriculture remains one of the top sectors in many developing nations. In these nations, rice serves as a vital crop for millions of people, particularly in Bangladesh and other developing nations [1]. More than a source of food, rice is key to ethnic identity and economic stability. Several diseases, such as brown spot, bacterial leaf blight, and tungro virus, have been responsible for the worst productivity losses in rice crops [2]. Early and accurate detection of these diseases will facilitate timely intervention to protect crop productivity and quality [3]. Advanced tools for automated identification and diagnosis of crop diseases based on image analyses have revolutionized the agriculture sector in the last few years, thanks to machine learning (ML) and deep learning (DL) methods. By providing farmers with precisely timed data, these advancements have the potential to radically transform the operations of traditional agricultural farms [4].

Despite the impressive results, the application of ML and DL in agriculture still faces numerous challenges. The challenge lies in the complexity and variability of agricultural data, which often resembles the characteristics of a farm setting [5]. In that point, A further issue is the inadequate data quantity, which may limit deep network models in their learning and generalization capabilities. Still, insufficient data on rice leaves may hinder the model's ability to accurately represent the many variations and phases of disease expression. Conventional convolutional neural networks (CNNs) have extensively investigated plant disease identification; however, the numerous factors linked to image collection frequently complicate classification problems [6]. Variables such as different image resolutions or complex scenery with distinct lighting conditions will widely affect the accuracy and resilience of these models. This simple correlation still leaves a significant amount of uncertainty and necessitates the development of more sophisticated strategies to effectively handle these types of complications [7]. One such alternative is in the form of Vision Transformers (ViTs) that exploit self-attention [8]. ViTs are better at naturally recognizing intricate structures and correlations in the image data, which can be challenging for CNNs [9].

In previous studies, a number of ML and DL models were investigated for detecting plant diseases[10]. For instance, in the aspect of plant disease classification, transfer learning models such as EfficientNet and even InceptionNet have been able to achieve high precision. Still, the generalization of these models across different datasets is somewhat limited. One of the public datasets is the Plant Village dataset [11] which was a complex dataset with a skewed class problem. Researchers have improved the image feature extraction capabilities of ViT through the introduction of more efficient Transformer variants, including Convolutional Visual Transformer (CVIT) and Pyramid Layered Networks (PDN). Recent advancements in attention mechanisms [12], [13], [14] have been developed to thoroughly analyze image data, thereby enhancing target recognition accuracy. Recent years have seen a growing focus among scholars on the lightweighting research of Vision Transformer (ViT). One of the most notable efficient transformer models, which has garnered considerable attention, utilizes the shifted windows attention mechanism [15], facilitating

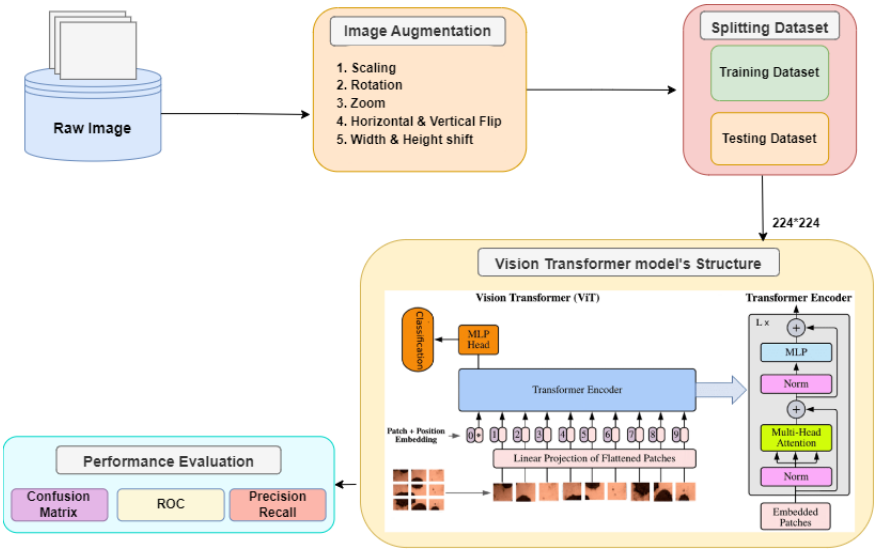

Fig. 1. Overall workflow of the proposed methodology

efficient processing of large-scale images via hierarchical attention [16]. In this situation, the model's light weight is one of the most important things in our study, especially when it comes to finding rice diseases. In addition, existing methods often do not account for the practical constraints faced by farmers in real settings, such as poor image quality and different device capabilities.

In this study, we use the Vision Transformer (ViT) model to offer a distinctive approach to identifying diseases in rice leaves. We conducted our study with the RiceLeafBD [27] dataset, which contains 1,555 rice leaf images captured from multiple environments across the Bangladesh growing season. To demonstrate the viability of the ViT model, we compare it with traditional transfer learning models like EfficientNet and InceptionNet. We have developed, for instance, a robust field disease detection system designed to fit the true agricultural environments. This diagnostic process is beneficial for Bangladeshi farmers because it allows for early and accurate disease identification. This technology can assist in encouraging sustainable farming practices, reducing crop losses, and improving farmer livelihoods. This study makes the following key contributions:

(1) A thorough comparative analysis was conducted between Transformer models and Transfer learning models for classifying rice leaf diseases, highlighting the strengths and limitations of each.

(2) We addressed the challenge of data by collecting real-world rice leaf datasets, ensuring that our deep learning models are trained on authentic, field-based data, which we proposed named "RiceLeafBD".

(3) We suggest the best-performing model and propose a user-friendly interface that accurately identifies rice leaf diseases, which is crucial for Bangladeshi farmers to protect their crops and improve agricultural productivity.

The remaining parts of the paper are organized as follows: Section II describes the detailed methodology, whereas Section III presented the experimental results. Finally, Section IV and V covers the discussion of outcomes and the conclusion of the study.

2. PROPOSED METHODOLOGY

The proposed methodology outlines the systematic framework for achieving accurate and reliable results in this study, encompassing data acquisition, preprocessing, model development, and evaluation procedures, as shown in Figure 1.

2.1 Data Acquisition

In this research, the dataset used was a derivative of the "RiceLeafBD,"[27] which consists on high-resolution images of rice leaves with different diseases. They feature some images shot in the rice fields of Bangladesh. The dataset consisted of data samples on various rice leaf states (figure 2), such as healthy leaves and those infected with different diseases.

2.2 Data Preprocessing

Effective data preparation is essential for improving the performance of deep learning models. The following procedures were taken to prepare the "RiceLeafBD" [27] dataset for the Vision Transformer:

2.2.1 Image Augmentation. Data augmentation refers to the process of increasing the size of a dataset by applying various transformations or modifications to the existing data, to improve the performance and generalization of machine learning models [18]. By using several data augmentations, we have enhanced the variety of the dataset and further improved the generalizability of our model [19]. The images have undergone extensive pre-processing using the ImageDataGenerator class, which includes rescaling to 224x224, rotations, zooms, horizontal and vertical flips, and shifts. These improvements facilitated the replication of different real-world scenarios, hence enhancing the model.

2.2.2 Train-Test Split. For making the model evaluation reliable, the dataset is split into training and testing sets with an 80:20 ratio.

Fig. 2. Sample images of datasets.

Table 1. Performance for ViT model

Leaf Symptoms	Precision	Recall	F1-Score
Healthy	0.86	0.95	0.90
Bacterial leaf blight	0.94	0.89	0.91
Brown Spot	0.91	0.84	0.87
Tungro virus	1.00	0.98	0.99

This means that I have trained this model on 80% of the images and tested it for performance to the rest of 20%. This split ensures that most of the data is used for training and a separate test set should be made available to learn from, in order to verify performance unbiasedly [20].

2.2.3 Label Encoding. The category labels which correspond to different diseases, and healthy leaves were encoded using LabelEncoder of the sci-kit-learn module. These encoded labels were then one-hot encoded as the model needs this information for training so that it knows which class of rice leaf condition to correctly predict[21].

2.2.4 Dataset Creation. TFRecord files and num_records were used for the training/test fold that was released plus the original flow_from_dataframe function from Keras on TensorFlow to create datasets for model training as well. For the training dataset, the images were further included that might provide various courses of action based on the manner in which rice leaves are available. However, the validation dataset was left as is so that we could properly evaluate how well our model would perform on real-world data.

2.3 Vision Transformer

The Vision Transformer (ViT) employs transformer-based architectures, which is a radical departure from the conventional convolutional neural networks (CNNs) [22]. For processing the incoming images, ViT segments them into fixed-size patches and forwards these through transformer layers. This allows the model to efficiently obtain global context data. ViT has no hierarchical feature extraction as in CNN, nonetheless, the self-attention mechanism of ViT is capable of learning long-range relationships and this makes it very successful for image classification [23]. The current study aims to evaluate the performance of ViT for disease detection on rice leaves, in comparison with other cutting-edge models and highlights its advantages as well as contribution.

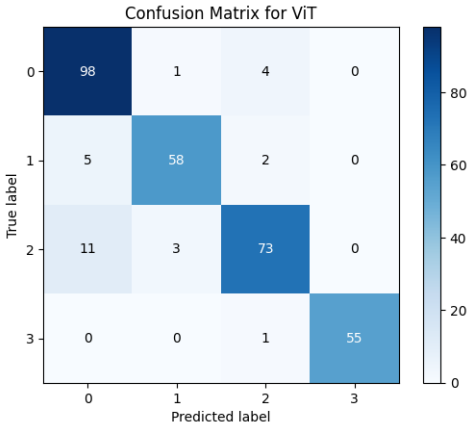

Fig. 3. Condusion Matrix of the ViT model

2.4 Model Configuration

The Vision Transformer (ViT) model has been set-up to achieve state-of-the-art performance in rice leaf disease detection. For this study, our workflow diagram are presented in Figure-1. They configure the ViT model with twelve transformer layers, each having twelve attention heads and a 768 hidden size. After resizing the input photos to 224 by 224 pixels, each image was split up into 16 by 16 patches, for a total of 196 patches per image. To preserve spatial information, positional embeddings were added after these patches were linearly embedded into vectors of size 768. A learning rate scheduler lowered the learning rate by a factor of 0.1 if the validation loss plateaued for more than three epochs while using the Adam optimizer, which had an initial learning rate of 0.001. The categorical cross-entropy loss function was used to train the model throughout 50 epochs with a batch size of 32. Dropout layers with a rate of 0.1 were included, and early halting with a patience of 5 epochs was used to prevent overfitting. To ensure a thorough evaluation of the model's classification skills, accuracy, and categorical cross-entropy loss were used to measure model performance.

3. RESULT ANALYSIS AND DISCUSSION

This section elaborates on the performance evaluation of the proposed model, analyzing quantitative metrics such as accuracy, precision, recall, and F1-score, followed by a critical discussion of observed trends and implications.

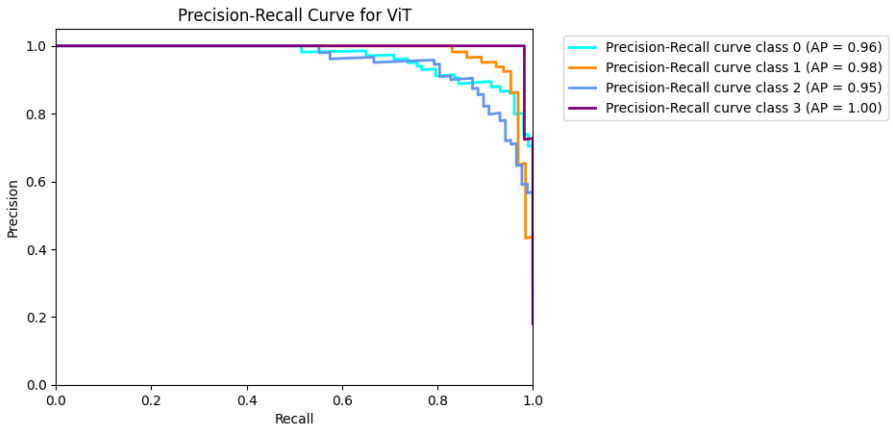

Fig. 4. Precision-Recall of the ViT model

3.1 Performance Measures

In this section, we perform a comprehensive performance evaluation with the RiceLeafBD for validating the Vision Transformer (ViT) model through experiments on three split datasets, including train, test and validation. These experiments were meticulously designed to evaluate the effectiveness of the ViT model in identifying rice leaf diseases. We assessed the model's performance through quantitative metrics, statistical analyses, and comparisons with state-of-the-art models.

The results in the tables show that our ViT method achieves Get to accuracy, precision, and F1-score is better than others, which means the proper capacity to classify disease categories correctly. We conducted similar performance evaluations using a pre-trained model on the RiceLeafBD agricultural datasets. As shown in Table 1, the recall, precision, and F1-score for each class in the RiceLeafBD dataset were impressive, with average values of 92.75%, 91.5%, and 91.75%, respectively.

Figure 2 illustrates the confusion matrices of the ViT model across the RiceLeafBD datasets. A higher number of misclassifications was observed in the "Brown Spot" category, which can be attributed to the visual similarity between early-stage diseases and healthy leaf images, leading to classification errors. Analysis of the confusion matrix reveals that the ViT model exhibits minimal false positives and false negatives. The misclassified samples primarily occur when leaves affected by different diseases exhibit similar visual characteristics. As depicted in Figures 3, even experienced farmers could struggle with accurately identifying these cases, underscoring the challenge presented by visually similar disease manifestations.

Figure 4 also shows the Precision-Recall curves for the ViT model across the RiceLeafBD datasets. The model accurately identified the tungro virus, achieving a perfect performance score of 100%. Following this, the model also performed well in detecting Bacterial Leaf Blight, healthy leaves, and Brown Spot, with precision values of 98%, 96%, and 95%, respectively. A comparative analysis of the precision-recall values across all classes reveals that the model consistently maintained high performance across the board.

Table 2. Comparision with other work using same dataset

Study	Datasets	Used models	Accuracy
Rimi et al.[17]	RiceLeafBD [27]	InceptionNet-V2, MobileNet-V2, and EfficientNet-V2	85%, 89.75%, and 91.50%
Proposed model	RiceLeafBD	ViT	92.75%

The performance of any model is influenced by various evaluation metrics, one of which is the ROC curve, which plots the true positive rate (TPR) against the false positive rate (FPR) at different threshold settings. Figure 5 depicts the ROC curve, showing that the ViT model performed exceptionally well across all classes. Notably, the model achieved a 99% accuracy in classifying the tungro virus, followed by 94% accuracy for both Bacterial Leaf Blight and healthy leaves. This analysis confirms that the ViT model offers significant performance advantages in rice leaf disease identification using the RiceLeafBD datasets.

3.2 Comparison with the same datasets for other work

Previous work on the RiceLeafBD datasets, such as that by Rimi et al. [17], utilized deep learning and transfer learning models to develop a system for early disease detection. Model accuracy is a critical factor in disease identification, as higher accuracy indicates a more reliable detection capability. Rimi et al. achieved notable performance with EfficientNet-V2, reaching 91.5% accuracy, followed by MobileNet-V2 at 89.75%, and InceptionNet-V2 at 85%, as illustrated in Table 2. In our study, the ViT model achieved an accuracy of 92.75% in correctly predicting rice leaf diseases, surpassing the models used by Rimi et al. [17]. This demonstrates that our model provides farmers with a more accurate and reliable tool for disease detection.

3.3 Comparison with the different datasets for other work

In this section, we compare our study with existing works that used different datasets for the transformer-based model. Han et al.[56] detect ligneous leaf diseases using Vision Transformer and Trans-

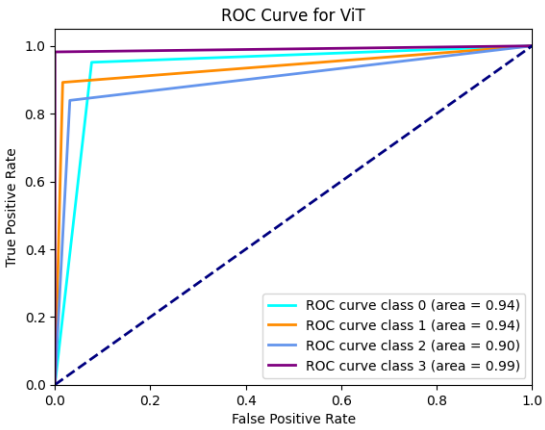

Fig. 5. ROC curve of the ViT model

Table 3. Comparision with other work using different dataset

Study	Datasets	Used models	Accuracy
Azim et al.[24]	Rice Leaf Dataset	XGBoost	86.58%
Mehnaz et al.[25]	Dhan-Shomadhan Dataset	ResNet50, InceptionNet-V3	91.50%, 87.50
Chakrabarty et al.[26]	Plan Village Dataset	ViT	90.33%
Proposed model	RiceLeafBD	ViT	92.75%

fer Learning where he achieved different performance and specially they achieved highest accuracy for Vision Transformer model. Table 3 demonstrates about transfer learning models and ViT model performance with our study.

4. DISCUSSION

The Vision Transformer (ViT) achieved 92.75% accuracy in classifying rice leaf diseases from the RiceLeafBD dataset, outperforming established convolutional and transfer learning models. The model delivered high precision, recall, and F1-scores across most classes, with perfect precision for the Tungro virus. Misclassifications were mainly between early-stage Brown Spot and healthy leaves due to close visual resemblance.

The ViT's self-attention mechanism captured fine details and long-range relationships, maintaining performance under varied lighting, resolution, and background conditions. It's relatively low model complexity makes it practical for use in field conditions where computing resources are limited. Comparison with previous studies on the same dataset confirmed improved accuracy, indicating strong potential for real-world agricultural use. Future work should extend the dataset with samples from different regions, seasons, and growth stages to improve generalization. Multi-label classification could enable the detection of multiple diseases on the same leaf, while object localization would allow the precise identification of affected areas. Optimized lightweight models can support deployment on mobile devices, and edge-based systems could deliver real-time feedback to farmers. Extending the approach to other crops and adding interpretability features would further enhance its usefulness and adoption.

5. CONCLUSIONS

Effectively identifying and managing rice plant diseases is crucial for modern agriculture. To ensure early disease prediction and on-time intervention for researchers & farmers, this research investigates image processing (IP), machine learning (ML), as well as deep learning concepts. We introduced a streamlined version of the novel Vision Transformer (ViT) model and compared it with convolutional-based architectures of similar complexity. The ViT-based model surpasses existing convolutional and transfer learning models in terms of accuracy, achieving a remarkable 92.75% accuracy on the RiceLeafBD datasets. An analysis of the confusion matrix revealed that the ViT model correctly classified nearly all categories, with particularly reliable and accurate identification of the tungro virus class. The precision, recall, and F1 scores for this class were 100%, 98%, and 99%, respectively. This performance underscores the effectiveness of the ViT model's architecture, as well as its lightweight nature, making it a promising candidate for real-world applications where computational resources are constrained. Looking ahead, future work will focus on advancing towards an automated disease detection system. This will require the improvement of object localization networks for plant leaf detection and multi-label classification to deal with multiple diseases. We will also further optimize the hyperparameters of the model, which provides us with both performance and a good balance between complexity.

6. REFERENCES

[1] Roy, D., Sarker Dev, D., & Sheheli, S. (2019). Food security in Bangladesh: insight from available literature. Journal of Nutrition and Food Security, 4(1), 66-75.

[2] Conde, S., Catarino, S., Ferreira, S., Temudo, M., & Monteiro, F. (2024). Rice Pests and Diseases Around the World: Who, Where and What Damage Do They Cause?. Rice Science.

[3] Deng, R., Tao, M., Xing, H., Yang, X., Liu, C., Liao, K., & Qi, L. (2021). Automatic diagnosis of rice diseases using deep learning. Frontiers in plant science, 12, 701038.

[4] Kansanga, M., Andersen, P., Kpienbaareh, D., Mason-Renton, S., Atuoye, K., Sano, Y., & Luginaah, I. (2019). Traditional agriculture in transition: Examining the impacts of agricultural modernization on smallholder farming in Ghana under the new Green Revolution. International Journal of Sustainable Development & World Ecology, 26(1), 11-24.

[5] Leroux, C., & Tisseyre, B. (2019). How to measure and report within-field variability: a review of common indicators and their sensitivity. Precision Agriculture, 20(3), 562-590.

[6] Lu, D., & Weng, Q. (2007). A survey of image classification methods and techniques for improving classification performance. International journal of Remote sensing, 28(5), 823-870.

[7] Rawat, W., & Wang, Z. (2017). Deep convolutional neural networks for image classification: A comprehensive review. Neural computation, 29(9), 2352-2449.

[8] Ahmed, S. T., Barua, S., Fahim-Ul-Islam, M., & Chakrabarty, A. (2024, March). Enhancing Precision in Rice Leaf Disease Detection: A Transformer Model Approach with Attention Mapping. In 2024 International Conference on Advances in Computing, Communication, Electrical, and Smart Systems (iCACCESS) (pp. 1-6). IEEE.

[9] Thai, H. T., Tran-Van, N. Y., & Le, K. H. (2021, October). Artificial cognition for early leaf disease detection using vision transformers. In

2021 International Conference on Advanced Technologies for Communications (ATC) (pp. 33-38). IEEE.

[10] Ahmed, I., & Yadav, P. K. (2023). Plant disease detection using machine learning approaches. Expert Systems, 40(5), e13136.

[11] Plant Village. (2020). Plant Village Dataset. Available online: https://www.kaggle.com/emmarex/plantdisease (accessed on 14 July 2024).

[12] H. Wang, Y. Zhu, B. Green, H. Adam, A. Yuille, and L.-C. Chen, "Axial-DeepLab: Stand-alone axial-attention for panoptic segmentation," in Proc. Eur. Conf. Comput. Vis. Springer, 2020, pp. 108–126.

[13] Y. Xu, Z. Zhang, M. Zhang, K. Sheng, K. Li, W. Dong, L. Zhang, C. Xu, and X. Sun, "Evo-ViT: Slow-fast token evolution for dynamic vision transformer," in Proc. AAAI Conf. Artif. Intell., 2022, vol. 36, no. 3, pp. 2964–2972.

[14] M.-H. Guo, C.-Z. Lu, Z.-N. Liu, M.-M. Cheng, and S.-M. Hu, "Visual attention network," Comput. Vis. Media, vol. 9, no. 4, pp. 733–752, Dec. 2023.

[15] X. Pan, T. Ye, Z. Xia, S. Song, and G. Huang, "Slide-transformer: Hierarchical vision transformer with local self-attention," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2023, pp. 2082–2091.

[16] Y. Fang, X.Wang, R.Wu, andW. Liu, "What makes for hierarchical vision transformer?" IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 10, pp. 12714–12720, Oct. 2023.

[17] Rimi, S. A., Chowdhury, M. J. U., Abdullah, R., Ahmed, I., Mim, M. A., & Rahman, M. S. (2025). Empowering Agricultural Insights: RiceLeafBD–A Novel Dataset and Optimal Model Selection for Rice Leaf Disease Diagnosis through Transfer Learning Technique. arXiv preprint arXiv:2501.08912.

[18] Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. Journal of big data, 6(1), 1-48.

[19] Krois, J., Garcia Cantu, A., Chaurasia, A., Patil, R., Chaudhari, P. K., Gaudin, R., & Schwendicke, F. (2021). Generalizability of deep learning models for dental image analysis. Scientific reports, 11(1), 6102.

[20] Rácz, A., Bajusz, D., & Héberger, K. (2021). Effect of dataset size and train/test split ratios in QSAR/QSPR multiclass classification. Molecules, 26(4), 1111.

[21] Jia, B. B., & Zhang, M. L. (2021). Multi-dimensional classification via decomposed label encoding. IEEE Transactions on Knowledge and Data Engineering, 35(2), 1844-1856.

[22] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 10012-10022).

[23] Liu, H., Shi, S., & Ma, T. (2024, August). Rock lithology classification algorithm based on improved self-attention mechanism VIT. In Journal of Physics: Conference Series (Vol. 2816, No. 1, p. 012049). IOP Publishing.

[24] Azim, M. A., Islam, M. K., Rahman, M. M., & Jahan, F. (2021). An effective feature extraction method for rice leaf disease classification. TELKOMNIKA (Telecommunication Computing Electronics and Control), 19(2), 463-470.

[25] Mehnaz, S., & Islam, M. T. (2025). Rice Leaf Disease Detection: A Comparative Study Between CNN, Transformer and Non-neural Network Architectures. arXiv preprint arXiv:2501.06740.

[26] Chakrabarty, A., Ahmed, S. T., Islam, M. F. U., Aziz, S. M., & Maidin, S. S. (2024). An interpretable fusion model integrating lightweight CNN and transformer architectures for rice leaf disease identification. Ecological Informatics, 82, 102718.

[27] Rimi, Sadia Afrin; Chowdhury, Md Jalal Uddin (2025), "RiceLeafBD: A Real-Field Image Dataset for Rice Leaf Disease Detection and Classification in Bangladesh", Mendeley Data, V1, doi: 10.17632/kx9rx8p2mz.1