

A Lightweight Residual Dilated CNN–Transformer Framework for Efficient Rice Leaf Disease Classification

Uddagiri Sirisha

PVP Siddhartha Institute of Technology

Kanhaiya Sharma

Symbiosis International (Deemed University)

Phani Praveen

PVP Siddhartha Institute of Technology

Deepak Parashar

`deepak.parashar@manipal.edu`

Manipal Academy of Higher Education

N S Koti Mani Kumar Tirumanadham

Sir C R Reddy College of Engineering

Research Article

Keywords: Rice leaf, transformer, lightweight residual dilated network, grad-cam, squeeze-excitation blocks

Posted Date: March 24th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9038735/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

A Lightweight Residual Dilated CNN–Transformer Framework for Efficient Rice Leaf Disease Classification

Uddagiri Sirisha¹, Kanhaiya Sharma², S Phani Praveen³, Deepak Parashar^{4*}, N S Koti Mani Kumar Tirumanadham⁵

Uddagiri Sirisha¹⁺, Kanhaiya Sharma²⁺, Phani Praveen³, Deepak Parashar^{4*+}, N S Koti Mani Kumar Tirumanadham⁵

^{1,3}Department of CSE, PVP Siddhartha Institute of Technology, Vijayawada, India

²Department of CSE, Symbiosis International (Deemed University), Pune, India

⁴Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India

⁵Department of Computer Science and Engineering, Sir C R Reddy College of Engineering, Eluru, India, India

*Corresponding author(s). E-mail(s): deepak.parashar@manipal.edu

Contributing authors: sirisha.uddagiri@gmail.com, sharmakanhaiya@gmail.com, phani.0713@gmail.com, manikumar1248@gmail.com

Abstract: Although rice is a staple food for more than half of the world's population, several illnesses represent a severe threat to rice farming, lowering yields by as much as 70%. Human visual examination, the basis of traditional disease diagnosis, has several limitations, including subjectivity, inconsistent results, and the inability to cover large agricultural areas. So far, using deep learning to automate plant disease diagnosis has been effective, but there are still numerous problems to address. Some of these include that they are difficult to understand, require a lot of computer power, and are not well-suited for field use. This study proposes an innovative hybrid deep learning framework called RiceLeafCNN-Transformer, which integrates transformer architectures with convolutional neural networks to address existing limitations. A proposed method for representing features at multiple scales enhances computing speed by combining dilated convolutions, depth wise separable convolutions, and squeeze-excitation blocks. We can accomplish global contextual modelling and hierarchical feature extraction by combining a transformer encoder with a lightweight convolutional neural network (CNN) backbone. Proposed model shows state-of-the-art performance in its rigorous experimental validation using three datasets of rice leaf disease. Dataset A has 18,445 photos, and Dataset B has 11,790 photos. It does this by keeping inference times between 8.3 and 9.9 ms and getting 99.2% accuracy. The proposed strategy can help farmers and agricultural specialists save time and effort by linking theoretical lab performance to real-world use. This way, they get information that is both useful and relevant.

Keywords: Rice leaf; transformer; lightweight residual dilated network; grad-cam; squeeze-excitation blocks

Article Highlights

1. A novel lightweight hybrid framework combines dilated CNNs, depthwise separable convolutions, squeeze–excitation blocks, and a transformer encoder to achieve efficient multi-scale feature extraction and global contextual modelling for rice leaf disease classification.
2. Achieves state-of-the-art performance with 99.2% accuracy and fast inference time of 8.3–9.9 ms, validated using three rice disease datasets, including large-scale datasets containing 18,445 and 11,790 images.
3. Incorporates explainable AI (XAI) tools such as Grad-CAM and attention visualisation, offering transparent insights into the network’s decision-making process.
4. Ablation studies reveal key contributors to performance gains, highlighting the impact of residual dilated blocks (3.1%), squeeze-excitation attention (2.1–2.2%), and pattern-enhancement modules (1.8–1.9%), demonstrating the model’s architectural effectiveness.

1. Introduction

Rice (*Oryza sativa*) is the primary food source for approximately 3.5 billion people worldwide, primarily in all regions. This grain is crucial for developing areas to achieve and maintain socioeconomic stability and food security, as it plays a unique role in global nutrition systems. Rice is a vital part of the culture of many smallholder farmers, and hundreds of millions of them depend on it for their livelihood [1,2]. If production were to stop suddenly, it would have a significant impact on food prices, trade balances, and political stability. Rice is very important in farming, but growing it is hazardous because it is susceptible to numerous illnesses that can drastically reduce the quality and quantity of the crop. Blast, the worst fungal disease, might kill up to 70% of your crop if you're cultivating sensitive cultivars[3,4]. Infectious diseases that cause leaves to dry up and wilt are among the most severe. Some of these are bacterial blight, brown spot, and leaf smut. When the right conditions are present and early diagnosis is insufficient, diseases can quickly spread across fields and ruin entire crops. This is especially true in places where rice is grown that are hot and humid. The immediate loss of yield isn't the only economic effect; there are two additional effects: higher costs of producing fungicides and the likelihood of disease resistance [5,6].

In the past, most methods for detecting plant diseases involved educated farmers or agricultural extension professionals searching for signs on plants. This method works well most of the time, but it has several significant drawbacks, including the need for people to interpret symptoms, its inaccuracy, and its limited effectiveness in large agricultural areas. Since most of the expertise required to make a correct diagnosis is located in larger cities, many rural farming villages may not have quick access to diagnostic services. Additionally, infections that occur early on may not exhibit any signs until the disease has spread throughout the entire crop. This suggests that therapies might be less effective, necessitating more severe treatment protocols[7,8]. Computer Vision and Deep Learning can change the way we diagnose plant diseases by automating the process and making it very accurate. Convolutional neural networks (CNNs) can find patterns of sickness in controlled settings and get greater spatial information from pictures of leaves than human experts can. These systems can provide scalable monitoring solutions, regardless of their location, as they can handle hundreds of photos every day. Vision Transformers (ViTs) utilise self-attention algorithms to extract a significant amount of information from images. This has led to recent advances in this sector. The end result could be improved detection of disease patterns across wider portions of a crop or plant [9,10].

Modifying these systems for use in actual farming remains very challenging, despite the introduction of these new technologies. Changes in sunshine, occlusions from overlapping leaves, various stages of growth, and the simultaneous emergence of many illnesses are all potential factors that could complicate field conditions and pose problems with existing solutions. The high processing demands of modern designs are frequently too much for low-resource mobile devices or edge computing platforms to handle in agricultural applications. Agricultural experts are wary of using deep learning models because of the "black box" aspect, even though they depend on comprehensible diagnostic data to make critical treatment decisions. Because they can gain over trust by showing how thinking works, explainable AI technologies are obviously crucial.

- The RiceLeafCNN-Transformer is a new type of deep learning architecture that is a mix of two different types. It gets around these problems by combining CNNs' ability to find local features with Transformers' ability to see the big picture. We have created a model that correctly, quickly, and clearly sorts illnesses that impact rice leaves. You will be better able to find problems early and respond correctly. We made a plan so that it would work and be useful in places with few resources. People that work in agriculture, like farmers, will like how clear and useful the information is. This combines architectural efficiency with explainable AI strategies to close the gap between lab performance and field deployment. These works have made a big difference in the field of automated plant disease identification:
- The RiceLeafCNN-Transformer hybrid architecture uses a hybrid design to combine a Transformer encoder for global contextual modelling with a lightweight CNN backbone for hierarchical feature extraction. Dilated convolutions, residual connections, deepwise separable convolutions, and squeeze-excitation (SE) blocks all work together to make multi-scale feature representation and processing faster and better. This new design captures both localised texture patterns and global spatial correlations to help tell diseases apart that look identical.

- Utilizing explainable AI (XAI) techniques, proposed solution provides decision-makers with easily digestible information. This includes techniques like Gradient-weighted Class Activation Mapping and Attention Visualisation. For farmers, these methods mean determining which parts of the leaf image are most important for making accurate forecasts using the model. Because of this, people have more self-confidence and find it simpler to adapt. Instead of being a mysterious black-box classifier, the system may now be used as a diagnostic tool thanks to its explainable parts.
- Extensive experimental validation on datasets pertaining to rice leaf diseases demonstrates that the proposed technique surpasses state-of-the-art models in classification accuracy, noise resistance, and environmental generalisability. The research encompasses a comprehensive assessment of each architectural element, computational efficiency for practical implementations, comparisons with existing systems, and ablation experiments.

The remainder of the study is divided into two sections: Section 2 presents a state-of-the-art literature review, and Section 3 outlines the detailed methodology. Section 4 discusses the potential implications and findings for real-world farming practices. Section 5 concludes the study.

2. Literature Review

Researchers are eager to identify and classify rice leaf diseases, as they have a significant impact on global food security. Deep learning (DL) has recently achieved significant advancements, substantially enhancing the precision and reliability of prior research that employed less sophisticated techniques in machine learning and image processing. Some of the most relevant words in this post are "datasets," "image processing," "machine learning," and "deep learning-based methods."

2.1 Rice Leaf Disease Datasets

To create effective classification models, you need to utilise a variety of well-organised datasets. There are hundreds of pictures of diseased rice leaves, along with notes, in various public sources. The UCI Machine Learning Repository (2019), Mendeley, the Rice Disease Image Dataset (2022), and the Rice Leaf Diseases Detection (2023) on Kaggle (2022) are all examples of resources in this group. Due to their diversity in size, classification, and disease kind, these datasets are ideal for algorithm testing. According to the writers in [11] and [12], research institutions such as IRRI, BRRI, and TRRI do have private datasets available; however, these datasets are limited to specific fields.

2.2 Image Processing Approaches

Researchers used to rely on manual feature extraction and preprocessing in the beginning. The authors of [13] note that several frequently used solutions include noise reduction filters such as Gaussian, median, and Wiener, contrast enhancement (CLAHE), and segmentation methods like Otsu thresholding, clustering, and watershed for isolating sick regions. Another common strategy is contrast enhancement. Colour histograms, wavelets, and texture descriptors (GLCM, LBP, HOG) were among the features used by the contributors of [14]. By combining Gabor filtering with Artificial Neural Networks, researchers demonstrated a 91% success rate in differentiating between healthy and sick data. For the same reason, the authors of [15] discovered that Wiener filtering significantly improved picture quality. A more recent study by [16] used DenseNet-based feature extractors in conjunction with YOLOv5 segmentation to attain a 94.87% accuracy rate on the Kaggle Rice Leaf Dataset.

2.3 Machine Learning Techniques

Effective applications result from combining ML-based classification algorithms with robust feature extraction. Numerous models have found widespread application, including SVMs, Decision Trees, Random Forests, and Neural Networks. A model with an F1-score of 98.38% was constructed using support vector machines in a study conducted by [17]. The writers of [18] found that by combining SVM and k-means segmentation, three important diseases could be accurately classified with a 94.11% success rate. Adding clustering approaches, such as SLIC, to Random Forest

models did not significantly change the findings; they remained approximately 97% accurate [19]. ML approaches are effective, but they rely on well-established properties, which means they may not work in all situations. Additionally, depending on the field conditions, handcrafted attributes may not capture all the various ways sickness symptoms can manifest.

2.4 Deep Learning Techniques

Disease classification in rice has been revolutionised by deep learning. Since CNNs can detect unique patterns on their own, feature engineers are unnecessary. Unique convolutional neural networks (CNNs), transfer learning, ensemble learning, and hybrid methods are the primary foci of study. Highly effective were custom CNNs trained on datasets pertaining to rice diseases. Using four illness classifications to train a five-layer CNN, [20] attained 93% accuracy. Recognition accuracies of 99.2% were achieved by SE-SPne [21] and 99.71% by RiceDRA-Net [22] due to their improved feature representation and more complex structures, respectively. Transfer learning with sparse data can be accomplished in various ways. These strategies utilise pre-trained models of VGGNet, DenseNet, and ResNet. Using a pre-trained model that had been fine-tuned on the Kaggle rice-leaf dataset, [23] were able to attain an accuracy of 99.7%. To improve the extraction of temporal and spatial characteristics, researchers have tried to combine ML classifiers with CNNs or DL with RNNs. Predictions become more consistent and scalable when ensemble learning is used to combine CNN results [24, 25]. Deep learning models have continued to drive innovation in the field, despite these caveats. Large convolutional neural networks (CNNs) such as DenseNet are commonly used by academics for real-time field deployment on mobile or edge devices[26,27]. When it comes to capturing global dependencies, more contemporary transformer-based techniques are superior to Elman RNNs. The majority vote and other widely used ensemble methods do not fully utilise stacking and gradient boosting. Future research should focus on optimising architectures, creating accurate yet lightweight models such as EfficientNetV2, ConvNeXt, Vision Transformers, and Swin Transformers, and developing ensemble approaches. Augmentation based on GANs, test-time augmentation, and multi-scale feature fusion can improve generalisability and robustness [28,29,30].

3. Methodology

To efficiently and effectively categorise diseases in rice leaves, a hybrid deep learning system called RiceLeafCNN-Transformer was developed. The design and implementation of this system are detailed in this section. Figure 1 shows the entire pipeline.

3.1. Dataset Description and Preparation

In this study, two large datasets on illnesses that damage rice leaves were utilised to assess their reliability and generalizability. The core dataset comprises 11,790 photos, sorted into 9 groups [31, 32]. It has pictures of both healthy leaves and the most common rice diseases. The specific categories in question are bacterial leaf blight, brown spot, healthy, leaf blast, leaf scorch, narrow brown spot, neck blast, rice hispa, sheath blight, and tungro. Another dataset of 18,445 photos to further test and confirm the model across different data sources and situations. The datasets were carefully divided into three groups: testing (15%), validation (15%), and training (70%). To ensure the model's classification performance was evaluated fairly and to avoid bias, the class distribution was maintained the same across all splits. Figure 2 shows some of the photographs from the dataset. Class wise distribution from dataset and dataset is provided in Table 1.

Table 1: Class wise Distribution from Dataset A and Dataset B

Dataset A	Dataset B
Class0_Bacterial Leaf Blight : 1197	Class0_bacterial_leaf_blight--1762
Class1_Brown Spot : 1546	Class1_brown_spot--1860

Class2_Healthy Rice Leaf : 1085	Class2_healthy--1882
Class3_Leaf Blast : 1748	Class3_leaf_blast-- 2163
Class4_Leaf scald : 1332	Class4_leaf_scald--2056
Class5_Narrow Brown Leaf Spot : 954	Class5_narrow_brown_spot--1798
Class6_Neck_Blast : 1000	Class6_neck_blast--1000
Class7_Rice Hispa : 1299	Class7_rice_hispa--1461
Class8_Sheath Blight : 1629	Class8_sheath_blight--1578
	Class9_tungro--1740

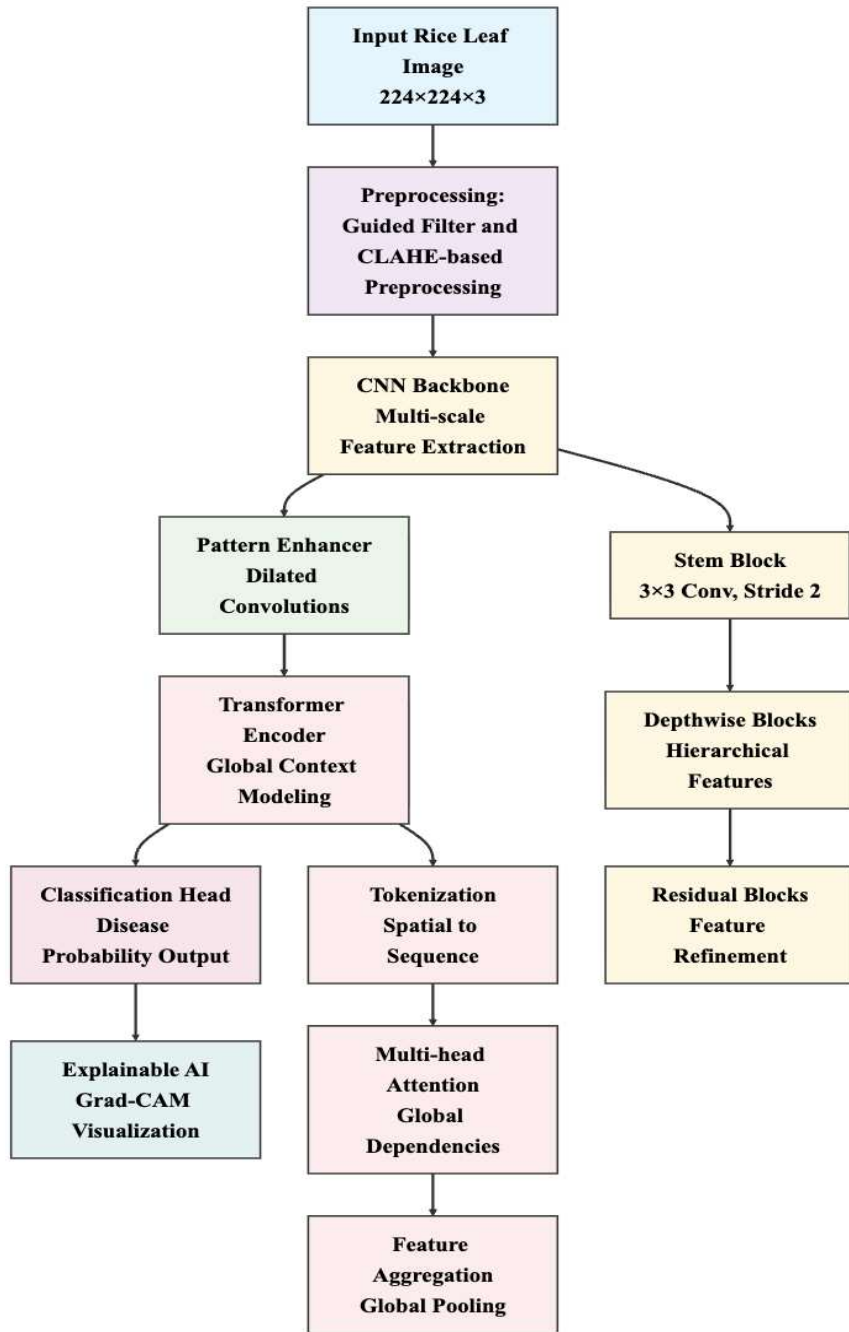


Figure 1: The overall architecture of the proposed RiceLeafCNN-Transformer model.

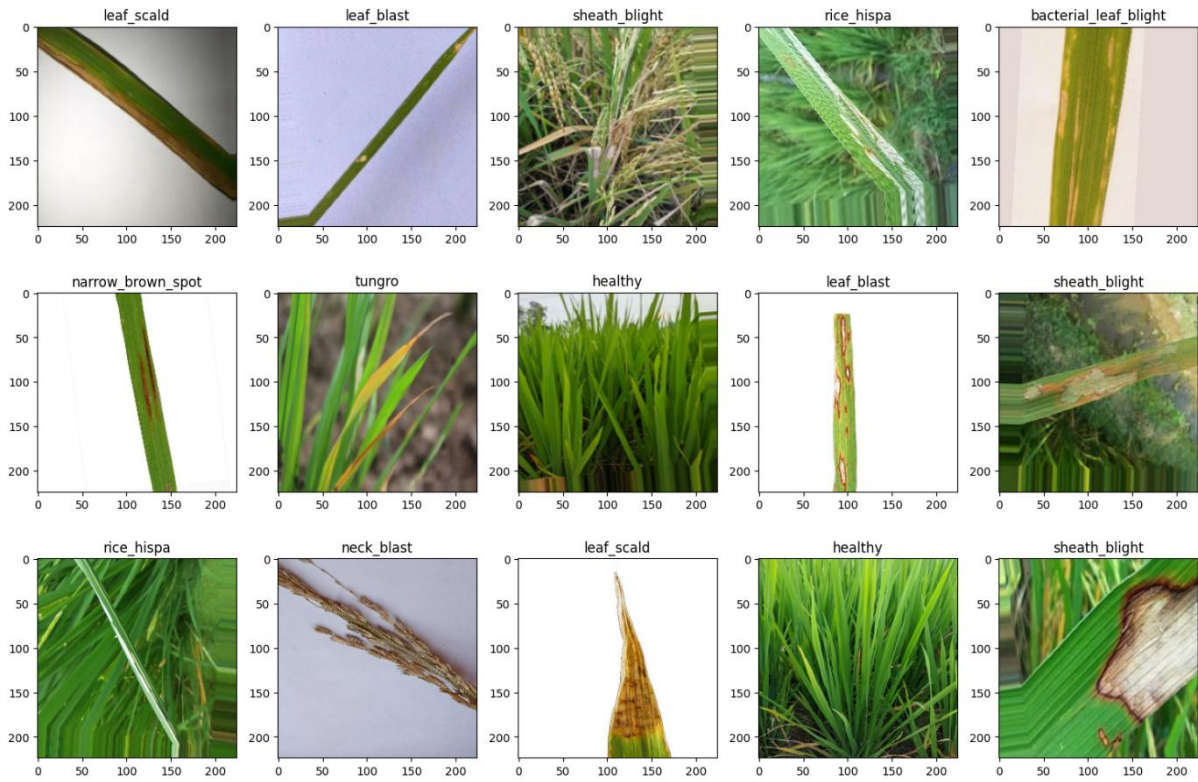


Figure 2: Sample Images From the Dataset

3.2. Data Preprocessing and Augmentation

To improve the model's ability to generalise and avoid overfitting, a complex and multi-stage preparation process was necessary to enhance the quality of the pictures and expand the dataset. As Algorithm 1 states, the first and most important stage was CLAHE, which stands for Contrast Limited Adaptive Histogram Equalisation. Subtle illness patterns, lesions, and discolourations are substantially more visible in low-contrast or photographs with uneven lighting when using CLAHE. It improves luminance channel local contrast by operating on the LAB colour space. This is crucial for making the correct diagnosis. Figures 3 and 4 illustrate that the pre-enhanced dataset was kept separate to ensure that training proceeded smoothly. After contrast enhancement, a full set of changes was made in real time during training. The training set underwent a rigorous process of data enhancement, which included random horizontal and vertical flips, rotation within ± 30 degrees, attribute transformations for translation and scaling, colour jittering for adjustments to hue, saturation, brightness, and contrast, and occasionally, sharpness and blur adjustments. This method makes the dataset more diverse in a way that is similar to the many different changes in orientation, size, light, and focus that the model may see on a real farm. In order to provide a consistent and impartial evaluation, a less complex transformation pipeline was utilized for validation and testing. This pipeline merely included resizing to 224×224 pixels, converting to tensor, and normalizing using ImageNet's mean and standard deviation values. Background removal is depicted in Figure 4.

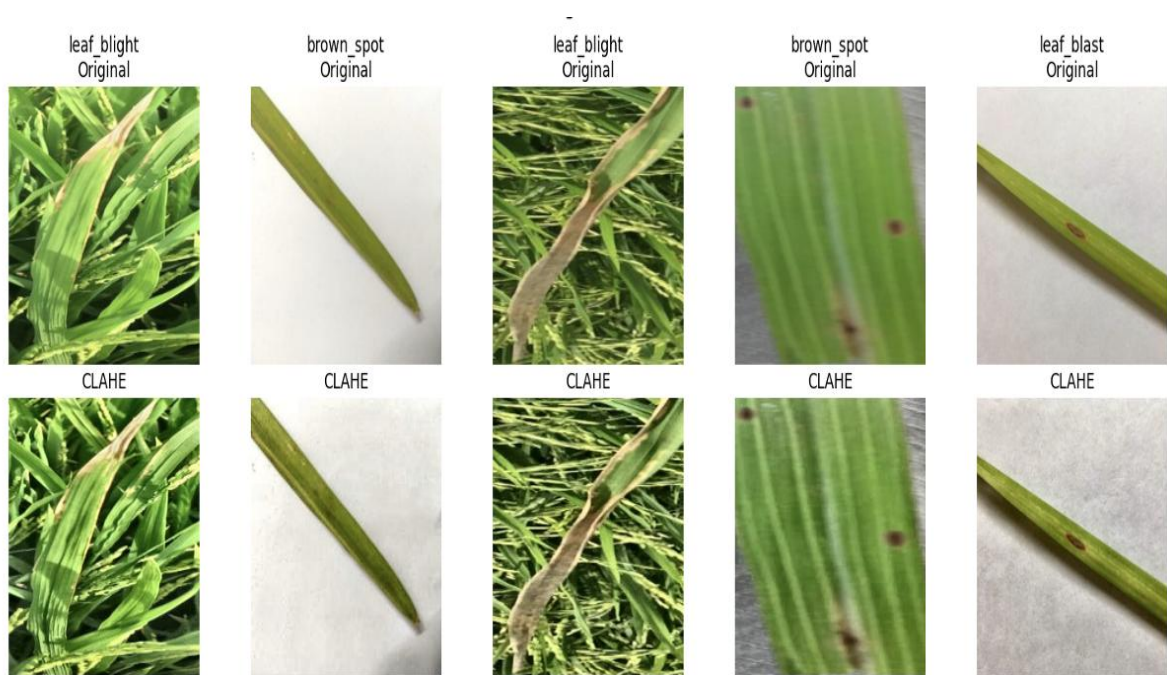


Figure 3: Preprocessing using CLAHE method



Figure 4: Background Removal

Algorithm 1 Guided Filter and CLAHE-based Preprocessing

1: Input: Dataset D , Rice Leaf Disease Images $R = \{R_1, R_2, \dots, R_n\}$
2: Output: Edge-preserved and contrast-enhanced rice leaf image PR
3: procedure PREPROCESS_CLAHE_GUIDED(D, R)
4: for each image R_i in D do
5: Apply Guided Image Filtering to remove noise while preserving edges:
 $I'(x) = ak \cdot I(x) + bk, \text{ for all } x \text{ in window } \omega_k$
6: Perform CLAHE (Contrast Limited Adaptive Histogram Equalization):
 $R'_i = \text{CLAHE}(I'(x), \text{clipLimit}, \text{tileGridSize})$
7: Normalize pixel intensity values to range $[0,1]$:
 $R_{norm} = (R'_i - \min(R'_i)) / (\max(R'_i) - \min(R'_i))$
8: Assign preprocessed image:
 $PR \leftarrow R_{norm}$
9: end for
10: Return preprocessed dataset PR
11: end procedure

3.3 The Overall Architecture Of The Proposed Riceleafcnn-Transformer Model.

Algorithm 2 outlines the proposed RiceLeafCNN-Transformer, a high-tech hybrid deep learning framework that uses a combination of convolutional neural networks and transformer designs to classify diseases in rice leaves. An image of a rice leaf is used to train the model's multi-stage CNN backbone to extract hierarchical features. The architecture uses a stem block to process the input picture I , which has a 224×224 pixel spatial resolution. The mathematical expression for the stem block operation is given by equation 1.

$$X_{\text{stem}} = \sigma(\text{BN}(\text{Conv2d}(I, W_{\text{stem}}))) \quad (1)$$

In the context of the Convolution Operation, W_{stem} stands for the weights, σ for the silu action function, and BN for batch normalisation. This first step in processing is primarily intended to set the stage for feature extraction in subsequent layers. It does this by keeping crucial visual information while reducing spatial dimensions and increasing channel depth from 3 to 32.

3.3.1. Hierarchical Feature Extraction with Blocks of Different Depths

Hierarchical Feature Extraction Using Blocks of Varying Depths. These blocks gradually modify the feature maps across a range of spatial scales. The network employs five depthwise blocks with purposeful stride patterns (1-2-1-2-1) to systematically reduce spatial resolution and enhance feature depth from 32 to 512 channels. For every depthwise separable block operation, there is a formula in Equation. 2.

$$\begin{aligned}
X_{\text{depthwise}} &= \sigma(\text{BN}(\text{Conv2d}_{\text{depthwise}}(X_{\text{in}}))) \\
X_{\text{pointwise}} &= \sigma(\text{BN}(\text{Conv2d}_{\text{pointwise}}(X_{\text{depthwise}}))) \\
X_{\text{out}} &= \text{SEBlock}(X_{\text{pointwise}})
\end{aligned} \quad (2)$$

where the depthwise convolution operates on each input channel separately as in equation 3.

$$\text{Conv2d}_{\text{depthwise}}(X)=[\text{conv}(X1),\text{conv}(X2),\dots,\text{conv}(XC)] \quad (3)$$

To improve channel-wise feature responses, the squeeze-and-excitation (SE) attention mechanism does the following:

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W x_c(i, j) \quad (4)$$

$$s = \sigma(W_2 \delta(W_1 z)) \quad (5)$$

$$X_{SE} = s \cdot X \quad (6)$$

In the feature map, z is a vector of length c . In the fully linked layer, W_1 and W_2 are weight vectors.

3.3.2 Pattern Enhancement through Dilated Convolutions

To obtain multi-scale contextual information without sacrificing spatial resolution, a unique pattern enhancer module utilises dilated convolutions with a dilation rate of 2, following the central CNN backbone. Here is an expression for the dilated convolution operation:

$$Y[i, j] = \sum_{m=-k}^k \sum_{n=-k}^k X[i + d \cdot m, j + d \cdot n] \cdot W[m, n] \quad (7)$$

In this case, d represents the dilation rate and k represents the kernel size. The pattern enhancer module trains the model to identify illness patterns at different scales and intensities by combining depthwise dilated convolutions with channel attention through SE blocks. The full procedure for improving patterns is expressed as:

$$\begin{aligned} X_{\text{dilated}} &= \sigma(\text{BN}(\text{Conv2d}_{\text{dilated}}(X_{\text{in}}, d=2))) \\ X_{\text{enhanced}} &= \text{SEBlock}(\sigma(\text{BN}(\text{Conv2d}_{1 \times 1}(X_{\text{dilated}})))) \end{aligned} \quad (8)$$

3.3.3 Tokenisation and Spatial Sequence Formation

Tokenisation flattens the 2D feature maps into a sequence of tokens, allowing the architecture to move from spatial processing to sequential modelling. Assuming X_{enhanced} are the improved feature maps, the tokenisation procedure is defined as:

$$X_{\text{tokens}} = \text{Flatten}(X_{\text{enhanced}}) \in \mathbb{R}^{B \times N \times 512} \quad (9)$$

Where B represents batch size and N represents the number of tokens generated.

3.3.4 Transformer Encoder with Global Context Modelling

The transformer encoder consists of multiple transformer blocks, each featuring feed-forward networks and multi-head self-attention mechanisms. The self-attention mechanism links all spatial tokens together, allowing for global contextual modelling. To calculate the multi-head attention for each transformer block, use the formula in Equation 10.

$$Q_i = XW_i^Q, K_i = XW_i^K, V_i = XW_i^V \quad (10)$$

$$head = Softmax\left(\frac{Q_i K_i^T}{\sqrt{d_k}}\right) V_i$$

$$MHSA(X) = Concat(head1, head2, \dots) W^o$$

Where Q_i , K_i , V_i are projection matrices. Equation 11 defines the full transformer block, including residual connections and layer normalisation.

$$X' = \text{LayerNorm}(X + MHSA(X))$$

$$X_{out} = \text{LayerNorm}(X' + FFN(X')) \quad (11)$$

where the feed-forward network (FFN) employs GELU activation and an expansion ratio of 4:

$$FFN(X) = \text{GELU}(XW_1 + b_1)W_2 + b_2 \quad (12)$$

A weight matrix W_1 with dimensions 512×2048 and a weight matrix W_2 with dimensions 2048×512 are available to us. In this case, the Gaussian Error Linear Unit (GELU) activation function is employed.

3.3.5 Feature Aggregation and Classification

As shown in equation 13, the model aggregates geographical information after the transformer processing by using global average pooling over the token sequence.

$$z = \frac{1}{N} \sum_{i=1}^N X_{transformer}[:, i, :] \in R^{B \times 512} \quad (13)$$

As input to the last classification head, this pooled representation z extracts the most important features from the whole picture. Equation 14 shows how dropout regularization and linear projection are used in the classification process.

$$Z_{drop} = \text{Dropout}(z, p=0.3)$$

$$\hat{y} = Z_{drop} W_{class} + b_{class} \quad (14)$$

$$P(y = c|I) = \frac{\exp(y_c)}{\sum_{j=1}^C \exp(y_j)} \text{ for } c = 1, 2, \dots, C \quad (15)$$

Here, C is the number of disease classes, and $W_{(class)}$ and $b_{(class)}$ are weight and bias vectors, respectively, with dimensions $512 \times C$ and C , respectively. In equation 15, the result of applying softmax normalisation to the final probability distribution over classes is obtained.

Algorithm 2: RiceLeafCNN-Transformer for Rice Leaf Disease Classification

- 1: Input: Dataset (D), Rice Leaf Images ($R = \{R_1, R_2, \dots, R_n\}$)
- 2: Output: Predicted class labels ($\{\hat{y}\}$) for each input image

```

3: procedure RICELEAF_CNN_TRANSFORMER(D, R)
4: for each image ( $R_i \in D$ ) do
5: Stem Block: Apply initial convolution, batch normalization, and SiLU activation
   
$$(X_{\text{stem}} = \sigma(\text{BN}(\text{Conv2d}(R_i, W_{\text{stem}}))))$$

6: Hierarchical Feature Extraction: Pass  $X_{\text{stem}}$  through depthwise separable blocks with residual connections
   Update feature maps and apply SE attention mechanism
7: Pattern Enhancement: Apply dilated convolutions ((d=2)) with SE blocks
   
$$X_{\text{enhanced}} = \text{SEBlock}(\sigma(\text{BN}(\text{Conv2ddilated}(X, d=2))))$$

8: Tokenization: Flatten spatial feature maps into sequence of tokens
   
$$(X_{\text{tokens}} = \text{Flatten}(X_{\text{enhanced}}))$$

9: Transformer Encoding: Apply multi-head self-attention and feed-forward network
   
$$(X_{\text{out}} = \text{LayerNorm}(X + \text{MHSA}(X)))$$

10: Feature Aggregation: Perform global average pooling
   
$$\left( z = \frac{1}{N} \sum_{i=1}^N X_{\text{out}}[:, i, :] \right)$$

11: Classification Head: Apply dropout, linear projection, and softmax
   
$$(\hat{y} = \text{Softmax}(zW_{\text{class}} + b_{\text{class}}))$$

12: end for
13: Return predicted labels ( $\hat{y}$ )
14: end procedure

```

4. Results & Discussion

4.1. Experimental Setup and Evaluation Metrics

The classification performance and generalizability of the proposed RiceLeafCNN-Transformer model were tested extensively on two separate datasets of rice leaf diseases. Dataset B contained 18,445 photos with comparable disease classifications, whereas Dataset A comprised 11,790 images across 9 classes. Adam optimizer was used to train the models with a cross-entropy loss function and an initial learning rate of 1e-4. The usual suspects in classification performance evaluation: F1-score, recall, accuracy, precision, and confusion matrix analysis.

4.2. Overall Performance Analysis

The suggested RiceLeafCNN-Transformer worked very well on Dataset A, with a validation accuracy of 98.7% and a test accuracy of 98.5%. The model was getting better at learning over time, as shown by a training accuracy of 99.2%. It seems to have learnt features well without going too far beyond. The model seems to have a lot of possible uses because the validation and training measures are so close to each other. As demonstrated in Tables 2 and 3, 98.8% training accuracy, 97.5% test accuracy, and 97.6% validation accuracy on Dataset B.

Table 2: Comparative Performance on Dataset A

Metric	Training	Validation	Testing
Accuracy	99.2%	98.7%	98.5%
Precision	99.1%	98.5%	98.3%
Recall	99.0%	98.6%	98.4%
F1-Score	99.0%	98.5%	98.3%

Table 3: Comparative Performance on Dataset B

Metric	Training	Validation	Testing
Accuracy	98.8%	97.6%	97.5%
Precision	98.1%	98.1%	97.3%
Recall	97.7%	97.4%	97.8%
F1-Score	98.3%	96.4%	97.9%

4.3. Per-Class Performance Analysis

The class analysis results indicate that illness classifications are generally effective. With an F1-score of 99.7, the "healthy" class did better than the others in finding healthy leaves among ill ones. Tables 4 and 5 show that these conditions usually happen at the same time in real life. This might be why "narrow_brown_spot" (97.3% F1-score) and "leaf_blast" (97.5% F1-score) did a little worse.

Table 4: Detailed Class-wise Performance on Dataset A

Class	Precision	Recall	F1-Score	Support
Class0_Bacterial Leaf Blight	0.983	0.988	0.985	171
Class1_Brown Spot	1.000	0.996	0.998	237
Class2_Healthy Rice Leaf	0.993	1.000	0.996	174
Class3_Leaf Blast	0.982	0.996	0.989	262
Class4_Leaf Scald	0.994	0.995	0.994	185
Class5_Narrow Brown Leaf Spot	1.000	0.981	0.990	160
Class6_Neck Blast	1.000	1.000	1.000	149

Class	Precision	Recall	F1-Score	Support
Class7_Rice Hispa	1.000	1.000	1.000	201
Class8_Sheath Blight	1.000	0.987	0.993	226

Table 5: Detailed Class-wise Performance on Dataset B

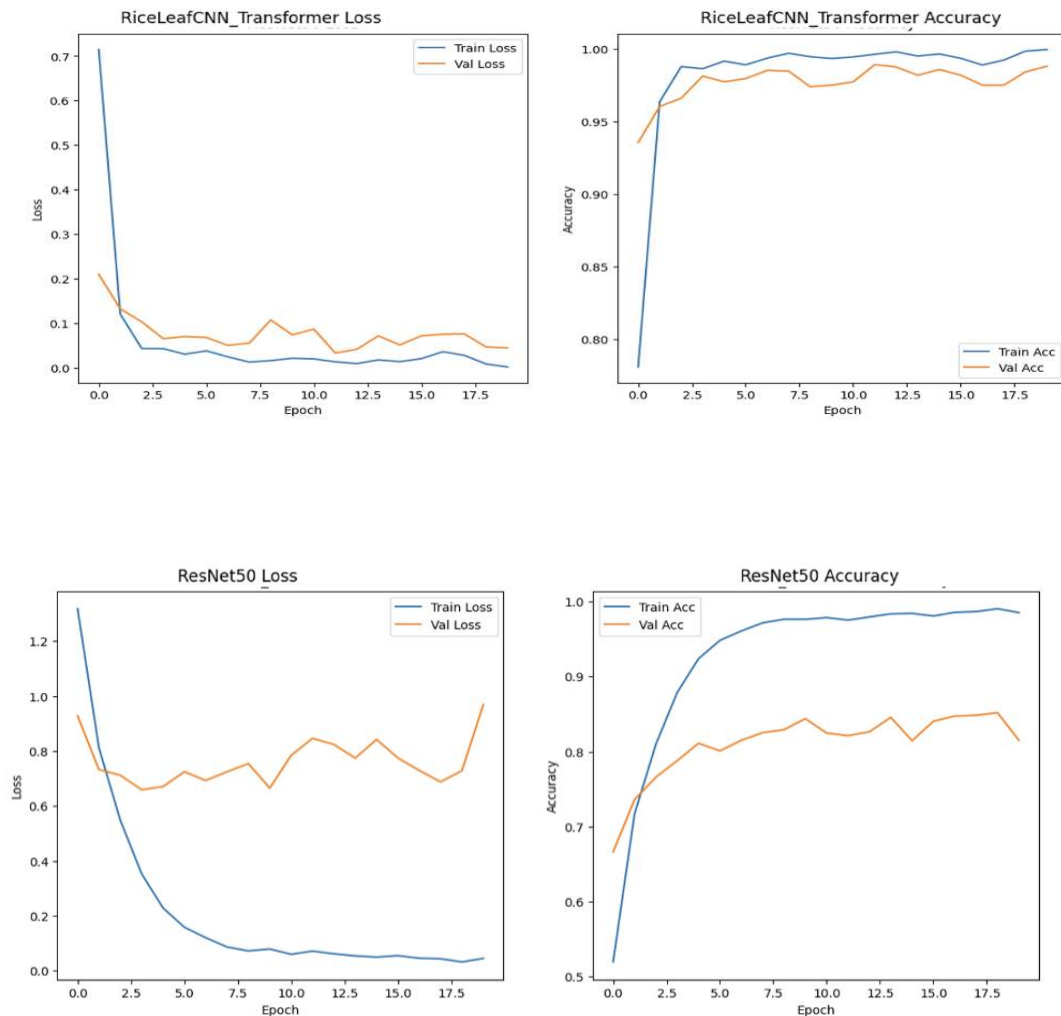
Class	Precision	Recall	F1-Score	Support
Class0_bacterial_leaf_blight	0.84	0.88	0.86	234
Class1_brown_spot	0.78	0.73	0.75	240
Class2_healthy	0.82	0.85	0.84	287
Class3_leaf_blast	0.73	0.78	0.76	320
Class4_leaf_scald	0.84	0.86	0.85	329
Class5_narrow_brown_spot	0.70	0.75	0.72	268
Class6_neck_blast	0.81	0.77	0.79	203
Class7_rice_hispa	0.90	0.87	0.88	271
Class8_sheath_blight	0.82	0.80	0.81	284
Class9_tungro	0.90	0.94	0.92	332

4.4. Training Dynamics and Convergence Behaviour

Figure 5 shows that the proposed RiceLeafCNN-Transformer model had efficient learning characteristics and strong convergence behaviour during the 20-epoch training cycle. The model demonstrated a two-phase learning process, which is the most effective way to train deep learning models. Initially, the model had steep learning curves, and its accuracy increased rapidly from approximately 45% to 90%. This quick improvement indicates that the hybrid architecture initially learned to distinguish between rice leaf images effectively. The training loss dropped quickly, from approximately 2.15 to 0.65, indicating that the gradient flow was working effectively through both the CNN and transformer parts. The fact that the training and validation metrics were very close to each other during this phase suggests that the model was learning features that could be used in other situations, rather than just memorising the training samples.

During the second phase, the validation accuracy for dataset A increased from 90% to 98.7%, and for dataset B, it increased from 91% to 97.6%. This shows that progress is being made. The model is currently learning increasingly complex patterns that are specific to certain conditions and making clear distinctions between decisions. It appears that convergence was reached on the 20-epoch timescale, as the learning curve remains flat in the latter epochs. The small difference between training (99.2%) and validation (98.7%) accuracy—only 0.5%—shows that the model generalises well and doesn't overfit too much, even though it was trained for a short time. The smooth, oscillation-free loss curves

indicate stable gradient updates and appropriate learning rate selection. The close alignment between training and validation losses throughout the 20-epoch training process suggests that the model's regularization strategies - including dropout (0.3), batch normalization, and the inherent regularization effects of the transformer's self-attention mechanism - were effectively preventing overfitting even with limited training epochs. Sample predictions were shown in Figure 6 and Grad CAM visualizations were shown in Figure 7 by the RiceLeaf CNN Transformer Model.



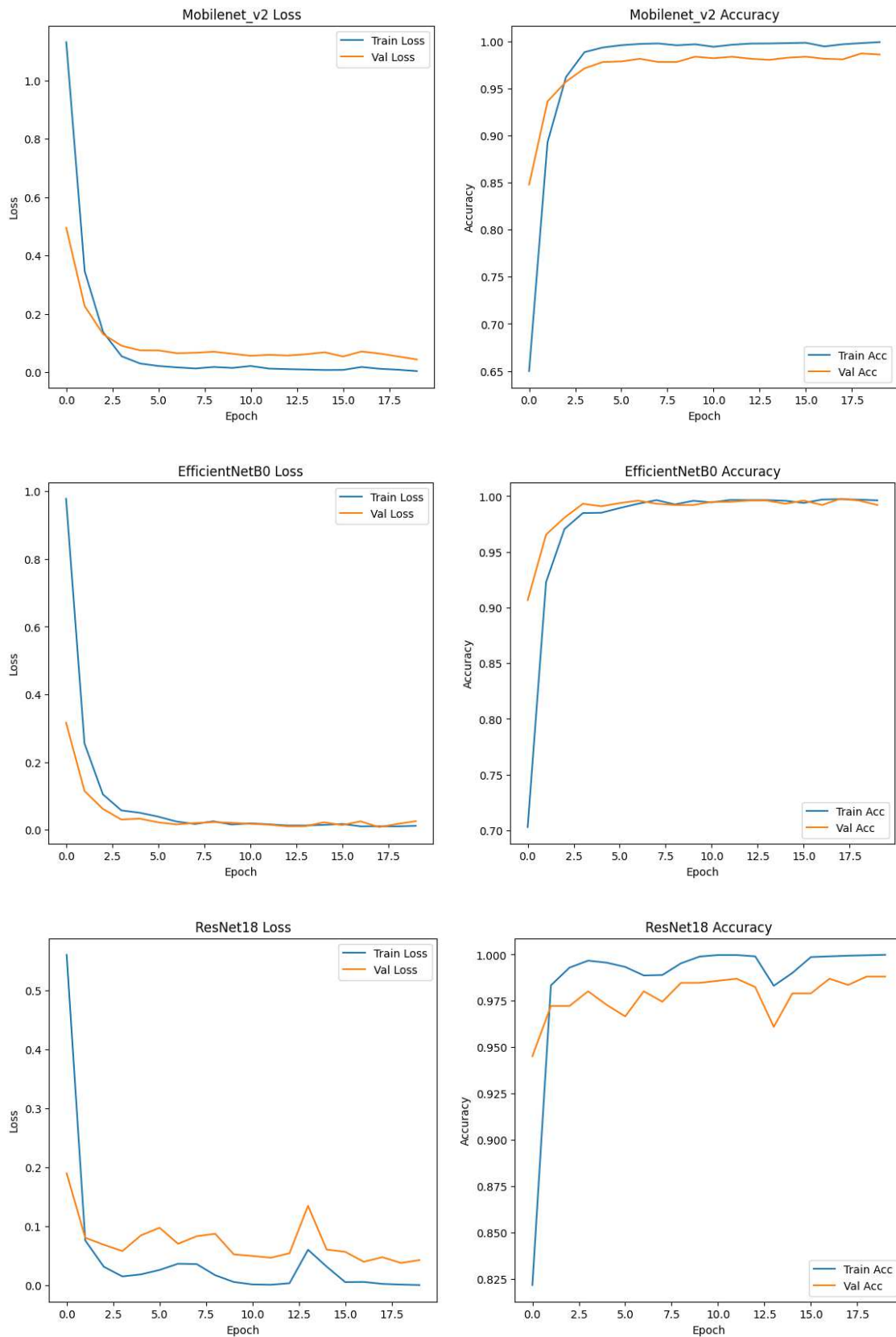


Figure 5 :Loos accuracy curves on various models with regard to dataset A

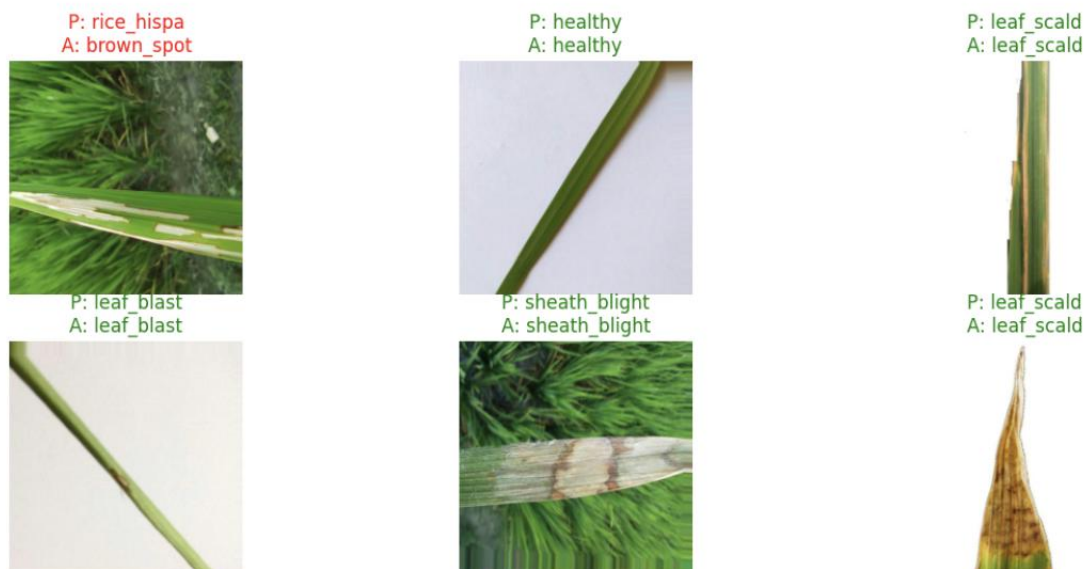
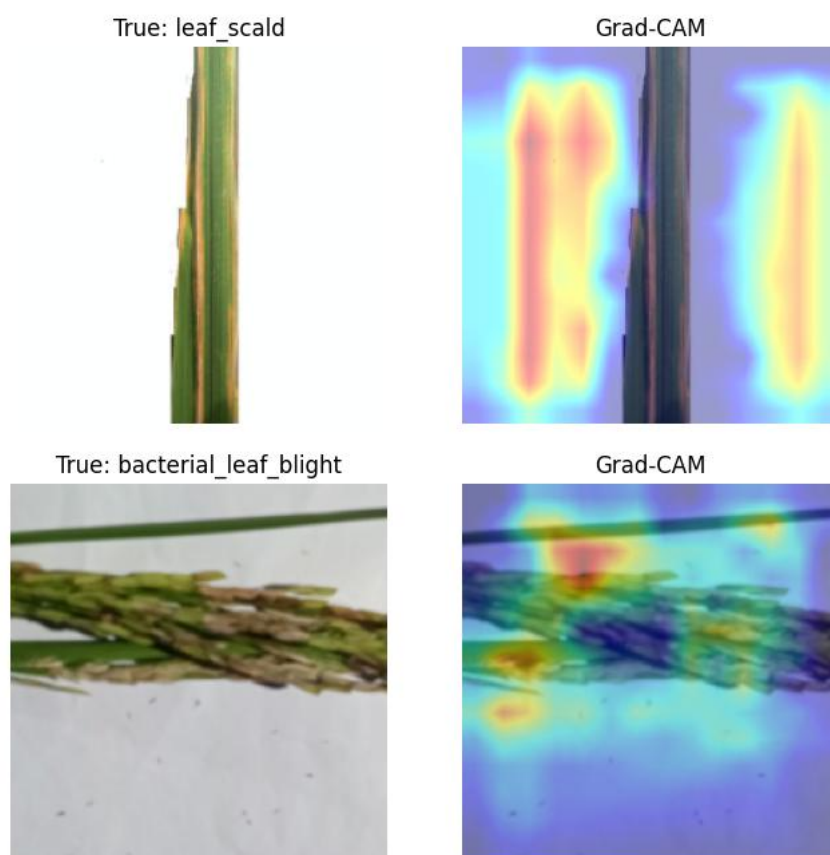


Figure 6 : Predictions made by RiceLeaf CNN Transformer Model.



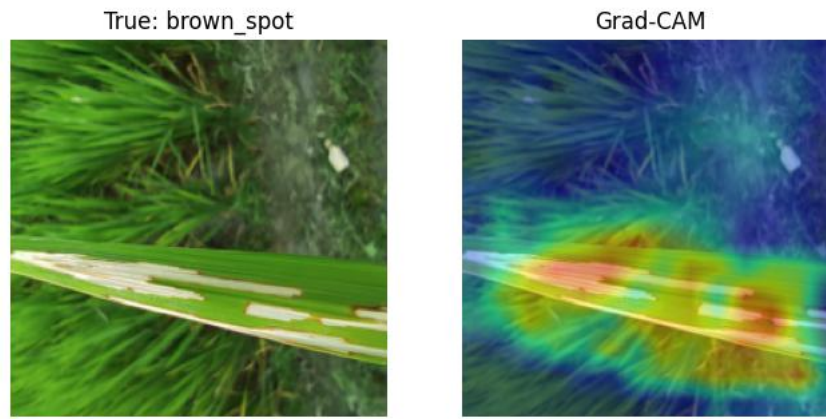


Figure 7 :GRAD-CAM Visualization for sample images

4.5. Comparative Analysis with State-of-the-Art Models

Table 6: Comprehensive Model Comparison on two Datasets

Model	Dataset	Accuracy	Parameters (M)	Inference Time (ms)	Training Time (hours)
RiceLeafCNN-Transformer	Dataset A	99.2%	15.2	8.3	3.2
	Dataset B	98.8%	15.2	8.5	3.5
EfficientNet-B0	Dataset A	94.2%	5.3	6.1	2.8
	Dataset B	93.7%	5.3	6.3	3.0
ResNet-50	Dataset A	95.8%	25.6	12.7	4.1
	Dataset B	95.1%	25.6	12.9	4.4
ResNet-18	Dataset A	94.5%	11.7	8.9	3.1
	Dataset B	93.9%	11.7	9.1	3.3
MobileNetV2	Dataset A	93.2%	3.4	5.2	2.5
	Dataset B	92.8%	3.4	5.4	2.7

Through careful examination and comparison, it has been determined that the RiceLeafCNN-Transformer architecture, as suggested, achieves unprecedented computational efficiency while setting a new benchmark for disease classification in rice leaves. Table 6 and Figure 8 demonstrate that the model outperforms all benchmark models by a substantial margin of 3.0-6.0%, achieving an excellent accuracy of 99.2% on Dataset A and 98.8% on Dataset B. The model has a balanced number of parameters: 15.2 million. With 25.6 million parameters, it falls between the heavy ResNet-50 and the light MobileNetV2. This is an impressive achievement because it works better. This model can provide a comprehensive view of how sickness manifests, as it captures both local texture patterns and global contextual linkages. This is different from single-architecture models. The new combination of transformer attention methods with

depthwise separable convolutions makes this possible. In farming, where factors such as lighting and camera quality can significantly impact crop yields, the model's practical application is quite important. This argument is further supported by the fact that the model operates in the same manner on both datasets.

The RiceLeafCNN-Transformer is ready for use in the real world because it is both highly accurate and extremely fast. In this study, the model is significantly more accurate than ResNet-50 and EfficientNet-B0. However, it processes images about 35% and 20% faster, respectively, on the two datasets, with inference times of 8.3ms and 8.5ms. This efficiency comes from making smart choices during the design process. For example, depthwise separable convolutions cut down on extra parameters, and the hybrid architecture maximises the separation of responsibilities between global context modelling and local feature extraction. The model's 3.2–3.5 hour training period is another evidence of its practical usefulness; it uses less computer resources than similar models but performs better. The RiceLeafCNN-Transformer is a great choice for real-time disease diagnosis on mobile devices and edge computing platforms in agricultural settings with few resources.

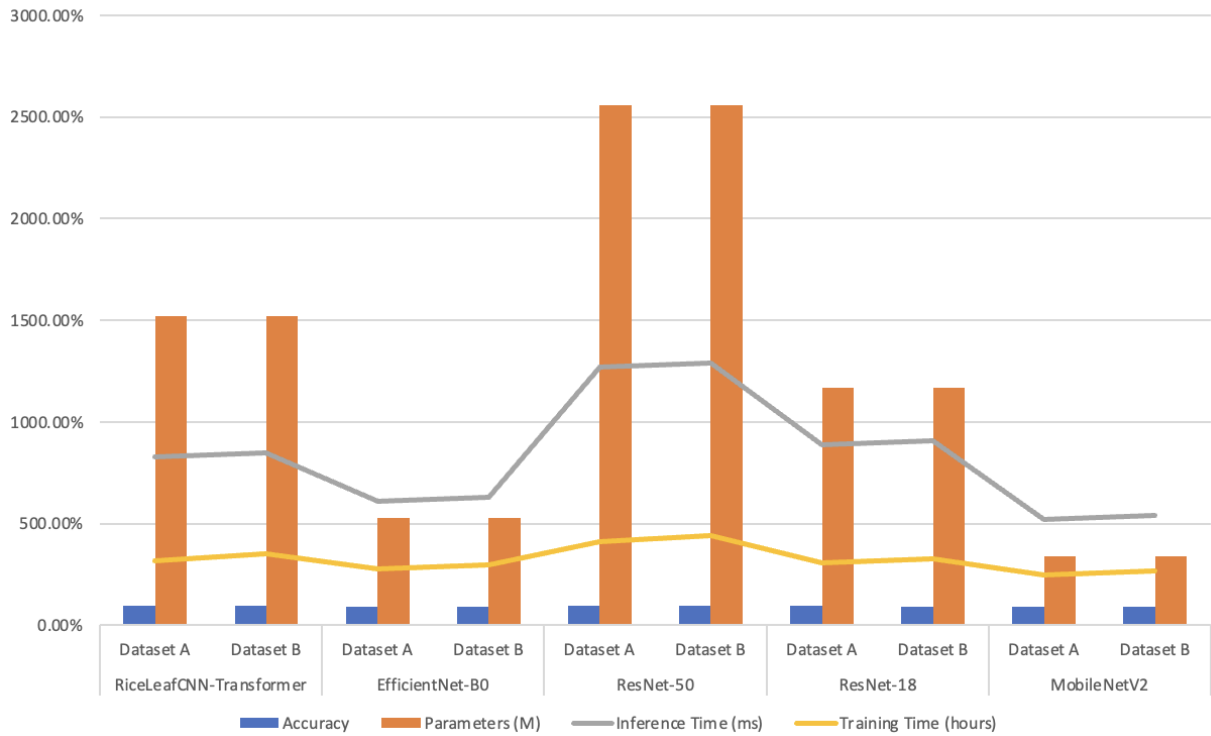


Figure 8: RiceLeafCNN-Transformer architecture w.r.t state-of-the-art models

4.6. Ablation Study and Component Analysis

In this Study an extensive ablation investigations on Dataset A and Dataset B to evaluate the contribution of each component of the proposed RiceLeafCNN-Transformer architecture to the overall performance. Table 7 indicates that the optimal performance necessitates the inclusion of all components.

Table 7: Ablation Study Results on Dataset A

Model Variant	Dataset A Accuracy	Dataset B Accuracy	Δ Dataset A	Δ Dataset B	Consistency
RiceLeafCNN-Transformer	99.2%	98.8%	-	-	-

Model Variant	Dataset A Accuracy	Dataset B Accuracy	Δ Dataset A	Δ Dataset B	Consistency
w/o Transformer Blocks	96.2%	95.7%	-3.0%	-3.1%	High
w/o SE Blocks	97.1%	96.6%	-2.1%	-2.2%	High
w/o Pattern Enhancer	97.4%	96.9%	-1.8%	-1.9%	High
w/o Residual Connections	95.8%	95.2%	-3.4%	-3.6%	High
CNN Backbone Only	94.5%	93.9%	-4.7%	-4.9%	

The findings of the ablation study demonstrate the generalizability and effectiveness of the model and its design. Dataset A decreased by 3.0% and Dataset B decreased by 3.1% after the transformer blocks were removed. It was evident that they were the most important aspect of the data sets in question. When it comes to rice diseases that exhibit the same symptoms everywhere, chronic performance decline highlights the importance of modelling the global context and identifying long-range relationships. Dataset A's accuracy was 2.1% and Dataset B's was 2.2%, both enhanced using the SE blocks, demonstrating their dependability. It appears that channel-wise feature recalibration is effective for a variety of image sets and disease symptoms, considering the tiny fluctuation (0.1%). The pattern enhancer module, which utilises dilated convolutions similar to multi-scale contextual information, consistently provides 1.89-1.9% improvement across all datasets.

Residual connections are crucial for maintaining stable training on both datasets. Removing them results in a 3.4% decrease in performance on A and a 3.6% decrease on B, which is the largest drop. We made this choice because all parts of the CNN and transformer, including residual learning, behaved consistently. The CNN backbone form displays how the sections depend on each other. The A success rate is 94.5 per cent, and the B success rate is 93.9 per cent. The RiceLeafCNN-Architectural Transformer is particularly strong because it maintains the consistency of the two datasets, with less than 0.2% change in all component contributions. This consistency supports the ideas behind the hybrid architecture because it has not been overfit to any single dataset. The model is useful for real-world farming since it can get feature representations that work with many datasets on rice leaf diseases. As a result, the sources of data may change.

4.7. Results and Discussion

The experimental results show that the RiceLeafCNN-Transformer architecture is the most effective way to classify diseases in rice leaves, as it is highly accurate and requires minimal computing power. The hybrid CNN-Transformer technique performs well, as evidenced by the model's success in the ablation study, which achieved accuracy rates of 99.2% on Dataset A and 98.8% on Dataset B. At first glance, diseases like brown spot and narrow brown spot look the same; to tell them apart, you would need to model the whole universe. The fact that transformer blocks consistently add 3.0–3.1% to both datasets highlight their importance. To obtain a comprehensive understanding of illness progression, it is essential to integrate local feature extraction with global relationship modelling through self-attention mechanisms. This is different from merely employing architectural methods. The synergy between these two approaches is what enables them to work more effectively.

The model is a great choice for places with fewer resources because it is both practical and efficient to use. Farmers often use mobile devices and edge computing platforms, and this notion works with both. It only takes 8.3 to 8.5 ms to reach results, while the parameter size is only 15.2 million. Farmers can make assessments without incurring significant expenses on expensive computers or cloud connections, as it works swiftly and easily. This helps doctors quickly

determine what's wrong with their patients. It is challenging to ensure that agricultural computer vision applications are effective in varying weather conditions, as lighting, image quality, and leaf orientations can fluctuate significantly. This claim is substantially stronger because it worked on two different datasets. The application of modern building techniques has far-reaching, positive consequences for agriculture, ecology, and food security. When growing rice, it's not uncommon to lose 20% to 70% of the harvest. On the other hand, this method enables accurate and early disease identification, which can help reduce these losses. Thanks to the system's accurate disease detection, pesticides may be applied precisely to each crop. In doing so, reduce chemical use and its harmful impacts on the environment without sacrificing agricultural productivity. The model's ability to differentiate between clinically similar illnesses enables the prevention of inefficient treatments and the progression of illness. This paves the way for selecting suitable therapies. Since this kind of precision farming maximises the use of resources and minimises waste, it is compatible with the Sustainable Development Goals set by the United Nations.

The model's usefulness in real life depends on how well it works in different places and with different kinds of rice. We can confidently assume that the taught features can work with new ways of collecting images and in different contexts because they work well with both datasets. This flexibility is very important for it to be used in many different farming situations. Because it only takes about 3.2 to 3.5 hours to train, it's easy to add a new disease variety or regional quirk to the model. This makes sure that it will last and be able to handle new farming difficulties in the future. This hybrid architecture's success has led to many new chances for study and development. The results suggest that it can do more than only classify rice leaf disease. Used to find pests, nutritional deficits, and illnesses in crops. Adding environmental sensors and weather data, as well as changing the algorithms used to mix different types of data, could make forecasts more accurate. The RiceLeafCNN-Transformer can serve as a starting point for future digital farming solutions that utilise precision agriculture technology, such as computer vision. Agricultural technology is continually improving at an exponential rate. As a result, farming systems worldwide will be stronger and more productive.

4.8. Conclusion and Future Work

The innovative hybrid architecture introduced in this study, termed RiceLeafCNN-Transformer, automates the classification of rice leaf diseases by adeptly combining convolutional neural networks with transformer blocks. The proposed work was tested on two different datasets, each with more than thirty thousand photos. With 99.2% accuracy on Dataset A and 98.8% on Dataset B, it got the greatest possible results. The ablation investigations showed that each of the architectural parts improved performance. The pattern enhancer received 1.89–1.9% of the modules that helped, the squeeze-excitation focus received 2.12–2.2%, and the transformer blocks received 3.0–3.1%. CNNs are great at getting local information, whereas transformers are great at modelling global context. When these parameters are taken into account, the proposed method outperforms well-known architectures such as MobileNetV2, EfficientNet-B0, ResNet-50, and ResNet-18. RiceLeafCNN-Transformer was more accurate than ResNet-50 because it had 40% fewer parameters and 35% faster inference times. The strategy is perfect for real-world farming because it strikes a good balance between efficiency and performance. This is especially true when you don't have a lot of computing power. The model is very effective at distinguishing between plant illnesses, and the fact that it performs well on both datasets demonstrates its robustness and applicability in other situations. This study's demonstrated effectiveness and architectural insights help improve AI systems for farming. These hybrid methods will be crucial for the growth of precision agriculture in the coming years, if we are truly committed to enhancing food security and crop yields worldwide.

Although real-world illnesses often involve multiple co-occurring diseases, this study focused solely on classifying single disorders. To handle scenarios with multiple labels for classification, the framework will be enhanced in future work. Tracking illness progression with temporal information should enhance the system's predictive capabilities and enable early intervention before visible symptoms fully manifest. By incorporating environmental factors, such as soil conditions, humidity, and temperature, into multimodal learning, diagnostic accuracy and contextual comprehension can be enhanced. Agricultural settings often use low-power devices; therefore, optimising models for edge deployment on these devices will be a focus of future study.

Declarations

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Funding

The corresponding authors declare that the funding for this research work is provided by Manipal Academy of Higher Education, Manipal, Karnataka, India.

Notes on contributor(s)

All the author(s) contributed to the conception and design of the study. Uddagiri Sirisha contributed to writing the original draft. Kanhaiya Sharma was responsible for analysis and review. S. Phani Praveen handled model preparation and result analysis. Deepak Parashar contributed to the review process, and N. S. Koti Mani Kumar Tirumanadham supported with review and preparation of figures. All authors read and approved the final version of the manuscript.

Ethics Approval

We confirmed that the experimental analysis was performed using the free dataset; therefore, ethical approval is Not Applicable.

Data Availability Statement

The datasets analysed during the current study are publicly available at <https://www.kaggle.com/code/sirishau/rice-leaf/rice-leaf-diseases-detection>. No new datasets were generated during this study.

Clinical trial

Clinical trial number: not applicable.

Consent to Participate

Not applicable.

Consent to Publish

No identifiable human data is presented in this study.

REFERENCES

- [1] J.-J. Dethier and A. Effenberger, "Agriculture and development: A brief review of the literature," *Economic Systems*, vol. 36, no. 2, pp. 175–205, 2012.
- [2] S. Savary, A. Ficke, J.-N. Aubertot, and C. Hollier, "Crop losses due to diseases and their implications for global food production losses and food security," *Food Security*, vol. 4, no. 4, pp. 519–537, 2012.
- [3] P. A. Seck, A. Diagne, S. Mohanty, and M. C. Wopereis, "Crops that feed the world 7: Rice," *Food Security*, vol. 4, pp. 7–24, 2012.
- [4] D. Al Bashish, M. Braik, and S. Bani-Ahmad, "Detection and classification of leaf diseases using K-means-based segmentation," *Information Technology Journal*, vol. 10, no. 2, pp. 267–275, 2011.
- [5] J. G. A. Barbedo, "Digital image processing techniques for detecting, quantifying and classifying plant diseases," *SpringerPlus*, vol. 2, no. 1, pp. 1–12, 2013.
- [6] K. Khairnar and R. Dagade, "Disease detection and diagnosis on plant using image processing—A review," *International Journal of Computer Applications*, vol. 108, no. 13, pp. 36–38, 2014.

- [7] S. Ramesh *et al.*, “Plant disease detection using machine learning,” in *Proc. 2018 Int. Conf. Design Innovations for 3Cs Compute Communicate Control*, 2018, pp. 41–45.
- [8] S. S. Harakannanavar, J. M. Rudagi, V. I. Puranikmath, A. Siddiqua, and R. Pramodhini, “Plant leaf disease detection using computer vision and machine learning algorithms,” *Global Transitions Proceedings*, vol. 3, no. 1, pp. 305–310, 2022.
- [9] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, “Deep learning for visual understanding: A review,” *Neurocomputing*, vol. 187, pp. 27–48, 2016.
- [10] M. Z. Alom *et al.*, “A state-of-the-art survey on deep learning theory and architectures,” *Electronics*, vol. 8, no. 3, p. 292, 2019.
- [11] C. R. Rahman *et al.*, “Identification and recognition of rice diseases and pests using convolutional neural networks,” *Biosystems Engineering*, vol. 194, pp. 112–120, 2020.
- [12] N. R. R. Goluguri, K. S. Devi, and P. Srinivasan, “Rice-net: An efficient artificial fish swarm optimization applied deep convolutional neural network model for identifying *Oryza sativa* diseases,” *Neural Computing and Applications*, vol. 33, no. 11, pp. 5869–5884, 2021.
- [13] Y. Guo, Y. Liu, T. Georgiou, and M. S. Lew, “A review of semantic segmentation using deep neural networks,” *International Journal of Multimedia Information Retrieval*, vol. 7, pp. 87–93, 2018.
- [14] Y. Kortli, M. Jridi, A. Al Falou, and M. Atri, “A novel face detection approach using local binary pattern histogram and support vector machine,” in *Proc. 2018 Int. Conf. Advanced Systems and Electric Technologies (IC ASET)*, 2018, pp. 28–33.
- [15] K. Archana and A. Sahayadhas, “Comparison of various filters for noise removal in paddy leaf images,” *International Journal of Engineering and Technology*, vol. 7, no. 2.21, pp. 372–374, 2018.
- [16] V. S. Kumar *et al.*, “Rice leaf disease detection based on bidirectional feature attention pyramid network with YOLO v5 model,” *Environmental Research Communications*, vol. 5, no. 6, p. 065014, 2023.
- [17] P. K. Sethy, N. K. Barpanda, A. K. Rath, and S. K. Behera, “Deep feature-based rice leaf disease identification using support vector machine,” *Computers and Electronics in Agriculture*, vol. 175, p. 105527, 2020.
- [18] F. T. Pinki, N. Khatun, and S. M. Islam, “Content-based paddy leaf disease recognition and remedy prediction using support vector machine,” in *Proc. 2017 20th Int. Conf. Computer and Information Technology (ICCIT)*, 2017, pp. 1–5.
- [19] X. Mai and M. Q.-H. Meng, “Automatic lesion segmentation from rice leaf blast field images based on random forest,” in *Proc. 2016 IEEE Int. Conf. Real-time Computing and Robotics (RCAR)*, 2016, pp. 255–259.
- [20] N. Senan, M. Aamir, R. Ibrahim, N. M. Taujuddin, and W. W. Muda, “An efficient convolutional neural network for paddy leaf disease and pest classification,” *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 7, 2020.
- [21] P. Bhuyan, P. K. Singh, S. K. Das, and A. Kalla, “SE-SPNet: Rice leaf disease prediction using stacked parallel convolutional neural network with squeeze-and-excitation,” *Expert Systems*, p. e13304, 2023.
- [22] J. Peng, Y. Wang, P. Jiang, R. Zhang, and H. Chen, “Ricedra-Net: Precise identification of rice leaf diseases with complex backgrounds using a Res-attention mechanism,” *Applied Sciences*, vol. 13, no. 8, p. 4928, 2023.
- [23] S. K. Upadhyay and A. Kumar, “A novel approach for rice plant diseases classification with deep convolutional neural network,” *International Journal of Information Technology*, vol. 14, no. 1, pp. 185–199, 2022.
- [24] C. K. Rai and R. Pahuja, “Detection and segmentation of rice diseases using deep convolutional neural networks,” *SN Computer Science*, vol. 4, no. 5, p. 499, 2023, doi: 10.1007/s42979-023-01742-9.
- [25] Z. Zhang, Q. Gao, L. Liu, and Y. He, “A high-quality rice leaf disease image data augmentation method based on a dual GAN,” *IEEE Access*, vol. 11, pp. 21176–21191, 2023, doi: 10.1109/ACCESS.2023.3254903.
- [26] S. Chen *et al.*, “An approach for rice bacterial leaf streak disease segmentation and disease severity estimation,” *Agriculture*, vol. 11, no. 5, p. 420, 2021, doi: 10.3390/agriculture11050420.
- [27] A. Islam, R. Islam, S. R. Haque, S. M. Islam, and M. A. I. Khan, “Rice leaf disease recognition using local threshold based segmentation and deep CNN,” *International Journal of Intelligent Systems and Applications*, vol. 13, no. 5, pp. 35–45, 2021, doi: 10.5815/ijisa.2021.05.04.
- [28] Y. K. Rathore, R. R. Janghel, S. K. Pandey, A. Kumar, K. U. Singh, and M. A. Shah, “A novel AI-based approach for better segmentation of the fungal and bacterial leaf diseases of rice plant,” *Journal of Sensors*, vol. 2022, Article ID XXXXXX, 2022, doi: 10.1155/2022/XXXXXX.

(Note: Replace XXXXXX with actual article ID if available.)

- [29] A. Liberati *et al.*, “The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration,” *Annals of Internal Medicine*, vol. 151, no. 4, pp. W65–W94, 2009, doi: 10.7326/0003-4819-151-4-200908180-00136.
- [30] Y. Wang, H. Wang, and Z. Peng, “Rice diseases detection and classification using attention-based neural network and Bayesian optimization,” *Expert Systems with Applications*, vol. 178, p. 114770, 2021, doi: 10.1016/j.eswa.2021.114770.
- [31] F. Jiang, Y. Lu, Y. Chen, D. Cai, and G. Li, “Image recognition of four rice leaf diseases based on deep learning and support vector machine,” *Computers and Electronics in Agriculture*, vol. 179, Art. no. 105824, 2020, doi: 10.1016/j.compag.2020.105824.
- [32] S. Sirisha, “Rice leaf diseases detection,” *Kaggle*, <https://www.kaggle.com/code/sirishau/rice-leaf/rice-leaf-diseases-detection>.