



**IJIRCCCE**

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 5, May 2024

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

**Impact Factor: 8.379**



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

# Machine Learning Methods for Identification of Leaf Disease and Predicting Crop

Aniket Patole<sup>1</sup>, Ankit Nagwade<sup>2</sup>, Rushikesh Rokade<sup>3</sup>, Prof. Anupama Ghadge<sup>4</sup>

UG Student, Dept. of E & TC Engg., HSBPVT'S GOI Faculty of Engineering, Kashti, Maharashtra, India<sup>1-3</sup>

Assistant Professor, Dept. of E & TC Engg., HSBPVT'S GOI Faculty of Engineering, Kashti, Maharashtra, India<sup>4</sup>

**ABSTRACT:** The timely identification of rice leaf diseases is critical to the continued viability of rice as a staple meal on a worldwide scale. Therefore, the key to reducing the expense of chemical testing and manual visual identification is to effectively detect diseases after they have already manifested. The identification in the recent past the majority of leaf diseases in crops are treated manually with specialized equipment, which is laborious and ineffective. This work provides a cure by effectively identifying and categorizing rice leaf diseases through the use of Deep Learning (DL) and transfer learning techniques. Regardless of the degree of illness spread throughout the leaves, a comprehensive dataset consisting of 5932 self-generated photos of rice leaves was assembled in conjunction with the benchmark datasets. The images were classified into 9 classes.

**KEYWORDS:** Rice, leaf disease, Deep learning, CNN, Crop yield, Agriculture, Food security

## I. INTRODUCTION

According to the Food and Agriculture Organization (FAO) (Asibi et al., 2019; Taw, 2022), rice is the primary food source for more than half of the world's population and damage to crops result in substantial losses to farmers worldwide. These losses can have serious economic, social, and environmental impacts, as they affect food security, poverty reduction efforts, and sustainable development. Further, International Rice Research Institute (IRRI) states that, rice diseases can deteriorate yields by up to 80%, leading to a decline in farm incomes, lower food availability, and higher prices for consumers (Shew et al., 2019). Which further results to the use of harmful pesticides and chemicals, which can have negative impacts on human health and the environment (Jones, 2021). Therefore, effective management of rice leaf diseases is crucial for ensuring sustainable rice production and food security for millions of people around the world (Damalas and Koutroubas, 2016; Tudi et al., 2021) Spotting the severity of the disease in rice plants is often determined by the extent and spread of the infection over the surface area of the leaves (Tudi et al., 2021; Bock et al., 2022) and (Liu et al., 2008). A mild case of the disease is characterized by less than 10% damage to the leaf surface, while severe cases are defined by more than 10% damage. It is significant because it directly impacts the growth, development, and yield of rice crops (Shoaib et al., 2023) Moreover, spotting this rice disease manually seems tiring and does not provide early detection of diseases, which can lead to significant yield losses (Andrew et al., 2022). Accurate, real-time identification and categorization of crop diseases are made possible by using machine learning and deep learning techniques, improving agricultural productivity and quality. It simultaneously helps the process become more accurate overall and reduces labour costs (Aggarwal et al., 2022). Recent advances in DL, particularly in the field of image processing, have shown great potential in the early identification and classification of security for millions of people around the world (Damalas and Koutroubas, 2016; Tudi et al., 2021) Spotting the severity of the disease in rice plants is often determined by the extent and spread of the infection over the surface area of the leaves (Tudi et al., 2021; Bock et al., 2022) and (Liu et al., 2008). A mild case of the disease is characterized by less than 10% damage to the leaf surface, while severe cases are defined by more than 10% damage. It is significant because it directly impacts the growth, development, and yield of rice crops (Shoaib et al., 2023) Moreover, spotting this rice disease manually seems tiring and does not provide early detection of diseases, which can lead to significant yield losses (Andrew et al., 2022). Accurate, real-time identification and categorization of crop diseases are made possible by using machine learning and deep learning techniques, improving agricultural productivity and quality. It simultaneously helps the process become more accurate overall and reduces labour costs (Aggarwal et al., 2022). Recent advances in DL, particularly in the field of image processing, have shown great potential in the early identification and classification of plant diseases. Convolutional Neural networks (CNN) have successful in identifying and classifying plant diseases based on images of plant leaves. The use of CNNs in the early identification and classification of rice leaf diseases can provide a reliable and quick way of detecting and diagnosing plant diseases, enabling timely interventions to minimize yield loss (Lu et al., 2017). Compared to

traditional approaches, DL models offer several advantages, including high accuracy, speed, and the ability to process large amounts of data. This has significant implications for agriculture, as the early detection and management of plant diseases are crucial for maintaining crop yield and ensuring food security. Several studies support this fact that DL models outperform traditional methods in identifying and classifying plant diseases (Javidan et al., 2023; Kumar et al., 2023). These models can effectively detect plant diseases at an early stage, providing a quick and reliable way of detecting and diagnosing plant diseases [Andreas et al.2018]. In addition to this, the precision and resilience of the DL models can be further improved using pre-trained models and data augmentation approaches, thereby rendering them more trustworthy for use in practical applications. Therefore, (Wang et al., 2023) in his paper, explores the use of deep learning techniques, specifically CNNs, to detect and classify the severity of rice leaf diseases and evaluate the performance analysis of the transfer learning approaches. Furthermore, (Aggarwal et al., 2023a) in his work, he proposed InceptionResNet V2 for the classification of rice plant disease. He evaluated the following categories of rice leaf diseases namely bacterial leaf blight, blast, Brown spot using his proposed model and achieved an accuracy of 88% in classifying those diseases. In this article, we propose a novel architecture of VGG 16 mitigating the limitations observed in prior works in plant disease detection and classification. In this proposed customized VGG 16 architecture, we introduce critical changes, including the strategic incorporation of supplementary dropout layers, dense layers, and the fine-tuning of filter counts in the convolutional layers. Through this meticulous customization, our proposed architecture demonstrates significantly improved generalization capability and heightened accuracy, effectively addressing common challenges such as overfitting and elevating the model's capacity to recognize intricate patterns and abstract features within the images. Consequently, our proposed approach presents a substantiated solution with the existing state-of-art models for advancing reliable and efficient plant disease detection in diverse agricultural settings.

## II. RELATED WORK

Following literature survey is conducted with different technologies used for the system. The study shows the problems which could be faced by system and the solution is given for every problem.

A 37% annual drop in rice yield is a consequence of rice plant diseases. Although there hasn't yet been a suitable application created that is capable of precisely diagnosing these rice plant diseases and controlling those diseases, it may occur primarily because of a lack of understanding in identifying and controlling rice plant diseases. However, there have only been a few studies done on the diagnosis of illnesses affecting rice plants. This section provides an insight to the recent literature summary on the existing deep learning models in rice disease identification and the state-of-art which exists in classifying the rice plants disease. (Bari et al., 2021) used Faster R-CNN algorithm to distinguish blast, brown spot and Hispa and identified healthy leaves with 99.25% accuracy. The proposed models were evaluated using self-generated database and Kaggle with a total of 2400 images. Caffe DL approach were used where feature maps of infected leaves are used for training purposes; An original Rice Leaf Disease Dataset (RLDD) was constructed from both an online database and their own dataset.(Bari et al., 2021; Pandian et al., 2022) developed a ResNet19 and employed evolutionary search technique for optimising its layers. Augmentation techniques such as scaling, cropping, flipping, padding rotation, etc. were used to create additional data from the existing images. The ResNet197 model detects various plant leaf diseases with the input image of size  $224 \times 224 \times 3$  pixels. This ResNet197 model used six blocks of layers and it was trained on a combined dataset consisting of 154,500 images from 22 plants which included both healthy and diseased leaves. While training this model in a GPU environment for up to 1000 epochs resulted in an average classification accuracy rate of 99.58% which is better than other existing architectures or transfer learning methods.

(Roy and Bhaduri, 2021) designed a model to address early disease detection and optimized both speed and accuracy in detecting apple diseases under complex orchard scenarios. The mean average precision (mAP) achieved was 91.2% with an F1 score of 95.9%, at a rate of 56.9 FPS (frames per second). Compared to existing models, the proposed model for multi-class plant disease detection showed significant improvement in mAP by 9%, 0.05% and F1 Score by 7.6%. (Saber Anari, 2022) used Model Engineering (ME) learning to classify diseased leaves. The ME learning technique employed a combination of deep transfer learning, multiple support vector machine (SVM) models as shown in Fig. 4, and radial basis function to extract features from the images. The combination of multiple techniques helped to improve the recognition of leaf diseases, including k-NN, DT, NN, SVM-L, SVM-RBF, and ensemble models. All the techniques were used in combination to extract features from the data and classify the images. The deep transfer learning model used was a modified version of a deep CNN that was used to extract features from images of leaves on various fruits. In a study, (Deng et al., 2021) collected 33,026 images with six varieties of rice leaf diseases



and used to train and test five sub-models. The information about the methods used, the dataset used, the camera used to capture data, the number of observations, the learning rate, the number of iterations, and the performance of the model are also briefed.

(Deng et al., 2021; Li et al., 2017) used a Faster-RCNN model to diagnose rice leaf diseases. The dataset used was from rice fields in Anhui, and Hunan Province in China and was captured using a mobile phone camera and a Sony DSC-QX10 camera. The number of observations, the learning rate and the number of iterations was 5320, 0.002, and 50,000 respectively. (Prajapati et al., 2017) used a SVM model to diagnose rice leaf diseases. The dataset used was from a farm field and was captured using a NIKON D90 digital SLR camera. The number of observations was 120, the learning rate and the number of iterations were not reported. The performance of the model was 93.33%, 73.33%, 83.80% and 88.57% during training, testing, 5-fold cross-validation, and 10-fold cross-validation respectively. (Rahman et al., 2020) used a simple CNN model to diagnose rice leaf diseases. Recently to bring precise classification of the rice plant disease. (Velusamy et al., 2023) proposed a hand-crafted feature engineering in database by performing segmentation, augmentation, and pre-processing. Followed by the hand-crafted featuring techniques the classification accuracy improved with an improvement of 3.1% producing an accuracy of 90.63%. Another approach presented by (Aggarwal et al., 2023a, 2023b, 2023c) CNN's watershed and graph cut segmentation algorithm and improved the classification of rice leaf blight, rice blast and spot with 94% of classification accuracy but lacked scope in terms of severity-based classification. However, made a significant mark in terms of accuracy. In our study we try to overcome this problem of severity-based classification while actively sustaining the robustness of the DL model by relying on Transfer Learning (TF) and Deep-Feature Extraction methodology ensuring accuracy and reliability so that the proposed work can be adopted and implemented in any similar agricultural scenario with ease. The authors of (Aggarwal et al., 2023b) focused exclusively on classifying different types of leaf diseases, without considering severity-based classification. While their work achieved notable accuracy, it lacked the dimension of assessing disease severity. In our research, we aim to address this limitation by incorporating severity-based classification.

### III. PROPOSED METHODOLOGY

#### A. System Architecture:

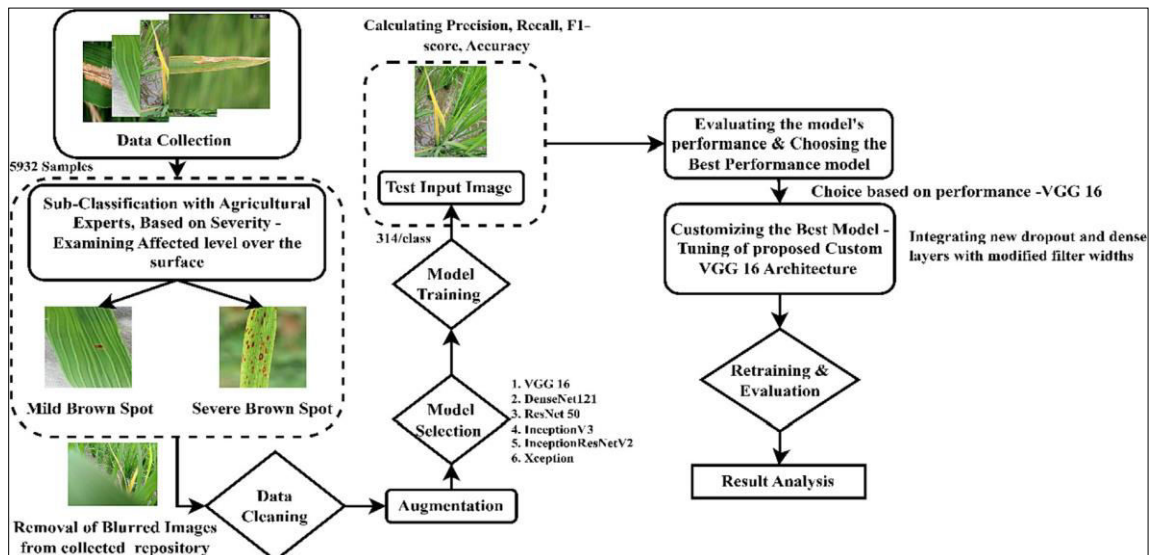


Fig 1 : System Architecture



Fig 2 : Sample Dataset

B. Proposed Approach:

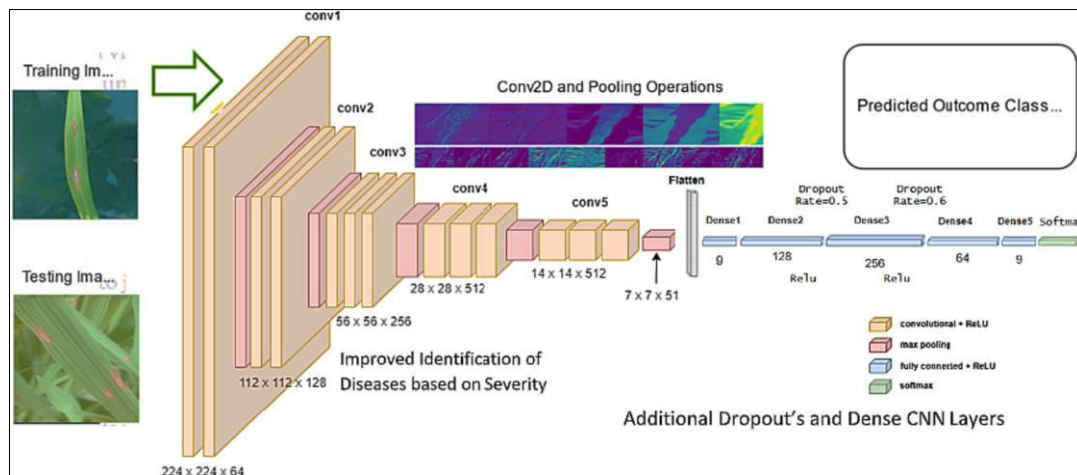


Fig 2 : Proposed Approach

C. Working:

The productivity and quality of the crop will be adversely affected by any illness that affects the rice crop. With the goal to take measures that can minimize rice yield loss, improve rice quality, and increase farmer income, it is important to accurately diagnose rice leaves disease at an early stage. Presently, farmers must diagnose and categorize diseases by hand, which takes lengthier. To get around this, automated techniques can be used to spot plant leaf disease early on. Data accessibility therefore becomes a crucial component of Convolutional Neural Networks (CNN) training since they are capable of detecting rice plant illness on their own. Because CNNs are accurate at classifying and identifying images, many researchers have employed them to identify plant diseases. Therefore, a detailed workflow of our proposed model is depicted in Fig. 1 wherein the evaluation is conducted with our own self-generated database.

These pre-trained CNN's makes use of the spatial correlations between pixels in the picture, enabling them to learn helpful features for image classification through convolutional layers and pooling layers and we will be performing training and validation tests with the pre-trained CNN's and proposed custom VGG 16 model.

1.1. Data gathering

For our evaluation we have used the dataset proposed by (Sethy et al., 2020), in their work on SVM-based deep feature-based disease diagnosis for rice leaf. The collected rice leaf dataset initially contained 5932 images of 4 types of diseases such as Tungro with a collection of 1308 images, Blast with 1440 sample images, Bacterial Blight with 1584 images, and Brown spot with 1600 images. This dataset is made available in the Mendeley archive data. Apart from this, 900 images of healthy leaves images were collected from the Kaggle website and UCI Machine Learning Repository so that the proposed model possibly distinguishes the healthy rice leaf plants from the different categories of rice plant diseases. Fig. 2 depicts the sample images of the dataset with different categories of rice leaves.

1.2. Data pre-processing and cleaning

The collected dataset contained several noisy images, Blurred and unwanted. They had to be cleaned so that they do not hinder the performance of the model. By cleaning these images, we were able to improve our classification. The process of data cleaning includes,

- *Noise and Blur Removal:* Identifying the noisy and blurred images within the dataset.
- *Duplicate Removal:* Identifying and removing any duplicate entries or images within the dataset to prevent bias or redundancy.
- *Data Augmentation:* Generating additional data samples through techniques like rotation, cropping, or adding noise to improve the dataset's diversity.
- *Label Validation:* Checking the accuracy and consistency of the labels assigned to each data point. Ensure that labels are correct and correspond to the data accurately.
- *Balancing Classes:* In classification tasks, ensuring that each class or category has enough samples to prevent class imbalance issues.
- *Data Splitting:* Dividing the dataset into training, validation, and test sets to assess model performance accurately and avoid data leakage.

Therefore, the process of data cleaning holds a pivotal role within the data preparation pipeline because the dataset's quality has a direct impact on the effectiveness and dependability of machine learning models trained on it. A meticulously cleaned dataset is inclined to produce precise and significant outcomes in various data-driven analyses, ranging from the classification of rice leaf diseases to other analytical tasks.



Fig. 3. Sample images after data augmentation.

IV. SIMULATION RESULTS

```

Out[276]: <matplotlib.image.AxesImage at 0x190cb2f28e0>
0
25
50
75
100
125
150
175
200
0 50 100 150 200
folders
['train\\Healthy',
'train\\Mild Bacterial blight',
'train\\Mild Blast',
'train\\Mild Brownspot',
'train\\Mild Tungro',
'train\\Severe Bacterial blight',
'train\\Severe Blast',
'train\\Severe Brownspot',
'train\\Severe Tungro']

In [282]: a=model.predict(x)
1/1 [=====] - 0s 105ms/step

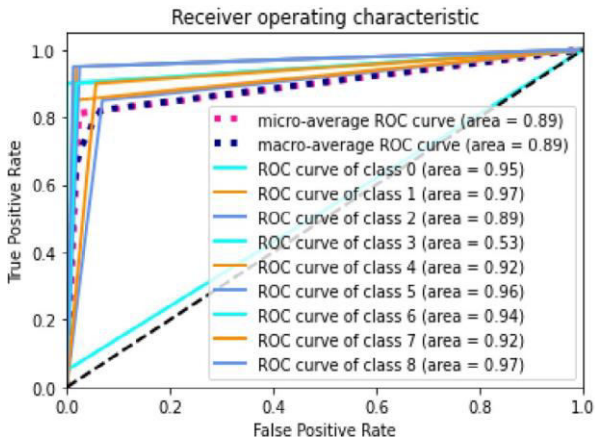
In [283]: a.argmax()
Out[283]: 0
    
```

Fig. 4. Simulation Result

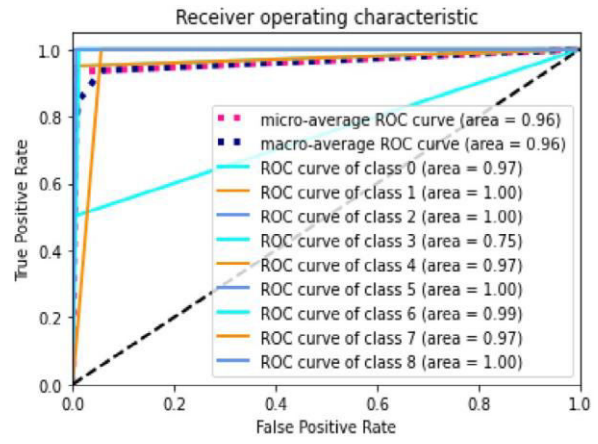
Table 1  
Performance Measure Analysis of all the transfer learning approach.

Model Name	Precision	Recall	F1_score	Accuracy
Xception	0.977	0.977	0.977	97.7%
DenseNet121	0.938	0.938	0.938	93.8%
InceptionResnetV2	0.933	0.933	0.933	93.3%
Resnet 50	0.811	0.811	0.811	81.1%
InceptionV3	0.955	0.955	0.955	95.5%
VGG 16	0.988	0.988	0.988	98.8%

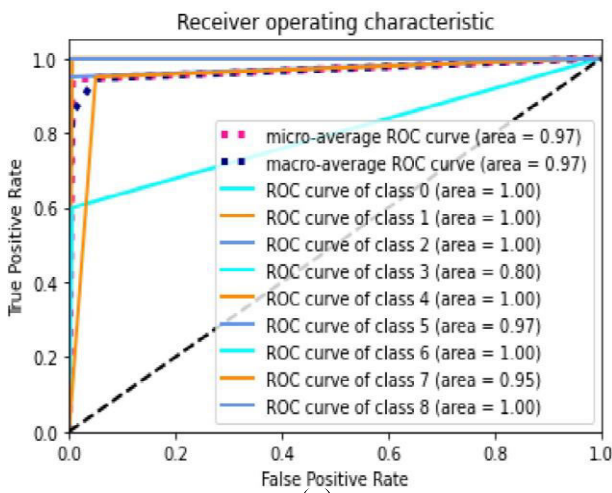




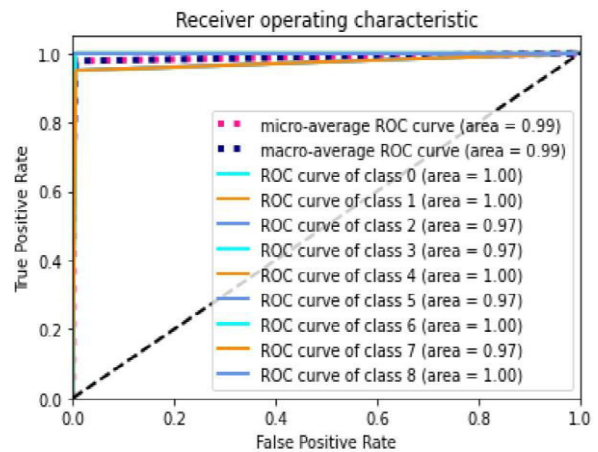
(a)



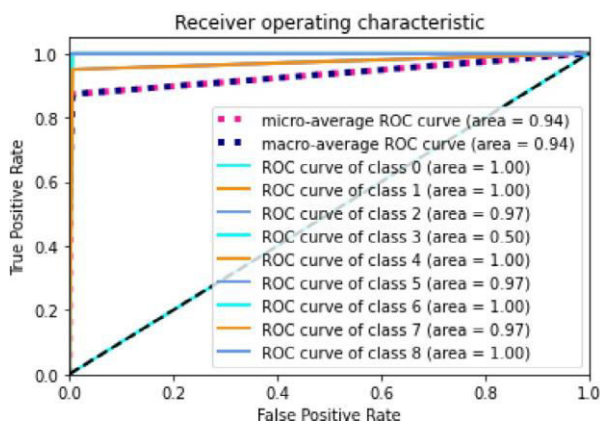
(b)



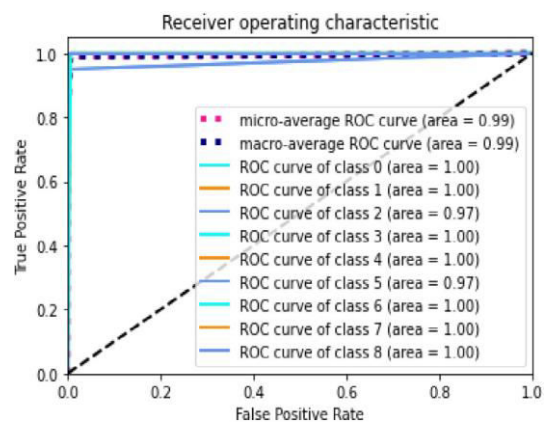
(c)



(d)



(e)



(f)

## V. CONCLUSION AND FUTURE WORK

The outcomes of this investigation show the potential of Deep Learning, and more specifically CNNs, in the early detection and categorization of illnesses affecting rice leaves. The proposed custom VGG 16 model showed high performance in identification and categorization of 9 different disease class labels of rice leaf with improved accuracy of 99.94% and maximum precision and recall scores, providing a quick and reliable way of detecting and diagnosing plant diseases. Six transfer models were selected based on criteria such as model size, parameter size (with a major focus on real-time deployment), and subsequently retrained. Among these models, VGG16 exhibited the most outstanding performance in terms of precision, recall, and accuracy. Following this, we conducted further parameter fine-tuning, which involved the addition of extra filters, dense layers, and dropouts, optimized through trial and error. As a result of these efforts, the proposed CNN demonstrated an exceptional accuracy rate. This high-performing model holds promise for implementation in various agricultural scenarios, where it can be utilized to identify crop diseases based on their severity levels. The study also involved the analysis of confusion matrices and ROC plots for all six selected models, illustrating key metrics such as true positives (TP), true negatives (TN), false positives (FP), false negatives (FN), and the relationship between true positive rate (TPR) and false positive rate (FPR). Moreover, the model's robustness and reliability were ensured by using data augmentation techniques and pre-trained models. The contribution of this research to agriculture is significant, as it can help farmers detect plant diseases at an early stage, allowing them to take necessary measures and avoid crop losses.

In future, the proposed custom VGG 16 architecture can be applied to other crop diseases, contributing to the development of more effective and efficient disease diagnosis tools in agriculture. The study also highlighted the negative impacts of rice leaf diseases on crop yields, food security, and poverty, emphasizing the urgency to find effective solutions. The traditional methods of disease detection have been found to be time-consuming, unreliable, and often require the expertise of trained personnel, making them inaccessible to small-scale farmers. The proposed model's high accuracy and efficiency make it a promising tool for disease diagnosis, ultimately benefiting farmers and promoting sustainable crop production. This study's findings, therefore, contribute significantly to the emerging field of precision agriculture, where DL-based approaches are increasingly being used to improve crop yield and quality.

## REFERENCES

1. Aggarwal, M., Khullar, V., Goyal, N., 2022. Contemporary and futuristic intelligent Technologies for Rice Leaf Disease Detection. 2022 10th international conference on reliability, Infocom technologies and optimization (trends and future directions), ICRITO 2022. <https://doi.org/10.1109/ICRITO56286.2022.9965113>.
2. Aggarwal, M., Khullar, V., Goyal, N., 2023a. Exploring classification of Rice leaf diseases using machine learning and deep learning. Proceedings of 2023 3rd international conference on innovative practices in technology and management, ICIPTM 2023. <https://doi.org/10.1109/ICIPTM57143.2023.10117854>.
3. Aggarwal, M., Khullar, V., Goyal, N., Alammari, A., Albahar, M.A., Singh, A., 2023b. Light-weight federated learning for Rice leaf disease classification using non independent and identically distributed images. Sustainability 15 (16), 12149. <https://doi.org/10.3390/SU151612149>.
4. Aggarwal, M., Khullar, V., Goyal, N., Singh, A., Tolba, A., Thompson, E.B., Kumar, S., 2023c. Pre-trained deep neural network-based features selection supported machine learning for Rice leaf disease classification. Agriculture 13 (5), 936. <https://doi.org/10.3390/agriculture13050936>.
5. Andrew, J., Eunice, J., Popescu, D.E., Chowdary, M.K., Hemanth, J., 2022. Deep learning-based leaf disease detection in crops using images for agricultural applications. Agronomy 12 (10), 2395. <https://doi.org/10.3390/AGRONOMY12102395>.
6. Asibi, A.E., Chai, Q., Coulter, J.A., 2019. Rice blast: A disease with implications for global food security. Agronomy 9 (8), 451.
7. Bari, B.S., Islam, M.N., Rashid, M., Hasan, M.J., Razman, M.A.M., Musa, R.M., Nasir, A.F.A., Majeed, A.P.P.A., 2021. A real-time approach of diagnosing rice leaf disease using deep learning-based faster R-CNN framework. PeerJ Computer Science 7, e432. <https://doi.org/10.7717/PEERJ->



CS.432/SUPP-1.

8. Bock, C. H., Chiang, K. S., & Del Ponte, E. M. (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. <https://doi.org/10.1007/S40858-021-00439-Z/TABLES/3>.
9. Damalas, C.A., Koutroubas, S.D., 2016. Farmers' exposure to pesticides: toxicity types and ways of prevention. *Toxics* 4 (1), 1.
10. Deng, R., Tao, M., Xing, H., Yang, X., Liu, C., Liao, K., Qi, L., 2021. Automatic diagnosis of Rice diseases using deep learning. *Front. Plant Sci.* 12, 701038. <https://doi.org/10.3389/FPLS.2021.701038/BIBTEX>.
11. Gopi, S.C., Kishan Kondaveeti, H., 2023. Transfer learning for Rice leaf disease detection. *Proceedings of the 3rd International Conference on Artificial Intelligence and Smart Energy, ICAIS 2023*, pp. 509–515. <https://doi.org/10.1109/ICAIS56108.2023.10073711>.
12. 10073711.
13. Haridasan, A., Thomas, J., Raj, E.D., 2023. Deep learning system for paddy plant disease detection and classification. *Environ. Monit. Assess.* 195 (1). <https://doi.org/10.1007/s10661-022-10656-x>.
14. He, Y., Zhang, W., Ma, Y., Li, J., Ma, B., 2022. The classification of Rice blast resistant seed based on Raman spectroscopy and SVM. *Molecules* 27 (13). <https://doi.org/10.3390/molecules27134091>.
15. Hossain, S.M.M., Tanjil, M.M.M., Ali, M.A., Bin Islam, M.Z., Islam, M.S., Mobassirin, S., Sarker, I.H., Islam, S.M.R., 2020. Rice Leaf Diseases Recognition Using Convolutional Neural Networks. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12447 LNAI, pp. 299–314. [https://doi.org/10.1007/978-3-030-65390-3\\_23](https://doi.org/10.1007/978-3-030-65390-3_23).
16. Javidan, S.M., Banakar, A., Vakilian, K.A., Ampatzidis, Y., 2023. Diagnosis of grape leaf diseases using automatic K-means clustering and machine learning. *Smart Agricultural Technology* 3. <https://doi.org/10.1016/j.atech.2022.100081>.
17. Jones, R.A.C., 2021. Global plant virus disease pandemics and epidemics. *Plants (Basel, Switzerland)* 10 (2), Kaur, P., Harnal, S., Tiwari, R., Upadhyay, S., Bhatia, S., Mashat, A., Alabdali, A.M., 2022. Recognition of leaf disease using hybrid convolutional neural network by applying feature reduction. *Sensors* 22 (2). <https://doi.org/10.3390/s22020575>.
18. Kumar, V.S., Jaganathan, M., Viswanathan, A., Umamaheswari, M., Vignesh, J., 2023. Rice leaf disease detection based on bidirectional feature attention pyramid network with YOLO v5 model. *Environmental Research Communications* 5 (6), 065014. <https://doi.org/10.1088/2515-7620/ACDECE>.
19. Li, X., Fang, J., an, & Li, H., 2017. Exponential adaptive synchronization of stochastic memristive chaotic recurrent neural networks with time-varying delays. *Neurocomputing* 267, 396–405. <https://doi.org/10.1016/j.neucom.2017.06.049>.
20. LIU, Z.Y., HUANG, J.F. and TAO, R.X., 2008. Characterizing and estimating fungal disease severity of rice brown spot with hyperspectral reflectance data. *Rice Science*, 15(3), pp.232-242.
21. Lu, Y., Yi, S., Zeng, N., Liu, Y., Zhang, Y., 2017. Identification of rice diseases using deep convolutional neural networks. *Neurocomputing* 267, 378–384. <https://doi.org/10.1016/J.NEUCOM.2017.06.023>.
22. Mekha, P., Teeyasuksaet, N., 2021. Image classification of Rice leaf diseases using random Forest algorithm. *2021 Joint 6th International Conference on Digital Arts, Media and Technology with 4th ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunication Engineering, ECTI DAMT and NCON 2021*, pp. 165–169. <https://doi.org/10.1109/ECTIDAMTNCN51128.2021.9425696>.



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  [ijircce@gmail.com](mailto:ijircce@gmail.com)



[www.ijircce.com](http://www.ijircce.com)

Scan to save the contact details