**Sources of Data (experimental design)**

In the study, the researcher will be gathering rice disease datasets that consist of images of bacterial leaf blight, rice tungro, sheath blight, and rice blast. Gathered datasets were utilized by previous studies about leaf disease detection which are located from publicly available source from Kaggle. Researchers will further investigate the gathered images from public available datasets via conducting interview to an expert to verify that collected datasets are according to what type of rice disease it claimed. In addition with the gathered public datasets, researchers will seek assistance to the research aides of the research participants to capture high resolution images of rice crop that has the concern rice diseases which are bacterial leaf blight, rice tungro, sheath blight and rice blast.

Datasets will be divided into three sets which are: training set that contains the most number of images for the model to learn from; validation set to tune model’s hyperparameters and evaluate performance during training to prevent overfitting; and testing set for evaluation of final performance of the model. Rice leaf datasets are significant to the study in training models in classifying healthy and unhealthy leaves to gather performance data from each model that will be further analyze to check the significant difference in the performance in each tool that utilized different evolutionary computation as feature selector.

**Research Instrument**

The study will utilized various evaluation metrics to determine the performance of different models consist of a evolutionary computation as feature selector and an SVM classifier. To generate performance data, once the model was developed, accuracy, precision recall, f-score and true positive ratio and true negative ratio will be collected. Experiment paper will be utilized as a tool to collect and further analyzed gathered performance data.

TP (true positive) – if predicted positive match actual positive

FP (false positive) – if predicted positive (healthy) does not match actual negative

TN (true negative) – if predicted negative match actual negative

FN (false negative) – if predicted negative (diseased) does not match actual positive

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Accuracy | Precision | Recall | F-score | True Positive Ratio and True Negative Ratio | | |
| Predicted | Actual | Evaluation |
| GA + SVM |  |  |  |  |  |  |  |
| PSO + SVM |  |  |  |  |  |  |  |
| ABC + SVM |  |  |  |  |  |  |  |
| ACO + SVM |  |  |  |  |  |  |  |