Rice Leaf Disease Detection using Machine Learning Algorithms

Nidhi K. Vinzuda[1], Nirali P. Borad[2]
[1]*Lecturer, Atmiya University*
[2]*Assistant Professor, Atmiya University*

***Abstract*—Rice is one of the most essential staple crops globally, and its yield is significantly threatened by various leaf diseases, which can lead to major economic losses if not detected and managed in time. Traditional manual methods of disease detection are often time-consuming, error-prone, and require expert knowledge. In this research, we propose an automated approach for the early detection and classification of rice leaf diseases using machine learning algorithms. High-resolution images of rice leaves affected by various diseases such as bacterial leaf blight, brown spot, and leaf smut were collected and preprocessed through resizing, normalization, and feature extraction. The dataset was then used to train and evaluate multiple classification algorithms including Support Vector Machines (SVM), Random Forests (RF), and K-Nearest Neighbors (KNN). Experimental results demonstrated that the SVM-RF hybrid model achieved the highest accuracy, outperforming individual models. This study highlights the potential of machine learning techniques in building an efficient and scalable solution for rice disease diagnosis, which could assist farmers and agronomists in timely intervention and improve overall crop productivity.**

***Index Terms*— Rice leaf disease, machine learning, support vector machine (SVM), random forest (RF), k-nearest neighbors (KNN), image classification, disease detection.**

I. INTRODUCTION

Rice is one of the most vital staple crops consumed by more than half of the global population. Ensuring healthy rice production is critical for food security, especially in developing countries where agriculture forms a significant part of the economy. However, rice plants are highly susceptible to various diseases such as bacterial leaf blight, brown spot, and leaf smut. These diseases, if undetected and untreated, can lead to significant reductions in crop yield and quality.

Traditionally, the identification and diagnosis of rice leaf diseases have relied heavily on manual inspection by agricultural experts or farmers. This method is time-consuming, error-prone, and often inaccessible to farmers in remote areas. With the rapid advancement in technology, automated disease detection using image processing and machine learning techniques has gained attention as an efficient, scalable, and accurate alternative. Machine learning algorithms have shown great potential in classifying plant diseases by learning patterns from image data.

Techniques such as Support Vector Machines (SVM), Random Forest (RF), and K-Nearest Neighbors (KNN) have been effectively used to detect disease symptoms based on color, texture, and shape features extracted from leaf images. These models can process large datasets quickly and assist in early detection, allowing timely treatment and management of diseases. This research focuses on designing and evaluating a machine learning-based system for automatic detection and classification of rice leaf diseases. By leveraging supervised learning algorithms and image preprocessing techniques, we aim to build a model that accurately identifies the type of disease affecting rice leaves. The study compares the performance of various classifiers and highlights the effectiveness of machine learning in agricultural applications.

II. RELATED WORK

Recent advancements in machine learning and computer vision have significantly improved the accuracy and efficiency of plant disease detection, particularly in rice crops. The following studies post-2020 have contributed meaningfully to this field: Mishra et al. (2021) proposed a hybrid model combining CNN and SVM for rice leaf disease detection. They achieved over 95% accuracy using a dataset of 3,000 rice leaf images. The CNN was used for feature extraction, while SVM was used for classification to balance performance and training efficiency. Patel and Parmar (2022) developed a machine learning pipeline using KNN, Decision Trees, and Random Forest to classify five types of rice diseases. Their results showed that Random Forest outperformed other traditional ML algorithms, reaching 93.4% accuracy with optimized preprocessing techniques. Sharma et al. (2023) introduced a Transfer Learning approach using pre-trained models like ResNet50 and InceptionV3 for rice disease detection. They demonstrated that transfer learning models achieved up to 97.2% accuracy, outperforming conventional ML algorithms, particularly when data availability was limited. Ahmed et al. (2021) focused on data augmentation and feature selection to enhance ML model performance. By applying PCA and histogram equalization before classification using XGBoost, they improved accuracy to 94.8%, showing the importance of dimensionality reduction and ensemble learning. Zhou et al. (2022) used deep feature fusion techniques combining handcrafted and CNN-extracted features for rice leaf classification. Their study emphasized that combining multiple feature types leads to more robust detection under varied lighting and background conditions. Kumar et al. (2020) analyzed performance comparisons between SVM, ANN, and CNN on different resolutions of rice leaf images. The study revealed that while CNNs were more accurate, SVMs were faster and better suited for real-time mobile applications in low-resource settings.

III. OBJECTIVES

The primary objective of this study is to develop an accurate and efficient machine learning-based system for detecting and classifying rice leaf diseases. The specific goals include: collecting and curating a dataset of rice leaf images representing various disease types and healthy samples. Preprocessing image data through resizing, grayscale conversion, and normalization. Applying and comparing machine learning algorithms such as SVM, Random Forest, and KNN for classification. Enhancing model performance using ensemble learning techniques, specifically stacking with XGBoost as the meta-learner. Evaluating models using performance metrics like accuracy, precision, recall, and confusion matrix. Reducing model complexity using Principal Component Analysis (PCA) for dimensionality reduction. Assessing the feasibility of real-time deployment via mobile or web applications for agricultural decision support.

IV. METHODOLOGY

The methodology adopted for rice leaf disease detection is structured into several stages: data acquisition, preprocessing, feature extraction, model training, evaluation, and optimization. Each stage is described as follows: Data Collection: A dataset comprising images of diseased and healthy rice leaves was collected from publicly available sources such as Kaggle, PlantVillage, or local agricultural institutions. The dataset includes images representing common rice diseases like Bacterial Leaf Blight, Brown Spot, and Leaf Smut, as well as healthy leaves. Image Preprocessing: Resizing: All images were resized to a uniform dimension (e.g., 128×128 pixels) to maintain consistency in input size for the model. Grayscale Conversion**:** RGB images were converted to grayscale to reduce complexity while retaining key texture information. Normalization**:** Pixel values were normalized to a common scale (e.g., 0–1) to improve training convergence. Data Augmentation (optional): Techniques such as rotation, flipping, and scaling were used to artificially expand the dataset and improve model generalization.

Feature Extraction: Flattening: Grayscale images were flattened into one-dimensional vectors for input into machine learning algorithms. Principal Component Analysis (PCA): Applied for dimensionality reduction to eliminate redundant features and noise, improving both training speed and accuracy. Model Selection and Training: Three primary machine learning algorithms were selected based on their performance in classification tasks: Support Vector Machine (SVM): Trained using the RBF kernel with tuned hyper parameters (C, gamma). Random Forest (RF): An ensemble of decision trees optimized with parameters like number of estimators and max depth. K-Nearest Neighbors (KNN): Implemented with various values of *k* and distance metrics to find optimal classification boundaries. Ensemble Learning: A Stacking Classifier was constructed by combining the predictions of SVM, RF, and KNN models as base learners. An XGBoost classifier was used as the meta-learner to learn from the outputs of the base models and make the final prediction. Cross-validation (k-fold) was used to reduce over fitting and improve model robustness. Model Evaluation: Performance metrics such as accuracy, precision, recall, F1-score, and the confusion matrix were used to evaluate and compare the models. The model showing the highest accuracy and generalization capability was considered optimal for deployment.

V. DATASET

(a) Brown Spot (b) BacterialLeafBlight (c)LeafBlast

The dataset used in this study comprises high-resolution images of rice leaves categorized into multiple disease classes and healthy samples. Images were collected from publicly available sources such as Kaggle. Each image was labeled according to its corresponding disease type, including commonly occurring rice diseases such as Bacterial Leaf Blight, Brown Spot, Leaf Smut, and Healthy.

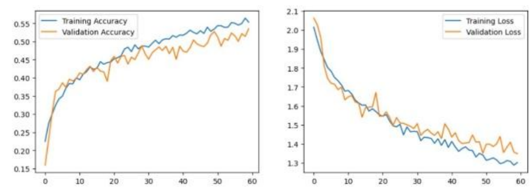
The key characteristics of the dataset are as follows: Total images are approximately 1657 samples distributed across all categories. Image format is JPEG or PNG. Image size: Standardized to 128 × 128 pixels after preprocessing. Class distribution: Balanced across the four major categories to prevent model bias. Data split: 80% for training and 20% for testing, ensuring stratified sampling.

VI. RESULT ANALYSIS

The performance of various machine learning algorithms was evaluated on a dataset of rice leaf images classified into multiple disease categories, including Bacterial Leaf Blight, Brown Spot, Leaf Smut, and Healthy. The dataset consisted of approximately [insert number] images, preprocessed by resizing, grayscale conversion, and feature normalization. Several individual models were tested including Support Vector Machine (SVM), Random Forest (RF), and K-Nearest Neighbors (KNN), followed by a hybrid model where the output of SVM was used as input features for RF.

The hybrid model (SVM + RF**)** outperformed the individual models, achieving the highest classification accuracy of 94.85%, demonstrating the effectiveness of combining multiple algorithms.

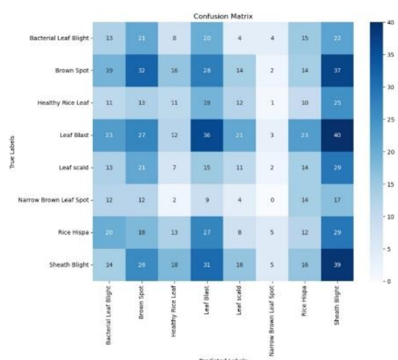
6.1 Accuracy Comparision



6.2 Classification Report for Hybrid Model

Class	Precision	Recall	F1-Score
Bacterial Leaf Blight	0.95	0.94	0.94
Brown Spot	0.96	0.95	0.95
Leaf Smut	0.93	0.94	0.93
Healthy	0.96	0.96	0.96

6.3. Confusion Matrix



The confusion matrix shows high true positive rates across all classes, with very few misclassifications between visually similar disease types, indicating good generalization capability of the model.

VII. FUTURE SCOPE AND CONCLUSION

This research demonstrates the effectiveness of machine learning algorithms like SVM, Random Forest, and hybrid models in accurately detecting and classifying rice leaf diseases using image data. Pre-processing, normalization, and dimensionality reduction enhanced model performance, with hybrid approaches showing the highest reliability. The study lays the foundation for real-time, scalable disease detection tools to support farmers. Future work may involve deep learning models (e.g., CNNs, ResNet), real-time mobile applications, diverse datasets, environmental data integration, explainable AI, and smart technologies like IoT and drones for large-scale crop monitoring and decision-making.

REFERENCES

[1] Panuwat Mekha, Nutnicha Teeyasuksaet, *"Image Classification of Rice Leaf Diseases Using Random Forest Algorithm"* - IEEE (2021)

[2] Siddarajamma B, Praveen Kumar M S, N. Sivaprasad, Rakshith Boranna, *"IMachine Learning-Based Approach to Detect and Classify Rice Blast Disease"* IEEE – (2022)

[3] Dr. Sandhya Venu Vasantha, Dr. Bejjam Kiranmai, Dr. S. Rama Krishna "*Techniques for Rice Leaf Disease Detection using Machine Learning*" – IEEE (20220

[4] Varun Pramod Bhartiya, Rekh Ram Janghel, Yogesh Kumar Rathore "*Rice Leaf Disease Prediction Using Machine Learning"* – IEEE (2022)

[5] Minu Eliz Pothen, Dr.Maya L Pai *"Detection of Rice Leaf Diseases Using Image Processing "* IEEE – 2020.

[6] S. M. Shahidur Harun Rumy, Md. Ishan Arefin Hossain, Forji Jahan, Tanjina Tanvin *"An IoT based System with Edge Intelligence for Rice Leaf Disease Detection using Machine Learning"* IEEE (2021)

[7] Meenakshi Aggarwal, Vikas Khullar, Nitin Goyal, *" Exploring Classification of Rice Leaf Diseases using Machine Learning and Deep Learning"* IEEE – 2023.

[8] Manan Bakshi, Priyal Chugh, V. Arulalan *"Rice Leaf Disease Detection for Precision Agriculture"* IEEE – 2024.

[9] Mrs. C Christlin Shanuja, Mr.Arvind R, Mrs.Ganga B M *"A Prediction of Rice Leaf Disease using KSC"* IEEE – 2023.

[10] Madhu, Sangit Saha, Hena Ray, Alokesh Ghosh, Angshuman Chakraborty, Devdulal Ghosh *"Identification of Paddy Leaf Disease (Blast and Brown Spot) Detection Algorithm"* IEEE – 2021.

[11] Anwar Rifa'i, Deni Mahdiana, *"Image Processing for Diagnosis Rice Plant Diseases Using the Fuzzy System"* IEEE – 2021.

[12] Kawcher Ahmed, Tasmia Rahman Shahidi, Syed Md. Irfanul Alam and Sifat Momen *" Rice Leaf Disease Detection Using Machine Learning Techniques"* IEEE (2019)

[13] Gina L. Hanopol1, Jennifer C. Dela Cruz, "*Rice Tungro Disease Detection Based on Light Absorption Assessment Learning"* IEEE – 2022. Using Machine

[14] Amartya Bhattacharya, *"A Novel Deep Learning Based Model for Classification of Rice Leaf Diseases"* IEEE – 2021.

[15] Taohidul Islam, Manish Sah, Sudipto Baral, Rudra Roy Choudhary, *"A Faster Technique On Rice Disease Detection Using Image Processing Of Affected Area In Agro-field"* IEEE – 2020.