

A Comparative Analysis of Transfer Learning, LeafNet, and Modified LeafNet Models for Precise Classification of Rice Leaf Diseases

¹Dr. A. Sherin, ²Mrs. Jayakeerthi M & ³Mrs. J. Divya Jose ⁴Dr. R. Rajeswari

¹Assistant Professor & Head, Department of DCFs, Nehru Arts and Science College, Coimbatore – 641105, TamilNadu, India

²Assistant Professor, Department of Computer Science, Nehru Arts and Science College, Coimbatore – 641105, TamilNadu, India

³Assistant Professor, Department of Computer Science, Nehru Arts and Science College, Coimbatore – 641105, TamilNadu, India

⁴Professor, Department of Computer Applications, Bharathiar University. Coimbatore – 641 046

Abstract:- This research focuses on the timely identification of plant diseases, crucial for efficient crop disease management to mitigate yield loss. The study introduces a methodology for classifying diseases in rice leaves utilizing four distinct deep learning models and a dataset comprising 2658 images of both healthy and diseased rice leaves. The compared models include LeafNet, Modified LeafNet, MobileNetV2, and Xception. Modifications to LeafNet's architectural parameters were implemented in the Modified LeafNet model, while transfer learning techniques were applied to pretrained MobileNetV2 and Xception models. Optimal training hyperparameters were determined by considering various factors such as batch size, data augmentation, learning rate, and optimizers. Notably, the Modified LeafNet model demonstrated the highest accuracies, achieving 97.44% and 87.76% for the validation and testing datasets, respectively. Comparatively, LeafNet achieved 88.92% and 71.84%, Xception achieved 88.64% and 71.95%, and MobileNetV2 achieved 82.10% and 67.68% for validation and test accuracies on the same datasets, respectively. This research significantly contributes to the advancement of automated disease classification systems for rice leaves, thereby enhancing agricultural productivity and sustainability.

Keywords: Plant diseases, Rice leaves, Deep learning models

1. Introduction

Agriculture has been a pivotal development in the progression of human civilization, enabling urban living by facilitating surplus food generation through crop cultivation. However, agricultural productivity and financial performance have often been hampered by diseases affecting crops, leading to significant challenges in large-scale agricultural operations and human sustenance [1]. Moreover, environmental factors and climate variations in tropical and temperate regions have further exacerbated these challenges, posing significant threats to crop production [2], [3].

Timely detection of plant diseases is paramount for ensuring healthy food production and plays a crucial role in ecological research on plant dynamics. Farmers often face difficulties in accurately diagnosing plant diseases due to their subtle manifestations [4]. According to a report by the Foreign Agricultural Service of the United States Department of Agriculture (USDA), rice cultivation covered approximately 163.99 million hectares in 2020 and 2021, yielding an average of 4.57 metric tons per hectare and a total production of 502.10 million metric tons.

However, there was a 2.16% decline in production compared to previous years [5]. This decline in rice production can be attributed to various diseases affecting crop growth and productivity, with pathogens such as bacteria, fungi, and viruses being major contributors [6]. Common rice diseases include brown spot, hispa, and leaf blast [7]. Implementing effective disease management strategies, including crop rotation, utilizing resistant varieties, and timely fungicide application, is crucial for minimizing the adverse effects of these diseases on rice production [8].

Chemical pesticides serve as a common method for controlling rice diseases [9], typically applied to paddy fields to eliminate or inhibit the growth of pathogens like bacteria, fungi, and viruses responsible for diseases. These chemicals are administered as foliar sprays, seed treatments, or soil drenches, depending on the disease type and stage [8]. While chemical pesticides can effectively manage rice diseases, they also pose potential environmental and health risks [10].

Early identification and diagnosis of rice diseases are vital for effective disease management. Timely recognition of the disease and its severity allows farmers to implement measures to contain its spread and minimize yield losses. Moreover, early diagnosis enables the adoption of targeted and less hazardous control methods, such as cultural practices or the use of less toxic pesticides. Rapid and accurate diagnosis involves a combination of visual inspection, laboratory tests, and diagnostic tools like molecular assays [11]. Modern diagnostic tools, including remote sensing and machine learning-based algorithms, also facilitate early detection of rice diseases [12].

Convolutional neural networks (CNNs) are adept at performing convolutional operations on input data, offering a robust system architecture capable of addressing complex problems. Comprising input layers, alternating convolutional and pooling layers, fully-connected layers, and an output layer, CNN models leverage autonomous learning and feature extraction capabilities to automatically extract image features for classification and identification [13]. Convolutional layers facilitate feature extraction, with the extracted features forwarded to the fully-connected layer for classification. CNN stands out as one of the most utilized deep learning architectures, boasting significant model capacity and the ability to handle complex information [13], [14].

In recent years, computer vision techniques have become prevalent in detection and classification tasks [15]. These techniques involve employing image processing algorithms to analyze digital plant images and detect signs of disease [16]. Computer vision aids in detecting diseases at early stages, often before symptoms are visible to the naked eye. Common computer vision techniques for plant disease detection include:

1. Image Segmentation: Dividing images into distinct regions or segments with similar characteristics enables the identification and isolation of specific plant parts, such as leaves, for further analysis [17].
2. Feature Extraction: Extracting pertinent information from images, including texture, color, and shape, facilitates the classification of images as healthy or diseased [18].

Machine learning (ML) algorithms, such as support vector machines, offer a valuable approach for classifying images based on extracted features. These algorithms leverage labeled datasets of healthy and diseased plant images for training and can subsequently classify new plant images [19].

Transfer learning, a technique in deep learning, utilizes convolutional neural networks (CNNs) pretrained for specific tasks as the foundation for models intended for related tasks. Pretrained networks, initialized with weights learned from large labeled datasets like ImageNet, can significantly reduce the training time and effort compared to starting from scratch. Notably, models like VGGNet, ResNet, Inception V4, DenseNets, and SqueezeNet, initially trained for plant disease classification, heavily rely on transfer learning [20], [21].

Our study focused on classifying diseases in rice leaves, employing the LeafNet model—a CNN-based network proposed by Barré et al. [22]—as the base model. LeafNet is specifically designed to analyze images of plant leaves and classify them into different species.

In this study, we enhanced the LeafNet model's ability to classify rice diseases using rice leaf images through modifications. We compared the performance of the modified model based on critical parameters such as batch size, learning rate, precision, recall, F1-score, and accuracy. Two optimizers, Adaptive Moment Estimation

(Adam) and Root Mean Square Propagation (RMSprop), were utilized for error minimization. Additionally, we incorporated two pretrained models—Xception and MobileNetV2—for comparison. Xception, known for its complex architecture, and MobileNetV2, recognized for its lightweight design, were both subjected to transfer learning using rice leaf images. These pretrained models, initially trained on ImageNet, had their convolutional layers frozen, while the fully-connected layers were trained using the rice leaf image dataset. Hyperparameters were fine-tuned, and the models' performance was evaluated.

The main contributions of this paper are as follows:

(a) **Modified LeafNet Model:** We developed a convolutional neural network capable of accurately classifying rice leaf diseases from images. Compared to the LeafNet model, our modified version exhibited higher classification accuracy (both in validation and testing) on the rice leaf disease dataset.

(b) **Transfer Learning:** We leveraged state-of-the-art models known for superior performance in classification tasks. Xception and MobileNetV2 were utilized in this study, with transfer learning applied to these models by freezing their convolutional layers and employing pretrained weights.

(c) **Explainable Artificial Intelligence (AI):** To provide insights into the model's decision-making process, we utilized the intermediate class activation map technique to visualize the pixels that exerted the most influence on the model's predictions for specific classes.

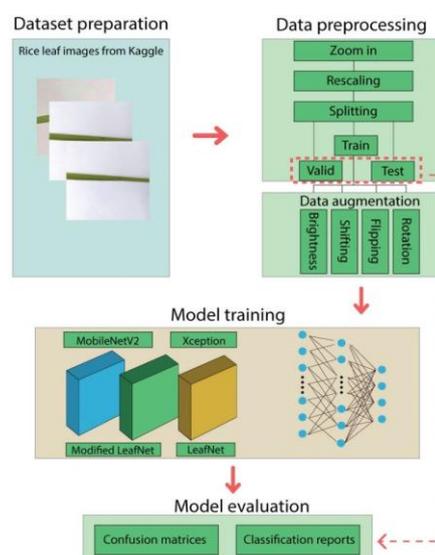
2. Objectives

The objective of this study is to conduct a comparative analysis of three different models for precise classification of rice leaf diseases. Transfer learning, LeafNet, and a modified version of LeafNet will be evaluated to determine their effectiveness in accurately identifying various diseases affecting rice plants based on leaf images. The aim is to assess the performance, robustness, and efficiency of each model in this specific task, considering factors such as classification accuracy, computational resources required, and suitability for practical deployment in agricultural settings. This analysis will provide insights into selecting the most suitable approach for disease diagnosis in rice plants using machine learning techniques.

3. Methods

The proposed framework for disease classification from rice leaf images is depicted in Fig. 1. The primary objective of this study is to categorize rice leaf diseases into four classes: brown spot, leaf blast, hispa, and healthy. The key stages of our methodology include dataset preparation, data preprocessing (including augmentation of leaf images), model training, and model evaluation based on disease classification in the provided images.

FIGURE 1. Overview of the methodology used in this study.



A. DATA COLLECTION

In this investigation, a collection of rice leaf images was employed to classify rice leaf diseases, comprising four categories: brown spot, leaf blast, hispa, and healthy. A total of 2658 labeled images were obtained from publicly accessible datasets accessible via the Kaggle platform [36], [37]. Subsequently, the dataset was divided into training, validation, and testing subsets.

The training and validation subsets were utilized during the model training process, while the testing subset remained unseen to the model. Model assessment was conducted using the validation and testing subsets. Table 1 provides an overview of the dataset utilized in this study. Additionally, Figure 2 displays sample images representing each class within the dataset. It's worth noting that the leaf images were originally in JPEG format upon retrieval from the Kaggle platform.

TABLE 1. Data splitting details.

Classes	Train	Valid	Test	Total
Brown Spot	360	90	86	536
Hispa	512	128	113	753
Leaf Blast	361	91	151	603
Healthy	499	125	142	766

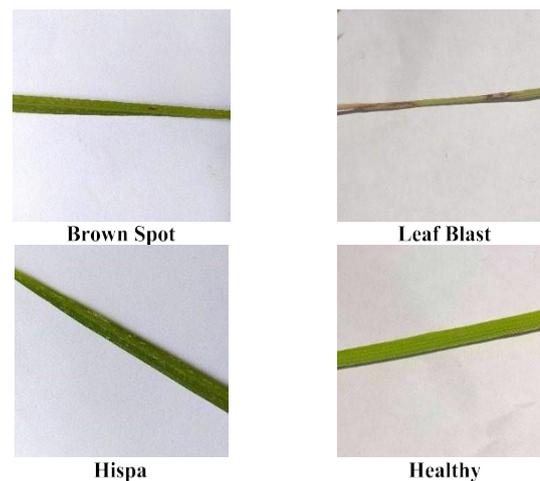


FIGURE 2. Samples of images obtained from the Kaggle platform.

B. IMAGE PROCESSING

The initial step in data preprocessing involved zooming in on the images to enhance the visibility of spots or infected regions on the rice leaves, as depicted in Fig. 3, which illustrates images from the original dataset alongside their zoomed-in versions. Following this, the preprocessed images underwent rescaling and resizing. During rescaling, the pixel values of the images were adjusted by multiplying them by a factor of $1/255$, ensuring that each pixel value fell within the range of 0 to 1. This normalization process aimed to standardize the dataset to a consistent scale.

Addressing real-world challenges in rice leaf disease detection is crucial, particularly given the diversity of imaging conditions that can introduce variability and potential artifacts into images. To mitigate this, we employed data augmentation techniques to enhance the robustness and real-world applicability of our models. This involved applying various augmentation techniques to the training images, including rotation by 30 degrees, horizontal and

vertical flipping, adjusting the height and width within a range of 80% to 120%, and modifying the brightness within a range of 80% to 120%.

Rigorous validation and testing procedures were conducted to evaluate the models' ability to handle potential artifacts or alterations introduced during the preprocessing phase. These preprocessing techniques can be integrated into the system to facilitate the automated processing of real-world unprocessed images before their utilization in disease classification.

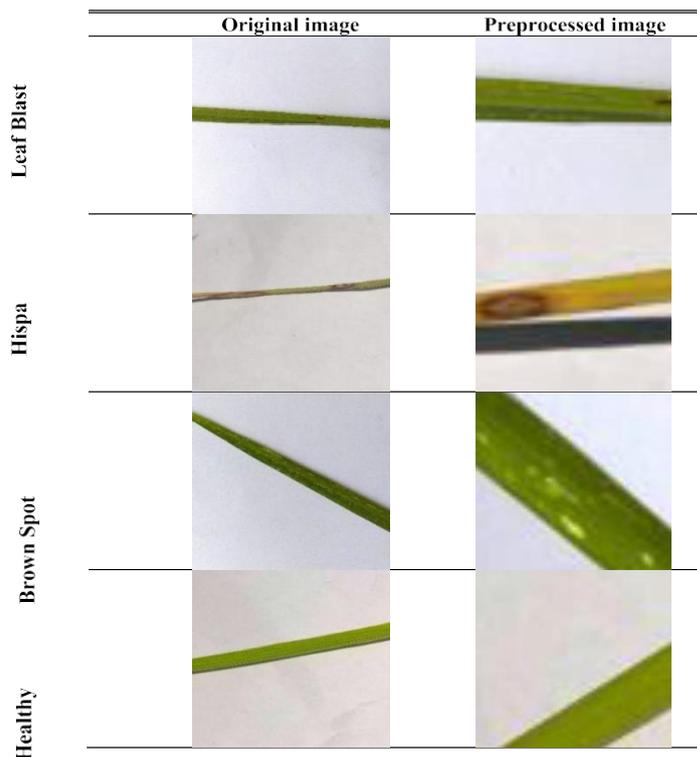


FIGURE 3. Examples of original (left) and preprocessed (right) images used in this study.

C. LEAFNET ARCHITECTURE

In this study, we adopted the LeafNet architecture proposed by Barré et al. [22] as the foundational model. The LeafNet model consists of 11 convolutional layers followed by three fully-connected layers. The architecture of the LeafNet model is illustrated in Fig. 4. During training, the LeafNet model processed inputs of size 256×256 and utilized a max pooling filter of size 3×3 . Each convolutional layer employed varying kernel filters, with the first two layers utilizing a filter shape of 9×9 , the subsequent two layers using a filter shape of 5×5 , and the remaining layers employing a kernel filter shape of 3×3 . The fully-connected layers of the model comprised two dense layers, each consisting of 2048 neurons, followed by an output layer with a Softmax activation function. Notably, the output layer was configured to classify inputs into four classes, unlike the original LeafNet model, which was designed to classify inputs into 185 classes.

D. Modified Leafnet Architecture

To enhance the classification of rice leaf diseases, we proposed a modified version of the LeafNet model. The architecture and parameters of each layer in the Modified LeafNet model are detailed in Table 2, and the schematic representation of the Modified LeafNet model is depicted in Fig. 5.

Several modifications were made to the original LeafNet model architecture to improve its performance. Notably, we standardized the kernel size to 3×3 across all convolutional layers, whereas the original model employed kernel sizes of 9×9 , 5×5 , and 3×3 . Additionally, we utilized a 3×3 max-pooling layer, whereas the original LeafNet

model employed a 2×2 max-pooling layer. These alterations were essential for achieving enhanced results tailored to our specific application. In the Modified LeafNet model, the output layer comprised four neurons representing four distinct classes, unlike the original LeafNet model, which featured 185 neurons representing 185 different classes. These adjustments contributed to improved model performance by effectively extracting crucial features for better classification.

Table 3 provides an overview of the input parameters for all models evaluated in this study, while Table 4 offers a comprehensive summary of these models. The process flow structure is outlined in Table 5, elucidating the steps involved in data preprocessing, model construction, hyperparameter tuning, and model evaluation.

Hyperparameter tuning was conducted using a grid search method to optimize parameters such as batch size, learning rate, and optimizers. Upon training the model for a specified number of epochs, we saved the best weights in the HDF5 format based on the minimum loss value observed in the validation data. These trained weights were subsequently utilized for classifying rice leaf images on a local machine.

E. Transfer Learning With Pretrained Models

In this study, we employed the Xception and MobileNetV2 models to showcase and compare the distinctions between complex and lightweight models. Both models are deep convolutional neural networks with differing input sizes and architectures. The Xception model had an input size of $299 \times 299 \times 3$, while MobileNetV2 had an input size of $224 \times 224 \times 3$. For consistency, we standardized the input size to $224 \times 224 \times 3$ in this study. Both models had undergone extensive training on the ImageNet dataset for the classification of 1000 categories. These pretrained models can be readily imported with their weights using the Keras Application Programming Interface (API).

Transfer learning was implemented on the MobileNetV2 and Xception models by loading the weights obtained from their prior training on the ImageNet dataset and subsequently freezing the convolutional layers. For the Xception model, the convolutional layers were frozen from block1_conv1 (Conv2D) to block14_sepconv2_act (Activation), while for MobileNetV2, the freezing spanned from Conv1 (Conv2D) to out_relu (ReLU) layers. The original fully-connected layers were replaced with Flatten and Dense layers. The Flatten layer converted the feature map obtained from the max-pooling layer into a format compatible with Dense layers, which were responsible for classifying the input. Subsequently, the rice leaf dataset was utilized to train new fully-connected layers for classification.

Table 6 delineates the final layers of the Xception and MobileNetV2 models. Leveraging pretrained models from extensive datasets via transfer learning facilitated the extraction of valuable features without the need for an extensive amount of training data. This not only reduced the computational resources required for training but also led to improved performance, as pretrained models typically acquired a rich set of features during their initial training phase.

Furthermore, transfer learning enables the reuse of meticulously crafted architectures like Xception and MobileNetV2, mitigating the necessity for labor-intensive model design and experimentation. This approach capitalizes on training with large-scale datasets, amplifying the models' capacity to extract meaningful features and enhancing overall performance.

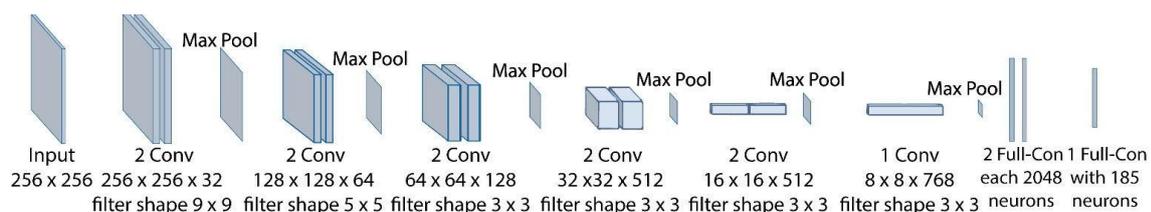


FIGURE 4. Architecture of the original LeafNet model [22].

TABLE 2. Model summary of the modified LeafNet model.

Layers	Output Shape	Parameters
conv2d	224,224,32	155
conv2d_1	220,220,32	1,344
conv2d_2	218,218,32	1,344
Max_Pooling2d	109,109,32	0
Batch_Normalization	109,109,32	128
conv2d_3	107,107,64	2,400
Max_Pooling2d_1	52, 52, 64	0
conv2d_4	105, 105, 64	4,763
Batch_Normalization_1	52, 52, 64	256
conv2d_5	50, 50, 128	8,896
conv2d_6	48, 48, 128	17,664
Max_Pooling2d_2	24, 24, 128	0
Batch_Normalization_2	24, 24, 128	512
conv2d_7	22, 22, 256	34,176
conv2d_8	20, 20, 256	68,096
Max_Pooling2d_3	6, 6, 256	0
Batch_Normalization_3	6, 6, 256	1024
Flatten	9216	0
Dense	128	1,179,776
Dense_1	64	8,256
Dense_2	4	260

F. Visualizing Intermediate Class Activation Maps

In CNN models, visualizing intermediate class activations during training provides deeper insights into the feature extraction process, especially for image-based datasets. These activations represent the output of different layers in the network, with feature maps generated by convolutional and pooling layers. The purpose of visualizing these activations is to understand how the network decomposes an input image using learned filters.

The Intermediate Class Activation Map (ICAM) serves as a visual aid to interpret the decision-making process of CNN models. It showcases the extracted features of important regions in an input image that contribute to the final prediction. By visualizing these regions, one can understand which features the model focuses on and whether they are relevant for making accurate predictions.

To generate an ICAM, the model is modified to produce intermediate activations of specific layers during the forward pass. These activations are then utilized to compute a weighted sum of activations from the last convolutional layer, resulting in a heatmap representation. The weights for this sum are determined by computing the gradient of the model's output concerning the activation of the last convolutional layer. The resulting heatmap highlights the regions in the input image that significantly influence the model's prediction.

Table 7 presents the ICAM for both the LeafNet and modified LeafNet models, illustrating the feature extraction process undertaken by these models. The original image contained affected areas with pale yellow dots indicating leaf blast disease. Subsequent columns display the extracted pixels from both models. While the LeafNet model failed to extract the affected pixels, as seen in column conv2d_8, the Modified LeafNet model successfully captured the most influential pixels. These extracted features, as revealed by ICAM, demonstrate the modified

LeafNet's ability to focus on the most significant and informative pixels, laying the groundwork for accurate classification.

g. model hyperparameters

The performance of the models is influenced by various hyperparameters, including the optimizer, learning rate, metrics, batch size, and epochs. The models minimize the loss function using the Adam or RMSprop optimizer, with learning rates set to 0.001 or 0.0001. The batch size, referring to the number of images fed into the model at a time, was varied between 16 and 32 to assess model performance. Table 8 provides a summary of the hyperparameters used in the study.

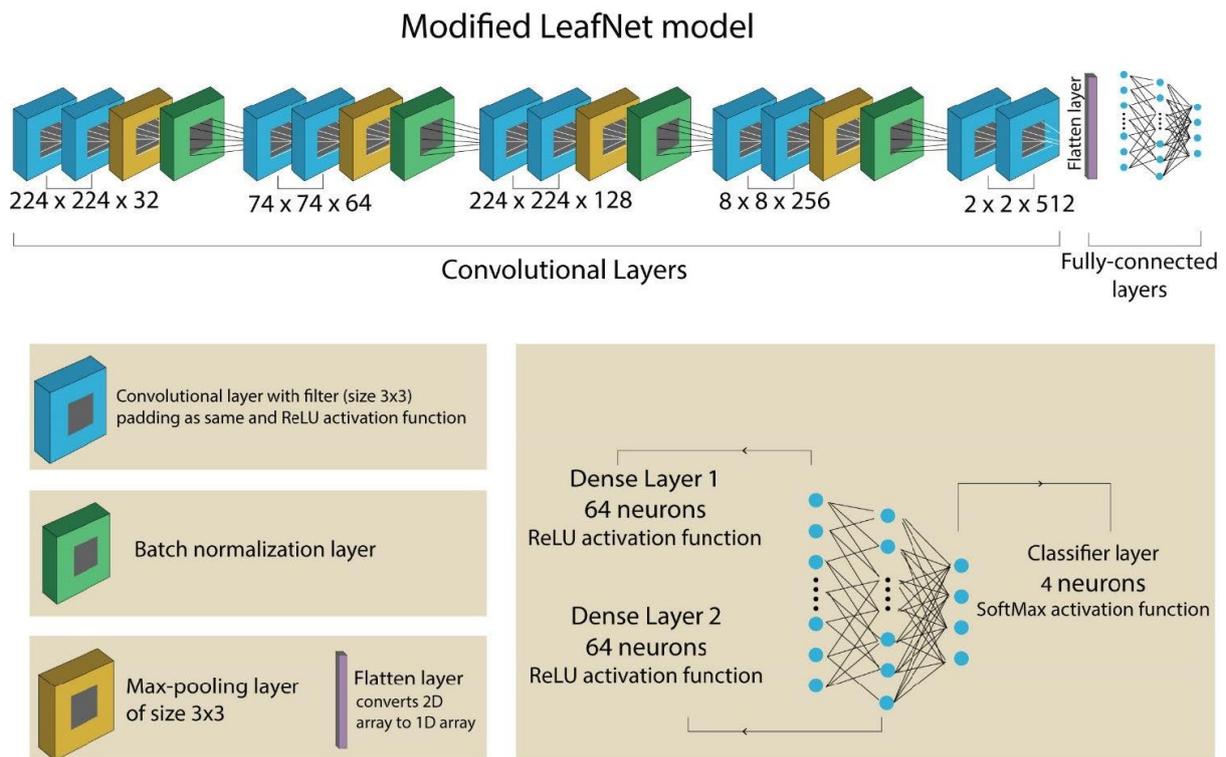


FIGURE 5. Architecture of the Modified LeafNet model.

TABLE 3. Models' parameters used in this study.

Models	Convolutional Layers	Input layer size	Output layer size
LeafNet	11	(224, 224, 3)	(4,1)
Modified LeafNet	10	(224, 224, 3)	(4,1)
MobileNetV2	53	(224, 224, 3)	(4,1)
Xception	71	(299, 299, 3)	(4,1)

TABLE 4. Summary of the models.

Models	Total parameters	Trainable parameters	Non-Trainable parameter	Size (MB)
LeafNet	4,747,236	4,746,254	982	11
Modified LeafNet	1,329,063	1,328,023	960	6
MobileNet V2	6,272,388	4,014,404	2,257,904	14
Xception	27,284,332	6,422,852	20,861,480	88

TABLE 5. Process flow used in this study.

Data Preprocessing	
1:	Importing Libraries
2:	Importing dataset from device
3:	Zooming and saving the new pictures
4:	Train-Valid-Test splitting
5:	Hyperparameters selection
6:	Data Augmentation using image data generator library
Building Models' Architecture	
LeafNet and Modified LeafNet	MobileNetV2 and Xception
Building CNN model from scratch	1: Importing models from Keras libraries
Adding convolutional and Maxpooling layers	Importing ImageNet weights
Adding Flatten and Dense Layers	3: Adding Flatten and Dense Layers
4: Model summary	4: Model summary
Model Training	
1:	Training the model on the training dataset and validation dataset
2:	Print the loss and accuracy graphs
Model Evaluation	
1:	Print confusion metrics on validation and testing dataset

4. Results

The experiments in this study were conducted utilizing Google Colab, employing Python version 3.8, TensorFlow version 2.9.2, and Keras version 2.9.0. The hardware infrastructure employed was an NVIDIA Tesla T4 GPU with driver version 460.32.03.

PERFORMANCE EVALUATION

The evaluation of the LeafNet, Modified LeafNet, MobileNetV2, and Xception models was conducted using both the testing and validation datasets. The Adam optimizer was utilized with a learning rate set to 0.0001. To compare the performance of each model effectively, key evaluation metrics including recall, precision, F1-score, and accuracy were computed. These metrics were derived using equations (1), (2), (3), and (4).

Performance metrics such as True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) were obtained from 4×4 confusion matrices. The confusion matrix with the corresponding terms for each case is presented in Table 9.

TABLE 6. Summary of last layers in the pretrained models.

Models	Layers	Output Shape	Parameters
Xception	Flatten	100,352	0
	Dense	64	6,422,592
	Dropout	64	0

	Dense_1	4	260		
MobileNetV2	Flatten	62,720	0		
	Dense	64	4,014,144		
	Dropout Dense_1	64	4	0	260

INTERPRETATION OF CONFUSION MATRIX

Interpreting a 4x4 confusion matrix can be challenging due to the variation in terms like True Positive (TP) and True Negative (TN) for each class. Table 10 provides the breakdown of TP and TN samples for each class, facilitating a clearer understanding.

Furthermore, False Positive (FP) and False Negative (FN) are crucial components of the confusion matrix, indicating misclassifications. Table 11 presents the FP and FN samples for each class, aiding in the assessment of model performance.

ACCURACY

Accuracy quantifies the model's overall ability to correctly classify both TP and TN instances across all classes. It is calculated using Equation (1):

$$\text{Accuracy} = \frac{TP+TN}{FP+TP+TN+FN} \times 100\%$$

PRECISION

Precision measures the ratio of correctly predicted positive samples (TP) to all positively predicted samples for a given class. It is computed using Equation (2):

$$\text{Precision} = \frac{TP}{TP+FP}$$

RECALL

Recall, also known as sensitivity, assesses the model's ability to correctly identify positive samples (TP) out of all actual positive samples for a particular class. It is calculated using Equation (3):

$$\text{Recall} = \frac{TP}{FN+TP}$$

F1-SCORE

The F1-score is a vital metric that differs from accuracy as it considers both precision and recall. While accuracy measures the ratio of correctly predicted samples to all samples, the F1-score provides a balanced measure of the model's performance in terms of precision and recall. It is calculated as the harmonic mean of precision and recall, as shown in Equation (4):

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

5. Discussion

The training performance of the models was visualized through training and validation curves spanning 400 epochs. Figures 6 and 7 depict the curves for the Modified LeafNet, LeafNet, Xception, and MobileNetV2 models, illustrating their performance under the Adam optimizer with a learning rate of 0.0001 and a batch size of 32. The curves indicated satisfactory performance without signs of overfitting or underfitting.

Confusion matrices and classification reports for the LeafNet and Modified LeafNet models on the testing dataset are presented in Figure 8. The Modified LeafNet model exhibited superior accuracy on the testing dataset, outperforming the LeafNet model with an accuracy of 97.44% on the validation dataset and 87.76% on the testing dataset. Notably, both models showed the most misclassified cases in the leaf blast class.

A comparison of classification accuracies between the Xception and MobileNetV2 models for rice leaf diseases is shown in Figure 9, which includes confusion matrices, classification reports, and accuracy, precision, recall, and F1-score metrics for each model. The Xception model performed well with a batch size of 32, while MobileNetV2 performed better with a batch size of 16. Models with complex architectures, like Xception, benefitted from larger batch sizes, accelerating the training process. Conversely, the simpler architecture of MobileNetV2 required a smaller batch size for effective learning.

Figure 10 displays the validation accuracies achieved by the models, with the Modified LeafNet model attaining the highest accuracy of 97.44% for rice leaf disease classification. However, the LeafNet, MobileNetV2, and Xception models also demonstrated commendable performance, achieving accuracies of 88.92%, 88.64%, and 82.10%, respectively.

TABLE 7. Intermediate class activation map (ICAM).

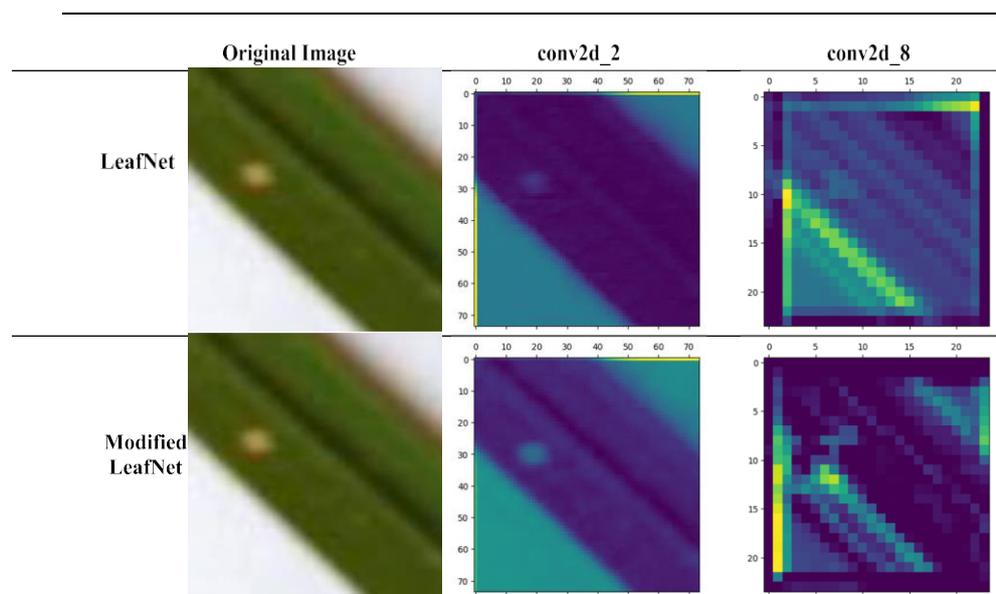


TABLE 8. Hyperparameters used in this study.

Parameters	Value		
Optimizer	Adam		RMSprop
Learning Rate			0.0001
Batch Size	0.001	0.0001	0.001
	16	32	16
			32

TABLE 11. False positive and false

Class	False Positive (FP)	False Negative (FN)
Brown Spot	$F_{BA} + F_{CA} + F_{DA}$	$F_{AB} + F_{AC} + F_{AD}$
Hispa	$F_{AC} + F_{BC} + F_{DC}$	$F_{CA} + F_{CB} + F_{CD}$
Leaf Blast	$F_{AB} + F_{CB} + F_{DB}$	$F_{BA} + F_{BC} + F_{BD}$
Healthy	$F_{AD} + F_{BD} + F_{CD}$	$F_{DA} + F_{DB} + F_{DC}$

TABLE 9. Confusion matrix for four classes.

		Predicted Classes			
		Brown Spot	Leaf Blast	Hispa	Healthy
Actual classes	Brown Spot	TP_{BS}	F_{AB}	F_{AC}	F_{AD}
	Leaf Blast			F_{BA}	TP_{LB}
	Hispa	F_{CA}	F_{CB}	TP_{HISPA}	F_{CD}
	Healthy	F_{DA}	F_{DB}	F_{DC}	TP_{NORMAL}

TABLE 10. True positive and true negative for four classes.

Class	True Positive (TP)	True Negative (TN)
Brown Spot	TP_{BS}	$TP_{LB} + F_{BC} + F_{BD} + F_{CB} + TP_{HISPA} + F_{CD} + F_{DB} + F_{DC} + TP_{NORMAL}$
Hispa	TP_{LB}	$TP_{BS} + F_{AC} + F_{AD} + F_{CA} + TP_{HISPA} + F_{CD} + F_{DA} + F_{DC} + TP_{NORMAL}$
Leaf Blast	TP_{HISPA}	$TP_{BS} + F_{AB} + F_{AD} + F_{BA} + TP_{LB} + F_{BD} + F_{DA} + F_{DB} + TP_{NORMAL}$
Healthy	TP_{NORMAL}	$TP_{BS} + F_{AB} + F_{AC} + F_{BA} + TP_{LB} + F_{BC} + F_{CA} + F_{CB} + TP_{HISPA}$

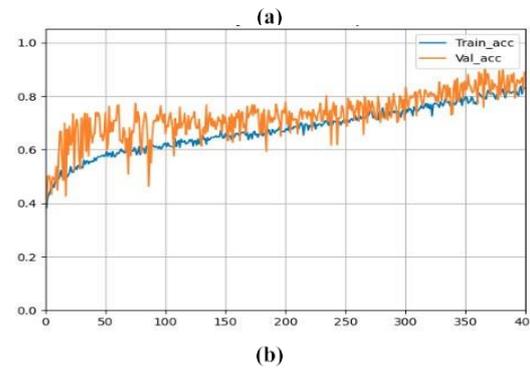
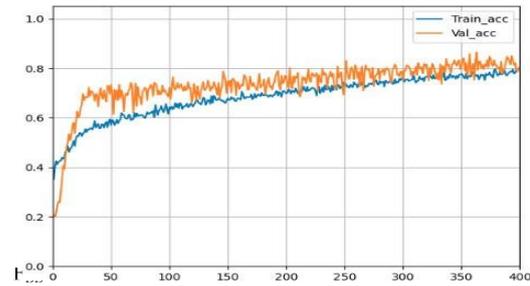
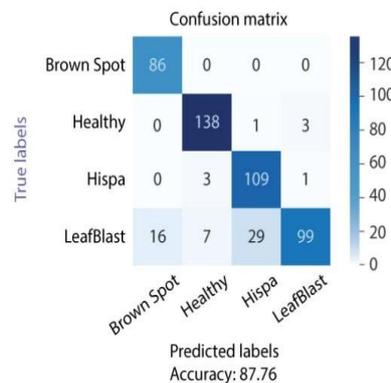


FIGURE 6. Training and validation curves.

(a) LeafNet. (b) ModifiedLeafNet.

Additionally, the models underwent evaluation on the test dataset, which remained unseen during the model training and validation phases. This allowed for an accurate assessment of their real-world performance. As depicted in Figure 10, the Modified LeafNet model exhibited superior performance compared to the other models, achieving a test accuracy of 87.76%. In contrast, the LeafNet, Xception, and MobileNetV2 models achieved test accuracies of 71.84%, 71.95% and 67.68%, respectively.

Figure 7. Training And Validation Curves. (A) Xception. (B) Mobilenetv2.



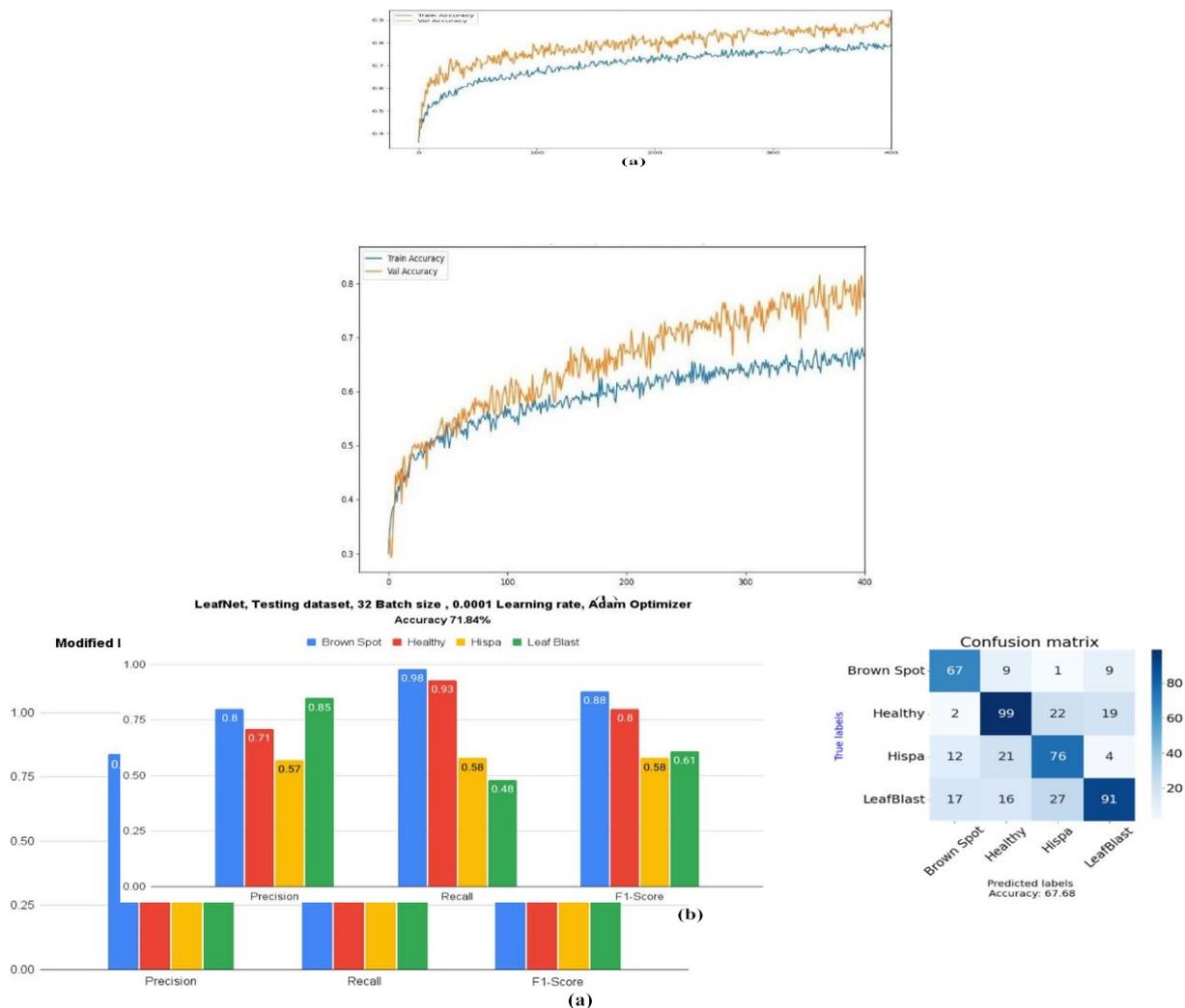


Figure 8. Model Performance On The Testing Dataset. (A) Modified Leafnet Classification Report And Confusion Matrix. (B) LeafnetClassification Report And Confusion Matrix.

One possible explanation for the exceptional performance of the Modified LeafNet model could be its specialized design tailored specifically for classifying rice leaf diseases. This design might incorporate more robust features adept at discerning subtle distinctions among various types of rice leaf ailments. Additionally, the heightened accuracy of the Modified LeafNet model could be attributed to its extensive parameterization, enabling it to capture intricate patterns within the dataset effectively.

In contrast, the MobileNetV2 and Xception models demonstrated respectable performance despite not being explicitly crafted for rice leaf disease classification. Their effectiveness suggests that transfer learning can serve as a potent strategy for categorizing images of rice leaf diseases, even when the models were initially trained on dissimilar image datasets.

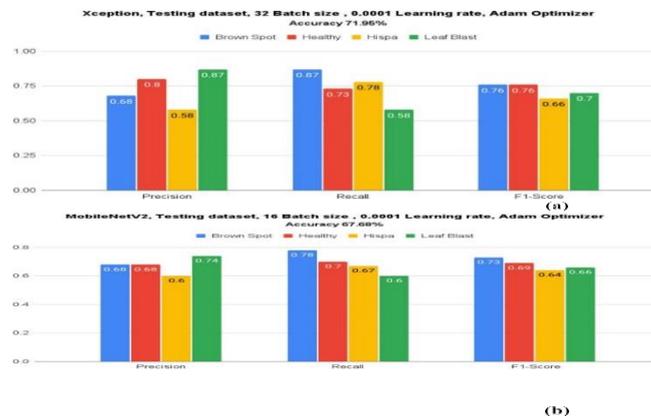


FIGURE 9. Model performance on the testing dataset. (a) Xception classification report and confusion matrix. (b) MobileNetV2 classification report and confusion matrix.

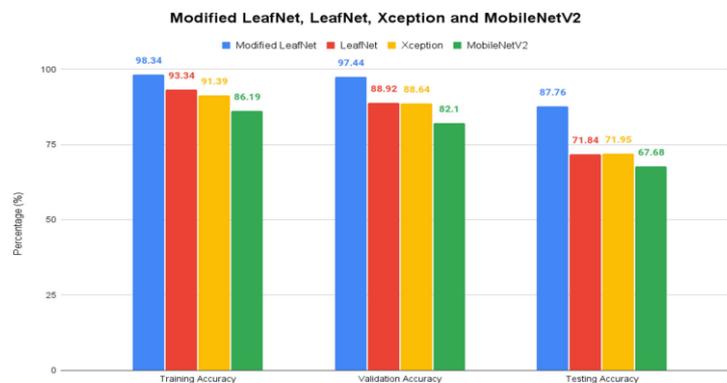


FIGURE 10. Performance comparison among all four models (all models used a batch size of 32, except MobileNetV2, which used a batch size of 16).

The primary objective of this research was to develop a model capable of accurately identifying rice leaf diseases, which presented challenges due to imaging conditions. The small size of the rice leaves and the consistent white background in the images posed difficulties for detection. Initially, all models struggled with low performance due to these conditions. To address this issue, data preprocessing techniques were implemented. This involved zooming in on the images to enhance the visibility of the rice leaves and applying data augmentation, including flips, shifts, and rotations, to the training set. The experiment highlighted the significant impact of image conditions on model performance, emphasizing the importance of image preprocessing.

CNN deep learning models emerged as suitable candidates for rice leaf disease classification. Transfer learning proved beneficial in improving model accuracy, leveraging pretrained CNN models trained on general image datasets like ImageNet. However, successful application of transfer learning required appropriate replacement of the last few layers to align with the study's objective of rice leaf disease classification. While the LeafNet model demonstrated proficiency in detecting and recognizing leaf types, the Modified LeafNet model specifically tailored for rice leaf disease classification yielded improved accuracy in this study.

For future researchers venturing into leaf classification studies, three recommendations are proposed. Firstly, consider utilizing CNN models with transfer learning to capitalize on pretrained models' knowledge. Secondly, explore LeafNet as a viable option, given its strong performance in leaf type recognition. Lastly, prioritize the use of the Modified LeafNet model, as it offers enhanced accuracy specifically tailored for classifying rice leaf diseases.

TABLE 12. Comparative analysis of studies with the same dataset.

Authors	Classes	Algorithm	Accuracy
Zhang [32]	Healthy, Brown Spot, Hispa, and Leaf Blast	WS-DAN	Testing Accuracy: 87.60% Validation Accuracy: N.A.
Putra et al. [33]	Brown Spot, Hispa, and Leaf Blast	HTL	Testing Accuracy: N.A. Validation Accuracy: 91%
Verma et al. [34]	Healthy, Brown Spot, Hispa, and Leaf Blast	Lightweight CNN model	Testing Accuracy: 73.02% Validation Accuracy: N.A.
Bhowmik et al. [35]	Healthy, Brown Spot, Hispa, and Leaf Blast	Ensemble Model (VGG16+Light GBM)	Testing Accuracy: N.A. Validation Accuracy: 96.49%
This study	Healthy, Brown Spot, Hispa, and Leaf Blast	Modified LeafNet model	Testing Accuracy: 87.76% Validation Accuracy: 97.44%

Table 12 provides a comparative analysis of studies utilizing the same dataset as ours for classifying various rice leaf diseases. Four algorithms from the literature are evaluated, with most studies testing on four classes: healthy, brown spot, hispa, and leaf blast. Zhang [32] employed the WS-DAN algorithm, achieving a testing accuracy of 87.60% across all classes. Putra et al. [33] utilized the HTL algorithm, reporting a validation accuracy of 91% for three classes: brown spot, hispa, and leaf blast. Verma et al. [34] employed a lightweight CNN model, yielding a testing accuracy of 73.02% for all classes. Bhowmik et al. [35] implemented an ensemble model (VGG16 Light GBM), achieving a validation accuracy of 96.49% for all classes. In contrast, our proposed Modified LeafNet model achieved the highest validation accuracy of 97.44% among all algorithms. These findings underscore the efficacy of the Modified LeafNet and ensemble models in plant leaf disease classification.

In summary, our results highlight the superior performance of the Modified LeafNet model in classifying rice leaf diseases. Moreover, transfer learning emerges as a potent strategy for leveraging pretrained models to achieve accurate classification. Future research endeavors should explore additional deep learning models and image processing techniques to further enhance the accuracy of rice leaf disease classification.

In conclusion, our study successfully demonstrates the feasibility of classifying rice leaf diseases, including brown spot, hispa, and leaf blast, using various models trained on rice leaf images. While the Xception and MobileNetV2 models achieved testing accuracies of 71.95% and 67.68%, respectively, the LeafNet model emerged as a state-of-the-art solution for leaf classification. However, our investigation extended to both the LeafNet model and a modified version, revealing that the Modified LeafNet model surpassed all others in performance, achieving outstanding classification accuracies of 97.44% on the validation set and 87.76% on the testing set.

While achieving higher accuracy is commendable, future research endeavors should aim to enhance the dependability and robustness of the model across diverse datasets. This entails addressing challenges posed by complex surroundings and varying lighting conditions commonly encountered in real-world scenarios. Moreover,

there is a need to prioritize the development of interpretable CNN models that provide insights into disease classification in a comprehensible manner.

Acknowledging the limitations of our study, particularly concerning dataset size and variety, future work should focus on collecting more extensive and diverse datasets to improve the robustness and practical utility of the proposed models. Additionally, efforts to optimize model size and inference speed, through techniques like quantization and pruning, will be crucial for deployment on resource-constrained edge devices.

Furthermore, external validation using independent datasets is essential to validate the generalizability of our models. By subjecting them to various real-world scenarios, we can strengthen their credibility and demonstrate their efficacy beyond the confines of our specific dataset. Finally, exploring alternative interpretability techniques will enhance the transparency and interpretability of our deep learning models, further advancing their utility in practical applications.

6. References

- [1] M. Mahmuti, J. S. West, J. Watts, P. Gladders, and B. D. L. Fitt, "Controlling crop disease contributes to both food security and climate change mitigation," *Int. J. Agricult. Sustainability*, vol. 7, no. 3, pp. 189–202, Aug. 2009.
- [2] Y. Elad and I. Pertot, "Climate change impacts on plant pathogens and plant diseases," *J. Crop Improvement*, vol. 28, no. 1, pp. 99–139, Jan. 2014.
- [3] S. Chakraborty and A. C. Newton, "Climate change, plant diseases and food security: An overview," *Plant Pathol.*, vol. 60, no. 1, pp. 2–14, Feb. 2011.
- [4] M. D. Nirmal, P. P. Jadhav, and S. Pawar, "Pomegranate leaf disease detection using supervised and unsupervised algorithm techniques," *Cybern. Syst.*, pp. 1–12, Jan. 2023. [Online]. Available: <https://www.tandfonline.com/toc/ucbs20/0/0?startPage=1>, doi: 10.1080/01969722.2023.2166192.
- [5] World Agricultural Production. Accessed: Jul. 1, 2023. [Online]. Available: <https://apps.fas.usda.gov/psdonline/circulars/production.pdf>
- [6] M. A. A. Elfri, F. H. Rahman, S. H. S. Newaz, W. S. Suhaili, and T. W. Au, "Determining paddy crop health from aerial image using machine learning approach: A Brunei Darussalam based study," in *Proc. 8TH BRUNEI Int. Conf. Eng. Technol.*, 2023, pp. 040031-1–040031-8.
- [7] Rice Knowledge Bank. Accessed: Jul. 1, 2023. [Online]. Available: <http://www.knowledgebank.irri.org/training/fact-sheets/pest-management/diseases>
- [8] Rice Disease Management. Accessed: Jul. 1, 2023. [Online]. Available: <https://extension.missouri.edu/programs/rice-extension/rice-diseases/rice-disease-management>
- [9] W.-C. Chen, T.-Y. Chiou, A. L. Delgado, and C.-S. Liao, "The control of Rice blast disease by the novel biofungicide formulations," *Sustainability*, vol. 11, no. 12, p. 3449, Jun. 2019.
- [10] J. M. Al-Khayri, R. Rashmi, R. S. Ulhas, W. N. Sudheer, A. Banadka, P. Nagella, M. I. Aldaej, A. A.-S. Rezk, W. F. Shehata, and M. I. Almaghasla, "The role of nanoparticles in response of plants to abiotic stress at physiological, biochemical, and molecular levels," *Plants*, vol. 12, no. 2, p. 292, Jan. 2023.
- [11] R. R. Patil and S. Kumar, "Rice-fusion: A multimodality data fusion framework for Rice disease diagnosis," *IEEE Access*, vol. 10, pp. 5207–5222, 2022.
- [12] L. Wei, Y. Luo, L. Xu, Q. Zhang, Q. Cai, and M. Shen, "Deep convolutional neural network for Rice density prescription map at ripening stage using unmanned aerial vehicle-based remotely sensed images," *Remote Sens.*, vol. 14, no. 1, p. 46, Dec. 2021.
- [13] S. B. Jadhav, V. R. Udipi, and S. B. Patil, "Identification of plant diseases using convolutional neural networks," *Int. J. Inf. Technol.*, vol. 13, no. 6, pp. 2461–2470, Dec. 2021.