Paddy Doctor: Paddy Disease Classification

Project Report

AMMI SENEGAL 2022, Mbour

Team: AFES Paddy **Public score:** 0.97577 **Private score:** NA

Team members :

- Annine Duclaire Kenne,
- Fenosoa Randrianjatovo,
- Elysee Manimpire Gasana,
- Abiodun Saheed Ademola.

1 Introduction

Agriculture is the most important way to feed the population around the world by farming crops such as Rice which is one of the staple foods worldwide and the most consumed. The plantations such as rice, cassava, etc. have often been affected by certain diseases. However, these problems have attracted most of the experts in the machine learning field. Thus, the main goal of this kaggle project is to develop a machine or deep learning-based model to classify the given paddy leaf images accurately. In this project, we have focused specifically on rice plant (Oryza sativa) disease. The images of the diseased symptoms in leaves and stems have been captured from the rice field. The different methods used, our experiments, and the results obtained are presented. The best score was obtained using **Resnet34** with 100 epochs accompanied by some deep layers. In the following sections, we provided additional details about our experiments.

2 Data and EDA

2.1 Data

The datasets used for the project were obtained from the kaggle platform. The data consist of 10,407 (75%) datasets for training and 3,469 (25%) datasets for testing. The labeled paddy leaf images have across ten classes (nine diseases and normal leaf). Those classes are: (a) bacterial leaf blight, (b) bacterial leaf streak, (c) bacterial, (d) panicle blight, (e) blast, (f) brown spot, (g) dead heart, (h) downy mildew, (i) hispa, (j) normal, (k) tungro.

2.2 Exploratory Data nalysis (EDA)

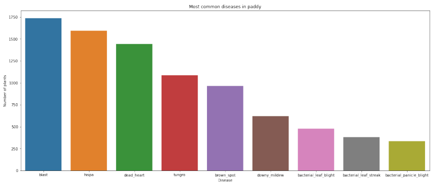
In order to understand the data and identify some patterns, we performed the EDA by plotting some graphs (available in EDA notebook [1]) as below:

The figure 1b shows that most of the diseased crops are over 60 days old, which may mean that the older paddy crops are getting diseased.

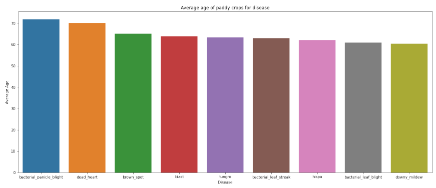
Hence our assumption about older crops being diseased was correct according to figure 1e. Mostly older crops are diseased but to an exception, a few younger crops are also diseased, figure 1e.

From the figure 2, we can notice that ADT45 is the most infected variety in the paddy crops, it may be because ADT45 is the most common choice of the farmers to grow. Let's look more at ADT 45 (figure 1f).

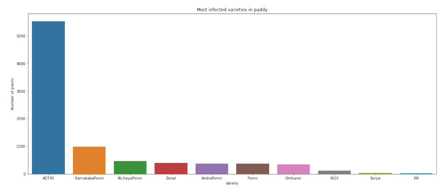
[1] `https://colab.research.google.com/drive/1c-Ixyp-SK5C-Y_zCtSNqEmZdPdlMA5S1?usp=sharing`



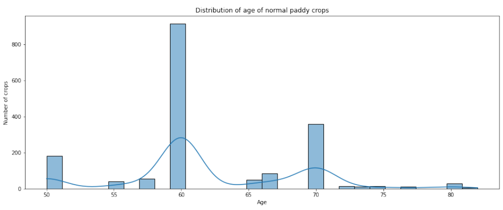
(a) Most common Disease



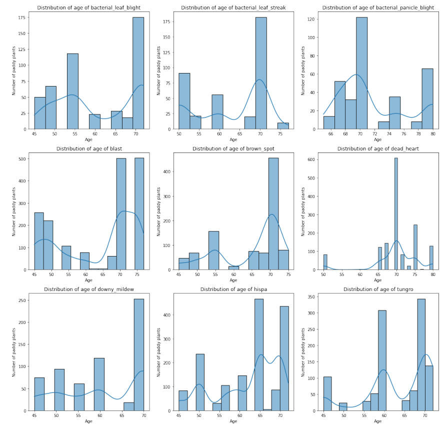
(b) Diseases with average age



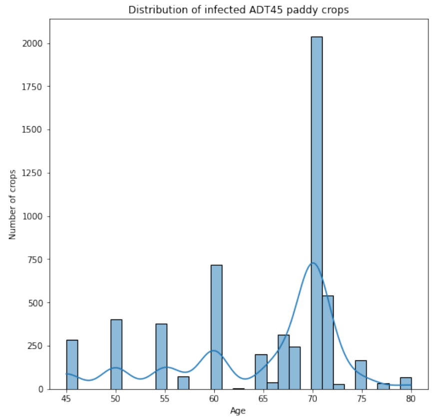
(c) Most Infected variety



(d) Distribution of age of normal paddy crops



(e) Distribution of age of diseased crops



(f) Distribution of infected ADT45 paddy crops

Figure 1

The figure 3 shows some images of diseased crops infected with each disease (figure 3a) and some examples of disease crops in each variety (figure 3b).

3 Model Architecture & Methods

The figure (4) represents the system structure of the model that we are using during this competition.

3.1 Image Preprocessing

To avoid over-fitting of our model, the image data was preprocessed and also data enhancement was done. The short side of the image was scaled to 224 pixels and the long side was scaled proportionally to reduce the computational complexity of the model that will facilitate the model speed. We then apply a random affine transformation to the image, which could be randomly translated, rotate, scaled, deform, and cut the image. Also, Gaussian blur and image flipping were applied randomly. Finally, the resized image was randomly cropped to a 224×224 pixels square area as the actual training image. These processes favored expanding the data set and reducing the over-fitting of the model on the original dataset without modifying the characteristics of rice diseases.

3.2 Model Structure

The structure of the convolutional neural network has a crucial influence on the performance of the final model. It was necessary to compare the performance of different networks in the diagnosis of rice diseases. Two network structures were selected and tested, and they were: ResNet and DenseNet. These networks are described below.

ResNet is a widely used network model, which uses residual blocks to enhance the depth of the CNN. By directly connecting the input and the output, ResNet can reduce the problems of gradient disappearance and gradient explosion, thus deepening the number of network layers and achieving better effects.

DenseNet uses a dense connection, which connects each layer to every other layer. Since DenseNet allows features to be reused, this can generate many features with a small number of convolution kernels. As a result, it can reduce gradient loss and enhance the propagation of features, and the number of parameters is greatly reduced.

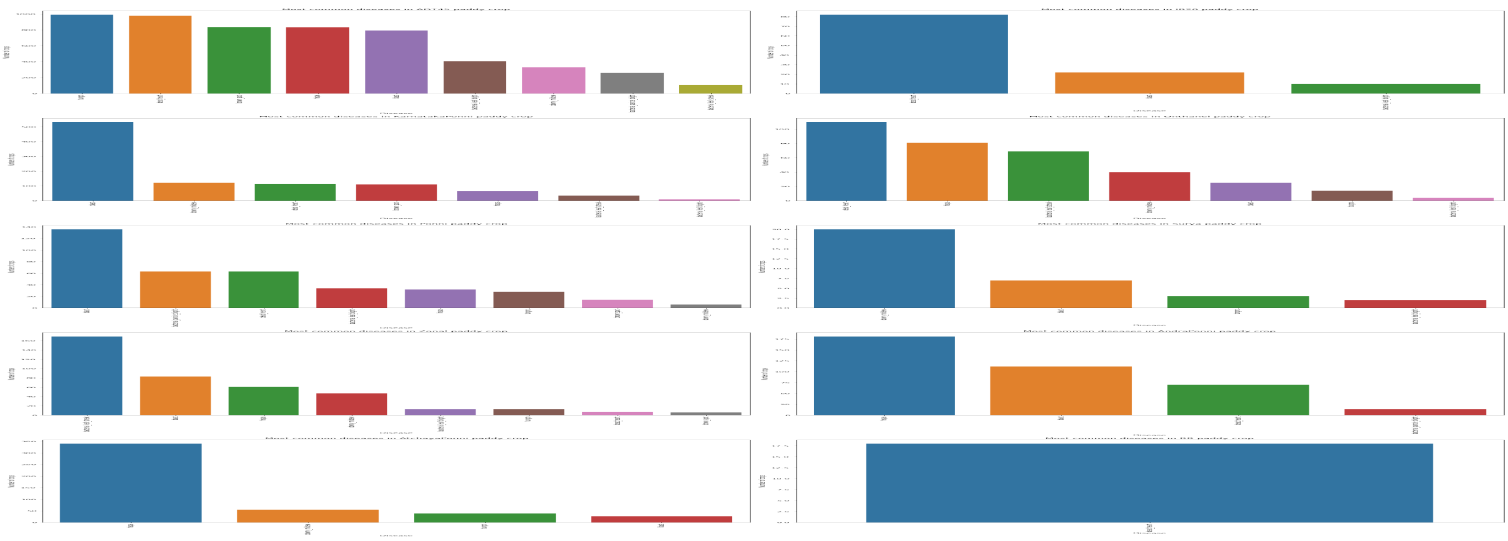


Figure 2: Proportion of diseases in each variety of paddy

(a) Some images of diseased crops infected with each disease

(b) Examples of diseased crops in each variety

Figure 3

4 Results

Table 1 provides the performance of the models, in terms of accuracy at the stated number of epochs.

The ResNet34, 152 and DenseNet121 models have been trained (see Notebook [2]), and from the table 1 we can conclude that those models performed approximately well but the **ResNet34** performed better i.e. was able to classifier with small error the rice plant (Oryza sativa) disease with a testing accuracy of **0.97577**, as it appears in the table 1 and on the kaggle website Leaderboard [3]

[2] `https://colab.research.google.com/drive/1It72IvWwmgPj3RhhZFxB2IzZN2GcBtug?usp=sharing`
[3] `https://www.kaggle.com/competitions/paddy-disease-classification/leaderboard`

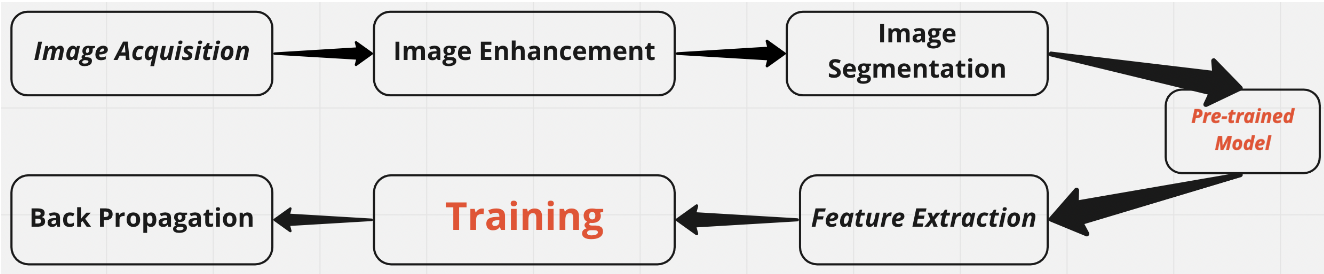


Figure 4: System structure of the model

Models	Epochs	Batch Size	Learning Rate	Loss	Optimizer	Train/Test Accuracy
Resnet34	100	32	0.001	Cross Entropy	Adam	0.9896/0.97577
Resnet152	100	32	0.001	Cross Entropy	SGD	0.9839/0.96539
Densenet121	74	32	0.001	Cross Entropy	Adam	0.9826/0.96270

Table 1: Summary of model performance

5 Conclusion

Finally, we reached the goal of our work, where a Deep Learning model to classify the rice plant disease has been created (Oryza sativa). After performing some exploratory data analysis (EDA) and preprocessing, we tested newly created models but their performance was lower. We made a decision of trying different pre-trained models such as VGG19, Inception V3, Resnet34, Resnet152, and DenseNet152. After many trials, **Resnet34** and **Resnet152** provided the best performance for our task. The higher accuracy was provided by the ResNet34 model with training and testing accuracy of **0.9896/0.97577**, respectively. Therefore, this accuracy could be improved by doing more hyperparameters tuning of our model, training and testing others pre-trained models such as VGG19, Inception V3, Alexnet, Squeezenet, etc. Our source code is available on Github [4].

[4] https://github.com/FenosoaRandrianjatovo/Paddy_Disease_Classification

References

[1] Shrivastava, Vimal K and Pradhan, Monoj K and Minz, Sonajharia and Thakur, Mahesh P *Rice plant disease classification using transfer learning of deep convolution neural network* pages:631-635, 2019.

[2] Goluguri, NV and Devi, K Suganya and Srinivasan, P *Rice-net: an efficient artificial fish swarm optimization applied deep convolutional neural network model for identifying the Oryza sativa diseases* Pages:5869-5884, 2021.

[3] Orillo, John William and Cruz, Jennifer Dela and Agapito, Leobelle and Satimbre, Paul Jensen and Valenzuela, Ira *Identification of diseases in rice plant (Oryza Sativa) using back propagation Artificial Neural Network* Pages:1-6, 2014.