Multimedia Tools and Applications (2025) 84:48145–48173
https://doi.org/10.1007/s11042-025-20979-9

RiceNet: a robust ensemble attention mechanism for automated rice plant disease classification

Muhammad Hanif Tunio[1,2] · **Jian ping Li**[1] · **Awais Ahmed**[3] · **Syed Attique Shah**[4] · **Xiaoyang Zeng**[1] · **Ubaidullah alias Kashif**[5] · **Yingling Li**[6] · **Abdullahi Yahya Imam**[7]

Received: 7 May 2024 / Revised: 2 April 2025 / Accepted: 6 June 2025 /

Abstract

Rice is a widely cultivated crop in Asia and is paramount in ensuring national and global food security. However, rice plants are susceptible to various diseases that negatively impact crop quality and quantity to meet the needs of the world's growing population. Automated rice plant disease classification ensures food security and agricultural sustainability. Although traditional deep learning approaches have shown promising results in rice plant disease classification, the challenges posed by the heterogeneity of the data set and the feature imbalance persist. This research introduces a robust and novel Ensemble Attention Mechanism (EAM) that uses fine-tuning transfer learning to address these challenges and pre-trained (VGG16, VGG19, and customized ResNet called RiceNet, which comprises ResNet18 and ResNet50) as baseline models, specifically tailored to improve rice plant disease classification within heterogeneous datasets. The main contribution of this paper is to introduce a RiceNet framework that incorporates ensemble learning principles and attention mechanisms to adaptively balance data heterogeneity and feature representation by effectively integrating every representation to mitigate inherent class distribution imbalances. Comprehensive ablation studies validate the effectiveness of each component in the framework, demonstrating significant improvements in classification performance compared to traditional methods. Furthermore, the evaluation of RiceNet on two extensive publicly available datasets (close environment and field environment) shows its superior performance, achieving an impressive F1 score of 100% and a balanced precision of 100% on both large and small datasets. This research sets a new benchmark for rice disease classification and provides a versatile framework applicable to agricultural precision, contributing to food security and sustainability.

Keywords Rice disease classification · RiceNet · Ensemble attention mechanism · Transfer learning

---

Extended author information available on the last page of the article

1 Introduction

Rice is globally considered an important food crop, and it significantly feeds more than half the world's population [19]. China is at the forefront of rice consumption and production, covering 35.6% of the total cultivated land area for food crops [45]. At the same time, India has emerged as a formidable rice harvester. Vietnam and Bangladesh closely follow, contributing significantly to the global rice market [38], Fig. 1 shows the statistics of rice consumption by country. Rice plants are affected by various diseases that severely impact the quality and quantity of rice production [3, 37]. Approximately 30 different types of plant diseases that adversely affect rice plants have been identified, leading to a decline in the quantity and quality of the rice yield. Bacterial, leaf smut, brown spot, rice blast, bacterial blight, and fungal infections are the primary culprits behind these plant diseases [5]. Detecting these diseases at an early stage is crucial for minimizing crop losses and ensuring the well-being of plants, which is of utmost importance in attaining food sustainability and security within the agricultural sector.

Traditionally, farmers heavily relied on personal visual observation experience or experimental methods with the guidance of experts to diagnose plant diseases [30]. However, these conventional approaches are time-consuming and can lead to the further spread of plant diseases. Additionally, this manual approach often hindered the timely selection and application of suitable pesticides. Farmers had to invest significant resources in observing and identifying symptoms, comparing them with limited reference materials, and drawing on their own experiences to make informed decisions. This subjective and labor-intensive method often delays identifying the specific disease affecting their crops and subsequently choosing the appropriate course of action, resulting in significant losses for farmers. Therefore, early detection of diseases is crucial to minimize disease spread and overall loss [22].

Automatic plant disease detection is pivotal in ensuring plant quality, which is crucial for promoting high crop production. Recent advancements in deep learning networks and image processing techniques have opened new avenues for enhancing crop protection [3, 29, 42].

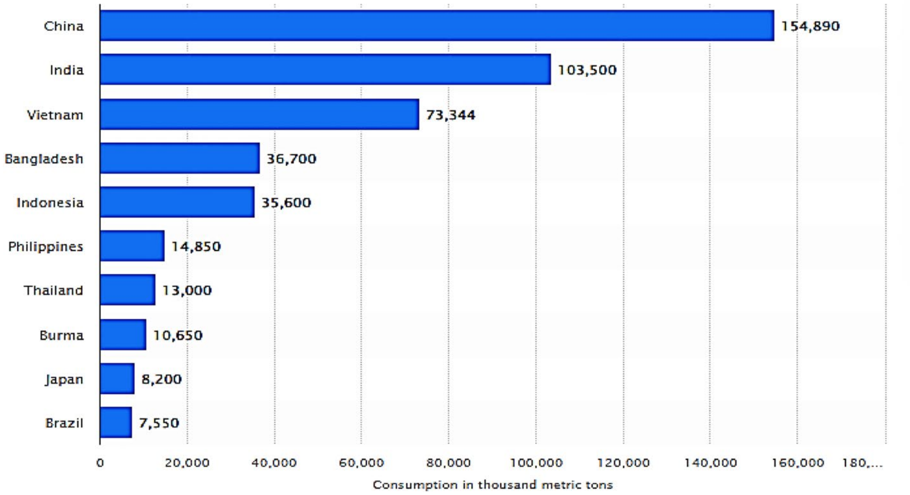


Fig. 1 Country-wise Rice Consumption as 1^{st} April-2022 [38]

Table 1 Summary of deep learning research in rice disease detection

Ref.	YoP	Dataset	Model	Metrics	Advantage	Limitation
[21]	2022	Custom dataset of rice plant images	Improved CNN	Accuracy, Precision	High disease detection accuracy	Limited by the specificity of dataset conditions
[24]	2017	Public and proprietary rice disease images	Deep CNN	Accuracy, F1 Score	Efficient processing of large image sets	Generalization issue across different rice varieties
[26]	2021	Field-acquired dataset of paddy leaves	Optimized DNN	Precision, Recall	Optimized for mobile deployment	Computationally intensive
[7]	2021	Real-time captured images of rice leaves	Faster R-CNN	Real-time detection rate	Suitable for real-time applications	Requires high-performance computing resources
[15]	2020	Diverse dataset including common rice diseases	VGG16 with transfer learning	Accuracy, Speed	Transfer learning enhances model efficiency	Limited by the diversity of the transfer source dataset
[20]	2020	Coffee leaf dataset	CNN with transfer learning	Accuracy	Demonstrate the effectiveness in plant diseases	Depends on large amounts of labeled data for training
[10]	2023	Mixed dataset from various sources	Rider Water Wave-enabled deep learning	Accuracy, Efficiency	Innovative approach to model training	Novelty of method may limit diverse datasets and comparative analysis
[12]	2023	High-resolution images of brown spot disease	VGG19	Accuracy, Specificity	High specificity for brown spot disease	Not as effective for other rice diseases
[27]	2023	Smartphone-captured rice plant images	DenseNet201, Xception, MNetV2, ResNet50	Usability, Accuracy	Accessible for non-experts	Image quality variability impacts performance
[16]	2021	Images of Oryza sativa with various diseases	Deep CNN with fish swarm optimization	Accuracy, Efficiency	High accuracy with computational efficiency	Generalizability and scalability and Optimization process complexity
[44]	2021	Rice leaf images	Attention-based NN with Bayesian optimization	Accuracy, Generalization	Improved disease classification accuracy	Bayesian optimization requires careful tuning
[8]	2021	Large-scale dataset of diseased and healthy rice plants	Lightweight attention networks	Accuracy	Efficient on edge devices	limited diversity of the real-life image dataset
[14]	2024	Three benchmark datasets	Multi-Level Handcrafted Feature Extraction, SVM	Accuracy	Handles dataset imbalance, improves feature representation, enhances classification accuracy	Relies on handcrafted features, lacks deep learning feature extraction, dataset generalizability.
[2]	2024	Three benchmark datasets	Multi-Level Handcrafted Feature Extraction, Color Correlogram, SVM	Accuracy	Combines texture and color features, achieves high accuracy, and addresses illumination problems	Lacks deep learning-based feature extraction, dataset dependence may limit generalizability

Table 1 (continued)

Ref.	YoP	Dataset	Model	Metrics	Advantage	Limitation
[35]	2025	PlantVillage, PlantDoc, Turkey Disease, Rice Disease, RoCole, NLB	IX-CNN (Inception-Xception CNN)	Accuracy: 98.64% - 100%	High classification accuracy, effective feature extraction	Possible overfitting, lack of real-field validation, high model complexity.

For image processing, images of paddy plants are captured from the fields and subsequently undergo preprocessing. Further, disease detection methods can be broadly categorized into two main types: object detection and classification. Several deep-learning approaches have been developed to specifically focus on segmentation to differentiate healthy plants from those affected by diseases [36]. Following segmentation [41], relevant features are extracted from the segmented images, and deep learning techniques are employed to classify various rice plant diseases [18]. On the other hand, rice plant disease classification refers to categorizing and identifying different diseases that affect rice plants according to their visual symptoms and characteristics [31].

Recent advances in rice plant disease classification techniques have greatly improved the accuracy and efficiency of identifying and categorizing different diseases that affect rice plants [1, 6, 23, 33, 39, 40]. These techniques use state-of-the-art technologies to analyze visual symptoms and characteristics of diseased rice plants. By extracting relevant features from the images, these methods enable the use of advanced deep learning-based classification models such as Convolutional Neural Networks (CNNs) [17], Recurrent Neural Networks (RNNs), [28] and Support Vector Machines (SVMs) [32]. The models are trained on single large datasets of labeled rice plant images, allowing them to learn complex patterns and accurately classify different types of diseases. Moreover, the significance of fine-tuning transfer learning lies in its ability to further enhance the performance of disease classification in rice plants by leveraging pre-trained models on large-scale datasets, improving accuracy even with limited labeled data.

In this research paper, we present a novel framework for the automated classification of rice plant diseases to assist farmers in enhancing the quality and quantity of their rice crops. The primary objective of this study is to develop an efficient and reliable rice plant disease detection model that can prevent unhealthy farming practices and minimize crop losses. Our research article introduces several significant contributions and advancements in rice plant disease detection and classification.

These contributions are succinctly summarized as follows:

- This research study presents a novel framework that uses fine-tuning transfer learning, and ensemble attention heads mechanisms for autonomously balanced feature representation precise classification, even in challenging scenarios, like where the disease data exhibits variations and complexities.
- This paper presents a robust and generalized RiceNet framework that effectively handles the data heterogeneity which contains different samples and features.
- This study presents rigorous experiments, using diverse rice disease datasets, and compares RiceNet's performance with benchmark Deep Learning Algorithms and state-of-the-art methods.

This research paper is structured into several sections, each serving a specific purpose. Section 1 introduces the significance and motivation of the research study, and Section 2 provides an overview of related work in the field, setting the foundation for the investigation. Section 3 provides a comprehensive analysis of the dataset employed in the study and the methodology utilized in the research in detail, which explores insights into the research strategies. Section 4 presents the study's potential findings, and a comparison with baselines and SOTA is evaluated comprehensively. Section 5 provides an understanding of the study's conclusions and critical analysis of the findings and their implications. Finally, it proposes potential future directions for further exploration in the field.

2 Related work

Detecting plant diseases is vital for maintaining agricultural productivity and ensuring food security. Numerous computational techniques, such as conventional machine learning, approaches based on manually crafted features, and deep learning methods, have been investigated to tackle this challenge. In particular, the detection of rice diseases has attracted considerable focus, prompting researchers to create a variety of strategies aimed at enhancing both accuracy and efficiency. These approaches span from traditional image processing methods to sophisticated neural network architectures, each playing a role in improving automated disease identification within precision agriculture.

The study [11] explores the application of image processing methods to identify and classify leaf diseases of plants, analyzing various techniques proposed between 1997 and 2016. Although the study emphasizes the effectiveness of image processing in disease recognition by improving automated plant disease classification, reducing the reliance on manual processes, and highlighting the limitations of traditional manual visual inspection methods, such as their unpredictability and time-consuming nature, it has certain limitations. For instance, it does not extensively discuss the challenges associated with image acquisition, variability in disease symptoms, and the need for extensive and diverse datasets for robust classification models. In addition, the study [13] introduces the MBi-LSTM model for classification of leaf disease of paddy plants. However, it is important to note that the generalizability of the proposed method can be limited to paddy plants, and its effectiveness may vary when applied to plant species or crops in the field environment. More research is needed to assess the adaptability and performance of the methodology in diverse agricultural contexts and types of disease. There is a pressing need to explore the potential drawbacks or trade-offs of using image processing techniques regarding computational complexity or real-time applicability. These limitations warrant further investigation and consideration in future research.

Furthermore, the study [21] explores the application of deep learning techniques in agriculture, specifically for detecting and classifying rice leaf diseases. The authors propose a modified approach based on a Deep Convolutional Neural Network (DCNN) and transfer learning, precisely a modified VGG19-based method, to accurately detect and diagnose six distinct classes of rice leaf diseases. The results demonstrate high accuracy, precision, recall, specificity, and F1-score, outperforming similar approaches reported in the literature. This research highlights the potential of combining machine learning algorithms, drone technology, and the Internet of Things (IoT) to provide a comprehensive solution for disease detection in agriculture. However, the proposed study heavily relies on digital image processing and recognition methods, which may introduce challenges related to variations in lighting conditions, image quality, and the need for extensive labeled datasets.

The research study [24] presents a novel approach for identifying rice diseases using deep learning techniques. The authors address the need for accurate and efficient diagnosis of rice diseases, which significantly threaten rice production and food security. By employing deep convolutional neural networks (CNNs) and training them on a dataset of 500 natural images of diseased and healthy rice leaves and stems, the proposed model achieves an impressive accuracy of (95.48%) under the 10-fold cross-validation strategy. This accuracy surpasses conventional machine learning models and demonstrates the feasibility and effectiveness of the proposed method. The research underscores the immense promise of utilizing deep learning in agricultural information processing, specifically focusing on convolutional neural networks (CNNs). It vividly demonstrates the remarkable capacity of this technology to enhance the precision of disease diagnosis outcomes, bringing about significant advancements in the field.

Nalini et al. [26] introduced an innovative methodology that revolutionizes identifying and classifying paddy leaf diseases. Their approach utilizes a deep neural network (DNN) model, fine-tuned with a crow search algorithm (CSA), to tackle the pressing issue of early disease detection in plants–an imperative aspect for bolstering agricultural productivity. The proposed model achieves remarkable accuracy using image processing and machine learning by effectively reducing classification errors and optimizing key parameters such as weights and biases. Notably, the performance of this model surpasses that of a support vector machine (SVM) algorithm. While the proposed approach shows promise in accurately identifying and classifying paddy leaf diseases, the study does not comprehensively evaluate the model's performance on a large-scale dataset. Further validation on diverse datasets and real-world scenarios would be beneficial to assess its robustness and generalizability.

The study [7] Presents a real-time disease diagnosis system for rice leaves using the Faster R-CNN algorithm. The research addresses the limitations of conventional methods by utilizing machine-driven diagnosis systems, which are more efficient and accurate. By employing the Faster R-CNN model, the study achieves high accuracy in diagnosing three discriminative rice leaf diseases, namely rice blast, brown spot, and spa, and identifying healthy rice leaves. However, despite its advantages, certain limitations, such as the fact that it requires high-performance computing resources and strong image backgrounds, vague symptoms, and variations in symptoms from the same infection, still mar the efficacy of the system and its usage.

The research study [15] employed VGG16 with transfer learning to achieve an accuracy of 92.4% in identifying rice plant diseases. Kumar et al. [20] utilized Fuzzy Logic Based Expert System, Radial Basis Function Neural Network, and CNN with transfer learning, obtaining a remarkable accuracy of (97.61%) for coffee leaf disease detection. These studies demonstrate the effectiveness of deep learning models and transfer learning in accurately classifying plant diseases. However, one limitation of relying solely on deep learning models is their dependence on large amounts of labeled data for training. Acquiring and annotating such datasets can be time-consuming and labor-intensive. Therefore, while deep learning shows promise in disease diagnosis, addressing the data requirements remains a significant challenge.

Research contribution by [9] presents a compelling approach for disease detection in rice plants using a deep neural network (DNN) model. The study addresses the crucial challenge of timely identification and diagnosis of diseases in rice plants, which can significantly impact agricultural productivity. By incorporating texture and deep features extracted from plant images, the proposed DNN model achieves impressive results in accurately detecting various diseases. The paper demonstrates the model's effectiveness through extensive experimentation and evaluation using a diverse dataset. Deep learning techniques enhance the model's ability to capture intricate patterns and subtle differences indicative of different diseases. However, one limitation of this study is the lack of comparison with other state-of-the-art disease detection methods. A comprehensive analysis and comparison with alternative approaches would have provided valuable insights into the strengths and weaknesses of the proposed DNN model.

The Study [16] introduces an innovative method for detecting rice diseases by harnessing the deep convolutional neural networks (DCNN) with various optimization techniques. The researchers conduct a comprehensive analysis, comparing the efficacy of DCNN against support vector machine (SVM), artificial neural network (ANN), and long short-term memory (LSTM) models. To enhance the accuracy of LSTM, they ingeniously integrate particle swarm optimization, artificial fish swarm optimization (AFSO), and efficient artificial fish swarm optimization (EAFSO) algorithms to optimize the determination of optimal weights. The results show that the proposed DCNN-LSTM (EAFSO) technique outperforms other methods, achieving a remarkable (97.5%) accuracy in identifying rice diseases. The study highlights the importance of image recognition and artificial intelligence in precision agriculture and disease detection. However, one limitation of the study is the limited discussion on the generalizability and scalability of the proposed technique.

The research [10] presents a novel approach for detecting diseases in rice plants at an early stage using the Rider Water Wave-based neural network (RWW-NN). The proposed method combines the Rider Optimization algorithm (ROA) and Water wave optimization (WWO) to train a deep neural network for disease recognition. The paper highlights the importance of early disease detection in rice plants to mitigate economic losses and improve agricultural productivity. The RWW-based NN achieved impressive results with a high accuracy of 0.908, F-measure of 0.907, sensitivity of 0.862, and specificity of 0.947 based on the K-value using the Rice disease dataset. This research contributes to rice plant disease detection by introducing a new approach that utilizes deep learning and optimization techniques. However, one drawback of the proposed method is that it did not perform well with diverse datasets and identification of grain storage warehouses.

The research study [12] proposed an innovative approach to accurately identify and classify rice leaf diseases using a Convolutional Neural Network (CNN) and Visual Geometry Group (VGG)19 model. The authors employ transfer learning based on VGG19 to identify the brown spot class of rice leaf diseases, achieving an impressive accuracy of (93.0%) on the dataset. The proposed technique outperforms existing baseline models regarding sensitivity, specificity, precision, and F1-score. However, a potential limitation of the study is that it focuses solely on detecting only brown spot rice leaf disease (a single disease) rather than detecting multiple diseases on a single rice leaf.

Nayak et al. [27] addresses the pressing concern of early detection of plant diseases in the farming community. The study demonstrates the accuracy of disease detection in rice plants by utilizing advanced digital cameras and smartphones equipped with improved image acquisition modes and deep learning techniques. Various image segmentation techniques, such as foreground extraction, were employed to identify affected portions accurately. The researchers optimized models and procedures to enable offline smartphone working capabilities, contributing to developing a dynamic framework. The study compares different models for image classification and identifies the top-performing ones, including DenseNet201, Xception, MobileNetV2, and ResNet-50. While the ResNet50 model excels for cloud architectures, MobileNetV2 is the best model for smartphone applications. Furthermore, an android application, "Rice Disease Detector," incorporating the MobileNetV2 model, was successfully tested for multiple disease occurrences. However, a drawback of the study is the need for model evaluation using diverse datasets with varying configurations.

The research study [44] focuses on developing an attention-based depthwise separable neural network with Bayesian optimization (ADSNN-BO) for detecting and classifying rice diseases from rice leaf images. The authors address the significant economic losses caused by rice diseases and the need for rapid and accurate disease identification. The ADSNN-BO model, based on the MobileNet structure and augmented attention mechanism, achieves a remarkable test accuracy of 94.65%, outperforming all tested state-of-the-art models. The proposed attention-based mechanism effectively guides the model to learn informative features, enhancing its interpretability. The study also highlights the importance of visualizing and understanding the learning process of neural networks through techniques such as activation visualization and filter visualization. However, one limitation of the study is the lack of implementation of the proposed model for rice disease detection in diverse and real-world scenarios, which could provide more insights into its practical applicability.

The research study [8] presents a comprehensive study on applying deep learning techniques, specifically convolutional neural networks (CNNs), for identifying rice plant diseases. The authors highlight the significance of rice crops in ensuring global food security and the detrimental impact of diseases on agricultural productivity. They propose a novel approach that combines the MobileNet-V2 model with an attention mechanism to improve the learning capability for minute lesion features. The proposed procedure achieves remarkable identification accuracy through extensive experiments, surpassing state-of-the-art methods, with an average accuracy of (99.67%) on a public dataset. Furthermore, even under complex backdrop conditions, the average accuracy reaches (98.48%) for identifying rice plant diseases. The study emphasizes the importance of early detection and provides a simple, fast, less expensive, and accurate system for automatic disease identification. However, despite all these advantages, a potential limitation of the research lies in the limited diversity of the image database, which mainly consists of laboratory images rather than real-life diverse agricultural fields. This limitation may affect the generalizability of the proposed approach to real-world scenarios.

The research study [14] presents the comparative analysis of different deep learning architectures for rice leaf disease classification, demonstrating the effectiveness of CNN-based models. Despite the promising results in terms of high classification accuracy the proposed study primarily relies on a single dataset sourced from Kaggle, which may not fully represent real-world variations in environmental conditions, leaf orientations, and disease severity. A more extensive dataset incorporating images from diverse sources would improve the model's generalizability, while the [2] presents a robust classification system for rice plant disease detection using a multi-level handcrafted feature extraction approach. Recent studies have shown that transformer-based architectures and self-supervised learning techniques can significantly improve feature representation in plant disease detection. Furthermore, the study primarily focuses on handcrafted feature extraction, which, despite its interpretability, may not fully leverage the hierarchical feature learning capability of deep neural networks. The dependency on benchmark datasets, which may not capture real-world variations in lighting, disease severity, and leaf occlusion, also raises concerns about model generalizability. Furthermore [25] presents the recent comprehensive review of 48 rice plant disease identification studies and discusses the various machine learning algorithms challenges in data heterogeneity, model, and generalizability.

while the study [35] demonstrating high accuracy, the proposed IX-CNN model raises concerns about overfitting due to its near-perfect performance on benchmark datasets without real field validation. The lack of discussion on dataset imbalance, computational complexity for low-memory devices, and the absence of comparisons with state-of-the-art architectures limit its generalizability. Furthermore, the study does not address explainability, making it unclear whether decisions are based on meaningful features or background artifacts. Without validation in diverse environmental conditions and consideration of early disease stages, the real-world applicability of the model remains uncertain.

One of the key challenges in the classification of rice plant disease is encountered when the representation of characteristics in a data set is highly imbalanced, with certain characteristics significantly underrepresented compared to others. This feature imbalance poses a difficulty for classification algorithms in learning and accurately classifying minority features effectively. Furthermore, the presence of multiple diseases within a single image further compounds the challenge, as classification algorithms may struggle to detect and identify all diseases in such heterogeneous datasets comprehensively. To overcome these limitations, there is a growing need to adopt target disease identification methodologies that specifically address the issues of feature imbalance and multi-disease detection. These targeted approaches aim to improve the performance and accuracy of disease classification algorithms, enabling more robust and reliable identification of various rice plant diseases in real-world scenarios.

3 Material and methods

This section of the paper delves into the details of two distinct datasets of rice plant disease (large and small) used in this study, accompanied by comprehensive details of the methods and techniques used.

3.1 Data acquisition and pre-processing

3.1.1 Rice leaf disease image samples (Large Dataset)

This dataset is publicly accessible and sourced from [34]. It comprises 5,932 high-resolution images of diseased rice leaves, including bacterial blight, blast, brown spot, and tungro varieties of the plant disease. However, we used the custom subset of the large dataset, which contains 4624 images, in connection with the consistency of the diverse small dataset, shown in Table 2. Furthermore, these images were originally captured from various rice fields of western Odisha using a Nikon DSLR-D5600 with an 18–55 mm lens. The diseased

Table 2 Statistics of rice leaf disease datasets

Name of Disease	No. of Samples in Large Dataset	No. of Samples in Small Dataset
Bacterial Blight	1584	40
Blast	1440	40
Brown Spot	1600	40
Total	4624	120

patch areas were cropped from the original large images. Some images of rice disease were also collected from the source of the agricultural pest and insect pest picture database from [4] to make it a standard dataset.

3.1.2 Rice leaf diseases (Small Dataset)

This publicly accessible dataset was sourced from [43] and comprises 120 images with three disease class categories. Each disease category contains 40 images in JPEG format. The dataset was initially collected from the farm fields, and the images of rice plant leaves were captured with a white background under direct sunlight using a NIKON D90 DSLR camera with 12.3 megapixels. The images have the dimensions of 2848x4288 pixels. The Table 2 presents the disease names and corresponding quantities of the small and large datasets utilized for experimentation.

Comprehensive data preprocessing and augmentation techniques were applied to prepare both datasets for training a robust and generalized deep-learning model for rice plant disease classification. The images in the dataset were resized to a fixed size of 224x224 pixels, ensuring uniform dimensions across all samples. Additionally, data augmentation was performed by randomly applying horizontal flips to the images, introducing variations, and enhancing the model's ability to generalize to different orientations. The image data was transformed into a tensor transformation to facilitate the training process.

$$\mathrm{NV} = \frac{\text{original\_value} - \frac{1}{N}\sum_{i=1}^{N}\text{original\_value}}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\text{original\_value} - \frac{1}{N}\sum_{i=1}^{N}\text{original\_value})^2}} \tag{1}$$

Here, N represents the number of samples. $original_value$ represents the pixel value of the image before normalization. Moreover, normalization was applied to the image tensors by subtracting the mean values [0.485, 0.456, 0.406] and dividing by the standard deviation values [0.229, 0.224, 0.225] for each RGB channel as shown in (1), where NV indicates the normalized value. This normalization step helps achieve convergence and stability during training and ensures consistent scaling across the datasets. The dataset was carefully prepared by employing effective preprocessing and augmentation techniques, providing a diverse and representative set of training samples for the deep transfer learning model. This meticulous preparation enhanced the model's ability to accurately classify and detect various rice plant diseases, even with diverse datasets.

Algorithm 1 RiceNet: EAM for automated rice plant disease classification.

```
Input: Dataset of rice plant images with labels
Output: Trained model for rice disease classification
Initialization: Initialize Pretrained $M_1$ and Pretrained $M_2$ models, Transformer encoder with encoder layer and num_layers = 1, linear classification layer, optimizer as Adam optimizer and disable gradient calculation for all parameters.
Prepossessing: apply data Integration, augmentation, Reduction, and Data Denoising.
Data Split: split into the train, validation, and test sets
Define: encoder layer with $d\_model = 512$ and $nhead$ (number of attention heads), linear layer with input size of 512 and num_classes output units, and Define loss function as cross-entropy loss
Create: model list as ModuleList containing pretrained $M_1$ and pretrained $M_2$
Create combine_model as ModuleList [combining model_list, transformer encoder, and linear.]
for each epoch in num_epochs do
    Training:
    for each batch in training_loader do
        Forward pass through ensemble models and Transformer encoder
        Compute loss and gradients, update model parameters
    end for
    Validation:
    for each batch in validation_loader do
        Forward pass through combine ensemble models and Transformer encoder
        Compute validation loss and accuracy
    end for
end for
Testing:
for each batch in test_loader do
    Forward pass through combine ensemble models and Transformer encoder
    Compute test loss and accuracy
end for
Evaluation: Compute precision, recall, F1-score, balanced accuracy, Cohen's kappa
Visualization: Plot loss and accuracy curves, confusion matrix, and sample predictions.
```

3.2 Ensemble attention mechanism approach

This section introduces the Ensemble Attention Mechanism (EAM) in RiceNet, an advanced network for classifying rice plant diseases. The model architecture diagram, depicted in Fig. 2, visually represents the RiceNet model employed in accurately classifying rice plant

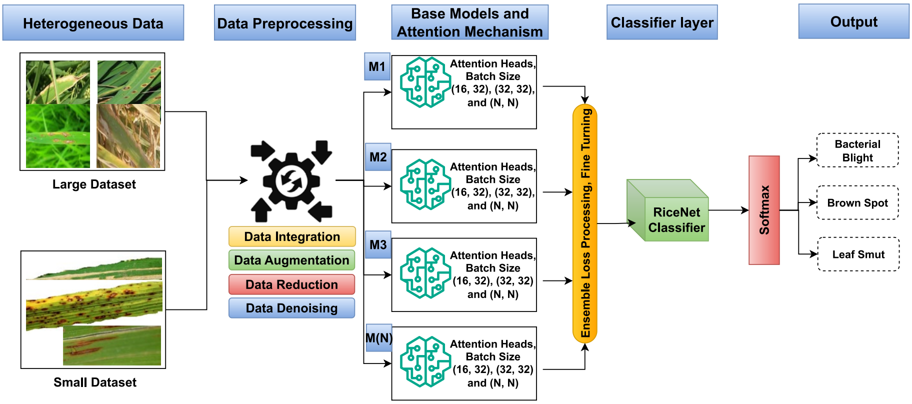


Fig. 2 RiceNet model architecture

diseases. The Ensemble Attention Mechanism (EAM) is a pivotal component of the RiceNet framework, as it serves as a key element in achieving precise classification and effectively mitigating the issue of feature representation. By incorporating inherent attention heads, the EAM ensures a balanced representation of classes and features, amplifying the model's overall performance. This approach employs a multi-model ensemble strategy, harnessing the collective predictive capabilities of individual models to enhance the classification accuracy, as exemplified by (2).

$$\mathrm{Attention}(x) = \mathrm{Softmax}(\mathrm{Linear}(x)) \tag{2}$$

Where the function *Attention*(*x*) signifies the application of the attention mechanism to the input *x*, *Linear*(*x*) denotes the linear transformation performed on the input *x*, and *Softmax* represents the activation function utilized.

3.3 Working of framework's attention layer

The scaled dot-product attention mechanism is significant in various deep learning models, particularly natural language processing. It assumes a vital role in transformer architectures, driving their success. The mathematical formulation of the attention mechanism can be expressed by considering a set of input vectors or sequences as:

- A set of query vectors, denoted as Q, characterized by dimensions such as batch size, seq_length, and model.
- A group of key vectors, denoted as K, also possesses dimensions such as batch size, seq_length, and d_model.
- A collection of value vectors, represented as V, and sharing dimensions such as batch size, seq_length, and model.

The self-attention mechanism calculates attention scores for every combination of query and key vectors. Subsequently, these attention scores are employed to determine the importance of the value vectors, leading to the generation of the desired output. The mathematical expression for the scaled dot-product attention mechanism is presented as follows. (1) Calculate the raw (unscaled) attention scores, denoted as A:

$$[A = Q \cdot K^{\mathsf{T}} / \sqrt{d_k}] \tag{3}$$

In (3), the symbol *T* signifies the transpose operation performed on matrix *K*, while $sqrt(d_k)$ represents the square root of the dimensionality of the key vectors, denoted as d_k. This scaling factor prevents the dot products from becoming excessively large, potentially resulting in vanishing gradients during training. A softmax function ensures proper normalization of the attention scores across the sequence length. This normalization process yields the attention weights, denoted as *W*, as depicted in (4).

$$[W = \mathrm{softmax}(A)] \tag{4}$$

The attention score A is transformed into attention weights W by applying the softmax function. This transformation guarantees that the attention weights add up, accurately indicating the significance assigned to each element in the sequence. Using these attention weights, the value vectors V are weighted to calculate the output O, as illustrated in (5).

$$O = W \cdot V \tag{5}$$

The attention layer in a neural network computes a weighted sum by multiplying input vectors with a weight matrix, combining these scores with the encoder's output. An attention function, such as softmax, is subsequently applied to produce a probability distribution that signifies the importance of each element. This mechanism allows the model to calculate relevance weights for each input piece by combining the input with the encoder's output. These procedures are applied to each query in the batch, resulting in an output tensor O with the same shape as the input value tensor V. The self-attention method is used concurrently for every query in the batch, enabling the model to learn which elements in the input sequence are most relevant to each query.

$$\text{EnsemblePrediction} = \frac{1}{N}\sum_{i=1}^{N}\text{Attention}(M_i) \tag{6}$$

Here, *EnsemblePrediction* represents the ensemble prediction, N denotes the total number of models in the ensemble, while $Attention(M_i)$ indicates the attention mechanism applied to the i^{th} model. The ensemble prediction, as defined by (6), combines the predictions from the individual models, integrating diverse insights captured by each (M_i), resulting in a more comprehensive and robust classification outcome. The final output is computed by taking the mean prediction from all the individual models in the ensemble.

3.4 Attention layer encoder and decoder process

The proposed methodology incorporates an ensemble attention layer consisting of encoder and decoder processes to enhance the classification performance. The encoder process captures the salient features of the input data using a transformer encoder layer. The decoder transforms the encoded features into class predictions using a linear layer. The ensemble is formed by integrating multiple pretrained models with the encoder and decoder layers. The encoder process utilizes a transformer encoder layer with a d_model and heads, as shown in (7)

$$\text{Encoder: } \mathrm{z} = \text{TransformerEncoder}(\mathrm{x}), \mathrm{x} \quad \in \mathbb{X} \subset \mathbb{R}^{d_o}, \quad \mathrm{z} \in \mathbb{Z} \subset \mathbb{R}^{d_k} \tag{7}$$

The encoder network converts an input $x \in X \subset \mathbb{R}^{d_0}$ into a feature representation $z \in Z \subset \mathbb{R}^{d_k}$. On the other hand, the decoder takes this feature representation z as input, processes it, and generates an output $y \in Y \subset \mathbb{R}^{d_d l}$. The decoder and encoder contain the same number of layers k to maintain symmetry in the network architecture. Additionally, the encoder layer's input dimensions correspond to the decoder layer's output dimensions, denoted as D^1. This symmetrical network design ensures a balanced relationship between the encoding and decoding processes, contributing to the overall performance and coherence of the network. Mathematically, the mappings of the encoder and decoder layers can be represented as $\sigma : \mathbb{R}^{dl-1} \rightarrow \mathbb{R}^{d_l}, \quad \mathrm{d}_l : \mathbb{R}^{d_l} \rightarrow \mathbb{R}^{d_{l-1}}$, respectively.

Here, the notation $l \in [k]$ with $[n]$ represents a transpose operation, and $1, \ldots, n$ pertains to the input from the j^{th} channel, which possesses a dimension of d_0. As a result, the cumulative input dimension is denoted as q_{l-1}. At the encoder's l^{th} layer, the convolution operation is employed to generate an output with channels. $\sigma_{l-1} = [\sigma_{l-1^T 1}, \ldots, \sigma_{l-1^T q_{l-1}}] \in \mathbb{R}^d_{l-1}$

The transformer encoder is formed by employing the transformer encoder layer with a specific number of layers (in this case, one layer), as in the (8).

$$\mathrm{TE} = \mathrm{TransformerEncoder}(z, num\_layers = 1) \tag{8}$$

The decoder process involves a linear layer that transforms the encoded features into class predictions. The linear layer has an input size of 512 (matching the input feature dimensions of the transformer encoder) and an output size of the number of classes, as shown in (9)

$$\text{Decoder: } \mathrm{y} = \mathrm{Decoder}(\mathrm{z}), \quad \mathrm{z} \in \mathbb{R}^{d_k}, \quad \mathrm{y} \in \mathbb{R}^{d_d} \tag{9}$$

Here d_1 represents the input dimensionality, d_k represents the feature space dimensionality, and d_d represents the output dimensionality. k is the number of layers in the encoder and decoder networks. The decoder layer is formed as shown in the (10), $d_l - 1$ and d_l represent each layer's input and output dimensions, respectively.

$$\text{Decoder Layer: } \mathbf{d_l} : \mathbb{R}^{d_l} \rightarrow \mathbb{R}^{d_{l-1}} \tag{10}$$

The combined ensemble model (CEM) is constructed using the ModuleList, which includes the individual models like (VGG, RegNet, EnceptionNet, and ResNet), the transformer encoder, and the linear decoder layer, as shown in the following equation

$$CEM = ModuleList([M1, M2, M3, TE, y]) \tag{11}$$

3.5 Model training and network settings

In our pursuit of effectively classifying rice diseases, we conducted a systematic evaluation of several prominent CNN baseline models, including VGG, RegNet, and InceptionNet. Our RiceNet model outperformed all others, showcasing its superiority. We carefully examined different training batch sizes (4, 16, and 32) with varying training split ratios to optimize the model's performance. We determined that batch size 32 with a large dataset and 4 with a small yielded the best results. Furthermore, we used the Adam (Adaptive Moment Estimation) is an optimization algorithm that combines the advantages of both the momentum and RMSprop optimizers. It adapts the learning rate for each parameter, leading to faster convergence and improved performance in deep learning models.

For loss function selection, we compared Categorical Cross-Entropy as shown in (12) and Kullback-Leibler Divergence as shown in (13), ultimately finding that Categorical Cross-Entropy was the most efficient choice.

$$\text{Cross-Entropy} = -\sum_{i=1}^{N} y_i \cdot \log(p_i) \tag{12}$$

$$\text{Kullback-Leibler Divergence} = \sum_{i=1}^{N} P(i) \log\left(\frac{P(i)}{Q(i)}\right) \tag{13}$$

Our investigation also extended to kernel sizes, considering 3x3 and 5x5 kernels. Notably, the 3x3 kernel size emerged as the optimal choice for our objectives. The learning rate played a crucial role in our model's performance, and after testing different values (0.1, 0.001, 0.0001), we determined that a learning rate of 0.0001 produced the most favorable outcomes.

3.6 Evaluation metrics

In this study, we utilize a comprehensive set of evaluation metrics to ensure a thorough and reliable assessment of RiceNet's performance. These metrics include precision, recall, F1 score, balanced accuracy, and Cohen's Kappa. They provide valuable insights into different aspects of classification accuracy.

Precision measures the model's ability to accurately identify positive instances, while recall evaluates its sensitivity in detecting true positive cases. The F1 score balances precision and recall, offering a unified assessment of overall effectiveness. Balanced accuracy considers the class distribution, objectively evaluating RiceNet's performance across diverse categories. Cohen's Kappa measures the agreement between predicted and true labels, offering a robust measure of reliability. By incorporating these metrics, we comprehensively understand RiceNet's performance across various network settings.

$$\text{Precision (Micro)} = \frac{TP_{\text{total}}}{TP_{\text{total}} + FP_{\text{total}}} \tag{14a}$$

$$\text{Precision (Macro)} = \frac{1}{N}\sum_{i=1}^{N} \frac{TP_i}{TP_i + FP_i} \tag{14b}$$

$$\text{Precision (Weighted)} = \frac{\sum_{i=1}^{N}(TP_i + FP_i) \cdot \text{support}_i}{\sum_{i=1}^{N}(TP_i + FP_i)} \tag{14c}$$

$$\text{Recall (Micro)} = \frac{TP_{\text{total}}}{TP_{\text{total}} + FN_{\text{total}}} \tag{14d}$$

$$\text{Recall (Macro)} = \frac{1}{N}\sum_{i=1}^{N} \frac{TP_i}{TP_i + FN_i} \tag{14e}$$

$$\text{Recall (Weighted)} = \frac{\sum_{i=1}^{N}(TP_i + FN_i) \cdot \text{support}_i}{\sum_{i=1}^{N}(TP_i + FN_i)} \tag{14f}$$

$$\text{F1 (Micro)} = 2 \cdot \frac{\text{Precision (Micro)} \cdot \text{Recall (Micro)}}{\text{Precision (Micro)} + \text{Recall (Micro)}} \tag{14g}$$

$$\text{F1 (Macro)} = \frac{1}{N} \sum_{i=1}^{N} \frac{2 \cdot \text{Precision}_i \cdot \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i} \tag{14h}$$

$$\text{F1 (Weighted)} = \frac{\sum_{i=1}^{N} (2 \cdot \text{Precision}_i \cdot \text{Recall}_i \cdot \text{support}_i)}{\sum_{i=1}^{N} (\text{Precision}_i + \text{Recall}_i \cdot \text{support}_i)} \tag{14i}$$

$$\text{Balanced Accuracy} = \frac{1}{N} \sum_{i=1}^{N} \frac{TP_i}{TP_i + FN_i} \tag{14j}$$

$$\text{Cohen's Kappa} = \frac{P_o - P_e}{1 - P_e} \tag{14k}$$

These equations represent the formulas for precision, recall, F1 score, balanced accuracy, and Cohen’s Kappa. Each equation calculates the respective evaluation metric using the true positive (TP), false positive (FP), false negative (FN), and support values.

4 Experimental results and discussion

The Table 3 presents the performance metrics of different ensembles with attention mechanisms on a large dataset. The table comprehensively evaluates various backbone models, network selection strategies, and split ratios, along with their corresponding precision, recall, F1 score, balanced accuracy, Cohen’s kappa, and average time per epoch. The backbone models used in the experiment include VGG (16 and 19), InceptionNet (V1 and V3), and RegNet (Y-16 and Y-32). Each backbone model combines an Ensemble Attention Head and Batch (EAHB) mechanism with different attention heads and batch size configurations. The split ratio indicates the division of the dataset into training (Tr), validation (Vl), and test (Ts) sets. The precision, recall, and F1 scores are reported in three different metrics: micro, macro, and weighted. Micro metrics consider all classes equally, while macro metrics compute the average per class. Weighted metrics take into account class or feature imbalance. The balanced accuracy and Cohen’s kappa provide additional measures of classification performance. Lastly, the average time per epoch indicates the computational efficiency of the models during training.

In the first experiment using the large dataset, the VGG-16 and VGG-19 baseline models were used in an ensemble configuration with the Ensemble Attention Head 16 and batch size 32. The large dataset was split into [1156, 1156, 2312] for training, validation, and testing. Further experiments also showed competitive performance for both VGG ensemble models in the table, achieving the highest precision (macro) score among all models. This indicates that VGG ensembles accurately classify rice plant diseases across multiple classes.

In further experiments, the InceptionNet models, specifically InceptionNet V1 and InceptionNet V3, were employed as backbone models in the ensemble configuration using EAHB

Table 3 Performance metrics of ensembles with attention mechanism on large dataset

Backbone	NS	SR (Tr,Vl,Ts)	P (Mi)	P (Ma)	P (W)	R (Mi)	R (Ma)	R (W)	F1 (Mi)	F1 (Ma)	F1 (W)	BA	CK	Avg. T
VGG(16,19)	EAHB(16,32)	1156, 1156, 2312	0.9771	0.9771	0.9771	0.9771	0.9765	0.9771	0.9771	0.9767	0.9771	0.9765	0.9656	3.23
VGG(16,19)	EAHB(16,32)	416, 3745, 463	0.9438	0.9465	0.9445	0.9438	0.9431	0.9438	0.9438	0.9445	0.9440	0.9431	0.9153	1.14
VGG(16,19)	EAHB(32,32)	1156, 1156, 2312	0.9658	0.9654	0.9658	0.9658	0.9651	0.9658	0.9658	0.9652	0.9658	0.9651	0.9487	3.20
InceptionNet	EAHB(32,32)	416, 3745, 463	0.9633	0.9631	0.9637	0.9633	0.9631	0.9633	0.9633	0.9629	0.9634	0.9631	0.9449	3.06
InceptionNet	EAHB(32,32)	3744, 417, 463	0.9978	0.9980	0.9979	0.9978	0.9977	0.9978	0.9978	0.9978	0.9978	0.9977	0.9968	77.36
RegNet	EAHB(32,32)	416, 3745, 463	0.9784	0.9784	0.9784	0.9784	0.9791	0.9784	0.9784	0.9784	0.9787	0.9784	0.9791	8.40
RiceNet	EAHB(32,4)	1156, 1156, 2312	0.9641	0.9639	0.9666	0.9641	0.9657	0.9641	0.9641	0.9638	0.9644	0.9657	0.9461	7.8
RiceNet	EAHB(32,4)	3744, 417, 463	0.9978	0.9980	0.9979	0.9978	0.9978	0.9978	0.9978	0.9979	0.9978	0.9978	0.9968	8.2
RiceNet	EAHB(32,32)	1156, 1156, 2312	0.9849	0.9850	0.9849	0.9849	0.9842	0.9849	0.9849	0.9845	0.9848	0.9842	0.9773	2.9
RiceNet	EAHB(32,32)	415, 47, 4162	0.9118	0.9187	0.9225	0.9118	0.9139	0.9118	0.9118	0.9123	0.9133	0.9139	0.8680	1.05
RiceNet	**EAHB(32,32)**	**3744, 417, 463**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**3.02**

The bold entries indicate the best performance results and are intended to highlight key findings

32 and batch size 32, the dataset was split into [416, 3745, 463] for training, validation, and testing. However, in various settings, InceptionNet ensembles showcased superior performance in terms of recall (weighted), indicating their effectiveness in capturing relevant features from the input data. This suggests that InceptionNet can correctly identify instances of different rice plant diseases.

The comprehensive experiments using RegNet Y-16 and RegNet Y-32 backbone models were utilized in the ensemble configuration with EAHB 32 and batch size 32. The dataset was divided into [1156, 1156, 2312] for training, validation, and testing. RegNet ensembles exhibited superior balanced accuracy, indicating their ability to maintain a balanced classification performance with different network settings across all disease classes. This suggests that RegNetis is well-suited for accurately classifying various rice plant diseases.

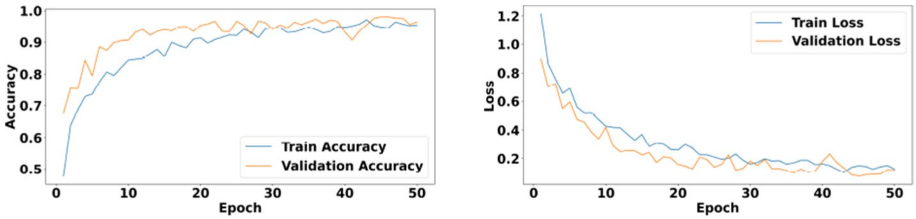
The RiceNet, a modified ResNet-18 and ResNet-50 architecture with ensemble learning principles and attention mechanisms, was evaluated utilizing a large dataset with field environment complexities. The ensemble configuration employed the EAH 32 and batch size 4, and the dataset was divided into [1156, 1156, 2312] for training, validation, and testing. RiceNet consistently outperformed the baseline backbone models across various network settings and metrics, including precision, recall, F1 score, balanced accuracy, and Cohen's kappa. This demonstrates the effectiveness of the RiceNet framework in accurately classifying rice plant diseases. Additionally, RiceNet demonstrated computational efficiency, with an average time per epoch comparable to the baseline models. These findings suggest that RiceNet offers a promising rice plant disease classification solution.

The numerical results in the table showcase the performance of different backbone models and the RiceNet architecture for rice plant disease classification. The baseline models exhibited competitive performance in various metrics, each with strengths. However, RiceNet, with its modified ResNet architecture, ensemble learning principles, and attention mechanisms, emerged as the top performer. It achieved superior precision, recall, F1 score, balanced accuracy, and Cohen's kappa compared to the baseline models. Furthermore, RiceNet demonstrated computational efficiency, making it a promising approach for accurate and efficient rice plant disease classification.

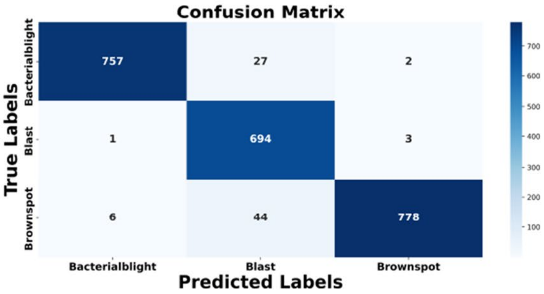
Furthermore, visualization results of extensive experiments using two different datasets with various baseline models, data split, batch size, and ensemble attention heads are visualized in Figs. 3, 4, 5, 6, 7, 8, 9 and 10. The comprehensive analysis of the performance of the RiceNet framework for automated rice plant disease classification provides valuable insights into RiceNet's effectiveness across various network settings and datasets. These results are crucial in assessing the effectiveness and robustness of the framework in addressing the challenges posed by dataset heterogeneity and feature imbalance.

The Table 3 defines the acronyms as NS = Network Selection, SR = Split Ratio, Tr = Train, Vl = Validation, Ts = Test, P = Precision, R = Recall, Mi = Micro, Ma = Macro, W = Weighted, BA = Balanced Accuracy, CK = Cohen's Kappa, Avg. T = Average Time Per epoch, and EAHB = Ensemble Attention Head and Batch. The computed numerical values in Table show the performance of VGG(16, 19) on a large dataset. The table showcases the results obtained using ensemble attention heads with a count of 16, a batch size of 32, and a specific split ratio of [1156, 1156, 2312] for the train, validation, and test sets.

The RiceNet's visualizations include a) accuracy, b) loss, and c) confusion matrix. The accuracy metric provides insights into the proportion of correctly classified samples, while the loss metric measures the dissimilarity between predicted and true labels associated with

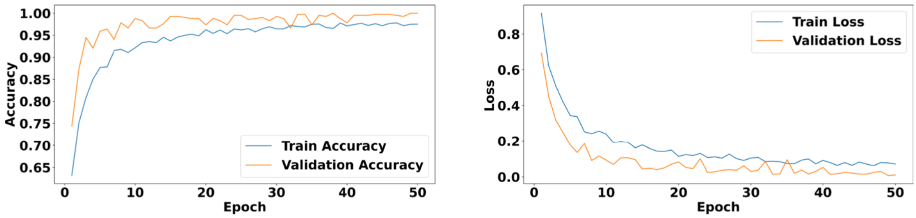


(a) Accuracy (b) Loss

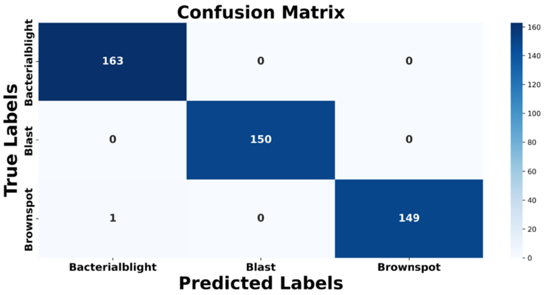


(c) Confusion Matrix (CM)

Fig. 3 RiceNet Performance Visualization on Large Dataset using Ensemble Attention Heads (32), Batch Size (4), and Split Ratio [1156, 1156, 2312] (Train, Validation, and Test)

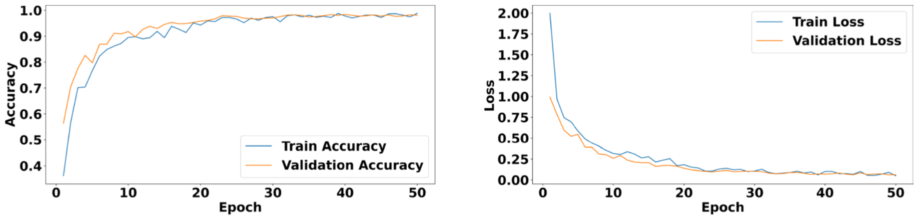


(a) Accuracy (b) Loss

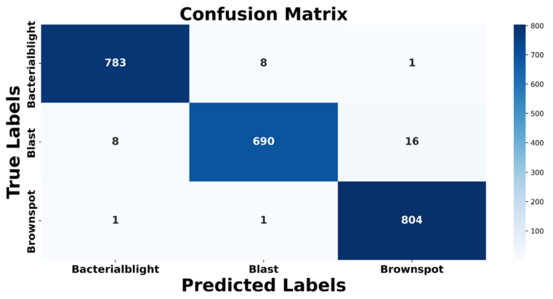


(c) Confusion Matrix (CM)

Fig. 4 RiceNet Performance Visualization on Large Dataset using Ensemble Attention Heads (32), Batch Size (4), and Split Ratio [3744, 417, 463] (Train, Validation, and Test)

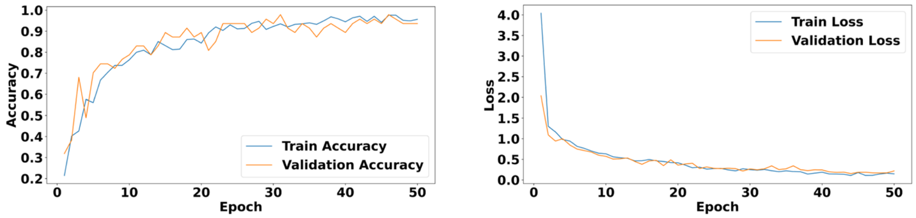


(a) Accuracy (b) Loss

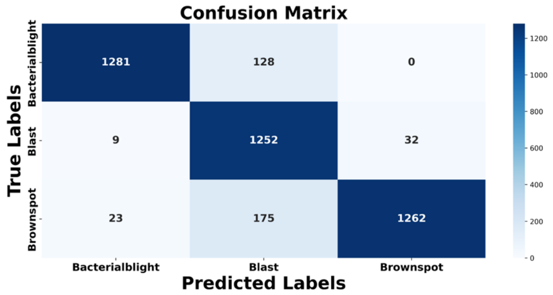


(c) Confusion Matrix (CM)

Fig. 5 RiceNet Performance Visualization on Large Dataset using Ensemble Attention Heads (32), Batch Size (32), and Split Ratio [1156, 1156, 2312] (Train, Validation, and Test)

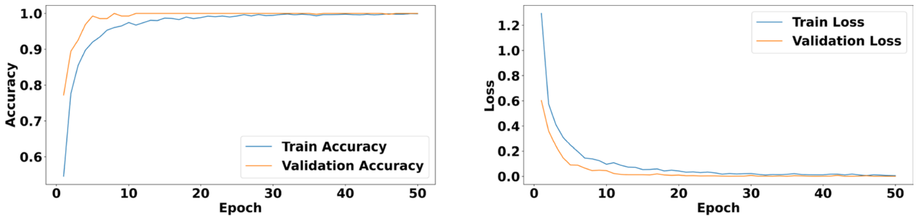


(a) Accuracy (b) Loss

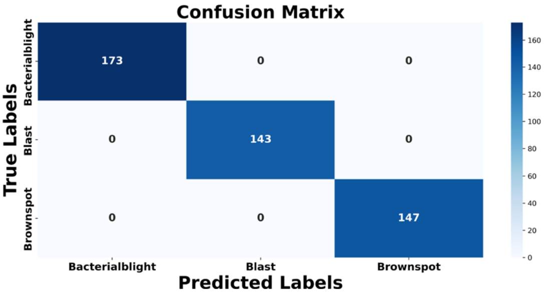


(c) Confusion Matrix (CM)

Fig. 6 RiceNet Performance Visualization on Large Dataset using Ensemble Attention Heads (32), Batch Size (32), and Split Ratio [415, 47, 4162] (Train, Validation, and Test)

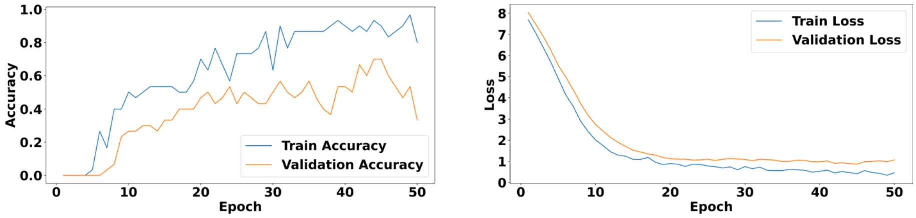


(a) Accuracy (b) Loss

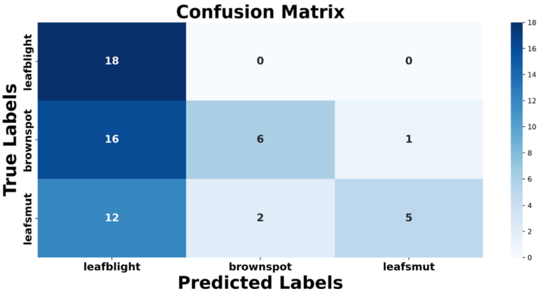


(c) Confusion Matrix (CM)

Fig. 7 RiceNet Performance Visualization on Large Dataset using Ensemble Attention Heads (32), Batch Size (32), and Split Ratio [3744, 417, 463] (Train, Validation, and Test)

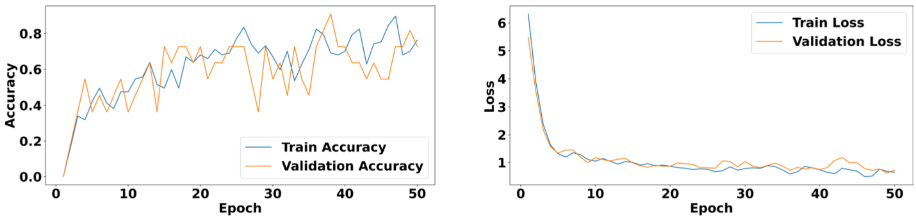


(a) Accuracy (b) Loss

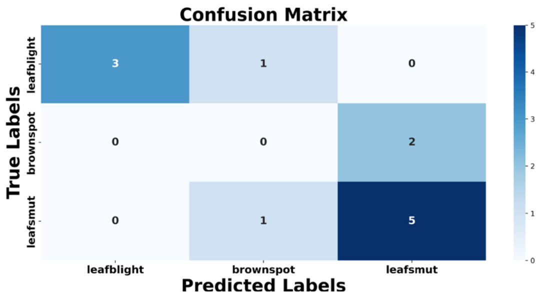


(c) Confusion Matrix (CM)

Fig. 8 RiceNet Performance Visualization on Small Dataset using Ensemble Attention Heads (32), Batch Size (32), and Split Ratio [30, 30, 60] (Train, Validation, and Test)

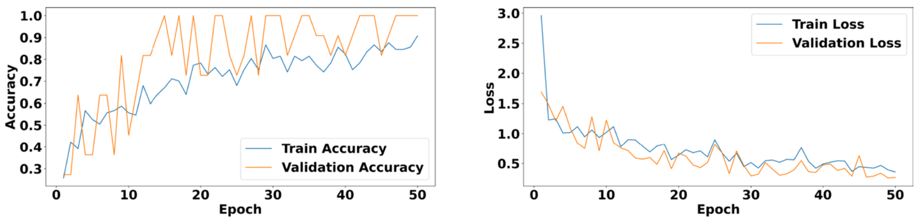


(a) Accuracy (b) Loss

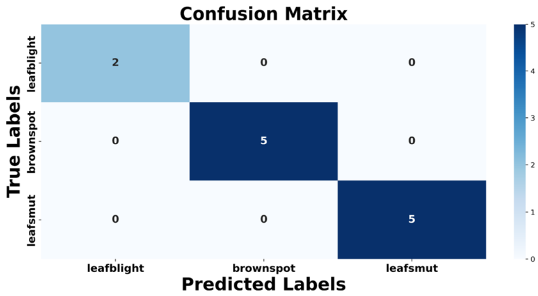


(c) Confusion Matrix (CM)

Fig. 9 RiceNet Performance Visualization on Small Dataset using Ensemble Attention Heads (32), Batch Size (32), and Split Ratio [97, 11, 12] (Train, Validation, and Test)



(a) Accuracy (b) Loss



(c) Confusion Matrix (CM)

Fig. 10 RiceNet Performance Visualization on Small Dataset using Ensemble Attention Heads (32), Batch Size (4), and Split Ratio [97, 11, 12] (Train, Validation, and Test) the samples in the dataset. The confusion matrix offers a detailed breakdown of the classification performance for individual classes, such as Bacterial blight, Blast, and Brownspot. Similarly, Figs. 3 to 7 present the performance visualization of RiceNet on a large dataset, using different ensemble attention heads, batch size, and with a different split ratio as indicated into the caption of each figure.

The subsequent figures (Figs. 8 to 10) continue to visualize and evaluate the performance of RiceNet on small dataset using different ensemble attention heads, batch sizes, and split ratio. These figures showcase the results obtained on a small dataset, comprehensively evaluating the RiceNet framework's performance across different settings. The visualizations highlight the framework's ability to handle dataset heterogeneity and feature imbalance, resulting in improved classification performance compared to traditional methods. The findings also validate the adaptability and robustness of the RiceNet framework, making it a promising approach for automated rice plant disease classification tasks. The comprehensive evaluation of the RiceNet framework on large and small datasets strengthens its potential as a valuable approach for agricultural researchers and practitioners.

Table 4 The study presents the performance evaluation of various baseline models, including VGG-16 and VGG-19 ensembles, and the RiceNet model, for rice leaf disease classification. The experiments were conducted on a small dataset captured in a controlled environment, with a single rice leaf per sample. In the first experiment, using a dataset split of [30, 30, 60] for training, validation, and testing, the models achieved precision scores of 0.7167 (micro), 0.7194 (macro), and 0.7194 (weighted). The recall scores were 0.7167 (micro, macro), and 0.7167 (weighted), with F1 scores of 0.7167 (micro), 0.7154 (macro), and 0.7154 (weighted). The balanced accuracy was 0.7167, and Cohen's Kappa reached 0.5750. The average time per epoch was 0.12 seconds. In a subsequent experiment with a larger dataset split of [97, 11, 12], the precision scores improved to 0.7500 (micro), 0.6944 (macro), and 0.7708 (weighted). The recall scores were 0.7500 (micro), 0.8111 (macro), and 0.7500 (weighted), with F1 scores of 0.7500 (micro), 0.7222 (macro), and 0.7500 (weighted). The balanced accuracy increased to 0.8111, and Cohen's Kappa was 0.5814. However, when trained on a smaller dataset split of [10, 2, 108], the model's performance decreased, resulting in precision scores of 0.5278 (micro), 0.6302 (macro), and 0.6302 (weighted). The recall scores were 0.5278 (micro, macro, weighted), and F1 scores were 0.5278 (micro), 0.4815 (macro), and 0.4815 (weighted). The balanced accuracy dropped to 0.2917, and Cohen's Kappa was 0.07. The average time per epoch for this configuration was 0.07 seconds.

Additionally, the study evaluated the performance of InceptionNet (V1, V3) ensemble models using different configurations. When combined with EAHB 32 and a batch size of 4, and trained on a dataset split of [30, 30, 60] for training, validation, and testing, the model achieved impressive precision scores of 0.8500 (micro), 0.8608 (macro), and 0.8639 (weighted). Similarly, the recall scores were 0.8500 (micro), 0.8524 (macro), and 0.8500 (weighted), while the F1 scores reached 0.8500 (micro), 0.8417 (macro), and 0.8477 (weighted). The model exhibited a balanced accuracy of 0.8524 and Cohen's Kappa of 0.7756. The average time per epoch for this configuration was 1.09 seconds. However, the model's performance decreased notably when trained on a smaller dataset [10, 2, 108]. The precision scores dropped to 0.5000 (micro), 0.3939 (macro), and 0.3830 (weighted), while the recall scores were 0.5000 (micro), 0.5143 (macro), and 0.5000 (weighted). The F1 scores experienced a significant decrease, reaching 0.5000 (micro), 0.04089 (macro), and 0.3975

Table 4 Performance Metrics of ensembles with attention mechanism on small dataset

Backbone	NS	SR (Tr,Vl,Ts)	P (Mi)	P (Ma)	P (W)	R (Mi)	R (Ma)	R (W)	F1 (Mi)	F1 (Ma)	F1 (W)	BA	CK	Avg. T
VGG(16,19)	EAHB(32, 4)	30, 30, 60	0.7167	0.7194	0.7194	0.7167	0.7167	0.7167	0.7167	0.7154	0.7154	0.7167	0.5750	0.12
VGG(16,19)	EAHB(32, 4)	97, 11, 12	0.7500	0.6944	0.7708	0.7500	0.8111	0.7500	0.7500	0.7222	0.7500	0.8111	0.5814	0.42
VGG(16,19)	EAHB(32, 4)	10, 2, 108	0.5278	0.6302	0.6302	0.5278	0.5278	0.5278	0.5278	0.4815	0.4815	0.5278	0.2917	0.07
InceptionNet	EAHB(32, 4)	30, 30, 60	0.8500	0.8608	0.8639	0.8500	0.8524	0.8500	0.8500	0.8417	0.8477	0.8524	0.7756	1.09
InceptionNet	EAHB(32, 4)	10, 2, 108	0.5000	0.3939	0.3830	0.5000	0.5143	0.5000	0.5000	.04089	0.3975	0.5143	0.2603	0.66
RegNet	EAHB(32, 4)	30, 30, 60	0.8833	0.9008	0.8982	0.8833	0.8638	0.8833	0.8833	0.8670	0.8772	0.8638	0.8238	0.99
RegNet	EAHB(32, 4)	97, 11, 12	0.9167	0.8889	0.9444	0.9167	0.9524	0.9167	0.9167	0.9077	0.9218	0.9524	0.8621	3.17
RegNet	EAHB(32, 4)	10, 2, 108	0.6389	0.7027	0.7114	0.6389	0.6546	0.6389	0.6389	0.6386	0.6358	0.6546	0.4662	0.36
RiceNet	EAHB(32,32)	10, 2, 108	0.4074	0.1946	0.2530	0.4074	0.3092	0.4074	0.4074	0.2246	0.2939	0.4123	0.1211	0.53
RiceNet	EAHB(32,32)	30, 30, 60	0.4833	0.6582	0.6688	0.4833	0.5080	0.4833	0.4833	0.4499	0.4438	0.5080	0.2482	1.02
RiceNet	EAHB(32,32)	97, 11, 12	0.6667	0.5714	0.6905	0.6667	0.5278	0.6667	0.6667	0.5421	0.6703	0.5278	0.4419	1.54
RiceNet	**EAHB(32,4)**	**97, 11, 12**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**0.47**

The bold entries indicate the best performance results and are intended to highlight key findings (weighted). The balanced accuracy was measured at 0.2603, and Cohen's Kappa reached 0.66. The average time per epoch for this configuration was recorded as 0.66 seconds.

The study also examined the performance of RegNet (Y-16, Y-32) ensemble models combined with EAHB 32 and a batch size of 4 using different dataset split sizes. When trained on a dataset split of [30, 30, 60] for training, validation, and testing, the model achieved precision scores of 0.8833 (micro), 0.9008 (macro), and 0.8982 (weighted). The recall scores were 0.8833 (micro), 0.8638 (macro), and 0.8833 (weighted), while the F1 scores reached 0.8833 (micro), 0.8670 (macro), and 0.8772 (weighted). Additionally, the balanced accuracy was measured at 0.8638, and the Cohen's Kappa coefficient was 0.8238. In another experiment with a dataset split of [97, 11, 12], the precision scores improved to 0.9167 (micro), 0.8889 (macro), and 0.9444 (weighted). The recall scores were 0.9167 (micro), 0.9524 (macro), and 0.9167 (weighted), while the F1 scores were 0.9167 (micro), 0.9077 (macro), and 0.9218 (weighted). The balanced accuracy increased to 0.9524, and the Cohen's Kappa coefficient was 0.8621. However, when trained on a split of the data set of [10, 2, 108], the performance of the model decreased, resulting in precision scores of 0.6389 (micro), 0.7027 (macro) and 0.7114 (weighted). The recall scores were 0.6389 (micro), 0.6546 (macro), and 0.6389 (weighted), while the F1 scores were 0.6389 (micro), 0.6386 (macro), and 0.6358 (weighted). The balanced accuracy dropped to 0.6546, and the Cohen's Kappa coefficient was 0.4662. The average time per epoch for this configuration was 0.36 seconds.

RiceNet, which customizes ResNet-18 and ResNet-50, also emerges as the outperformed ensemble model among all baseline models evaluated on a small dataset in terms of its overall performance across different network settings. When trained ensemble attention heads EAH 32 and batch size of 32 on a dataset split size of [30, 30, 60], RiceNet's performance improved significantly, with precision scores of 0.4833 (micro), 0.6582 (macro), and 0.6688 (weighted). This indicates the model's capability to generalize and extract meaningful patterns from the data. Furthermore, when evaluated in an even larger split size of data sets [97, 11, 12], RiceNet continued to exhibit strong performance, achieving precision scores of 0.6667 (micro), 0.5714 (macro) and 0.6905 (weighted). This demonstrates the robustness and adaptability of the model to different dataset sizes. When combined with EAHB 32 and batch size 4 on the same dataset split, RiceNet achieved perfect precision, recall, and F1 scores of 1.0000, highlighting its exceptional ability to classify instances in the given task accurately. RiceNet consistently demonstrates its effectiveness and versatility across different configurations, making it the standout model for rice classification.

5 Conclusion and future work

This paper introduces RiceNet, a novel framework for the automated classification of diseases of rice plants. The study addresses the challenges of heterogeneity in dataset and features imbalance commonly found in traditional deep-learning approaches. The proposed framework incorporates fine-tuning transfer learning with ensemble learning principles and attention mechanisms to adaptively balance data heterogeneity and feature representation, effectively mitigating inherent class distribution imbalances. Through comprehensive ablation studies and experiments with baseline models, the effectiveness of each component in the RiceNet framework demonstrated significant improvements in classification performance compared to baselines. The evaluation of RiceNet on extensive publicly available datasets achieves superior performance of impressive F1-Score of (100%) and a Balanced Accuracy of (100%) on large and small datasets. The RiceNet framework significantly advances the automated classification of rice plant diseases. Despite the framework's superior performance, it shows a comparable limitation of the average time per epoch. Future directions for further exploration in this field include expanding the framework's applicability to other crops and plant diseases, conducting more extensive experiments on diverse datasets, and exploring the integration of advanced techniques such as domain adaptation-based transfer learning and Bayesian optimization by focusing on the time efficiency of the framework. These advancements will further enhance the accuracy, efficiency, robustness, and scalability of automated disease classification systems, ultimately benefiting farmers and promoting sustainable agricultural practices.

Author Contributions We sincerely thank all contributors who made this research possible. Dr. Muhammad Hanif Tunio conceptualized the study, designed the methodology and data collection, conducted the experiments, and was responsible for writing, reviewing, and editing. Prof. Jian Ping Li supervised the study. Dr. Awais Ahmed contributed to the conceptualization, review, and editing. Dr. Syed Attique Shah contributed to the conceptualization, review, and editing. Dr. Xiaoyang Zeng participated in the designing, review and editing. Dr. Ubaidullah alias Kashif contributed to the review and editing. Dr. Yingling Li assisted with the review and editing. Dr. Imam Abdullahi Yahya contributed to the review, writing, and editing.

Funding The authors acknowledge that this article received financial support from Birmingham City University, United Kingdom.

Data Availability The datasets used in this study are publicly available online and can be accessed via links: https://data.mendeley.com/datasets/fwcj7stb8r/1 https://www.kaggle.com/datasets/vbookshelf/rice-leaf-diseases

Declarations

Competing Interests The authors declare that they have no any conflict of interest.

References

1. Aggarwal M, Khullar V, Goyal N, Alammari A, Albahar MA, Singh A (2023) Lightweight federated learning for rice leaf disease classification using non independent and identically distributed images. Sustainability 15(16):12149
2. Alsakar YM, Sakr NA, Elmogy M (2024) An enhanced classification system of various rice plant diseases based on multi-level handcrafted feature extraction technique. Sci Rep 14(1):30601
3. Anandhan K, Singh AS (2021) Detection of paddy crops diseases and early diagnosis using faster regional convolutional neural networks. In: 2021 international conference on advance computing and innovative technologies in engineering (ICACITE). IEEE, pp 898–902

4. Anhui Agricultural Network (2025) Anhui agricultural big data center. Accessed: 22-Feb-2024
5. Archana K, Srinivasan S, Bharathi SP, Balamurugan R, Prabakar T, Britto A SF (2022) A novel method to improve computational and classification performance of rice plant disease identification. J Supercomput 1–21
6. Arya A, Mishra PK (2024) Mobilenetv2-incep-m: a hybrid lightweight model for the classification of rice plant diseases. Multimed Tools Appl 1–28
7. Bari BS, Islam MN, Rashid M, Hasan MJ, Razman MAM, Musa RM, Ab Nasir AF, Majeed APA (2021) A real-time approach of diagnosing rice leaf disease using deep learning-based faster r-cnn framework. PeerJ Comput Sci 7:e432
8. Chen J, Zhang D, Zeb A, Nanehkaran YA (2021) Identification of rice plant diseases using lightweight attention networks. Expert Syst Appl 169:114514
9. Daniya T, Vigneshwari S (2022) Deep neural network for disease detection in rice plant using the texture and deep features. Comput J 65(7):1812–1825
10. Daniya T, Vigneshwari S (2023) Rider water wave-enabled deep learning for disease detection in rice plant. Adv Eng Softw 182:103472
11. Dhingra G, Kumar V, Joshi HD (2018) Study of digital image processing techniques for leaf disease detection and classification. Multimed Tools Appl 77:19951–20000
12. Dogra R, Rani S, Singh A, Albahar MA, Barrera AE, Alkhayyat A (2023) Deep learning model for detection of brown spot rice leaf disease with smart agriculture. Comput Electr Eng 109:108659
13. Dubey RK, Choubey DK (2024) An efficient adaptive feature selection with deep learning model-based paddy plant leaf disease classification. Multimed Tools Appl 83(8):22639–22661
14. Dutta M, Islam Sujan MR, Mojumdar MU, Chakraborty NR, Marouf AA, Rokne JG, Alhajj R (2024) Rice leaf disease classification—a comparative approach using convolutional neural network (cnn), cascading autoencoder with attention residual u-net (caar-u-net), and mobilenet-v2 architectures. Technologies 12(11):214
15. Ghosal S, Sarkar K (2020) Rice leaf diseases classification using cnn with transfer learning. In: 2020 IEEE calcutta conference (CALCON). IEEE, pp 230–236
16. Goluguri NRR, Devi KS, Srinivasan P (2021) Rice-net: an efficient artificial fish swarm optimization applied deep convolutional neural network model for identifying the oryza sativa diseases. Neural Comput Appl 33(11):5869–5884
17. Hasan MM, Rahman T, Uddin AS, Galib SM, Akhond MR, Uddin MJ, Hossain MA (2023) Enhancing rice crop management: disease classification using convolutional neural networks and mobile application integration. Agriculture 13(8):1549
18. Jiang Z, Dong Z, Jiang W, Yang Y (2021) Recognition of rice leaf diseases and wheat leaf diseases based on multi-task deep transfer learning. Comput Electron Agri 186:106184
19. Khush GS (2005) What it will take to feed 5.0 billion rice consumers in 2030. Plant Mole Biol 59:1–6
20. Kumar M, Gupta P, Madhav P et al (2020) Disease detection in coffee plants using convolutional neural network. In: 2020 5th International conference on communication and electronics systems (ICCES). IEEE, pp 755–760
21. Latif G, Abdelhamid SE, Mallouhy RE, Alghazo J, Kazimi ZA (2022) Deep learning utilization in agriculture: detection of rice plant diseases using an improved cnn model. Plants 11(17):2230
22. Li J-Y, Wang J, Zeigler RS (2014) The 3,000 rice genomes project: new opportunities and challenges for future rice research. Gigascience 3(1):2047–217X
23. Lu Y, Tao X, Jiang F, Du J, Li G, Liu Y (2024) Image recognition of rice leaf diseases using atrous convolutional neural network and improved transfer learning algorithm. Multimed Tools Appl 83(5):12799–12817
24. Lu Y, Yi S, Zeng N, Liu Y, Zhang Y (2017) Identification of rice diseases using deep convolutional neural networks. Neurocomputing 267:378–384
25. Mukherjee R, Ghosh A, Chakraborty C, De JN, Mishra DP (2025) Rice leaf disease identification and classification using machine learning techniques: a comprehensive review. Eng Appl Artif Intell 139:109639
26. Nalini S, Krishnaraj N, Jayasankar T, Vinothkumar K, Britto ASF, Subramaniam K, Bharatiraja C (2021) Paddy leaf disease detection using an optimized deep neural network. Comput Mater Contin 68(1):1117–1128
27. Nayak A, Chakraborty S, Swain DK (2023) Application of smartphone-image processing and transfer learning for rice disease and nutrient deficiency detection. Smart Agricul Technol 4:100195
28. Patel B, Sharaff A (2023) Automatic rice plant's disease diagnosis using gated recurrent network. Multimed Tools Appl 1–20
29. Patil RR, Kumar S (2022) Rice-fusion: a multimodality data fusion framework for rice disease diagnosis. IEEE Access 10:5207–5222

30. Pertot I, Kuflik T, Gordon I, Freeman S, Elad Y (2012) Identificator: a web-based tool for visual plant disease identification, a proof of concept with a case study on strawberry. Comput Electron Agri 84:144–154
31. Phadikar S, Sil J, Das AK (2013) Rice diseases classification using feature selection and rule generation techniques. Comput Electron Agri 90:76–85
32. Prabu M, Chelliah BJ (2023) An intelligent approach using boosted support vector machine based arithmetic optimization algorithm for accurate detection of plant leaf disease. Pattern Anal Appl 26(1):367–379
33. Rahman CR, Arko PS, Ali ME, Khan MAI, Apon SH, Nowrin F, Wasif A (2020) Identification and recognition of rice diseases and pests using convolutional neural networks. Biosyst Eng 194:112–120
34. Sethy PK (2020) Rice leaf disease image samples. Mendeley Data 1:2020
35. Shafik W, Tufail A, Liyanage De Silva C, Apong Awg Haji Mohd RA (2025) A novel hybrid inception-xception convolutional neural network for efficient plant disease classification and detection. Sci Rep 15(1):3936
36. Shoaib M, Shah B, Ullah I, Ali F, Park SH (2022) Deep learning-based segmentation and classification of leaf images for detection of tomato plant disease. Front Plant Sci 13:1031748
37. Sudhesh K, Sowmya V, Kurian S, Sikha O (2023) Ai based rice leaf disease identification enhanced by dynamic mode decomposition. Eng Appl Artif Intell 120:105836
38. Talukder MSH, Sarkar AK (2023) Nutrients deficiency diagnosis of rice crop by weighted average ensemble learning. Smart Agri Technol 4:100155
39. Tholkapiyan M, Aruna Devi B, Bhatt D, Saravana Kumar E, Kirubakaran S, Kumar R (2023) Performance analysis of rice plant diseases identification and classification methodology. Wirel Person Commun 130(2):1317–1341
40. Tunio MH, Jianping L, Butt MHF, Memon I (2021) Identification and classification of rice plant disease using hybrid transfer learning. In: 2021 18th International computer conference on wavelet active media technology and information processing (ICCWAMTIP). IEEE, pp 525–529
41. Tunio MH, Jianping L, Butt MHF, Memon I, Magsi Y (2022) Fruit detection and segmentation using customized deep learning techniques. In: 2022 19th International computer conference on wavelet active media technology and information processing (ICCWAMTIP). IEEE, pp 1–5
42. Tunio MH, Li JP, Zeng X, Akhtar F, Shah SA, Ahmed A, Yang Y, Heyat MBB (2024) Meta-knowledge guided bayesian optimization framework for robust crop yield estimation. J King Saud Univ Comput Inf Sci 36(1):101895
43. vbookshelf (2018) Rice leaf diseases. Kaggle dataset
44. Wang Y, Wang H, Peng Z (2021) Rice diseases detection and classification using attention based neural network and bayesian optimization. Expert Syst Appl 178:114770
45. Yang H, Deng X, Shen H, Lei Q, Zhang S, Liu N (2023) Disease detection and identification of rice leaf based on improved detection transformer. Agriculture 13(7):1361

Authors and Affiliations

Muhammad Hanif Tunio[1,2] · Jian ping Li[1] · Awais Ahmed[3] · Syed Attique Shah[4] · Xiaoyang Zeng[1] · Ubaidullah alias Kashif[5] · Yingling Li[6] · Abdullahi Yahya Imam[7]

✉ Jian ping Li
jpli2222@uestc.edu.cn

✉ Syed Attique Shah
syedattique.shah@bcu.ac.uk

Muhammad Hanif Tunio
hanif.tunio@salu.edu.pk

Awais Ahmed
ahmedawais@cwnu.edu.cn

Xiaoyang Zeng
202011081605@std.uestc.edu.cn

Ubaidullah alias Kashif
kashif@saus.edu.pk

Yingling Li
liyingling@swun.edu.cn

Abdullahi Yahya Imam
ayimam.it@buk.edu.ng

[1] School of Computer Science and Engineering, University of Electronic Science and Technology of China - UESTC, Sichuan 611731, China

[2] Institute of Computer Science, Shah Abdul Latif University, Khairpur 66111, Sindh, Pakistan

[3] School of Computer Science, China West Normal University, Nanchong, China

[4] School of Computing and Digital Technology, Birmingham City University, STEAMhouse, B4 7RQ Birmingham, United Kingdom

[5] Department of Computer Science, The Shaikh Ayaz University, Shikarpur, Sindh, Pakistan

[6] College of Computer Science and Artificial Intelligence, Southwest Minzu University, Chengdu, China

[7] Faculty of Computing, Bayero University Kano, Kano, Nigeria