International Journal of Electrical and Computer Engineering (IJECE)
Vol. 16, No. 3, June 2026, pp. 1307~1318
ISSN: 2088-8708, DOI: 10.11591/ijece.v16i3.pp1307-1318

GAN-augmented vision transformer with balanced synthetic data generation for robust rice leaf disease detection

Saiful Islam[1,2], Md. Nasim Akhtar[2], M. Mahadi Hassan[2], A. N. M. Rezaul Karim[2], Israt Binteh Habib[2]
[1]Department of Computer Science and Engineering, Dhaka University of Engineering and Technology, Gazipur, Bangladesh
[2]Department of Computer Science and Engineering, International Islamic University of Chittagong, Chittagong, Bangladesh

Article Info

Article history:

Received Nov 7, 2025
Revised Feb 10, 2026
Accepted Mar 16, 2026

Keywords:

Data augmentation agriculture
Deep convolutional generative adversarial network
Generative adversarial networks
Rice leaf disease detection
Vision transformer

ABSTRACT

Early and accurate identification of rice leaf diseases is essential for sustainable crop management; however, many existing convolutional neural networks (CNNs) based solutions struggle with class imbalance and limited robustness when applied to real-field data. In this work, a generative adversarial network (GAN) augmented vision transformer (ViT) framework is introduced to overcome these limitations. A deep size representative samples for underrepresented disease categories, resulting in a more balanced training dataset and achieving a Fréchet inception distance (FID) score of 18.6. The balanced dataset is then used to train a vision transformer model that leverages self-attention to capture global contextual features of rice leaf images. Experimental evaluation across ten disease classes shows that the proposed approach attains an overall classification accuracy of 96.5%, exceeding the performance of several established CNN architectures. Additionally, the model demonstrates strong generalization capability on an external field dataset, achieving 94.8% accuracy. To validate real-world applicability, the trained model is deployed on a Jetson Nano edge device, where it delivers efficient inference performance suitable for practical agricultural applications. The findings indicate that combining GAN-based data augmentation with transformer-based learning provides a reliable and scalable solution for rice leaf disease detection.

Corresponding Author:

Saiful Islam
Department of Computer Science and Engineering, International Islamic University of Chittagong
Chittagong, Bangladesh
Email: saifulcse@iiuc.ac.bd

1. INTRODUCTION

For approximately fifty percent of people worldwide, rice serves as their primary nutritional energy source and the principal food source for over 3.5 billion people in Asia, Latin America, and portions of Africa, making rice one of the major staple foods consumed around the world [1], [2]. Although rice plays an important role, various leaf diseases pose serious problems for rice production. These diseases can lower yields by as much as 70% and damage grain quality [3]. Disease diagnosis has historically depended on farmers' or agronomists' manual visual investigation, which costs money, takes a long time, and is prone to human mistakes, especially for smallholder farmers who have limited access to professional assistance [4]. These restrictions frequently result in incorrect diagnoses, postponed treatments, and the overuse of pesticides, all of which worsen agricultural losses and environmental damage. Recent developments in deep learning (DL) and artificial intelligence (AI) have shown promise in automating the diagnosis of plant diseases. Convolutional neural networks (CNNs) have demonstrated efficacy in detecting rice diseases from images of leaves; however, models such as VGG16 or ResNet frequently suffer from overfitting on small

datasets, struggle to capture long-range spatial features, and demand significant computational resources. Generative adversarial networks (GANs) have been designed to produce artificially created images in order to balance datasets and enhance model generalization [5]. To overcome these limitations, vision transformers (ViTs) have garnered attention because of their ability to identify global contextual linkages and long-range dependencies in images, outperforming CNNs in certain tasks [6]. Furthermore, ViTs offer multi-task learning, which enables the classification of disease types, and this feature can benefit precision farming. This research proposes a novel framework for the detection and classification of rice leaf diseases based on vision transformers. The ViT model is contrasted with conventional CNN architectures, and GAN-based augmentation is incorporated to overcome class imbalance. The goal of this comparative analysis is to provide an automated rice disease management solution that is more precise, effective, and scalable. This research makes the following contributions: i) A novel GAN-based (ViT framework is proposed for rice leaf disease detection (RLDD), where GAN-generated synthetic images address class imbalance, reduce anomalies during oversampling, and prevent redundant data generation and ii) A reproducible training pipeline is presented, including explicit ViT and deep convolutional generative adversarial network (DCGAN) hyperparameters, dataset split strategy, and optimization settings.

The remainder of this paper is organized as follows: section 2 reviews related work, section 3 describes the proposed methodology, section 4 presents experimental results and discussion, and section 5 concludes the paper with future research directions.

2. LITERATURE REVIEW

For the purpose of identifying simultaneous occurrence of leaf diseases in rice and wheat, Jiang *et al.* [7] suggested a multitask deep transfer learning model built on an enhanced VGG16 architecture. Using ImageNet-pretrained weights with fine-tuning and data augmentation, the model demonstrated an accuracy of 97.22% for rice and 98.75% for wheat, outperforming single-task models, ResNet50, and DenseNet121. A real-time plant disease dataset was created by Joseph *et al.* [8], who also suggested a deep learning (DL) based detection technique. They gathered and annotated a sizable collection of disease photos in authentic field settings with the aim of overcoming the drawbacks associated with the current datasets. Their research, which used convolutional neural networks (CNNs) for classification, showed resilience in real-time applications and accomplished excellent accuracy in diagnosing a variety of plant diseases. Bijoy *et al.* [9] presented a lightweight deep CNN for rice leaf disease detection (RLDD), achieving 99.81% accuracy with only 0.18M parameters. To facilitate practical agricultural application, they continued to create a crop health monitoring system with an open application programming interface (API), web interface, and Android application. Roy and Kukreja *et al.* [10] presented a vision transformer (ViT)-based model for rice leaf disease identification and severity estimation. Ayyappan *et al.* [11] applied CNNs for rice disease detection, comparing DenseNet121, EfficientNetB4, Xception, and MobileNetV3 on a dataset of 10,407 images across 10 classes. The study outperformed earlier models like VGG19 and ResNet101 by using a larger dataset and advanced architecture. Study [12] presents an optimized lightweight neural model for RLDD, addressing the challenge of limited computational resources in agricultural settings. The authors compared CNN-based models on a dataset of rice leaf images and found their proposed lightweight model.

A DL-driven approach for recognizing and categorizing five important rice diseases was created by Haridasan *et al.* [13]. Srinivas *et al.* [14] proposed an improved CNN framework for RLDD, resolving problems associated with uneven lighting, intricate background patterns, and similarity in disease symptoms. The study used an enhanced CNN model with optimized feature extraction and classification, evaluated on multiple rice disease datasets. An agricultural Internet of Things (IoT) system using a hybridized deep learning technique for rice crop disease diagnosis was presented by Wang *et al.* [15]. Their technique leverages IoT-based data gathering for real-time monitoring by employing CNNs to capture discriminative features and support vector machines (SVMs) for classification. Images of rice leaves were used to test the system, and it showed excellent accuracy in identifying and categorizing common rice diseases. A lightweight ResNet-9 classifier and an improved vision transformer (IVT) are used in a hierarchical framework for recognizing diseases in the foliage of plants by Vallabhajosyula *et al.* [16]. The model effectively captures both global and local features with reduced computational cost and achieved up to 99.7% accuracy on PlantVillage datasets, outperforming InceptionV3, MobileNetV2, and ResNet50. A convolutional neural network-based image classification model combined with a chatbot interface is used in the automatic rice disease detection system described by Temniranrat *et al.* [17]. The device can identify common rice illnesses including leaf smut and gives farmers easily accessible real-time feedback. A twin CNN framework for rice leaf disease identification was proposed by Pai *et al.* [18]. It combines multi-scale features from VGG16, ResNet50, and InceptionV3 with principal component analysis (PCA) for dimensionality reduction. Radhakrishnan *et al.* [19] introduced an enhanced machine learning (ML)

framework for detecting rice blast disease in paddy crops, leveraging the PlantVillage dataset comprising approximately a set of 60,000 rice leaf images covering both healthy and infected cases. This approach addresses limitations in traditional diagnostic methods, offering potential benefits for early disease detection in agriculture.

A lightweight quantized CNN called RiceLeafClassifier-v1.0 was proposed by Martins *et al.* [20] to identify five different types of rice leaf conditions. The model was created from scratch using data augmentation and optimization techniques to improve generalization, and it was trained using 2,807 photos. The paper emphasizes how it could be used in precision agriculture for low-cost, real-time rice disease monitoring. RDRM-YOLO, a lightweight YOLOv5-based model for rice disease detection in challenging field circumstances, was created by Li *et al.* [21]. Using Hor-BNFA, SPD-Conv, GsConv, and weighted intersection over union (WIoU) loss, it outperformed YOLOv6–v8 and Faster R-CNN, achieving 94.3% precision, 89.6% recall, and 93.5% mean average precision (mAP) on 5,930 images. Using photos gathered from several parts of India, Kumar *et al.* [22] created a CNN-based model for the automatic diagnosis of rice leaf diseases. Three convolutional feature extraction layers along with two fully connected layers were part of the architecture. For real-time field application, the study recommends integration with mobile and IoT devices. With an emphasis on Bangladeshi agriculture, Chowdhury *et al.* [23] examined the use of machine learning (ML) and deep learning (DL) for plant leaf disease identification. They draw attention to disease risks that affect food security and gross domestic product (GDP) in crops like rice, tomatoes, potatoes, and bell peppers. The focus of this study is on early identification to reduce the usage of pesticides. Rahman *et al.* [24] proposed a work that presents a DL-based real-time plant diagnosis system for sustainable agriculture. A total of 30,945 photos from eight plant species and 35 disease classes were combined from various sources, including the PlantVillage dataset, and utilized in the study. The findings demonstrate the great potential of deep learning-based systems for better crop management and precise, real-time plant disease diagnostics.

Despite the remarkable progress achieved through various CNN, hybrid, and transformer-based architectures, the reviewed studies reveal several persistent challenges that justify the development of a new model. Based on the limitations identified in existing CNN and transformer-based approaches, this study introduces a GAN-augmented vision transformer framework designed to achieve balanced, accurate, and edge-compatible rice leaf disease detection.

3. METHOD

Figure 1 presents the overall flowchart of the proposed methodology for RLDD. The framework integrates data collection, preprocessing, GAN-based augmentation, model training, and performance evaluation. A dataset of rice leaf images was compiled, annotated, and preprocessed through resizing, normalization, and traditional augmentation techniques to ensure consistency. To address class imbalance, synthetic samples were generated using a GAN. The ViT model was then trained and compared with several CNN-based architectures under identical conditions.

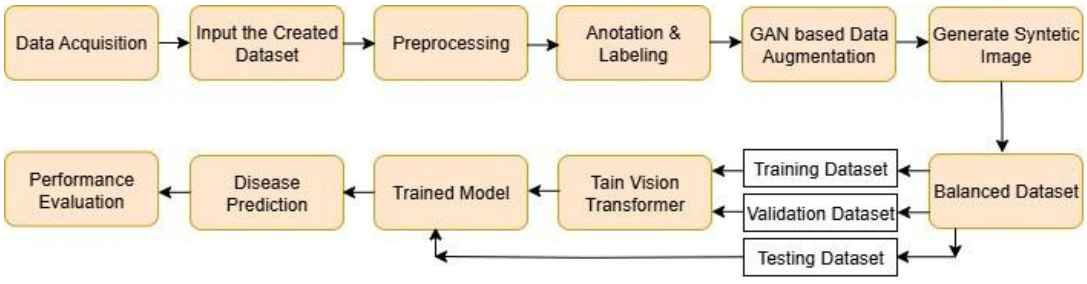

Figure 1. End-to-end pipeline of the proposed GAN-augmented vision transformer framework, including data acquisition, preprocessing, GAN-based synthetic image generation, dataset balancing, ViT training, and performance evaluation

3.1. Dataset and annotation

A total of 13,876 rice leaf images were collected from Kaggle [25]. Each image was carefully annotated with expert consultation to identify one of ten major rice leaf conditions and healthy leaves. To address class imbalance, a deep convolutional generative adversarial network (DCGAN) was employed to generate synthetic images for underrepresented disease categories. The GAN produced 3,676 synthetic samples for underrepresented disease categories. This augmentation resulted in a final dataset of

17,552 images (13,876 real and 3,676 synthetic), thereby improving class distribution and diversity. The quality of generated images was verified through visual inspection by agricultural experts and a Fréchet inception distance (FID) score of 18.6, indicating strong similarity to real samples. The dataset was divided into 70% for training, 15% for validation, and 15% for testing, making sure that all illness types are fairly represented for fair model evaluation.

As shown in Table 1, the original training dataset exhibits substantial class imbalance, with several disease categories having significantly fewer samples. To mitigate this issue, GAN-based augmentation was applied selectively to underrepresented classes. A total of 3,676 synthetic images were generated in proportion to the level of class imbalance, resulting in improved class representation while avoiding excessive over-sampling of majority classes.

Table 1. Class-wise distribution of paddy leaf images before and after GAN-based augmentation

Class name	Original samples	After GAN augmentation
Normal	1764	1764
Blast	1738	1751
Hispa	1594	1680
Dead heart	1442	1606
Tungro	1088	1432
Brown spot	965	1371
Downy mildew	620	1201
Bacterial leaf blight	479	1132
Bacterial leaf streak	380	1083
Bacterial panicle blight	337	1063
Total	10,407	14,083

3.2. Preprocessing

Every image was subjected to a number of preprocessing procedures before training in order to guarantee uniformity and enhance model generalization. To standardize pixel intensity values, each rice leaf image was resized to a fixed spatial resolution of 224×224 pixels and normalized using:

$$I_{norm} = \frac{I-\mu}{\sigma} \quad (1)$$

where I denotes the input image, μ is the mean, and σ is the standard deviation of the dataset. This normalization ensured that all images had zero-centered pixel distributions, improving numerical stability during training. To prevent overfitting and increase dataset diversity, traditional data augmentation techniques were applied. Each image underwent a set of random transformations such as rotation, horizontal and vertical flipping, and scaling:

$$I' = R\theta(I)\ (rotation\ by\ angle\ \theta) \quad (2)$$

$$I'' = Fh(I') or\ Fv(I')\left(\frac{horizontal}{vertical} flip\right) \quad (3)$$

$$I_{aug} = S\alpha(I'')(scaling\ by\ factor\ \alpha) \quad (4)$$

where θ, h, v, and α represent the rotation, flipping, and scaling operators respectively. This preprocessing pipeline not only enhanced the robustness of the ViT model but also ensured reliable performance across diverse disease classes by creating varied and realistic input conditions for model training.

3.3. Training framework and tools

The vision transformer was selected due to its ability to capture long-range spatial dependencies through self-attention, which is critical for distinguishing visually similar rice leaf diseases. DCGAN was chosen over more complex generative models such as StyleGAN or diffusion models because it provides a favorable balance between image quality, training stability, and computational efficiency for agricultural datasets. Images were resized to 224×224 to match standard ViT input resolution while preserving disease-related texture details. The Adam optimizer was adopted for both ViT and DCGAN training due to its fast convergence and robustness in deep vision models.

In order to investigate the efficacy of the suggested ViT model for RLDD, an experimental framework was created. Because of its capacity to capture long-range spatial dependencies, ViT was chosen

as the main architecture. A number of baseline CNN models, including VGG16, ResNet50, MobileNetV2, DenseNet121, DenseNet169, Xception, and ResNet50V2, were employed for comparative study. GANs were used to generate synthetic image for underrepresented classes illness classes for mitigating class distribution disparities and improve generalization, guaranteeing a more robust and balanced training dataset. The ViT model divides each image $x \in \mathrm{R}^{\mathrm{H}\times\mathrm{W}\times\mathrm{C}}$ into N patches of size $P\times P$, where $N = \frac{HW}{P^2}$. Each flattened patch is embedded as:

$$z_0^i = E\left(x_p^i\right) + E_{pos}^i, i = 1, 2, \ldots, N \tag{5}$$

where $E\left(x_p^i\right)$ is a linear projection and E_{pos} represents positional encoding. Within each transformer encoder, the self-attention mechanism computes:

$$Attention(Q, K, V) = softmax\left(\frac{Q\,K^T}{\sqrt{d_k}}\right)V \tag{6}$$

where $Q = z_l W_Q$, $K = z_l W_K$, and $V = z_l W_V$ are the query, key, and value projections. The encoder output is updated as:

$$z_{l+1} = MLP\left(LayerNorm\left(z_l + Attention(z_l)\right)\right) \tag{7}$$

After several layers, the class token representation z^{CLS} is passed to a classification head:

$$\hat{y} = softmax(W_c z_L^{CLS} + b_c) \tag{8}$$

The training of the model utilized categorical cross-entropy loss:

$$L^{cls} = -\frac{1}{M}\sum_{i=1}^{M}\sum_{c=1}^{C} y_{i,c}\, log\left(\hat{y}_{i,c}\right) \tag{9}$$

where $y_{i,c}$ and $\hat{y}_{i,c}$ denote the true and predicted probabilities for class c of sample i, respectively. The Adam optimizer was used to train each model in the same way, with a batch size of 32, a learning rate of 0.0001, and 50 epochs. Accuracy, precision, recall, and F1-score were used to assess the model's performance; confusion matrix analysis was used to provide additional understanding of performance per class. A Deep Convolutional GAN (DCGAN) architecture was employed to generate synthetic rice leaf images for underrepresented disease classes. The GAN framework consists of a generator G and a discriminator D, trained with the employing a minimax optimization criterion:

$$min_G\, max_D V(D, G) = \mathbb{E}_{x\sim p_{(data)(x)}}[\log D(x)] + \mathbb{E}_{z\sim p_{z(z)}}\left[\log\left(1 - D(G(z))\right)\right] \tag{10}$$

where $z \sim p_{z(z)}$ is the input noise vector. The generator and discriminator loss functions are:

$$L_G = -\mathbb{E}_{z\sim p_{z(z)}}[\log D(G(z))] \tag{11}$$

$$L_D = -\mathbb{E}_{x\sim p_{(data)(x)}}[\log D(x)] - \mathbb{E}_{z\sim p_{z(z)}}\left[\log\left(1 - D(G(z))\right)\right] \tag{12}$$

Training of the GAN spanned 200 epochs, employing the Adam optimizer at a learning rate of 0.0002, with batches of 64 samples and binary cross-entropy loss. Both the generator and discriminator were trained simultaneously on 224×224 images. To verify the effect of augmentation, we conducted a per-class accuracy analysis showing an average 4.2% improvement in minority disease categories, confirming that GAN-based augmentation effectively enhanced data diversity and class balance. The detailed configuration and hyperparameters of the Vision Transformer used in this study are presented in Table 2.

The DCGAN architecture consists of a convolutional generator and discriminator. The generator comprises four transposed convolutional layers with batch normalization and ReLU activations, followed by a Tanh output layer. The discriminator consists of four convolutional layers with LeakyReLU activations and a Sigmoid output layer. Both networks were trained using the Adam optimizer with learning rate of 0.0002, β_1= 0.5, and β_2= 0.999. The generator and discriminator were updated alternately with a 1:1 update ratio to ensure stable adversarial training.

Table 2. ViT configuration used in the proposed framework

Parameter	Value
Input image size	224×224
Patch size	16×16
Number of patches	196
Embedding dimension	768
Number of encoder layers	12
Number of attention heads	12
MLP hidden dimension	3072
Dropout rate	0.1
Optimizer	Adam
Learning rate	0.0001
Batch size	32
Number of epochs	50
Loss function	Categorical cross-entropy

3.4. Feature selection

Feature selection in this study was performed automatically by the deep learning models rather than through manual extraction. ViT captured both local details and global contextual relationships using its self-attention mechanism, while CNN baselines primarily focused on localized spatial features. In the ViT architecture, feature extraction is governed by the attention-weighted aggregation of patch embeddings:

$$F_i = \sum_{j=1}^{N} a_{ij} V_j \tag{13}$$

where F_i denotes the i-th output feature vector, V_j is the value embedding of patch j, and α_{ij} represents the attention weight computed as:

$$a_{ij} = \frac{exp\frac{Q_i K_j}{\sqrt{d_k}}}{\sum_{m=1}^{N} \exp\left(\frac{Q_i K_m}{d_k}\right)} \tag{14}$$

Here, Q_i and K_j are the query and key vectors corresponding to patches i and j, respectively, and d_k is the dimensionality of the key space. This formulation allows the model to capture both short and long-range dependencies among different regions of the rice leaf image. To enhance feature diversity, GANs were used to produce synthetic samples for underrepresented disease categories. This augmentation expanded the overall feature space:

$$F_{total} = F_{real} \cup F_{synthetic} \tag{15}$$

where $real$ and $synthetic$ denote the feature sets extracted from real and GAN-generated images, respectively. This combination enabled the models to learn more discriminative and generalizable representations. Through this automatic process, the most relevant and high-level semantic features were retained, ensuring robust detection of rice leaf diseases despite variations in appearance and environmental factors.

3.5. Disease prediction and performance evaluation

Once training was completed, the optimized ViT architecture was employed for inference on unseen test images to evaluate its practical applicability in real-world scenarios. Each input image was carefully processed through the model to predict the disease type. The predicted outputs $\hat{y}$ were systematically compared with the associated true labels y for the purpose of measuring accuracy and reliability. The performance of the model was assessed based on commonly used evaluation criteria, such as accuracy, precision, recall, F1-score, and an analysis of the confusion matrix, to provide both overall and class-wise insights. These metrics are formally defined as (16)-(19):

$$Accuracy = \frac{TP + TN}{TP+TN+FP+FN} \tag{16}$$

$$Precision = \frac{TP}{TP + FP} \tag{17}$$

$$Recall = \frac{TP}{TP + FN} \tag{18}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recal} \quad (19)$$

where TP, TN, FP, and FN denote the number of true positives, true negatives, false positives, and false negatives, respectively. These quantitative measures ensure a balanced evaluation of both positive and negative classifications across disease categories. To analyze the training dynamics, model convergence was monitored by tracking the classification loss L_{cls} and accuracy over epochs *t*:

$$L_{cls(t+1)} < L_{cls(t)}, \forall t \in [1, T] \quad (20)$$

$$Acc(t+1) > Acc(t), until\ convergence. \quad (21)$$

A smooth and monotonic decrease in *CLS* with a corresponding increase in *Acc* confirmed stable learning and effective parameter optimization of the ViT model. Compared to conventional CNN-based architectures, the ViT model exhibited superior performance in both accuracy and generalization, achieving an overall accuracy of 96.5% as presented in Table 3. Moreover, the integration of GAN-augmented data mitigated skewed class frequencies and enhanced the capability of the model to accurately classify underrepresented disease categories.

Table 3. Performance comparison of different deep learning models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
VGG16	91.5	90.8	90.2	90.5
ResNet50	92.7	92.0	91.6	91.8
MobileNetV2	93.2	92.5	92.1	92.3
DenseNet121	92.0	91.4	91.0	91.2
DenseNet169	92.3	91.7	91.2	91.5
Xception	91.8	91.0	90.6	90.8
ResNet50V2	92.5	91.8	91.4	91.6
ViT	96.5	95.8	95.4	95.6

4. EXPERIMENT AND RESULT

4.1. Experiment setup

All experimental evaluations were performed on a high-capacity computational setup equipped with an Intel® Core™ i9-13900K CPU @ 3.00 GHz, NVIDIA GeForce RTX 4090 GPU (24 GB VRAM), and 64 GB RAM, running Ubuntu 22.04 LTS. The implementation was carried out using Python 3.10 with deep learning frameworks TensorFlow 2.15 and PyTorch 2.2. A consistent configuration was maintained for both the proposed ViT and all baseline CNN models. For all models, a batch size of 32 and a learning rate of 0.0001 were used during training, and 50 epochs using the Adam optimizer. The categorical cross-entropy loss function was used for multi-class classification. To prevent overfitting, early stopping and learning rate scheduling were applied during training. To examine model generalization, the proposed ViT was tested on an external field dataset of 500 unseen images, achieving 94.8% accuracy, confirming robustness beyond the training distribution. The trained model requires 65 MB of storage and demonstrates an average processing time of 38 ms per image on an NVIDIA RTX 4090 and 210 ms on a Jetson Nano device, demonstrating suitability for real-time and edge deployment.

4.2. Evaluation method

Model evaluation was conducted using the performance metrics defined in section 3.5., namely accuracy, precision, recall, and F1-score. To assess the individual contribution of GAN-based augmentation, an ablation study was performed under three experimental configurations such as ViT without any data augmentation, achieving 93.4% accuracy, ViT with traditional augmentation only, achieving 94.1% accuracy, ViT with GAN-based augmentation, achieving 96.5%. The quantitative improvement obtained by the proposed GAN augmented ViT was statistically validated using a paired t-test. Let x_i and y_i denote the accuracy values for the baseline and GAN-augmented models over n experimental folds. The paired t-statistic is defined as:

$$t = \frac{\bar{d}}{s_d \sqrt{n}} \quad (22)$$

The null hypothesis H_0 assumes no significant difference between configurations. Results showed $p < 0.01$, indicating that the performance gain of the GAN-augmented ViT is statistically significant. In addition to

overall accuracy improvement (2.4% over traditional augmentation and 3.1% over baseline), per-class F1-scores increased by an average of 4.2% for minority disease categories, confirming that synthetic data generation effectively enhanced feature diversity and balance. To further verify generalization capability, the trained ViT model was evaluated on an external field dataset of 500 unseen rice leaf images, achieving 94.8% accuracy under natural illumination and background variability.

Finally, computational efficiency was analyzed in terms of average inference time T_{avg}, defined as:

$$T_{avg} = \frac{1}{N}\sum_{i=1}^{N}\left(t_i^{end} - t_i^{start}\right) \quad (23)$$

where N is the number of test images, and t^{start}, t^{end} denote the start and end timestamps of prediction for image. The model demonstrated an average inference duration of 38 ms per image on an NVIDIA RTX 4090 GPU and 210 ms on a Jetson Nano device, demonstrating suitability for real-time and edge deployments in precision agriculture.

5. RESULTS AND DISCUSSION

As shown in Figure 2, the ViT achieved a smooth and consistent increase in training and validation accuracy, eventually converging to an overall accuracy of 96.5%. In contrast, CNN architectures such as MobileNetV2 and ResNet50 achieved lower performance, with accuracies of 93.2% and 92.7%, respectively. GAN-based data augmentation had a substantial impact on the results, particularly in balancing underrepresented classes. This enhancement led to an average performance gain of 4%–5% across minority disease categories and reduced misclassification rates for severe cases. The training and validation loss steadily decrease, indicating improved model optimization, while both training and validation accuracy increase, demonstrating enhanced generalization after fine-tuning.

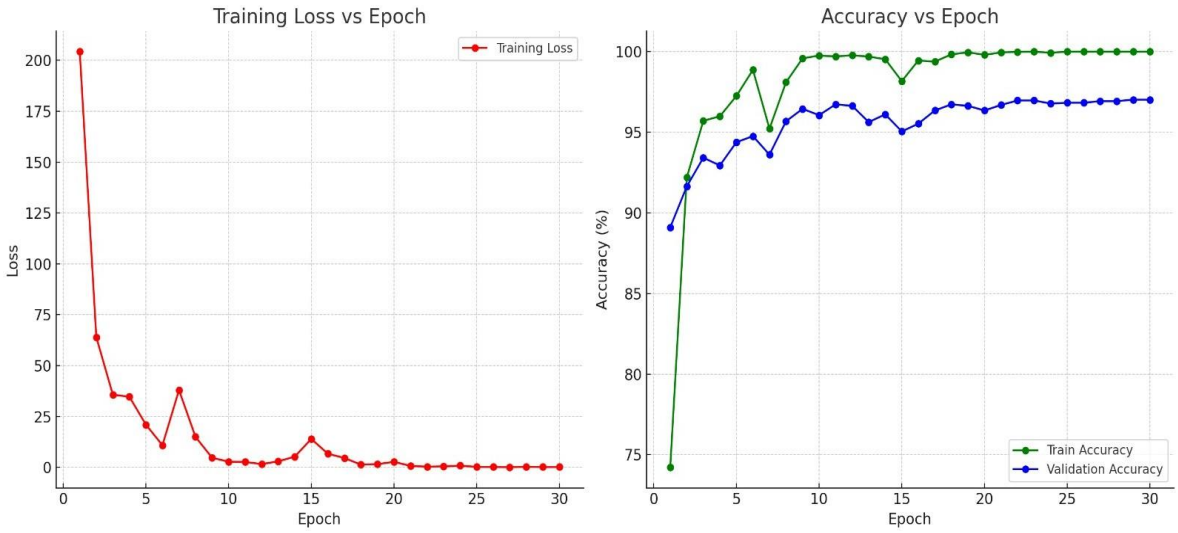

Figure 2. Train vs validation accuracy

For the early and precise detection of rice diseases, the ViT model produced lower false-negative rates than the CNN baselines, as shown by the confusion matrix in Figure 3. The matrix exhibits strong diagonal dominance across all ten classes, indicating a high correct classification rate. Most disease categories, including bacterial leaf blight, bacterial leaf streak, bacterial panicle blight, dead heart, and normal leaves, show very limited misclassification, demonstrating that the model learns highly discriminative features for these classes. However, minor confusion is observed among visually similar disease categories. In particular, blast is occasionally misclassified as tungro and hispa, which can be attributed to overlapping lesion patterns, discoloration, and texture similarities in advanced disease stages. Similarly, downy mildew shows limited confusion with blast and brown spot, as these diseases share diffuse lesion boundaries and comparable color variations under natural field conditions. The hispa class exhibits some confusion with normal leaves and blast, which is expected since early-stage hispa infestations produce subtle visual

symptoms that closely resemble healthy leaf textures. Tungro is also occasionally confused with blast and hispa due to common yellowing and mosaic-like patterns on the leaf surface. Importantly, the number of false negatives for minority disease classes remains low, indicating that the integration of GAN-based data augmentation effectively improves class balance and recall.

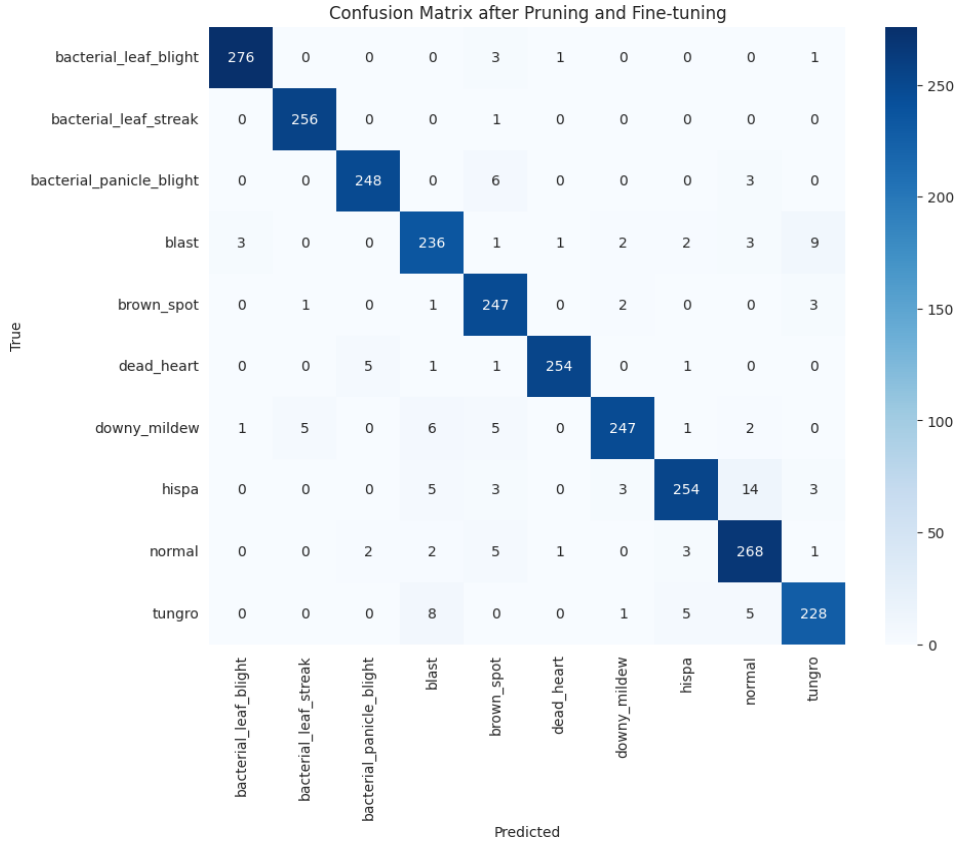

Figure 3. Confusion matrix after fine tuning

The confusion matrix confirms that the proposed framework not only achieves high overall accuracy but also maintains robust class-wise discrimination, which is essential for reliable rice leaf disease diagnosis in real-world agricultural settings. Table 4 presents the class-wise precision, recall, and F1-score. High precision and recall values are observed for most disease categories, indicating reliable classification performance. Slightly lower recall values for visually similar classes such as blast, hispa, and tungro can be attributed to overlapping lesion patterns and color similarities.

Finally, the results show that the ViT architecture combined with GAN-augmented data offers a dependable and scalable method for automated rice disease diagnosis. This ensures not only improved accuracy but also practical applicability for deployment in real time in precision agriculture.

Table 4. Per-class precision, recall, and F1-score

Class	Precision (%)	Recall (%)	F1-score (%)
Bacterial leaf blight	98.57	98.22	98.40
Bacterial leaf streak	97.71	99.61	98.65
Bacterial panicle blight	97.25	96.50	96.88
Blast	91.12	91.83	91.47
Brown spot	90.81	97.24	93.92
Dead heart	98.83	96.95	97.88
Downy mildew	96.86	92.51	94.64
Hispa	95.49	90.07	92.70
Normal	90.85	95.04	92.89
Tungro	93.06	92.31	92.68

6. CONCLUSION

This research proposed a GAN-augmented ViT framework for the early identification of diseases of rice leaves, offering a novel solution for precision agriculture. By utilizing DCGAN-based synthetic data generation, the framework effectively addressed dataset imbalance and improved model generalization across multiple rice disease categories. The ViT architecture, enhanced through self-attention mechanisms, successfully captured both global and local contextual relationships, enabling accurate recognition of disease type and severity even under practical and variable field conditions. Observations from the experiments suggested that the proposed ViT model achieved a 96.5% overall accuracy, outperforming all benchmark CNN models such as ResNet50, MobileNetV2, and DenseNet121. The inclusion of GAN-augmented data led to significant improvements in minority class detection, reduced false negatives, and enhanced reliability for early diagnosis. Furthermore, evaluation on an external field dataset confirmed the framework's strong generalization ability, while its efficient inference time on edge devices indicates readiness for real-time deployment. Future research will concentrate on extending the dataset across broader Environmental circumstances, incorporating multi-crop disease detection, and integrating the model with UAVs, IoT devices, and mobile platforms to support large-scale automated monitoring.

FUNDING INFORMATION

This research is funded by the International Islamic University of Chittagong and Dhaka University of Engineering and Technology.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Saiful Islam	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Md. Nasim Akhtar	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓
Mahadi Hasan	✓	✓			✓	✓	✓	✓	✓				✓	✓
A. N. M. Rezaul Karim	✓				✓		✓			✓	✓		✓	✓
Israt Binteh Habib													✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

REFERENCES

[1] E. Flaherty, "Around the world in eight commodities, episode three: Rice," *Global Landscapes Forum*, Dec. 2024. https://thinklandscape.globallandscapesforum.org/71338/around-the-world-in-eight-commodities-episode-three-rice/ (accessed Feb. 10, 2026).

[2] IFT Next, "About 3.5 billion people eat rice, which supplies nearly 20% of global human per capita energy," *IFT Food Technology Magazine*, Nov. 2018, [Online]. Available: https://www.ift.org/news-and-publications/food-technology-magazine/issues/2018/november/columns/iftnext-rice-production

[3] S. Miryala and K. Rasane, "Enhancing sugarcane leaf disease classification using vision transformers over CNNs," *Discover Artificial Intelligence*, vol. 5, no. 1, p. 89, 2025, doi: 10.1007/s44163-025-00340-7.

[4] C. G. Simhadri, H. K. Kondaveeti, V. K. Vatsavayi, A. Mitra, and P. Ananthachari, "Deep learning for rice leaf disease detection: A systematic literature review on emerging trends, methodologies and techniques," *Information Processing in Agriculture*, vol. 12, no. 2, pp. 151–168, 2025, doi: 10.1016/j.inpa.2024.04.006.

[5] S. K. M, A. R, S. Kurian. P, and S. O.K, "Interpretable multitask deep learning model for detecting and analyzing severity of rice bacterial leaf blight," *Scientific Reports*, vol. 15, no. 1, p. 27313, 2025, doi: 10.1038/s41598-025-12276-0.

[6] M. Z. Uddin, M. N. Mahamood, A. Ray, M. I. Pramanik, F. Alnajjar, and M. A. R. Ahad, "E2ETCA: End-to-end training of CNN and attention ensembles for rice disease diagnosis," *Journal of Integrative Agriculture*, vol. 25, no. 2, pp. 756–768, 2026, doi: 10.1016/j.jia.2024.03.075.

[7] Z. Jiang, Z. Dong, W. Jiang, and Y. Yang, "Recognition of rice leaf diseases and wheat leaf diseases based on multi-task deep

transfer learning," *Computers and Electronics in Agriculture*, vol. 186, p. 106184, 2021, doi: 10.1016/j.compag.2021.106184.
[8] D. S. Joseph, P. M. Pawar, and K. Chakradeo, "Real-time plant disease dataset development and detection of plant disease using deep learning," *IEEE Access*, vol. 12, pp. 16310–16333, 2024, doi: 10.1109/ACCESS.2024.3358333.
[9] M. H. Bijoy *et al.*, "Towards sustainable agriculture: a novel approach for rice leaf disease detection using dCNN and enhanced dataset," *IEEE Access*, vol. 12, pp. 34174–34191, 2024, doi: 10.1109/ACCESS.2024.3371511.
[10] P. S. Roy and V. Kukreja, "Vision transformers for rice leaf disease detection and severity estimation: a precision agriculture approach," *Journal of the Saudi Society of Agricultural Sciences*, vol. 24, no. 3, p. 3, 2025, doi: 10.1007/s44447-025-00007-w.
[11] A. B. Ayyappan, T. Gobinath, M. Kumar, and A. Sivaramakrishnan, "Rice plant disease detection using convolutional neural networks," *Discover Artificial Intelligence*, vol. 5, no. 1, p. 50, 2025, doi: 10.1007/s44163-025-00277-x.
[12] R. Raman and S. Jayaraman, "Artificial intelligence for sustainable farming with dual branch convolutional graph attention networks in rice leaf disease detection," *Scientific Reports*, vol. 15, no. 1, p. 10595, 2025, doi: 10.1038/s41598-025-94891-5.
[13] A. Haridasan, J. Thomas, and E. D. Raj, "Deep learning system for paddy plant disease detection and classification," *Environmental Monitoring and Assessment*, vol. 195, no. 1, p. 120, 2023, doi: 10.1007/s10661-022-10656-x.
[14] S. T. Y. Ramadan, M. S. Islam, T. Sakib, N. Sharmin, M. M. Rahman, and M. M. Rahman, "Image-based rice leaf disease detection using CNN and generative adversarial network," *Neural Computing and Applications*, vol. 37, no. 1, pp. 439–456, 2025, doi: 10.1007/s00521-024-10572-w.
[15] Y. Wang *et al.*, "A hybrid approach for rice crop disease detection in agricultural IoT system," *Discover Sustainability*, vol. 5, no. 1, p. 99, 2024, doi: 10.1007/s43621-024-00285-4.
[16] S. Vallabhajosyula, V. Sistla, and V. K. K. Kolli, "A novel hierarchical framework for plant leaf disease detection using residual vision transformer," *Heliyon*, vol. 10, no. 9, 2024, doi: 10.1016/j.heliyon.2024.e29912.
[17] P. Temniranrat, K. Kiratiratanapruk, A. Kitvimonrat, W. Sinthupinyo, and S. Patarapuwadol, "A system for automatic rice disease detection from rice paddy images serviced via a Chatbot," *Computers and Electronics in Agriculture*, vol. 185, p. 106156, 2021, doi: 10.1016/j.compag.2021.106156.
[18] P. Pai, S. Amutha, M. Basthikodi, B. M. Ahamed Shafeeq, K. M. Chaitra, and A. P. Gurpur, "A twin CNN-based framework for optimized rice leaf disease classification with feature fusion," *Journal of Big Data*, vol. 12, no. 1, p. 89, 2025, doi: 10.1186/s40537-025-01148-z.
[19] R. Sreevallabhadev, "An improved machine learning algorithm for predicting blast disease in paddy crop," *Materials Today: Proceedings*, vol. 33, pp. 682–686, 2020, doi: 10.1016/j.matpr.2020.05.802.
[20] O. O. Martins, C. C. Oosthuizen, and D. A. Desai, "RiceLeafClassifier-v1.0: A Quantized Deep Learning Model for Automated Rice Leaf Disease Detection and Edge Deployment," *Engineering Reports*, vol. 7, no. 6, p. e70231, 2025, doi: 10.1002/eng2.70231.
[21] P. Li, J. Zhou, H. Sun, and J. Zeng, "RDRM-YOLO: a high-accuracy and lightweight rice disease detection model for complex field environments based on improved YOLOv5," *Agriculture (Switzerland)*, vol. 15, no. 5, p. 479, 2025, doi: 10.3390/agriculture15050479.
[22] K. K. Kumar *et al.*, "A convolutional neural network approach for rice leaf disease detection in India using deep learning," *Journal of Neonatal Surgery*, vol. 14, no. 4s, pp. 1024–1028, 2025, doi: 10.52783/jns.v14.1909.
[23] M. J. U. Chowdhury, Z. I. Mou, R. Afrin, and S. Kibria, "Plant leaf disease detection and classification using deep learning: a review and a proposed system on Bangladesh's perspective," *International Journal of Science and Business*, vol. 28, no. 1, pp. 193–204, 2023, doi: 10.58970/ijsb.2214.
[24] K. N. Rahman, S. C. Banik, R. Islam, and A. Al Fahim, "A real time monitoring system for accurate plant leaves disease detection using deep learning," *Crop Design*, vol. 4, no. 1, p. 100092, 2025, doi: 10.1016/j.cropd.2024.100092.
[25] P. Doctor, "Paddy disease classification," *Kaggle.com*, 2022. https://www.kaggle.com/competitions/paddy-disease-classification/data (accessed Dec. 02, 2026).

BIOGRAPHIES OF AUTHORS

Saiful Islam is assistant professor at the Department of Computer Science and Engineering, International Islamic University of Chittagong, Bangladesh. He has achieved B.Sc. in computer science and engineering from Dhaka University of Engineering and Technology. He holds a M.Sc. degree in computer science and engineering from Chittagong University of Engineering and Technology. His research areas are image processing, computer vision, cyber security, natural language processing, and optimization. He can be contacted at email: saifulcse@iiuc.ac.bd.

Md. Nasim Akhtar earned his B.Sc. Tech. and M.Tech. in CSE from the National Technical University of Ukraine and a Ph.D. in information technology from Moscow State Academy, Russia. He is a professor in the Department of CSE at DUET and has served as dean, head of department, and director of the computer center. He is also a member of IEB and BCS and has contributed to national IT projects like the Tier IV data center at High-tech Park, Gazipur. His research interests include distributed and cloud computing, algorithms, and modern operating systems. He can be contacted at email: drnasim@duet.ac.bd.

M. Mahadi Hassan is an associate professor in the Department of Computer Science and Engineering (CSE). He earned his B.Sc. (Hons) in computer science and engineering from the International Islamic University Chittagong (IIUC) and completed his M.Sc. in engineering from Multimedia University (MMU), Malaysia. He is currently pursuing his Ph.D. at Chittagong University of Engineering and Technology (CUET). His research interests include machine learning (ML), deep learning (DL), cyber security, and image processing. With strong academic and research experience, he is committed to fostering innovation and excellence in the field of computer science. He can be reached via email at mahadi_cse@yahoo.com.

A. N. M. Rezaul Karim obtained B.Sc. (honors) and M.Sc. degree in mathematics from University of Chittagong (CU), Bangladesh, PGD degree in ICT (information and communication technology) from IICT, BUET and obtained Ph.D. degree from Islamic University, Kushtia, Bangladesh. Currently, he is working as a full-time professor in the Department of Computer Science and Engineering, International Islamic University Chittagong, Bangladesh. His research interests include modeling and simulation, mathematical analysis for computer science, function approximation, and optimization. He can be contacted at email: zakianaser@yahoo.com.

Israt Binteh Habib earned her B.Sc. and M.Sc. degrees in computer science and engineering (CSE) from the University of Chittagong, Bangladesh. She is currently serving as a lecturer in the Department of Computer Science and Engineering at the International Islamic University Chittagong (IIUC). Her research interests include belief rule-based expert systems, machine learning, and image processing. She can be contacted at: israthabib.cse@gmail.com.