Rice leaf disease prediction using CNN

24EEE431 - AI and Edge computing

Report

Submitted by

CB.EN.U4EEE22041	SAYA SAI KOWSHALYA
CB.EN.U4EEE22062	V SIVA DURGA SANAKR
CB.EN.U4EEE22064	M DHEERAJ REDDY
CB.EN.U4EEE22066	UNAIS I

DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING

AMRITA SCHOOL OF ENGINEERING,

AMRITA VISHWA VIDYAPEETHAM,

COIMBATORE – 641112.

MARCH 2025

CONTENTS:

Introduction:

Rice is one of the most important staple crops globally, feeding more than half of the world's population. However, rice cultivation is highly vulnerable to various diseases that significantly impact yield and quality. Among these, Leaf Blast, Bacterial Blight, and Brown Spot are the most common and destructive rice leaf diseases. Traditional methods of disease detection rely on manual inspection by farmers and agricultural experts, which can be time-consuming, subjective, and prone to errors. Hence, there is a need for an automated and accurate system to identify rice leaf diseases efficiently.

Machine learning (ML) has emerged as a powerful tool in precision agriculture, enabling rapid and reliable disease detection using image processing techniques. By leveraging ML models, particularly deep learning architectures like Convolutional Neural Networks (CNNs), it is possible to classify rice leaf diseases with high accuracy. These models analyze images of infected leaves, extract key features, and predict the type of disease, providing farmers with timely insights for disease management.

This study focuses on developing an ML-based system for detecting and classifying rice leaf diseases using image data. The proposed approach includes image preprocessing, feature extraction, model training, and evaluation using various machine learning techniques. The ultimate goal is to create a robust and scalable solution that can be deployed as a web or mobile application, helping farmers diagnose diseases in real time.

Problem Statement:

Rice is a crucial staple crop that supports the livelihoods of millions of farmers worldwide. However, rice plants are highly susceptible to various diseases, such as Leaf Blast, Bacterial Blight, and Brown Spot, which can cause significant yield losses if not detected and managed in time. Traditional disease detection methods rely on manual inspection by farmers and agricultural experts, which is often subjective, time-consuming, and prone to human error. Additionally, early-stage symptoms of these diseases can be difficult to distinguish, leading to delayed or incorrect treatment.

With the advancement of artificial intelligence and machine learning, automated disease detection systems have the potential to revolutionize rice farming. However, there is still a gap in developing an efficient, accurate, and user-friendly machine learning model that can classify rice leaf diseases with high precision. The challenge lies in preprocessing the image data, selecting the most effective ML model, and ensuring real-time deployment for practical field use.

This study aims to address these challenges by developing a machine learning-based system for rice leaf disease detection. The proposed solution leverages deep learning techniques to classify different rice leaf diseases accurately, providing farmers with an automated tool for early disease identification. By implementing such a system, this research seeks to reduce crop losses, improve decision-making in disease management, and promote sustainable agricultural practices.

Overview of the Dataset:

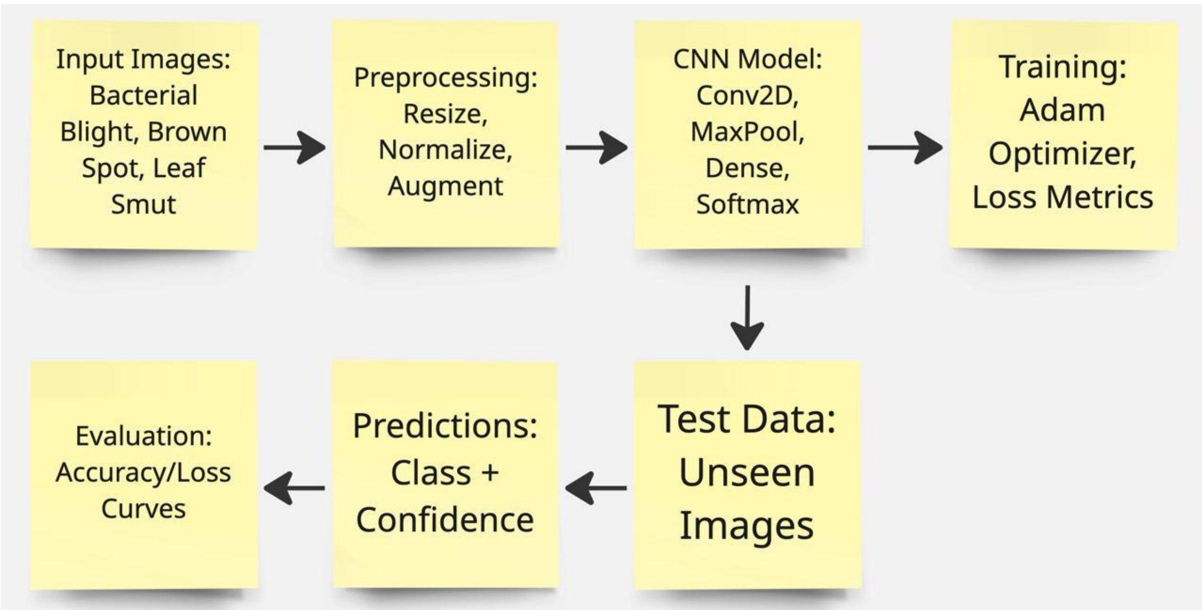
The dataset used in this study consists of rice leaf images categorized into three major disease classes: Bacterial Leaf Blight, Brown Spot, and Leaf Smut. Each category contains 40 images, making a total of 120 images in the dataset. These images capture visible symptoms of the diseases, which can be utilized for training and testing machine learning models for classification. The dataset provides a balanced distribution across all three classes, ensuring that the model does not develop bias toward a specific disease type.

The images are organized into separate folders for each disease category, making it convenient for preprocessing and model training. Since early detection of these diseases is critical for preventing crop losses, this dataset serves as an essential resource for developing an automated rice leaf disease prediction system using machine learning.

The dataset used in this study consists of 120 images of rice leaves categorized into three major disease classes: Bacterial Leaf Blight, Brown Spot, and Leaf Smut, with each category containing 40 images. These images depict visible symptoms of the diseases, making them valuable for training machine learning models for automatic classification. The dataset is well-structured, with each disease type stored in a separate folder, simplifying the preprocessing and labeling process.

The images in this dataset vary in terms of lighting conditions, background noise, and leaf orientations, which helps in making the model more robust for real-world applications. To improve the model’s performance, preprocessing techniques such as resizing, normalization, and data augmentation can be applied. The balanced distribution of images across the three disease categories ensures that the model does not develop bias towards any particular class, leading to a more reliable prediction system.The dataset is partitioned into train (80%), validation (10%), and test (10%) sets.

Methodology:



Problem statement:

- To prepare a complete data analysis report on the given data.
- To create a model which can classify the three major attacking diseases of rice plants like leaf blast, bacterial blight and brown spot.
- Analyze various techniques like Data Augmentation etc and create a report on that.

Domain analysis and Data Information:

Rice is one of the most important staple crops worldwide, providing sustenance to over half of the world's population. However, rice plants are vulnerable to several diseases that can severely impact the yield and quality of rice production. Some of the major rice diseases that affect rice plants include bacterial leaf blight, brown spot, and leaf smut.

This dataset contains 120 jpg images of disease-infected rice leaves. The images are grouped into 3 classes based on the type of disease. There are 40 images in each class.

Classes :

● Leaf smut ● Brown spot ● Bacterial leaf blight

Leaf smut caused by the fungus Entyloma oryzae, is a relatively less common rice disease. The symptoms of leaf smut include the formation of small, round, and reddishbrown spots on the leaves. These spots later turn black and produce powdery spores. Although leaf smut does not usually cause significant yield losses, it can affect the quality of rice grains by reducing their weight and size .

Brown spot, caused by the fungus Cochliobolus miyabeanus, is another major rice disease that can cause significant yield losses. The disease is characterized by small, oval to elliptical spots on the leaves, which turn brown with a yellow halo. In severe cases, the spots can coalesce and cause the leaves to wither and die. Brown spot can also affect the panicles, leading to a reduction in grain quality and yield .

Bacterial leaf blight, caused by the bacterium Xanthomonas oryzae pv. oryzae, is a serious disease that can cause extensive damage to rice plants. The symptoms of bacterial leaf blight include water-soaked lesions on the leaves, which later turn brown and dry up. In severe cases, the disease can cause wilting and death of the plant, leading to a significant reduction in crop yield .

Creating function to split the data for training, testing and validation:

Dividing dataset into 3 parts. Namely:

- Training: Dataset to be used while training
- Validation: Dataset to be tested against while training
- Test: Dataset to be tested against after we trained a model.

Chache, shuffle and prefetch the dataset:

Catche -

Instead of recomputing these values for each training iteration, the CNN can access and reuse them, which significantly speeds up training. This is especially helpful in CNNs where data augmentation or complex preprocessing is involved.

Prefetch -

During CNN training, there are often periods where the model is computing gradients while waiting for the next batch of data to load. Prefetching allows us to start loading the next batch in the background while the model is training on the current batch. This overlapping of operations helps keep the GPU or CPU busy and can lead to faster training times.

Shuffle -

If we feed the CNN consecutive batches of data that are highly correlated (e.g., all images of one class followed by another), the model may learn patterns based on this order, which can lead to poor generalization. Shuffling ensures that each batch contains a random mix of samples from different classes, making the training process more robust.

Creating a Layer for Resizing and Normalization:

Before we feed our images to network, we should be resizing it to the desired size. Moreover, to improve model performance, we should normalize the image pixel value (keeping them in range 0 and 1 by dividing by 256). This should happen while training as well as inference. Hence we can add that as a layer in our Sequential Model.

You might be thinking why do we need to resize (256,256) image to again (256,256). You are right we don't need to but this will be useful when we are done with the training and start using the model for predictions. At that time somone can supply an image that is not (256,256) and this layer will resize it.

Data Augmentation:

Data augmentation is a technique used to increase the size and diversity of a dataset by applying various transformations to the existing data. This technique has become an essential tool in computer vision and image processing tasks, such as object recognition and classification, due to its ability to enhance the generalization ability of machine learning models and prevent overfitting.

In the context of our rice leaf disease image classification project, data augmentation can play a crucial role in improving the performance and robustness of our model. By generating new images with different variations such as rotations, flips, zooms, and other transformations, we can increase the diversity of our dataset and provide our model with more examples to learn from, which can lead to better classification accuracy and robustness to variations in the real-world data.

Model Architecture:

- Here use a CNN coupled with a Softmax Activation in the output layerand also add the initial layers for Resizing, Normalization and Data Augmentation.

Compiling the Model:

- We use adam Optimizer, SparseCategoricalCrossentropy for losses, accuracy as a metric.

Summary:

- The project involved building a deep learning model to classify images of rice leaf disease. The dataset contained images from 3 type of rice leaf disease namely bacterial bligh, brown spot, and leaf smut. The project was divided into several steps, including data exploration, data preprocessing, building and training deep learning models, and evaluating model performance.
- During data exploration, we analyzed the dataset and visualized the images to get a better understanding of the

data. We observed that the dataset was balanced, with an equal number of images for rice leaf diseases. We also noticed that the images were of different sizes and needed to be resized to a uniform size before being used for training. We have rescaled them to uniform size of 256 X 256.

- For data preprocessing, We have normalized the training dataset . We used Keras' ImageDataGenerator to generate augmented images to increase the size of the dataset and reduce overfitting. We also resized the images and divided them into training, validation, and testing sets.
- We built deep learning models by applying transfer learning technique, we trained model on augmented datasets and evaluated their performance using accuracy, loss, and time taken per epoch.
-

Problem Challenge :

Limited amount of data:

- One of the biggest challenges in the project was the limited amount of data available for training the models. We had only 119 images from 3 classes. This could lead to overfitting, where the model memorizes the training data and does not generalize well to new data. To deal with this challenge, data augmentation techniques were used to artificially increase the size of the dataset. This helped to improve the performance of the models and reduce overfitting.

Complexity of deep learning models:

- Deep learning models can be very complex and have a large number of parameters, which makes them difficult to train and optimize. To deal with this challenge, various techniques such as regularization, and early stopping were used to prevent overfitting and improve the performance of the models.

Conclusion:

• Rice leaf diseases pose a significant threat to global rice production, affecting crop yield and food security. Traditional disease detection methods are often time-consuming, subjective, and prone to human error, making early identification and management challenging. In this study, a machine learning-based approach has been proposed for the automated detection and classification of rice leaf diseases, focusing on Bacterial Leaf Blight, Brown Spot, and Leaf Smut. By utilizing image preprocessing techniques, deep learning models, and performance evaluation metrics, the system aims to provide an accurate and efficient disease classification solution.

• The expected outcomes of this research include the development of a robust CNN-based model capable of classifying diseases with high accuracy. Furthermore, the deployment of this model as a web or mobile application will enable farmers to quickly identify diseases and take necessary preventive actions. By integrating artificial intelligence into precision agriculture, this system can help minimize crop losses, reduce excessive pesticide use, and improve overall rice productivity.

Reference:

[1] Prajapati, H. B., Shah, J. P., & Dabhi, V. K., “Detection and Classification of Rice Plant Diseases”, Intelligent Decision Technologies, IOS Press 2017, 357–373.

[2] Q. Yao, Z. Guan, Y. Zhou, J. Tang, Y. Hu and B. Yang, "Application of Support Vector Machine for Detecting Rice Diseases Using Shape and Color Texture Features," 2009 International Conference on Engineering Computation, Hong Kong, 2009, pp. 79-83.

[3] Srdjan Sladojevic, Marko Arsenovic, Andras Anderla, Dubravko Culibrk, and Darko Stefanovic, “Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification,” Computational Intelligence and Neuroscience, vol. 2016, Article ID 3289801, 11 pages, 2016.

[4] Konstantinos P. Ferentinos, “Deep learning models for plant disease detection and diagnosis,” Computers and Electronics in Agriculture,Volume 145, 2018, Pages 311-318.

[5] Hsieh, Jia-You, Wei Huang, Hsin-Tieh Yang, Chia-Chieh Lin, YoChung Fan and Huan Chen. “Building the Rice Blast Disease Prediction Model based on Machine Learning and Neural Networks.” (2019).

[6] Dharmar, Vydeki. (1970), “Application of machine learning in detection of blast disease in South Indian rice crops.” Journal of Phytology. 11. 31-37. 10.25081/jp.2019.v11.5476.

[7] Xu, Guili, Fengling Zhang, Syed Ghafoor Shah, Yongqiang Ye, and Hanping Mao. “Use of leaf color images to identify nitrogen and potassium deficient tomatoes.” Pattern Recognition Letters 32, no. 11 (2011): 1584-1590.