FINAL PROJECT REPORT

PADDY DISEASE CLASSIFICATION

C0887501 - Anju Krishna Thekkeveettil
C0887502 - Athish Jobin
C0884599 - Jiya Maria Eapen
C0883113 - Roshin Baby
C0885273 - Roy Joseph

PROBLEM STATEMENT

Rice is one of the basic foods across the world. Paddy, the raw grain before removal of husk, is cultivated in tropical climates, primarily in Asian countries.
Paddy cultivation requires consistent supervision because several diseases and pests might affect the paddy crops, leading to up to 70% yield loss. Expert supervision is usually necessary to mitigate these diseases and prevent crop loss. With the limited availability of crop protection experts, manual disease diagnosis is tedious and expensive. Thus, it is increasingly important to automate the disease identification process by leveraging computer vision-based techniques that achieve promising results in various domains.

OBJECTIVE

The main objective is to develop a deep learning-based model to classify the given paddy leaf images accurately. We provide a training dataset of labeled images across ten classes (nine disease categories and normal leaf).

INTRODUCTION

The backbone of India's economy is agriculture, with the country ranking as the world's second-largest producer and largest exporter of rice. However, India faces several growing concerns such as a rising population and significant climate changes that impact agricultural output. With the population expanding, there's a pressing need for advancements in agriculture using the latest technologies to ensure better crop yields. A highly efficient and cost-effective method is required for the rapid detection of crop diseases to maintain productivity.

Rice is a major crop in India, susceptible to various diseases caused by parasites, bacteria, and viruses. Farmers struggle to identify these diseases on rice crops, leading to a lack of proper preventive measures. Several medical interventions such as pesticides or insecticides are available to manage crop diseases and boost production. However, accurate diagnosis and effective treatment require expert assessment or prior knowledge. Diseases on plants manifest through specific symptoms on leaves, serving as indicators of the actual disease. Leveraging these indicators, we developed an AI model, an advanced and cost-effective technology that provides accurate results in a shorter time frame for detecting rice crop diseases and offering appropriate remedies like insecticides or pesticides. We focused on nine rice crop diseases that frequently lead to significant losses in paddy crop production. Additionally, our model can also identify healthy leaves.

Some common rice crop diseases that frequently lead to significant losses in paddy crop production are

1. Leaf Blast

Leaf blast is considered the most severe and damaging disease in rice plants, affecting different parts of the leaves including the neck and leaf blades. It is particularly common in areas with intermittent rainfall, cool temperatures, and low soil moisture. Leaf blast can impact a paddy crop at any growth stage and has the potential to destroy entire leaves once developed. Images (a), (b), (c), and (d) illustrate leaf blast-affected rice plant leaves.

Fig. 1 Leaf Blast affected leaves

2. Hispa

This disease is often identified when the mining of the grubs is visibly observed on the leaves. When the area is heavily infested, the paddy crop fields appear consumed. Severe infection in rice plants leads to shriveling of the damaged leaves. Typically, this disease affects plants in their early stages of growth. Managing this disease often involves avoiding excessive fertilization of the area. The following figures depict leaves affected by the Hispa disease.

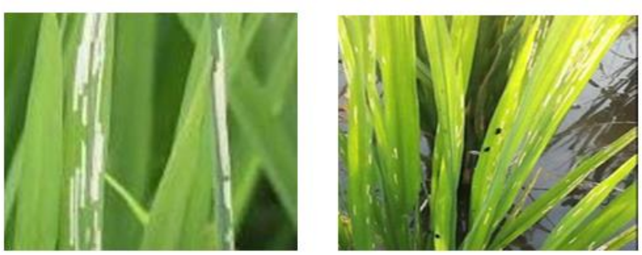


Fig. 2 Hispa affected Rice plant Leaf

3. Brown Spot

Brown spot - In Figure (c), the rice plant leaf exhibits symptoms of brown spot disease, characterized by dark spots on the leaves of the paddy crop. This fungal disease is often identified by signs such as seedling death, large areas of leaf death, and the presence of brown or dark spots. It leads to both quantitative and qualitative losses, with a 5% yield reduction observed across South and Southeast Asia. Preventing this disease involves ensuring the crop receives the correct balance of nutrients and avoiding water stress. Additionally, treating seeds with chemicals can help reduce the risk of infection.

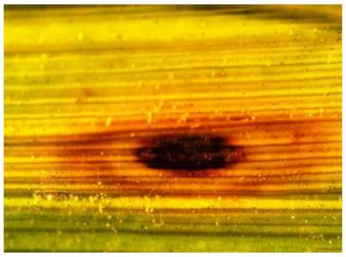


Fig. 3 Brown Spotted Rice Plant Leaf

Literature Review

This model is designed to detect diseases in paddy crops such as rice blast and Bacterial leaf blight. The system consists of two main phases: one for training the model and the other for detecting diseases in input images. During the first phase, the model is trained using an image dataset comprising both healthy and diseased paddy leaf images. The dataset includes 10407 images of paddy leaf images sourced from the Kaggle website. The training process employs the Convolutional Neural Network (CNN) Algorithm to process and learn from these images.

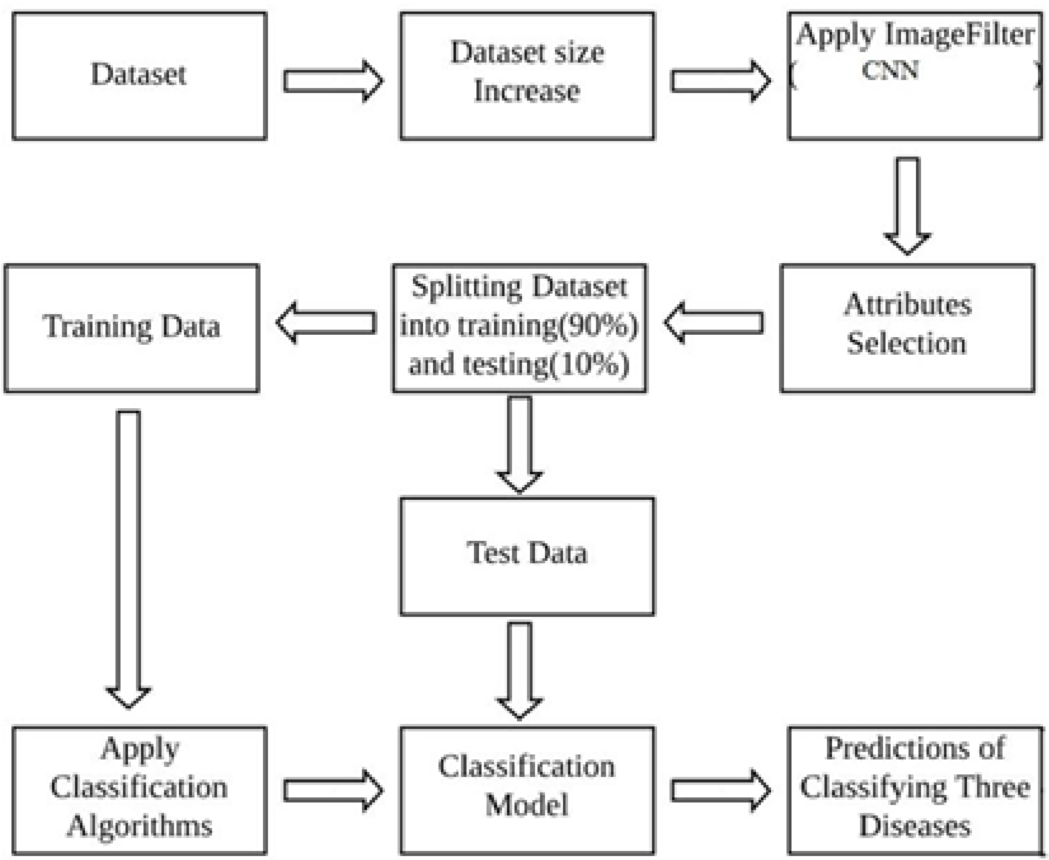
The model 1 focuses on detecting and categorizing plant diseases using a machine learning technique. This paper introduces the use of transfer learning with deep CNN to classify diseases in paddy crops. The researchers divided the entire dataset into various ratios for training and testing purposes. Their model achieved a classification accuracy of 97% with an 80%-20% training-validation partition.

The complex CNN Model of paddy leaf diseases successfully classifies accuracy of 97% with an 80%-20% training-validation partition.

Overview of the Methodology Used

The project follows a structured methodology for classifying paddy diseases using convolutional neural networks (CNNs). Initially, the dataset is loaded from a CSV file and preprocessed to handle missing values and duplicates. Visualizations are created to understand the distribution of plant varieties and disease labels in the dataset. Subsequently, the dataset is split into training and validation sets. Two models are built using TensorFlow/Keras: Model 1 consists of multiple dense layers with ReLU activation functions, while Model 2 is a CNN comprising convolutional layers, max-pooling layers, dropout regularization, and fully connected dense layers. Both models are compiled with appropriate loss functions and optimizers before being trained on the training dataset. Early stopping is utilized to prevent overfitting during training. The trained models are then evaluated using the validation dataset, and Model 2 is saved for future use. Finally, inference is performed on a sample image using the saved model to make predictions. This comprehensive methodology covers data loading, preprocessing, model building, training, evaluation, and deployment, ensuring a systematic approach to paddy disease classification.

A convolutional neural network is a feed-forward neural organization that is generally used to breakdown visual pictures by handling information with a network like geography. It's otherwise called a ConvNet. A convolutional neural organization is utilized to distinguish and characterize objects in the picture.



A convolutional neural network contains several hidden layers that aid in extracting information from an image. The four key layers in CNN are as follows:

- Convolution layer
- ReLU layer
- Pooling layer
- Completely Associated layer

Convolution Layer: This marks the starting stage in the process of extracting key components from an image. A convolutional layer contains multiple channels that carry out convolution operations. Each image is viewed as a matrix of pixel values.

ReLU layer : ReLU stands for rectified linear unit. Once the feature maps have been extracted, the next step is to pass them through a ReLU layer.ReLU conducts an element-wise operation, setting all negative pixels to zero. It introduces non-linearity into the network, and the resulting output is the rectified feature map.

Pooling Layer: Pooling is a downsampling operation that reduces the dimensionality of the feature map. The rectified feature map then undergoes pooling to produce a pooled feature map.

Fully Connected Layer: The pooled feature map is flattened and fed into a fully connected layer to obtain the final output.

DATA PREPROCESSING

Data Loading:

The dataset is loaded from the 'train.csv' file, which contains information about images, disease labels, plant varieties, and age.

Handling Missing Values and Duplicates:

The dataset is checked for missing values, and none are found. Duplicates are also checked and removed if present.

```
[6]: data.isnull().sum()
```

```
[6]: image_id    0
     label       0
     variety     0
     age         0
     dtype: int64
```

```
[7]: # Check for duplicates
     data.duplicated().sum()
```

```
[7]: 0
```

Dataset Split:

The dataset is split into training and validation sets using TensorFlow's image_dataset_from_directory utility. This ensures separate datasets for training the model and evaluating its performance.

Normalization:

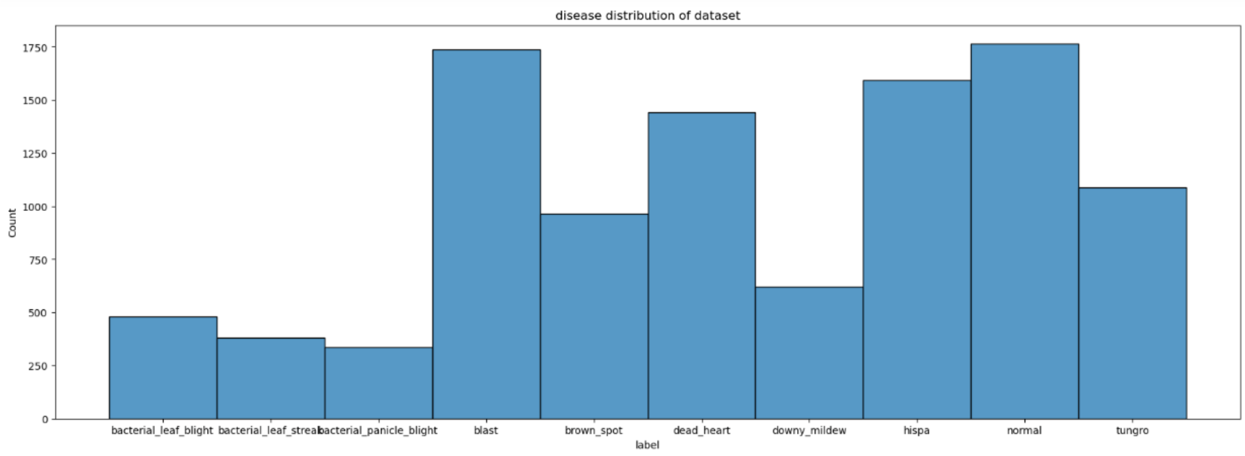
This normalization is achieved using a Rescaling layer from TensorFlow's Keras API, which rescales the pixel values of the images. The normalization layer is defined with a scaling factor of 1./255, which divides all pixel values by 255. By applying this normalization, the pixel values of the input images are transformed from the original range of [0, 255] to the normalized range of [0, 1]. This process ensures that the input pixel values are on a consistent scale, facilitating better convergence and stability during the training of machine learning models. After applying the normalization layer to the dataset (train_ds), the first batch of normalized images and their corresponding labels are retrieved for verification. The minimum and maximum pixel values of the first normalized image are printed, confirming that the pixel values now fall within the desired range of [0, 1]. This normalization step enhances the effectiveness and efficiency of subsequent model training and optimization processes classification.

VISUALIZATION

Variety Distribution:
A histogram is created to visualize the distribution of plant varieties in the dataset. This provides insights into the frequency of different plant varieties present, helping to understand the dataset's composition and potential biases.

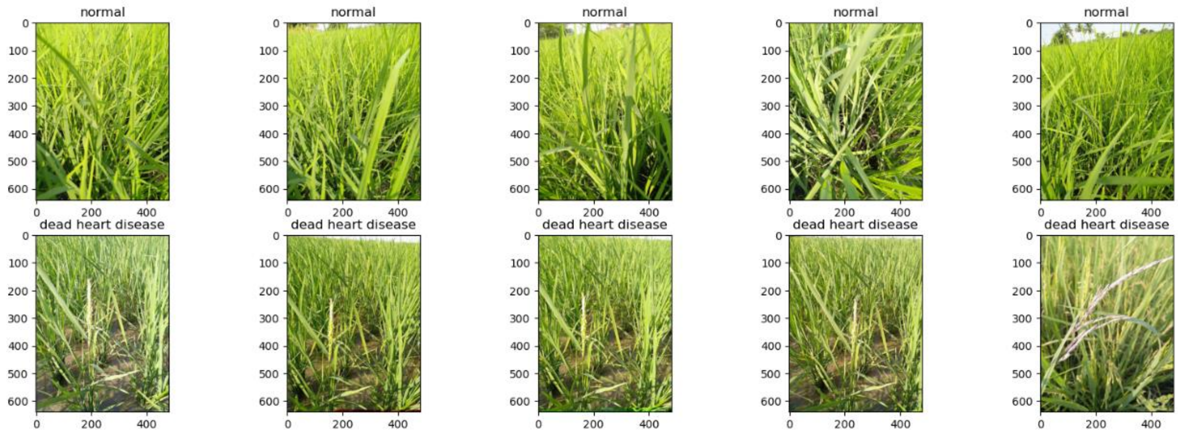
Disease Distribution:

Another histogram is generated to visualize the distribution of disease labels in the dataset. This allows for an understanding of the prevalence of different diseases across the dataset, aiding in the identification of any imbalances or biases that may exist.



Visualizing Images:

Images of both normal paddy plants and plants affected by dead heart disease are displayed for visual inspection. These images help in understanding the characteristics of different classes and can aid in identifying patterns or features relevant to disease.



PARAMETER TUNING

The parameter tuning was performed for optimizing the efficiency of the data processing and training pipeline using TensorFlow's AUTOTUNE constant, caching, and prefetching techniques.

AUTOTUNE Constant: TensorFlow's autotune constant is utilized to enable automatic tuning of parameters related to dataset processing and training. By dynamically adjusting these parameters based on available computational resources, TensorFlow optimizes the efficiency of the overall pipeline without manual intervention.

Caching the Training Dataset: The cache() method is applied to the training dataset (train_ds). Caching allows TensorFlow to store the elements of the dataset in memory, avoiding redundant data loading and preprocessing operations during training. This optimization significantly improves training speed, particularly with large datasets.

Prefetching the Dataset: The prefetch() method is used to prefetch elements from both the training and validation datasets (train_ds and val_ds, respectively) into memory. Prefetching overlaps the data preprocessing and model execution phases, enabling TensorFlow to fetch data in parallel while the model is training. This overlapping of computation and data transfer reduces training time and maximizes GPU utilization.

Overall, by leveraging AUTOTUNE, caching, and prefetching techniques, TensorFlow optimizes the efficiency of the data processing and training pipeline, leading to faster model convergence and improved overall performance.

CONCLUSIONS

This project introduces an AI approach to detect nine distinct paddy leaf diseases such as: hispa, bacterial leaf blight, and brown spot disease etc. The image classification technique described in this paper leverages the basic architecture of a CNN, which includes several convolutional layers, a pooling layer, and a final fully connected layer. The goal of this project is to aid farmers in the early detection of diseases in rice plants using image processing with convolutional neural networks. Having identified a nearly optimal algorithm, we aim to expand this research further as larger datasets of rice plant diseases become available in the future.

REFERENCES

1. Prajwal Gowda B.S, Nisarga M A, Rachana M, Paddy Crop Disease Detection using Machine Learning (International Journal of Engineering Research & Technology (IJERT), 2020)
2. Siddharth Swarup Rautaray, Manjusha Pandey, Mahendra Kumar Gourisaria, Ritesh Sharma, Sujay Das, Paddy Crop Disease Prediction- A Transfer Learning Technique, (International Journal of Recent Technology and Engineering (IJRTE), March 2020)
3. Dr. Sandhya Venu Vasantha, Dr. Bejjam Kiranmai, Dr. S. Rama Krishna, Techniques for Rice Leaf Disease Detection using Machine Learning Algorithms, (IJERT, 2021)
4. Classification Using Transfer Learning Of Deep Convolution Neural Network, (The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2019)
5. https://www.kaggle.com/code/mubeenalvi/rice-diseases-classification-using-cnn