Business case

Prepare a complete data analysis report on Rice Leaf Diseases and Create a model which can classify the three major attacking diseases of rice plants like leaf blast, bacterial blight and brown spot. By analyzing various techniques like Data Augmentation, etc.

1.Context

Diseases of crops are the main reason for the decrease in yield and product quality. Rice diseases can cause an **annual loss of 10–30% of rice production**, which is the equivalent of feeding 60 million people. Other effects of rice diseases are **Increases rice prices, Economic impact, Threat to food security.** Fungal diseases alone are estimated to reduce annual rice production by 14% globally whereas Bacterial diseases can cause significant losses in rice production, with estimates ranging from 50% to 75% of the crop.

2.Introduction

In this Dataset we have some major and minor rice leaf disease which affect the quality and overall production of rice. By creating a model that can detect the rice disease using the CNN model farmers can separate the diseased crop before the harvesting and take the necessary measure to prevent further.

1. Bacterial leaf blight is a destructive disease which affects millions of rice variety throughout Asia. In Japan alone, annual losses are estimated to be between 22,000 and 110,000 tons. In the Philippines, susceptible varieties lose up to 22.5% of the total harvest during wet seasons and up to 7.2% in the dry season.
2. Brown spot is a fungal disease that can infect both seedlings and mature plants. The disease causes blight on seedlings, which are grown from heavily infected seeds, and can cause 10-58% seedling mortality. Famine” in 1942 resulting to yield losses of 50% to 90% and caused the death of 2 million people.
3. Leaf smut is a minor disease that occurs late in the growing season. It's widespread where rice is grown, but it doesn't often cause serious damage. However, it can make rice vulnerable to other diseases, which can reduce yield

3. Data Overview

This dataset includes 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

* **Bacterial leaf blight** - *Xanthomonas oryzae is a bacterium that* causes a potentially devastating disease **Bacterial Leaf Blight** in rice crops all over the world. Found worldwide in temperate and tropical regions, it can destroy up to 80 % of a crop if the disease develops early. Even if it develops late, it can severely diminish the quality and yield of the grain.
* **Brown spot** - *Cochliobolus miyabeanus* is an important plant pathogen which causes a common and widespread rice disease of **Brown Spot** which in result causes high level of crop yield losses. It was a major cause of the [Bengal famine of 1943](https://en.wikipedia.org/wiki/Bengal_famine_of_1943), where the crop yield was dropped by 40% to 90% and the death of 2 million people was recorded. Other known severe crop loss cases caused by *Cochliobolus miyabeanus* are globally distributed. In India and Nigeria, it can reduce total crop yield by up to 40%.
* **Leaf smut** - It is caused by the fungus Entyloma oryzae, it is a widely distributed but somewhat minor disease of rice as compared to Bacterial Leaf Blight and Brown spot. Smuts are cereal and crop [pathogens](https://en.wikipedia.org/wiki/Pathogen) that most notably affect members of the grass family. Economically important hosts include [maize](https://en.wikipedia.org/wiki/Maize), [barley](https://en.wikipedia.org/wiki/Barley), [wheat](https://en.wikipedia.org/wiki/Wheat), [oats](https://en.wikipedia.org/wiki/Oats), [sugarcane](https://en.wikipedia.org/wiki/Sugarcane), and [forage grasses](https://en.wikipedia.org/wiki/Forage_grasses). Leaf smut is a minor disease that occurs late in the growing season. It's widespread where rice is grown, but it doesn't often cause serious damage. However, it can make rice vulnerable to other diseases, which can reduce yield.

4. EDA

**Bar plot** - By using the bar plot checked if the data is balanced.

**Visualization** – visualised the images from each class.

4. Data Preparation

* **Resizing** - Resized the Imaged to a square shape we ensure that the input dimensions are consistent, i.e. all the input images that need to fed while training the images must be of equal size to make it easier to train the CNN model.
* **Scaling** – Rescaled every image pixel by dividing it with 255. As every pixel value is represented by a range between 0-255, then dividing it with 255 is doing the Min-Max scaling, i.e. it will convert it to range of 0-1.

7. Data Augmentation

Data augmentation is a crucial technique in machine learning, particularly in Convolutional Neural Networks (CNNs), used to artificially expand the training dataset. It generates new training examples by applying random transformations to the original data. Data augmentation is especially beneficial in cases where labelled data is scarce or expensive to obtain.

Augmenting the data can help improve the model's ability, reduce overfitting, and increase robustness.

Transformation Used

* **Rotation**: Randomly rotating images by small angles (e.g., -15° to +15°) helps the model learn rotational invariance.
* **Flipping**: Horizontal or vertical flipping of images helps the model generalize to different object orientations.
* **Translation**: Shifting images along the X or Y axis mimics changes in object position and helps models handle shifting objects in the image.
* **Scaling**: Resizing or zooming into an image enables the model to focus on different levels of object detail.

Advantages

* **Enhanced Performance on Unseen Data**: By exposing the CNN model to a broader range of variations during training, it becomes better at handling input during inference.
* **Cost-Effective**: Instead of collecting new data, augmentation enables the reuse of existing data, saving time and costs associated with manual labelling collection.
* **Flexibility**: Data augmentation can be customized to the specific needs of the model, dataset, and task. Techniques like random rotation, flipping, and scaling can be adjusted to introduce the appropriate level of variability.
* **Reduction in Overfitting**: By adding noise or varying the data, augmentation helps avoid the model from memorizing patterns in the training data, which can lead to poor generalization.

8. Conclusion

Data augmentation is an effective and essential strategy for improving the performance of CNN models in the rice leaf disease classification task. By artificially expanding the dataset through transformations such as rotation, flipping, zooming, and colour adjustments, the model becomes more robust to variations in input data.