Ensemble of deep convolutional neural networks for predicting blast disease severity in paddy

Girish Saunshi[1,2], Rajesh Yakkundimath[1,2*], Guruprasad Konnurmath[3]

(1. Department of Computer Science and Engineering, K. L. E. Institute of Technology, Hubballi 580027, Karnataka, India;
2. Visvesvaraya Technological University, Belagavi 590018, Karnataka, India;
3. School of Computer Science and Engineering, K L E Technological University, Hubballi 580031, Karnataka, India)

Abstract: Blast disease poses a significant threat to paddy crops worldwide, causing substantial yield losses under favorable environmental conditions. This study aims to develop an efficient ensemble model for predicting blast disease severity in paddy plants using convolutional neural network (CNN) models. Three popular CNN architectures, namely VGG16, ResNet50, and InceptionV3, are employed in this work. Initially, the ImageNet dataset is used to pre-train these models for an identification task and then fine-tuned on a dataset for blast disease severity classification based on the Percentage Disease Index (PDI) score. The final model is constructed as an ensemble of the three CNN networks, combining their outputs using a weighted averaging method. The efficiency of the suggested ensemble model is demonstrated by experimental data, achieving impressive training accuracy, validation accuracy, and testing accuracy of 96.09%, 94.44%, and 88%, respectively, using a dataset of 18,865 labeled images. The developed ensemble model of deep CNN architectures achieved good training and testing accuracies in classifying blast disease severity. These results would be helpful for early, automated detection, precision agriculture and yield protection. The findings show that the ensemble CNN model predicts blast disease severity in paddy crops with higher classification accuracy. The proposed approach can support early disease detection, precision agriculture, and yield protection. Future work may focus on real-time field deployment and integration of advanced imaging techniques.

Keywords: blast disease severity prediction, convolutional neural networks, VGG16, ResNet50, Inception-V3, transfer learning, ensemble learning, Percentage Disease Index (PDI), precision agriculture.

Citation: Saunshi, G., R. Yakkundimath, and G. Konnurmath. 2026. Ensemble of Deep Convolutional Neural Networks for Predicting Blast Disease Severity in Paddy. CIGR Journal, 28(2):302-316.

1 Introduction

Rice paddy (Oryza sativa L.) is a vital staple food that plays a crucial role in feeding over 50% of the world population, with Asia and Africa being the largest consuming regions (Straits Research, 2025). Cultivated over 162 million hectares worldwide, paddy production reaches an annual yield of 543 million tonnes (FAO, 2024). India, with a harvest area of about 40 million hectares and projected production of 150–178 million tonnes accounting for 27% of global rice production (Shahbandeh, 2024). Despite its significant production, the cultivation of rice paddy faces numerous challenges, including unsuitable environmental conditions and the presence of both biotic and abiotic stresses. These factors collectively contribute to substantial yield losses. Among the various challenges, diseases emerge as the

Received date: 2025-05-20 **Accepted date:** 2025-12-25

***Corresponding author: Rajesh Yakkundimath, Associate Professor.** Department of Computer Science and Engineering, K. L. E. Institute of Technology, Hubballi 580027, Karnataka, India. Tel: 9481126276. E-mail:rajeshymath@gmail.com.

primary limiting factor, causing significant reductions in yields (Divya et al., 2014). Therefore, early and accurate detection of blast disease severity is essential for effective disease management and yield protection.

Paddy is highly susceptible to bacterial, fungal, and viral infections, posing a significant threat to its productivity. These diseases can lead to substantial yield reductions of 20% to 70% (Sabri et al., 2023; Bag et al., 2023). The paddy blast disease caused by the filamentous ascomycete Magnaporthe oryzae is a serious issue in global paddy production due to its wide distribution and destructive nature, especially under favorable conditions. Global reports indicate that paddy blast alone can result in significant yield losses of up to 85% (Divya et al., 2014). Annually, paddy blast destroys a significant quantity of paddy that could have fed over 60 million people (Upadhyay and Bhatta, 2020) Minimizing these losses is crucial to support marginalized and impoverished farmers in developing countries. Cultivating paddy varieties resistant to biotic stresses is a practical and readily adoptable genetic approach to address these challenges and enhance productivity.

Farmers and agricultural experts have been adopting various strategies to manage paddy blast disease and minimize yield losses (Khadka et al., 2025). While fungicides are commonly used, their cost and environmental impact pose challenges. In response, a more eco-friendly approach utilizing computer technology has gained traction. The application of Artificial Intelligence and Machine Learning (AIML) has emerged as a highly effective and cost-efficient method for controlling paddy blast disease. This technology offers an economical solution, requiring no additional expenses for farmers, while providing effective disease management. By leveraging AIML, farmers can mitigate paddy blast disease and safeguard their yields in an environmentally sustainable manner (Pai et al., 2025).

The field of AIML has gained widespread adoption across numerous industries, including business, biological science, material science, and medical science, among others. In computer vision, AIML has emerged as a promising tool to address various challenges. In this study, our focus lies on the processing of images related to paddy blast disease. Traditionally, disease identification in paddy has relied on manual assessment by experts (Ning et al., 2023). However, the cost of consulting experts can be prohibitive for farmers, particularly when expert availability is limited (Parasa et al., 2023). Consequently, alternative approaches are being explored to reduce the use of harmful chemicals and preserve the environment. Early detection of blast disease can significantly help farmers prevent substantial losses (Padhi et al., 2025). Researchers are leveraging technology support, including image processing, machine learning, and artificial intelligence (Cruz et al., 2021). The following summary provides an overview of relevant literature papers related to various aspects of the present study. Bock et al. (2022) provided a comprehensive historical perspective on visual severity assessment in plant diseases, emphasizing key concepts, tools, and methods to enhance accuracy. Best-operating practices and future research directions are presented. Liu et al. (2023) proposed a novel hyperspectral imaging method for monitoring and predicting disease severity in avian paramyxoviruses infected apple leaves. The study established a strong correlation between LCC average, coefficient of variation, and disease severity, achieving 98.89% accuracy in identifying disease severity on the validation set. Mandal et al. (2023) employed hyperspectral remote sensing to assess paddy blast disease severity. Models like support vector machines, partial least squares, random forests, and multivariate adaptive regression splines are used in field tests comprising 20 rice genotypes. Support vector machine outperforms other models, achieving 94% accuracy in severity estimation. Palma et al. (2022) introduced an efficient image-based method for quantitatively assessing pathogenic disease severity. Non-linear system is employed to isolate disease patterns in leaf images, demonstrating potential for generalized symptom detection. Pujari et al. (2013) proposed statistical techniques for identifying and categorizing fungal diseases based on severity levels. Various statistical features are extracted from images of disease-affected fruits and achieved high accuracies using the nearest neighbor classifier. Pujari et al. (2016) developed quantitative detection methods for soybean rust at different disease stages. Color range and pixel relationships in rust affected leaf images is analyzed, categorized severity into grades, and calculated the Percentage Disease Index (PDI) over time. Shi et al. (2023) conducted a comprehensive review of CNN-based plant disease severity assessment, analyzing different CNN segmentation networks, architectures, and frameworks. Dataset acquisition methods, performance evaluation metrics, and challenges in practical applications is discussed, proposing research ideas and potential solutions. Sibiya and Sumbwanyambe (2019) proposed an algorithm using fuzzy logic for estimating plant leaf disease severities, with applications for precision agriculture. Fuzzy logic inference rules and image segmentation, distinguishing it from existing methods is integrated in this approach. Wang et al. (2017) diagnosed the severity of apple black rot disease using CNN. Training and fine tuning is done with shallow networks and deep models through transfer learning, finding that the VGG16 model achieves the highest accuracy of 90.4%. Yakkundimath and Saunshi (2023) focused on blast disease identification in paddy crops using transfer learning multi-layer CNN models. Field images are classified based on severity levels and reported the testing efficiencies of CapsNet, EfficientNet-B7, and ResNet50 models.

The existing literature predominantly focuses on the detection of plant disease severity, with a majority of studies concentrating on generic plant disease detection rather than specifically addressing the diagnosis of plant disease severity based on PDI (Percent Disease Index) scores. Consequently, there is a notable gap in comprehensive studies that specifically address these aspects in the context of paddy blast disease, particularly during the early stages of infection. The present study attempts to fill this gap and provides insights into the motivation behind the conducted research. The remaining portion of the paper is structured as follows; Section 2 discusses the methodologies employed for detecting the presence of rice blast disease. Section 3 focuses on the paddy blast disease severity classification based on PDI scores. Finally, the conclusion of the work undertaken and outlines potential avenues for future research is presented in Section 4, including opportunities for improving the developed methodologies.

2 Materials and method

The proposed approach is divided into two phases. Dataset preparation of paddy leaf images affected by blast disease in the first phase and the utilization of a CNN for classification based on PDI scores in the second stage. Figure 1 illustrates the phases involved in automating the identification of paddy blast disease.

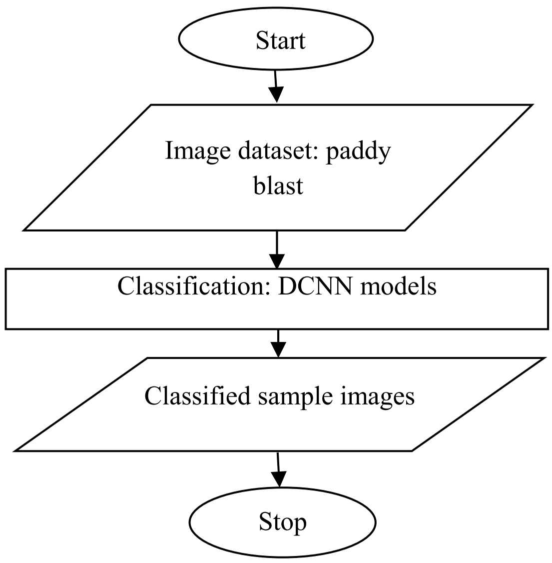

Figure 1 Schematic flow diagram of the proposed method

2.1 Image dataset preparation

The Agricultural Research Station (A.R.S.) Mugad, University of Agricultural Sciences, Dharwad, Karnataka, India, provided the dataset for this study during the kharif season. Mugad has been identified as a paddy blast disease hotspot due to favorable microclimate conditions in the irrigated rice belt of Dharwad. Field images of paddy crops from various paddy varieties are captured through a roving survey (Yakkundimath and Saunshi, 2023).

A 0–9 rating system based on the paddy standard evaluation system (SES) developed by the International Rice Research Institute (IRRI) in Manila, Philippines, is used to determine the severity of blast disease in ten randomly chosen rice plants in the paddy fields. Table 1 provides a representation of the severity scores used (Saha, 2022).

Table 1 Scoring for blast

Score	Affected leaf area
0	Spots observed/ No lesions
1	Pin point size brown speaks
2	About 1 -2 mm in diameter small roundish to slightly elongated
3	Appears on upper leaves as scale 2
4	< 4%
5	4% to 10%
6	11% to 25%
7	26% to 50%
8	51% to 75%
9	> 75%

In this study, Nikon D3300 digital SLR camera with a resolution of 24 megapixels is used to capture images of blast affected leaf images. Each image has a resolution of 4624×3472 pixels. To ensure consistency, the camera is positioned at a 1.0m of focal distance from the paddy crop varieties. A tripod stand is used to stabilize the camera and facilitate accessibly. The camera lens is set at a 45^0 angle with respect to the object's axis. The images are captured during near solar noon in the paddy fields, utilizing natural lighting conditions.

The dataset comprises a total of 2,666 images, all of which are labeled by experts to indicate the presence of blast disease. In this study, blast disease severity scores ranging from 0 to 4 are considered for experimentation. It is within this range that the detection of blast disease becomes challenging. The severity scores are categorized as score 0, score 1, score 2, score 3, and score 4. Figure 2 provides an overview of the blast affected paddy leaf images based on their corresponding disease severity scores.

(a) Score 0 (b) Score 1 (c) Score 2

(d) Score 3 (e) Score 4

Figure 2 Sample leaf images of rice blast severity

To enhance the size of the initial image dataset for processing CNN models, several classical image augmentation techniques are utilized. These techniques include horizontal flip and vertical flip, shear transform and random rescaling. By applying these augmentation techniques, the original dataset of 2,666 images is expanded to a larger dataset consisting of 18,865 labeled images. The augmented images display noticeable variations from the original images, thereby enhancing the diversity and richness of the dataset.

2.2 Architecture and configuration of CNN models

In this study, three widely recognized architectures of CNN are utilized for the categorization of blast severity based on severity scores. The selected architectures are VGG16 (Simonyan and Zisserman, 2015), InceptionV3 (Szegedy et al., 2014), and ResNet50 (He et al., 2016). These architectures are pre-trained initially on the ImageNet dataset (Krizhevsky et al., 2017), which is a large-scale dataset used for general image classification tasks. Subsequently, the pre-trained models on the blast dataset are fine-tuned to adapt them specifically for the classification of blast severity scores. This process allows the models to learn and specialize in detecting patterns and features relevant to the specific problem of blast severity classification.

At the first-level, ResNet50, VGG16, and InceptionV3 CNN models are trained independently on the dataset. Relevant features are fetched from each model to make identification based on the input images. In the second level, an ensemble model is constructed by combining the outputs of these three CNN models. This is achieved using a weighted averaging method, where the predictions of each individual model are combined based on predefined weights. The ensemble model leverages the collective insights of the individual models to make a final prediction, aiming to increase the overall accuracy of the classification task. By employing an ensemble approach, the proposed model takes leverage of the diverse strengths and capabilities of the individual CNN models, providing more robust and accurate classification system for blast severity assessment.

The CNN models used in this study are designed to recognize patterns and features in the input images and make corresponding predictions. Each CNN model consists of several layers including an input, convolutional, rectified linear unit (ReLU), max-pooling, and a fully connected output layers. To optimize the performance of the models, transfer learning is employed. This involves initializing the CNN models with pre-trained weights obtained from the ImageNet database. By doing so, the models are able to capture generic image features and accelerate the training process. After initialization, the network is further fine-tuned using the stochastic gradient descent (SGD) algorithm with the Adam optimizer. This optimization process aims to minimize the loss function and improve the accuracy of the predictions. At the second level, a weighted averaging method is employed to combine the outputs of the individual CNN models from the first-level. The weights assigned to each model's output are carefully chosen to enhance the overall performance of the final prediction. By leveraging the strengths of each individual model, the ensemble approach improves the robustness and accuracy of the predictions. Figure 3 illustrates the ensemble architecture of CNN models deployed for prediction of blast disease severity scores.

The augmented images are fed into each CNN model, and both the base layers and upper layers of the pre-trained models are utilized to extract features. This allows the models to learn meaningful representations of the input images at different levels of abstraction. Using the learned features, the models in the first-level, with their specific architectures and hyper-parameters, independently classify the images. The outputs of these models serve as predictions for each image. To obtain a consolidated prediction, the predictions from the first-level models are combined using a weighted averaging approach. This ensemble technique assigns different weights to each model's prediction and computes a final prediction that aims to achieve higher accuracy. To estimate the performance of the classification on the test dataset, a consolidated confusion matrix is computed. This matrix provides an overview of the predictions made by the model including true positives, true negatives, false positives, and false negatives. Furthermore, the efficacy of the test dataset and the cross-entropy loss are used to assess the performance and effectiveness of the CNN models. These metrics provide insights into the accuracy and efficiency of the models in their classification task. By employing these evaluation measures, the performance and reliability of the CNN models can be analyzed, aiding in the assessment of their classification capabilities.

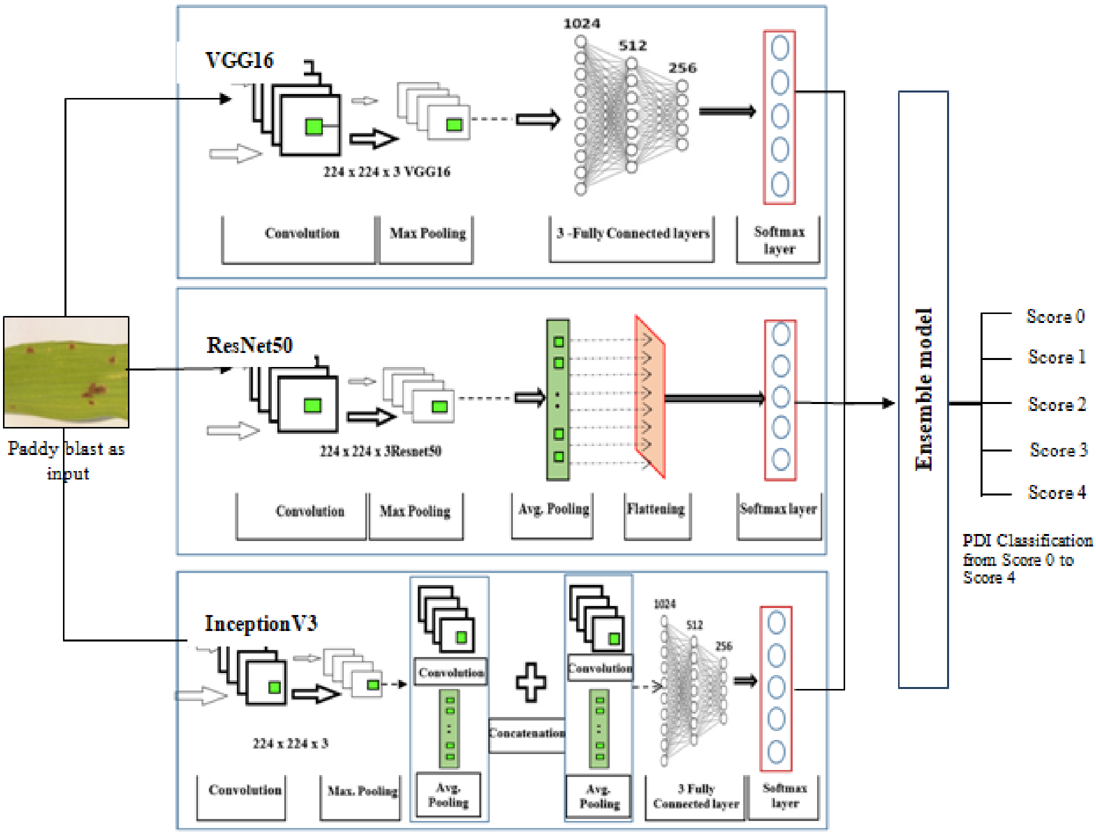

Figure 3 Schematic architecture of the proposed method

The models are implemented using the Keras deep learning framework, which offers a user-friendly interface for building and training neural networks. The experiments are executed on an Ubuntu Linux server equipped with a GTX 1070 GPU (8 GB memory) and a 3.40 GHz i7-3770 CPU (16 GB memory).

The augmented image dataset, consisting of 18,865 blast disease severity images is divided into three sets such as training, validation and testing. The training set contains 15,400 images, the validation set contains 3,080 images, and the testing set contains 385 images. This division allows for evaluating the efficiency of the trained models on unseen data with the testing set representing approximately 10% of the training dataset. The identification efficiency of the pre-trained VGG16, ResNet50, InceptionV3, and ensemble models is computed using Expressions 1 and 2, which measure the accuracy and efficiency of the classification process. The CNN identification methodology based on transfer learning is described in Algorithm 1. This algorithm outlines the steps involved in training and evaluating the models, including data preparation, model construction, training, and performance assessment.

$$\text{Efficiency of classification (\%)} = \frac{\text{Sample images correctly classified}}{\text{Sample images in total for testing}} \times 100 \quad (1)$$

$$\text{Efficiency of average classification(\%)} = \frac{\text{Correctly classified sum of sample images}}{\text{Total sample images}} \times 100 \quad (2)$$

2.3 Algorithm 1: CNN classification of blast disease severity scores

Description: Let x^i represents the i^{th} input image from the dataset, where i ranges from 1 to N, representing the total number of images in the dataset. The outputs from all the models are denoted by y^i, which is a vector representing the predictions for M=5 labels, indicating the total number of labels to predict. The outputs for the i^{th} image from the first-level models, namely VGG16, ResNet50, and InceptionV3 CNN models form a vector, which combines the outputs of all the first-level models for a given image x^i. The task of the first-level models is to learn a function $f(x^i) \rightarrow y$, where x^i is the input image and y^i is the output prediction and thus output of first-level models is $y^i = f(x^i)$. The second-level ensemble model learn a function $g(y^i) \rightarrow y$, where y^i is the input consisting of the outputs from the first-level models for the i^{th} image. The second-level model generates outputs representing the predictions from the second-level models based on the combined outputs of the first-level models.

Parameters:

ε: epochs, ξ: Iteration step, β: batch size; η: CNN learning rate=1e-3; w: ImageNet weights, n: number of training examples for every iteration, N: number of images in dataset, D=dataset, P:momentum= 0.9, learning rate decay= 1e-5, batch size=30.

// Input: Training dataset D= x^i, where i= {1, 2, 3 …, N}

First-level models: C_1, C_2, C_3

Second-level model: ĉ

// Output: Trained first level and second-level models.

Step 1. Select sample images randomly from the labeled datasets for training and validation

Step 2. Configure input, hidden, and output layers of CNN models

Step 3. Configure convolution, pooling, flatten, and dropout layers of CNN models.

Step 4. Configure parameters:η,ε,β.

Step 5: Set network weights (w_1, w2…, w_n) and the pre-trained CNN models.

Step 6: Configure the network parameters and determine weights.

Step 7: Train the first-level models and calculate the initial weights

for ξ =1 to ε do

Randomly select mini-batch from (size: β) from training data sets

Forward propagation and calculate loss (E) using Expression 3.

$$E(W) = -\frac{1}{n}\sum_{xi=1}^{n} \sum_{k=1}^{k}[y_{ik}\log P(x_i = k) + (1 - y_{ik})\log(1 - P(x_i = k))] \quad (3)$$

Back-propagation and adjust weights with SGD using Expression 4.

$$W = W_{k-1} - \eta(\partial E(W)/\partial W) \quad (4)$$

Update weights with adam optimizer

end

Step 8: Store computed weights in the database.

Step 9: Train the ensemble classifier using step 7.

Step 10: Applying weighted averaging method using Expression 5.

Averaging = w1 * output_VGG16 + w2 * output_ResNet50 + w3 * output_InceptionV3 (5)

where, w1, w2, w3 are the weights allocated to each model's output

Step 11: Calculate CNN models' accuracy using test dataset.

Step 12: Calculate CNN models' assessment metrics using the test dataset.

Stop.

3 Results and discussion

Experiments are conducted using VGG16, ResNet50, InceptionV3, and ensemble CNN models. The models performance is improved to ensure adaptability with the image dataset. The efficacy of each CNN model is assessed through independent training and testing.

The VGG16 model's training and validation efficiency curves are shown in Figure 4. The curve indicates that the model achieved the highest training efficiency of 95.93% and the highest validation efficiency of 84.56% after 30 epochs.

The ResNet50 model's training and validation efficiency curves are shown in Figure 5. The curve demonstrates that the model attained the highest training efficiency of 93.09% and the highest validation efficiency of 88.09% after 30 epochs.

The InceptionV3 model's training and validation efficiency curves are shown in Figure 6. The curve indicates that the model achieved the highest training efficiency of 93.72% and the highest validation efficiency of 86.11% after 30 epochs.

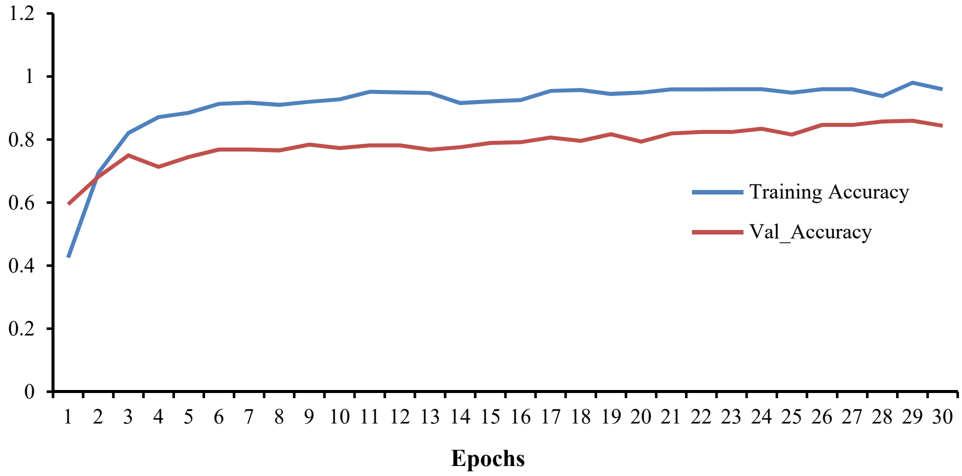

Figure 4 Training and validation efficiency curve for VGG16 CNN model

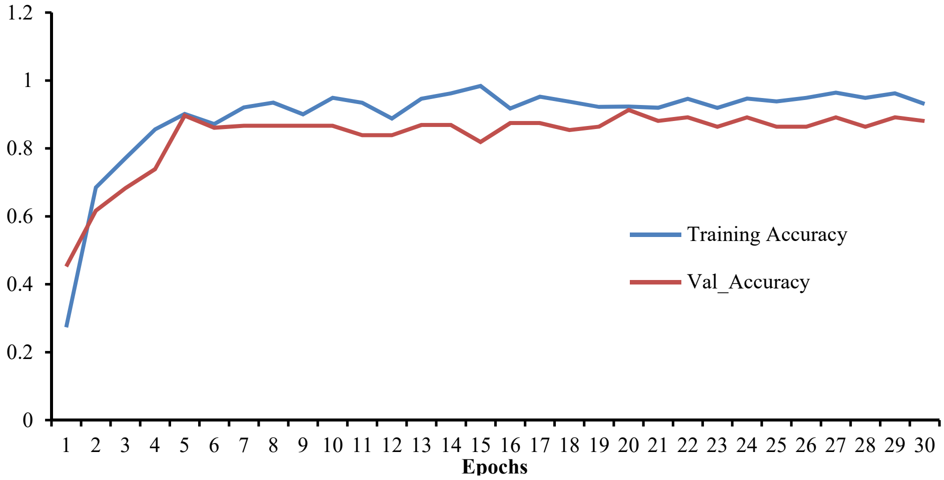

Figure 5 Training and validation efficiency curve for Resnet50 CNN model

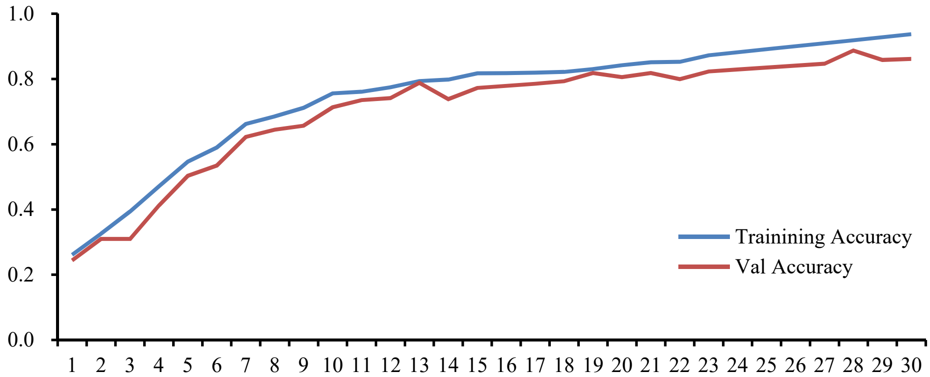

Figure 6 Training and validation efficiency curve for InceptionV3 CNN model

The Figure 7 shows the plot of the ensemble model's training and validation efficiency curves. The ensemble model achieved the highest training efficiency of 96.09% and the highest validation efficiency of 94.44% for 30 epochs.

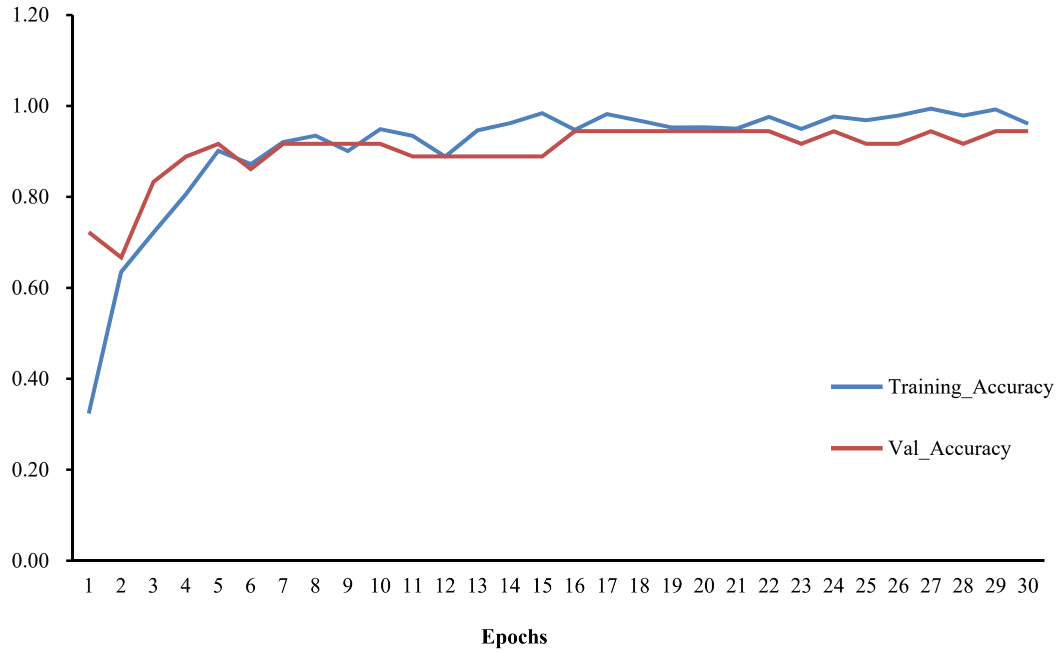

Figure 7 Training and validation efficiency curve for ensemble CNN model

The Mean Squared Error (MSE) and Nash–Sutcliffe Efficiency (NSE) metrics were used to assess the performance of the CNN models. The MSE (Eq. 6) measures the mean squared difference between the predicted (Pi) and observed (Oi) values. Models with lower MSE values show better performance. The predictive skill of the model is evaluated by the NSE (Equation 7) through a comparison of residual variance and variance of the observed data. An NSE value that is near to 1 indicates that the predicted and observed values are in agreement with each other, while negative values shows that the predictive performance is poor. Table 2 presents the comparative analysis of various CNN architectures—VGG16, ResNet50, InceptionV3, and the Ensemble model— showed significant differences in the generalization performance of the models. The Ensemble CNN presented the highest validation efficiency of 94.44%, the lowest MSE of 2.72, highest R^2 value of 0.96 and the highest NSE of 0.95 among all models, and possessing a strong generalization capability. It also showed consistency between training and validation accuracies. In contrast, the VGG16 model showed a larger gap between training and validation efficiencies of 11.56%, resulting in a high MSE value of 133.63, negative NSE value of –1.31, and comparatively lower R^2 value of 0.68, indicating overfitting and poor generalization performance. ResNet50 showed moderate performance with MSE of 25.30 , R^2 value of 0.84 and NSE of 0.56, followed by InceptionV3 with MSE of 57.91, R^2 value of 0.79 and NSE approximately equals to 0.00, showing consistency but enhanced stability in comparison to VGG16. In general, the Ensemble model outperformed the individual CNNs, demonstrating superior predictive accuracy and reliability for the classification of paddy leaf disease severity.

$$MSE = \frac{1}{n} \sum (P_i - O_i)^2 \quad (6)$$

Where,

$P_i =$ training efficiencies (predicted);

$O_i =$ validation efficiencies (observed).

$$NSE = 1 - \frac{\sum (O_i - P_i)^2}{\sum (O_i - \bar{O})^2} \quad (7)$$

Where,

$\bar{O}$ is mean of observed data.

Table 2 Training and validation performance results from the different CNN models

Sl No	CNN Model	Image Size in Pixels	Training (Time: hh:mm:ss)	Training Efficiency (%)	Validation Efficiency (%)	R^2	MSE (per model)	NSE
1	VGG16	244 x 244	00:56:45	95.93	84.37	0.68	133.63	-1.31
2	ResNet50	244 x 244	01:16:23	93.09	88.06	0.84	25.30	0.56
3	InceptionV3	244 x 244	00:52:37	93.72	86.11	0.79	57.91	~0.00
4	Ensemble	244 x 244	01:07:39	96.09	94.44	0.96	2.72	+0.95

The obtained results are consistent compared to earlier studies on plant disease severity classification using deep learning methods. Wang et al. (2017) reported an accuracy of 90.4% using the VGG16 model for apple disease severity estimation, while the proposed Ensemble CNN achieved a validation efficiency of 94.44% for paddy blast severity classification. Mandal et al. (2023) reported nearly 94% accuracy for rice blast severity estimation using hyperspectral imaging and machine learning techniques. In comparison, the present study achieved similar performance using RGB images and ensemble CNN architectures. The higher NSE value of 0.95, lower MSE value of 2.72, and improved R^2 value of 0.96 show that the ensemble approach provided better prediction consistency and generalization performance compared to the individual CNN models and previously reported methods.

The test image dataset is employed to assess the performance of the CNN models with confusion matrices plotted to visualize the classification performance. These matrices provide insights into each model's accuracy by identifying the number of correct and incorrect predictions for every class. Figure 8 illustrates the confusion matrix for the VGG16 model, Figure 9 for the ResNet50 model, Figure 10 for the InceptionV3 model, and Figure 11 for the ensemble model. By analyzing these matrices, the models' abilities to correctly and incorrectly classify each class are evaluated. This analysis provides a clearer understanding of the strengths and limitations of each model in accurately identifying various classes within the test dataset used in the present study.

Table 3 presents the comparative performance metrics for the VGG16, ResNet50, InceptionV3, and Ensemble model. By examining these performance metrics, accuracy, completeness, and balance of the classification results obtained by each model are assessed. Higher values for precision, recall, and F1-score indicate better overall performance in terms of correctly classifying different classes.

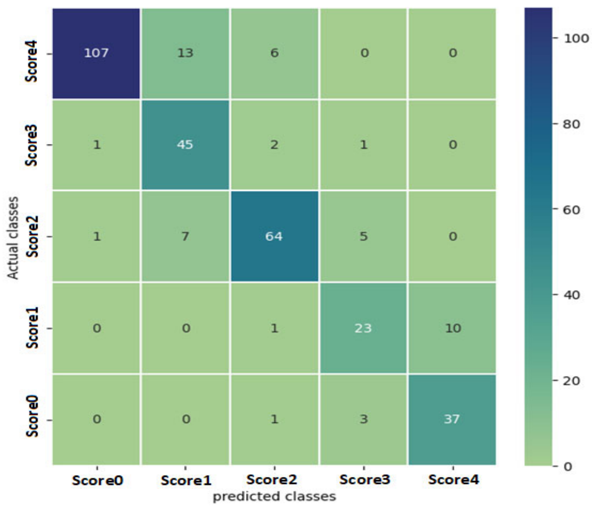

Figure 8 Confusion matrix for blast disease severity identification using the VGG 16 CNN model

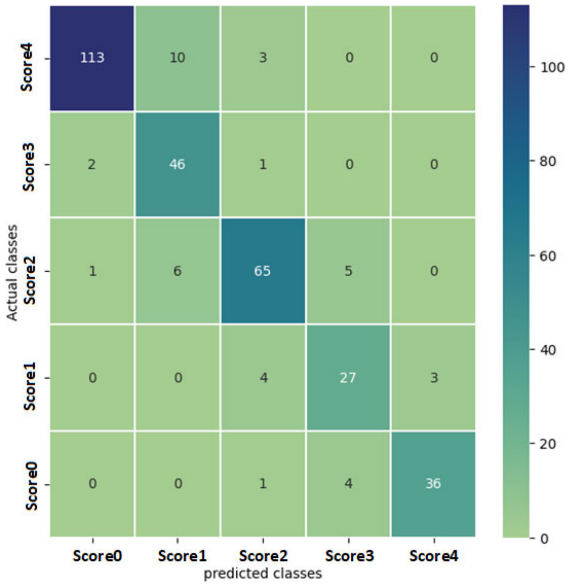

Figure 9 Confusion matrix for blast disease severity identification using the ResNet50 CNN model

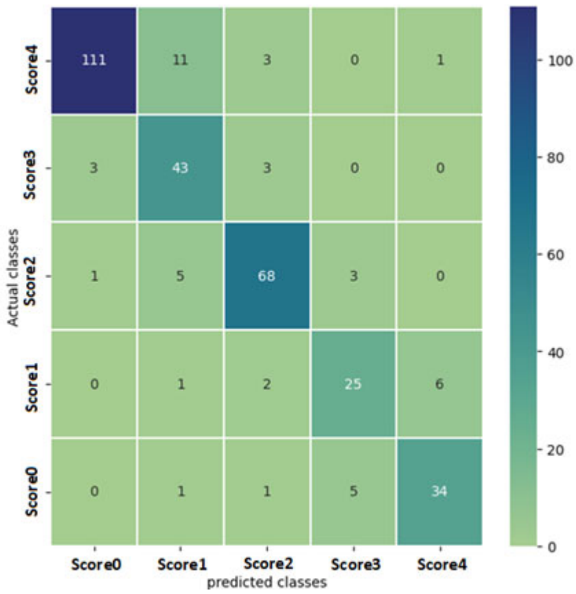

Figure 10 Confusion matrix for blast disease severity identification using the InceptionV3 CNN model

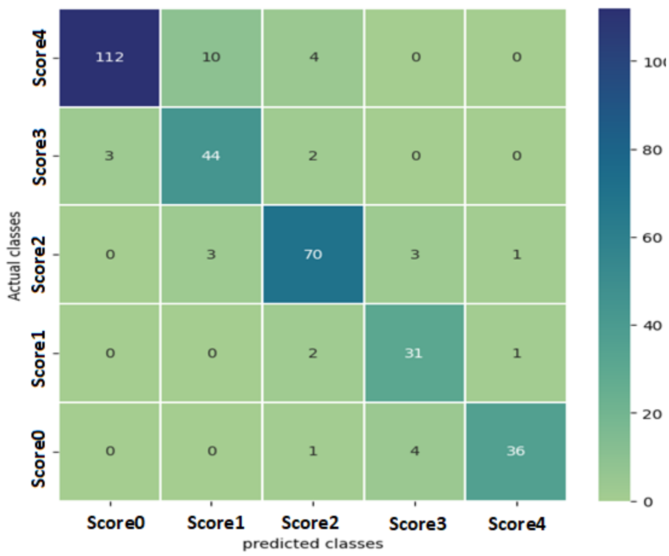

Figure 11 Confusion matrix for blast disease severity identification using the Ensemble CNN model

Table 3 Comparative performance metrics for blast disease severity identification using VGG16, ResNet50, InceptionV3, and Ensemble CNN models

Severity Score	VGG16			Resnet 50			Inception V3			Ensemble		
	Precision	Recall	F1 Score	Precision	Recall	F1 Score	Precision	Recall	F1 Score	Precision	Recall	F1 Score
Score 0	98	85	91	97	90	93	97	88	92	97	89	93
Score 1	69	92	79	74	94	83	70	88	78	77	90	83
Score 2	86	92	79	88	84	86	88	88	88	89	91	90
Score 3	72	68	70	75	79	77	76	74	75	82	91	86
Score 4	79	90	84	92	88	90	83	83	83	95	88	91
Avg (%)	81	84	84	85	87	86	83	84	83	88	90	89

In Figure 12, the graph showcases the classification efficiencies for different disease severity scores using the VGG16, ResNet50, InceptionV3, and ensemble CNN models. Each model's performance is evaluated based on its ability to accurately classify the severity levels of paddy blast disease. For the VGG16 model, it is observed that the highest classification efficiency of 91% is achieved for disease severity score 0, indicating accurate identification of this severity level. The lowest classification efficiency of 70% is obtained for disease severity score 3, indicating some misclassifications in identifying this severity score. Similarly, for the ResNet50 model, the highest classification efficiency of 93% is achieved for disease severity score 0, while the lowest classification efficiency of 77% is obtained for disease severity score 3. For the InceptionV3 model, the highest classification efficiency of 92% is obtained for disease severity score 0, and the lowest classification efficiency of 75% is observed for disease severity score 3. Lastly, for the ensemble model, the highest classification efficiency of 93% is achieved for disease severity score 0, and the lowest classification efficiency of 83% is obtained for disease severity score 1. These results indicate variations in the performance of the different models for different disease severity scores, with some models demonstrating higher accuracy in classifying certain severity levels compared to others. The present study primarily focused on RGB imagery for disease severity classification without incorporating spectral or thermal information. The inclusion of such multimodal data could potentially improve early disease detection and reduce misclassification among visually similar symptoms. Therefore, future research should focus on integrating spectral and thermal imaging modalities to improve the model's discriminative capability and robustness under diverse field conditions.

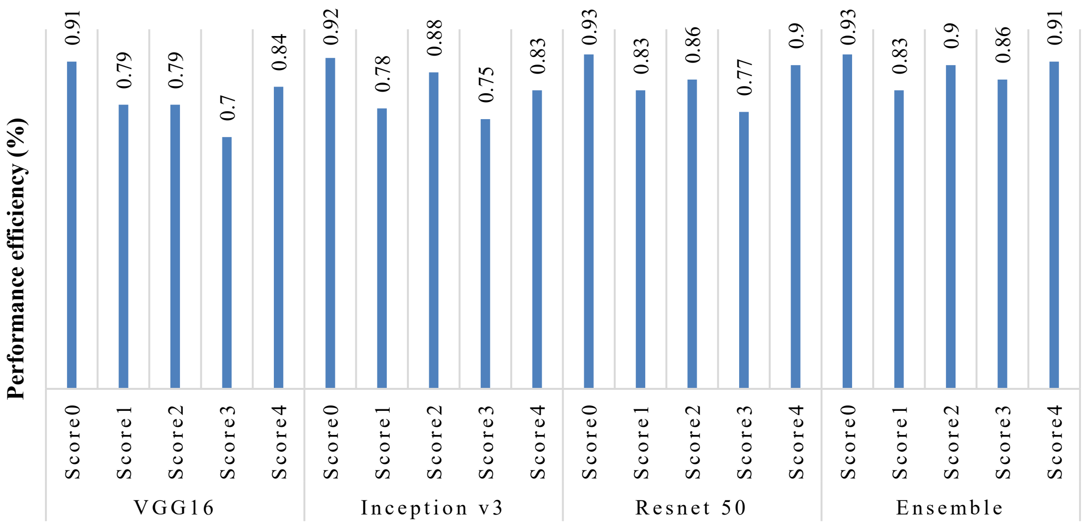

Figure 12 Blast disease severity identification efficiencies using the CNN models.

The study has some limitations that should be considered while interpreting the results. The CNN models were trained and tested using images collected from a specific region and under controlled field conditions. This may affect their performance under different climatic conditions, paddy varieties, and field environments. The study used RGB images for blast severity classification and did not include hyperspectral, multispectral, or thermal imaging data. In addition, image augmentation techniques may have introduced similarities among samples due to repeated transformations of the original images. Moreover, the experiments were carried out mainly under offline conditions, and extensive real-time field testing under varying lighting, occlusion, and complex backgrounds was not performed. Despite these limitations, the proposed approach showed good potential for automated paddy blast disease severity classification and can support future work in multimodal and real-time precision agriculture systems.

4 Conclusions

This study developed models for predicting the severity of blast disease in paddy plants. By utilizing deep convolutional neural networks, specifically VGG16, ResNet50, InceptionV3, and an Ensemble model, which were pre-trained on the ImageNet dataset and fine-tuned on a blast disease dataset, accurate predictions of disease severity were achieved. The ensemble model, combining the outputs of these networks using fusion methods, demonstrated the highest performance on the hold-out dataset. These results highlight the potential of deep learning techniques in agricultural disease control. Accurate assessment of blast disease severity enables timely intervention and effective management strategies. The implementation of these models empowers farmers and agricultural professionals to make informed decisions, preventing or mitigating the spread of the disease and improving crop yield, thus ensuring food security. The success of this study emphasizes the importance of leveraging advanced technologies, such as deep learning, in modern agricultural practices. With further refinement and optimization, these CNN models can be extended to other plant diseases, enabling proactive disease management and reducing crop losses. Overall, the findings of this study demonstrate the potential of deep learning models in blast disease severity prediction, providing valuable insights and tools for disease management in paddy cultivation.

Based on the findings of this study, the future research efforts can focus on expanding the dataset to include more diverse environmental conditions and disease variations to enhance model generalization. The integration of multimodal imaging techniques, such as hyperspectral, thermal, and UAV-based imaging , can improve early-stage disease detection and reduce misclassification of similar symptoms. Additionally, optimizing lightweight deep learning architectures and deploying the developed model on embedded platforms, such as drones or mobile-based systems, can enable real-time monitoring and decision support for farmers.

Acknowledgements

We gratefully acknowledge the assistance and cooperation extended by Dr. Surendra Palaiah, Principal Scientist and Head, Department of Genetics and Plant Breeding, All India Coordinated Rice Improvement Project (AICRIP), Mugad and Sirsi, Karnataka State, India, and Dr. Satish, R. G., Assistant Professor, Department of Genetics and Plant Breeding, AICRIP, Mugad, Karnataka State, India, who kindly assisted us in gathering datasets and shared insightful agricultural knowledge for the present study.

Funding

This work did not receive any specific grant from funding agencies in the public, commercial, or nonprofit sectors.

Competing Interests

The corresponding author certifies that none of the other authors have any competing interests.

Availability of Data and Materials

Dataset associated with this article are available at https://www.kaggle.com/datasets/girishkleit/blast-severity-from-the-scale-of-0-to-4

References

He, K., X. Zhang, S. Ren, and J. Sun. 2016. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90

Khadka RB, Manandhar HK, Shrestha S, Acharya B, Sharma P, Baidya S, Luu VS and Joshi KD (2025) Defending rice crop from blast disease in the context of climate change for food security in Nepal. Front. Plant Sci. 16:1511945. doi: 10.3389/fpls.2025.1511945

Pai, P., Amutha, S., Patil, S. et al. Deep learning-based automatic diagnosis of rice leaf diseases using ensemble CNN models. Sci Rep 15, 27690 (2025). https://doi.org/10.1038/s41598-025-13079-z

Jagamohan Padhi, Kunal Mishra, Ashoka Kumar Ratha, Santi Kumari Behera, Prabira Kumar Sethy, Aziz Nanthaamornphong, Enhancing paddy leaf disease diagnosis -a hybrid CNN model using simulated thermal imaging, Smart Agricultural Technology, Volume 10, 2025, 100814, ISSN 2772-3755, https://doi.org/10.1016/j.atech.2025.100814.

Bag, M. K., S. Raghu, A. Banerjee, S. R. Prabhukarthikeyan, M. S. Baite, and M. Yadav. 2023. Durable resistance of rice to major and emerging diseases: Current status. *The Open Agriculture Journal*, 17(1): e187433152212301.

Bock, C. H., K. S. Chiang, and E. M. Del Ponte. 2022. Plant disease severity estimated visually: a century of research, best practices, and opportunities for improving methods and practices to maximize accuracy. *Tropical Plant Pathology*, 47(1): 25-42.

Cruz, Y. J., M. Rivas, R. Quiza, A. Villalonga, R. E. Haber, and G. Beruvides. 2021. Ensemble of convolutional neural networks based on an evolutionary algorithm applied to an industrial welding process. *Computers in Industry*, 133: 103530.

Divya, B., S. Robin, R. Rabindran, S. Senthil, M. Raveendran, and A. J. Joel. 2014. Marker assisted backcross breeding approach to improve blast resistance in Indian rice (*Oryza sativa*) variety ADT43. *Euphytica*, 200(1): 61-77.

FAO. 2024. FAOSTAT. Rome, Italy: Food and Agriculture Organization of the United Nations.

Khadka, R. B., H. K. Manandhar, S. Shrestha, B. Acharya, P. Sharma, S. Baidya, V. S. Luu, and K. D. Joshi. 2025. Defending rice crop from blast disease in the context of climate change for food security in Nepal: A Review. *Frontiers in Plant Science*, 16: 1511945.

Krizhevsky, A., I. Sutskever, and G. E. Hinton. 2017. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6): 84–90.

Liu, Y., Y. Zhang, D. Jiang, Z. Zhang, and Q. Chang. 2023. Quantitative assessment of apple mosaic disease severity based on hyperspectral images and chlorophyll content. *Remote Sensing*, 15(8): 2202.

Mandal, N., S. Adak, D. K. Das, R. N. Sahoo, J. Mukherjee, A. Kumar, V. Chinnusamy, B. Das, A. Mukhopadhyay, H. Rajashekara, and S. Gakhar. 2023. Spectral characterization and severity assessment of rice blast disease using univariate and multivariate models. *Frontiers in Plant Science*, 14: 1067189.

Ning, H., S. Liu, Q. Zhu, and T. Zhou. 2023. Convolutional neural network in rice disease recognition: accuracy, speed and lightweight. *Frontiers in Plant Science*, 14: 1269371.

Padhi, J., K. Mishra, A. K. Ratha, S. K. Behera, P. K. Sethy, and A. Nanthaamornphong. 2025. Enhancing paddy leaf disease diagnosis – a hybrid CNN model using simulated thermal imaging. *Smart Agricultural Technology*, 10: 100814.

Pai, P., S. Amutha, S. Patil, T. Shobha, M. Basthikodi, B. M. A. Shafeeq, and A. P. Gurpur. 2025. Deep learning-based automatic diagnosis of rice leaf diseases using ensemble CNN models. *Science Reports*, 15(1): 27690.

Palma, D., F. Blanchini, and P. L. Montessoro. 2022. A system-theoretic approach for image-based infectious plant disease severity estimation. *PloS ONE*, 17(7): e0272002.

Parasa, G., M. Arulselvi, and S. Razia. 2023. Identification of diseases in paddy crops using CNN. *International Journal of Intelligent Systems and Applications in Engineering*, 11(6s): 548-557.

Pujari, J. D., R. S. Yakkundimath, S. Jahagirdar, and A. M. Byadgi. 2016. Quantitative detection of soybean rust using image processing techniques. *Journal of Crop Protection*, 5(1): 75-87.

Pujari, J. D., R. S. Yakkundimath, and A. M. Byadgi. 2013. Statistical methods for quantitatively detecting fungal disease from fruits' images. *International Journal of*

Intelligent Systems and Applications in Engineering, 1(4): 60-67.

Sabri, S., M. Z. Ab Wahab, Z. Sapak, and I. S. Mohd Anuar. 2023. A review of bacterial diseases of rice and its management in Malaysia. *Food Research*, 7(S2): 120-133.

Saha, M. 2022. Forecasting of rice blast disease severity in West Bengal, India based on PDI values and cumulative logit model. *Journal of Mycopathological Research*, 60(4): 523-530.

Shahbandeh, M. 2024. Research department and content philosophy — Our commitment to empowering your fact-based decisions with reliable data. Statista. Available at: https://www.statista.com/aboutus/our-research-commitment/1239/m-shahbandeh. Accessed 25 06 2025

Shi, T., Y. Liu, X. Zheng, K, Hu, H. Huang, H. Liu, and H. Huang. 2023. Recent advances in plant disease severity assessment using convolutional neural networks. *Scientific Reports*, 13(1): 2336.

Sibiya, M., and M. Sumbwanyambe. 2019. An Algorithm for Severity Estimation of Plant Leaf Diseases by the Use of Colour Threshold Image Segmentation and Fuzzy Logic Inference: A Proposed Algorithm to Update a "Leaf Doctor" Application. *AgriEngineering*, 1(2): 205-219.

Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. CoRR, abs/1409.1556.

Straits Research. 2025. Top 10 rice consuming countries in 2025. StraitsResearch. Available at: https://straitsresearch.com/statistic/top-10-rice-consuming-countries-in-2025. Accessed 25 06 2025.

Szegedy, C., W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. 2014. Going Deeper with Convolutions. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition 2015*, 1-9. Boston, MA, USA, 7-12 June.

Upadhyay, K., and B. Bhatta. 2020. Rice blast (Magnaporthe oryzae) management: A review. *Agricultural Journal*, 15(3): 42–48.

Wang, G., Y. Sun, and J. Wang. 2017. Automatic image-based plant disease severity estimation using deep learning. *Computational Intelligence and Neuroscience*, 2017(1): 2917536.

Yakkundimath, R., and G. Saunshi. 2023. Identification of paddy blast disease field images using multi-layer CNN models. *Environmental Monitoring and Assessment*, 195(6): 646.

Ning, H., Liu, S., Zhu, Q., & Zhou, T. (2023). Convolutional neural network in rice disease recognition: accuracy, speed and lightweight. Frontiers in Plant Science, 14, 1269371.

Parasa, G. ., Arulselvi, M. ., & Razia, S. . (2023). Identification of Diseases in Paddy Crops Using CNN. International Journal of Intelligent Systems and Applications in Engineering, 11(6s), 548–557.

Enhancing paddy leaf disease diagnosis -a hybrid CNN model using simulated thermal imaging (2025). https://doi.org/10.1016/j.atech.2025.100814

Cruz, Y. J., Rivas, M., Quiza, R., Villalonga, A., Haber, R. E., & Beruvides, G. (2021). Ensemble of convolutional neural networks based on an evolutionary algorithm applied to an industrial welding process. Computers in Industry, 133, 103530.