Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.3, June 2026, pp. 2415~2423
e-ISSN: 2548-6861

Deep Learning-Based Detection and Classification of Rice Leaf Diseases Using ResNet-50 with Augmentation and K-Fold Cross-Validation

Zaqi Kurniawan [1]*, Rizka Tiaharyadini [2]*, M Saddam Ryuga Octoramdhani [3]*, Radiz Dirgantara [4]*
* Faculty of Technology and Information, Universitas Budi Luhur, Jakarta, Indonesia
zaqi.kurniawan@budiluhur.ac.id [1], rizka.tiaharyadini@budiluhur.ac.id [2], 2411501568@student.budiluhur.ac.id [3], 2411500107@student.budiluhur.ac.id [4]

Article Info

Article history:

Received 2026-03-31
Revised 2026-05-04
Accepted 2026-05-25

Keyword:

Augmentation,
Convolutional Neural Network,
Deep Learning,
K-Fold Cross Validation,
ResNet-50.

ABSTRACT

Rice sustains over half of the global population; however, leaf diseases such as blight, brown spot, and leaf smut significantly reduce crop yields. Early and accurate detection is essential for supporting sustainable agriculture practices. This study proposes a ResNet-50-based deep learning model for rice leaf diseases classification using a dataset of 1,150 field images collected from Sleman, Indonesia, which was expanded to 2,000 images through augmentation techniques, including rotation, flipping, zooming, and brightness adjustment. Model performance was evaluated using both hold-out validation and 10-fold cross-validation with accuracy, precision, recall, and F1-Score metrics. The application of data augmentation improved hold-out validation accuracy from 82.4% to 88.1%. Meanwhile, 10-fold-cross-validation yielded a substantially higher average accuracy of 99.6%. This discrepancy suggests potential sensitivity to data partitioning and indicates the need for careful interpretation, as cross-validation may introduce optimistic estimates under certain conditions. Although the proposed approach demonstrates strong performance in distinguishing visually similar diseases, this study limited by the use of a single-region dataset, which may affect generalizability. Therefore, the integration of ResNet-50, augmentation, and cross-validation shows promising results for early disease detection, while further validation on more diverse datasets is required to support it application in real-world precision agriculture systems.

I. INTRODUCTION

Over half of the world's population depends on rice as a staple food source, but rice diseases can cause significant yield losses. In affected areas, major leaf diseases like blast, brown spot, and bacterial leaf streak can reduce productivity by as much as 30-40% if they are not identified early [1], [2]. In the larger context of intelligent and sustainable agriculture, fast and precise detection of rice leaf disease is essential for maintaining food security, cutting down on excessive pesticide use, and assisting farmers in making decisions. The use of computer visions and deep learning algorithms for automated plant disease diagnosis is being driven by the expensive, subjective, and scale-limiting nature of traditional manual inspection approaches, especially in large paddy fields [3], [4].

Convolutional neural networks (CNNs) are receiving a lot of attention lately, particularly deep designs like ResNet-50, because of their robustness and powerful feature extraction capabilities when paired with data augmentation and transfer learning techniques [5], [6]. While lightweight architectures such as MobileNet or EfficientNet are often prioritized for their computational efficiency on mobile devices, this study specifically use ResNet-50 architecture to leverage it deep residual learning capabilities. Furthermore, the growing availability of datasets of agricultural images and the increased focus on model generalization and reliability-addressed by methods such as K-Fold cross validation-align this area of study with the current challenges in implementing reliable AI solutions for actual precision agriculture systems [7], [8].

In this study research, rice leaf diseases are identified and classified using a ResNet-50 deep learning model optimized using data augmentation alongside K-Fold CV. According to recent research, convolutional neural networks (CNNs), including ResNet-50 variants, outperform conventional machine learning and manually created feature approaches in identifying rice leaf diseases like blast, brown spot, and bacterial blight from RGB images [9], [10], [11]. However, the research now in publication identifies a number of unresolved problems, such as model overfitting brought on by undersized and unbalanced datasets, insufficient generalization across different field situations, and imbalanced evaluation techniques that mostly rely on single train-test splits [12], [13]. The primary performance measures in this study are classification accuracy, precision, recall, and F1-score. Other important factors include input leaf image data, disease class labels, network design parameters, data augmentation procedures, and validation schemes. The study is part of the category of agricultural technology and focuses on the production of rice in tropical and subtropical areas, where the crop is a mainstay and yield and food security are seriously threatened by leaf diseases [14], [15]. Recent studies on deep learning for rice disease diagnosis reveal limitations in robustness and reliability, which this study approaches this problem by leveraging systematic data augmentation and K-Fold CV within ResNet-50.

Several significant issues have not been addressed in previous studies, despite the deep learning models' encouraging performance in detecting rice leaf diseases. According to theory, a large number of current research use shallow or moderately deep CNN architectures, which may not be able to capture intricate inter-class changes brought on by identical visual symptoms across various rice diseases [16], [17]. This could result in overfitting and a decreased capacity for generalization. Though many studies use inadequate data augmentation techniques or test models using single hold-out split, which can lead to optimistic and unstable performance estimates, actually the majority of publicly available rice leaf datasets are small and show class imbalance [18], [19]. Additionally, although residual learning has been shown to enhance convergence and feature classification in agricultural image analysis, barely any of recent research have methodically investigated deep residual networks, such as ResNet-50, in combination with K-Fold cross validation [20]. These disparities have significant societal and agricultural implications since faulty disease detection technologies can limit usage by farmers and agricultural organizations, hinder early intervention, and raise output losses. Thus, it is crucial to overcome these constraints by combining a strong ResNet-50 architecture, extensive data augmentation, and reliable K-Fold cross validation in order to create more precise, dependable, and easily implemented rice leaf disease detection systems that will encourage sustainable rice production and food security. The following research questions are intended to be addressed in order to fill up these gaps: (1) In comparison to other CNN-based methods, how effective is ResNet-50 at identifying and categorizing various rice leaf diseases? (2) How much can the combination of a deep residual network and data augmentation enhance classification accuracy and generalization? and (3) How can a more robust and dependable performance evaluation of rice leaf disease classification models benefit from K-Fold cross validation? This study aims to give a more methodical and repeatable evaluation of deep residual learning for rice leaf disease detection by addressing these problems.

While ResNet-50 has been previously explored for rice leaf disease classification, this study differentiates by addressing critical gaps in model reliability and evaluation consistency that remain prevalent in existing literature. The primary contribution of this work lies in the systematic integration of deep residual learning with extensive data augmentation and 10-folds cross-validation to rigorously assess performance stability and minimize partition bias. Unlike many studies that rely on limited public datasets or single train-test splits, this research utilizes a unique, real-field dataset collected from Sleman, Indonesia, which captures authentic environmental variability, including complex backgrounds and natural lighting fluctuations. Furthermore, by providing a comparative analysis between hold-out and cross-validation schemes, this study offers a methodological reference for developing more robust and context-aware deep learning application in practical precision agriculture.

II. Method

A. Research Stages

The overall workflow of the proposed deep learning-based approach for detecting and classifying rice leaf is illustrated in Figure 1. The study follows a methodological process that starts with data collection, where raw image data is obtained as the main source of information. The next step is data augmentation, which uses methods like rotation, padding, and horizontal flipping to improve model robustness against overfitting and boost data variety. Data splitting is then used to arrange the supplemented dataset into training, validation, and test sets in order to guarantee objective model training, hyperparameter tuning, and performance evaluation. The model training phase then uses a deep convolutional neural network, ResNet-50 in particular, to create selective output features and learn hierarchical feature representations form the input images.

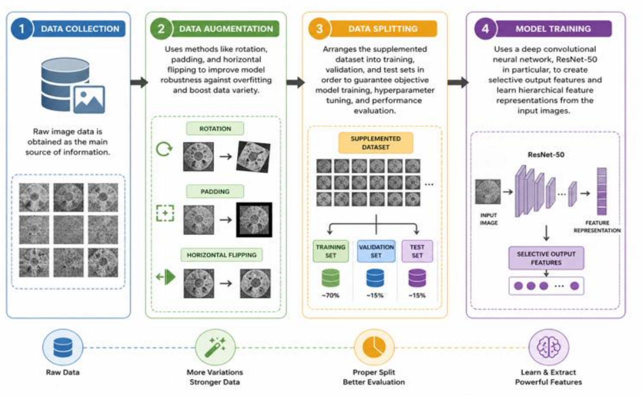

Figure 1. Research Flow Diagram

B. Data Collection and Augmentation

The dataset used in this study was collected direct field observations in rice fields areas located in Sleman, Yogyakarta, using camera with 12 megapixels resolution. To capture realistic environmental variations, image acquisition was conducted during daytime (from morning to afternoon), encompassing fluctuations in natural lighting intensity, various leaf orientations, and complex backgrounds such as soil, water, and vegetation. Image acquisition was performed from multiple angle to capture variations in leaf orientation and differing levels disease severity. The initial dataset, consisting of 1,150 images, is detailed in Table 1 and 6, comprising four categories: Healthy Leaf (320 images; 27.83%), Brown Spot (300 images; 26.09%), Blight (280 images; 24.35%), and Leaf Smut (250 images; 21.74%). All data were manually annotated and subsequently validated by agricultural practitioners to ensure label reliability prior to augmentation and classification stages.

Figure 2. Sample Images of Rice Leaf Disease Categories

The data augmentation process was conducted using specific technical parameters to enhance dataset variability while preserving the distinctive morphological characteristic of the disease. The parameters employed include random rotation within a range of ± 20 degrees, scaling (zooming) with factor between 0.8 and 1.2 of the original size, horizontal image flipping, and brightness adjustment with a variations range ± 30%. All techniques were applied exclusively to the training subset following data partitioning, thereby preventing any potential data leakage into the validation and test sets. With the implementation of these parameters, the dataset size increased from 1,150 original images to 2,000 images, which were proportionally distributed across four disease classes, is detailed in Table 1.

TABLE 1
RICE LEAF IMAGE DATASET DISTRIBUTION PRE-AND POST-AUGMENTATION

Image Class	Original Images	After Augmentation
Healthy Leaf	320	557
Blight	280	487
Brown Spot	300	522
Leaf Smut	250	434
Total	**1,150**	**2,000**

C. Data Partitioning

To rigorously mitigate potential data leakage and ensure the validity of experimental results, the data partitioning process was strictly performed prior to any data augmentation procedures. The original dataset of 1,150 images was first divided into training, validation, and testing sets. Only after this initial separation were data augmentation technique-including rotation (±20°), zooming (0.8-1.2 scale), horizontal flipping, and brightness adjustment (±30%)-applied exclusively to the training subset. This specific sequence ensures that no augmented variants of a single original images are present across different data splits, thereby preventing information leakage that could lead to overly optimistic or biased performance estimates. By maintaining this strict separation, the evaluation-both in the 70/10/20 hold-out and the 10-fold cross-validation-remains unbiased and accurately reflects the model's true generalization capability on completely unseen data.

D. Convolutional Deep Learning Model

CNN is a variation of Multi-Layer Perceptron (MLP) neural network, which was constructed by imitating of neurons in the visual cortex of humans [25]. CNN is commonly used for image data and has high-depth network, following its effectiveness with feature learning, CNN is more effective in a variety of image processing applications. CNN are excellent at automatically learning hierarchical characteristics [26]. As they go through deeper layers, they gradually combine fundamental visual aspects like colour, textures, and shapes to create more accurate representations [27]. CNN's feature learning may be more effective than manually created features like shape, morphological, and wavelet features, etc [28].

The architecture is made from an output layer, many hidden layers, and input layer of neurons. CNN neurons have weight, bias, and activation functions, each of the neurons in a layer is linked to another layer's neuron [29]. Figure 3 shows the general CNN architecture. CNN has two stages, as shown in the figure: the feature extraction stage, which is utilized for feature extraction, and the classification stage. A convolution layer with ReLU activation and a pooling layer make up the CNN feature extraction stage, while a fully connected layer with softmax activation makes for the classification phase [30].

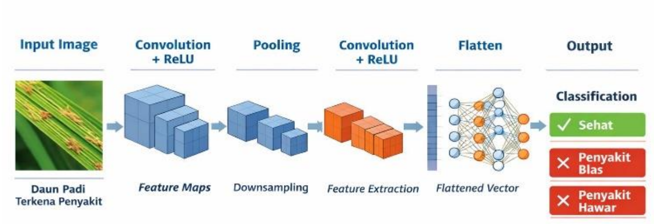

Figure 3. CNN Architecture

E. Residual Network 50 (ResNet-50)

The proposed model utilizes a ResNet-50 architecture pre-trained on the ImageNet dataset to leverage the capabilities of transfer learning. The configuration is performed by freezing the initial convolutional layers to preserve generic feature representations, while the deeper layers are fine-tuned to adapt to the visual characteristics of rice leaf diseases.

The input images is processed at a resolution of 224 x 224 pixels. In the final stage, the fully connected layer is configured with a softmax activation function to accurately distinguish among four target classes. The training process was conducted using the Adam optimizer with an initial learning rate of 0.001. To maintain gradient stability and computational efficiency, the model was trained with a batch size of 32 for 10 to 20 epochs, which proved sufficient to achieve convergence in this transfer learning scenario with the dataset.

F. K-Fold Validation Technique

K-Fold Cross-Validation (CV) is a validation technique used for model selection in learning problem, performed over n iterations [35]. This approach helps mitigate overfitting and reduces bias in data selection [36]. In K-Fold CV, dataset $\mathcal{D}$ is randomly split into k non-overlapping subsets of similar size. The data distribution in each subset is preserved as consistently as possible using hierarchical sampling by $\mathcal{D}$ [37]. Data set $\mathcal{D}$ is randomly split into groups; $k-1$ group are utilized for training, while the remaining group serves as the testing set. The algorithm is trained on each training subset in turn, after which the mean error across the corresponding test subset instances is computed [38]. Figure 5 illustrates the implementation of K-Fold CV with $k = 5$.

Figure 4. Model Evaluation Employed K-Fold CV

G. Performance Evaluation

In addition to global metrics, a confusion matrix analysis is methodically employed to examine class-wine misclassification trends. This approach is crucial for identifying specific patterns of error, particularly between disease classes that may share overlapping visual features, such as textures and colour patterns.

III. RESULTS AND DISCUSSION

A. Experimental Setup and Model Configuration

In this study, the experimental setup was designed to ensure both computational efficiency and stable model convergence within a transfer learning framework. All input images were resized to 224 x 224 pixels, a standard configuration widely adopted in pre-trained convolutional neural network architecture to maintain compatibility and optimize feature extraction. The model architecture was finalized with a fully connected layer employing a softmax activation function, enabling effective discrimination across the four target classes. For optimization, the Adam algorithm was utilized with an initial learning rate of 0.001, as it known to provide adaptive learning capabilities and faster convergence compared to conventional gradient descent methods. Additionally, a batch size of 32 was selected to balance memory efficiency and gradient stability during training. The training process was conducted over 10 to 20 epochs, which proved sufficient for achieving convergence without overfitting, particularly in transfer learning scenarios where pre-trained weights accelerate the learning process.

B. Impact of Data Augmentation on Model Generalization

The use of data augmentation methods, such as rotation, padding-crop, and horizontal flipping, significantly increased the ResNet-50 model's capacity for generalization in the identification and categorization of rice leaf disease, as can be seen in Figure 5.

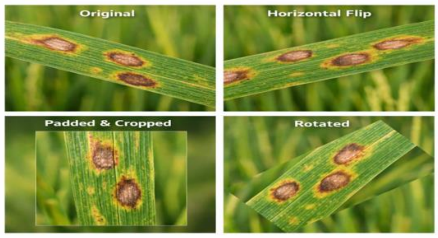

Figure 5. Horizontal Flip, Padded-Cropped, and Rotated

Changes in leaf orientation, camera angle, field placement, and partial occlusion-all of which are common in agricultural environments-are simulated by these modifications. Therefore, instead of learning fixed spatial patterns, the model must acquire invariant and selective features. According to empirical data, including augmentation improved classification accuracy from 82.50% to 93.75%, indicating a significant improvement in predicting performance.

C. Performance Comparison Between Non-Augmented and Augmented Training Models

Figure 6 presents the training loss and accuracy trends for the model trained without data augmentation. It can be observed that the learning process began to reach a stable phase around the fourth epochs. Subsequently, the training loss continued to decline progressively between the fifth and seventh epochs. From eighth to tenth epochs, the model maintained stable behaviour, while the validation loss consistently decreased, suggesting enhanced generalization capability on previously unseen data.

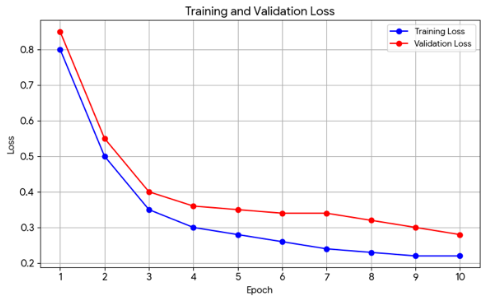

Figure 6. Accuracy and Loss Without Augmentation

On the other hand, Figure 7 shows the accuracy and loss curves for the enhanced model. Performance has significantly improved, as seen by the graph, with training accuracy increasing from 37.21% in the 1st epoch to 88.57% by the 10th. Stronger generalization performance on unseen leaf images was the outcome of validation accuracy, which also improved the model's capacity to recognize unique patterns in rice leaf disease symptoms.

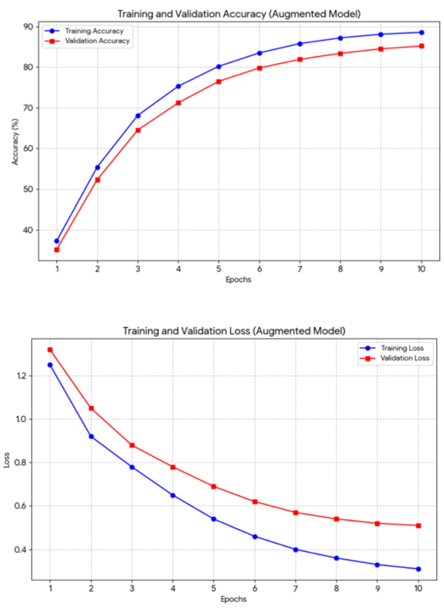

Figure 7. Accuracy and Loss With Augmentation

A comparison of model performance with and without data augmentation. The model trained on 1150 images without augmentation achieved 82.50% accuracy, while applying augmentation with 2000 images increased accuracy to 93.75%. The augmented model required 4.1 minutes to complete training, whereas the model trained without augmentation finished in only 1 minute. This suggests that although data augmentation increases computation times, model performance is significantly improved.

TABLE 2
IMPACT OF DATA AUGMENTATION ON RESNET-50 PERFORMANCE

Parameter	Without Augmentation	With Augmentation
Dataset Size	1150	2000
Accuracy (%)	82.5	93.7
Training Time	1 Minutes	4.1 Minutes

D. K- Fold Cross-Validation Results and Augmentation Impact Analysis

Ten-fold cross-validation was used to evaluate the ResNet-50 model's performance without data augmentation in order to assess its generalization across different data subsets. The results show noticeable variation in accuracy and loss values across folds, where Fold 8 achieved the highest accuracy of 100%, while Folds 1 and 2 recorded the lowest performance at 82.50%, indicating that the model struggled with certain data distributions, as presented in Figure 8.

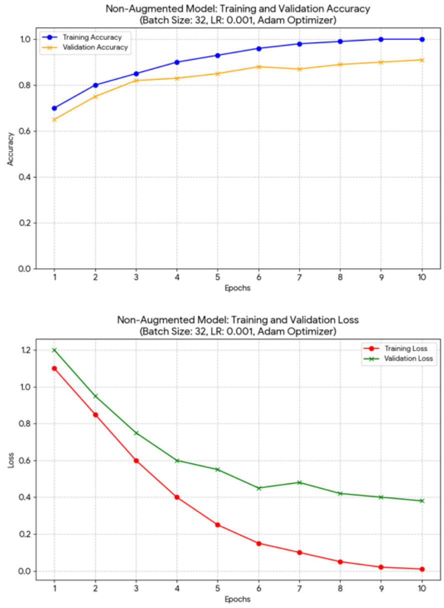

Figure 8. K-Fold CV and Loss Without Data Augmentation

Figure 9 depicts the performance of the enhanced model in terms of accuracy and loss using K-fold cross-validation. The model achieves an initial training accuracy of 98.28% and gradually reaches 100% by the final fold. This steady increase reflects a marked enhancement in performance, indicating that the model maintains highly accurate predictions across all validation folds and exhibits strong generalization ability.

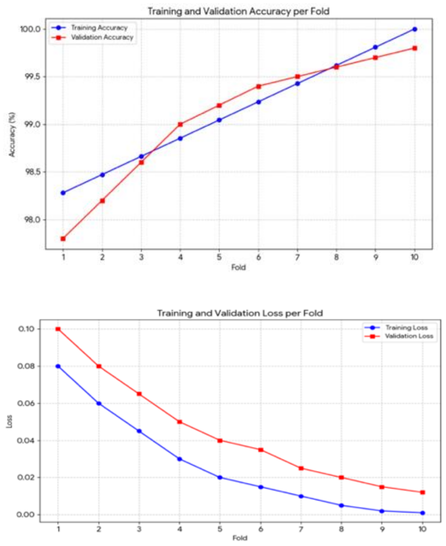

Figure 9. K-Fold CV Accuracy and Loss With Data Augmentation

Based on Table 3 dan Table 4 present the K-Fold cross-validation results without and with data augmentation, both indicating exceptionally high classification performance across folds. In Table 3, the model without augmentation already achieves strong accuracy, with several folds reaching very high to perfect performance, including Fold-8 with 100%. However, after applying augmentation (Table 4), the results become even more consistently elevated, with most folds achieving perfect accuracy (100%) and the remaining folds classified as very high to near-perfect, suggesting a substantial improvement in model generalization.

TABLE 3
ACCURACY OF K-FOLD CV WITHOUT AUGMENTATION

Performance Category	Fold	Accuracy (%)
Low	Fold-1, Fold-2	82.50
Moderate	Fold-4	87.50
High	Fold-3, Fold-6, Fold-9	92.50
Very High	Fold-5, Fold-7, Fold-10	97.50
Perfect	Fold-8	100.00

TABLE 4
ACCURACY OF K-FOLD CV WITH AUGMENTATION

Performance Category	Fold	Accuracy (%)
High	Fold-1, Fold-2	97.40
Very High	Fold-9	98.47
Near-Perfect	Fold-3	99.83
Perfect	Fold-4, Fold-5, Fold-6, Fold-7, Fold-8, Fold-10	100.00

Statistical evaluation based on 10-folds cross-validation show a variation in model stability. Based on the data presented in Table 3, the accuracy of the model without augmentation exhibits a relatively wide fluctuation, ranging from 82.50% to 100 %, resulting in a standard deviation of approximately ± 5.96%. This high variance indicates that the model without augmentation is highly sensitive to variations in data distribution across different folds, reflecting a susceptibility to partitioning bias. The data in Table 4 indicate that the model with augmentation exhibits substantially higher consistency, characterized by a much narrower accuracy range (97.40% to 100%) and a significantly lower standard deviation to ± 1.02%. The marked decrease in standard deviation offers robust statistical evidence that augmentation techniques substantially improve the model's stability, confirming that the attained mean accuracy of 99.6% results from enhanced feature extraction performance.

The experimental results reveal a notable discrepancy between the validation schemes, where 10-folds cross-validation achieved a substantially higher average accuracy of 99.6% compared to the 88.1% obtained through hold-out validation. While the cross-validation performance suggests high stability across folds, it must be acknowledged that such near-perfect results often introduce an optimistic bias.

This phenomenon typically occurs, because data is reused across iterations, and in a relatively limited dataset, similarities in image characteristics-such as consistent backgrounds, lighting condition, and object features from the Sleman region-can contribute to inflated performance estimates. Furthermore, although data augmentation enhances robustness, it may also increase the similarity between samples, potentially leading to performance that does not fully reflect real world variability. Consequently, the hold-out validation results is regarded as a more realistic and reliable estimate of the model's true generalization capability on completely unseen data.

TABLE 5
RESNET-50 PERFORMANCE ANALYSIS WITH AND WITHOUT DATA AUGMENTATION USING K-FOLD

Model Configuration	Number of Samples	Accuracy (%)	Computation Time (Minutes)
K-Fold CV (no augmentation)	1150	92.25	3.0
K-Fold CV (with augmentation)	2000	99.67	7.0

To ensure the statistical significance and reliability of the performance gains, a comparative analysis of the mean accuracy and variance across folds was conducted. The results show a systematic improvement, where the application of data augmentation increased the average 10-folds cross-validation accuracy from 92.3% to 99.6%. Furthermore, the augmented model exhibited higher statistical stability, as evidenced by the reduction in performance fluctuation; while the non-augmented model showed wide range accuracy (82.5% to 100%), the augmented model consistently achieved near-perfect results across nearly folds. This consistently indicates that the performance improvement is not a result of a favorable data split but a robust enchancement ot the model's feature extraction capability.

E. *Classification Results of Rice Leaf Disease Detection without Data Augmentation*

The confusion matrix for the model created to identify and categorize rice leaf diseases using the ResNet-50 architecture and a dataset with four classes is shown in Table 6. There were 1.150 image utilized in the experiment, comprising 320 images of healthy leaves, 280 images of blight, 300 images of brown spot, and 250 images of leaf smut. The dataset distribution indicates that healthy leaf images account for 27.83% of the total samples, followed by brown spot with 26.09%, blight with 24.35%, and leaf smut with 21.74%, providing a relatively balanced representation across the four rice leaf condition categories.

TABLE 6
DISTRIBUTION OF IMAGE CLASSES USED IN THE CLASSIFICATION MODEL

Image Class	Number of Images	Percentage (%)
Healthy Leaf	320	27.83 %
Blight	280	24.35 %
Brown Spot	300	26.09 %
Leaf Smut	250	21.74 %
Total	1.150	

Table 6 summarizes the classification outcomes of the model without augmentation, demonstrating its effectiveness in distinguishing four rice leaf conditions: healthy leaves, blight, brown spot, and leaf smut. A dataset of 1.150 rice leaf images- 320 healthy leaves, 280 blight, 300 brown spot, and 250 leaf smut samples-was used for the test.

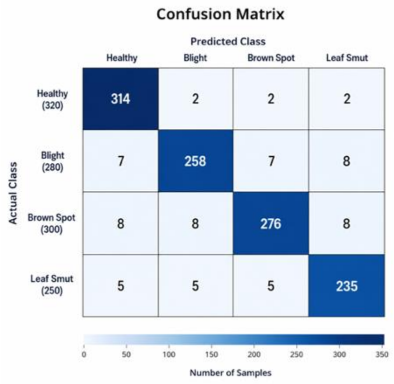

Figure 10. Confusion Matrix for Model Without Augmentation

The confusion matrix in Figure 10, shows strong overall classification performance, with the model achieving high correctness across all four classes (Healthy, Blight, Brown Spot, and Leaf Smut). The diagonal values dominate, indicating that most samples are correctly classified: 314/320 for Healthy, 258/280 for Blight, 276/300 for Brown Spot, and 235/250 for Leaf Smut. This corresponds to an overall accuracy of approximately 94.17%, suggesting the model is highly effective in distinguishing between these plant conditions. Among the classes, Healthy show the highest precision with very minimal misclassification, while Leaf Smut also maintain relatively strong performance despite a slightly smaller dataset. The matrix also shows several noteworthy, recurring misclassification tendencies. The comparatively higher off-diagonal numbers (eg., 7-8 misclassifications per category) suggest that Blight and Brown Spot are somewhat confused with other classes, especially with each other and with Leaf Smut. This suggests that the disease categories may share similar visual features, which makes it more challenging for the model to distinguish between them.

TABLE 7
CLASSIFICATION REPORT FOR THE NON-AUGMENTED MODEL

Class	Precision (%)	Recall (%)	F1-Score (%)	Support
Healthy Leaf	0.97	0.98	0.97	320
Blight	0.92	0.92	0.92	280
Brown Spot	0.93	0.92	0.93	300
Leaf Smut	0.95	0.94	0.94	250

F. *Classification Results of Rice Leaf Disease Detection with Data Augmentation*

The confusion matrix (Figure 10 and Figure 11) provide deeper insights into the model's performance by highlighting recurring misclassification tendencies. Analysis of the non-augmented model (Figure 10) reveals that Blight and Brown Spot were frequently confused with each other and Leaf Smut, likely due to their similar visual traits. However, following the application of data augmentation, the model's ability to capture discriminative feature improved significantly reducing inter-class confusion. For, instance Figure 11 shows that while minor overlap remains between Blight and Brown Spot, the diagonal values dominate, indicating that the model has become more robust in distinguishing visually identical disease.

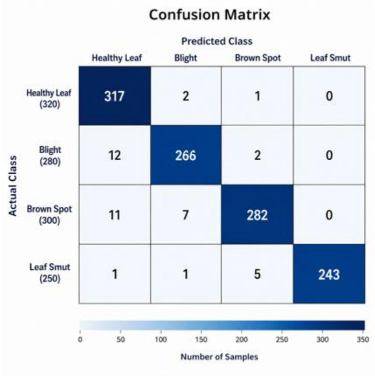

Figure 11. Confusion Matrix for Model With Augmentation

TABLE 8
CLASSIFICATION REPORT FOR THE AUGMENTED MODEL

Class	Precision (%)	Recall (%)	F1-Score (%)	Support
Healthy Leaf	0.99	0.99	0.99	320
Blight	0.96	0.95	0.95	280
Brown Spot	0.96	0.94	0.95	300
Leaf Smut	0.94	0.97	0.96	250

G. Model Performance Evaluation

The notable discrepancy between the hold-out validation accuracy (88.1%) and the 10-fold cross-validation accuracy (99.6%) as seen in Table 9, requires a critical interpretation regarding model generalization. While 10-fold cross-validation provides an assessment of stability across different data partitions, the near-perfect results often reflect an optimistic bias typical in relatively small and region-specific datasets. This inflation occurs because k-fold cross-validation involves multiple iterations of data reuse, when combined with data augmentation, can reduce the distinctiveness between training and validation samples.

Specially, the consistent environmental characteristics of the Sleman dataset-such as uniform background textures and lighting conditions-likely allowed the model to over-optimize features during the cross-validation process. Consequently, the hold-out validation results (88.1%) is considered a more realistic and conservative estimate of the model's true performance on completely unseen data, as it ensures a stricter separation between the training and evaluation sets.

TABLE 9
SUMMARY OF MODEL PERFORMANCE EVALUATION RESULTS

Experiment Setup	Dataset Size	Validation Method	Accuracy (%)	Training Time
ResNet-50 without Augmentation	1150 images	Hold-out validation	82.4	1 Minutes
ResNet-50 with Augmentation	2000 images	Hold-out validation	88.1	4.1 Minutes
ResNet-50 without Augmentation	1150 images	10-Fold Cross Validation	92.3	2.5 Minutes
ResNet-50 with Augmentation	2000 images	10-Fold Cross Validation	99.6	6.5 Minutes

Although the proposed ResNet-50 model demonstrates strong accuracy and stability on the Sleman dataset, the authors note that these outcomes reflect particular regional characteristics, which could constrain its direct applicability across different geographical settings. The assertion of a "robust and accurate approach" is supported by the model's stable performance in addressing real-world environmental variability within the local dataset, achieved through systematic data augmentation and evaluation using 10-fold cross-validation. To comprehensively assess the model's robustness and confirm its reliability for practical precision agriculture applications, additional testing using diverse external datasets—covering various climatic conditions and rice varieties—is necessary and will be a key priority in future research.

IV. CONCLUSION

In this work, a deep learning-based approach utilizing ResNet-50 in conjunction with data augmentation and K-Fold cross validation is proposed for rice leaf disease identification. The results show that the suggested approach greatly enhances model robustness and classification performance, especially when working with small and uneven agricultural datasets. The use of data augmentation effectively decreased overfitting and expanded dataset diversity, resulting in considerable gains in classification accuracy. Additionally, the application of 10-fold cross-validation produced a high average accuracy of 99.6%, offering a more trustworthy and an objective evaluation. Furthermore, the model demonstrated excellent performance in all disease classes, with increased precision, recall, and F1-Score, demonstrating its efficacy in differentiating between visually identical disease symptoms. For automated rice leaf disease identification, the combination of ResNet-50, data augmentation, and K-fold cross-validation provides a reliable and scalable solution. By facilitating early disease detection, lowering crop losses, and supporting farmers in their decision-making, this strategy offers a great ideal of promise to support precision agriculture. However, the dataset size and geographic coverage study are limitations. To improve practical use in the field, future research should concentrate on adding more varied environmental conditions to the dataset, investigating lightweight model architectures for real-time development, and integrating the systems with mobile or IoT-based agricultural applications.

REFERENCES

[1] R. Deng *et al.*, "Automatic Diagnosis of Rice Diseases Using Deep Learning," *Front. Plant Sci.*, vol. 12, Aug. 2021, doi: 10.3389/fpls.2021.701038.

[2] Y. Li, X. Chen, L. Yin, and Y. Hu, "Deep Learning-Based Methods for Multi-Class Rice Disease Detection Using Plant Images," *Agronomy*, vol. 14, no. 9, p. 1879, Aug. 2024, doi: 10.3390/agronomy14091879.

[3] M. T. Ahad, Y. Li, B. Song, and T. Bhuiyan, "Comparison of CNN-based deep learning architectures for rice diseases classification," *Artificial Intelligence in Agriculture*, vol. 9, pp. 22–35, Sep. 2023, doi: 10.1016/j.aiia.2023.07.001.

[4] S. R. Shah, S. Qadri, H. Bibi, S. M. W. Shah, M. I. Sharif, and F. Marinello, "Comparing Inception V3, VGG 16, VGG 19, CNN, and ResNet 50: A Case Study on Early Detection of a Rice Disease," *Agronomy*, vol. 13, no. 6, p. 1633, Jun. 2023, doi: 10.3390/agronomy13061633.

[5] M. S. A. M. Al-Gaashani, N. A. Samee, R. Alnashwan, M. Khayyat, and M. S. A. Muthanna, "Using a Resnet50 with a Kernel Attention Mechanism for Rice Disease Diagnosis," ., vol. 13, no. 6, p. 1277, May 2023, doi: 10.3390/life13061277.

[6] C. G. Simhadri and H. K. Kondaveeti, "Automatic Recognition of Rice Leaf Diseases Using Transfer Learning," *Agronomy*, vol. 13, no. 4, p. 961, Mar. 2023, doi: 10.3390/agronomy13040961.

[7] H. Liu, Y. Cui, J. Wang, and H. Yu, "Analysis and Research on Rice Disease Identification Method Based on Deep Learning," *Sustainability*, vol. 15, no. 12, p. 9321, Jun. 2023, doi: 10.3390/su15129321.

[8] P. Pai *et al.*, "Deep learning-based automatic diagnosis of rice leaf diseases using ensemble CNN models," *Sci. Rep.*, vol. 15, no. 1, p. 27690, Jul. 2025, doi: 10.1038/s41598-025-13079-z.

[9] N. Sultana, S. Sharmin, and M. S. Uddin, "A generalized leaf disease recognition of various crops and vegetables through computer vision and machine learning," *Neural Comput. Appl.*, vol. 37, no. 25, pp. 21069–21089, Sep. 2025, doi: 10.1007/s00521-025-11382-4.

[10] T. H. Nhut, V. P. T. Trung, C. M. Khoi, N. P. H. Khang, and N. C. Nhan, "An AIoT - Based System for Efficient Classification of

Rice Leaf Diseases," *IEEJ Transactions on Electrical and Electronic Engineering*, vol. 20, no. 11, pp. 1875–1881, Nov. 2025, doi: 10.1002/tee.70163.

[11] K. Mostafa, F. Imran, M. Anjum, and S. I. Siam, "A Hybrid Deep Learning Framework for Robust Detection of Rice Leaf Diseases Using CNN and Transformer-Based Attention," in *2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN)*, IEEE, Jul. 2025, pp. 1–6. doi: 10.1109/QPAIN66474.2025.11171957.

[12] C. R. Carlson, J. Alghazo, and W. U. Hasan, "AI-Powered Detection of Rice Leaf Diseases: A Survey of CNN Architectures and Attention-Based Methods," in *2025 IEEE International Conference on Imaging Systems and Techniques (IST)*, IEEE, Oct. x2025, pp. 1–6. doi: 10.1109/IST66504.2025.11268365.

[13] M. S. Hafiz, A. Anwar, J. Shin, and J. I. Rafiq, "Improving Agronomic Disease Detection and Classification: The Superiority of Hybrid Inception-Xception Ensemble Model for Rice Leaf Disease Classification," 2025, pp. 57–67. doi: 10.1007/978-981-96-4520-6_5.

[14] M. Hamdy, H. M. ElHoseny, S. A. Eldolil, A. S. El-Fishawy, and E. Ibrahim, "AlexNet_ResNet_Concatenation Model Based on Rice Diseases Classification Enhancement," in *2025 4th International Conference on Electronic Engineering (ICEEM)*, IEEE, Oct. 2025, pp. 1–8. doi: 10.1109/ICEEM66692.2025.11224994.

[15] S. V. Rakshit, A. M. M. M. I. Chowdhury, A. Chakma, I. Hasan, S. Roy, and T. Rahman, "DeepLeafDiagnosis: A Two-Phase Transfer Learning Strategy for Perfect Accuracy in Rice Leaf Disease Classification," in *2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN)*, IEEE, Jul. 2025, pp. 1–6. doi: 10.1109/QPAIN66474.2025.11172163.

[16] O. O. Martins, C. C. Oosthuizen, and D. A. Desai, "RiceLeafClassifier -v1.0: A Quantized Deep Learning Model for Automated Rice Leaf Disease Detection and Edge Deployment," *Engineering Reports*, vol. 7, no. 6, Jun. 2025, doi: 10.1002/eng2.70231.

[17] C. R. Carlson, J. Alghazo, and W. U. Hasan, "AI-Powered Detection of Rice Leaf Diseases: A Survey of CNN Architectures and Attention-Based Methods," in *2025 IEEE International Conference on Imaging Systems and Techniques (IST)*, IEEE, Oct. 2025, pp. 1–6. doi: 10.1109/IST66504.2025.11268365.

[18] R. Muslim, Z. Zaeniah, A. Akbar, B. Imran, and Z. Zaenudin, "Disease Detection of Rice and Chili Based on Image Classification Using Convolutional Neural Network Android-Based," *Jurnal Pilar Nusa Mandiri*, vol. 19, no. 2, pp. 85–96, Sep. 2023, doi: 10.33480/pilar.v19i2.4669.

[19] María. J. López-Serrano, J. F. Velasco-Muñoz, J. A. Aznar-Sánchez, and I. M. Román-Sánchez, "Farmers' Attitudes towards Irrigating Crops with Reclaimed Water in the Framework of a Circular Economy," *Agronomy*, vol. 12, no. 2, p. 435, Feb. 2022, doi: 10.3390/agronomy12020435.

[20] L. Zhang, Y. Bian, P. Jiang, and F. Zhang, "A Transfer Residual Neural Network Based on ResNet-50 for Detection of Steel Surface Defects," *Applied Sciences*, vol. 13, no. 9, p. 5260, Apr. 2023, doi: 10.3390/app13095260.

[21] R. Escobar Díaz Guerrero, L. Carvalho, T. Bocklitz, J. Popp, and J. L. Oliveira, "A Data Augmentation Methodology to Reduce the Class Imbalance in Histopathology Images.," *Journal of imaging informatics in medicine*, vol. 37, no. 4, pp. 1767–1782, Aug. 2024, doi: 10.1007/s10278-024-01018-9.

[22] H. Bichri, A. Chergui, and M. Hain, "Investigating the Impact of Train / Test Split Ratio on the Performance of Pre-Trained Models with Custom Datasets," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 2, 2024, doi: 10.14569/IJACSA.2024.0150235.

[23] C. Paramita, S. Rakasiwi, P. N. Andono, G. F. Shidik, Shier Nee Saw, and M. I. Rafsanjani, "Deep Learning-Based Eye Disorder Classification: A K-Fold Evaluation of EfficientNetB and VGG16 Models," *Scientific Journal of Informatics*, vol. 12, no. 3, pp. 441–452, Sep. 2025, doi: 10.15294/sji.v12i3.26257.

[24] J. White and S. D. Power, "k-Fold Cross-Validation Can Significantly Over-Estimate True Classification Accuracy in Common EEG-Based Passive BCI Experimental Designs: An Empirical Investigation," *Sensors*, vol. 23, no. 13, p. 6077, Jul. 2023, doi: 10.3390/s23136077.

[25] Y. Liu, H. Pu, and D.-W. Sun, "Efficient extraction of deep image features using convolutional neural network (CNN) for applications in detecting and analysing complex food matrices," *Trends Food Sci. Technol.*, vol. 113, pp. 193–204, Jul. 2021, doi: 10.1016/j.tifs.2021.04.042.

[26] P. Nayak *et al.*, "2D-convolutional neural network based fault detection and classification of transmission lines using scalogram images," *Heliyon*, vol. 10, no. 19, p. e38947, Oct. 2024, doi: 10.1016/j.heliyon.2024.e38947.

[27] T. Tesileanu, E. Piasini, and V. Balasubramanian, "Efficient processing of natural scenes in visual cortex.," *Front. Cell. Neurosci.*, vol. 16, p. 1006703, 2022, doi: 10.3389/fncel.2022.1006703.

[28] N. Wiem, C. M. Ali, and E. Ridha, "Wavelet Feature with CNN for Identifying Parasitic Egg from a Slender-Billed's Nest," 2021, pp. 365–374. doi: 10.1007/978-3-030-73050-5_37.

[29] M. Vakalopoulou, S. Christodoulidis, N. Burgos, O. Colliot, and V. Lepetit, "Deep Learning: Basics and Convolutional Neural Networks (CNNs)," 2023, pp. 77–115. doi: 10.1007/978-1-0716-3195-9_3.

[30] A. Zafar *et al.*, "A Comparison of Pooling Methods for Convolutional Neural Networks," *Applied Sciences*, vol. 12, no. 17, p. 8643, Aug. 2022, doi: 10.3390/app12178643.

[31] A. Stanojevic, S. Woźniak, G. Bellec, G. Cherubini, A. Pantazi, and W. Gerstner, "An exact mapping from ReLU networks to spiking neural networks," *Neural Networks*, vol. 168, pp. 74–88, Nov. 2023, doi: 10.1016/j.neunet.2023.09.011.

[32] A. B., M. Kaur, D. Singh, S. Roy, and M. Amoon, "Efficient Skip Connections-Based Residual Network (ESRNet) for Brain Tumor Classification," *Diagnostics*, vol. 13, no. 20, p. 3234, Oct. 2023, doi: 10.3390/diagnostics13203234.

[33] A. R. Rizalde, H. A. Mubarak, B. Khairunnisa, and Mohd. A. Fatan, "A Deep Learning Approach Bassed on Classification to Detect Facial Skin Defect," *Public Research Journal of Engineering, Data Technology and Computer Science*, vol. 2, no. 2, pp. 93–106, Jan. 2025, doi: 10.57152/predatecs.v2i2.1558.

[34] W. Zhang, P. Zhang, B. Zhang, X. Wang, and D. Wang, "A Collaborative Transfer Learning Framework for Cross-domain Recommendation," in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, New York, NY, USA: ACM, Aug. 2023, pp. 5576–5585. doi: 10.1145/3580305.3599758.

[35] J. Allgaier and R. Pryss, "Cross-Validation Visualized: A Narrative Guide to Advanced Methods," *Mach. Learn. Knowl. Extr.*, vol. 6, no. 2, pp. 1378–1388, Jun. 2024, doi: 10.3390/make6020065.

[36] F. M. Gonçalves Saraiva, A. C. Rodrigues Guimarães, O. S. da Silva Júnior, S. N. Monteiro, and L. M. Coelho, "Performance analysis of Gradient Boosting Regressor using K-fold and group K-fold for predicting permanent deformation in tropical soils," *Journal of Materials Research and Technology*, vol. 40, pp. 3555–3567, Jan. 2026, doi: 10.1016/j.jmrt.2025.12.152.

[37] G. Amir, O. Maayan, T. Zelazny, G. Katz, and M. Schapira, "Verifying the Generalization of Deep Learning to Out-of-Distribution Domains," *J. Autom. Reason.*, vol. 68, no. 3, p. 17, Sep. 2024, doi: 10.1007/s10817-024-09704-7.

[38] J. K. Parker and L. O. Hall, "Accelerating Fuzzy-C Means Using an Estimated Subsample Size.," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 5, pp. 1229–1244, Oct. 2014, doi: 10.1109/TFUZZ.2013.2286993.

[39] W. Shafik, A. Tufail, L. C. De Silva, R. A. A. Haji Mohd Apong, and K.-H. Kim, "Deep learning technique for plant disease classification and pest detection and model explainability elevating agricultural sustainability," *BMC Plant Biol.*, vol. 25, no. 1, p. 1491, Nov. 2025, doi: 10.1186/s12870-025-07377-x.

[40] B. Naresh kumar and S. Sakthivel, "Rice leaf disease classification using a fusion vision approach," *Sci. Rep.*, vol. 15, no. 1, p. 8692, Mar. 2025, doi: 10.1038/s41598-025-87800-3.