Policy Research Journal
ISSN (E): 3006-7030 ISSN (P) : 3006-7022 Volume 4, Issue 4, 2026

DEEP CONVOLUTIONAL NEURAL NETWORK FOR AUTOMATED CLASSIFICATION OF RICE LEAF DISEASES USING A PUBLICLY AVAILABLE DATASET

Nazir Ahmad[*1], Malik Asim Rajwana[2], Asim Abdul Qadir[3]

[*1,2]*Department of Computer Science, National College of Business Administration & Economics Lahore, Multan Sub Campus, 60000, Pakistan*
[3]*Department of Software Engineering, National University of Modern Languages, Multan, 60000, Pakistan*

[*1]nazirgohar79@gmail.com, [3]asim.qadir@numl.edu.pk

DOI: https://doi.org/10.5281/zenodo.19678353

Keywords
Rice leaf disease, deep learning, convolutional neural network, tagging, diagnosis, smart farming

Article History
Received: 24 February 2026
Accepted: 04 April 2026
Published: 21 April 2026

Corresponding Author: *
Nazir Ahmad

Abstract
Rice leaf diseases are a major issue for the whole world's food security because they lower crop harvest and grain yield. Detecting these diseases at the initial stage is significant for dealing with crops effectively. Traditional detection methods depend on physical observation. This process takes time and there can be mistakes. A convolutional neural network (CNN), a deep learning model, is used in this research paper on a publicly available Rice diseases images dataset at Kaggle to classify the rice leaf diseases. The dataset consists of 5932 images and 04 folders named as healthy leaves, brown leaf spot (BLS), bacterial leaf blight (BLB), and Rice leaf smut (RLS). Before training, every image was normalized and resized to (128 x 128) pixels. The model is a grouping of six convolutional layers with max pooling, Rectified linear unit activation, and dropout regularization. These are followed by fully connected layers using softmax classification. Training used the Adam optimizer with a learning rate of 0.001 for 50 epochs. Accuracy, recall, precision, and F1 score were used to measure performance. The proposed model attained a validation accuracy of 92.60%. This was higher than K Nearest Neighbors at 56.84% and Support Vector Machine at 61.47%. These results show that the model can properly detect different rice disease types under numerous field situations. This study proposes that deep learning can make a substantial contribution to precision farming by providing a consistent and efficient method for disease diagnosis. Anticipated steps will enhance the dataset, implement the model in its intended environment, and design lightweight versions appropriate for mobile and Internet of Things (IoT) applications.

1. INTRODUCTION

Rice stands as an essential food source in the world that offers nutrition to a large part of the global population [0]. Its production is often condensed by some leaf diseases that affect both yield and quality. Identifying these diseases in time is important for controlling and promoting healthier crop management. Traditional methods depend on physical inspection which takes time, needs skilled knowledge and may cause miscalculations [0]. New developments in deep learning, especially in convolutional neural networks, have made it possible to identify plant diseases from images with high accuracy.

These networks can learn difficult features from leaf images without physical processing and are suitable for large-scale farming [0]. This article presents a CNN model for automatic detection of rice leaf diseases using the Rice Diseases Image Dataset from Kaggle. It aims to generate an accurate and real-world system that can help farmers manage rice crops more efficiently and help with smart farming practices [0]. Figure1 below illustrate the rice plants diseases ratio in graphical form. While Figure2 shows random images selected from the dataset.

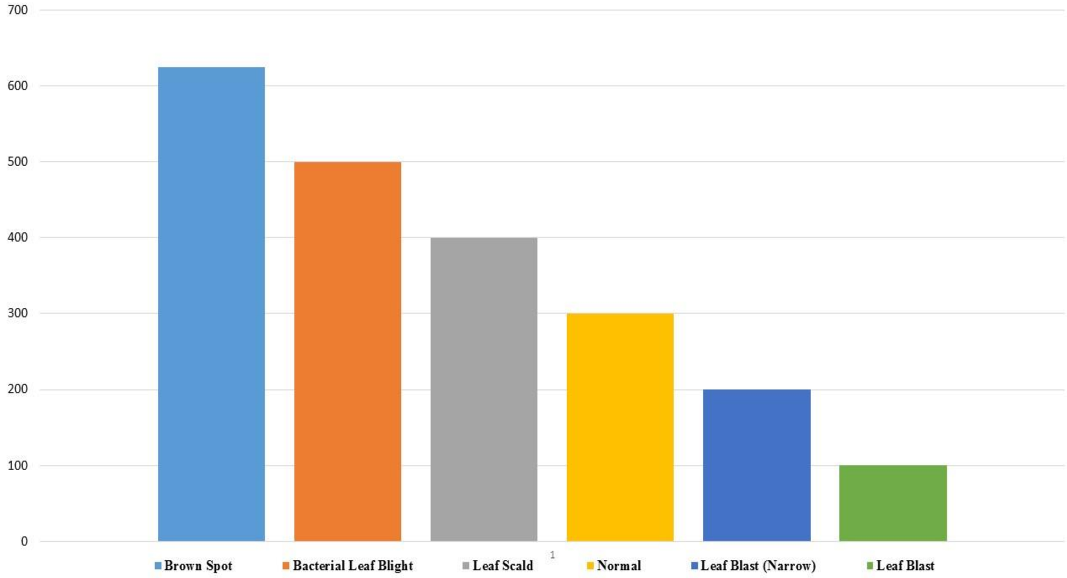


Figure 1: Column Chart Showing the Different Diseases ratio of Rice plants

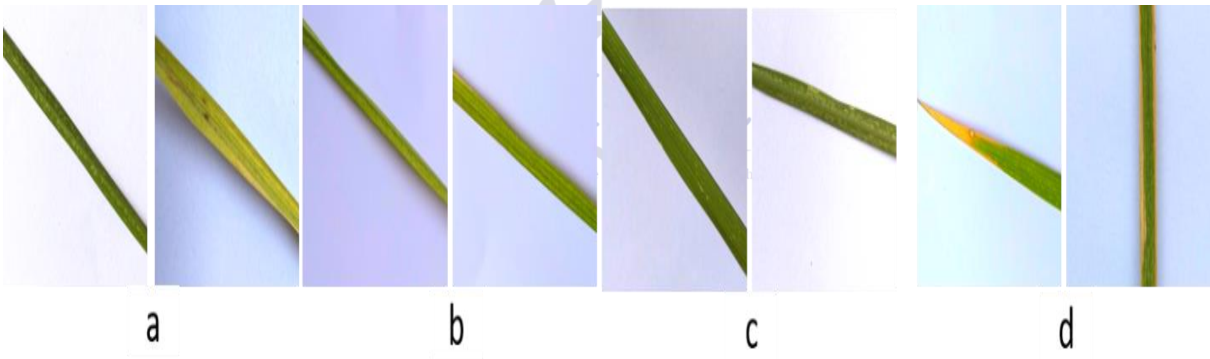


Figure 2: Random images from the dataset

1.1 Dataset Description

This Dataset is publicly available on Kaggle and consists of 5932 images grouped into four further classes. The Bacterial Leaf Blight class has 1613 images and is produced by the bacteria Xanthomonas oryzae pv oryzae. Brown Spot class consists of 1578 images and is a fungal disease produced by Bipolaris oryzae. Leaf Smut class contains 1627 images and is a fungal infection initiated by Entyloma oryzae. Healthy has 1114 images and represents rice leaves without infection. The images show variances in lighting, color, texture, leaf position and background which show real field conditions. Each image in these classes was resized to 128 by 128 pixels and normalized formerly training. The dataset was range into 70.00% for training, 20.00% for validation and 10.00% for testing.

2. Related Work

Deep learning methods display solid performance in identifying and grouping agricultural diseases [0]. Previous research suggests that models based on convolutional neural networks (CNN) perform better than traditional machine learning techniques when working with complex image data [0]. CNNs have been used to classify 44 plant species with high accuracy by learning feature patterns without supervision [0].

Networks such as Google Net and VGGNet have also been efficiently used for categorizing plant leaf images [0]. Deeper networks, like a 26-layer residual model, have classified 100 ornamental plant species with accuracy above 91 .00% [0]. In field analysis, encoder-decoder CNN models have been applied for separating crops, weeds and background from RGB images [0].

For rice disease detection, some new studies trained CNN models using datasets such as the Rice Leaf Disease Dataset from the UCI repository and achieved an accuracy above 90 .00% [0]. However, these studies often relied on small or private datasets which reduced reproducibility. This study uses the publicly available Rice Diseases Image Dataset from Kaggle to provide a transparent and available foundation for future research.

The study presented in [0]"Applying Augmentation Techniques for better Accuracy" (2026) determines the working of data augmentation on deep learning for vegetal ailment discovery. By means of convolutional neural networks trained on augmented datasets, the model achieved enhanced organizational accuracy, highlighting the importance of data variety in increasing generality across hidden trials.

The proposed of CNN founded framework for rice leaf disease sorting by means of image datasets of diseased crops. Their model attained modest accuracy by [0]leveraging feature extraction competencies of deep networks, showing strength in finding numerous disease classes under changing ecological situations.

This paper focused on optimizing dataset [0] variety to advance rice flash illness detection. By exercising a deep learning model on varied databases, the study stated improved accuracy and model stability, highlighting that various training data meaningfully recover real-world situations to present.

This article showed an inclusive review of deep learning techniques [0] for rice disease detection, with datasets, architectures, and evaluation metrics. The study acknowledged CNN-founded models as leading, with accuracy and F1 score being the most frequently used metrics, while also highlighting difficulties such as dataset inequity and imperfect field inconsistency.

This study established a CNN-based model for rice leaf disease detection using ground-composed datasets. Their method attained [0] high classification accuracy, representing the success of deep learning in taking disease-based graphic designs under real farming circumstances.

This paper presented a hybrid [0]feature-improved CNN model that mixes feature collection with convolutional learning. The model enhanced forecast accuracy and minimized computational complications, demonstrating the worth of hybrid optimization approaches in farming disease sorting.

This study shared CNN with [0]generative adversarial networks (GAN) to improve rice leaf disease finding. The GAN factor made artificial training samples, which better model strength and classification accuracy, mainly in situations with partial labeled data.

This research planned a hybrid CNN model joining imitation thermal imagery for [0]paddy disease diagnosis. Their approach improved feature illustration and attained improved accuracy, illustrating the possibility of multimodal data integration in precision farming.

This paper offered a complete rice leaf image dataset intended for machine learning submissions. The dataset contains [0]varied disease groups and environmental differences, giving a good standard for evaluating deep learning models in farming diagnosis.

This article defines a hybrid CNN ResNet50 BiLSTM architecture for paddy leaf disease detection. The model leveraged [0] 3-D and chronological feature extraction, attaining high accuracy and better organization performance related to standalone CNN models.

This work conducted a relative study between already trained and custom CNN architectures for rice disease recognition. Their results showed that pretrained models usually outperform custom networks in terms of accuracy and training efficiency when practically used in limited datasets.

This research established a hybrid deep learning [0]framework for rice leaf disease sorting. The model shared multiple learning tactics to better discover accuracy, indicating success in handling complex disease designs and differences in leaf images.

3. Methodology

This section will describe the data processing and baseline models detail as below with images segmentation process.

3.1 Data Preprocessing

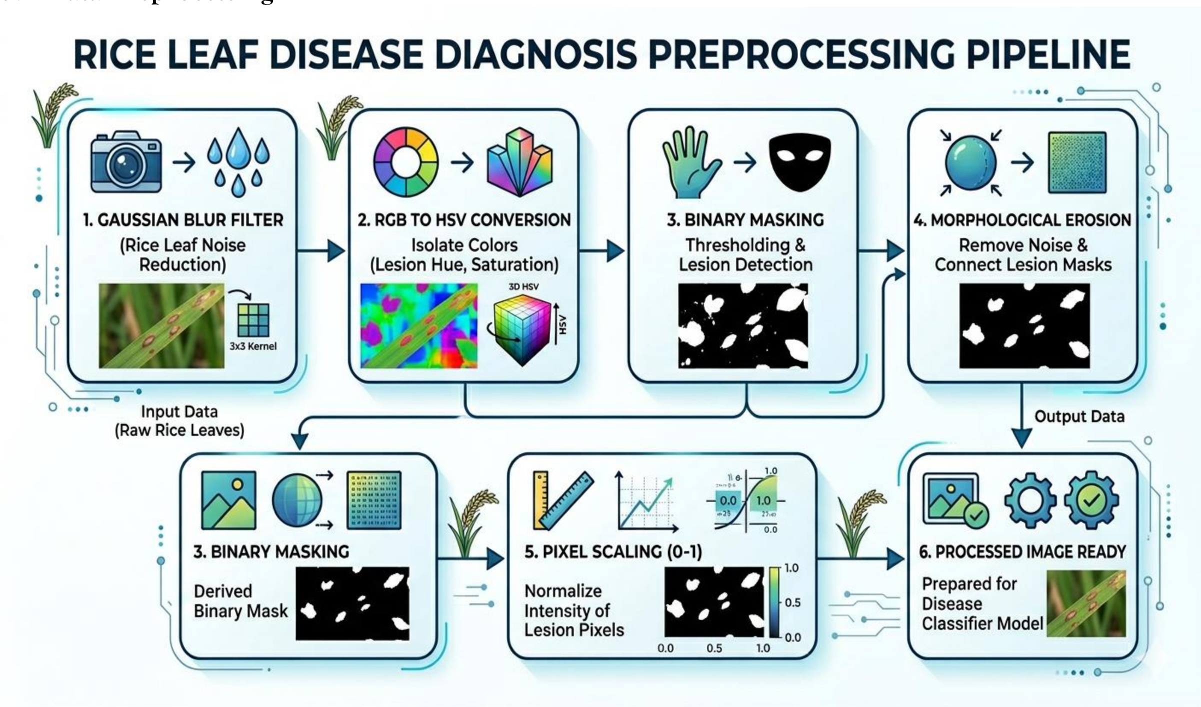


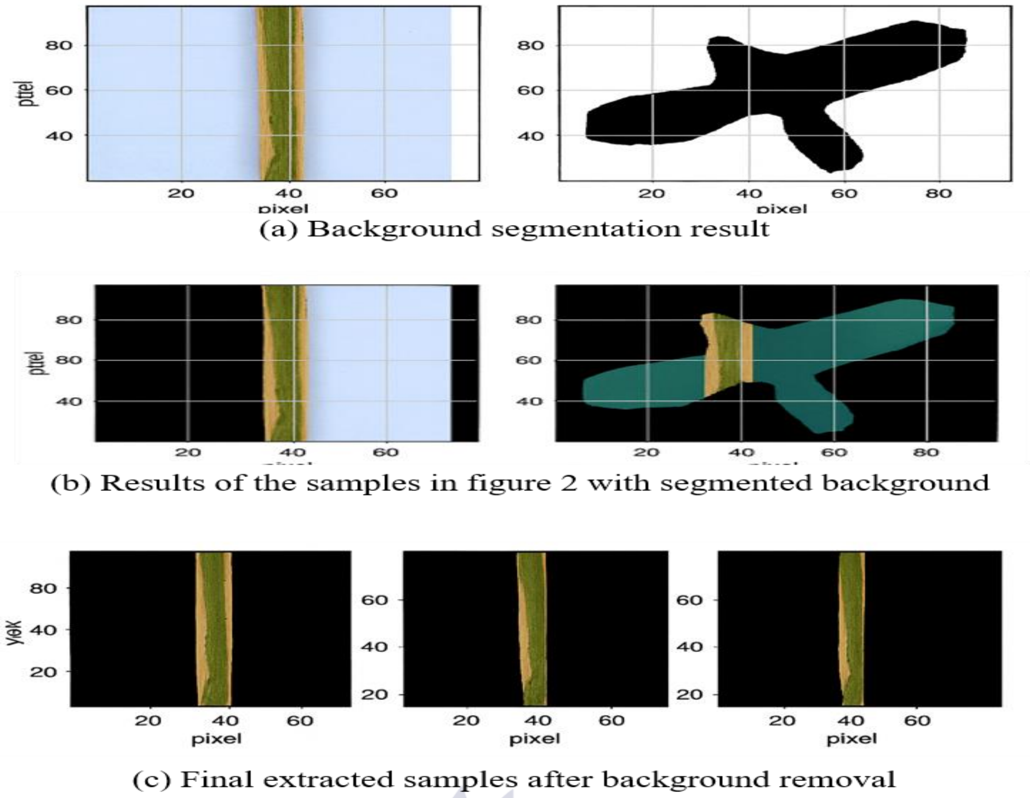
Figure 3: Rice leaf data preprocessing

Each image in the dataset was managed with a Gaussian Blur filter to decrease noise and highlight vital parts of the leaf. The images were then converted into the HSV color space to improve feature extraction under different lighting situations. A binary mask was made using the HSV range of rice leaves and small background noise was removed through morphological erosion with an 11 by 11 kernel. The pixel values were then scaled between 0 and 1 uniformly as illustrated in Figure3

3.2 Baseline Models

Two traditional machine learning models were used.

K Nearest Neighbors was improved through grid search to find the best number of neighbors. The finest result was obtained with k =5 using uniform weights which achieved 56.84% accuracy on validation data.



(a) Background segmentation result

(b) Results of the samples in figure 2 with segmented background

(c) Final extracted samples after background removal

Figure 4: Images of the segmentation process

Support Vector Machine (SVM) used a linear kernel with a penalty value of C=5 and reached 61.47% accuracy.

These results show that traditional models cannot fully learn the complex visual features existing in rice leaf images which supports the use of convolutional neural networks for good performance.

4. Convolutional Neural Network (CNN): Results and Comparisons

This part of the study will discuss about the architecture model, Evaluation Metrics and Experimental Results.

4.1 Model Architecture

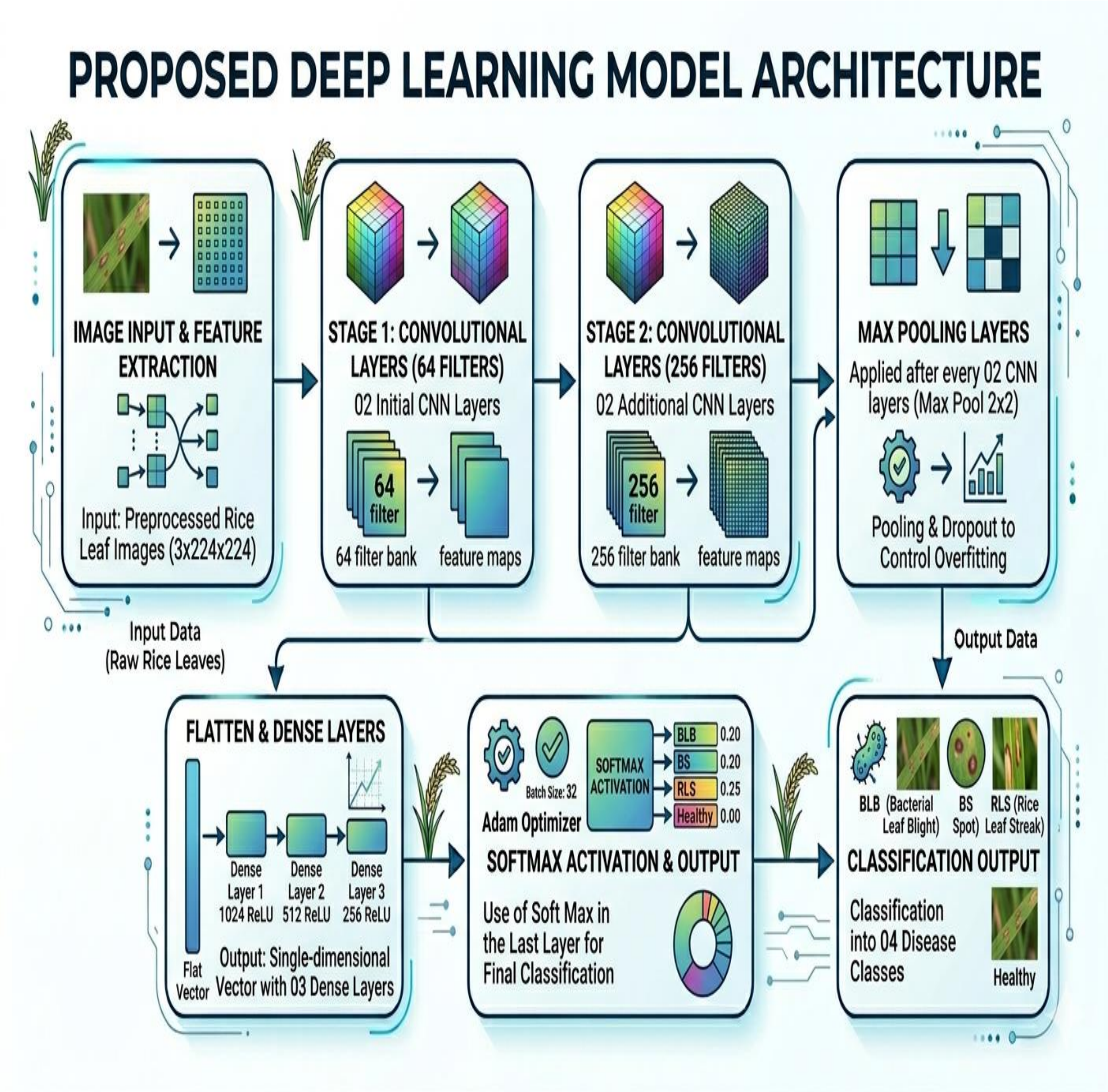


Figure 5: Proposed model architecture

The proposed convolutional neural network has six convolutional layers followed by ReLU activation and zero padding. The initial two layers hold 64 filters, while the next two have 128 filters and the rest two enclose 256 filters. Max pooling is applied after every two convolutional layers to decrease the feature diagram size. A dropout of 10 percent follows each pooling layer to prevent overfitting. The output is redesigned into a one-dimensional vector and nursed into three dense layers. The last layer utilizes a softmax activation function to classify images into four classes which are BLB, BS, RLS and Healthy. Training was sustained with a set of 32 by means of the Adam optimizer at a learning rate of 0.001 and definite logarithmic loss as the loss function for 50 epochs. This is explained in Figure 5 below:

4.2 Evaluation Metrics

The growth stage was measured through confusion matrix, accuracy, precision, recall and F1 score. These measures deliver a complete view of working and help to assess the model even when the data classes are not balanced.

4.3 Evaluation Matrices

The **accuracy** $S_{Acc}^{(r)}$ is defined as:

$$S_{Acc}^{(r)} = \frac{t_p^{(r)}+t_n^{(r)}}{t_p^{(r)}+f_p^{(r)}+f_n^{(r)}+t_n^{(r)}} \quad (1)$$

where $t_p^{(r)}$, $f_p^{(r)}$, $f_n^{(r)}$, and $t_n^{(r)}$ represent true positives, false positives, false negatives, and true negatives, respectively.

The **recall (sensitivity)** $S_{Rec}^{(r)}$ is computed as:

$$S_{Rec}^{(r)} = \frac{t_p^{(r)}}{t_p^{(r)}+f_n^{(r)}} \quad (2)$$

The **specificity** $S_{Spec}^{(r)}$, which measures correct identification of normal cases, is expressed as:

$$S_{Spec}^{(r)} = \frac{t_n^{(r)}}{t_n^{(r)}+f_p^{(r)}} \quad (3)$$

The **precision** $S_{Prec}^{(r)}$, indicating the reliability of positive predictions, is defined as:

$$S_{Prec}^{(r)} = \frac{t_p^{(r)}}{t_p^{(r)}+f_p^{(r)}} \quad (4)$$

The **F1-score** $S_{F1}^{(r)}$, representing the harmonic mean of precision and recall, is computed as:

$$S_{F1}^{(r)} = \frac{2\times S_{Prec}^{(r)}\times S_{Rec}^{(r)}}{S_{Prec}^{(r)}+S_{Rec}^{(r)}} \quad (5)$$

The **Matthews Correlation Coefficient (MCC)** $S_{MCC}^{(r)}$ is formulated as:

$$S_{MCC}^{(r)} = \frac{\left(t_p^{(r)}t_n^{(r)}-f_p^{(r)}f_n^{(r)}\right)}{\sqrt{\left(t_p^{(r)}+f_p^{(r)}\right)\left(t_p^{(r)}+f_n^{(r)}\right)\left(t_n^{(r)}+f_p^{(r)}\right)\left(t_n^{(r)}+f_n^{(r)}\right)}} \quad (6)$$

4.4 Experimental Results

The model was practiced under two setups. The first used raw RGB images without any preprocessing. The second used preprocessed images that were segmented using OpenCV before training.

Table1: Table showing the performance of the algorithms

Algorithm	Description	Accuracy (%age)
KNN	Traditional baseline	56.84
SVM	Traditional baseline	61.47
CNN Raw images	Deep CNN without preprocessing	80.21
CNN Preprocessed images	Deep CNN with OpenCV segmentation	92.60

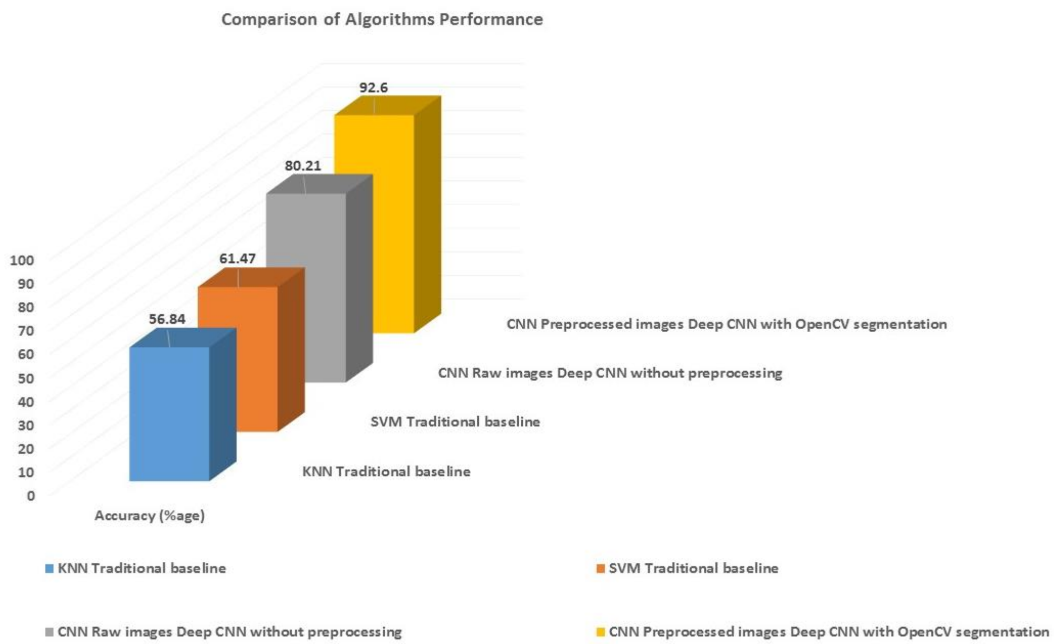


Figure 6. Graphical representation of algorithm performance

The best CNN model achieved 98.30% accuracy in training and 92.60% accuracy in validation. It performed much better than the baseline models. The confusion matrix showed the best results for Brown Spot and Bacterial Leaf Blight, with some overlap between Leaf Smut and Healthy classes. The average precision was 91.80%, the recall was 92.30%, and the F1 score was 91.50% across all classes, which checks the stability and consistency of the model for rice disease detection. To understand easily and make comparison among the algorithms performance that is graphically shown in Figure 6.

5. Conclusion

This paper presented a CNN model for automatic finding of rice leaf diseases using a publicly available dataset. The model yielded a validation accuracy of 92.60% which was greater than traditional models such as K-Nearest Neighbors algorithm (KNN) AND Support Vector Machine (SVM). The outcomes show that convolutional neural networks can learn complex visual features from rice leaves and provide accurate classification under different field conditions. By using the publicly available Rice Diseases Image Dataset, this article promotes replicability and gives a solid start for future agricultural research. The results show that deep learning can keep precision farming by improving disease detection and helping to increase crop production.

6. Future Work

Future work regarding this study is proposed as:

I. Expanding the dataset by adding more disease types and environmental conditions to improve model accuracy and flexibility.

II. Testing the model with field images captured by drones or mobile devices to check its performance in real situations.

III. Integrating the trained model into smart farming systems such as IoT devices and mobile applications for real-time monitoring.

IV. A lightweight version of the models, like EfficientNet and MobileNet, for use in limited online service areas.

These suggestions will help to make a reliable and automatic system for disease detection that can help the farmer to make quick decisions to enhance productivity and minimize losses.

REFERENCES:

Bouman, B., Barker, R., Humphreys, E., Tuong, T. P., Atlin, G., Bennett, J., ... & Wassman, R. (2013). Rice: feeding the billions. In *Water for Food Water for Life* (pp. 515-550). Routledge.

Johnson, T. L., Fletcher, S. R., Baker, W., & Charles, R. L. (2019). How and why we need to capture tacit knowledge in manufacturing: Case studies of visual inspection. *Applied ergonomics*, *74*, 1-9.

Wang, D., Cao, W., Zhang, F., Li, Z., Xu, S., & Wu, X. (2022). A review of deep learning in multiscale agricultural sensing. *Remote Sensing*, *14*(3), 559.

Alfred, R., Obit, J. H., Chin, C. P. Y., Haviluddin, H., & Lim, Y. (2021). Towards paddy rice smart farming: a review on big data, machine learning, and rice production tasks. *Ieee Access*, *9*, 50358-50380.

Zhou, J., Li, J., Wang, C., Wu, H., Zhao, C., & Teng, G. (2021). Crop disease identification and interpretation method based on multimodal deep learning. *Computers and Electronics in Agriculture*, *189*, 106408.

Liu, Y., Pu, H., & Sun, D. W. (2021). Efficient extraction of deep image features using convolutional neural network (CNN) for applications in detecting and analysing complex food matrices. *Trends in Food Science & Technology*, *113*, 193-204.

Lee, S. H., Chan, C. S., Mayo, S. J., & Remagnino, P. (2017). How deep learning extracts and learns leaf features for plant classification. *Pattern recognition*, *71*, 1-13.

Syarief, M., & Setiawan, W. (2020). Convolutional neural network for maize leaf disease image classification. *Telkomnika (Telecommunication Computing Electronics and Control)*, *18*(3), 1376-1381.

Vibe, I. M. Ø. (2007). *San Personal Ornaments from the Later Stone Age at Blombos Cave and Blomboschfontein, southern Cape, South Africa* (Master's thesis, The University of Bergen).

Khan, A., Ilyas, T., Umraiz, M., Mannan, Z. I., & Kim, H. (2020). Ced-net: crops and weeds segmentation for smart farming using a small cascaded encoder-decoder architecture. *Electronics*, 9(10), 1602.

Bijoy, M. H., Hasan, N., Biswas, M., Mazumdar, S., Jimenez, A., Ahmed, F., ... & Momen, S. (2024). Towards sustainable agriculture: a novel approach for rice leaf disease detection using dCNN and enhanced dataset. *IEEE Access*, *12*, 34174-34191.

Applying Augmentation Techniques for Improved Accuracy: A Deep Learning Approach. In 2026 7th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI) (pp. 1155-1160). IEEE.

Perez, J. G., & Perez, E. S. (2026, January). CNN-Based Rice Leaf Disease Classification Alfred, R., Leo, J., & Kaijage, S. F. (2025). Optimizing dataset diversity for a robust deep-learning model in rice blast disease identification to enhance crop health assessment across diverse conditions. Smart Agricultural Technology, 10, 100726

Ismail, U. I., Chua, H. N., Nordin, R., & Ahmed, M. K. (2025). A comprehensive review of deep learning approaches for rice disease detection: Datasets, methodologies, and future directions. Smart Agricultural Technology, 11, 100976.

Kumar, K. K., Sasikala, S., Maheswari, G., Suthar, M. K. B., Sridevi, P., Kavitha, D., & Vengatesh, T. (2025). A convolutional neural network approach for rice leaf disease detection in India using deep learning. Journal of Neonatal Surgery, 14(4s).

Vijayan, S., & Chowdhary, C. L. (2025). Hybrid feature optimized CNN for rice crop disease prediction. Scientific Reports, 15(1), 7904.

Ramadan, S. T. Y., Islam, M. S., Sakib, T., Sharmin, N., Rahman, M. M., & Rahman, M. M. (2025). Image-based rice leaf disease detection using CNN and generative adversarial network. Neural computing and applications, 37(1), 439-456.

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

Hasan, A., Layes, T. A., Afridi, A. S., Rifat, S. H., Nur, F. N., & Moon, N. N. (2025). A comprehensive dataset of rice leaf images for disease detection using machine learning. Data in Brief, 111977.

Basitur Rahman Bappi, M., Masfequier Rahman Swapno, S. M., Akhter, S., & Rabbi, M. F. (2025). Deploying CNN-ResNet50-BiLSTM for paddy leaf disease detection. In Machine vision in plant leaf disease detection for sustainable agriculture (pp. 131-143). Singapore: Springer Nature Singapore.

Mridul, A. H., Armin, J. F., Saleh, M. A., Das, A., Islam, M. M., Shakil, G. M., & Rakib, M. A. (2025, November). Optimizing Rice Health: A Comparative Evaluation of Pretrained and Custom Convolutional Neural Networks for Disease Recognition. In Artificial Intelligence and Applications.

Raigonda, M. R. (2025). Identification And Classification of Rice Leaf Disease Using Hybrid Deep Learning. Journal of Scientific Research and Technology, 93-101.