

## Article

# Classification of Basal Stem Rot Disease in Oil Palm Plantations Using Terrestrial Laser Scanning Data and Machine Learning

Nur A. Husin <sup>1</sup>, Siti Khairunniza-Bejo <sup>1,2,3,\*</sup>, Ahmad F. Abdullah <sup>1,4</sup>,  
Muhamad S. M. Kassim <sup>1,3</sup>, Desa Ahmad <sup>1,2</sup> and Mohd H. A. Aziz <sup>5</sup>

<sup>1</sup> Department of Biological and Agricultural Engineering, Faculty of Engineering, Universiti Putra Malaysia, UPM, Serdang 43400, Selangor, Malaysia; nurazuan@upm.edu.my (N.A.H.); ahmadfikri@upm.edu.my (A.F.A.); saufi@upm.edu.my (M.S.M.K.); desa@upm.edu.my (D.A.)

<sup>2</sup> Smart Farming Technology Research Centre, Universiti Putra Malaysia, UPM, Serdang 43400, Selangor, Malaysia

<sup>3</sup> Institute of Plantation Studies, Universiti Putra Malaysia, UPM Serdang 43400, Selangor, Malaysia

<sup>4</sup> International Institute of Aquaculture and Aquatic Sciences, UPM, Port Dickson 71050, Negeri Sembilan, Malaysia

<sup>5</sup> Department of Science and Technology, Faculty of Humanities, Management and Science, Universiti Putra Malaysia, Bintulu Campus, Bintulu 97008, Sarawak, Malaysia; m\_hamim@upm.edu.my

\* Correspondence: skbejo@upm.edu.my; Tel.: +60-397-694-332

Received: 8 September 2020; Accepted: 9 October 2020; Published: 22 October 2020



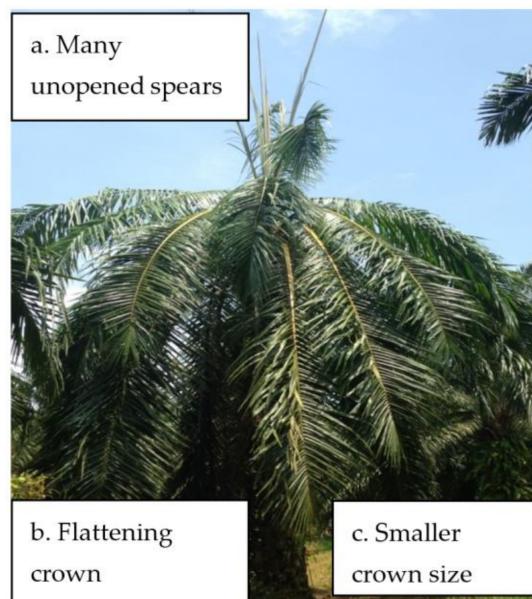
**Abstract:** The oil palm industry is vital for the Malaysian economy. However, it is threatened by the *Ganoderma boninense* fungus, which causes basal stem rot (BSR) disease. Foliar symptoms of the disease include the appearance of several unopened spears, flat crowns, and small crown size. The effect of this disease depends on the severity of the infection. Currently, the disease can be detected manually by analyzing the oil palm tree's physical structure. Terrestrial laser scanning (TLS) is an active ranging method that uses laser light, which can directly represent the tree's external structure. This study aimed to classify the healthiness levels of the BSR disease using a machine learning (ML) approach. A total of 80 oil palm trees with four different healthiness levels were pre-determined by the experts during data collection with 40 each for training and testing. The four healthiness levels are T0 (healthy), T1 (mildly infected), T2 (moderately infected), and T3 (severely infected), with 10 trees in each level. A terrestrial scanner was mounted at a height of 1 m, and each oil palm was scanned at four positions at a distance of 1.5 m around the tree. Five tree features were extracted from the TLS data: C200 (crown slice at 200 cm from the top), C850 (crown slice at 850 cm from the top), crown area (number of pixels inside the crown), frond angle, and frond number. C200 and C850 were obtained using the crown stratification method, while the other three features were obtained from the top-down image. The obtained features were then analyzed by principal component analysis (PCA) to reduce the dimensionality of the dataset and increase its interpretability while at the same time minimizing information loss. The results showed that the kernel naïve Bayes (KNB) model developed using the input parameters of the principal components (PCs) 1 and 2 had the best performance among 90 other models with a multiple level accuracy of 85% and a Kappa coefficient of 0.80. Furthermore, the combination of the two highest PC variance with the most weighted to frond number, frond angle, crown area, and C200 significantly contributed to the classification success. The model also could classify healthy and mildly infected trees with 100% accuracy. Therefore, it can be concluded that the ML approach using TLS data can be used to predict early BSR infection with high accuracy.

**Keywords:** machine learning; TLS; BSR; kernel naïve Bayes; oil palm crown; oil palm frond

## 1. Introduction

The oil palm (*Elaeis guineensis*) is a species of palm that is extensively planted in Southeast Asia, which is currently the main palm oil-producing region. In Malaysia, oil palm is the most important commodity crop, and the country is one of the world's largest palm oil producers. Palm oil and palm-based products are among the country's top 10 exports, with annual exports increasing steadily over the last 30 years. White-rot fungus, identified as *Ganoderma boninense*, is the causal pathogen for basal stem rot (BSR) disease [1]. It has been found that a *Ganoderma boninense* attack can lead to yield reduction in fresh fruit bunches (FFBs) of up to 4.3 tonnes per hectare, and it was estimated that more than 400,000 ha could be affected in 2020, amounting to 1.74 million tonnes of FFB yield reduction [2]. According to [1,2], the *Ganoderma boninense* species is the most devastating, having a significant detrimental effect on the palm oil industry and the economy in Southeast Asia.

Healthy trees are considered to have a larger crown size and a better developed canopy compared with infected trees [3–5]. BSR infection can cause a change in the physical appearance and growth of oil palm trees. This change is due to the internal tissue damage caused by the *Ganoderma boninense* fungus, which restricts the level of water and nutrient consumption. Consequently, this affects the ability of the plant to perform normal photosynthesis, disrupting growth and degenerating the oil palm tree's physical condition [6]. Nutrient deficiency results in impaired new leaf growth [7] and, in severe cases, the non-development of new leaves or bunches has been observed [8]. Stunted leaf growth also leads to smaller crowns [9,10]. The effect of the disease on the tree's physical structure is more pronounced and detectable when the infection is more severe. The foliar symptoms of infected trees are flattening of the crown, a high presence of unopened spear leaves, and a smaller crown size (Figure 1). The lack of standards, coupled with error-prone methods, has led to contradictory assessments in the literature [11,12]. Laboratory-based methods are reliable for early detection; however, they are costly, complex, and ill-suited for outdoor conditions. Sensor-based techniques can distinguish between healthy and unhealthy oil palm trees with varying levels of accuracy. However, these techniques are not sufficiently able to distinguish between the different levels of infection severity.



**Figure 1.** Foliar symptoms of an oil palm crown severely infected with basal stem rot (BSR) disease: (a) many unopened spears, (b) flattening crown, (c) smaller crown size.

Light detection and ranging (LiDAR) is an active ranging method that measures the distance or range to a target using pulsed laser light. It can directly represent the external structure and profile of an object, such as a tree. Extensive biometric data have been used by researchers and field site workers

to estimate tree properties while reducing inventory costs. Previous studies have demonstrated that terrestrial LiDAR can be used to obtain canopy vegetation profiles and other structural tree properties from an understory perspective [13–19]. These researchers used point cloud data from terrestrial laser scanning (TLS) and extracted data for various features, such as tree height, diameter at breast height, crown height, width, area, and plant area index. The results show that point cloud data from a terrestrial scanner can be used for the extraction of various tree features with high correlation. On the basis of this literature, we can conclude that TLS is well adapted for the intensive in situ study of tree geometry. However, very few TLS studies have focused on the detailed oil palm tree architecture. Recently, the authors of [20] studied the changes in the oil palm architecture, owing to the BSR disease, whereas [21] used single and combined features extracted from oil palm trees using the TLS method. The latter study used the raw data and a regression approach to measure how close it fits the expected model's line and curve. The same study also discussed the importance of the extracted features, whereas a machine learning approach has not yet been used.

Research into image processing for plant disease detection has grown rapidly over the past decade [22]. Machine learning (ML) approaches have been applied in various fields, including bioinformatics, aquaculture, food, and precision farming, which is now also known as digital farming [23]. ML has emerged to facilitate the monitoring and early information on plant health for strategic management strategies. To study nutrient disease in oil palms, a kernel-based support vector machine (SVM) classifier was applied to 420 oil palm leaf samples [24]. Three hundred images were used for training and 120 images were used for testing, comprising 40 images from each of the following nutrients: magnesium, nitrogen, and potassium. The images were pre-processed using a median filter, and then color histogram-based features and gray level co-occurrence were extracted from the images. An SVM with a polynomial kernel (soft margin) executed the classification with 95% accuracy. Meanwhile, the naïve Bayes (NB) method was used to diagnose oil palm disease in Indonesia [25] on the basis of various symptoms identified in the leaves, spear, stem, and fruits. According to the results, the diagnosis of oil palm disease was achieved with 80% accuracy. Furthermore, a thermal imaging technique was employed to detect BSR disease using 53 healthy and 53 infected palms [26]. Four values were extracted from the images: maximum, minimum, mean, and standard deviation of pixel intensity. Further analysis was conducted using principal component analysis (PCA), followed by two classification techniques, SVM and *k*-nearest neighbor (kNN). The SVM results achieved better results than kNN, with 89.2% accuracy during training and 84.4% during testing. The proposed techniques are capable of distinguishing between healthy and BSR-infected oil palm trees, but they are unable to detect the infection's level of severity.

Multispectral Quickbird satellite images were used by [27] for BSR disease classification in Sumatra, Indonesia. The site consisted of 144 oil palm trees of different ages, ranging from 10 to 21 years old, which were divided into two classes: 99 healthy trees and 45 unhealthy trees. The results showed that the random forest (RF) classifier was the best classifier in comparison with SVM and regression tree (CART) models, with the highest accuracy in the producer (91%), user (83%), and overall (91%) categories. Subsequently, images from the WorldView-3 satellite, which has a panchromatic resolution of 131 cm, eight bands with 1.24 m resolution, and a revisit interval of less than 1 day, were used for BSR severity classification [28]. Oil palm trees were selected on the basis of four severity labels: healthy, initially unhealthy, moderately unhealthy, and severely unhealthy. Similar ML algorithms were applied, with CART being replaced by decision tree (DT), while a stepwise variable selection was used to obtain significant variables to separate the classes. As a result, the SVM approach was the best classifier for distinguishing all four classes, with a moderate overall accuracy of 54%. In addition, a neural network analysis method was used to isolate and classify the spectral data of healthy and infected oil palm trees [29]. A total of 1016 oil palm leaflet samples (416 samples from the first trial and 600 samples from the second trial) were obtained from frond numbers 9 and 17. Spectral data of the foliar samples were scanned using a portable spectroradiometer at 1.45-nm intervals, at a range of 273 to 1100 nm, with a resolution of 5 nm. The neural network method used in that study was the

back-propagation and multilayer method, owing to its ability to determine non-linear combinations of raw, first, and second derivative spectral datasets. The best results in distinguishing between T1 and T2 infection levels occurred in the visible green wavelength range, with accuracies of 83.3% and 100% for 540 and 550 nm, respectively.

The detection of BSR in oil palms was also explored in North Sumatra, Indonesia, using PP-SYSTEMS at a range of 310 to 1130 nm (256 bands, 10 nm resolution) [11]. The system was mounted on a 2 m shaft on top of a scaffold to measure the canopy of 95 oil palm trees consisting of the following categories: healthy (36 trees), level 1 (18 trees), level 2 (38 trees), and level 3 (3 trees), with level 3 being the most severely infected. The classification method using partial least squares discriminant analysis (PLS-DA) was applied and the results showed that the proposed method could distinguish between healthy and infected trees with 98% accuracy and between the four classification levels with 94% accuracy. In addition, the authors of [30] used a handheld portable hyperspectral spectroradiometer to collect the leaf reflectance data from frond number 17 for 47 healthy, 55 slightly damaged, 48 moderately damaged, and 40 severely damaged oil palm trees. PCA, followed by the kNN classification model, resulted in an average overall accuracy of 97% with the second derivative dataset. These techniques have the ability to differentiate between healthy and non-healthy trees with varying levels of accuracy. However, further improvements are needed to accurately distinguish between the levels of severity, especially between levels T2 and T3.

The researchers used electrical properties, such as impedance, capacitance, dielectric constant, and dissipation factor to detect BSR disease in oil palm trees at an early stage [31]. Only 56 samples from mature oil palm trees were selected, with 14 trees in each of the four infection levels. Leaflets from frond number 17 were randomly collected, with 224 samples gathered in total. PCA was performed to reduce the dimensionality of the data, followed by classification models using linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), kNN, and NB. QDA achieved the highest accuracy, at 80.79%, and impedance was the best parameter, with an overall accuracy ranging from 82% to 100%. Meanwhile, a handheld e-nose sensor was used to detect BSR disease by taking *Ganoderma boninense* basidiocarp samples to the laboratory for testing without conducting field tests [32]. Data processing consisted of PCA, hierarchical cluster analysis, and LDA, which were used to find the separation between the samples. The results showed that a possible approach is to segregate the infected and healthy trees; however, with different levels of infection, further research is needed. Moreover, ALOS PALSAR 2, a synthetic aperture radar sensor emitting L-band microwave radio waves, was used to classify four severity levels in 92 oil palm trees [33]. For the reception and emission of radar acquisition, the researchers employed two polarisations of the satellite images, namely, HH (horizontal–horizontal) and HV (horizontal–vertical). The data were pre-processed to filter out noise using the sentinel application platform (SNAP). The multilayer perceptron (MP) classifier for HV polarization achieved better results than the KStar classifier, with 77% accuracy.

A summary of the ML approaches to classifying the severity of BSR infection in oil palms is given in Table 1. Different input data were used, and various levels of accuracy were obtained using the diverse methods of ML. However, none of them used TLS as input data. Therefore, further research is needed to determine the capabilities of TLS combined with ML techniques to detect BSR disease using oil palm tree crown properties.

**Table 1.** List of machine learning (ML) techniques used for *Ganoderma boninense* detection.

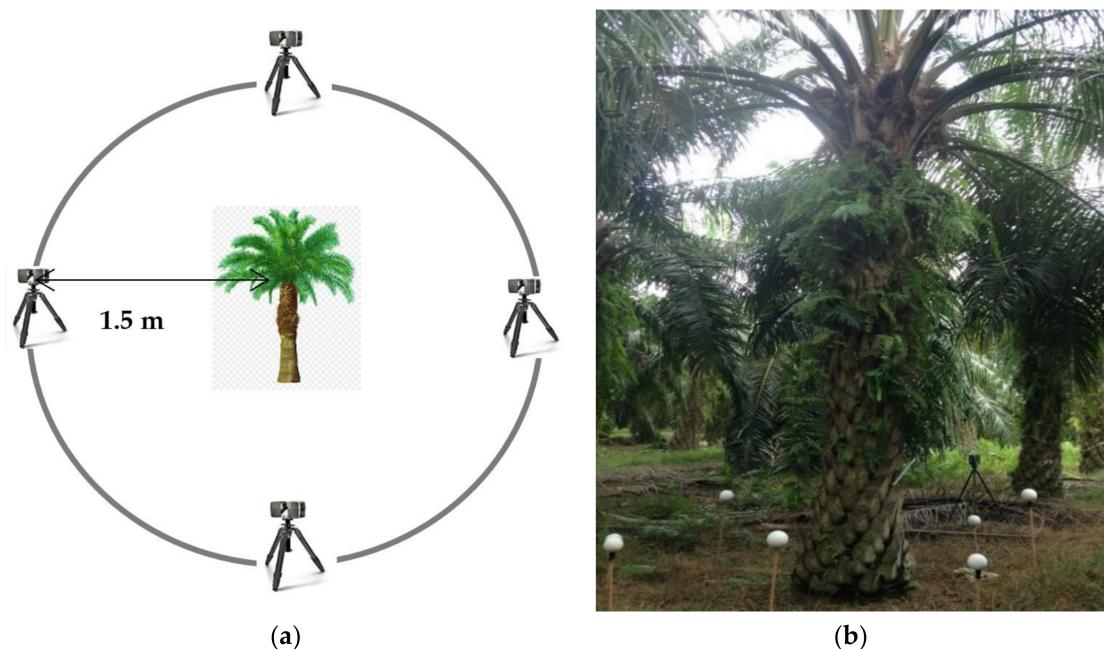
Input Data	Machine Learning Techniques	Accuracy	References
Spectral image	Random forest (RF)	91%	[27]
	SVM	54%	[28]
	Multilayer perceptron (MP)	77%	[32]
	Neural network	100%	[29]
	SVM	89.2%	[26]
	PLS-DA	94%	[11]
Electrical impedance	kNN	97%	[30]
	QDA	80.79%	[31]
	LDA	100%	[32]
	Kernel—SVM	95%	[24]
Fuzzy logic of symptoms	Naïve Bayes (NB)	80%	[25]

It is hypothesized that the BSR infection causes physiological changes in the oil palm trees, which are enhanced and can be easily detected over time. *Ganoderma boninense* fungus produces enzymes that damage the xylem and phloem tissues, which play a crucial role in the storage and transport of water and carbohydrates [34]. Severe water deficits and low carbohydrate intake may limit metabolic functions, thus impeding the tree growth and deteriorating the physical condition of the oil palm trees. Therefore, the use of ML techniques is expected to improve the efficiency and accuracy of the results.

## 2. Materials and Methods

### 2.1. Data Collection

The study area was located at an oil palm plantation in Seberang Perak, Malaysia. The oil palm trees in this replanted plantation area were all approximately 9 years old, when palm oil production is at its peak. The trees were bred from D × P (*Dura0020* × *Psifera*) and the soil type was peat. The study area was located at an average altitude of 19.36 m above sea level and had a topology with a slope variation of less than 1°. Maintenance of the oil palm planting, including fertilization regime, fruit harvesting, pruning, and weed management followed that of the commercial oil palm estates. Adequate pruning of mature palms was required to remove dead or senescing leaves and to allow access to the FFBs at the correct harvesting time. The planting density of the plot was 142 palms/ha and the palms were planted in an equilateral triangle pattern at a distance of 9 × 9 × 9 m. The data used to develop the classification model were collected in early July 2017. This model was later tested using data from different plot collected in September 2018. In order to ensure the validity of the oil palm tree samples during the testing stage, the same experts predetermined the healthiness level of the trees using manual inspections in early July 2017. There was some difference between data collection in early July 2017 (Figure 2b) and September 2018 due to upgrading of the SCENE software (FARO Technologies, Inc., Florida, USA), which allows for data collection without reference spheres. The scanner was mounted on a surveying tripod at a height of 1 m and a distance of 1.5 m from an oil palm tree [35,36]. The depth image from a single viewpoint of the tree did not provide sufficient information to construct a complete, three-dimensional (3D) point cloud model; therefore, each tree was scanned at four different positions [18,35,37,38] (Figure 2).



**Figure 2.** Data collection setup: (a) four scan positions around an oil palm tree; (b) example of the setup at the site with reference spheres around the tree.

A total of 80 oil palm trees used in this study, with 40 each for training and testing process being pre-determined by the experts from the MPOB using manual inspections based on visible symptoms appearing on the tree [39]. For each process, the oil palm trees were categorized into 4 healthiness levels with 10 trees for each level: T0 (healthy palm, no foliage symptoms (0%), no fruiting body), T1 (mild infection, minimal foliage symptoms (0–25%), fruiting body), T2 (moderate infection, foliage symptoms (25–50%), fruiting body), and T3 (severe infection, foliage symptoms (50–75%), fruiting body) [26,37]. Subsequent results and discussions in this paper were based on these levels. The occurrence of *Ganoderma boninense* in the oil palm trees was then confirmed by *Ganoderma* selective medium analysis [40].

A FARO laser scanner (Focus 3D X 120; FARO Technologies, Inc., Lake Mary, USA) was used for data acquisition. It uses a continuous wave (phase-based) laser operating at 905 nm in the infrared section of the light spectrum. The scanner operated at a pulse repetition frequency of 97 Hz and a maximum range of 120 m. The scanner uses high-speed laser technology to acquire millions of 3D laser points for detailed measurement and documentation in a significantly shorter time period. It sends an infrared laser beam to the center of the rotating mirror before projecting it outward from the scanner. Once the infrared light comes into contact with the object, it is then reflected back to the scanner. The output is a 3D set of data in space, known as a point cloud. The scanner used in this research had a beam divergence of 0.27 mrad, resulting in an increased beam diameter of 27 mm per 100 m distance. The vertical field-of-view (FOV) was 305°, limited only by the base of the scanner; the horizontal FOV was 360°.

## 2.2. Data Pre-Processing

The point clouds were pre-processed using the “registration” process to match the multiple scan positions and synchronize the laser point data. SCENE software (FARO Technologies, Inc., Florida, USA) is FARO’s point cloud processing software, which is specifically designed to process and manage scan data simply and efficiently. Two versions of SCENE software were used in this study. No differences were observed between the 2 versions, except for the method using reference spheres in data collection. The data collected in July 2017 were processed using version 5.2, where the

registration process needed to use a reference sphere. Meanwhile, the data collected in September 2018 were processed using version 6.0, where the registration process did not use a reference sphere. Each tree was extracted using a clipping box for further processing steps.

### 2.3. Features

This study aimed to fill the research gap on the capability of ML for BSR classification using TLS data, which has not been explored. Therefore, the same 5 features used in [21] were also used in this study: C200 (crown slice located 200 cm from the top), C850 (crown slice located 850 cm from the top), crown area, frond number, and frond angle. The top is defined as the highest vegetation visible on the tree. C200 and C850 were obtained using the crown stratification method. Both slices were identified using analysis of variance, which differed significantly between slices and health levels. Meanwhile, the crown area, frond number, and frond angle were obtained from the top-down view image. A detailed description of each operation is described in the following sections.

#### 2.3.1. Crown Stratification

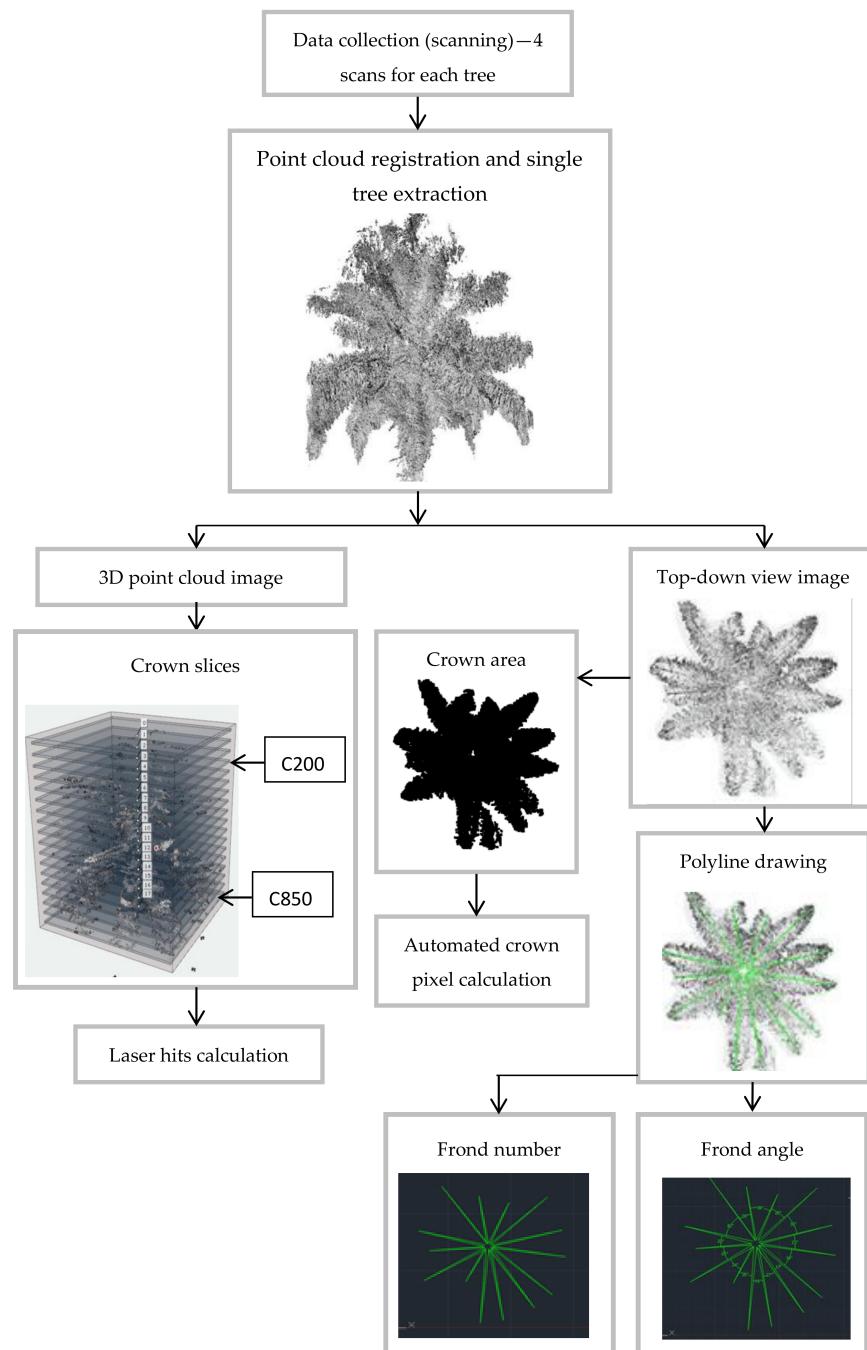
The stratification method was used to segment the data for data reduction [19,40–42]. A series of horizontal layers (known as slices), with equal intervals and size, perpendicular to the vertical direction, were created. The process was started by creating a clipping box that covered the whole area of the canopy to separate the tree from its surroundings. Then, a blue circle was placed at the center of the crown, which was indicated by a hole in the top-down view. The length and width of the clipping box were both set to 6 m to minimize the number of overlapping fronds from neighboring trees. A series of horizontal planes along the vertical axis below the top part of the crown section were created using the “create clipping boxes along an axis” function in SCENE. The thickness of the clipping box was set to 0.1 m and the space between the boxes was set to 0.5 m [43,44]. The 0.5 m space between the boxes was selected arbitrarily but was aimed at reducing the number of laser points associated with the conversion from 3D real space to 2D model space. The number of clipping boxes was set to 17 because the point clouds in the slices above 17 were either empty or mixed with ground vegetation. Then, the laser points in the active clipping boxes were exported to a PTS file that contained the number of laser hits.

#### 2.3.2. Top-Down View Image

The crown area for each tree was determined by counting the number of pixels inside the crown image. A top-down view of the crown image was used, and this image was saved in a JPEG form. For standardization, the image was clipped to a size of  $10 \times 10$  m to ensure that the entire crown area was included. Then, the image was cropped using Paint software (Microsoft Corp., Redmond, USA) to remove unwanted features, such as fronds from other trees, and was later processed using MATLAB software (Mathworks Inc., Natick, USA). Otsu's algorithm was used to separate the crown image from the background. The process involved portioning the crown area by dividing the image into similar areas of connected pixels. Otsu is a global thresholding that separates objects of interest from the background of an image on the basis of its gray level distribution. The algorithm assumes that the image contains two classes of pixels following a bi-modal histogram (foreground pixels and background pixels). It then calculates the optimum threshold that separates the two classes so that their combined spread (intra-class variance) is minimal. The method is widely used because it is simple and effective [45]. Then, the morphological operation (opening) was used to remove imperfections in the structure of the crown image. The opening operation smooths the outline of the object, clears the narrow bridges, and eliminates minor extensions present in the crown image by performing erosion followed by dilation. Both operations use a small, disc-shaped matrix structure with a radius of 11 cm (distance from the origin to the edge of the disc).

The same image of the crown area was used for frond counting. The image was uploaded and analyzed using AutoCAD software (Autodesk, Inc., San Rafael, USA). The “layers” tab was selected

to create a new layer, and this was then set as the current layer. Then, the “draw” tool was selected, and the “polyline” function was used to draw the frond. The lines were drawn following the shape of the frond on the basis of the 2D top-down image only. Therefore, the fronds that overlapped or were located below, which were not visible in the top-down image, were not considered. The same image was also used for frond angle measurement. The “dimension” tab was selected to determine the angle between each frond, which measures the space between the fronds. Angle measurement was continued to the adjacent fronds until all frond angles were measured. Finally, the number of fronds and the angle between the fronds were recorded for each tree. Figure 3 shows a step-by-step schematic view of the methods used for feature extractions such as crown area, crown density, and frond parts.



**Figure 3.** Schematic of the method used for feature extraction.

#### 2.4. Principal Component Analysis

Principal component analysis (PCA) is a transformation of the original data into a set of linearly uncorrelated variables, called principal components (PCs), to reduce the dimensionality of dataset and increase its interpretability while at the same time minimizing information loss. PCA is the most common technique in environmental and biological sciences, where the datasets are invariably small due to high cost of the samples collection and analysis, as it can visualize important underlying data structures [46]. Therefore, PCA can be useful when the discrimination problem exists, and statistical tests cannot be performed due to the relatively small sample size [47]. Moreover, PCA allows the use of variables that are not measured in the same units, thus revealing the general covariance structure inherent in the data matrix. In addition, PCA data transformation allows the determination of a subset of a feature with better classification accuracy than that achieved when using all the features because the presence of redundant features with reduced discriminative power can obscure the classifier [48]. In this study, JMP software (SAS Institute, North Carolina, USA) was used to run the PCA of the five features, i.e., C200, C850, crown area, frond number, and frond angle. The formal criterion was used to determine the number of PCs, requiring at least 80% of the total variation explained by the number of components [49]. The scree plot was used to maintain the number of PCs, which involved eye-balling of the eigenvalues plot for a break, turning point, or a sharp drop (also referred to as an “elbow”) [50]. Bartlett’s test was also used to check the hypothesis that the correlation matrix is an identity matrix, where insignificant values would indicate that the variables are unrelated and that the original data reduction may not be useful [51].

#### 2.5. Machine Learning

The ML approach is not only used for the classification of large number of samples but also for small sample numbers [52,53]. In addition, according to [54], the number of samples used in this study was sufficient for implementing the ML approach, while PCA was used to transform all the five oil palm features, consisting of C200, C850, crown area, frond number, and frond angle into numbers of PCs. In this study, the combinations of PCs were used as the inputs for the ML. The classification learner app toolbox in the MATLAB software (Mathworks Inc., Massachusetts, USA) was used to fit different models using ML approaches, and six ML methods available in the software were used, as described below:

1. **Decision tree (DT)**—splits a dataset into a number of tree-like models, consisting of nodes for testing attributes, edges for branching by value of selected attributes, and leaves for labelling classes, where each unique leaf is attached.
2. **Discriminant analysis (DA)**—assigns observations to previously defined groups, by examining the differences between two or more groups of objects with respect to several variables simultaneously.
3. **Naïve Bayes (NB)**—assigns each object to the class with the highest conditional probability, with a strong assumption of independence between the parameters.
4. **Support vector machine (SVM)**—creates a line or hyperplane in dimensional space to distinctly separate the data into classes.
5. **Nearest neighbor (NN)**—stores existing cases and classifies new cases on the basis of similarity measures (e.g., distance functions).
6. **Ensemble modelling (EM)**—constructs a set of classifiers and then classifies new data by taking a (weighted) vote of their combined predictions.

Due to the small number of samples, we performed a cross-validation process to evaluate the model performance. K-fold cross-validation function in MATLAB was used to split the data into training and validation datasets and was used to obtain independent evaluation of the model accuracy. Thus, the resulting model was more valid and not limited to only one set of data. The developed models used 5 repeated cross-validations (5 folds), where the original samples were randomly partitioned into 5 equal-sized subsamples. A total of 30 classification models were developed using different

combinations of PCs with various types of kernels, as shown in Table 2. Further explanations of each model and kernels are provided in [55]. To limit the number of classification models, we only selected models with accuracies greater than 70%. Every selected model was then run at 5 iterations to check its stability. The model accuracy, which is defined by averaging a total value of model accuracy in 5 iterations, was calculated, and the models with accuracies reduced by more than 5% were excluded. Afterwards, the selected models were tested using new data collected in September 2018. The best model from each combination of PCs was determined using the results of model accuracy, the accuracies obtained for healthy trees (T0), and the overall accuracies for 4 healthiness levels (also known as multiple levels).

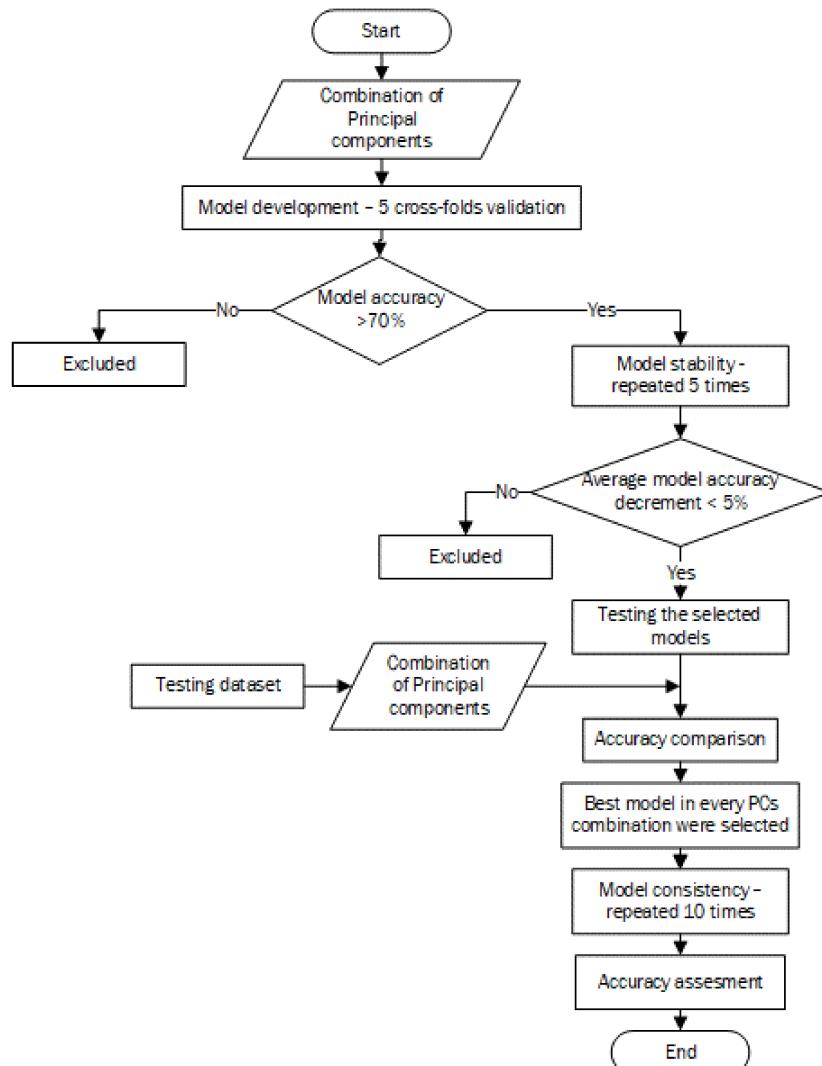
**Table 2.** List of classification models and kernel.

Classification Model	Types of Kernel
Decision tree (DT)	Fine tree, medium tree, coarse tree, all trees
Discriminant analysis (DA)	Linear discriminant, quadratic discriminant, all discriminant
Naïve Bayes (NB)	Gaussian NB, kernel NB, all NB
Support vector machine (SVM)	Linear SVM, quadratic SVM, cubic SVM, fine Gaussian SVM, medium Gaussian SVM, coarse Gaussian SVM, all SVMs
Nearest neighbor (NN)	Fine kNN, medium kNN, coarse kNN, cosine kNN, cubic kNN, weighted kNN, all kNNs
Ensemble modelling (EM)	Boosted trees, bagged trees, subspace discriminant, subspace kNN, RUSBoosted trees, all ensembles

Furthermore, in order to determine the consistency of the best models with different combinations of PCs, we ran the models at 10 iterations. The accuracy of each model, determined by averaging the total model accuracy in 10 iterations, was calculated. On the basis of the model accuracy, accuracy assessment [56], and Kappa coefficient value [57], we chose only one model as the best. In the accuracy assessment, we presented the overall accuracy in the confusion matrix [58] for the classification of 4 levels of healthiness. In the confusion matrix, the columns correspond to the true class, and the rows correspond to the predicted class. The diagonal cells correspond to observations that were correctly classified, and the off-diagonal cells correspond to incorrectly classified observations. The Kappa coefficient or Cohen's Kappa is a standardized statistical measurement for inter-rater or intra-rater reliability testing, and the coefficient calculation was based on the study of [57]. In addition, the user's and producer's accuracies were also stated. The user's accuracy is the class percentage that actually represents the right class, and it was calculated by taking the total number of correct classifications for a particular class and dividing it by the row total. Meanwhile, the producer's accuracy is the percentage of reference class being properly classified, and it was calculated by taking the total number of correct classifications for a particular class and dividing it by the column total. Figure 4 illustrates the ML process using combinations of PCs as inputs.

Of note, the contributions of the features to a PC can be assessed to determine if any features strongly influence a particular PC. Additionally, when interpreting the component loadings (eigenvector components multiplied by the square root of the associated eigenvalue), we can use the squared cosine between features and a PC to determine if real correlations exist or if there is only an apparent relation due to the projection onto a low-dimensional subspace [59]. To infer a correlation, there should be clustering on a two-dimensional (2D) loading plot, and the squared cosine should be greater than one half. As in all statistical interpretations, the best practice is to examine multiple sources of information, i.e., the scree plot, eigenvalues, loadings, 2D scatter plots of observations projected on the PCs, and squared cosines for features and partial contributions of the features. We examined the component loadings, which are the product of square root of eigenvalue with eigenvector. The correlation matrix also revealed the correlation coefficients (cosines) between the variables and factors (PCs). Additionally, the squared cosine, also called "relative contributions", which are defined as the ratios between the vector norms in the PC space and the original norms, was computed [60].

Meanwhile, the partial contributions enable us to observe the percentage that each variable contributes to each principal component.



**Figure 4.** Machine learning process.

### 3. Results and Discussion

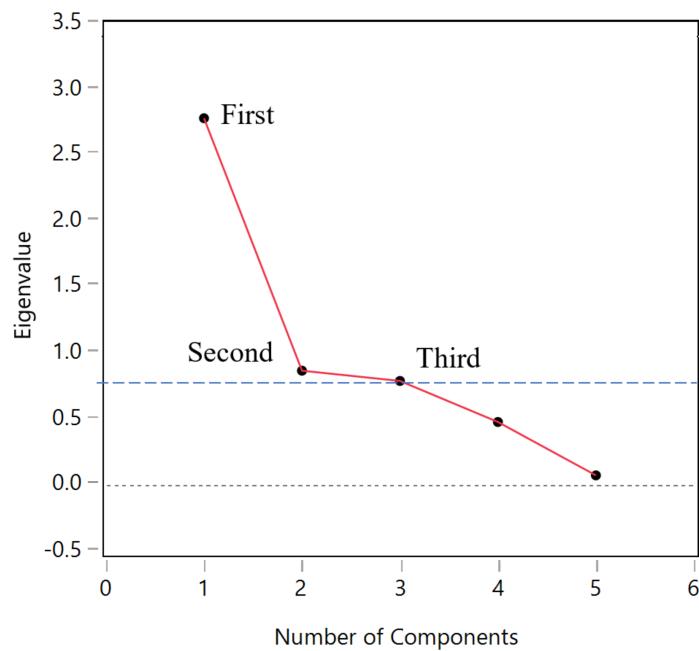
#### 3.1. Principal Component Analysis

Table 3 shows the eigenvalues, percentage, and Bartlett's sphericity test of PC1 to PC5. The results were significant with a *p*-value of less than 0.0001. In PCA, the eigenvectors of the PCs determine the direction of the new feature space, and the eigenvalues determine their magnitude. In other words, the eigenvalues explain the variance of the data along the new feature axes. Following the criteria set for the use of all PCs before the drastic change of the slope, excluding the drastic change of the first PC (Figure 5a) and at least 80% of variation data, we obtained PCs that summed to 88.83% (Figure 5b), with a total variation that consisted of a first PC with an eigenvalue of 2.780 and a variance of 55.61%, followed by a second PC with an eigenvalue of 0.8693 and a variance of 17.39% and a third PC with an eigenvalue of 0.792 and a variance of 15.84% (Table 3). On the basis of these data, we used three PC combinations in the ML classification model. The first was a combination of PC1 and PC2, the second was a combination of PC1 and PC3, and the third was a combination of PC1, PC2, and PC3.

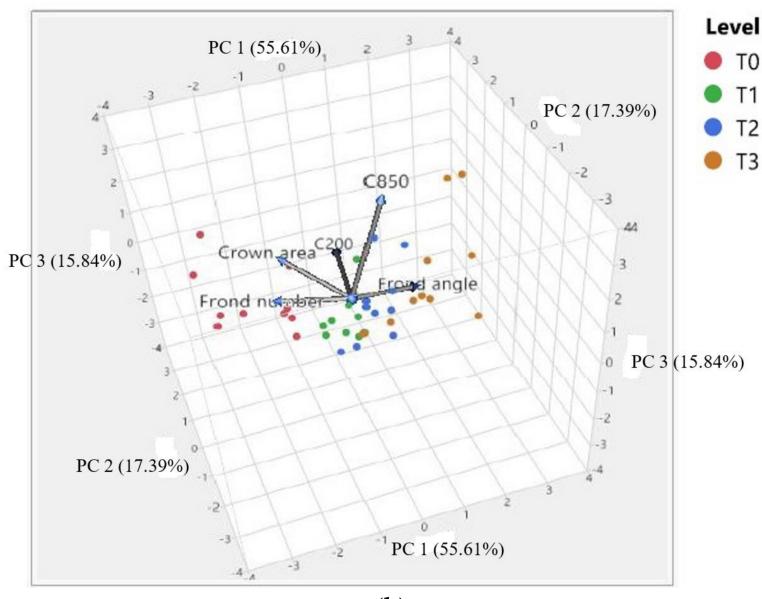
**Table 3.** Eigenvalues, percentage, and Bartlett's sphericity test of the principal components (PCs).

Number	Eigenvalue	Variance Percent	Cumulative Percent	Chi Square	df	Bartlett's Test Prob > ChiSq
1	2.780	55.609	55.609	99.127	10.433	<0.0001 *
2	0.869	17.387	72.996	48.254	9.294	<0.0001 *
3	0.792	15.837	88.833	40.902	5.759	<0.0001 *
4	0.481	9.622	98.455	26.298	2.989	<0.0001 *
5	0.077	1.545	100.000	0.000	0.842	1.0000

\* significant at 5% level.



(a)



(b)

**Figure 5.** (a) Scree plot of the principal component analysis (PCA) eigenvalues and (b) principal component score 3D plot.

### 3.2. Classification Model

A total of 30 classification models were developed using different kernel types (Table 2), as stated in Section 2.2. The results of the model classification with different PC combinations are shown in Tables 4–6, where the left columns present models with an accuracy higher than 70% and a reduction in the model accuracy of less than 5% when ran in five iterations. Meanwhile, the right columns present the accuracy of the same models when tested using different types of data taken from a different plot in September 2018.

**Table 4.** Results of the classification model using PC1 and PC2 (variance of 73%).

Type of Model and Accuracy (%)	Accuracy When Tested Using Different Dataset (%)	
	T0 Classification	Multiple Level Classification (T0, T1, T2, T3)
1. Decision tree—fine (72.50)	90.00	80.00
2. Quadratic discriminant (75.00)	100.00	75.00
3. Kernel naïve Bayes (77.50) *	100.00 *	85.00 *
4. SVM—linear (72.50)	100.00	80.00
5. Fine kNN (75.00)	100.00	82.50
6. Ensemble bagged trees (72.50)	90.00	80.00

Note: (\*) Shows the highest accuracy.

**Table 5.** Results of the classification model using PC1 and PC3 (variance of 71.45%).

Type of Model and Accuracy (%)	Accuracy When Tested Using Different Dataset (%)	
	T0 Classification	Multiple Level Classification (T0, T1, T2, T3)
1. Decision tree—coarse (82.50)	80.00	57.50
2. Quadratic discriminant (85.00)	20.00	55.00
3. Gaussian NB (77.50)	80.00	70.00
4. Medium Gaussian SVM (82.50) *	80.00 *	77.50 *
5. Fine kNN (85.00)	80.00	67.50
6. Ensemble bagged trees (82.50)	80.00	62.50

Note: (\*) Shows the highest accuracy.

**Table 6.** Results of the classification model using PC1, PC2, and PC3 (variance of 88.83%).

Type of Model and Accuracy (%)	Accuracy When Tested Using Different Dataset (%)	
	T0 Classification	Multiple Level Classification (T0, T1, T2, T3)
1. Decision tree—fine (70.00)	90.00	80.00
2. Linear discriminant (70.00)	100.00	80.00
3. Gaussian NB (70.00)	100.00	80.00
4. Linear SVM (72.50)	90.00	80.00
5. Ensemble subspace discriminant (72.50) *	100.00 *	85.00 *

Note: (\*) shows the highest accuracy.

Table 4 shows the results of models developed using a combination of PC1 and PC2, which held the variance from the original data of 73%. On the basis of the data of this table, we clearly see that six models were selected with the model accuracy between 72.5 and 77.5%. Quadratic discriminant, kernel NB (KNB), linear SVM, and fine kNN achieved 100% accuracy for healthy trees (T0) when tested using a different dataset. In contrast, the fine decision tree and ensemble bagged trees showed

only 90% accuracy for healthy trees. Among the classification models that achieved 100% accuracy for healthy trees, KNB achieved the highest model accuracy with 77.5%. Furthermore, the KNB model had the highest multiple level classification accuracy for the four healthiness levels (T0 to T3) at 85%.

Results of the classification models using PC1 and PC3, which held constant the variance of the original data at 71.45%, is shown in Table 5. As shown in this table, the accuracies of all six selected models were higher than 80%, except for the Gaussian NB model (77.5%). The highest model accuracy was achieved by the quadratic discriminant and fine kNN models, which had an 85% accuracy. These were followed by the coarse decision tree, medium Gaussian SVM (MGS), and ensemble bagged trees models, with 82.5% accuracy. Although higher model accuracy was achieved using PC1 and PC3, all models were only able to classify healthy trees with 80% accuracy when tested using a different dataset, except for quadratic discriminant, which had the lowest score at 20%. For the multiple level classification (T0 to T3), in general, the accuracies of the models were lower than models using PC1 and PC2. The highest accuracy was only achieved at 77.5% using the MGS model.

The classification models using PC1, PC2, and PC3, which held the variance from the original data at 88.83%, are shown in Table 6. Linear SVM and ensemble subspace discriminant (ESD) had the highest model accuracy, with 72.5%, while fine decision tree, linear discriminant, and Gaussian NB achieved 70% accuracy. When tested using a different type of dataset, the linear discriminant, Gaussian NB and ESD achieved 100% accuracy for healthy trees, while fine decision tree and linear SVM achieved 90% accuracy. For multiple level classification, ESD gave the best accuracy (85%), while the other models achieved 5% less accuracy, which was only 80%.

### 3.3. Best Model Comparison

The best models identified with different types of PC combinations were KNB with PC1 and PC2 inputs; MGS with PC1 and PC3 inputs; and ESD with PC1, PC2, and PC3 inputs. Their performance in terms of model accuracy, multiple levels accuracy, and Kappa coefficient is presented in Table 7, where it is clear that after 10 iterations, the model accuracy of KNB decreased by as much as 3% from 77.5% to 74.5%. There was a slight decrease in accuracy for MGS of 0.25% from 82.5% to 82.25% and a 0.5% decrease for ESD from 72.5% to 72%. The MGS had the highest model accuracy, followed by the KNB and ESD. Meanwhile, the KNB and ESD shared the same multiple levels accuracies (85%), and they only differed in the Kappa coefficient, where KNB was 0.80 and ESD was 0.85. The MGS served the lowest multiple levels accuracy and Kappa coefficients compared with the KNB and ESD.

**Table 7.** Classification accuracy.

Accuracy	Machine Learning Model		
	KNB	MGS	ESD
Model accuracy	74.5%	82.25%	72%
Multiple levels accuracy	85%	77.50%	85%
Kappa coefficient	0.80	0.77	0.85

Even though the average model accuracy of the KNB (74.5%) decreased after 10 iterations, it was still higher compared with the ESD (72%). However, the ESD gave the best Kappa coefficient with a value of 0.85. This value indicates that the model has the same strong agreement, which is a slightly higher value than that of the KNB. The KNB results were similar to those of the ESD, and the model accuracy for the KNB was higher than for the ESD, at 74.5% compared with 72.5%, respectively.

In addition, on the basis of the confusion matrix shown in Tables 8–10, the classification of the BSR disease, which was developed using the KNB model (Table 8), gave a better result compared with the MGS (Table 9), with 100% user's accuracies at the T0 and T1 levels, whereas the ESD model could only classify the T0 level with 100% user's accuracy (Table 10). It was also found that although the MGS had the highest model accuracy (82.25%), it could not classify a healthy tree (T0) or mild infection (T1), thus providing only 80% user's accuracy (Table 9). Moreover, the KNB model could classify the

healthiness level at a very early stage, i.e., T1 at 100% user's accuracy, which is important for the early detection problem of the BSR disease [29]. In addition, the perfect T0 classification is important for evading a false positive error [30]. It also showed that the high percentage of variance from the original data taken from PCs resulted in higher Kappa coefficient values, i.e., ESD (with PC variance of 88.83%) > KNB (with PC variance of 73%) > MSG (with PC variance of 71.45%). However, it could not perfectly detect the mild infection as well as the KNB. Therefore, the KNB that used the PC1 and PC2 as input parameters was identified as the best model for the BSR detection in this study. In the multiple level classification, MGS and ESD performed better for the T2 level than KNB, with 70% user's accuracy, whereas for the T3 level, KNB and ESD performed better than MGS with 80% user's accuracy. In terms of producer's accuracy, all models afforded 100% accuracy for the T0 level. In contrast, ESD (90%) and KNB (75%) gave the highest accuracies for the T1 and T2 levels, respectively, while for the T3 level, KNB and ESD had the same accuracy (80%), which was higher than that of MGS (77.78%).

**Table 8.** Confusion matrix of kernel naïve Bayes (KNB; input of PC1 and PC2).

		Reference Data					
		T0	T1	T2	T3	Row Total	User's Accuracy
Classified Data	T0	10	0	0	0	10	100%
	T1	0	10	0	0	10	100%
	T2	0	2	6	2	10	60%
	T3	0	0	2	8	10	80%
	Column total	10	12	8	10	40	
Producer's accuracy		100%	83.33%	75%	80%	Overall accuracy = 85%	

**Table 9.** Confusion matrix of medium Gaussian support vector machine (MGS; input of PC1 and PC3).

		Reference Data					
		T0	T1	T2	T3	Row Total	User's Accuracy
Classified Data	T0	8	2	0	0	10	80%
	T1	0	9	1	0	10	90%
	T2	0	1	7	2	10	70%
	T3	0	0	3	7	10	70%
	Column total	8	12	11	9	40	
Producer's accuracy		100%	75%	63.64%	77.78%	Overall accuracy = 77.50%	

**Table 10.** Confusion matrix of ensemble subspace discriminant (ESD; input of PC1, PC2, and PC3).

		Reference Data					
		T0	T1	T2	T3	Row Total	User's Accuracy
Classified Data	T0	10	0	0	0	10	100%
	T1	0	9	1	0	10	90%
	T2	0	1	7	2	10	70%
	T3	0	0	2	8	10	80%
	Column total	10	10	10	10	40	
Producer's accuracy		100%	90%	70%	80%	Overall accuracy = 85%	

The difference in accuracy results is due to the different classifiers having different characteristics, where different kinds of classifiers are sensitive to various parameter optimizations. The KNB model that is developed on the basis of naïve Bayes showed better performance during model assessment, probably due to its simple probabilistic classifier based on Bayes' theorem, which interprets the probability of an event on the basis of previous knowledge of conditions that may be associated with that event. The naïve description of NB is due to the assumption of mutual independence between the input characteristics. It works well in this study due to its capability to use a small number of training statistics to estimate the means and variances of the variables important for class [61,62] and performs well when the data cannot support a more complex classifier [63]. Furthermore, kernels used

in kernel density estimation to estimate random variables' density features in these non-parametric estimation strategies were performed by controlling the classifier model predictor distribution by changing the parameters for model flexibility and smoothing the density. Non-parametric estimators have no constant shape and rely on all the record points to reach an estimate. In addition, the lower accuracy during the model development that could be attributed to the independence assumption between the features in this study was not entirely valid. However, the naïve Bayesian classifier is remarkably effective, even if the independent assumption seldom holds [64]. KNB is simpler and better understood, and the classification algorithm is effective for biological problems [65], i.e., the disease classification in oil palm trees. In the field of remote sensing, NB classifier is often used as another popular method for classification problems, where it does not require a complicated iterative parameter estimation scheme and produces a much higher accuracy [66,67]. Studies have also shown that the classification accuracies slightly increased when the training set size was increased [68,69]. The NB classifier was also used to classify oil palm disease on the basis of its symptoms with high accuracy: 80% [25], 84% [70], and 92.25% [71], which showed that the NB classifier is well suited with the oil palm disease classification. Moreover, the classification model using the ML approach gave better results compared to the classification model using a combination of features [21], where the accuracy for multiple levels was 80%.

### 3.4. Contribution of PCs to the ML Models

The statistical analysis shown in Table 11 using the Kruskal–Wallis test for all the features showing significant differences between the levels of healthiness at the 5% significance level. Crown area, frond angle, and frond number presented the lowest *p*-value, less than 0.01%, followed by C200, with *p*-value of 0.0166, and C850, with *p*-value of 0.0471. The highest chi-squared value was obtained by frond number at 33.428, and the lowest was obtained by C850 at 8.058.

**Table 11.** Kruskal–Wallis test for each feature.

Features	Chi-Squared Value	<i>p</i> -Value
C200	10.248	0.0166 *
C850	8.058	0.0471 *
Crown area	23.058	<0.0001 *
Frond angle	32.666	<0.0001 *
Frond number	33.428	<0.0001 *

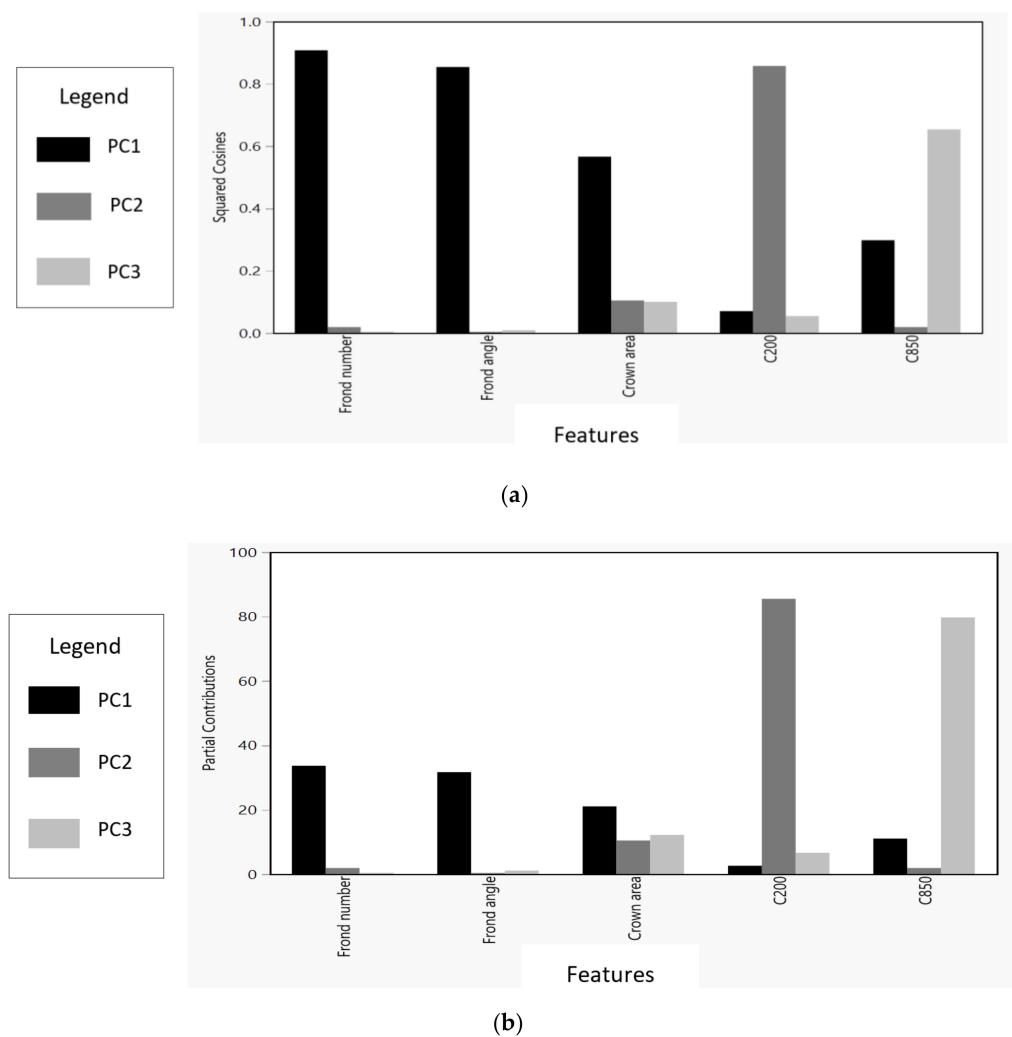
\* significant at 5% level.

Table 12 shows the loading matrix, which is a correlation between the original variables (features) and the PCs. The closer the value is to one, the greater the effect of the PC on the feature. The loading of PC1 attributes the most weight to all high significant features (*p* < 0.0001), i.e., frond number, frond angle, and crown area. The loading of PC2 attributes the most weight to C200, whereas the loading of PC3 attributes the most weight to C850. On the basis of the contribution of each PC and the correlation with the significant features, different combinations of PCs will exhibit different results when run using different models of ML. Despite the different algorithms in each classification model, the input using different combinations of PCs influence the accuracy of the results. For example, for the best classification model, KNB used inputs from the combination of PC1 and PC2, where it is most weighted in terms of frond number, frond angle, crown area, and C200. The best classification accuracy with the most weighted to frond number, frond angle, crown area, and C850 was achieved by MGS. However, it gave the lowest classification accuracy among the best identified models, suggesting that the model needs more loading of C200 than with C850 for better results.

**Table 12.** Loading matrix of the features on the five significant PCs.

Features	PC 1	PC 2	PC 3	PC 4	PC 5
Frond number	−0.93154	0.03556	0.19416	−0.22877	0.20233
Frond angle	0.91567	0.06992	−0.12020	0.32764	0.18672
Crown area	−0.74945	0.17597	0.34380	0.53670	−0.03335
C200	0.50580	0.73751	0.41017	−0.17804	−0.01753
C850	0.50667	−0.53695	0.67326	−0.04113	0.00270

Figure 6 shows the squared cosines and partial contributions of the features that can be used to evaluate the achievement results. The correlations can be inferred to the features with a squared cosine value greater than one half [59]. The greater the squared cosine, the greater the relationship with the corresponding PC. As shown in Figure 6a, the PC1 clearly captured the behavior of the first three features, i.e., frond number, frond angle, and crown area. However, the PC2 captured the behavior of the C200 feature, and the PC3 captured the behavior of the C850 feature. Similar behaviors of the partial contribution can also be seen in Figure 6b, where the PC2 had a negligible contribution to the feature of the C850 and the PC3 had a negligible contribution to the feature of the C200.

**Figure 6.** (a) Squared cosines of the features and (b) partial contributions of the features.

The results have shown that the input features play an important role in ensuring accurate classification. Architectural features usually have a more significant impact on performance than the

choice of the individual algorithm in an ML approach [72]. As pointed out in [73], it is fair to say that there is probably no model that generally outperforms the others. For practical problems, the choice of approach will depend on constraints, namely, the required accuracy, the size of the data, time available for development and training, and the nature of the classification problem. A simple, efficient, and accurate analysis provides a good choice for multi-class classification problem solving [74].

### 3.5. Application in the Plantation

Manual observation is laborious and subjectively depends on the individual workers who are susceptible to tiredness. Although the proposed method in this study is time-consuming in terms of the data acquisition process, it has a bright future when considering the evolution of laser scanning technology, which continually has improved scanning times, spatial coverage, and data quality and provides various platform types. This approach of scanning a single tree can thus be improved by scanning a group of trees at a time, where the scanning time and processing data can be decreased. Mobile laser scanning (MLS) can be considered while attaching the sensor with a trailer or tractor to speed up the data acquisition process by running simultaneously with plantation operations, such as fertilizing and fruit collecting. Moreover, the approach can be expanded by using different type of platform, i.e., aerial imagery or drone, which can significantly cover a larger area. The application could be based on three features extracted from the top view, namely, the crown pixel, frond number, and frond angle. As shown in Table 13, ESD delivered the best results using these three features that could be seen from the aerial view with an overall accuracy of 70%. Consequently, ESD showed a great potential of aerial application for BSR detection. However, a detailed study is required because the ground-based and aerial imagery are varied in terms of data capture mode, scanning mechanism, typical project size, resolution, and accuracy that can be obtained.

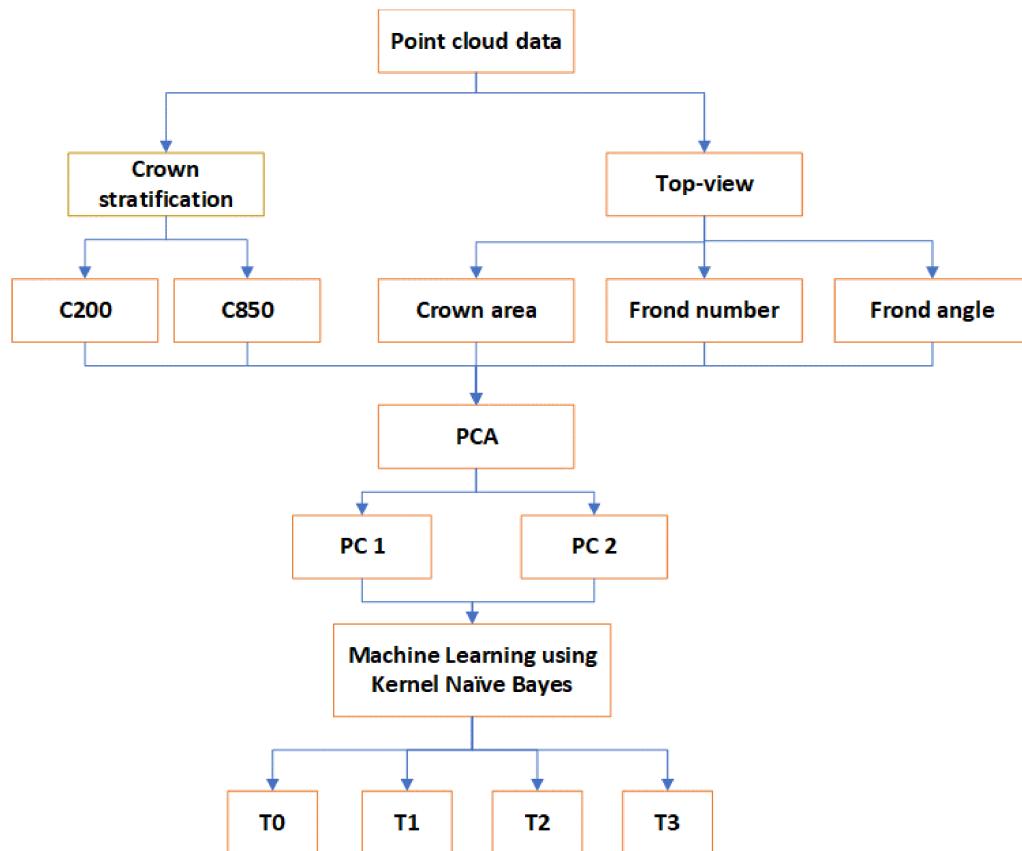
**Table 13.** Confusion matrix of ESD (input of PC1, PC2, and PC3) using three features: crown pixel, frond number, and frond angle.

Class	Reference Data				Row Total	User's Accuracy
	T0	T1	T2	T3		
Classified Data	T0	10	0	0	10	100%
	T1	0	7	3	10	70%
	T2	0	3	6	10	60%
	T3	0	0	5	10	50%
Column total		10	10	14	40	
Producer's accuracy		100%	70%	43%	83%	Overall accuracy = 70%

## 4. Conclusions

In this study, the physical properties of oil palm trees were extracted using TLS data and divided into four levels of health: healthy, mild infection, moderate infection, and severe infection. Five features were used, namely, C200, C850, crown area, frond number, and frond angle. Three different types of PC combinations were used as inputs for ML classification. For each type of PC combination, we developed 30 classification models using six ML methods: decision trees, discriminant analysis, the NB algorithm, SVM, nearest neighbor, and ensemble modelling with different types of kernels. Among the classification techniques, when using PC1 and PC2; PC1 and PC3; and PC1, PC2, and PC3 classification models, kernel naïve Bayes (KNB), medium Gaussian SVM (MGS), and ensemble subspace discriminant (ESD) achieved the highest accuracies, respectively. In conclusion, the KNB was identified as the best model with a model accuracy of 74.5%, multiple levels accuracies of 85%, and Kappa coefficients of 0.80. This is might have been due to the advantages of the effectiveness of the biological application and the smaller quantity of training data. The used features in the study, i.e., frond number, frond angle, crown area, and C200 provided significant contributions and helped in providing accurate classifications. The major benefit derived from this study is the development of a model suitable for the early detection and severity level classification of BSR disease in oil palm trees caused by

*Ganoderma boninense* using ML techniques and point cloud data of crown properties. In future studies, the database will also be broadened to improve the classification accuracy. Figure 7 summarizes the best method (KNB) for the detection of BSR disease and classification into the four different severity levels.



**Figure 7.** Flow of the KNB method for detecting and classifying the BSR into four different severity levels.

**Author Contributions:** Conceptualization, N.A.H. and S.K.B.; methodology, N.A.H. and S.K.-B.; software, N.A.H. and S.K.-B.; validation, S.K.-B.; formal analysis, N.A.H.; investigation, N.A.H. and M.H.A.A.; resources, S.K.-B. and D.A.; data curation, N.A.H. and S.K.-B.; writing—original draft preparation, N.A.H.; writing—review and editing, S.K.-B., A.F.A., and M.S.M.K.; visualization, N.A.H. and S.K.-B.; supervision, S.K.-B.; project administration, S.K.-B.; funding acquisition, S.K.-B. and D.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Malaysia Ministry of Higher Education (MOHE) and Universiti Putra Malaysia (UPM) under grant number LRGS-NANOMITE/5526305.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Naher, L.; Yusuf, U.K.; Ismail, A.; Tan, S.G.; Mondal, M.M.A. Ecological status of '*Ganoderma*' and basal stem rot disease of oil palms ('*Elaeis guineensis*' Jacq.). *Aust. J. Crop Sci.* **2013**, *7*, 1723.
2. Chong, K.P.; Dayou, J.; Alexander, A. Pathogenic Nature of *Ganoderma boninense* and Basal Stem Rot Disease. In *Detection and Control of Ganoderma boninense in Oil Palm Crop*; SpringerBriefs in Agriculture; Springer: Cham, Switzerland, 2017; pp. 5–12. [CrossRef]
3. Barnes, C.; Balzter, H.; Barrett, K.; Eddy, J.; Milner, S.; Suárez, J.C. Airborne laser scanning and tree crown fragmentation metrics for the assessment of *Phytophthora ramorum* infected larch forest stands. *For. Ecol. Manag.* **2017**, *404*, 294–305. [CrossRef]
4. Vossen, P.M. *Organic Olive Production Manual*; UCANR Publications: Oakland, CA, USA, 2007.

5. Waring, R.H. Characteristics of trees predisposed to die. *Bioscience* **1987**, *37*, 569–574. [[CrossRef](#)]
6. Horbach, R.; Navarro-Quesada, A.R.; Knogge, W.; Deising, H.B. When and how to kill a plant cell: Infection strategies of plant pathogenic fungi. *J. Plant Physiol.* **2011**, *168*, 51–62. [[CrossRef](#)]
7. Srinivasan, N. Diseases and disorders of coconut and their management. In *Plant Pathology*; Trivedi, P.C., Ed.; Pointer Publishers: Jaipur, India, 2001; pp. 194–254.
8. Srinivasulu, B.; Aruna, K.; Krishna, P.; Rajamannar, M.; Sabitha, D.; Rao, D.V.R.; Hameed, K.H. Prevalence of basal stem rot disease of coconut in coastal agro ecosystem of Andhra Pradesh. *Indian Coconut J.* **2002**, *XXXIII*, 23–26.
9. Broschat, T.K. Palm Morphology and Anatomy. The Environmental Horticulture Department, Florida Cooperative Extension Service, Institute of Food and Agricultural Sciences, University of Florida. ENH1212, 2016. Available online: <https://edis.ifas.ufl.edu/ep473> (accessed on 24 June 2018).
10. Corley, R.H.V.; Tinker, P.B. *The Oil Palm*, 4th ed.; Blackwell Science: Oxford, MS, USA, 2008.
11. Lelong, C.C.; Roger, J.M.; Brégand, S.; Dubertret, F.; Lanore, M.; Sitorus, N.A.; Raharjo, D.A.; Caliman, J.P. Evaluation of oil-palm fungal disease infestation with canopy hyperspectral reflectance data. *Sensors* **2010**, *10*, 734–747. [[CrossRef](#)]
12. Nisfariza, M.N.; Shafri, Z.H.; Idris, A.; Steven, M.; Boyd, D.; Mior, M. Hyperspectral sensing possibilities using continuum removal index in early detection of *Ganoderma* in oil palm plantation. In Proceedings of the World Engineering Congress 2010, Conference on Geomatics and Geographical Information Science, Kuching, Malaysia, 2–5 August 2010; pp. 233–239.
13. Fritz, A.; Kattenborn, T.; Koch, B. UAV-based photogrammetric point clouds—Tree stem mapping in open stands in comparison to terrestrial laser scanner point clouds. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2013**, *40*, 141–146. [[CrossRef](#)]
14. Kankare, V.; Holopainen, M.; Vastaranta, M.; Puttonen, E.; Yu, X.; Hyppä, J.; Vaaja, M.; Hyppä, H.; Alho, P. Individual tree biomass estimation using terrestrial laser scanning. *ISPRS J. Photogramm. Remote Sens.* **2013**, *75*, 64–75. [[CrossRef](#)]
15. Lin, Y.; Herold, M. Tree species classification based on explicit tree structure feature parameters derived from static terrestrial laser scanning data. *Agric. For. Meteorol.* **2016**, *216*, 105–114. [[CrossRef](#)]
16. Moorthy, I.; Miller, J.R.; Berni, J.A.J.; Zarco-Tejada, P.; Hu, B.; Chen, J. Field characterization of olive (*Olea europaea* L.) tree crown architecture using terrestrial laser scanning data. *Agric. Meteorol.* **2011**, *151*, 204–214. [[CrossRef](#)]
17. Palace, M.; Sullivan, F.B.; Ducey, M.; Herrick, C. Estimating Tropical Forest Structure Using a Terrestrial Lidar. *PLoS ONE* **2016**, *11*, e0154115. [[CrossRef](#)] [[PubMed](#)]
18. Raumonen, P.; Kaasalainen, M.; Åkerblom, M.; Kaasalainen, S.; Kaartinen, H.; Vastaranta, M.; Holopainen, M.; Lewis, P. Fast automatic precision tree models from terrestrial laser scanner data. *Remote Sens.* **2013**, *5*, 491–520. [[CrossRef](#)]
19. Trochta, J.; Krůček, M.; Vrška, T.; Král, K. 3D Forest: An application for descriptions of three-dimensional forest structures using terrestrial LiDAR. *PLoS ONE* **2017**, *12*, e0176871. [[CrossRef](#)] [[PubMed](#)]
20. Azuan, N.H.; Khairunniza-Bejo, S.; Abdullah, A.F.; Kassim, M.S.M.; Ahmad, D. Analysis of Changes in Oil Palm Canopy Architecture From Basal Stem Rot Using Terrestrial Laser Scanner. *Plant Dis.* **2019**, *103*, 3218–3225. [[CrossRef](#)]
21. Husin, N.A.; Khairunniza-Bejo, S.; Abdullah, A.F.; Kassim, M.S.; Ahmad, D.; Azmi, A.N. Application of Ground-Based LiDAR for Analysing oil palm canopy properties on the occurrence of Basal Stem Rot (BSR) Disease. *Sci. Rep.* **2020**, *10*, 1–16. [[CrossRef](#)]
22. Golhani, K.; Balasundram, S.K.; Vadimalai, G.; Pradhan, B. A review of neural networks in plant disease detection using hyperspectral data. *Inf. Process. Agric.* **2018**, *5*, 354–371. [[CrossRef](#)]
23. Liakos, K.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine learning in agriculture: A review. *Sensors* **2018**, *18*, 2674. [[CrossRef](#)]
24. Asraf, H.M.; Nooritawati, M.T.; Rizam, M.S. A comparative study in kernel-based support vector machine of oil palm leaves nutrient disease. *Procedia Eng.* **2012**, *41*, 1353–1359. [[CrossRef](#)]
25. Nababan, M.; Laia, Y.; Sitanggang, D.; Sihombing, O.; Indra, E.; Siregar, S.; Purba, W.; Mancur, R. The diagnose of oil palm disease using Naive Bayes Method based on Expert System Technology. *J. Phys. Conf. Ser.* **2018**, *1007*, 012015. [[CrossRef](#)]

26. Bejo, S.; Abdol-Lajis, G.; Abd-Aziz, S.; Abu-Seman, I.; Ahamed, T. Detecting Basal Stem Rot (BSR) disease at oil palm tree using thermal imaging technique. In Proceedings of the 14th International Conference on Precision Agriculture, Montreal, QC, Canada, 24–27 June 2018; International Society of Precision Agriculture: Monticello, IL, USA, 2018.
27. Santoso, H.; Tani, H.; Wang, X. Random Forest classification model of basal stem rot disease caused by Ganoderma boninense in oil palm plantations. *Int. J. Remote Sens.* **2017**, *38*, 4683–4699. [[CrossRef](#)]
28. Santoso, H.; Tani, H.; Wang, X.; Prasetyo, A.E.; Sonobe, R. Classifying the severity of basal stem rot disease in oil palm plantations using WorldView-3 imagery and machine learning algorithms. *Int. J. Remote Sens.* **2019**, *40*, 7624–7646. [[CrossRef](#)]
29. Ahmadi, P.; Muhamram, F.M.; Ahmad, K.; Mansor, S.; Abu Seman, I. Early detection of Ganoderma basal stem rot of oil palms using artificial neural network spectral analysis. *Plant Dis.* **2017**, *101*, 1009–1016. [[CrossRef](#)]
30. Liaghat, S.; Ehsani, R.; Mansor, S.; Shafri, H.Z.; Meon, S.; Sankaran, S.; Azam, S.H. Early detection of basal stem rot disease (Ganoderma) in oil palms based on hyperspectral reflectance data using pattern recognition algorithms. *Int. J. Remote Sens.* **2014**, *35*, 3427–3439. [[CrossRef](#)]
31. Khaled, A.Y.; Aziz, S.A.; Bejo, S.K.; Nawi, N.M.; Seman, I.A. Spectral features selection and classification of oil palm leaves infected by Basal stem rot (BSR) disease using dielectric spectroscopy. *Comput. Electron. Agric.* **2018**, *144*, 297–309. [[CrossRef](#)]
32. Abdullah, A.H.; Adom, A.H.; Shakaff, A.Y.M.; Ahmad, M.N.; Zakaria, A.; Saad, F.S.A.; Isa, C.M.N.C.; Masnan, M.J.; Kamarudin, L.M. Hand-held electronic nose sensor selection system for basal stamp rot (BSR) disease detection. In Proceedings of the 2012 Third International Conference on Intelligent Systems Modelling and Simulation, Kota Kinabalu, Malaysia, 8–10 February 2012; IEEE: New York, NY, USA, 2012; pp. 737–742.
33. Hashim, C.I.; Rashid, M.S.A.; Bejo, S.K.; Muhamram, F.M.; Ahmad, K. Severity of Ganoderma boninense disease classification using SAR data. In Proceedings of the 39th Asian Conference on Remote Sensing (ACRS 2018), Kuala Lumpur, Malaysia, 15–19 October 2018; pp. 2492–2499.
34. Oliva, J.; Stenlid, J.; Martínez-Vilalta, J. The effect of fungal pathogens on the water and carbon economy of trees: Implications for drought-induced mortality. *New Phytol.* **2014**, *203*, 1028–1035. [[CrossRef](#)]
35. Graham, M.; Davies, A. *3D Point Cloud Tree Modelling*; Intelligence Surveillance and Reconnaissance Division DSTO Defence Science and Technology Organisation: Edinburgh, Australia, 2010; 18p, Available online: <https://apps.dtic.mil/dtic/tr/fulltext/u2/a526083.pdf> (accessed on 26 February 2018).
36. Wu, D.; Phinn, S.; Johansen, K.; Robson, A.; Muir, J.; Searle, C. Estimating changes in leaf area, leaf area density, and vertical leaf area profile for mango, avocado, and macadamia tree crowns using terrestrial laser scanning. *Remote Sens.* **2018**, *10*, 1750. [[CrossRef](#)]
37. Khairunniza-Bejo, S.; Vong, C.N. Detection of basal stem rot (BSR) infected oil palm tree using laser scanning data. *Agric. Agric. Sci. Procedia* **2014**, *2*, 156–164. [[CrossRef](#)]
38. Wilkes, P.; Lau, A.; Disney, M.; Calders, K.; Burt, A.; de Tanago, J.G.; Bartholomeus, H.; Brede, B.; Herold, M. Data acquisition considerations for terrestrial laser scanning of forest plots. *Remote Sens. Environ.* **2017**, *196*, 140–153. [[CrossRef](#)]
39. Corchado, E.S.; Yin, H. *Intelligent Data Engineering and Automated Learning-IDEAL*; Springer: Berlin/Heidelberg, Germany, 2009.
40. Ayrey, E.; Fraver, S.; Kershaw Jr, J.A.; Kenefic, L.S.; Hayes, D.; Weiskittel, A.R.; Roth, B.E. Layer stacking: A novel algorithm for individual forest tree segmentation from LiDAR point clouds. *Can. J. Remote Sens.* **2017**, *43*, 16–27. [[CrossRef](#)]
41. Tang, S.; Dong, P.; Buckles, B.P. Three-dimensional surface reconstruction of tree canopy from lidar point clouds using a region-based level set method. *Int. J. Remote Sens.* **2013**, *34*, 1373–1385. [[CrossRef](#)]
42. Yang, B.; Dai, W.; Dong, Z.; Liu, Y. Automatic forest mapping at individual tree levels from terrestrial laser scanning point clouds with a hierarchical minimum cut method. *Remote Sens.* **2016**, *8*, 372. [[CrossRef](#)]
43. Bienert, A.; Maas, H.G.; Scheller, S. Analysis of the information content of terrestrial laser scanner point clouds for the automatic determination of forest inventory parameters. In Proceedings of the Workshop on 3D Remote Sensing in Forestry, Vienna, Austria, 14–15 February 2006; pp. 1–7.
44. Lovell, J.L.; Jupp, D.L.B.; Newnham, G.J.; Culvenor, D.S. Measuring tree stem diameters using intensity profiles from ground-based scanning lidar from a fixed viewpoint. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 46–55. [[CrossRef](#)]

45. Hongzhi, W.; Ying, D. An improved image segmentation algorithm based on Otsu method. International symposium on photoelectronic detection and Imaging. *Proc. SPIE* **2008**, *66*, 196–202. [[CrossRef](#)]
46. Shaukat, S.S.; Rao, T.A.; Khan, M.A. Impact of sample size on principal component analysis ordination of an environmental data set: Effects on eigenstructure. *Ekológia* **2016**, *35*, 173–190. [[CrossRef](#)]
47. Adler, N.; Yazhemsky, E. Improving discrimination in data envelopment analysis: PCA–DEA or variable reduction. *Eur. J. Oper. Res.* **2010**, *202*, 273–284. [[CrossRef](#)]
48. Subasi, A. *Practical Guide for Biomedical Signals Analysis Using Machine Learning Techniques: A MATLAB Based Approach*; Academic Press: Cambridge, MA, USA, 2019.
49. Bonate, P.L.; Steimer, J.L. *Pharmacokinetic-Pharmacodynamic Modeling and Simulation*; Springer: New York, NY, USA, 2011; pp. 157–186.
50. Cattell, R.B.; Jaspers, J. A general plasmode (No. 30-10-5-2) for factor analytic exercises and research. *Multivar. Behav. Res. Monogr.* **1967**, *67*, 212.
51. Bartlett, M.S. A note on the multiplying factors for various  $\chi^2$  approximations. *J. R. Stat. Soc. Ser. B (Methodol.)* **1954**, *296–298*. [[CrossRef](#)]
52. Shao, Y.; Zhao, C.; Bao, Y.; He, Y. Quantification of nitrogen status in rice by least squares support vector machines and reflectance spectroscopy. *Food Bioprocess Technol.* **2012**, *5*, 100–107. [[CrossRef](#)]
53. Yang, W.; Sigrimis, N.; Li, M.; Sun, H.; Zheng, L. Correlations between nitrogen content and multispectral image of greenhouse cucumber grown in different nitrogen level. In *International Conference on Computer and Computing Technologies in Agriculture*; Springer: Berlin/Heidelberg, Germany, 2012; Volume 393, pp. 456–463. [[CrossRef](#)]
54. Solomonoff, R.J. Machine learning-past and future. In Proceedings of the Dartmouth Artificial Intelligent Conference, Dartmouth, NH, USA, 13–15 July 2006.
55. Ciaburro, G. *Matlab for Machine Learning*; Packt Publishing Ltd.: Birmingham, UK, 2017.
56. Congalton, R.G. Putting the map back in map accuracy assessment. In *Remote Sensing and GIS Accuracy Assessment*; Lunetta, R.S., Lyon, J.G., Eds.; CRC Press, Inc.: Boca Raton, FL, USA, 2004; pp. 1–12.
57. McHugh, M.L. Interrater reliability: The kappa statistic. *Biochem. Med. Biochem. Med.* **2012**, *22*, 276–282. [[CrossRef](#)]
58. Hay, A.M. The derivation of global estimates from a confusion matrix. *Int. J. Remote Sens.* **1998**, *9*, 1395–1398. [[CrossRef](#)]
59. David, C.C.; Jacobs, D.J. Principal component analysis: A method for determining the essential dynamics of proteins. In *Protein Dynamics*; Humana Press: Totowa, NJ, USA, 2014; pp. 193–226. [[CrossRef](#)]
60. Vichi, M.; Monari, P.; Mignani, S.; Montanari, A. (Eds.) New Developments in Classification and Data Analysis. In Proceedings of the Meeting of the Classification and Data Analysis Group (CLADAG) of the Italian Statistical Society, Bologna, Italy, 22–24 September 2003; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2006.
61. Sangulagi, P.; Sutagundar, A.V.; Stelvarani, S. Storage of Mobile Sensor Data in Clouds using Information Classification Algorithms. *Int. J. Adv. Netw. Appl.* **2018**, *10*, 3893–3897. [[CrossRef](#)]
62. Sebastiani, P.; Gussoni, E.; Kohane, I.S.; Ramoni, M.F. Statistical challenges in functional genomics. *Stat. Sci.* **2003**, *18*, 33–60. [[CrossRef](#)]
63. Friedman, N.; Geiger, D.; Goldszmidt, M. Bayesian network classifiers. *Mach. Learn.* **1997**, *29*, 131–163. [[CrossRef](#)]
64. Domingos, P.; Pazzani, M. On the optimality of the simple Bayesian classifier under zero-one loss. *Mach. Learn.* **1997**, *29*, 103–130. [[CrossRef](#)]
65. Xue, Y.; Chen, H.; Jin, C.; Sun, Z.; Yao, X. NBA-Palm: Prediction of palmitoylation site implemented in Naive Bayes algorithm. *BMC Bioinform.* **2006**, *7*, 458.
66. Wu, X.; Kumar, V.; Quinlan, J.R.; Ghosh, J.; Yang, Q.; Motoda, H.; McLachlan, G.J.; Ng, A.; Liu, B.; Philip, S.Y.; et al. Top 10 algorithms in data mining. *Knowl. Inf. Syst.* **2008**, *14*, 1–37. [[CrossRef](#)]
67. Zhang, Y.; Chu, C.H.; Chen, Y.; Zha, H.; Ji, X. Splice site prediction using support vector machines with a Bayes kernel. *Expert Syst. Appl.* **2006**, *30*, 73–81. [[CrossRef](#)]
68. Abdelwahab, O.; Bahgat, M.; Lowrance, C.J.; Elmaghhraby, A. Effect of training set size on SVM and Naive Bayes for Twitter sentiment analysis. In Proceedings of the 2015 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), Abu Dhabi, UAE, 7–10 December 2015; IEEE: New York, NY, USA, 2015; pp. 46–51.

69. Hashim, U.K.M.; Ahmad, A. The effects of training set size on the accuracy of maximum likelihood, neural network and support vector machine classification. *Sci. Int. -Lahore* **2014**, *26*, 1477–1481.
70. Afdal, M.; Purnianda, S. Expert System of Palm Oil Plant Diagnosis Using Bayesian Network Method. *Sci. J. Inf. Syst. Eng. Manag.* **2019**, *5*, 218–223.
71. Sidauruk, A.; Pujianto, A. Expert System for Diagnosis of Palm Oil Diseases Using Bayes Theorem. *Data Manag. Inf. Technol.* **2017**, *18*, 51–56.
72. Rasekhshaffee, K.C.; Jones, R.C. Machine Learning for Stock Selection. *Financ. Anal. J.* **2019**, *75*, 70–88. [[CrossRef](#)]
73. Schölkopf, B.; Smola, A.J.; Bach, F. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*; MIT Press: Cambridge, MA, USA, 2002.
74. Li, Y.; Anderson-Sprecher, R. Facies identification from well logs: A comparison of discriminant analysis and naïve Bayes classifier. *J. Pet. Sci. Eng.* **2006**, *53*, 149–157. [[CrossRef](#)]

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).