Internasional Journal of Integrative Sciences (IJIS)
Vol.4, No.1, 2025: 141-150

Application of Random Forest for Rice Plant Disease Classification

Perani Rosyani[1*], Anang Muhamad Lutfi[2], Eko Purwadi[3], Kamaluddin[4], Yusuf Ali Hanaan[5], Ines Heidiani Ikasari[6]
Universitas Pamulang
Corresponding Author: Perani Rosyani dosen00837@unpam.ac.id

A R T I C L E I N F O

Keywords: Rice Leaf Disease, Random Forest, Machine Learning, Disease Lassification, Agricultural Productivity

Received : 2 November
Revised : 20 December
Accepted: 19 January

A B S T R A C T

Indonesia's agricultural sector faces significant challenges in maintaining rice production due to land conversion, pest attacks, and poor irrigation. Early detection of rice leaf diseases is critical to mitigating these challenges. This study applies the Random Forest (RF) algorithm to classify three rice leaf diseases: Bacterial Leaf Blight, Brown Spot, and Leaf Smut. The proposed method achieved an accuracy of 75%, demonstrating its effectiveness in disease detection. This research provides a foundation for integrating machine learning to improve crop management and agricultural productivity

DOI: https://doi.org/10.55927/ijis.v4i1.13477
E-ISSN: 2985-9050
https://journal.formosapublisher.org/index.php/ijis

INTRODUCTION

Rice is one of the most important staple foods in the world, particularly in Asia. However, rice production is frequently threatened by various plant diseases, which can significantly reduce crop yields. Among these diseases, leaf diseases such as Bacterial Leaf Blight, Brown Spot, and Leaf Smut are major contributors to yield loss. According to studies, rice diseases are responsible for a substantial percentage of crop failure, with leaf diseases alone accounting for up to 30% of potential yield loss. Effective disease management often requires early detection, which can be challenging due to the visual similarity between different types of leaf diseases and the limited availability of trained experts in rural farming areas.

Recent advancements in machine learning, particularly Random Forest (RF) algorithms, have shown promise in automating the identification and classification of plant diseases. RF, known for its robustness and ability to handle large datasets with high dimensionality, has been successfully applied in various plant disease detection systems. The ability to automate disease identification in rice plants would allow farmers to address issues promptly, minimizing crop loss and ensuring food security.

This article aims to explore the application of the Random Forest algorithm in classifying rice leaf diseases. By leveraging the publicly available rice leaf disease dataset on Kaggle, this research will examine the efficacy of RF in distinguishing between Bacterial Leaf Blight, Brown Spot, and Leaf Smut. The purpose of this article is to demonstrate how machine learning techniques, specifically Random Forest, can aid in the rapid and accurate identification of these diseases, thus offering an innovative approach to improving rice crop management and mitigating the effects of plant diseases on rice production.

LITERATURE REVIEW

Research Findings

This study successfully implemented the Random Forest classification model to identify three types of rice leaf diseases: Bacterial Leaf Blight, Brown Spot, and Leaf Smut. The dataset consisted of 120 images, with 40 images for each class, resized to a uniform size of 128x128 pixels. The main objective was to classify these diseases effectively and evaluate the model's performance based on standard metrics such as accuracy, precision, recall, and F1-score.

The results showed that the Random Forest model achieved an overall accuracy of 75% on the test dataset. Detailed evaluation metrics are provided in the fo l lowing table:

Table 1. Detailed Evaluation Metrics

Class	Precision	Recall	F1-Score	Support
Bacterial leaf blight	0.78	0.88	0.82	8
Brown spot	0.86	0.60	0.71	10
Leaf smut	0.62	0.83	0.71	6
Overall Accuracy			**0.75**	**24**

1. Classification Performance

- The Random Forest model performed well in classifying Bacterial Leaf Blight, achieving the highest F1-score of 0.82. This indicates that the model is proficient in identifying this disease with both high precision and recall.
- For Brown Spot, the model demonstrated the highest precision of 0.86, but the recall was lower (0.60), suggesting that the model struggles to correctly identify all instances of this class.
- The classification of Leaf Smut showed the lowest precision (0.62) but achieved a high recall (0.83), indicating that the model identifies most of the Leaf Smut cases but includes some false positives.

2. Theoretical Insights

- Random Forest is known for its ability to handle classification problems with imbalanced datasets and noisy data. The results align with the theory, showing good overall accuracy despite the relatively small dataset size.
- The variability in precision and recall across classes suggests that some disease features are more distinguishable than others. For instance, the visual features of Bacterial Leaf Blight might be more distinct compared to Brown Spot or Leaf Smut.

3. Comparison with Previous Findings

- Previous studies using deep learning models like Convolutional Neural Networks (CNNs) reported higher accuracies, but they required larger datasets and computational resources.
- This study demonstrates that Random Forest, a simpler method, can still achieve satisfactory results when computational efficiency is a priority, especially in small-scale datasets.

METHODS

1. Random Forest

Random Forest is a machine learning algorithm based on ensemble learning that is often used for classification tasks, including detecting diseases in rice plants. This algorithm works by combining multiple decision trees trained on different data samples using bootstrap techniques. The final prediction is made based on majority voting (for classification) from the results of these trees (Rigatti, 2017).

One of the main advantages of Random Forest is its ability to handle data with numerous variables, automatically detect interactions between predictors, and reduce the risk of overfitting compared to a single decision tree (Schonlau & Zou, 2020). Moreover, this algorithm is flexible as it can work on non-linear data and deliver high accuracy (Anang et al., 2024)

In the context of rice plant disease classification, studies have shown that Random Forest achieves excellent accuracy, such as reaching 99.65% in classifying rice leaf diseases (Anang et al., 2024) This makes it a recommended method for early detection of rice diseases to support the improvement of quality and productivity in agricultural yields.

2. Research Instruments

The primary research instrument in this study was the Random Forest classifier, a supervised machine learning algorithm known for its robust performance in classification tasks. The dataset used was obtained from Kaggle, containing images of three rice leaf diseases: Bacterial Leaf Blight, Brown Spot, and Leaf Smut. Each class had 40 images, resulting in a total of 120 images.

3. Data Collection Processes

The dataset was downloaded from Kaggle and consisted of images classified into three categories: Bacterial Leaf Blight, Brown Spot, Leaf Smut.

1. Bacterial Leaf Blight

Bacterial Leaf Blight (BLB) is one of the most destructive diseases of rice, caused by Xanthomonas oryzae pv. oryzae. This disease is prevalent in tropical and temperate rice-growing regions, especially in Asian countries. BLB initially manifests as water-soaked lesions at the leaf tips and margins, which eventually turn yellow and necrotic, often resulting in severe yield losses. The disease can cause up to 60% yield reduction under favorable conditions (Naqvi et al., 2019)

The management of BLB involves a combination of cultural practices, resistant varieties, and chemical treatments. Biological control using plant extracts and eco-friendly agents has shown potential in suppressing the pathogen, offering an environmentally sustainable approach

2. Brown Spot

Brown spot disease in rice plants is caused by the fungus Drechslera oryzae (synonym: Helminthosporium oryzae). This disease can cause significant damage to both the quality and quantity of the harvest. Its characteristic symptoms include light reddish-brown lesions or spots with a gray center surrounded by a dark to reddish-brown margin and a bright yellow halo. The infection primarily affects leaves but can also spread to other parts of the plant, such as stems, panicle branches, and grains (Mau et al., 2020).

Environmental factors, such as water scarcity, nutrient-deficient soil, and climate change (e.g., prolonged drought), increase the prevalence and severity of this disease. Research has noted that brown spot has caused yield losses of up to 90%, such as during the Bengal famine of 1942–1943. Additionally, a positive linear relationship has been identified between disease severity and yield loss.

3. Leaf Smut

Leaf smut, caused by the fungus Ustilaginoidea virens, primarily affects the panicle and grains of rice plants. This disease is characterized by the formation of dark green to black smut-like spore balls that replace healthy grains within the panicle. These spore balls lead to significant yield losses, sometimes as high as 30%, and affect grain quality. The disease spreads through infected seeds, soil, or airborne spores, making early detection and prevention critical (Mannepalli et al., 2024).

Management strategies include using disease-free seeds, practicing crop , and maintaining proper field sanitation. Additionally, adequate plant spacing and ensuring good drainage can reduce disease incidence. Early and effective intervention is key to minimizing the economic impact of leaf smut on rice production.

4**. Data Analysis Processes**

The analysis process consisted of the following steps:

1. Image Preprocessing:
   - were resized to 128x128 pixels to reduce computational complexity while retaining disease features.
   - Images The pixel values of each image were flattened into one-dimensional arrays for compatibility with the Random Forest classifier.
2. Dataset Splitting:
   - The dataset was divided into training and testing sets using an 80:20 ratio to evaluate the model's performance.
3. Model Training:
   - A Random Forest classifier was trained using the training dataset.
   - Hyperparameters such as the number of estimators and maximum depth were set to their default values for simplicity.

4. Model Evaluation:

- The trained model was evaluated using the test dataset. Metrics such as accuracy, precision, recall, and F1-score were calculated to measure the model's performance.
- A confusion matrix was generated to visualize the classifier's predictions.

RESULTS

Step 1. Dataset Preparation

Description : The dataset used, sourced from Kaggle.com, consists of 120 images of rice plants divided into 3 disease classes (Bacterial leaf blight, Brown spot, and Leaf smut). Each class contains 40 images.

Action :

- Organize directories for each class.
- Resize images to 128x128 pixels to speed up the process.

Table 1. Summary of Dataset Structure

Class	Number of image
1. Bacterial leaf blight	40
2. Brown spot	40
3. Leaf smut	40
Total	120

Step 2. Image Preprocessing and Feature Extraction

Description : The images are converted into numerical arrays, and feature extraction is performed by flattening the images into one-dimensional vectors (flattened arrays).

Action :

Use NumPy to flatten the images and arrange them in a format that can be processed by machine learning models.

Resized Image: 128x128 pixel image.

Flattened Image: Vector with a length of 16,384 (128x128 = 16,384 pixels).

Step 3. Splitting the Dataset

Description : The dataset is divided into two parts : training data (80%) and testing data (20%).

Action :

- Use train_test_split from scikit-learn to split the data.

Table 2. Train-Test Split

Dataset	Number of image
Training	96
Testing	24
Total	120

Step 4. Model Training

Description: A Random Forest model with 100 trees is used to train the data.

Action :

- Train the model using the training data to classify images based on the disease.

Step 5. Model Evaluation

Description : The evaluation is performed using the testing data. The metrics used are accuracy, precision, recall, and F1-score.

Action :

- Calculate metrics for each class and for the overall model.

Table 3. Evaluation Metrics (Confusion Matrix Report)

Class	Precision	Recall	F1-Score	Support
Bacterial leaf blight	0.78	0.88	0.82	8
Brown spot	0.86	0.60	0.71	10
Leaf smut	0.62	0.83	0.71	6
Accuracy	**0.75**			**24**
Macro Average	**0.75**	**0.77**	**0.75**	**24**
Weighted Average	**0.77**	**0.75**	**0.75**	**24**

Summary of the Steps and result

- **Data Preparation:** The dataset was divided into 3 classes, with 120 processed images.
- **Model:** A Random Forest model with 100 trees successfully classified the data with 75% accuracy.
- **Evaluation:** The model performed well in identifying "Bacterial leaf blight" but showed lower performance on "Brown spot."
- **Future Work:** Improvements are needed through data augmentation, hyperparameter optimization, or employing deep learning models.

CONCLUSIONS AND RECOMMENDATIONS

Performance of the Random Forest Model

- The Random Forest model used for classifying rice plant diseases (Bacterial leaf blight, Brown spot, and Leaf smut) achieved an accuracy of 75%.
- Based on the evaluation, the model showed the best performance in classifying the Bacterial leaf blight class (F1-score: 0.82), while the lowest performance was observed in the Brown spot class (F1-score: 0.71).

Dataset Characteristics

- The dataset used was relatively small (120 images), which may have limited the model's ability to fully recognize patterns.

Impact of Image Resizing and Preprocessing

- Resizing the images to 128x128 pixels improved computational efficiency without compromising the model's ability to detect basic patterns in the images.

Applicability of the Model

- This model can be used as a tool for the rapid identification of rice plant diseases, especially in cases with similar image resolution.

Recommendations

Dataset Augmentation

- To improve performance, the dataset size can be increased using image augmentation techniques (such as rotation, flipping, and light intensity adjustments). This will help the model recognize variations in patterns more effectively.

Model Improvement

- The use of other machine learning methods, such as Convolutional Neural Networks (CNN), is recommended to improve classification accuracy, especially for image-based datasets. CNNs are specifically designed to better recognize visual patterns compared to Random Forests.

Feature Engineering

- Exploring additional features, such as texture, color, or leaf shape, could help the model understand the data in more depth.

Field Deployment and Validation

- The trained model should be tested on images captured directly from the field to ensure that it performs well on images outside the training dataset.

Integration into Agricultural Tools

- The results of this research can be implemented in mobile applications or web-based systems to assist farmers in detecting rice plant diseases in real-time.

Vol.4, No.1, 2025: 141-150

Collaboration with Experts

- Collaborating with agronomy experts and farmers to validate model predictions and provide feedback for further development.

FURTHER STUDY

This study has certain limitations, and future research could aim to overcome these by enhancing the dataset through augmentation techniques to improve size and variability. Investigating advanced deep learning methods, such as CNNs, could further refine accuracy, while optimizing the parameters of the Random Forest model may yield better results. Testing the model on real-world agricultural data, extending its capability to identify additional rice diseases, and implementing it into mobile or IoT-based tools could significantly boost its applicability. Collaboration with experts in agriculture is recommended to ensure the system's practicality in real-world scenarios.

ACKNOWLEDGMENT

We would like to express our gratitude to the Kaggle community for sharing and maintaining the dataset used in this research. The availability of this open-source dataset has been instrumental in achieving the objectives of our study.

We also extend our heartfelt thanks to our advisor, **Ms. Perani Rosyani**, for her invaluable guidance, support, and encouragement throughout this research.

REFERENCES

Mannepalli, P. K., Pathre, A., Chhabra, G., Ujjainkar, P. A., & Wanjari, S. (2024). Diagnosis of bacterial leaf blight, leaf smut, and brown spot in rice leafs using VGG16. Procedia Computer Science, 235, 193–200. https://doi.org/10.1016/j.procs.2024.04.022

Mau, Y. S., Ndiwa, A. S. S., & Oematan, S. S. (2020). Brown spot disease severity, yield and yield loss relationships in pigmented upland rice cultivars from east nusa tenggara, indonesia. Biodiversitas, 21(4), 1625–1634. https://doi.org/10.13057/biodiv/d210443

Muhamad Lutfi, A., Purwadi, E., Ali Hanaan, Y., & Rosyani, P. (2024). Literatur Review: Penerapan Random Forest untuk Klasifikasi Penyakit Tanaman Padi. OKTAL : Jurnal Ilmu Komputer Dan Science, 3(10). https://journal.mediapublikasi.id/index.php/oktal

Naqvi, S. A. H., Umar, U. ud D., Hasnain, A., Rehman, A. ur, & Perveen, R. (2019). Effect of Botanical Extracts: A Potential Biocontrol Agent for Xanthomonas Oryzae Pv. Oryzae, Causing Bacterial Leaf Blight Disease of Rice. Pakistan Journal of Agricultural Research, 32(1), 59–72. https://doi.org/10.17582/journal.pjar/2019/32.1.59.72

Rigatti, S. J. (2017). Random Forest. In JOURNAL OF INSURANCE MEDICINE Copyright C 2017 Journal of Insurance Medicine J Insur Med (Vol. 47). http://meridian.allenpress.com/jim/article-pdf/47/1/31/1736157/insm-47-01-31-39_1.pdf