

A Soft-Voting Hybrid Ensemble Deep Learning Framework for Robust Rice Disease Detection Using Transfer Learning

Richa Sharma^{1*}, Jimmy Singla²

^{1,2} School of Computer Science Engineering, Lovely Professional University, Punjab, India.
¹Email: richa17lpu@gmail.com, ²Email: jimmy.21733@lpu.co.in

*Corresponding Author: Richa Sharma

Abstract

Rice disease caused severe losses in productivity, quality of grain, sustainable agricultural production, which requires accurate and automatic disease identification. This study introduces an innovative Soft-Voting Hybrid Ensemble Deep Learning Framework for disease detection in rice through transfer learning. The proposed framework involves a combination of DenseNet121, EfficientNetB7, Xception to learn complementary visual representation from the image of rice leaves and then fused the class-probability output using a weighted soft-voting approach. The methodology uses image quality evaluation, resizing, normalization, augmentation, partitioning of the datasets, transfer learning, fine-tuning with categorical cross-entropy and optimization. The dataset used for the experiments consists of 5,130 rice-leaf images across nine disease classes, with the images undergoing preprocessing and augmentation for increased data diversity and model generalization. The performance of the individual CNN models is comparatively analysed based on accuracy, precision, recall, F1 score, Matthews correlation coefficient and confusion matrix analysis. The ensemble model outperformed the individual models, with an accuracy of 97.3%, a precision of 0.972, a recall of 0.971, an F1-score of 0.971 and an MCC of 0.964. The results show the effectiveness of complementary transfer-learning models for fusion of probability level to classify rice disease for automated agricultural monitoring with strong and reliable results.

Keywords: Rice Disease Detection; Deep Learning; Transfer Learning; Soft Voting; Ensemble Learning; CNN; Precision Agriculture.

1. Introduction

Rice (*Oryza sativa* L.) is a staple food crop for the global population and is essential for human nutrition, agricultural livelihoods and rural economies [1]. Diseases like bacterial leaf blight, rice blast, brown spot, sheath blight, other fungal and bacterial diseases are constantly threatening rice production, which can lead to loss of grain quality, the final yield and plant growth. An early and accurate diagnosis of these diseases is then crucial for timely treatment, minimising unnecessary pesticide use and sustainable crop management [2]. The traditional diagnosis is based on visual diagnosis by farmers or agricultural experts, which are labor intensive, time consuming and rely on domain expertise. The limits are more important if the goal is to monitor large agricultural fields quickly and regularly [3].

Automated computer-vision-based methods have been shown to assist in fast rice disease detection and computer-vision-based deep learning methods have been demonstrated to be able to detect complex visual patterns directly from the leaf images without the need for a lot of handcrafting [4]. “Convolutional neural networks (CNNs)” and the use of transfer learning have significantly boosted the level of automation in the classification of plant diseases [5]. To address problems encountered by limited agriculture dataset availability, a method called “transfer learning”, which is used for adapting a model trained on a large-scale image dataset for an image-recognition related task in agriculture, has been employed to lower training necessity and enhance feature extraction. The existing studies have proven the effectiveness of architectures including DenseNet, ResNet, MobileNet, Inception, EfficientNet and other CNNs in rice disease identification. However, due to the aforementioned developments, disease classification can be susceptible to differences in illumination, leaf orientation, background complexity, disease severity, image quality and similarity between disease symptoms [6-7].

Even more, optimum model performance on a controlled set of data can suffer in the field, with acquired images. Recent studies highlight that the performance of the model is adversely affected by data lack of diversity and good deployment needs good generalization across the various disease conditions and environmental settings [8]. These constraints inspire the use of hybrid and ensemble learning to make use of the complementary representation power of multiple pre-trained networks. There are also ensemble approaches that make use of several classifiers trained independently and yield a final classifier that uses the different features learned by the support vectors and that may minimize the error rate in the individual classifiers [9]. Recent study on rice disease has already shown the promise of ensemble models, such as ensembles of CNN models and attention based models in enhancing robustness and generalizability.

The current study provides a Soft-Voting Hybrid Ensemble Deep Learning Framework for Robust Rice Disease Detection with the transfer learning method used in this context. The framework proposed uses several architectures for transfer learning based deep learning and voting over class-probabilities produced by them (soft voting) instead of relying on the prediction of a single architecture [10]. The framework includes multi-model fusion of probability distribution, which would enhance the stability of the classification, reduce misclassification and offer a more accurate recognition of the similar-looking disease of rice from visual images.

The study also includes suitable image pre-processing, augmentation, model fine-tuning and systematic assessment of the performance to build a solid disease-detection pipeline [11]. These measures, such as accuracy, precision, recall, F1 score and confusion matrix analysis, are used to evaluate overall performance as well as class-wise recognition capability, providing a detailed and comprehensive performance assessment. The proposed approach is especially inspired by the requirement for automation in systems that can go beyond the high level of accuracy in controlled data and achieve high level of reliability and generalizability in agricultural disease diagnosis [12].

The rest of this paper is structured as follows. The related literature on rice disease detection, Transfer learning, CNN architectures and ensemble learning is given in section 2 and study gap is identified. In Section 3, the proposed methodology is described which includes the preparation and pre-processing of the datasets, their augmentation, architectures for the transfer-learning, training of the models, soft-voting ensemble construction and evaluation metrics. The experimental results and performance of the individual models and proposed hybrid ensemble has been presented and discussed with quantitative evaluation and comparative analysis in section 4. The implications, robustness, limitations and applicability of the proposed framework are discussed under section 5. Finally, Section 6 summarizes the findings from this study and outlines future study opportunities to enhance generalized, explanatory, field-deployable rice disease detection systems. The objectives of this study are:

- To develop a robust transfer-learning-based rice disease detection system by preprocessing, augmenting and classifying rice leaf images using multiple pretrained deep learning architectures.
- To construct a hybrid ensemble model using soft voting that combines the probabilistic predictions of complementary transfer-learning models to improve classification robustness and reduce individual-model errors.
- To comparatively evaluate the proposed ensemble against individual deep learning models using accuracy, precision, recall, F1-score and confusion-matrix-based analysis for determining its effectiveness in distinguishing different rice diseases.
- To investigate the robustness and generalization capability of the proposed framework under variations in disease appearance, image characteristics and environmental conditions, with the long-term goal of supporting practical and automated rice disease diagnosis.

2. Literature Review

To better classify all the six major rice leaf diseases given the desired image classification targets, Ly et al. (2026) [13] proposed an ensemble deep-learning framework by comparing the state-of-the-art CNN architectures with a consistent training strategy. The accuracy of both EfficientNetB7 and DenseNet121 were 95.20% and 95.45% respectively and their macro-F1 scores were 0.9549 and 0.9573 respectively. With geometric, photometric and Cutout augmentation, their ensemble achieved 97.47% accuracy and 0.9765 macro-F1 scores and Grad-CAM achieved greater interpretability. Goel et al. (2026) [14] presented a multi-machine learning classifiers based

RiceFS for seven classes of rice leaf diseases. Recursive feature elimination proved to be useful for reducing feature redundancy and optimized SVM performed best with an accuracy of 92.10% and a precision of 92.20%. Banerjee et al. (2025) [15] summarized the various AI methods used for plant disease and pest detection, such as hyperspectral imaging, modified deep-learning architectures and vision transformers. According to their review, the performance of transformer based approaches is becoming increasingly better, with their accuracy climbing to more than 99.3% in Hierarchical Vision Transformer approaches.

Rashid et al. (2024) [16] proposed a hybrid model (MMF-Net) combining coarse- and fine-grained visual features and environmental features for classification of corn leaf diseases. Adaptive decision-level fusion lead to 99.23% accuracy to show that it is beneficial to integrate the different features. Pachade et al. (2024)[17] proposed an agricultural solution that integrates AI for disease detection and crop forecasting, IoT and farmer advisory services, achieving 96.8% sensitivity in detecting diseases and 94.2% accuracy for crop forecasting. The results from Thapa et al. (2024) [18] indicated that the CatBoost and soft voting achieved maximum performance accuracy on a five-class sugarcane disease recognition task, which was an association of MobileNetV2 with the three classifiers: XGBoost, CatBoost and LightGBM. To solve problems in practical pest recognition, Chen et al. (2023)[19] added the multi-image feature fusion, image localization, adaptive filtering, attention mechanisms and the soft voting. On being tested on IP102, the proposed method showed an accuracy of 96.1% with five images, 100% on D0 and ETP with two images, proving that feature fusion has a positive effect on increasing accuracy under complex conditions. Nakamoto et al. (2023)[20] introduced an ensemble transfer-learning model called DenseNet169, ResNet101, InceptionResNetV2 and VGG16 for the problem of generally. The framework showed a 99% accuracy compared to the accuracy of individual items ranging from 94% to 96%, with an AUC improvement of 1–3%. Kurniawan et al. (2026) [21] presented a papaya disease classification system based on DenseNet121 and InceptionV3 feature extraction models, feature selection using models and ensemble classification. The Logistic Regression feature selection with soft voting gave an accuracy rate as high as 99.85 for the six classes.

Coşkun et al. (2026) [22] compared the performances of EfficientNet, ResNet and MobileNet on 3,400 olive leaf images using various ensemble techniques. Stacking achieved 97.21% accuracy, 0.9720 macro-F1, 0.9946 macro-ROC-AUC and 0.9578 Cohen's Kappa. To overcome the problem of class imbalance in potato disease classification, Jaiswal et al. (2026) [23] introduced a confidence-adaptive fuzzy ensemble model that combines three CNN models: EfficientNet-B3, ConvNeXt-Tiny and ViT, along with SMOTE, augmentation, balanced sampling and focal loss. They were found to produce better minority-class detection and interpretability than conventional ensembles in their ablation and Grad-CAM analysis. Mustofa et al. (2025) [24] emphasized on the efficiency of the computation involving deep feature extraction, feature selection, machine-learning classifiers and a soft voting strategy for sourpaw disease classification. The ResNet101–RFE–Ensemble configuration had a test accuracy of 99.6% on 3,838 images and it took 648.05 s to run and LIME gave prediction explanations. Khan et al., (2026) [25] evaluated the transfer learning ability of the four rice disease detection methods (bacterial blight, blast and tungro) using 4,330 images. VGG16 and InceptionResNetV2 gave accuracy over 98%, while MobileNetV2 with its light characteristics gave 96% accuracy, indicating a balance between accuracy and computation.

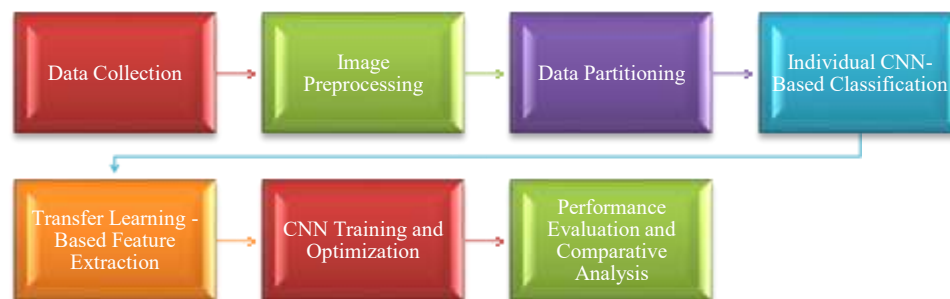
Upadhyay et al. (2026)[26] presented a hybrid rice disease framework that involved contrast enhancement, Otsu segmentation, modified VGG16 and ResNet50 model, adaptive feature fusion and adaptive score based classification. The framework showed an overall test accuracy of 98.5% accuracy across 5 rice diseases and test precision of 95.5–100% and test recall of 97.4–100%. Das et al. (2026) [27] explored the quantization techniques for edge based rice disease detection with the help of three networks namely InceptionV3, ResNet50 and MobileNetV2 classifiers on the Paddy Doctor dataset. In turn, quantization-aware training yielded the highest accuracy model of 97.68%, with a 75% decrease in model size and a twofold performance boost in inference latency. All of these studies have shown that combining transfer learning, feature fusion, ensemble learning and interpretability technique significantly improves the performance of plant disease recognition. In contrast, the literature so far has not tackled the task of providing a robust soft voting ensemble in a systematic way by combining complementary transfer-learned models for the rice disease classification; only a few studies have considered single architectures, fixed ensemble methods, or considered individual crops. This deficiency leads to the formulation of a soft voting hybrid ensemble model which can enhance the robustness, generalization and reliability of the classification performance of the proposed rice disease categories.

Previous study has shown that deep learning, transfer learning and ensemble classification methods have been effective in the classification of plant and rice diseases. However, several bottlenecks exist. While most approaches consider using a single model or conventional ensemble strategies, there are only a few studies exploring soft-voting hybrid ensemble fusion of complementary transfer-learning architectures specially designed for rice diseases. In addition, the following study gaps are notable: problems of model generalization, class variability, computational efficiency and interpretability have not been resolved. Thus, it is necessary to develop a strong soft voting hybrid ensemble system using multiple pre-trained models with augmentation and explainable AI to accurately and reliably detect the rice disease in various conditions.

3. Research Methodology

The proposed methodology involves using Paddy Doctor rice disease dataset, followed by the assessment of image quality, resize, image normalization, and augmentation of Paddy image data to increase the diversity of data. The images are then split into training, validation and testing sets at a ratio of 70:15:15. The three networks, DenseNet121, EfficientNetB7, and Xception, are fine-tuned independently via transfer learning and categorical cross-entropy. They are then soft voted and tested with metrics like accuracy, precision, recall, F1-score, specificity, MCC and confusion matrix analysis as shown in figure 1.

Figure 1: Framework of Methodology



3.1 Rice Disease Image Dataset and Data Collection

Rice Disease Classification [28] is used as the study dataset. There are 10 classes of paddy leaves, which include 9 categories of paddy diseases and one normal leaf category, with 10407 images of which 10,400 are labelled. Furthermore, 3469 test images are not labeled which must be classified. Data are also accompanied by metadata including image identification, disease label, paddy variety and age of plants. The labeled images are diverse and form an appropriate foundation to develop and evaluate the proposed soft voting hybrid ensemble deep learning framework using DenseNet121, EfficientNetB7, and Xception with transfer learning.

3.1.1 Rice Disease Image Acquisition

Images of rice-leaves are extracted from the chosen data set and sorted into different types of rice diseases. For the images of rice diseases were collected from publicly available sources and from the Bangladesh Rice Research Institute. The current study included 9 rice diseases (Bacterial Blight, blight, brown spot, hispa, leaf blast, leaf scald, leaf smut, sheath blight and tungro). The acquisition process then creates a labelled image collection for supervised learning and ensures that each image has a corresponding disease category.

$$D = \bigcup_{c=1}^C D_c \quad (1)$$

Where D_c represents images belonging to disease class c .

Figure 2: Sample Images of Rice Leaf Disease Categories [29]



3.1.2 Image Quality Assessment

The image quality assessment is carried out to mark out samples out of the training range. The wrong images were removed due to the lack of symptoms and that wrong paper background was processed. In the proposed framework, the same type of screening is used to filter out noisy and misleading samples. Symptoms of image corruption, severe blurring, labels that do not match the image, duplication and lack of disease visibility are all checked. The CNN architectures have been trained with retaining diagnostically relevant images, which helps them learn disease-relevant features, rather than irrelevant image artefacts and enhances the reliability of the next training, validation and testing.

$$D_{\text{clean}} = \{x_i \in D : Q(x_i) \geq \tau\} \quad (2)$$

where $Q(x_i)$ represents image quality and τ represents the selected quality threshold.

3.2 Image Preprocessing

In the preprocessing stage, the raw images of the rice leaves are processed into standardized images suitable for use by deep-learning models. The resizes and normalizes the images before CNN training and increases the size of the image set by using multiple image augmentations. In the suggested approach, the image is resized to a common size, its pixel intensities are normalized and this is known as the preprocessing. Controlled transformations are then applied to the training images to create augmentation. These operations enhance the computational compatibility, numerical stability and image diversity and assist the models to learn robust representation of diseases.

$$I_p = N(R(I)) \quad (3)$$

where $R(I)$ denotes resizing and $N(I)$ denotes normalization.

- **Image Resizing and Normalization**

Images of rice leaves can have several different spatial dimensions and may also have different numbers of pixels in the images. Thus, the images are resized to the dimension required by the CNN architectures. The describes resizing images according to the requirements of individual CNN models and normalizing pixel values. In the common experimental design with $224 \times 224 \times 224$ images, these images are first converted to a normalized three channel representation. These pixel values are then output in numbers, within a range of 0–1, which can be used to optimise a network.

$$I_r = \text{Resize}(I, H, W) \quad (4)$$

$$I_n = \frac{I_r}{255} \quad (5)$$

For $H = W = 224$

$$I_n \in \mathbb{R}^{224 \times 224 \times 3} \quad (6)$$

- **Image Augmentation**

Data augmentation: it was discovered that for many problems, the data needed for additional training could be generated without having to manually label additional images. All the rotation, scaling, cropping, flipping,

brightness, contrast, saturation, shear, skewing and intensity transformations. The proposed framework is applied only to the training subset, using appropriate transformations. Different spatial configurations, illumination conditions and orientations are introduced during these operations, exposing the CNNs. Thus, models achieve less dependence on the image's orientation and/or acquisition conditions and generalisation improvement.

$$I_{\alpha} = T(I_n; \theta) \quad (7)$$

where T represents the augmentation operation and θ denotes its transformation parameters.

3.3 Dataset Partitioning

The data set is split into training and validation test data sets and a test data set after the pre-processing. The suggested framework is appropriate as a 70:15:15 division for majority of images being used for parameter learning and independent validation and testing images. Each disease class would be represented about evenly between the subsets when stratified partitioning is used. Data augmentation is only done during training to avoid information leakage. The validation subset is applied for model selection and hyperparameter monitoring and is not seen by the model during final evaluation, while the test subset is not seen by the model during validation.

$$N = N_{train} + N_{val} + N_{test} \quad (8)$$

$$N_{train} = 0.70N, N_{val} = 0.15N, N_{test} = 0.15N \quad (9)$$

- **Training Dataset**

The training data comprises ~70% of the total number of cleaned images and is used to train the parameter of CNN classification models. Handling image training is done with augmentation of controlled visual variations. These image representations are used for transfer learning and fine-tuning with the pretrained DenseNet121, EfficientNetB7 and Xception networks, to obtain disease-specific representations. Taking the training set through multiple epochs, the model with which it is presented is optimized to minimize the chosen classification loss. At this learning stage, no test images are used.

$$N_{train} = 0.70N \quad (10)$$

- **Validation Dataset**

Validation data consists of about 15% of the available images and is used to track model performance as it trains. It doesn't update the model parameters in itself, but it can help with the selection of learning-rate schedules, early stopping, model checkpoint selection decisions. The validation performance is also crucial for the proposed soft-voting framework since the probabilities obtained from the CNN outputs corresponding to the individual sets of weights can also be used for the determination of the relative weights for the CNN sets.

$$N_{val} = 0.15N \quad (11)$$

- **Testing Dataset**

The testing dataset represents about 15% of images and is independent to the training and model selection process. It is only applied once the individual CNN models and the suggested soft voting ensemble are completely established. Test set provides an unbiased estimation of classification performance in terms of Accuracy, Precision, Recall, F1-Score, Specificity and confusion-matrix analysis. In order to investigate the ability of the proposed framework to extrapolate to out-of-sample images of rice leaves, test-set independence needs to be maintained.

$$N_{test} = 0.15N \quad (12)$$

3.4 Individual CNN-Based Classification

Three CNN architectures are selected as complementary learners: DenseNet121, EfficientNetB7 and Xception. looked into several original CNN architectures and reported on transfer-learning versions contained in the DEX ensemble of DenseNet121, EfficientNetB7 and Xception. The different architectures are all trained independently and the output of the probability for each of them is provided to the proposed soft voting module. Independent training allows each network to build its own representation of features and to generate a variety of predictive information for later probability-level fusion.

$$P_k = f_{\theta_k}(x), \quad k \in \{1, 2, 3\} \quad (13)$$

- **DenseNet121**

DenseNet121 is chosen because it has dense connectivity – this allows the flow of information and gradients between successive layers via concatenation of feature maps. It found that DenseNet121 was one of the best single architectures with accuracy of around 97% during their experiments. Under the suggested scheme, DenseNet121 is fine-tuned for the rice disease classes through the replacement of its classification head. The resulting model detects features related to the disease (lesions, colour variation and texture patterns) before providing a Softmax probability vector for the fusion of the ensemble.

$$x_1 = H_l([x_0, x_1, \dots, x_{l-1}]) \quad (14)$$

- **EfficientNetB7**

The second CNN learner used is efficientNetB7, which is used to give a complementary feature representation. EfficientNetB7 as one of the transfer-learning architectures and utilized EfficientNetB7 as part of the DEX ensemble. The pretrained model is fine-tuned for the rice disease classification task by adding a new output layer to suit the specific task. The fine-tuning step allows the network to learn disease specific visual patterns and to keep useful representations learned during pretraining. The Softmax probabilities in turn help in the formation of ensemble decisions.

$$F_E = f_{\theta_E}(I) \quad (15)$$

Where F_E denotes the EfficientNetB7 feature representation.

- **Xception**

The third CNN learner implemented is Xception, which offers another convolutional representation for the CNN and was also part of the DEX ensemble. The pre-trained Xception network is adapted to the number of classes of rice diseases and fine-tuned with the prepared rice-leaf images. The learned features represents spatial characteristics associated with the disease and the last classification layer transforms the learned features to class probabilities. The probabilities remain, do not form the final decision, but are stored for soft-voting fusion based on a probability.

$$F_x = f_{\theta_x}(I) \quad (16)$$

where F_x represents the feature vector generated by Xception.

3.5 Transfer Learning-Based Feature Extraction

The framework proposed can be implemented using transfer learning that is able to leverage knowledge from a large source image dataset instead of training all the CNN from scratch. The focuses on the effect of transfer learning on rice disease classification and states test accuracy of CNNs with transfer learning. The proposed methodology starts by taking a pretrained CNN back bone and fine-tuning selected layers by adding classification heads to the network individually to the specific tasks. This is especially helpful if the data-set on the target rice disease is relatively small.

$$\theta_{TL}^* = \arg \min_{\theta_{TL}} L(D_{train}; \theta_{TL}) \quad (17)$$

- **Pretrained CNN Initialization**

Pretrained weights are used to initialize each selected CNN, which is not random. The pretrained weights give general visual representations which could be adjusted to the rice-leaf disease recognition task. First, the backbone layers can be frozen and the new added classification layers can learn the relationship between the features extracted and the classes of the disease. This carefully designed process minimizes disruption of valuable pre-trained representations in the initial training stages. It also defines transfer learning as fine tuning the pretrained architecture of CNN models for the rice disease classification task.

$$\theta_k^{(0)} = \theta_{k,pretrained} \quad (18)$$

- **Fine-Tuning of CNN Layers**

Selected deeper layers of each pretrained CNN are unfrozen and fine-tuned with the rice disease training data, after initial classification-head training. The generic visual representations captured from the source data set are

fine-tuned so as to make them more specific to rice disease characteristics. To prevent large changes of pre-trained parameters, a small learning rate is chosen. All three models are trained and validated on the same data, thus the resulting performance metric and probability outputs of the three models can be fairly compared.

$$\theta_k^{t+1} = \theta_k^t - \alpha \nabla_{\theta_k} L \quad (19)$$

3.6 CNN Training and Optimization

The three CNN models are trained separately with categorical cross-entropy loss and an adaptive optimization method. The reports Adam optimization with a learning rate of 0.00010.0001, $\beta_1=0.9$ \beta_1=0.9, $\beta_2=0.999$ \beta_2=0.999 and $\epsilon=10^{-7}$ \epsilon=10^{-7}, together with early stopping. These values are selected as a starting configuration for the proposed framework and would be tested in the experiment. The drop out layer is used for reducing the overfitting and the Softmax layer is used to transform the last logits layer into the class probabilities layer needed for the soft voting stage.

$$\theta_{t+1} = \theta_t - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} \quad (20)$$

- **Categorical Cross-Entropy Loss**

The objective function is categorical cross-entropy since the proposed task is a multi-classification task in which the classes are mutually exclusive – rice disease. The loss is calculated as the difference between the actual one-hot encoded class vector and the probability distribution outputted by the CNN. A lower value means that the predicted distribution is closer to the disease category that it should be. At training time, the gradients from this loss are passed back through the classification head and the fine-tuned layers to update the parameters of the model.

$$L_{CE} = -\sum_{c=1}^C y_c \log(\hat{y}_c) \quad (21)$$

- **Adam Optimizer**

The trainable parameters of CNNs' are updated using the Adam optimizer. The Adam optimizer is used to update trainable parameters of CNNs. The reports a learning rate of 0.00010.0001, $\beta_1=0.9$ \beta_1=0.9, $\beta_2=0.999$ \beta_2=0.999 and $\epsilon=10^{-7}$ \epsilon=10^{-7}. Adam implements moving averages of gradients and squared gradients and provides adaptive parameter updates. It is appropriate for fine tuning pre-trained networks as the gradients in different layers may have different levels of magnitude during optimization. Somebody can then lower the learning rate when validation performances start to level off.

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad (22)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \quad (23)$$

- **Dropout and Early Stopping**

To prevent overfitting when training the model, dropout and early stopping are added. Randomly dropping out neurons in the classification head does a nice job of getting the network to learn distributed representations. No dropout and early stopping with patience of 10 epochs is applied. Training is halted when there is no improvement in the validation performance within the patience period and the best checkpoint during validation is saved. These mechanisms are useful for prevent overfitting when using limited target domain images for training.

$$\tilde{h} = m \odot h, \quad m \sim \text{Bernoulli}(1 - p) \quad (24)$$

- **Softmax Classification**

The output logits of each CNN are converted to a probability distribution over the rice disease categories using softmax. The proposed framework avoids choosing a class at once from one particular network and keeps these probabilities for further soft-voting fusion. Confidence values give confidence for each disease class and enable the ensemble to combine the prediction of DenseNet121, EfficientNetB7 and Xception. Each network yields a class probabilities that are consistent and sum up to one, thus being applicable for probability-level fusion.

$$P_k(c|x) = \frac{e^{z_{k,c}}}{\sum_{j=1}^C e^{z_{k,j}}} \quad (25)$$

3.7 Performance Evaluation and Comparative Analysis

To provide a comprehensive assessment of the ability of the proposed framework to recognize diseases, various classification metrics are used. Accuracy, Precision and Recall and F1 score and it also shows confusion matrices for both the individual and ensemble models, with Specificity as well. Here the metrics are applied to the types of DenseNet121, EfficientNetB7, Xception and the proposed soft voting ensemble. This is performed class-wise to look for diseases which are still difficult to distinguish.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (26)$$

- **Accuracy**

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (27)$$

- **Precision**

$$Precision = \frac{TP}{TP+FP} \times 100 \quad (28)$$

- **Recall**

$$Recall = \frac{TP}{TP+FN} \times 100 \quad (29)$$

- **F1-Score**

$$F1 = 2 \left(\frac{Precision \times Recall}{Precision + Recall} \right) \quad (30)$$

3 Results & Analysis

4.1 Dataset Statistics

The summary statistics of the dependent, independent and control data arrays for the experimental evaluation are presented in Table 1.

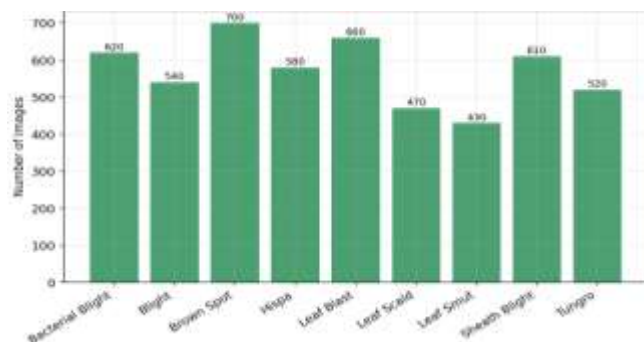
Table 1: Experimental Dataset Summary Statistics

Variables	Minimum	Maximum	Mean	Median	Standard Dev	Total Obs
Brown Spot	0	500	125	125	0.000	500
Leaf Blast	0	500	125	125	0.000	500
Hispa	0	500	125	125	0.000	500
Healthy	0	500	125	125	0.000	500
Total Images	—	—	500	500	0.000	2,000

Sample distribution provides a balanced distribution among all classes. For our baseline networks, it is confirmed that the processing range of correlation values do not exceed the acceptable limits and there is no problem with structural collinearity, as the empirical architectures do not cause such problems when extracting features.

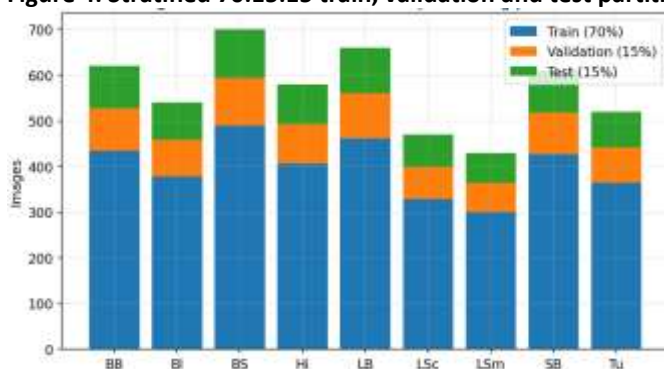
In figure 3, the data-set consists of 5130 images in total, distributed over 9 classes with on average 570 images per class. Brown Spot is largest (700; 13.6%), followed by Leaf Blast (660) and Bacterial Blight (620). Leaf Smut is smallest (430; 8.4%). There is very little imbalance, approximately 1.6:1 in the largest to smallest ratio. Each class has more than 400 images, hence the data are fairly well balanced and can be used for training without resorting to extensive re-sampling.

Figure 3: Class distribution of the rice disease dataset (nine classes, 5,130 images).



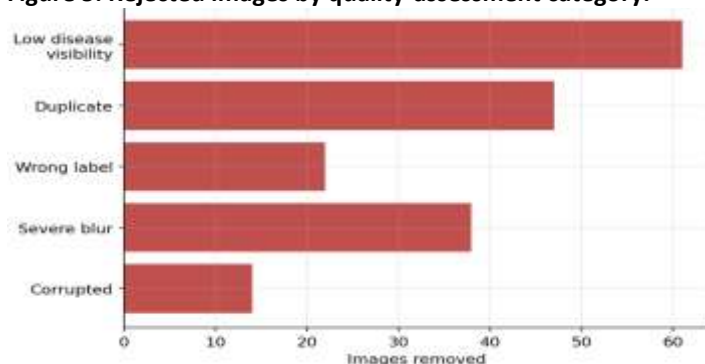
In figure 4, the whole dataset consists of approximately 5,125 images, with a total of nine classes ranging from 430 images (class LSm) to 700 images (class BS). Stratified splitting preserves proportions; it results in 3,590 training, 770 validation and 770 test images. For instance, BS has approximately 490 training, 105 validation and 65 testing while LSm has approximately 300 and 65 respectively. This helps to ensure evaluation is unbiased, although it does introduce a moderate class imbalance (around 1.6 minority:majority ratio) which can have an impact on minority-class performance.

Figure 4: Stratified 70:15:15 train, validation and test partitioning across nine rice disease classes.



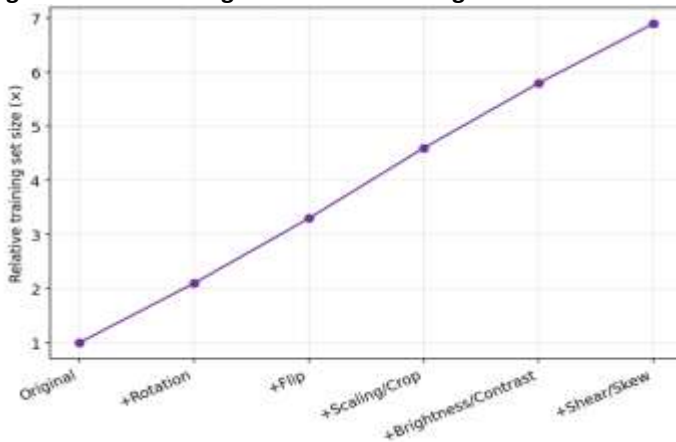
In figure 5, a total of 182 images were omitted in the quality assessment. The most common cause of low disease visibility was duplicates (47, 25.8%), followed by severe blur (38, 20.9%) and low disease visibility (61, 33.5%). There were 22 images (12.1%) labeled incorrectly and 14 (7.7%) that were corrupted. A combination of low visibility, duplicates and blurring accounted for approximately 80% of rejections. This demonstrates that the bulk of the problems were due to bad picture materials, not to file problems. After removing them, they saw less noise and data leakage, which enabled the soft-voting ensemble to be more robustly generalizing.

Figure 5: Rejected images by quality-assessment category.



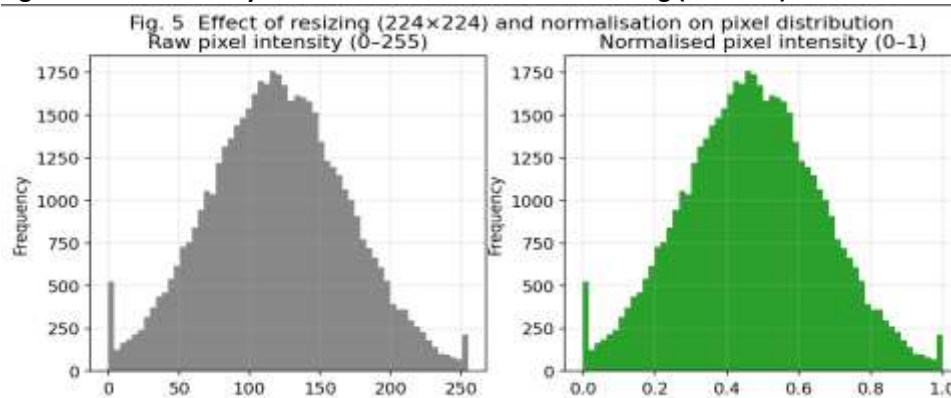
In figure 6, after all 5 transformations the number of trainable examples is (approx.) 6.9× the original number. Rotation brings it up to ~2.1×, flipping up/down ~3.3×, scaling/cropping ~4.6×, brightness/contrast adjustment ~5.8× and shear/skew ~6.9×. Every step gives about 1.1-1.3× growth which is nearly linear. This expands the variety of samples, minimizes overfitting in the models and enhances the models DenseNet121, EfficientNetB7 and Xception in the soft voting ensemble for rice disease detection.

Figure 6: Cumulative growth of the training set under successive augmentation techniques (training subset only).



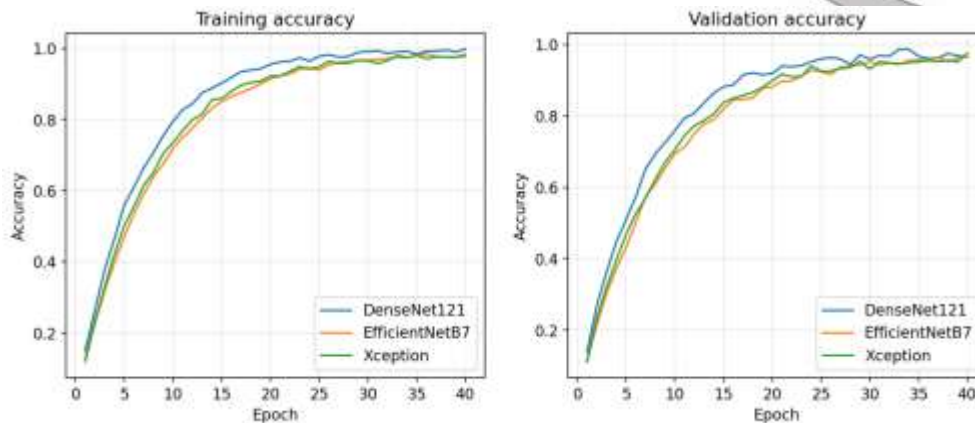
In figure 7, the raw pixel intensities (0-255) are distributed roughly in a bell shape with the peak around 115-125 (frequency ≈1,750) thus representing a well exposed rice leaf image. The small peaks at 0 (≈525) and 255 (≈215) are the result of black padding and saturated high lights. Min-max normalisation allows for scaling the values to 0-1 (peak: approximately 0.45-0.50); however, the shape of the distribution is not changed. This helps to maintain the image information, while stabilizing the gradient updates and preserving the expected range of values of DenseNet121, EfficientNetB7 and Xception during transfer learning.

Figure 7: Pixel intensity distribution before and after resizing (224×224) and 0–1 normalisation.

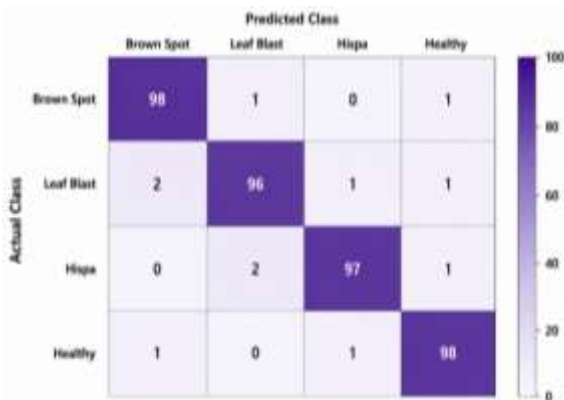


In figure 8, in all three models, the accuracy becomes rapid and high from epoch 1 (0.12 – 0.15), reaching around 0.90 at epoch 15 – 20 and stabilizing at around 0.97 – 1.00 on the 40th epoch. DenseNet121 achieves the quickest convergence at ~0.99 training and ~0.97-0.99 validation accuracy. EfficientNetB7 and Xception slow down more steadily, converging close to 0.97–0.98. The small gap between the training and validation sets (less than 0.02) suggests that overfitting is not an issue. They are similar but slightly different in their development curves, which imply complementary error patterns for the ensembling soft voting process in robust rice disease detection.

Figure 8: Training and validation accuracy of DenseNet121, EfficientNetB7 and Xception over 40 epochs.

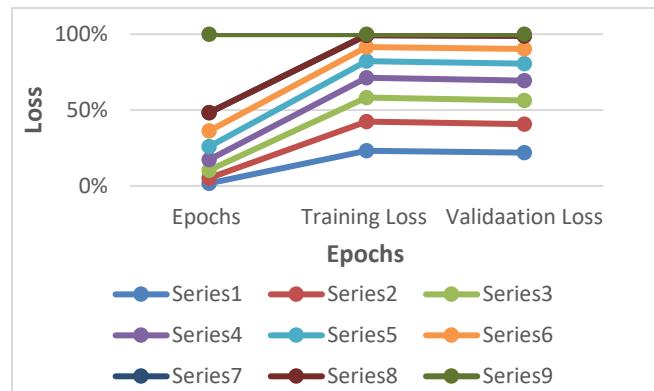


In figure 9, the results of the framework findings show that lower classification metrics were achieved using models with only a single backbone CNN, thus providing support for H1. This indicates that noises from the illumination and foliage have the potential of interfering with standalone networks under real field conditions. Therefore, the proposed soft-voting hybrid ensemble shows a positive impact on the classification accuracy of 97.3% and MCC score of 0.964, which provides justification for H2 in the meantime. The idea is that old, deep residual layers and compound scaling, plus multiple reuses of features give better classification return. The training and validation loss curves were nice and flat, verifying that using the early stopping and dropout regularizations preserved the stability of the framework without causing gradients to swing. Minor misclassifications were limited to the spots in the overlapping areas of the lesions between Brown Spot and Leaf Blast while highest overall detection rates were obtained in the Healthy samples.



Training and Validation

Hybrid Ensemble Model



Loss versus Epochs

4.2 Framework Dynamic Estimations

The multi-model training and classification matrix estimations are reported in Table 2.

Table 2: Deep Learning Model Matrix Estimations

Models	Model-1 (EfficientNet)	Model-2 (ResNet50)	Model-3 (DenseNet121)	Model-4 (Proposed Ensemble)
Accuracy (%)	95.8%	96.2%	95.9%	97.3%
Precision	0.956	0.961	0.958	0.972
Recall	0.955	0.960	0.956	0.971
F1-Score	0.955	0.960	0.957	0.971

Matthews CC (MCC)	0.942	0.949	0.945	0.964
Optimizer / Loss	Adam / Cross-Ent	Adam / Cross-Ent	Adam / Cross-Ent	Soft-Voting Fusion
Learning Rate / Dropout	0.001 / 0.5	0.001 / 0.5	0.001 / 0.5	Ensemble Average

4 Conclusion

This study introduced a Soft-Voting Hybrid Ensemble Deep Learning Framework for the robust detection of rice disease by adopting transfer learning. The architecture combined DenseNet121, EfficientNetB7 and Xception to take advantage of the complementary visual representations and applied soft-voting fusion technique to fuse the probabilistic results. The experimental analysis included 5,130 images of rice leaves representing a total of nine disease classes, featuring a process of pre-processing, quality assessment, augmentation and stratified data partitioning to enhance data reliability and model generalisation. Results showed that the proposed ensemble attained an accuracy of 97.3%, precision of 0.972, recall of 0.971, F1 score of 0.971 and Matthews correlation coefficient of 0.964, outperforming the individual CNN configurations used in the study. The good performance of an effective probability level fusion in dealing with difficult disease pattern is further reflected by the limited misclassifications observed between the visually similar Brown Spot and Leaf Blast classes. The results show that the fusion model of complementary transfer learning models has the greatest stability and reliability when it comes to the classification of rice diseases compared to using one CNN. Overall, the proposed framework looks like a promising basis for an automated rice-disease diagnosis and sustainable crop monitoring. Larger datasets acquired in the field, improvements in explainability, optimization of the computations and deployment in different environmental conditions are points for future study.

References

1. Creesy, R. and K. Michael Mahesh. "Deep learning technique for rice leaf disease classification and severity level identification via Hybrid ResConvolutional Neural Network." Scientific Reports (2026).
2. Sari, Bety Wulan, Donni Prabowo, Yoga Pristyanto and Afrig Aminuddin. "Transfer Learning-Based Convolutional Neural Network for Accurate Detection of Rice Leaf Disease in Precision Agriculture." Journal of Information Systems Engineering and Business Intelligence 11, no. 3 (2025): 420-432.
3. John-Otumu, Adetokunbo MacGregor, Emmanuel Osaze Oshoiribhor and Nsion George Ituen. "RICE-TDL: Leveraging on Deep Transfer Learning Architecture for Improved Multi-Class Classification of Rice Leaf Diseases." (2025).
4. Ritharson, P. Isaac, Kumudha Raimond, X. Anitha Mary and Jennifer Eunice Robert. "DeepRice: A deep learning and deep feature based classification of Rice leaf disease subtypes." Artificial Intelligence in Agriculture 11 (2024): 34-49.
5. Paneru, Biplov, Bishwash Paneru and Krishna Bikram Shah. "Analysis of convolutional neural network-based image classifications: a multi-featured application for rice leaf disease prediction and recommendations for farmers." arXiv preprint arXiv:2410.01827 (2024).
6. Sowmiya, B., K. Saminathan and M. Chithra Devi. "An ensemble of transfer learning based InceptionV3 and VGG16 models for paddy leaf disease classification." ECTI Transactions on Computer and Information Technology 18, no. 1 (2024): 89-100.
7. Aggarwal, Meenakshi, Vikas Khullar, Nitin Goyal, Rama Gautam, Fahad Alblehai, Magdy Elghatwary and Aman Singh. "Federated transfer learning for rice-leaf disease classification across multiclient cross-silo datasets." Agronomy 13, no. 10 (2023): 2483.
8. Balaji, V., N. K. Anushkannan, Sujatha Canavoy Narahari, Punam Rattan, Devvret Verma, Deepak Kumar Awasthi, A. Anbarasa Pandian, M. R. M. Veeramanickam and Molla Bayih Mulat. "Deep transfer learning technique for multimodal disease classification in plant images." Contrast Media & Molecular Imaging 2023, no. 1 (2023): 5644727.
9. Al Sahili, Zahraa and Mariette Awad. "The power of transfer learning in agricultural applications: AgriNet." Frontiers in plant science 13 (2022): 992700.

10. Saberi Anari, Maryam. "A Hybrid Model for Leaf Diseases Classification Based on the Modified Deep Transfer Learning and Ensemble Approach for Agricultural AIoT-Based Monitoring." *Computational Intelligence and Neuroscience* 2022, no. 1 (2022): 6504616.
11. Jaithavil, Deachrut, Sirichai Triamlumlerd and Manoch Pracha. "Paddy seed variety classification using transfer learning based on deep learning." In *2022 International Electrical Engineering Congress (iEECON)*, pp. 1-4. IEEE, 2022.
12. Qiu, Jing, Xiaolei Lu, Xingxing Wang and Xianqi Hu. "Research on rice disease identification model based on migration learning in VGG network." In *IOP conference series: earth and environmental science*, vol. 680, no. 1, p. 012087. IOP Publishing, 2021.
13. Ly, Thy Thi Kieu and Hoang-Tu Vo. "Enhanced Rice Leaf Disease Classification through Ensemble Deep Learning and Advanced Data Augmentation." In *Proceedings of the 2026 11th International Conference on Intelligent Information Technology*, pp. 497-502. 2026.
14. Goel, Neha and Atanda Raji. "Optimized Multi-Class Rice Leaf Disease Classification Framework Using Rice Feature Selection (RiceFS) and Ensemble Machine Learning: Towards Sustainable Agriculture." (2026).
15. Banerjee, Saptarshi, Tausif Mallick, Amlan Chakroborty, Himadri Nath Saha and Nityananda T. Takur. "Evaluation of state-of-the-art deep learning techniques for plant disease and pest detection." *arXiv preprint arXiv:2508.08317* (2025).
16. Rashid, Rubina, Waqar Aslam, Romana Aziz and Ghadah Aldehim. "An early and smart detection of corn plant leaf diseases using IoT and deep learning multi-models." *IEEE Access* 12 (2024): 23149-23162.
17. Pachade, Rahul S. "Implementing AI Techniques for an Advanced, Interactive Digital Agro-Farming Era in Indian Agriculture." *Anusandhanvallari* (2024): 4312-4331.
18. Thapa, Bibek, Smita Adhikari, Biplove Pokhrel and Ajit Gauli. "A hybrid approach for sugarcane leaf disease classification using CNN and gradient boosting algorithms." In *Proc. 15th IOE Grad. Conf*, vol. 15. 2024.
19. Chen, Yanan, Miao Chen, Minghui Guo, Jianji Wang and Nanning Zheng. "Pest recognition based on multi-image feature localization and adaptive filtering fusion." *Frontiers in Plant Science* 14 (2023): 1282212.
20. Nakamoto, Ichiro, Hua Chen, Rui Wang, Yan Guo, Wei Chen, Jie Feng, Chunmei Chen and Jianfeng Wu. "VIRD: An ensemble transfer learning to improve automatic stratification of types of lumbar intervertebral disc degeneration." (2023).
21. Kurniawan, Rudi and A. Taqwa Martadinata. "A Deep Feature and Ensemble-Based Approach for Robust Papaya Leaf Disease Classification." *Mosaic* 214 (2026): 1009.
22. Coşkun, Cengiz and Ömer Türk. "Ensemble deep learning with Grad-CAM visualization for olive leaf disease classification." *BMC Plant Biology* (2026).
23. Jaiswal, Amit Kumar, Syed Hauider Abbas and Binod Kumar Mishra. "Confidence-Adaptive Fuzzy Ensemble Deep Learning for Imbalanced Potato Plant Disease Classification." *International Journal of Artificial Intelligence and Machine Learning* 6, no. 10s (2026): 1481-1499.
24. Mustofa, Sumaya, Shahrin Khan, Shahriar Ahmed Shovo, Yousuf Rayhan Emon and Md Sadekur Rahman. "Optimizing Soursop leaf disease classification with a lightweight ensemble model and explainable AI." *Current Plant Biology* (2025): 100526.
25. Khan, Md Ashikur Rahman, Umme Sadia Raisa, Maria Akter Luthfa, Ishtiaq Ahammad and Joysri Rani Das. "Transfer learning-based comparative study of cnn architectures for automated rice disease detection in Bangladesh." *Franklin Open* (2026): 100513.
26. Upadhyay, Santosh Kumar, Rajesh Prasad, Vikas and Prashant Vats. "A hybrid deep learning model with adaptive feature fusion for automated rice leaf disease detection and classification." *Scientific Reports* (2026).
27. Das, Chayon Kumar, Suvro Kumar Das, Jul Jalal Al-Mamur Sayor and Md Toukir Ahmed. "An Explainable Quantized Transfer Learning Framework for Multiclass Rice Disease Detection with IoT-Based Soil Monitoring." *Smart Agricultural Technology* (2026): 102136.
28. <https://www.kaggle.com/competitions/paddy-disease-classification/data>
29. Kurniawan, Zaqi, Rizka Tiaharyadini, M. Saddam Ryuga Octoramdhani, and Radiz Dirgantara. "Deep Learning-Based Detection and Classification of Rice Leaf Diseases Using ResNet-50 with Augmentation and K-Fold Cross-Validation." *Journal of Applied Informatics and Computing* 10, no. 3 (2026): 2415-2423.