Identification of a suitable machine learning model for detection of asymptomatic Ganoderma boninense infection in oil palm seedlings using hyperspectral data

ABSTRACT

In Malaysia, oil palm industry has made an enormous contribution to economic and social prosperity. However, it has been affected by basal stem rot (BSR) disease caused by Ganoderma boninense (G. boninense) fungus. The conventional practice to detect the disease is through manual inspection by a human expert every two weeks. This study aimed to identify the most suitable machine learning model to classify the inoculated (I) and uninoculated (U) oil palm seedlings with G. boninense before the symptoms' appearance using hyperspectral imaging. A total of 1122 sample points were collected from frond 1 and frond 2 of 28 oil palm seedlings at the age of 10 months old, with 540 and 582 reflectance spectra extracted from U and I seedlings, respectively. The significant bands were identified based on the high separation between U and I seedlings, where the differences were observed significantly in the NIR spectrum. The reflectance values of each selected band were later used as input parameters of the 23 machine learning models developed using decision trees, discriminant analysis, logistic regression, naïve Bayes, support vector machine (SVM), k-nearest neighbor (kNN), and ensemble modelling with various types of kernels. The bands were optimized according to the classification accuracy achieved by the models. Based on the F-score and performance time, it was demonstrated that coarse Gaussian SVM with 9 bands performed better than the models with 35, 18, 14, and 11 bands. The coarse Gaussian SVM achieved an F-score of 95.21% with a performance time of 1.7124 s when run on a personal computer with an Intel® Core™ i7-8750H processor and 32 GB RAM. This early detection could lead to better management in the oil palm industry.

Keyword: Basal stem rot; Hyperspectral; Near-infrared; Machine learning; Asymptomatic disease; Early detection