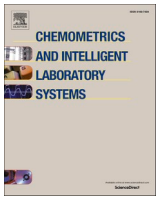
Chemometrics and Intelligent Laboratory Systems 232 (2023) 104718

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemometrics



Ganoderma boninense classification based on near-infrared spectral data using machine learning techniques

Mas Ira Syafila Mohd Hilmi Tan [a], Mohd Faizal Jamlos [a,*], Ahmad Fairuz Omar [b], Kamarulzaman Kamarudin [c], Mohd Aminudin Jamlos [d]

[a] Faculty of Electrical & Electronics Engineering Technology, Universiti Malaysia Pahang, 26600, Pekan, Malaysia
[b] School of Physics, Universiti Sains Malaysia, Penang, Malaysia, 11800, Malaysia
[c] Faculty of Electrical Engineering Technology, Universiti Malaysia Perlis, 02600, Pauh Putra, Perlis, Malaysia
[d] Faculty of Electronics Engineering Technology, Universiti Malaysia Perlis, 02600, Pauh Putra, Perlis, Malaysia

ARTICLE INFO

Keywords:
Oil palms
Near-infrared spectroscopy
Machine learning classification

ABSTRACT

Ganoderma boninense (G. boninense) infection reduces the productivity of oil palms and causes a severe threat to the palm oil industry. Early detection of G. boninense is vital since there is no effective treatment to stop the continuing spread of the disease unless ergosterol, a biomarker of G. boninense can be detected. There is yet a non-destructive and in-situ technique explored to detect ergosterol. Capability of NIR to detect few biomarkers such as mycotoxin and zearalenone (ZEN) has been proven to pave the way an effort to explore NIR's sensitivity towards detecting ergosterol, as discussed in this paper. A compact hand-held NIR with a measurement range of 900–1700 nm is utilized by scanning the leaves of three oil palm seedlings inoculated with G. boninense while the other three were non-inoculated from 16-weeks-old to 32-weeks-old. Significant changes of spectral reflectance have been notified occur at the wavelength of ~1450 nm which reflectance of infected sample is higher 0.2–0.4 than healthy sample which 0.1–0.19. The diminishing of the spectral curve at approximately 1450 nm is strongly suspected to happened due to the loss of water content from the leaves since G. boninense attacks the roots and causes the disruption of water supply to the other part of plant. However, a few overlapped NIRs' spectral data between healthy and infected samples require for further validation which chemometric and machine learning (ML) classification technique are chosen. It is found the spectra of healthy samples are scattered on the negative sides of PC-1 while infected samples tend to be on a positive side with large loading coefficients marked significant discriminatory effect on healthy and infected samples at the wavelength of 1310 and 1452 nm. A PLS regression is used on NIR spectra to implement the prediction of ergosterol concentration which shows good corelation of $R = 0.861$ between the ergosterol concentration and oil palm NIR spectra. Four different ML algorithms are tested for prediction of G. boninense infection: K-Nearest Neighbour (kNN), Naïve Bayes (NB), Support Vector Machine (SVM) and Decision Tree (DT) are tested which depicted DT algorithm achieves a satisfactory overall performance with high accuracy up to 93.1% and F1-score of 92.6% compared to other algorithms. High accuracy shows the capability of the classification model to correctly predict the G. boninense detection while high F1-score indicates that the classification is able to validate the detection of G. boninense correctly with low misclassification rate. The result represents a significant step in the development of a non-destructive and in-situ detection system which validated by both chemometric and machine learning (ML) classification techniques.

1. Introduction

In 2019, oil palm contributed 37.7% to the Gross Domestic Product (GDP) of the agriculture sector in Malaysia followed by other agriculture (25.9%), livestock (15.3%), fishing (12.0%), forestry & logging (6.3%) and rubber (3.0%) [1].

Oil palm industry has a major impact on Malaysia's economy, which generated profitable export earnings. As reported by the Malaysian Palm Oil Board (MPOB) in 2020, Malaysia produced 17.18 tonnes per hectare of oil palm fresh fruit bunch for 5.9 million hectares of the total planted

* Corresponding author.
E-mail address: mohdfaizaljamlos@gmail.com (M.F. Jamlos).

https://doi.org/10.1016/j.chemolab.2022.104718
Received 22 July 2022; Received in revised form 31 October 2022; Accepted 26 November 2022
Available online 1 December 2022

Table 1
Previous studies on metabolites detection using NIRS.

Sample	Metabolites	Instrument	Wavelength (nm)	Models/ Algorithms	Significant Result	Ref
Maize	Mycotoxin, Fusarium toxins, Fumonisins, Penicillium toxins	NIR spectrometer	1100–2500	RF	Classification accuracy: Mycotoxin: 82.2% R&E-Fusarium toxins: 77.9% E-Fusarium toxins: 82.4% Penicillium toxins: 95.1% Fumonisins: 70.5%	[36]
Brazilian maize	Fumonisins (FBs), Zearalenone (ZEN)	NIR spectrometer	400–2500	PLS	FBs: R = 0.809, R^2 = 0.899, RMSEP = 65.9, SEP = 68.2, RPD = 3.33 ZEN: R = 0.991, R^2 = 0.0.984, RMSEP = 69.4, SEP = 69.8, RPD = 2.71	[37]
Grapes	Ergosterol	VIS-NIR spectrometer	VIS: 450-850 NIR: 1050-1650	PLS	R^2 = 0.851 RMSEP = 3.61 RPD = 2.72	[38]
Wheat	Deoxynivalenol (DON), ergosterol	HSI-NIR spectrometer	900–1700	PLS, LDA	PLS performance of ergosterol raw spectra: Whole sample: R^2 = 0.88 Ground sample: R^2 = 0.62 PLS performance of DON raw spectra: Whole sample: R^2 = 0.62 Ground sample: R^2 = 0.62 LDA classification accuracy according to DON concentration: Whole sample: 82.52% Ground sample: 82.53%	[39]
Soil	Ergosterol	NIR spectrometer	700–2500	FCR, FLLR, FRER, FRR	FRER and FRR produce adequate result of CVPE = 0.91 and 0.89, respectively.	[40]

area. Sarawak is the biggest oil palm planted state with 1.59 million hectares, followed by Sabah with 1.54 million hectares. The total area of the other states in Peninsular Malaysia is 2.77 million hectares. Malaysia contributed 20.5% of world palm oil supplies making Malaysia the world's second-biggest palm oil manufacturer and exporter with total export revenue approximately RM67.52 billion [2].

After 30 months of planting, oil palm trees begin to produce fruit and can bear fruits for 20–30 years, thus ensuring a consistent supply of oils. Oil palm is the world's most efficient oil-bearing crop which produces one ton of oil in just 0.26 ha of land while rapeseed, sunflower and soybean require 1.52, 2 and 2.22 ha, respectively, to produce the same amount of oil [3]. However, a serious crop disease called Basal Stem Rot (BSR) disease threatens the sustainability of oil palm production [4]. The total area affected by G. boninense in 2020 was estimated to be about 443,430 ha or 65.6 million palm trees [5]. This disease is the main concern which severely affects Southeast Asia's oil palm plantations, particularly in North Sumatra and Malaysia (Flood et al., 2000). This fungal infection can reduce the yield of oil palm production by 80%, which indirectly causes great losses in oil palm production and affects the palm oil industry. BSR disease is caused by Ganoderma boninense (G. boninense), a white-rot fungus that is a type of pathogenic basidiomycete. However, early detection and control strategies for G. boninense are still undeveloped, although it was identified as the primary cause of death of oil palms. Therefore, early detection and identification of G. boninense infection are crucial to prevent production losses and reduce the cost of plantation management.

In earlier studies, the detection of G. boninense was done using several methods. For instance, visual inspection by evaluating the visible symptoms of the plant [6] and lab-based techniques such as enzyme-linked immunosorbent assay (ELISA) [7], Ganoderma selective medium (GSM) [8], multiplex PCR-DNA kits [9], GanoSken tomography [10], electrochemical DNA biosensors [11]. Visual inspection provides inaccurate information since the plant might show symptoms similar to malnutrition, lack of water or hyperacid soil. Lab-based techniques are somehow tedious, time-consuming, and complex which require sample preparation and DNA extraction of the sample. Recently, spectroscopy techniques to detect G. boninense have been explored since it meets the following criteria: non-invasive, rapid, sensitive and precise to diseases, which have been taken into consideration for the development and design for early detection [12]. Several spectroscopy techniques have been conducted for G. boninense detection, such as nuclear magnetic resonance (NMR) [13], mass spectroscopy (MS) [14], dielectric spectroscopy [15,16], Fourier transform Infrared (FTIR) spectroscopy [17–19], MIR spectroscopy [20], hyperspectral imaging spectroscopy [21] and visible to near-infrared (VIS-NIR) spectroscopy [22,23]. These numerous studies show the ability of spectroscopy techniques paired with classification algorithms led to promising results for the detection of G. boninense infection. However, NMR, MS, DS and FTIR spectroscopy were carried out in the laboratory and were destructive since the samples need to be processed before measurements. VIS/IR spectroscopy has higher accuracy in detecting plant disease than the other spectroscopy methods [24]. Based on the VIS-NIR spectroscopy study by Liaghat et al. [22], the spectral data in the NIR region portray significant differences between classes of samples compared to the VIS region. Liang et al. [25] stated that the VIS region is only useful for visual analysis; thus, it is not useful for asymptomatic detection. The VIS range is dominantly influenced by pigment concentration such as chlorophyll and carotenoids while NIR range is influenced by biomass composition in the infrared range [26]. NIR is an effective method for acquiring information on tree physiology and pathology [27]. Thus, a study based on only NIRS is proposed. NIRS has been used extensively for the rapid detection of organic components [28]. Fungi are the most common plant pathogens, accounting for 85% of all plant illnesses [29]. Rapid detection of fungal phytopathogens is crucial for effective disease control. Ergosterol is the most abundant sterol in fungi cell membranes which is absent in plants or microbial cells [30]. Ergosterol is a primary metabolite in cell wall produced by G. boninense. A study by Mohd As'wad et al. [31] found that ergosterol effectively detects BSR disease in oil palms. Ergosterol has been successfully used as an indicator for fungal biomass in oil palm tissues [31–33]. The unique existence of metabolites makes it a good marker for Ganoderma colonisation in oil palm [34]. Ergosterol can be detected as early as 6 h after inoculation which means that ergosterol can be detected once the fungus starts to colonize the root [35]. Therefore, it may be possible to detect infection based on ergosterol content before symptoms develop. Previous studies had demonstrated the ability of NIRS techniques for detection of metabolites including ergosterol, as summarised in Table 1. Capability/sensitivity of NIR to detect few metabolites biomarkers such as mycotoxin [36] and

zearalenone (ZEN) [37] including ergosterol in grapes [38] is proven to pave the way of an idea to explore/investigate its sensitivity towards ergosterol, a biomarker for G. Boninense.

The most recent study by Ghilardelli et al. [36]utilizes NIRS of 1100–2500 nm to detect the several metabolites incursion of maize. Random forest (RF) is used to classify the samples. Singularly, this approach can classify mycotoxin, regulated or emerging Fusarium-produced (R&E-Fusarium toxins), E-Fusarium toxins, Penicillium toxins, Fumonisins with the accuracy of 82.2%, 77.9%, 82.4%, 95.1% and 70.5%, respectively. This study shows several significant wavelengths at 1430, 1470, 1820, 2140, and 2180 nm which related to total fungal infection and could be assigned to the first overtone of the OH stretching modes of glucose, NH in most amino acids, and CH combination bands in unsaturated fatty acid [41].

Another study by Tyska et al. [37]applied NIRS ranging from 400 to 2500 nm to detect Fumonisins (FBs) and Zearalenone which are toxic secondary metabolites produced by some filamentous fungi in Brazillian maize. Based on FBs spectral data, the wavelength range 400–600 nm and 1900–2500 nm show significant band. As for ZEN, wavelengths between 400 and 500, 1200–1900 nm and 2100–2400 nm appear to be relevant. Partial least squares regression (PLS) regression is used in this study. Correlation coefficient (R), determination coefficient (R2), root mean square error of prediction (RMSEP), standard error of prediction (SEP) and residual prediction deviation (RPD) for FBs and ZEN are 0.809 and 0.991; 0.899 and 0.984; 659 and 69.4; 682 and 69.8; and 3.33 and 2.71, respectively.

A study by Femenias et al. [39]is conducted to estimate the deoxynivalenol (DON) and ergosterol in wheat via NIR hyperspectral image. The PLS performance for estimation of ergosterol for whole sample and ground sample are $R^2 = 0.85$ and 0.57, respectively while the PLS performance for estimation of DON for whole and ground sample are $R^2 =$ 0.56 and 0.49, respectively. Linear discriminant analysis (LDA) classification is only applied to DON samples spectra which correctly discriminate the samples up to 82%.

Detection of ergosterol in grapes for grape rot indication by utilizing VIS-NIR spectrometer is performed by Porep et al. [38]. The PLS model is obtained by combining raw VIS and NIR spectral data which produce good prediction performance with coefficient of determination (R^2), root mean square error of prediction (RMSEP) and ratio of prediction to deviation (RPD) of 0.851, 3.61 and 2.72, respectively. Further feasibility of NIRS for prediction of ergosterol in soil has been explored by Almanjahie et al. [40]. Four predictor models used are functional classical regression (FCR), functional local linear regression (FLLR), functional relative error regression (FRER) and functional robust regression (FRR). The models are evaluated based on the cross-validated prediction error (CVPE). FRER and FRR produce adequate result of CVPE = 0.91 and 0.89, respectively.

NIRS often being favoured over other spectroscopy and analytical methods as has the highest accuracy for disease detection on different types of plants compared to mid-infrared (MIR) and visible to near-infrared (VIS-NIR) spectroscopy [42]. NIRS is more precise and sensitive to diseases include G. boninense compared to VIS light [22]. The VIS region provides information based on colour whereas NIR region is principally attributed to C–H, O–H, and N–H vibrations. These vibrations contain information on the chemical elements, structures, and state of molecules. For early asymptomatic disease detection, NIR region is the main interest as NIR spectral data contain information on the interior tissue while VIS spectral data contain information on the exterior, such as colour and texture [43]. Compared to those in the MIR range, the shorter NIR wavelengths enable increased penetration depth and direct analysis of solid samples, requiring little to no sample preparation [42]. The recent advancement in NIRS instrument development now also enables in-field and on-site analysis with the availability of portable and compact instrumentation [44]. All these advantages along with being chemical-free, rapid, non-destructive and non-invasive make it possible for utilisation of NIRS for complete early detection systems in real-time. A review paper by Mohd Hilmi Tan et al. [45] discussed the utilisation of the NIRS technique and a machine learning classifier that demonstrated NIRS as a feasible method for early asymptomatic detection of G. boninense in oil palm. Previous study by Ahmadi et al. [23] able to discriminate between healthy and asymptomatic infected sample ML on VIS-NIR spectral data. However, the study is conducted on 12-years-old mature oil palm plants. G. boninense can infect oil palm trees at all stages, from seedlings to mature plants [46]. This fungus infects seedlings less than a year old in the nursery [47,48]. This paper proposes a novel, rapid and non-destructive detection of ergosterol in Ganoderma boninense via NIR spectroscopy as early as at nursery level with the utilisation of several machine learning classifiers such as kNN, NB, SVM and DT.

kNN is a simple classifier which widely used for pattern recognition problems technique. It is a lazy learning method based on learning by comparing a given test sample with the available training samples which are similar to it [49]. Its simplicity enables ease of classification [50]. Classification is achieved by i) identifying the nearest neighbours of the trained data, ii) calculating the distance between them and input data, and iii) predicting the class of input data [51]. This classifier is suitable to be implemented on multi-modal classes which a sample can have many class labels [52]. Liaghat et al. [53,54]employed a kNN classifier for G. boninense detection that classified four different classes of palm oil health conditions and generated the highest classification accuracy of 97.3%. kNN has also been implemented on NIR spectra to classify the severity of fungal infection in maize [55] and chestnuts [56].

NB is a simple Bayesian probabilistic classifier based on Bayes decision theorem. Bayes theorem is strong independence assumptions theorem [57]. This assumption is considered naïve as it assumes that the effect of an attribute on a class is statistically independent of all other attributes [58,59]. NB enable prediction of class membership probabilities which determine the probability of a given data item belongs to a particular class label [60]. NB has been increasingly applied for classification due to its efficiency, simplicity and satisfactory performance. NB has been increasingly applied for classification due to its efficiency, simplicity and good performance. Implementation of NB on spectral data has been tested for the detection of G. boninense [53,54].

The SVM has been used in many applications as this classifier is effective and sturdy to noise [61]. The SVM was originally intended for binary classification and was investigated to solve multi-class classification problems. It allows the SVM to classify samples in two classes or more. SVM construct or locate the optimal hyperplane as the decision line, separating the positive (+1) classes from the negative (−1) classes in the binary classification with the two classes largest margin [62]. If the samples are linearly separable, the SVM is used to find the optimal separating hyperplane. This is done by maximising the margin between the hyperplane and the training sample called support vectors [61,63]. SVM was successfully applied to detect and classify grape leaf diseases with an accuracy of 88.89% [64].

DT classifier is a predictive model which maps observations of data for determination of the class of a given feature [60]. It has a tree-like structure in which all sources are split into subsets based on its attribute values [65]. The leaves represent class labels, and the branches represent conjunctions of features leading to that class. This process split the data until no further splitting is possible or all has the same value of target variable. Many decision trees consist of random forest tree classifiers and output the category based on classes output by particular trees [66]. Sankaran et al. [67] investigated VIS-NIR spectroscopy as an approach to detect laurel wilt disease on avocado leaves by introducing four different classifiers, including a DT-based classifier. DT yielded high classification accuracies of over 94% when classifying asymptomatic leaves from infected plants. While the application of NIRS with the aid of ML has provided great insight into detection of plant disease and metabolites in plant, the reliable model for predictive analytics for early detection of G. boninense has yet been discovered. Thus, this proposed study is necessary to choose the best machine learning classifier which

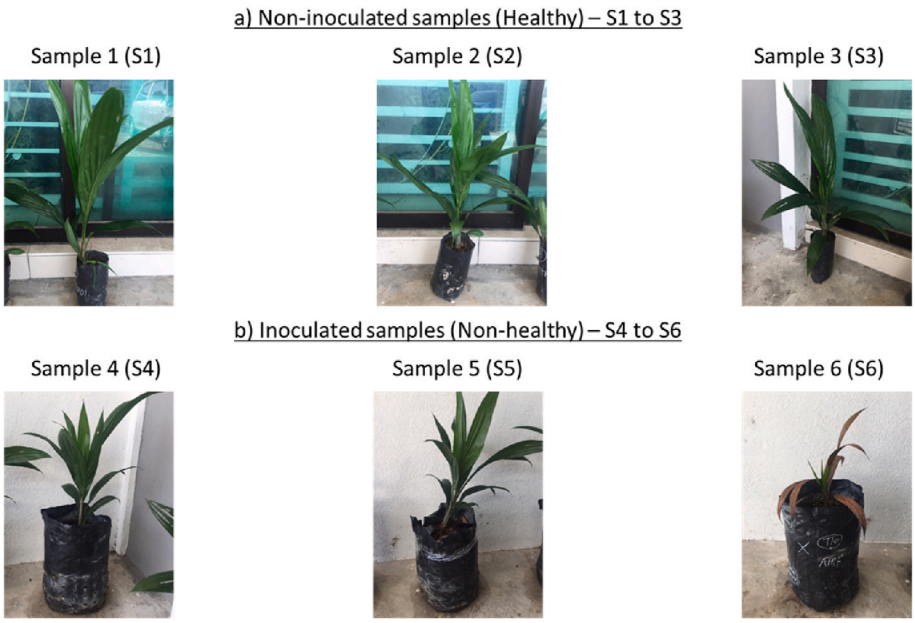


Fig. 1. Oil palm seedlings a) non-inoculated samples b) inoculated samples.

has the best overall performance and easy user-experience for development and deployment of G. boninense predictive model.

2. Methodology

The dataset and proposed methodology have been discussed in detail in this section.

2.1. Oil palm seedlings sample preparation

Six oil palm seedlings at 16-weeks-old are obtained from UniKL BMI, Bangi, Selangor. Three of the oil palm seedlings are inoculated with G. boninense while the other three were non-inoculated seedlings, as shown Fig. 1. The inoculated samples are planted in polypropylene bags containing rubberwood blocks with *Ganoderma* pathogens. Fresh basidiomata of G. boninense is isolated from artificially infected oil palm seedlings at Sime Darby Research Banting, Malaysia. Ganoderma Selective Media (GSM) [68] is used for the cultivation of G. boninense fruiting body. A pure culture of GSM G. boninense mycelium is then re-isolated onto potato dextrose agar (PDA) followed by incubation at 25 °C for 14 days. Then, the culture is transferred to Rubber Wood Block (RWB) for further cultivation. The RWB ($6 \times 6 \times 6$ cm^3) is used as a substrate for G. boninense inoculation cultivation. First, RWB is cleansed and oven-dried (70 °C–80 °C) to remove unwanted debris and natural saprophyte followed by sterilization at 121 °C for 15 min. Next, the RWB is packed in polyethylene bag containing 80 ml of malt extract broth (MEB) and then autoclaved for an hour to allow absorption of the media. The pure culture of G. boninense is then transferred into RWBs containing MEB and incubated (dark condition) at 25 °C for 30–60 days or until the wooden block was fully covered by mycelia fungal [69]. Fully colonized RWB is placed at the position of 2/3 of the polybags (size of 15×23 cm^2) filled with soil mixture. Palm seedling roots are allowed to have direct contact with the source of Ganoderma inoculums, RWB. The samples are divided into two classes based on inoculation status. Non-inoculated samples are Sample 1 to Sample 3 (S1–S3) labelled as 'healthy' and inoculated samples are Sample 4 to Sample 6 (S4–S6) labelled as 'infected'. The seedlings are maintained with regular watering and fertilizer application. One three-month-old oil palm seedling is obtained to assess the effect of water stress on plant and for comparison study with the G. boninense infected samples. The seedling is not watered since day 1 until day 4.

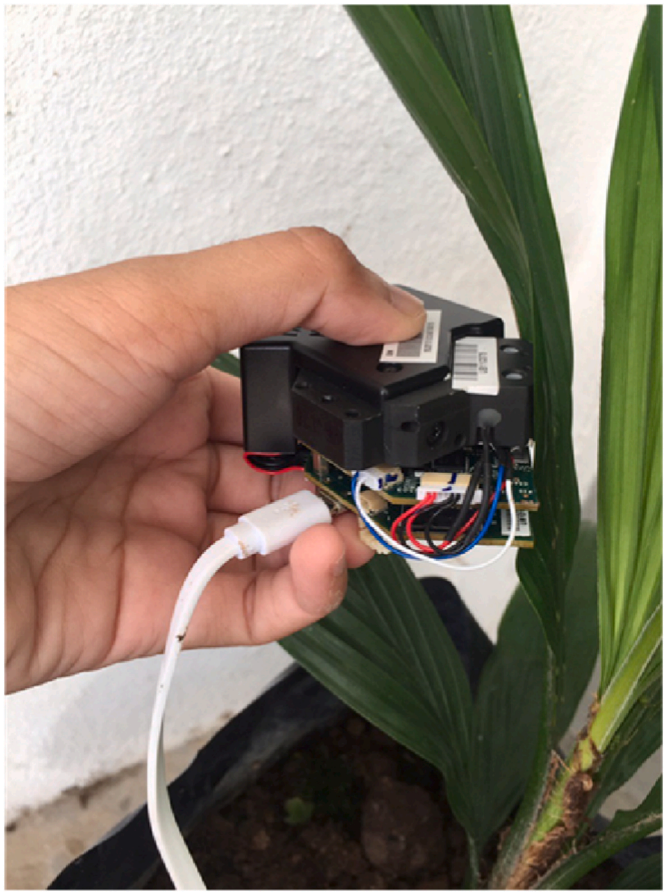


Fig. 2. Spectral acquisition of leaves by using DLP NIRscan Nano spectrometer. Each measurement is repeated three times at the same area to collect average and accurate data.

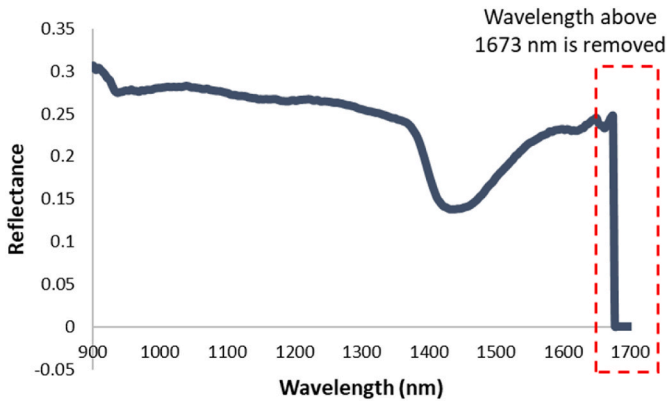


Fig. 3. Wavelength above 1673 nm is removed for further analysis and classification since it contains noise and irrelevant information.

2.2. High-performance liquid chromatography (HPLC) analysis

The high-performance liquid chromatography (HPLC) analysis is conducted to confirm the presence of ergosterol after inoculation procedure. High-performance liquid chromatography (HPLC) method is an analytical chemistry technique for separating, identifying, and quantifying every component of a mixture. The pump passes a pressurized liquid solvent with the sample mixture through a column with a solid adsorbent material. The different flow rate for each component leads to separation of the components as they flow out from the column. Ergosterol is extracted from the root tissue using microwave-assisted extraction (MAE) method. Oil palm root tissue is crushed in liquid nitrogen into powder then transferred into a Pyrex test tube with a Teflon screw cap. 2 ml of methanol and 0.5 mL of 2 M sodium hydroxide were added, and the tube is tightly sealed. The test tubes are microwaved at 70 °C at medium-high power in 30 s exposure time. After the solutions were cold, they are neutralised with concentrated hydrochloric acid. The solutions are extracted three times with 2 ml of pentane and then evaporated until dry by using Buchi Rotary Evaporator. The solution is quantified using HPLC with an ergosterol (≥95% purity, Sigma-Aldrich) standard. An ergosterol standard is prepared for constructing the standard curve. All ergosterol concentrations are reported on a per unit weight basis and each sample is analysed in triplicate.

2.3. Data collection by DLP NIR spectroscopy

The NIR spectral reflectance data is measured using a hand-held DLP NIRscan Nano spectrometer (DLP2010NIR DLP® NIRscan™ Light, Texas Instruments (TI), Dallas, Texas, United States) with a measurement range of 900–1700 nm (176–333 THz). The digital resolution of the NIR device was 228 which defines the number of captured wavelength points across the spectral range. The exposure time is set to 0.635 s with the scan of average of 6 per acquired spectrum. Spectral reflectance data are collected from leaves of the oil palm seedlings as shown in Fig. 2. The reflectance process involves reflected light from the sample, measuring its absorbent qualities by how much light is absorbed back. Sample collection is taken from 6 oil palm seedlings from 16-weeks-old to 32-weeks-old. Spectra for each sample are acquired for 3 times per week in 4 months duration. Meanwhile, the spectral collection for water stress assessment is only done from day 1 to day 4. The whole spectral data is saved in CSV file format and transferred into the computer for further analysis.

2.4. Data pre-processing

Prior to further analysis, pre-processing is applied to the raw collected data. Data pre-processing can influence results for spectral

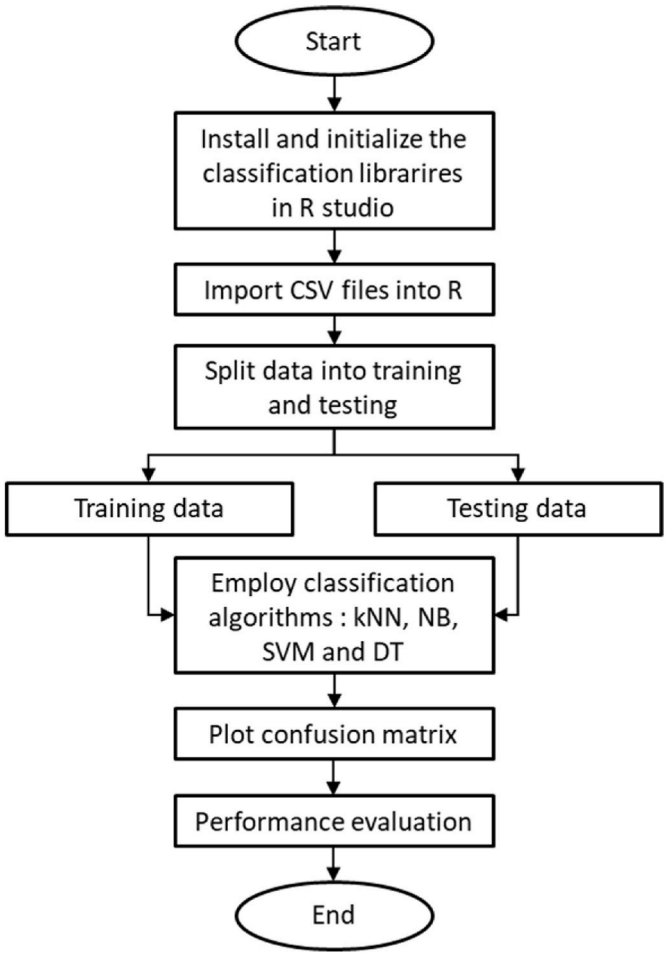
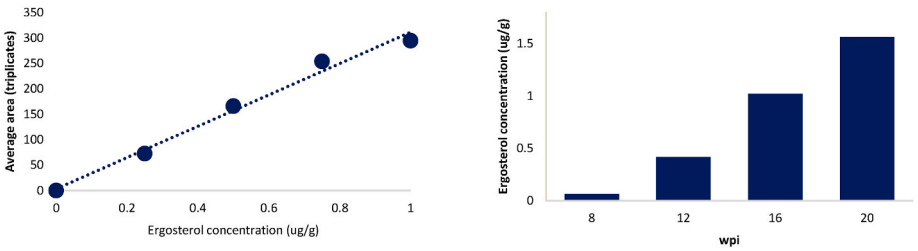


Fig. 4. Flowchart classification model integration on R studio.

data as raw spectral data consist high correlated variables, noises and irrelevant information [70,71]. The raw data contained a serial number, wavelength, intensity, absorbance, and reflectance measurement from wavelengths ranging from 901 to 1701 nm. The first step in data pre-processing is acquiring the relevant dataset. Only reflectance and wavelength are required for this study. Then, the spectral data above 1673 nm is eliminated since they contained noise and contributed to irrelevant readings. The spectral has sudden decline peak as shown in Fig. 3. The raw exported data also had missing data or blank rows. Thus, data cleaning is done by removing blank rows, unrequired columns and irrelevant wavelengths using Macro function on Microsoft Excel. Failing to do this might draw inaccurate and faulty conclusions. Data labels of 'infected' or 'healthy' are added to data to differentiate between known infected and healthy samples.

2.5. Chemometrics data analysis

Chemometrics are employed to extract and concentrate the connotative information for further discrimination of healthy and infected sample. Principal component analysis (PCA) is applied to the spectra sample to extract relevant information of the NIR spectra. PCA spectral loadings can determine feature importance of the spectral wavelength. Maxima and minima from the important bands in the loadings were used to draw the annotations. Each principal component (PC) is a linear sum of variables multiplied by corresponding weighted coefficients. The X-loadings of PCs demonstrate the importance of different variables which aid to finding optimal wavelengths for spectral evaluations [72]. The relationship between ergosterol concentration and NIR spectra of oil palm sample can be examined by employing a partial least square (PLS)



(a) (b)

Fig. 5. (a) Calibration curve of HPLC peak area versus ergosterol concentration; (b) Relationship of inoculation period and ergosterol concentration of oil palm germinated seedlings.

regression on the collected NIR spectral data. PLS implements a predictive model of ergosterol content. The prediction by PLS provides correlation coefficient (R), coefficient of determination (R^2), standard error of prediction (SEP) and root mean square error of prediction (RMSEP).

2.6. Machine learning classification

Several machine learning classifications are tested to find the most effective and suitable classifier model to distinguish between healthy and infected oil palm sample. R-studio software (version 4.1.0) is utilized to perform the classification as shown by the flow chart in Fig. 4.

Acquired spectral data is combined into one single CSV file for classification model development. Different machine learning (ML) algorithms such as K-Nearest Neighbour (kNN), Naïve Bayes (NB), Support Vector Machine (SVM) and Decision Tree (DT) are implemented in the R script to perform classification on spectral data. Some libraries are required to be installed for machine learning classifications execution. Data is imported into R Studio by executing command "iris = read.csv (file.choose(), header = T)" in R script.

In this study, 576 spectral data are used for classification. The whole pre-processed spectral data have been utilizing for the classification purpose with wavelength ranging from 900 to 1673 nm. 70% of dataset are assigned as training data (473 spectral data) while 30% of the remaining data are allocated as testing data (173 spectral data). This method also known as hold-out validation, where the dataset is divided into two different parts in order to create train and test set [73]. It is a reliable approach to employ existing dataset for validation [74]. Each classifier is trained and tested ten times on each dataset to randomise the data and to examine variability performances of the classification models. Thus, consistency and persistence of the model can be observed. The performance of the classification models is evaluated based on confusion matrix since the confusion matrix is more widely used to assess the efficiency of a machine learning classification model [75]. The number of correct and incorrect outputs in a classification problem are summarised and compared with the training model.

2.7. Performance evaluation

Performance evaluations such as accuracy, precision, recall, specificity, and F1-score are interpreted by calculating four statistical indices based on the confusion matrix. A confusion matrix is a summary of prediction results on a classification algorithm. The number of correct and incorrect predictions are summarised with count values and broken down by each class [76]. The different performance parameters can be computed to measure the validity of classification model by using four elements of confusion matrix: true positive (TP), true negative (TN), false positive (FP), false negative (FN), which is given in equations (1)–(5).

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$Specificity = \frac{TN}{TN + FP} \quad (4)$$

$$F1 - score = 2\left(\frac{Recall \times Precision}{Recall + Precision}\right) \quad (5)$$

The healthy sample is assigned as "positive" and the G. boninense infected sample is assigned as "negative". TP refers to the correctly predicted healthy sample, FP refer to incorrectly predicted healthy sample, TN refers to the correctly predicted infected sample, and FN refer to incorrectly predicted infected sample. Accuracy indicated the number of correct predictions made by the classification model. Precision is how many predicted healthy samples are correctly predicted. Recall indicates that from all actual healthy samples, how many are correctly predicted as healthy while specificity indicates that from all actual infected samples, how many are correctly predicted as infected. The F1-score indicates the average of precision and recall [77]. The best model has the maximum values of both recall and precision [78] can be used to comprehensively evaluate the prediction accuracy of a machine learning model [79]. High F1-score which consider both FP and FN indicate that the machine learning algorithm is unlikely to predict incorrectly. Thus, misprediction could be avoided.

3. Results and discussions

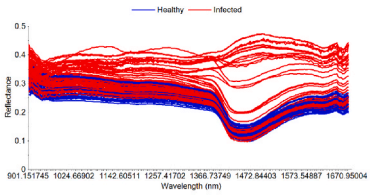
3.1. Ergosterol concentration

The calibration plot was obtained by using standard solutions of ergosterol and plotting the peak areas against the concentration of the standard solutions. Chromatogram shows the detection of pure ergosterol at 282 nm after separation by HPLC. The highest peak corresponding to ergosterol compound. The retention time is 6.673 min. The range of standard ergosterol concentrations used for obtaining the calibration plot is 0–1 μg/g. The plot shows a good linearity, with coefficient of determination, R^2 of 0.9865 as shown in Fig. 5(a). No ergosterol is detected on non-inoculated seedlings. This result corroborates the finding by Mille-Lindblom et al. [80], which reported that ergosterol is a fungal-specific as it is not found naturally in plants or other microbial cells. Ergosterol only can be detected in G. boninense infected oil palm sample. This finding is in agreement with the study by Toh Choon et al. [32] and Phin [81], which found that ergosterol is only detected in infected oil palm seedlings indicating the metabolites caused by G. boninense incursion. This significant finding validates ergosterol as a sensitive marker for detection of G. boninense infection for the detection of basal stem rot (BSR) in oil palm.

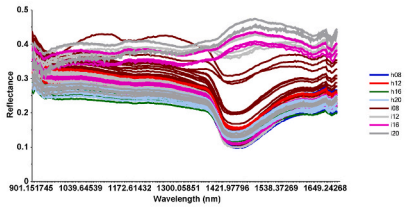
The concentration of ergosterol in samples by week post inoculation

Table 2
Summary of HPLC analysis on G. boninense infected oil palm sample.

Stage of infection (wpi)	Replicates	Area	Average area (triplicates)	Ergosterol concentration (ug/g) (Calculated based on equation y = 308.26x + 3.3651 from standard curve)
8	1	33.96	23.620	0.657
	2	20.37		
	3	16.53		
12	1	123.72	131.347	0.4152
	2	137.45		
	3	132.87		
16	1	303.53	317.853	1.0202
	2	279.91		
	3	370.12		
20	1	478.63	485.123	1.5628
	2	557.46		
	3	419.28		



(a) Raw reflectance spectral data of infected and healthy sample.



(b) Raw reflectance spectral data of infected and healthy sample based on wpi

Fig. 6. Raw reflectance spectral data in NIR region for healthy and infected sample. Noted that significant changes occur at wavelength ~1450 nm.

(wpi) is determined using the equation Y = 308.26x+3.3651 from standard curve calibration. This equation is used to calculate the ergosterol concentration based on peak area from HPLC analysis and is summarised in Table 2. Fig. 5(b) shows that there has been a steady incline in the ergosterol concentration from 8, 12, 16, and 20 wpi in the G. boninense inoculated seedlings. These results agree well with existing studies published from the study of Ganoderma infections in oil palm [82]. The concentration of ergosterol for wpi 8, 12, 16 and 20 are 0.0657, 0. 4152, 1.0202 and 1.5628 μg/g, respectively. A good positive correlation value of R^2 = 0.989 is observed between the inoculation period and ergosterol concentration.

3.2. NIR spectral data of healthy and infected oil palm sample

The acquired raw spectral data is observed in NIR wavelengths ranging from 900 to 1700 nm as shown in Fig. 6. Red line plots represent the infected sample while blue line plots represent the healthy sample. The observed reflectance for infected samples is higher than the healthy samples. The findings in this study mirror those of the previous study by Zhao et al. [83] that performed detection of fungus infection on grape-seed petals using NIR HIS ranging from 700 to 1700 nm. They have found that the reflectance of infected sample is higher than healthy sample which might explained by the decay phenomenon of the infected petals. Despite higher reflectance in infected samples, the spectral data sometimes overlapped between healthy and infected samples making it

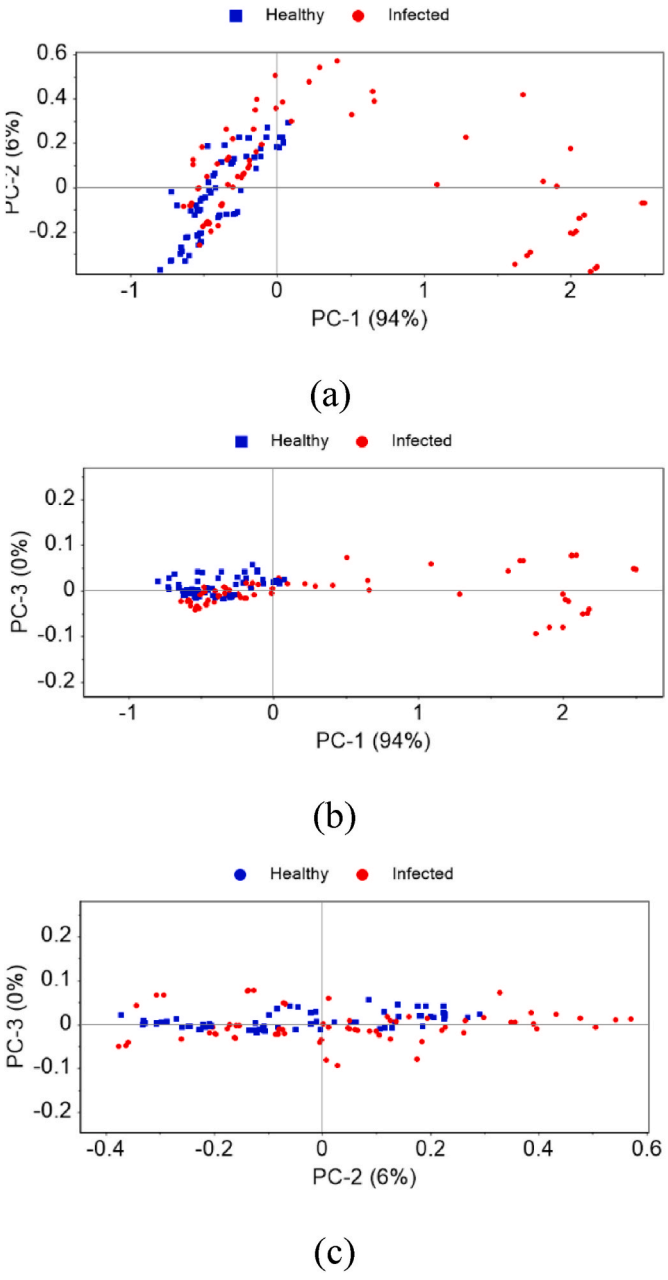


Fig. 7. Cluster plots based on the first three PCs (a) Cluster plot based on PC-1 and PC-2; (b) Cluster plot based on PC-1 and PC-3; (c) Cluster plot based on PC-2 and PC-3.

difficult to distinguish the infected sample from the spectra. Moreover, the visual interpretation of NIR spectra is complicated as specific bands in this region are overtone containing information molecular vibration and generic functional groups [44]. The diminishing of the spectral curve at approximately 1450 nm is strongly suspected to happened due to the loss of water content from the leaves. G. boninense commonly attacks the roots and causes the disruption of water supply to the other part of plant.

3.3. Operation of PCA and PLS chemometrics

PCA is employed to transform the full wavelengths (219 wavelengths) into several principal components (PCs). X-loadings of first PCs are applied for qualitatively identifying the optimal wavelengths that were responsible for the specific features. Fig. 6 shows the clustering of

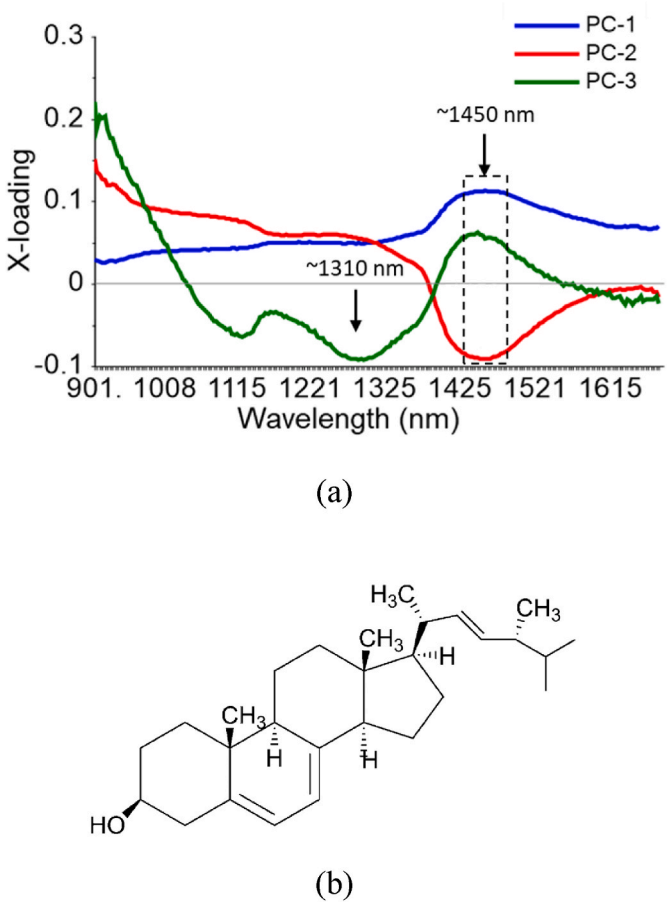


Fig. 8. (a) X-loadings of the PCs of the samples. 1310 and 1452 nm are selected as the optimal wavelength (b) Chemical structure of ergosterol [84].

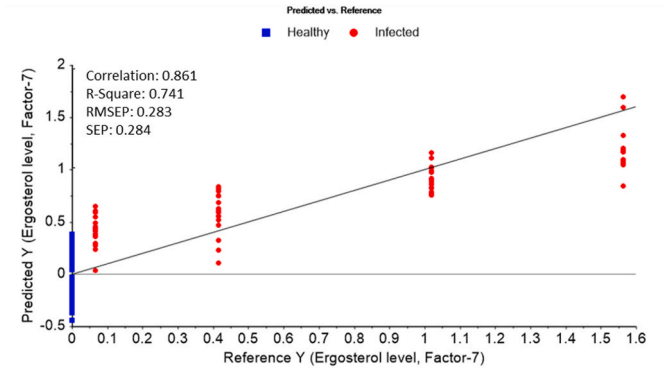


Fig. 9. Relationship between NIR predicted values and measured ergosterol content in oil palm sample.

two groups within the space of PC-1 and PC-2, PC1 and PC-3, PC-2 and PC3, respectively. The first three PCs explained 99% original variations and their score plots are displayed in Fig. 7. Each point in these scatter plots represented one spectrum from one sample. In Fig. 7(a) and (b), two group of samples provide an ostensible clustering. However, there is a slight cross between healthy and infected samples Fig. 7(c). Based on Fig. 7(a) and (b), the spectra of healthy samples are scattered on the negative sides of PC-1, while spectra of infected samples tend to be on a positive side of PC-1. In conclusion, the score plots successfully distinguished the healthy and infected sample.

Selection of optimal wavelengths is of critical significance for removing the redundant information from high-dimensional data,

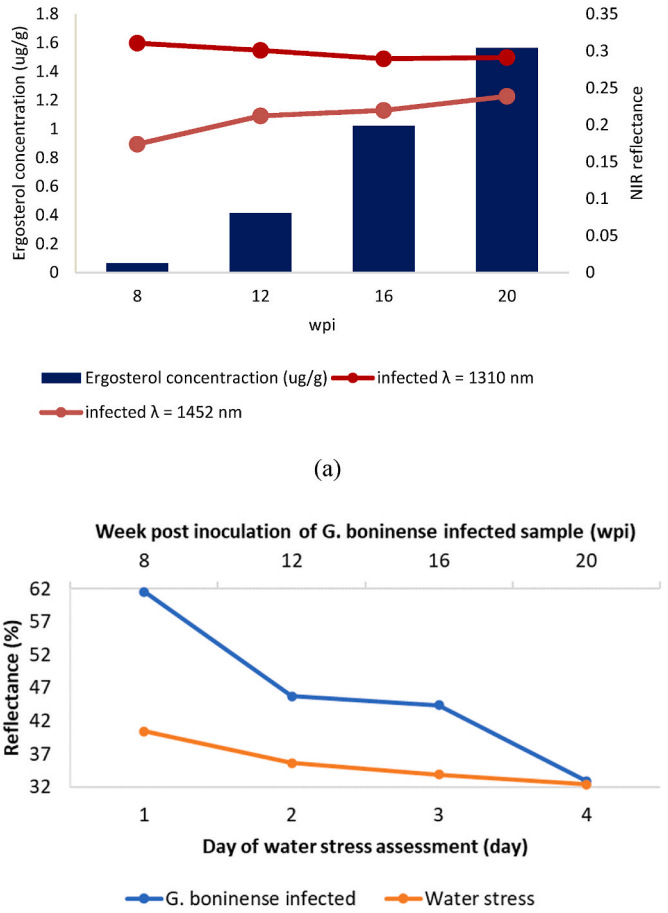


Fig. 10. (a) Correlation of ergosterol concentration and infected reflectance spectra at 1310 and 1452 nm by wpi (b) Comparison between reflectance (%) for water stress and G. boninense infected oil palm sample over time.

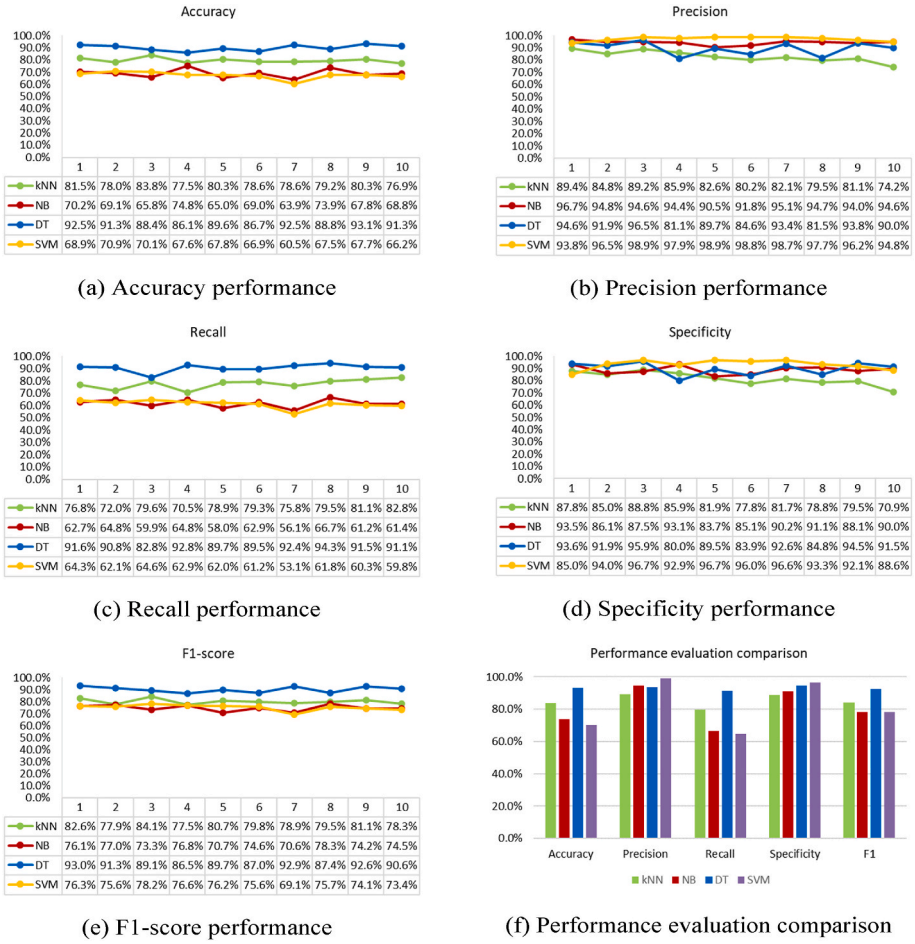
optimizing calibration models and producing excellent results [85]. Thus, identification of optimal wavelengths carrying the most valuable and authentic information is a challenging task in the current hyperspectral data analysis. The X-loadings of PC-1 to PC-3, which revealed the importance of the analysed variables, are shown in Fig. 8(a). Variable with large positive or negative loading coefficient was significant and considered as an optimal wavelength. The first three loading plots of PCA indicated that the reflectance at two wavelengths of 1310 and 1452 nm with the large loading coefficients had the greatest discriminatory effect on healthy and infected petals. The identified optimal wavelength at 1452 nm is approaching to the first overtones O–H stretching near 1450 nm, attributable to water content [86]. Whereas for the optimal wavelength at 1310 nm can be associated with the second overtone of CH_3 groups which is contributed by ergosterol that contains CH_3 components as shown in Fig. 8(b).

A PLS regression is used on NIR spectra to implement the prediction of ergosterol concentration. The PLS model is produced by non-pretreated, 2nd order-derived spectral data provided a prediction model of ergosterol in palm oil sample as shown in Fig. 9. However, the PLS model is characterized by a decent predictiveness with $R^2 = 0.741$, RMSEP = 0.283, and SEP = 0.284. The correlation coefficient, R = 0.861 shows good correlation between the ergosterol concentration and oil palm NIR spectra.

Based on the selected significant wavelengths of 1310 nm and 1452 nm, the relationship between ergosterol concentration and infected reflectance spectral can be observed as plotted in Fig. 10(a). As the

Table 3
Performance results are obtained from different machine learning algorithms.

Classifications	Confusion Matrix				Performance Result (%)				
	TP	FP	TN	FN	Accuracy	Precision	Recall	Specificity	F1-score
kNN	74	9	71	19	83.8	89.2	79.6	88.8	84.1
NB	72	4	41	36	73.9	94.7	66.7	91.1	78.3
DT	75	5	86	7	93.1	93.8	91.5	94.5	92.6
SVM	93	1	29	51	70.1	98.9	64.6	96.7	78.2



(a) Accuracy performance

(b) Precision performance

(c) Recall performance

(d) Specificity performance

(e) F1-score performance

(f) Performance evaluation comparison

Fig. 11. Performance evaluation of different classification techniques.

ergosterol concentration increases, the spectral reflectance shows decrement at 1310 nm and increment at 1452 nm as wpi goes by. The noticeable correlation between ergosterol concentration and infected reflectance spectra is suspected by the changing of water content. Further analysis is conducted to compare the reflectance for water stress and G. boninense infected oil palm sample over time as shown in Fig. 10 (b). Noted that the comparison is done at wavelength 901.15 nm. Water stress reflectance data is collected for day 1, 2, 3 and 4 while for G. boninense infected sample is wpi 8, 12, 16 and 20. As shown in Fig. 10 (b), the percentage of reflectance is decreasing over time for both water stress and G. boninense infected oil palm sample. G. boninense begins to infest the root area interrupting the water transportation to the leaves. This is validated by water stress testing and G. boninense infected reflectance spectra which shows similar reaction towards water depletion.

3.4. Evaluation of machine learning classifiers

Machine learning is employed on all collected spectral data to discriminate between healthy and infected samples. The different machine algorithms discussed earlier are implemented in R studio environment. The spectral datasets are trained and tested using machine learning techniques such as K-Nearest Neighbour (kNN), Naïve Bayes (NB), Support Vector Machine (SVM) and Decision Tree (DT). The results of the different machine algorithms based on confusion matrix are presented in Table 3.

The results show that the highest accuracy is obtained using the DT model with accuracy of 93.1%, followed by kNN and NB with accuracy of 83.8% and 73.9%, respectively. Thus, DT exhibits highest ability to classify between the healthy and infected sample. SVM shows the least accurate performance of only 70.1%. As for precision, all classifications models perform the ability to predict the healthy sample with the precision ranging from 89.2% to 98.9%. As briefed in methodology section, noted that TP refers to the correctly predicted healthy sample, FP refer to incorrectly predicted healthy sample, TN refers to the correctly predicted infected sample, and FN refer to incorrectly predicted infected sample. All classification models show low false positive implying low sample is incorrectly predicted as healthy which means; the infected

sample is being distinguished from healthy sample by all the models. Besides, the specificity for all classifiers determine that the classification models are able to correctly predict the infected sample. Low FN value resulted in higher recall which exposes the ability of the classifier to correctly predict a healthy sample. Lower recall resulting misclassification of infected sample being identified as healthy. DT computes the highest recall up to 91.5%. Despite accuracy, good F1-score that indicated there are low false positive and low false negatives thus correctly predicting the actual sample. DT produces the highest F-score performance of 92.6% compared to other classifiers. The performance of four machine learning techniques is plotted in Fig. 11(a–e) and the comparison for each classifier performance is illustrated into Fig. 11(f). The classification models are run ten times to evaluate the overall performance of the model.

While performing the classification, it is noticed that SVM took a longer time to process and not suitable for large datasets. Based on the performance evaluation, DT algorithm achieves the satisfactory overall performance with accuracy of about 93.1%, precision of 93.8%, recall of 91.5%, specificity of 94.5% and F1-score of 92.6% which indicates the reliability and feasibility to distinguish between G. boninense infected and healthy oil palm sample. Moreover, DT takes less time to execute which is suitable for a real-time prediction model. The results demonstrate a promising discrimination technique with several machine learning classifiers as a reliable classification of G. boninense disease based on extracting important and valuable information from large set of data. Compared to other studies, this integrated NIR with ML techniques may represent a rapid, non-destructive real-time detection method for asymptomatic disease at young plants.

4. Conclusion

Based on HPLC analysis, the inoculation oil palm seedlings indicated the presence of ergosterol which proof the presence of G. boninense in inoculated seedlings. No ergosterol can be found in healthy sample. The ergosterol concentration gradually increase over week post inoculation. The optimal wavelengths are identified at 1310 and 1452 nm where 1452 nm is attributed to water content and 1310 nm could be related to ergosterol content. Based on the PLS model, good correlation between the ergosterol concentration and oil palm NIR spectra is observed. In this study, several machine learning classifiers which are kNN, NB SVM and DT were performed on NIR spectral data to predict and classify healthy and infected oil palm. Nonetheless, it is identified that DT is a potential classifier for this research since DT exhibits consistent good performance compared to the other classifiers. DT produces high accuracy of 93.1% and F1-score of 92.6%. Accuracy and F1-score are the most common metrics used for binary classification in machine learning and significant since they portray the ability of the classification model to classify correctly. A direct comparison of the developed model is not possible due to a scarcity of research on the prediction of ergosterol in G. boninense infected oil palm using the NIRS approach. Although the findings of previous studies are not directly comparable to this present developed model, yet this DT classification model indicates that it has a promising potential for detection of G. boninense. Nevertheless, this is the first study incorporating ergosterol for detection of G. boninense in oil palm using NIR, which is confirmed as a reliable alternative methodology for the analysis of such metabolite's incursion.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgement

The authors would like to thank the Malaysian Ministry of Higher Education for providing largest financial support under Fundamental Research Grant Scheme (FRGS) No. FRGS/1/2019/STG02/UMP/01/1, (University reference RDU1901112), MTUN Matching Grant (RDU212802 and UIC211503), CREST No. T14C2-16, (university Reference UIC190810), and Universiti Malaysia Pahang for additional financial supportunder Internal Research grant RDU213307.

References

[1] M. Department of Statistics, Department of Statistics Malaysia, Selected Agricultural Indicators, 2021, p. 2021. Malaysia.
[2] Malaysian Palm Oil Board, Economics and Industry Development Division: Overview of Industry, 2019, p. 2020.
[3] Malaysian Palm Oil Council, The Oil Palm Tree, 2011.
[4] D. Ariffin, A.S. Idris, G. Singh, Status of Ganoderma in oil palm, in: Ganoderma diseases of perennial crops, 2000, p. 49.
[5] A. Roslan, A.S. Idris, Economic impact of Ganoderma incidence on Malaysian oil palm plantation-a case study in Johor, Oil Palm Ind. Econ. J. 12 (2012) 24–30.
[6] L.C. Wong, C.F.J. Bong, A.S. Idris, Ganoderma species associated with basal stem rot disease of oil palm, Am. J. Appl. Sci. 9 (2012) 879–885.
[7] T.W. Darmono, Detection of basal stem rot disease of oil palm using polyclonal antibody, Menara Perkebunan 67 (1999) 32–39.
[8] D. Ariffin, S. Idris, H. Khairudin, Conformation of ganoderma infected palm by drilling technique, PORIM Int. Palm Oil Cong. (1995) (No. L-0314), PORIM, https://repositorio.fedepalma.org/handle/123456789/78665.
[9] A.S. Idris, S. Rajinder, A.Z. Madihah, M.B. Wahid, Multiplex PCR-DNA kit for early detection and identification of Ganoderma species in oil palm, in: MPOB Information Series TS, 2010, p. 531.
[10] A.S. Idris, M.S. Mazliham, P. Loonis, M.B. Wahid, GanoSken for early detection of Ganoderma infection in oil palm, in: MPOB Information Series TT, 2010, p. 442.
[11] S.W. Dutse, N.A. Yusof, H. Ahmad, M.Z. Hussein, Z. Zainal, An electrochemical DNA biosensor for ganoderma boninense pathogen of the Oil palm utilizing a New ruthenium complex, [Ru (dppz) 2 (qtpy)]Cl2, Int. J. Electrochem. Sci. 7 (2012) 8105–8115.
[12] M.M. Lopez, E. Bertolini, A. Olmos, P.G. Caruso, P. Llop, R. Penyalver, M. Cambra, Innovative tools for detection of plant pathogenic viruses and bacteria, Int. Microbiol. 6 (4) (2003) 233–243.
[13] A. Isha, F.S. Akanbi, N.A. Yusof, R. Osman, W. Mui-Yun, S.N.A. Abdullah, An NMR metabolomics approach and detection of Ganoderma boninense-infected oil palm leaves using MWCNT-based electrochemical sensor, J. Nanomater. (2019) 1–12.
[14] A. Isha, N.A. Yusof, K. Shaari, R. Osman, S.N.A. Abdullah, M.Y. Wong, Metabolites identification of oil palm roots infected with Ganoderma boninense using GC-MS-based metabolomics, Arab. J. Chem. 13 (2020) 6191–6200.
[15] A.Y. Khaled, S.A. Aziz, S.K. Bejo, N.M. Nawi, I.A. Seman, Spectral features selection and classification of oil palm leaves infected by Basal stem rot (BSR) disease using dielectric spectroscopy, Comput. Electron. Agric. 144 (2018) 297–309.
[16] A.Y. Khaled, S.A. Aziz, S.K. Bejo, N.M. Nawi, I. Abu Seman, M.A. Izzuddin, Development of classification models for basal stem rot (BSR) disease in oil palm using dielectric spectroscopy, Ind. Crop. Prod. 124 (2018) 99–107.
[17] J. Dayou, A. Alexander, C.S. Sipaut, K.P. Chong, P.C. Lee, On the possibility of using FTIR for detection of Ganoderma boninense in infected oil palm tree, Int. J. Adv. Agricult. Environ. Eng. 1 (2014) 161–163.
[18] A. Alexander, C.S. Sipaut, K.P. Chong, P.C. Lee, J. Dayou, Sensitivity analysis of the detection of Ganoderma boninense infection in oil palm using FTIR, Trans. Sci. Technol. 1 (2014) 1–5.
[19] A.H. Abdullah, A.M. Shakaff, A.H. Adom, S.A. Ghani, I.A. Seman, P2. 1.7 Exploring MIP Sensor of Basal Stem Rot (BSR) Disease in Palm Oil Plantation, Proceedings of the Proceedings IMCS 2012Nuremberg, Germany, 2012, pp. 1348–1351.
[20] S. Liaghat, S. Mansor, R. Ehsani, H.Z.M. Shafri, S. Meon, S. Sankaran, Mid-infrared spectroscopy for early detection of basal stem rot disease in oil palm, Comput. Electron. Agric. 101 (2014) 48–54.
[21] H.Z. Shafri, M.I. Anuar, I.A. Seman, N.M. Noor, Spectral discrimination of healthy and Ganoderma-infected oil palms from hyperspectral data, Int. J. Rem. Sens. 32 (2011) 7111–7129.
[22] S.E. Liaghat, S. Mansor, H.Z. Shafri, S. Meon, S. Sankaran, S.H. Azam, Early detection of basal stem rot disease (Ganoderma) in oil palms based on hyperspectral reflectance data using pattern recognition algorithms, Int. J. Rem. Sens. 35 (2014) 3427–3439.
[23] P. Ahmadi, F.M. Muharam, K. Ahmad, S. Mansor, I. Abu Seman, Early detection of ganoderma basal stem rot of oil palms using artificial neural network spectral analysis, Plant Dis. 101 (2017) 1009–1016.
[24] A.Y. Khaled, S.A. Aziz, S.K. Bejo, N.M. Nawi, I.A. Seman, D.I. Onwude, Early detection of diseases in plant tissue using spectroscopy-applications and limitations, Appl. Spectrosc. Rev. 53 (2018) 36–64.
[25] P.S. Liang, D.C. Slaughter, A. Ortega-Beltran, T.J. Michailides, Detection of fungal infection in almond kernels using near-infrared reflectance spectroscopy, Biosyst. Eng. 137 (2015) 64–72.

[26] J.C. Marín-Ortiz, N. Gutierrez-Toro, V. Botero-Fernández, L.M. Hoyos-Carvajal, Linking physiological parameters with visible/near-infrared leaf reflectance in the incubation period of vascular wilt disease, Saudi J. Biol. Sci. 27 (1) (2020) 88–99.
[27] S. Fang, R. Cui, Y. Wang, Y. Zhao, K. Yu, A. Jiang, Application of multiple spectral systems for the tree disease detection: a review, Appl. Spectrosc. Rev. (2021) 1–27.
[28] Y. Lu, R. Garcia, B. Hansen, M. Gleicher, R. Maciejewski, The state-of-the-art in predictive visual analytics, Comput. Graph. Forum 36 (2017) 539–562.
[29] R. Kumar, S. Pathak, N. Prakash, U. Priya, A. Ghatak, Application of Spectroscopic Techniques in Early Detection of Fungal Plant Pathogens, IntechOpen, London, UK, 2021.
[30] M.L. Rodrigues, The multifunctional fungal ergosterol, mBio 9 (5) (2018) e01755-01718.
[31] A.M. As'wad, M. Sariah, R.R.M. Paterson, M.Z. Abidin, N. Lima, Ergosterol analyses of oil palm seedlings and plants infected with Ganoderma, Crop Protect. 30 (11) (2011) 1438–1442.
[32] R.L. Toh Choon, M. Sariah, M.N. Siti Mariam, Ergosterol from the soilborne fungus Ganoderma boninense, J. Basic Microbiol. 52 (5) (2012) 608–612.
[33] M.S.H.R. Bivi, A.S. Paiko, A. Khairulmazmi, M.S. Akhtar, A.S. Idris, Control of basal stem rot disease in oil palm by supplementation of calcium, copper, and salicylic acid, Plant Pathol. J. 32 (5) (2016) 396.
[34] R.L. Toh Choon, M. Sariah, M.N. Siti Mariam, Ergosterol from the soilborne fungus Ganoderma boninense, J. Basic Microbiol. 52 (2012) 608–612.
[35] M.M. Saad, N.S. Ali, S. Meon, Relationship between ganoderma ergosterol concentration and basal stem rot disease progress on Elaeis guineensis, Trop. Life Sci. Res. 31 (1) (2020) 19 (2020).
[36] F. Ghilardelli, M. Barbato, A. Gallo, A preliminary study to classify corn silage for high or low mycotoxin contamination by using near infrared spectroscopy, Toxins 14 (5) (2022) 323.
[37] D. Tyska, A.O. Mallmann, J.K. Vidal, C.A.A.D. Almeida, L.T. Gressler, C. A. Mallmann, Multivariate method for prediction of fumonisins B1 and B2 and zearalenone in Brazilian maize using Near Infrared Spectroscopy (NIR), PLoS One 16 (1) (2021), e0244957.
[38] J.U. Porep, M.E. Erdmann, A. Korzendorfer, D.R.C. Kammerer, Rapid determination of ergosterol in grape mashes for grape rot indication and further quality assessment by means of an industrial near infrared/visible (NIR/VIS) spectrometer-A feasibility study, Food Control 43 (2014) 142–149.
[39] A. Femenias, F. Gatius, A.J. Ramos, V. Sanchis, S. Marin, Near-infrared hyperspectral imaging for deoxynivalenol and ergosterol estimation in wheat samples, Food Chem. 341 (2021), 128206.
[40] I.M. Almanjahie, I. Ahmad, Z.C. Elmezouar, A. Laksaci, Soil quality analysis using modern statistics and NIR spectroscopy procedure, Pol. J. Environ. Stud. 28 (5) (2019) 3581–3588.
[41] N. Berardo, V. Pisacane, P. Battilani, A. Scandolara, A. Pietri, A. Marocco, Rapid detection of kernel rots and mycotoxins in maize by near-infrared reflectance spectroscopy, J. Agric. Food Chem. 53 (21) (2005) 8128–8134.
[42] M. Blanco, I. Villaroya, NIR spectroscopy: a rapid-response analytical tool, TrAC, Trends Anal. Chem. 21 (2002) 240–250.
[43] D. Wu, L. Feng, C. Zhang, Y. He, Early detection of Botrytis cinerea on eggplant leaves based on visible and near-infrared spectroscopy, Trans. ASABE (Am. Soc. Agric. Biol. Eng.) 51 (2008) 1133–1139.
[44] M. Manley, Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis of biological materials, Chem. Soc. Rev. 43 (2014) 8200–8214.
[45] M.I.S. Mohd Hilmi Tan, M.F. Jamlos, A.F. Omar, F. Dzaharudin, S. Chalermwisutkul, P. Akkaraekthalin, Ganoderma boninense disease detection by near-infrared spectroscopy classification: a review, Sensors 21 (9) (2021) 3052.
[46] L. Naher, U.K. Yusuf, A. Ismail, S.G. Tan, M.M.A. Mondal, Ecological status of Ganoderma'and basal stem rot disease of oil palms ('Elaeis guineensis' Jacq.), Aust. J. Crop. Sci. 7 (2013) 1723.
[47] G. Singh, Ganoderma-the scourge of oil palms [Elaeis guineensis] in the coastal areas [Peninsular Malaysia], Planter 67 (1991) 421–444.
[48] A. Susanto, Basal stem rot in Indonesia. Biology, economic importance, epidemiology, detection and control, in: Proceedings of International Workshop on Awareness, Detection and Control of Oil Palm Devastating DiseasesKuala Lumpur Convention Centre, Universiti Putra Malaysia Press, Malaysia, 2009.
[49] H. Bhavsar, A. Ganatra, A comparative study of training algorithms for supervised machine learning, Int. J. Soft Comput. Eng. 2 (2012) 2231–2307.
[50] T.M. Cover, P. Hart, Nearest neighbor pattern classification, IEEE Trans. Inf. Theor. 13 (1967) 21–27.
[51] T. Mitchell, Machine Learning, McGraw-Hill Higher Education, New York, 1997.
[52] X. Wu, V. Kumar, J.R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G.J. Mclachlan, A. Ng, B. Liu, P.S. Yu, Z. Zhou, M. Steinbach, D.J. Hand, D. Steinberg, Top 10 algorithms in data mining, Knowl. Inf. Syst. 14 (2008) 1–37.
[53] S. Liaghat, R. Ehsani, S. Mansor, H.Z. Shafri, S. Meon, S. Sankaran, S.H. Azam, Early detection of basal stem rot disease (Ganoderma) in oil palms based on hyperspectral reflectance data using pattern recognition algorithms, Int. J. Rem. Sens. 35 (2014) 3427–3439.
[54] S. Liaghat, S. Mansor, R. Ehsani, H.Z.M. Shafri, S. Meon, S. Sankaran, Mid-infrared spectroscopy for early detection of basal stem rot disease in oil palm, Comput. Electron. Agric. 101 (2013) 48–54.
[55] T.C. Pearson, D.T. Wicklow, Detection of corn kernels infected by fungi, Trans. ASABE (Am. Soc. Agric. Biol. Eng.) 49 (2006) 1235–1245.
[56] R. Moscetti, D. Monarca, M. Cecchini, R.P. Haff, M. Contini, R. Massantini, Detection of mold-damaged chestnuts by near-infrared spectroscopy, Postharvest Biol. Technol. 93 (2014) 83–90.
[57] U. Solanki, U.K. Jaliya, D.G. Thakore, A survey on detection of disease and fruit grading, Int. J. Inno. Emerg. Res. Eng. 2 (2015) 109–114.
[58] S.M. Kamruzzaman, Text classification using artificial intelligence, J. Electr. Eng. (2010).
[59] P. Langley, W. Iba, K. Thompson, Proceedings of the Tenth National Conference on Artificial IntelligenceSan Jose, An analysis of Bayesian Classifiers, CA, 1992.
[60] S.D. Jadhav, H.P. Channe, Comparative study of K-NN, naive Bayes and decision tree classification techniques, Int. J. Sci. Res. 5 (2016) 1842–1845.
[61] S.R. Gunn, Support Vector Machine for Classification and Regression, University of Southampton, 2005.
[62] S.N. Sabeh, Intelligent Computer Vision System Featuring Support Vector Machine with Wilk's Analysis and Unimodal Thresholding, Phd Thesis, Universiti Sains Malaysia, 2012.
[63] D.A. Ramli, Development of Multibiometric Speaker Identification Systems with Support Vector Machine Audio Reliability Estimation, Phd Thesis, Universiti Kebangsaan Malaysia, 2010.
[64] P.B. Padol, A.A. Yadav, SVM classifier based grape leaf disease detection, in: 2016 Conference on Advances in Signal Processing (CASP), IEEE, 2016, pp. 175–179.
[65] M. Suresha, K.S. Sandeep Kumar, G. Shiva Kumar, Texture features and decision trees based vegetables classification, Int. J. Comput. Appl. 975 (2012) 8878.
[66] S.R. Bandi, A. Varadharajan, A. Chinnasamy, Performance evaluation of various statistical classifiers in detecting the diseased citrus leaves, Int. J. Eng. Sci. Technol. 5 (2013) 298–307.
[67] S. Sankaran, R. Ehsani, S.A. Inch, R.C. Ploetz, Evaluation of visible-near infrared reflectance spectra of avocado leaves as a non-destructive sensing tool for detection of laurel wilt, Plant Dis. 96 (2012) 1683–1689.
[68] D. Ariffin, S. Idris, in: The ganoderma selective medium (GSM), PORIM information Series, Palm Oil Research Institute of Malaysia, 1993.
[69] A. Alexander, C.S. Sipaut, J. Dayou, K.P. Chong, Oil palm roots colonisation by Ganoderma boninense: an insight study using scanning electron microscopy, J Oil Palm Res. 29 (2017) 262–266.
[70] W. Wu, Y. Mallet, B. Walczak, W. Penninckx, D.L. Massart, S. Heuerding, F. Erni, Comparison of regularized discriminant analysis linear discriminant analysis and quadratic discriminant analysis applied to NIR data, Anal. Chim. Acta 329 (1996) 257–265.
[71] J.G. Tallada, D.T. Wicklow, T.C. Pearson, P.R. Armstrong, Detection of fungus-infected corn kernels using near-infrared reflectance spectroscopy and color imaging, Trans. ASABE (Am. Soc. Agric. Biol. Eng.) 54 (2011) 1151–1158.
[72] X. Li, P. Nie, Z.J. Qiu, Y. He, Using wavelet transform and multi-class least square support vector machine in multi-spectral imaging classification of Chinese famous tea, Expert Syst. Appl. 38 (9) (2011) 11149–11159.
[73] A. Verma, V. Ranga, Machine learning based intrusion detection systems for IoT applications, Wireless Pers. Commun. 111 (2020) 2287–2310.
[74] A. Vabalas, E. Gowen, E. Poliakoff, A.J. Casson, Machine learning algorithm validation with a limited sample size, PLoS One 14 (2019), e0224365.
[75] S. Samsir, J.H.P. Sitorus, Z. Ritonga, F.A. Nasution, R. Watrianthos, Comparison of machine learning algorithms for chest X-ray image COVID-19 classification, J. Phys. Conf. 1933 (2021), 012040.
[76] U. Sharma, S. Saran, S.M. Patil, Fake news detection using machine learning algorithms, Int. J. Creativ. Res. Thoughts 8 (6) (2020).
[77] S. Narkhede, Understanding Confusion Matrix, 2018.
[78] H. Huang, H.-H. Xu, X. Wang, W. Silamu, Maximum F1-score discriminative training criterion for automatic mispronunciation detection, IEEE/ACM Trans. Audio Speech Lang. Process. 23 (2015) 787–797.
[79] X. Xu, T. Lai, S. Jahan, F. Farid, Water and Sediment Analyse Using Predictive Models, 2022 arXiv preprint arXiv:2203.03422.
[80] C. Mille-Lindblom, H. Fischer, L.J. Tranvik, Litter-associated bacteria and fungi-a comparison of biomass and communities across lakes and plant species, Freshw. Biol. 51 (4) (2006) 730–741.
[81] C. Phin, An evaluation of Ganoderma fungal colonisation using ergosterol analysis and quantification, Planter 88 (1034) (2012) 311–319.
[82] M.M. Saad, N.S. Ali, S. Meon, Relationship between ganoderma ergosterol concentration and basal stem rot disease progress on Elaeis guineensis, Trop. Life Sci. Res. 31 (2020) 19.
[83] Y.R. Zhao, K.Q. Yu, X. Li, Y. He, Detection of fungus infection on petals of rapeseed (Brassica napus L.) using NIR hyperspectral imaging, Sci. Rep. 6 (1) (2016) 1–9.
[84] L. Luza, A. Gual, J. Dupont, The partial hydrogenation of 1, 3-dienes catalysed by soluble transition-metal nanoparticles, ChemCatChem 6 (3) (2014) 702–710.
[85] Z. Xiaobo, Z. Jiewen, M.J. Povey, M. Holmes, M. Hanpin, Variables selection methods in near-infrared spectroscopy, Anal. Chim. Acta 667 (1–2) (2010) 14–32.
[86] J.D. Campbell, L. Holder-Pearson, C.G. Pretty, C. Benton, J. Knopp, J.G. Chase, Development of a discrete spectrometric NIR reflectance glucometer, IFAC-PapersOnLine 53 (2) (2020) 15970–15975.