Classification of Rice Leaf Disease Using Convolutional Neural Network Models

Sokhey Kim[1]   **Hongly Va**[1]

[1]Cambodia Academy of Digital Technology, Phnom Penh, Cambodia

`Sokhey.kim@cadt.edu.kh, Hongly.va@cadt.edu.kh`

Abstract

Rice is a crucial crop globally, particularly in Asia, where it serves as a staple food. However, various diseases severely affect rice crop yields, and without proper detection, these diseases can spread, leading to a substantial decline in production. In extreme cases, diseases can result in a total crop loss, threatening food security. Deep learning, particularly Convolutional Neural Networks (CNNs), has become the standard method for image identification and classification tasks. Accurate diagnosis of rice diseases is essential to mitigate these impacts, yet current diagnostic methods are often inefficient, requiring specialized equipment. This study developed a deep learning–based automatic rice disease diagnosis method using an ensemble of CNN models. The method, built on deep learning, utilized a dataset of 17,500 images covering seven types of rice diseases, including bacterial blight, hispa, leaf blast, and others. The Ensemble Model, which combined several submodels, was the core of this method. Validation showed that EfficientNet-B0, DenseNet-121, and MobileNetV2 were the most effective submodels, achieving an overall accuracy of 96%. The Ensemble Model minimized confusion between disease types, reducing misdiagnosis and enhancing disease recognition accuracy, making it a reliable tool for rice disease detection.

Keywords: Rice leaf diseases; Deep learning classifier; CNNs; Ensemble learning;

1 Introduction

Rice is one of the most important staple crops worldwide, and its yield depends on multiple factors, including soil type, weather conditions, irrigation facilities, geography, seed selection, and biological threats [1]. Among these threats, plant diseases are a major cause of yield reduction and economic loss, particularly in Asia, where rice is a dominant food source [2]. Bacterial, fungal, and viral infections, including Bacterial Blight, Hispa, Leaf Blast, Sheath Blight, Tungro, and Brown Spot, frequently damage rice crops, leading to poor grain quality, lower production, and, in severe cases, complete crop failure. Protecting rice crops from these diseases is therefore critical for food security.

Traditional diagnosis of rice diseases relies on visual inspection, expert knowledge, and reference guides [4]. While these methods are helpful, they are time-consuming, labor-intensive, and prone to human error. They also often require trained specialists or specialized equipment that may not be accessible to farmers in rural areas. Consequently, automated rice disease detection methods have received increasing attention as a faster and more reliable alternative.

Early studies applied computer vision and machine learning techniques for crop disease detection, including handcrafted feature extraction and classifiers such as support vector machines (SVMs) and random forests [6]. Although these approaches achieved moderate success, they relied heavily on manual feature engineering, which limited their generalization to diverse field conditions.

More recently, deep learning has emerged as a powerful alternative for classifying plant diseases. CNNs in particular have demonstrated strong performance by automatically learning hierarchical features from raw images [7]. Models such as ResNet, DenseNet, and EfficientNet have achieved high accuracies across various crops. However, relying on a single CNN architecture may still lead to overfitting or misclassification, especially for visually similar diseases. To overcome these limitations, ensemble learning, which combines predictions from multiple models, has been proposed as an effective strategy to improve robustness and reduce errors [8].

In this study, we focus on the development

and evaluation of deep learning models for rice disease classification. A total of 63,889 images were initially collected from public sources, including Kaggle, Mendeley, and Roboflow, covering healthy leaves and ten rice diseases. To ensure balanced representation and avoid class bias, undersampling was applied, resulting in a final dataset of 17,500 images evenly distributed in seven categories (six major rice diseases and healthy). Three CNN architectures, including EfficientNet-B0, DenseNet-121, and MobileNetV2, were trained and evaluated, and their predictions were integrated through ensemble learning. The experimental results show that the ensemble model consistently outperformed individual CNNs, achieving an overall accuracy of 96%.

The main contributions of this paper are as follows.

- We trained a deep learning network model for diagnosing seven different types of rice diseases, including healthy samples.

- We evaluated three state-of-the-art CNN architectures, such as EfficientNet-B0, DenseNet-121, MobileNetV2, for rice disease detection.

- We designed an ensemble learning approach that integrates multiple CNN models to improve robustness and demonstrated a high accuracy rate of 96%, which is considered a good result.

The remainder of this paper is organized as follows. In the next section, we review related work that is pertinent to our own. Section III discusses the methodology of a model. Section IV describes the experiment and the analysis of detecting rice leaf disease. Finally, the discussion and conclusion are summarized at the end of this paper.

2 Related Work

To better position our approach, we review related research in rice disease detection and ensemble learning. Prior works can be broadly grouped into three categories: (1) traditional computer vision and machine learning techniques, (2) deep learning based approaches for plant disease detection, and (3) ensemble learning strategies.

Early work in crop disease detection relied on handcrafted feature extraction combined with classical classifiers. Techniques such as color analysis, texture descriptors, and shape features were commonly applied, followed by classifiers like SVM, k-nearest neighbors (KNN), and random forests [4][6]. While these methods achieved moderate success, they required domain expertise for feature design and performed poorly in complex field conditions where lighting and backgrounds varied significantly.

The emergence of CNNs revolutionized image-based plant disease detection by eliminating the need for manual feature engineering. CNNs have been widely used to classify diseases in crops such as rice, maize, wheat, and tomato. For example, Deng et al. [7] proposed an in-field rice disease detection system using deep CNNs, demonstrating strong performance under real-world conditions. Other studies have evaluated architectures such as ResNet, Inception, DenseNet, and EfficientNet for leaf disease classification, achieving higher accuracy compared to traditional machine learning methods [9]. Despite these improvements, individual CNNs often struggle to distinguish visually similar diseases, resulting in misclassification.

Ensemble learning has been increasingly adopted to address the limitations of single CNN models. By combining predictions from multiple classifiers, ensemble approaches enhance robustness, mitigate overfitting, and yield more reliable results. For instance, Deng et al. [7] employed ensemble CNNs to classify rice diseases, achieving higher accuracy than individual networks. Similarly, Feng et al. [8] demonstrated that ensemble methods improved yield prediction in alfalfa using hyperspectral imagery. In plant disease detection, ensemble strategies such as majority voting, soft probability averaging, and stacking have consistently shown advantages over single models.

While these approaches are promising, several challenges remain. Many datasets are imbalanced, with limited samples for certain diseases, which reduces generalization ability. Furthermore, most prior studies emphasize the performance of single models without fully exploiting ensemble methods to address inter-class similarity and misclassification among visually similar rice diseases. These limitations motivate our work on developing a balanced

dataset and integrating multiple CNN architectures, such as EfficientNet-B0, DenseNet-121, and MobileNetV2 to an ensemble model to enhance classification accuracy and robustness.

3 Methodology

3.1 Data acquisition

Deep learning requires a large number of training images to achieve good results [9]. In this study, a total of 63,889 rice leaf images were collected from public sources, including Kaggle, Mendeley, and Roboflow. The datasets included healthy leaves and ten rice diseases: Bacterial Blight, Brown Spot, Rice Hispa, Leaf Blast, Sheath Blight, Tungro, Leaf Scald, Narrow Brown Spot, Neck Blast, and Leaf Smut [10][11]. Table 1 summarizes eleven common rice diseases, their causal agents, symptoms, and impacts on the plant, while Fig. 1 shows the distribution of the initially collected samples. To ensure a well-rounded dataset, we incorporated two types of images. **White Background Images,** these images were taken in controlled settings with consistent lighting. This makes it easier to see the rice plants and their symptoms. The uniform background helps in training machine learning models by reducing distractions. **Field Background Images,** were captured in real agricultural environments. They show the rice crops in their natural conditions, with various backgrounds and lighting. This variety helps the model to learn how diseases appear in real life, making it more effective for practical use. However, to balance the dataset of each class, we apply undersampling, a technique that reduces the total number of images of some classes, resulting in a balanced dataset of 2,500 images with equal representation of healthy crops and each disease [12]. Each category, including healthy crops and six rice diseases, contains 2,500 images, with 1,700 field background images and 800 white background images per class. Therefore, the final dataset used for model training and evaluation comprised 17,500 images across seven categories (six rice diseases and one healthy category), with equal class representation.

3.2 Data preprocessing

Image Processing is performed to reduce overfitting of deep learning models during training and validation. Before the model reads the images, the pixels in the image are normalized. Then, we applied random affine transformations, including rotations, shifts, and shearing, to introduce the models to different perspectives. Brightness adjustments, Gaussian blur, and flipping were also applied to help the models learn from various conditions. Additionally, the images were resized to 224×224 as required by the architectures of selected deep learning models [13]. Figure 2 shows the processing of data preprocessing step by step of the models.

3.3 Model development

3.3.1 Convolutional neural network Models

The structure of the CNN has a crucial influence on the performance of the final model. To evaluate the effectiveness of different architectures, we selected three state-of-the-art CNN models, including EfficientNet-B0 [14], DenseNet-121 [15], and MobileNetV2 [16]. These models were chosen due to their balance between accuracy and computational efficiency, making them suitable candidates for agricultural applications, including potential deployment on mobile devices.

3.3.2 Ensemble learning

Ensemble learning combines multiple base models to enhance robustness and mitigate the risk of misclassification by leveraging complementary decision boundaries. In this study, the predictions of EfficientNet-B0, DenseNet-121, and MobileNetV2 were combined using a voting-based ensemble strategy [17]. Specifically, we implemented soft voting, in which the predicted class probabilities from each model were averaged, and the class with the highest probability was selected as the final output. This approach reduces bias from individual models and provides more stable performance compared to single CNNs.

3.3.3 Fine-tuning of the models

Fine-tuning was applied to optimize the CNN models for the rice disease dataset. Transfer learning was used to initialize each model with pre-trained ImageNet weights, enabling the knowledge learned from large-scale image classification to be adapted to the agricultural domain [18]. We then fine-tuned all layers of EfficientNet-B0, DenseNet-121, and MobileNetV2 on our dataset. Each model was
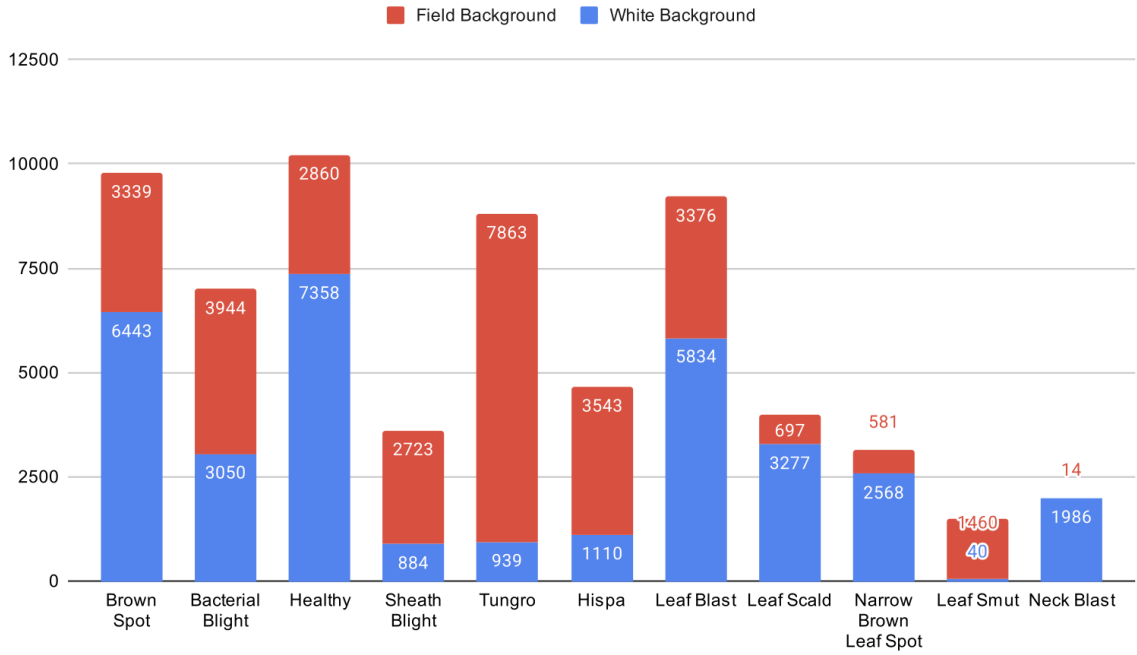
Figure 1: Overview of all the datasets before cleaning. The data consists of 11 types of rice disease, including healthy samples.

Table 1: Overview of rice viral diseases, including their causes, symptoms, and impact.

Disease Name	Cause	Description	Affected Plant Part
Healthy	None	Normal green leaves with no disease symptoms	Whole Plant
Bacterial Blight	Bacterial	Yellowish-brown streaks on leaves, leading to wilting and drying	Leaf
Neck Blast	Fungal	Collapse and decay of the neck area, affecting grain formation	Panicle
Leaf Scald	Fungal	Browning and wilting of leaf tips; reduced vigor	Leaf
Narrow Brown Leaf Spot	Fungal	Narrow, elongated brown lesions on leaves; leads to leaf death	Leaf
Leaf Smut	Fungal	Dark, smutty spores on leaves; affects photosynthesis	Leaf
Leaf Blast	Fungal	Gray-centered lesions with brown borders, affecting plant growth	Leaf
Sheath Blight	Fungal	Oval, water-soaked lesions on leaf sheaths, leading to plant lodging	Leaf Sheath, Leaf
Brown Spot	Fungal	Small brown spots with yellow halos, reducing grain quality	Leaf, Grain
Hispa	Parasitic	Small brown spots with yellow halos, reducing grain quality	Leaves
Tungro	Virus	Stunted growth and yellow-orange discoloration	Leaves, Whole Plant

trained for 20 epochs with early stopping criteria to prevent overfitting. This fine-tuning process enabled the networks to adapt to the specific visual features of rice diseases while reducing training time and computational cost.

4   Experimental results and analysis

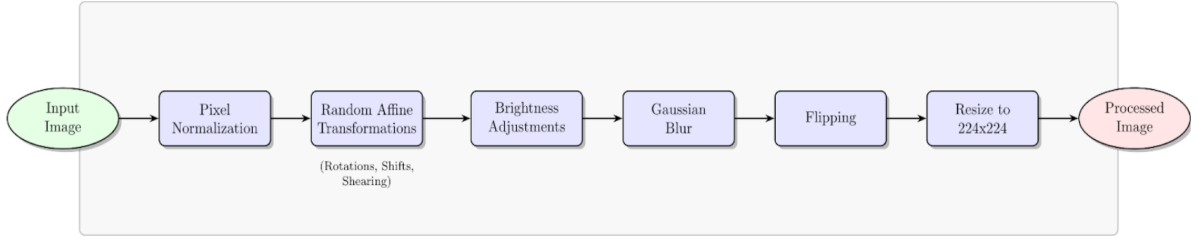This section mainly focuses on the empirical study of the proposed method.

Figure 2: Image preprocessing pipeline used in the models.

4.1 Evaluation metrics

To evaluate the performance of the proposed method, we used four standard classification metrics, such as Accuracy, Precision, Recall (Sensitivity), and F1-score.

Accuracy measures the overall proportion of correctly classified samples and is defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively.

Precision reflects the classification accuracy of the classifier, indicating all samples predicted to be positive examples. The proportion of correctly predicted samples.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Recall reflects the recall ability of the classifier, indicating that all positive samples are correctly predicted in the sample proportion tested.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

F1-score is the harmonic mean of Precision and Recall, providing a balanced measure of both metrics:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

4.2 Experimental dataset

The experiments were conducted using the proposed model applied to the rice disease dataset. The dataset was divided into training, validation, and testing subsets with a ratio of 7:2:1. This split was applied across all seven categories to ensure balanced representation. As a result, 70% of the samples (12,250 images) were used for training, 20% (3,500 images) for validation, and the remaining 10% (1,750 images) for testing. This stratified partitioning provided sufficient samples in each category for effective training and evaluation.

4.3 Results and analysis

All experiments were implemented in Python (3.10.11) using TensorFlow GPU 2.10 with CUDA 12.8.0 acceleration. Training was conducted on a system equipped with an AMD Ryzen 7 6800H CPU, an NVIDIA RTX 3060 GPU (6 GB), and 16 GB of RAM. The models were trained using the Adam optimizer with a learning rate of 0.0001 and a batch size of 32 for a total of 30 epochs. Specifically, 10 epochs were used for initial training followed by 20 epochs of fine-tuning. Sparse categorical cross-entropy loss was employed to optimize multi-class classification across the seven rice disease categories. The detailed software and simulation parameters are summarized in Table 2.

4.3.1 Models performance

As summarized in Table 3, the ensemble model outperformed all individual CNNs, achieving the highest accuracy, precision, recall, and F1-score (0.96). Table 4 further shows that this advantage extended across nearly all disease classes, with particularly strong results for Sheath Blight and Bacterial Blight (F1-scores of 0.98), while reducing errors in more challenging classes such as Brown Spot and Healthy.

- **EfficientNet-B0** achieved the highest single-model validation accuracy (0.95). Its confusion matrix (Figure 4a) shows high true positives for Bacterial Blight (240), Healthy (237), and Sheath Blight (245). Misclassifications included eight Brown Spot instances labeled as Leaf Blast and ten Healthy samples classified as Hispa. The classification heatmap (Figure 3a) reports

Table 2: Simulation software and parameters.

Software/Parameters	Value
Operating System	Windows 11, 64-bit
Programming Language	Python version = 3.10.11, TensorFlow-GPU = 2.10
CPU	Ryzen 7 6800H
GPU	NVIDIA RTX 3060 (6GB)
RAM	16 GB
Batch Size	32
Optimizer	Adam
Learning Rate	0.0001
Epochs	30 (10 initial trainings + 20 fine-tuning)

Table 3: Evaluation of the ensemble model's overall performance on the test set.

Model	Accuracy	Precision	Recall	F1-score
EfficientNet-B0	0.95	0.95	0.95	0.95
DenseNet-121	0.93	0.92	0.92	0.92
MobileNetV2	0.90	0.87	0.88	0.87
Ensemble Model	**0.96**	**0.96**	**0.96**	**0.96**

precision, recall, and F1-scores between 0.90 and 0.98.

- **DenseNet-121** obtained the second-highest accuracy (0.93). As shown in Figure 4b, it correctly identified most Sheath Blight (244) and Healthy (231) samples, but misclassified 16 Brown Spot images as Bacterial Blight and eight as Leaf Blast. The heatmap (Figure 3b) confirms strong results for Sheath Blight and Tungro (F1 = 0.96), while Brown Spot lagged (F1 = 0.90).

- **MobileNetV2** achieved an accuracy of 0.90. Its confusion matrix (Figure 4c) shows correct predictions for Sheath Blight (239) but notable errors, such as 38 Healthy samples labeled as Hispa and 22 Leaf Blast samples as Brown Spot. The heatmap (Figure 3c) shows an F1 score of 0.87, with strong results for Sheath Blight (0.95) and Tungro (0.96), but lower performance on Healthy (0.81) and Hispa (0.84). These findings illustrate MobileNetV2's trade-off between computational efficiency and predictive accuracy, relevant for mobile deployment scenarios.

- **Ensemble learning model** combined EfficientNet-B0, DenseNet-121, and MobileNetV2 using soft voting. As shown in Figure 4d, it achieved balanced results with minimal misclassifications (e.g., one sample of bacteria blight was classified as brown spot, and one healthy sample

was classified as hispa). The heatmap (Figure 3d) indicates consistently high precision (0.91–0.99) and recall (0.90–0.99), leading to per-class F1 scores between 0.90 (Brown Spot) and 0.98 (Sheath Blight). These results confirm that the ensemble delivers the most robust and stable performance in all categories.

### 4.3.2 Training and validation visualization

Figure 5 shows the comparison of the accuracy of all three models, including EfficientNet-B0, DenseNet-121, MobileNetV2 of training and validation loss.

During training, as shown in Figure 5(a) and Figure 5(b), EfficientNet-B0 consistently achieved the strongest results among the evaluated models. It reached approximately 95% training accuracy by the final epochs and maintained the lowest training loss, indicating its effectiveness in capturing complex patterns within the dataset without overfitting. DenseNet-121 also demonstrated robust learning behavior, showing steady improvements in accuracy alongside a decreasing loss curve. Although its loss values were slightly higher than those of EfficientNet-B0, the dense connectivity of its layers supported effective feature reuse and efficient learning of challenging patterns. In contrast, MobileNetV2 exhibited a slower increase in training accuracy and a higher loss, especially during the initial epochs.
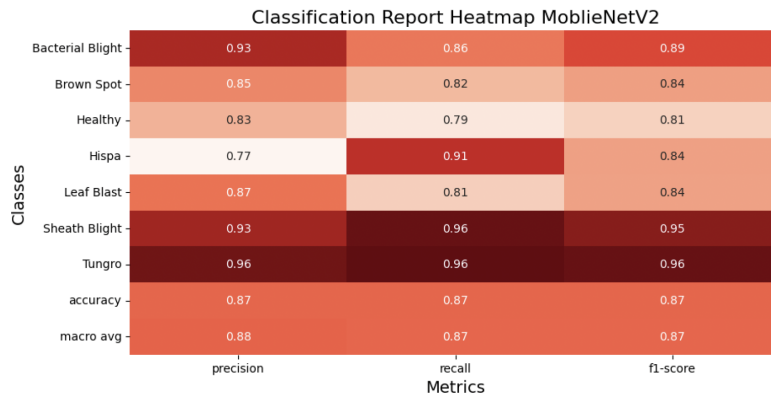
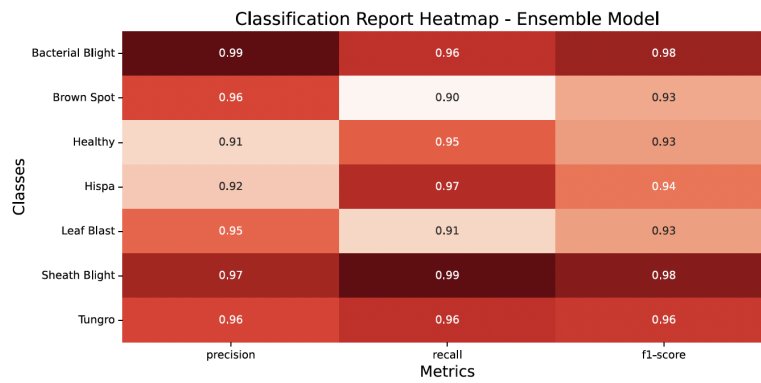Figure 5(c) and Figure 5(d) present the validation accuracy and loss for each model, illus-
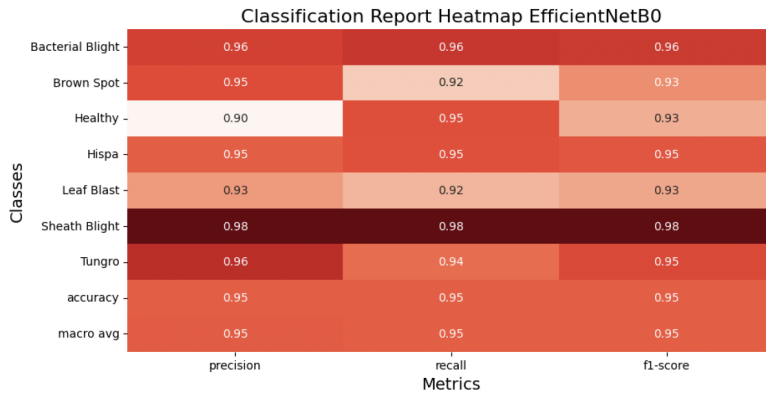
(a) EfficientNet-B0
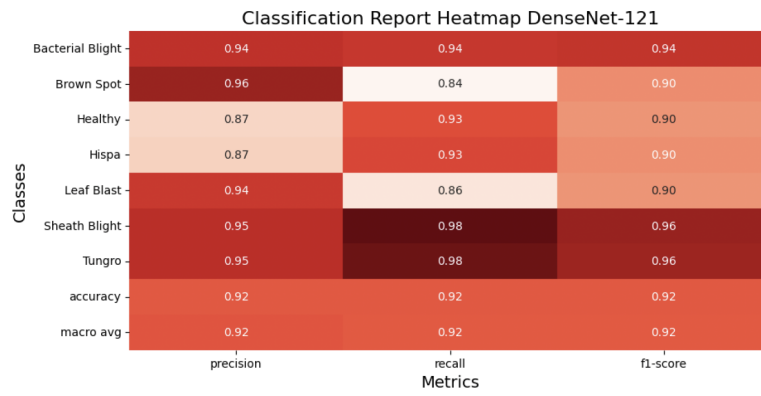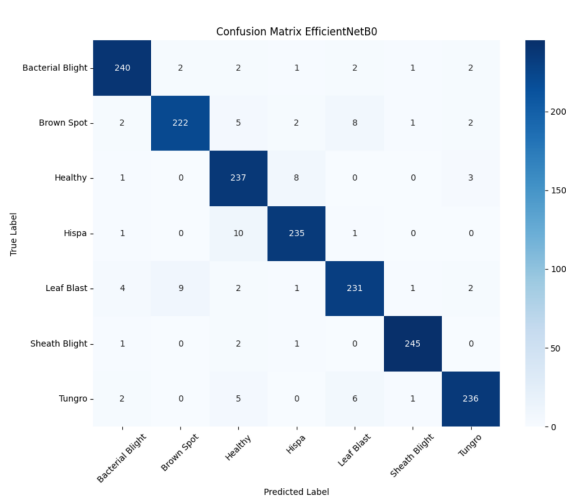

(b) DenseNet-121

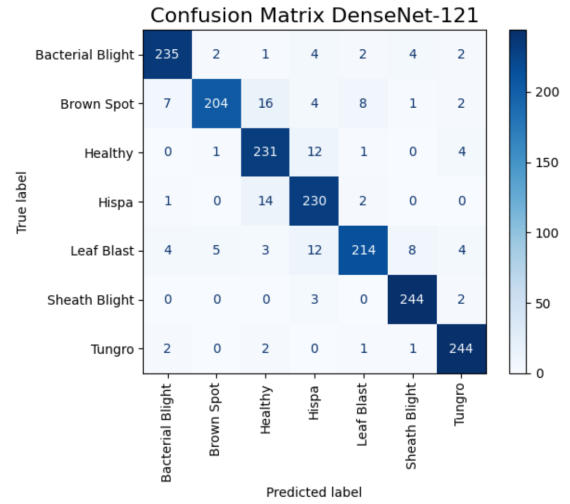

(c) MobileNetV2

(d) Ensemble Model

Figure 3: Heatmaps of the classification reports for individual models and the ensemble model.

Table 4: F1-scores of each model across all disease classes, with the ensemble model demonstrating the most balanced performance.
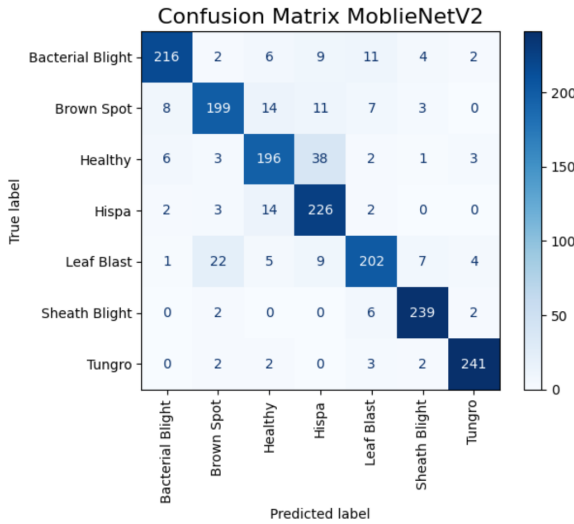
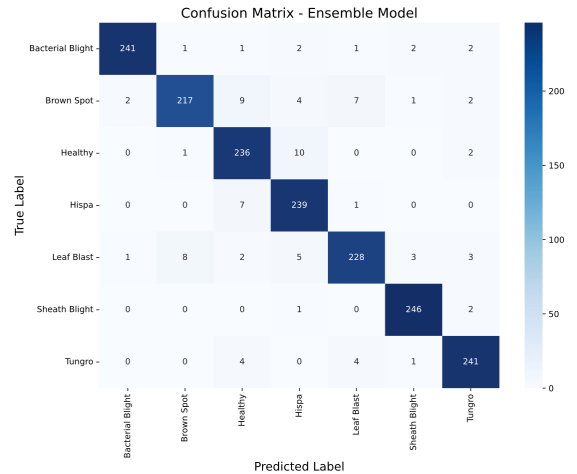| Disease	EfficientNet-B0	DenseNet-121	MobileNetV2	Ensemble Model
Bacterial Blight	0.96	0.94	0.88	**0.98**
Brown Spot	0.93	0.90	0.85	**0.93**
Healthy	0.94	0.91	0.81	**0.93**
Hispa	0.95	0.92	0.84	**0.94**
Leaf Blast	0.94	0.91	0.86	**0.93**
Sheath Blight	0.98	0.96	0.95	**0.98**
Tungro	0.96	0.96	0.96	**0.96**

(a) Confusion matrix of EfficientNet-B0

(b) Confusion matrix of DenseNet-121

(c) Confusion matrix of MobileNetV2

(d) Confusion matrix of Ensemble Model

Figure 4: Confusion matrices of the individual models and the ensemble model on the test set: (a) EfficientNet-B0, (b) DenseNet-121, (c) MobileNetV2, and (d) Ensemble Model. The ensemble demonstrates fewer misclassifications.

trating their ability to generalize to unseen data. EfficientNet-B0 again achieved the strongest results, maintaining the highest validation accuracy and the lowest validation loss. This consis-

tency across both training and validation phases confirms its robustness and ability to capture complex patterns without overfitting. DenseNet-121 displayed a similar trend, attaining com-

mendable validation accuracy and a gradual decrease in validation loss. These results demonstrate its capacity to generalize effectively to new data while maintaining stable learning dynamics. In contrast, MobileNetV2 performed well but lagged behind the other models in validation accuracy and showed a higher validation loss, mirroring its training behavior. This outcome reinforces that MobileNetV2's design prioritizes computational efficiency over representational depth, making it suitable for lightweight deployment but less optimal for highly complex classification tasks without further fine-tuning or architectural adjustments.

Overall, the training and validation curves highlight the strengths and weaknesses of individual CNNs, confirming EfficientNet-B0 as the most reliable base model, DenseNet-121 as a stable alternative, and MobileNetV2 as a lightweight but less expressive option, further justifying the use of an ensemble strategy to achieve balanced and robust performance.

5 Discussion

The experimental results demonstrate that CNNs are effective for rice disease detection, with EfficientNet-B0 and DenseNet-121 showing strong performance, and the ensemble model further improving overall robustness. In this section, we will discuss key observations, practical implications, and limitations of our work.

EfficientNet-B0 consistently achieved the highest accuracy among the single models (95%), reflecting its efficient scaling and ability to capture complex visual features even in relatively lightweight configurations. DenseNet-121 followed closely (92%), leveraging dense connections to promote feature reuse, which helped in capturing subtle lesion characteristics. MobileNetV2, while computationally efficient and suitable for mobile deployment, performed lower (87%) due to its reduced representational capacity. This highlights the trade-off between computational efficiency and classification accuracy, which is critical when selecting models for field deployment.

The ensemble model achieved the best overall accuracy of 96%, outperforming all individual CNNs. Its strength lies in combining the complementary decision boundaries of different models, which reduces the misclassification of visually similar diseases, such as Brown Spot vs.

Leaf Blast and Healthy vs. Hispa. These confusions were evident in the confusion matrices of the individual models, but significantly reduced in the ensemble predictions.

Despite promising results, some limitations remain. First, our dataset was restricted to seven categories, while rice crops are affected by a broader range of diseases. Extending the dataset to include additional disease types would improve generalizability. Second, the dataset balancing was achieved through undersampling, which reduced diversity in some classes. It is important to note that this balancing process did not involve artificial data generation or oversampling. As described in Section 3.1 (Data Acquisition), we applied undersampling by reducing the number of images in overrepresented classes to achieve an equal distribution of 2,500 samples per class, consisting of approximately 1,700 field background and 800 white background images. This approach ensured that each category contributed equally to training without introducing synthetic samples. However, undersampling inherently reduces data diversity, which may affect model generalization. Future work will explore advanced data augmentation and class-weighted learning strategies to preserve diversity while maintaining balance. Additionally, further verification revealed signs of model overfitting, likely caused by the limited dataset size and the use of a single dataset split (70/20/10) without cross-validation or an external test set. Although ensemble learning was adopted to improve robustness and reduce the risk of overfitting, its effectiveness is still constrained when data diversity is limited. While several techniques such as data augmentation, dropout, and early stopping were applied to reduce overfitting, these measures were insufficient to fully prevent it.

Future work will incorporate k-fold cross-validation, larger and more diverse datasets, and external validation to ensure better model generalization and reliability. Third, while the ensemble achieved strong results, its inference speed may be slower compared to lightweight single models, requiring further optimization for real-time deployment. Moreover, exploring optimization-based ensemble strategies, such as the Ensemble Genetic Algorithm and Convolutional Neural Network (EGACNN) [19], may provide an effective way to automatically optimize model combinations and improve general-
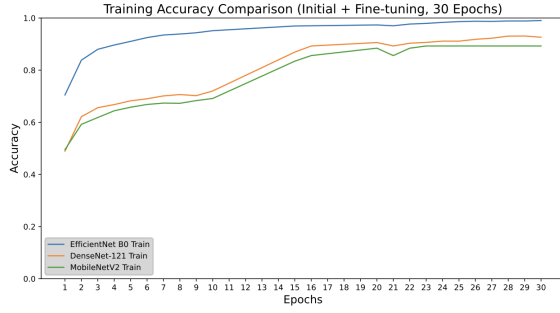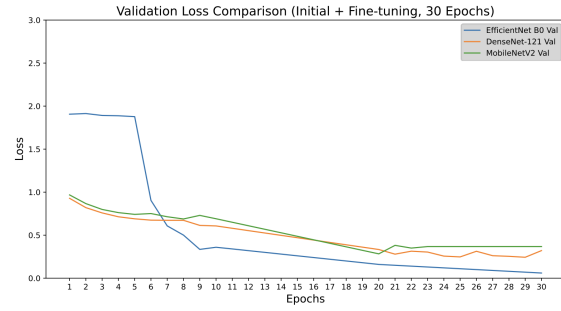
(a) Training accuracy curve.
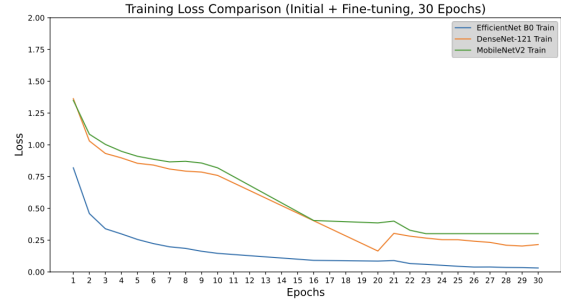
(b) Training loss curve.
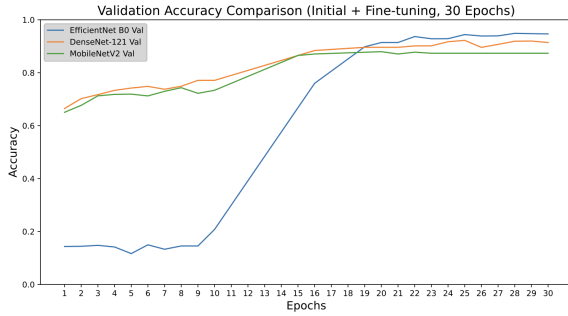
(c) Validation accuracy curve.

(d) Validation loss curve.

Figure 5: Comparison of the three models, EfficientNet-B0, DenseNet-121, and MobileNetV2 using training and validation metrics: (a) Training accuracy, (b) Training loss, (c) Validation accuracy, and (d) Validation loss.

ization. This represents a promising direction for integrating metaheuristic optimization with deep learning in agricultural image analysis.

6 Conclusion

This paper presented a deep learning framework for the detection of rice diseases using convolutional neural networks. By integrating EfficientNet-B0, DenseNet-121, and MobileNetV2, the ensemble model achieved superior performance, with an overall accuracy and F1-score of 96%. Compared to individual CNNs, the ensemble approach reduced misclassifications between visually similar diseases, demonstrating greater robustness and reliability for field use.

Beyond accuracy, this work emphasizes the development and evaluation of an effective deep learning–based ensemble model for rice disease classification. The findings highlight the potential of combining multiple CNN architectures to enhance performance and provide a foundation for future research in precision agriculture.

However, some limitations remain. The dataset used in this study was limited to seven categories and balanced through undersampling,

which reduced sample diversity. Future research should expand to a broader range of rice diseases, adopt advanced augmentation or generative techniques for data balancing, and explore optimization-based ensemble methods—such as the Ensemble Genetic Algorithm and Convolutional Neural Network (EGACNN)—to further improve model generalization and robustness.

In conclusion, the proposed ensemble approach represents a promising step toward practical, AI-driven solutions for precision agriculture. With further development and refinement, such models have the potential to support more accurate and reliable plant disease management, ultimately contributing to sustainable agricultural practices and global food security.

agement made this work possible.

References

[1] J.-J. Dethier and A. Effenberger, "Agriculture and development: A brief review of the literature," *Economic Systems*, vol. 36, no. 2, pp. 175–205, 2012.

[2] S. Savary, A. Ficke, J.-N. Aubertot, and C. Hollier, "Crop losses due to diseases and their implications for global food production losses and food security," *Food Security*, vol. 4, no. 4, pp. 519–537, 2012.

[3] P. A. Seck, A. Diagne, S. Mohanty, and M. C. S. Wopereis, "Crops that feed the world 7: Rice," *Food Security*, vol. 4, no. 1, pp. 7–24, 2012.

[4] P. K. Sethy, N. K. Barpanda, A. K. Rath, and S. K. Behera, "Image processing techniques for diagnosing rice plant disease: a survey," *Procedia Computer Science*, vol. 167, pp. 516–530, 2020.

[5] R. Deng *et al.*, "Deep learning-based automatic detection of productive tillers in rice," *Computers and Electronics in Agriculture*, vol. 177, p. 105703, 2020.

[6] L. C. Ngugi, M. Abelwahab, and M. Abo-Zahhad, "Recent advances in image processing techniques for automated leaf pest and disease recognition–A review," *Information Processing in Agriculture*, vol. 8, no. 1, pp. 27–51, 2021.

[7] R. Deng *et al.*, "Automatic diagnosis of rice diseases using deep learning," *Frontiers in Plant Science*, vol. 12, p. 701038, 2021.

[8] L. Feng *et al.*, "Alfalfa yield prediction using UAV-based hyperspectral imagery and ensemble learning," *Remote Sensing*, vol. 12, no. 12, p. 2028, 2020.

[9] J. G. A. Barbedo, "Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification," *Computers and Electronics in Agriculture*, vol. 153, pp. 46–53, 2018.

[10] Loki4514, "Rice Leaf Diseases Detection," *Kaggle*, 2021. [Online]. Available: https://www.kaggle.com/datasets/loki4514/rice-leaf-diseases-detection

[11] Mtech MIT, "Paddy leaf disease detection Dataset," *Roboflow Universe*, Jul. 2023. [Online]. Available: https://universe.roboflow.com/mtech-mit/paddy-leaf-disease-detection

[12] R. Mohammed, J. Rawashdeh, and M. Abdullah, "Machine learning with oversampling and undersampling techniques: overview study and experimental results," in *Proc. 11th Int. Conf. on Information and Communication Systems (ICICS)*, 2020, pp. 243–248.

[13] R. Szeliski, *Computer Vision: Algorithms and Applications*, Springer Nature, 2022.

[14] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. on Machine Learning (ICML)*, 2019, pp. 6105–6114.

[15] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 4700–4708, 2017.

[16] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 4510–4520.

[17] R. Caruana, A. Niculescu-Mizil, G. Crew, and A. Ksikes, "Ensemble selection from libraries of models," in *Proc. 21st Int. Conf. on Machine Learning (ICML)*, 2004, p. 18.

[18] J. Chen, J. Chen, D. Zhang, Y. Sun, and Y. A. Nanehkaran, "Using deep transfer learning for image-based plant disease identification," *Computers and Electronics in Agriculture*, vol. 173, p. 105393, 2020.

[19] Hussain, Wajahat and Mushtaq, Muhammad Faheem and Shahroz, Mobeen and Akram, Urooj and Ghith, Ehab Seif and Tlija, Mehdi and Kim, Tai-hoon and Ashraf, Imran, "Ensemble genetic and CNN model-based image classification by enhancing hyperparameter tuning," *Nature Publishing Group UK London*, vol. 15, p. 1003, 2025.