

Oil Palm Leaves Phenotyping using Biomarkers Derived from Raman Spectra

Mohd Syazwan Herman^a, Fazida Hanim Hashim^{a,b*}, Zuhaira Mohd Zulkifli^a,
Aqilah Baseri Huddin^a, Ghassan Maan Salim^a

^aDepartment of Electrical, Electronic and Systems, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, Bangi 43600, Selangor, Malaysia;

^bResearch Centre for Sustainable Process Technology (CESPRO), Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, Bangi 43600, Selangor, Malaysia

Abstract The efficient production of oil from oil palm trees is heavily dependent on their health status, reflected in the oil extraction rate (OER). The 17th frond of the oil palm trees contains a significant amount of organic compounds that directly influence the overall health of the tree. Achieving an optimal balance of essential nutrients such as nitrogen (N), phosphorus (P), and potassium (K) is crucial for classifying a tree as healthy, as it results in an increased oil to bunch and fruit to bunch ratio. To accurately assess the health level of oil palm trees, this study explores the application of Raman spectroscopy, a non-invasive technique in determining the molecular fingerprint of an organic sample. In this research, Raman spectroscopy is employed to determine the health level of oil palm trees, and a machine learning-based health level classification algorithm is developed. The algorithm analyzes the organic compounds found in oil palm leaves, which were collected from 20 different trees. The extracted spectral features from these leaves are used to classify them into two health levels: healthy and not healthy. For this purpose, 31 machine learning models are tested to identify the most accurate classifier. The findings reveal that the Tree and fine K-Nearest Neighbors (KNN) classifiers demonstrate the highest overall accuracy of 95% using three significant features from the 1046 cm^{-1} peak, namely the Raman intensity, Full Width at Half Maximum (FWHM), and area under the curve. This result signifies the potential of Raman spectroscopy as a reliable and promising method for non-invasively phenotyping oil palm leaves, enabling precise prediction of the health status of oil palm trees.

Keywords: Health level, Machine learning, Oil palm Leaves, Oil palm trees, Raman spectroscopy.

***For correspondence:**

fazida@ukm.edu.my

Received: 15 August 2023

Accepted: 4 Dec. 2023

©Copyright Herman. This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

Introduction

The oil palm tree, scientifically known as *Elaeis guineensis* Jacq., originated in West Asia and was introduced to Peninsular Malaysia (formerly known as Malaya) in 1870. The commercial cultivation of oil palm commenced in 1917, initiated by Henri Fauconnier, a Frenchman who imported oil palm seeds from Sumatra, Indonesia, and planted them at Tennamaram Estate, Batang Berjuntai, now known as Bestari Jaya. Since then, oil palm cultivation has witnessed rapid growth in Malaysia, with a vast crop area of 5.207 million hectares. Notably, Malaysia and Indonesia jointly contribute to 90% of global oil palm production and act as key exporters, accounting for 85% of the total global exports [1]. In 2020, Malaysia's crude palm oil (CPO) exports reached 17.4 million tons, representing 18.3% of the world's total exports of vegetable oils and fats.

Given the economic significance of the oil palm industry, it becomes imperative to monitor the health status of individual oil palm trees to ensure their survival and positive impact on farmers and the country's

economy as a whole. Optimal nutrient levels, achieved through a well-balanced combination of fertilizers, are crucial for enhancing fresh fruit bunches (FFB) production, resulting in larger, healthier, and more abundant fruit clusters [2]. Essential nutrients for oil palm trees, often referred to as NPK elements (Nitrogen, Phosphate, and Potassium), play a pivotal role in achieving a desirable oil to bunch ratio, increasing the oil extraction rate (OER), and elevating the overall productivity of the palm oil industry [3]. The health condition of oil palm trees directly influences their OER and yield. Healthy trees significantly contribute to higher oil palm tree productivity [4]. According to MPOB [5], a -1.4% decrease in OER resulted in a staggering loss of RM 1.15 billion in market value, considering the price of RM 1449.50 per ton that year. With the ever-increasing global demand for palm oil products, the market price continues to rise, reaching RM 4,045.50 per ton at the end of December 2022. In this context, maintaining optimal health levels of oil palm trees becomes paramount to ensure better fruit production. To achieve this, non-destructive research methods are sought, with a focus on non-destructive phenotyping using Raman spectroscopy. Raman spectroscopy has demonstrated capabilities for early disease detection, identification of nutrient concentrations such as NPK content in leaves, and discrimination of various stress conditions [6].

To maximize palm oil production, maintaining the health of oil palm trees is crucial. Nutrient deficiency in oil palm trees can lead to physical manifestations, such as stunted tree height, reduced foliage, and smaller frond leaf areas. Oil palm tree diseases can be categorized based on their symptoms in roots, basal stems, and leaves [7]. Leaves, as a critical biological system in plants, hold valuable nutrient data, including chlorophyll and NPK content, serving as a yardstick for assessing tree health [8].

Presently, methods for evaluating the health of oil palm trees include leaf analysis (LA), soil analysis (SA), and nutrient balance (NB). However, these methods have drawbacks as they take time to yield results and can only be conducted *in vitro*, not *in vivo*. Moreover, the lack of standardized methods or standard operating procedures (SOP) to assess nutrient concentrations further complicates the process. Physical examination of trees remains a common approach to assessing their health, yet this method proves less effective in detecting diseases like basal stem rot ((caused by *Ganoderma boninense*) in its early stages. Stem rot symptoms take time to manifest, making early diagnosis challenging. Consequently, oil palm trees cannot be saved once symptoms become visible, as the internal spread of the attacking fungus remains unnoticed until it is too late.

In pursuit of a more effective approach to categorizing the health level of oil palm trees, previous researchers have explored algorithms for image processing. These methods involve pre-processing, segmentation, feature extraction, and classification of leaf images [4]. However, this computer vision-based classification of health levels can be prone to errors due to variations in light intensities during image capture [8]. Thus, this study aims to establish a correlation between the health level of oil palm trees and organic compounds based on the Raman spectrum. By analyzing data from oil palm leaves and extracting information from the Raman spectrum, a machine learning-based classification system will be developed to accurately and swiftly monitor the health status of oil palm trees.

Materials and Methods

Flowchart of the Study

The study was conducted in four distinct phases, as depicted in Figure 1. In the initial phase, a batch of oil palm leaf samples (5 in total) was collected from the National University of Malaysia (UKM) oil palm plantation. Subsequently, the collected samples underwent Raman spectrometer scanning at i-CRIM UKM to obtain Raman scattering data.

The second phase involved the comprehensive assessment of N, P, and K content in the oil palm leaves, achieved through wet lab chemical analysis. Following this, the third phase consisted of collecting an additional batch of 5 oil palm leaf samples, totaling the samples to 10.

In the final phase, the acquired Raman spectra underwent meticulous analysis, including pre-processing to eliminate noise and baseline correction. Subsequently, relevant features were extracted from the Raman spectra to enable the accurate classification of oil palm leaves into healthy or under stress categories.

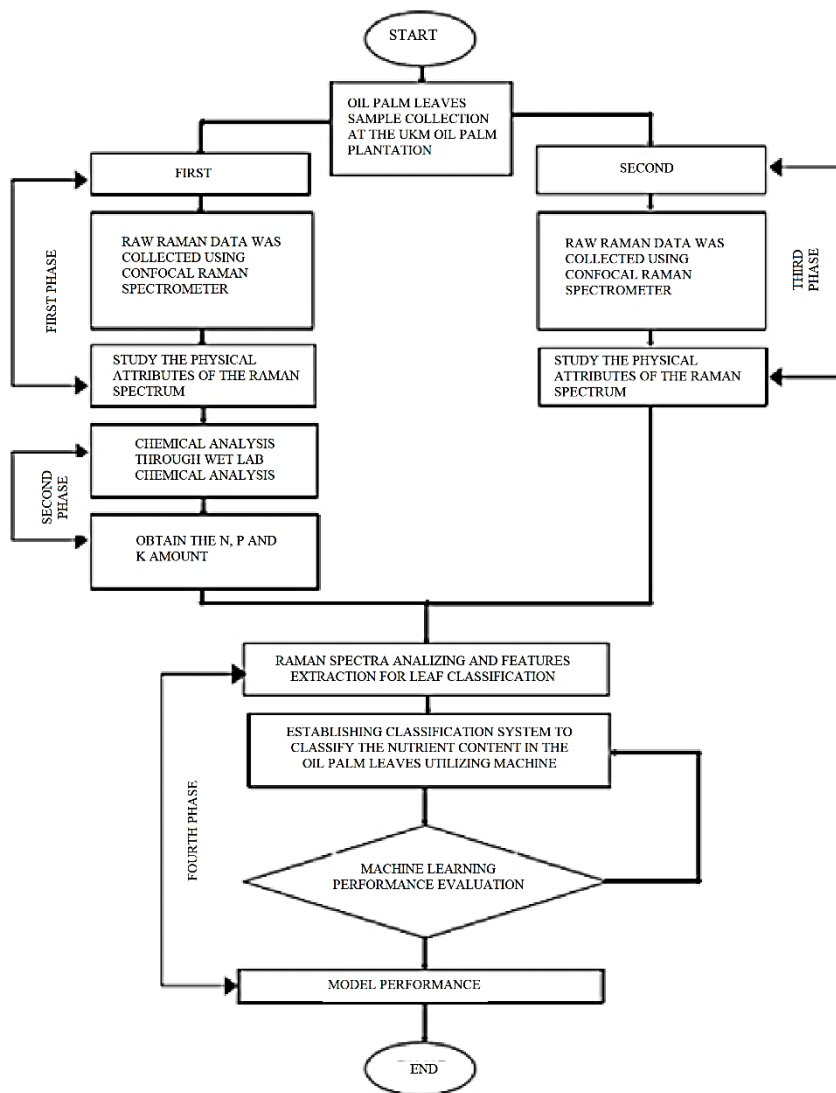


Figure. 1 Flowchart of the study

Oil Palm Leaves Samples Preparation

Samples were collected from the oil palm plantation owned and maintained by JANA@UKM, the commercial branch of UKM. The samples were taken from the *Elaeis guineensis* DxP species. Two batches of samples were collected around 10 days apart. Each batch consisted of 5 samples, making a total of 10 samples. These samples were obtained from the 17th frond of the oil palm trees. For each frond, 50 grams of oil palm leaves were carefully collected, labeled, and transported to the laboratory for subsequent NPK content analysis and Raman scattering measurements. A leaf sample from each frond was examined from both its top and bottom sides. Raman readings were taken using a Raman microscope, with two points measured on each side, resulting in a total of 4 readings per leaf sample as can be observed in Figure 2.

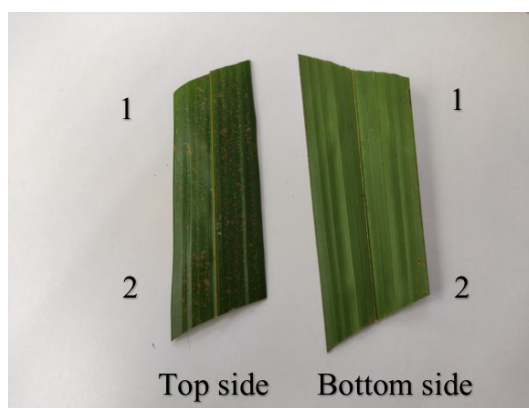


Figure 2. Top and bottom side of the leaf with 2 different sampling points

Raman Instrumentation

A Confocal Micro-Raman spectrometer (Thermo Scientific, DXR2xi Raman Imaging Microscope, Waltham, MA, USA) was employed to acquire the Raman spectra. The spectroscopy analysis utilized a 532 nm laser, a 50 m slit aperture, a 900 lines/mm grating, and a green filter. The samples underwent irradiation three times for a duration of 3 milliseconds each, using a 2.0 mW laser.

Chemical Analysis

A total of 50 grams of oil palm leaves samples were sent to UNIPEQ-UKM for chemical analysis to determine the N, P, and K content. The Kjeldahl method was used for this study, involving the use of an elemental analyzer, inductively coupled plasma mass spectrometry, and atomic absorption spectroscopy.

Raman Spectra Pre-processing

The pre-processing step of the Raman spectra is aimed to remove any noise or spikes from the raw data. Three filters were applied: a second-order Savitzky-Golay filter with a 21-point filter size was used to eliminate high-frequency noise, baseline correction using the rubber band algorithm was applied to correct zero offsets, and the spectra were segmented, focusing on the 1046 cm^{-1} area. This specific area was chosen based on the findings of a previous study [9], which demonstrated variations in Raman scattering responses corresponding to the N, P, and K content in the leaves.

Subsequently, a deconvolution process was conducted using OriginPro software to extract organic compound characteristics from the Raman spectrum. The Lorentz distribution was utilized for curve fitting during this process. Various features were extracted from each deconvoluted peak, including peak intensity, wavelength, full width at half maximum (FWHM), area, ratio of the second highest peak to the highest peak, and ratio of the third highest peak to the highest peak.

Statistical Analysis

Statistical analysis was conducted to study and analyze the extracted features obtained through curve fitting. Significant features were identified using one-way ANOVA and homogeneity tests in the IBM SPSS Statistics software. Features passing both tests were considered significant and selected for classification analysis, while insignificant features were removed to streamline the machine learning algorithm for developing the classification system.

Classification Analysis

The significant features were utilized as predictors in the classification analysis to establish classification models based on machine learning. MATLAB software was used to train the predictors with 31 machine learning models, including artificial neural network (ANN), k-nearest neighbors (KNN) and support vector machine (SVM) as classifiers. Accuracy results for each classifier were recorded. Data samples were imported from an Excel file into MATLAB, and the healthy and unhealthy samples were assigned as classes '1' and '2', respectively. The class information was placed at the right-end of the data table.

Results and Discussion

Chemical Analysis

Chemical analysis was performed to determine the organic compound content in the oil palm leaves. Based on Equation (1), a health level classification for the oil palm trees was established. The results of the organic compound content in the leaf samples are presented in Table 1, along with the classification of the samples into healthy and not healthy categories. The classification approach is derived from a study by [10], which revealed that at the Raman shift of 1046 cm⁻¹, variations occur due to changes in the concentration of nitrogen (N), while the Raman shift remains relatively unaffected by changes in the concentration of phosphorus (P). Therefore, according to Equation (1), if the value of the class is less than or equal to 6, the sample will be classified as healthy; otherwise, it will be classified as unhealthy. Table 2 illustrates the range of nutrient content in oil palm leaves and the corresponding classification into deficient, low, optimum, and excess. By comparing the nutrient content results from this study with Table 2, which was formulated by [11], it can be observed that the oil palm leaf samples fall under the nutrient-deficient class.

The results indicate that the oil palm trees analyzed in this study are experiencing nutrient deficiency. This finding underscores the importance of implementing appropriate nutrient management strategies to enhance the health of the oil palm trees, as nutrient deficiency can significantly impact the oil extraction rate and overall productivity [3].

In the following sections, we will discuss the outcomes of the Raman spectra pre-processing and the classification analysis based on machine learning models.

$$Class = \frac{P}{N} \times 100\% \tag{1}$$

Table 1. N, P and K content in oil palm leaves samples [11]

Sample	N content (%)	P content (%)	K content (%)	(P/N) x 100%	Class
1	0.75	0.042	0.395	5.574	Healthy
2	0.87	0.049	0.312	5.648	Healthy
3	0.94	0.038	0.364	4.008	Healthy
4	0.70	0.041	0.323	5.923	Healthy
5	0.96	0.043	0.333	4.511	Healthy
6	0.88	0.061	0.285	6.909	Not Healthy
7	0.81	0.055	0.279	6.815	Not Healthy
8	0.66	0.047	0.374	7.045	Not Healthy
9	0.85	0.067	0.315	7.847	Not Healthy
10	0.71	0.061	0.273	8.535	Not Healthy

Table 2. Range of nutrient content in oil palm leaves

Variables	Deficient	Low	Optimum	Excess
N (%)	<1.87	1.87-2.24	2.24-2.97	2.97-3.34
P (%)	<0.05	0.05-0.08	0.08-0.14	0.14-0.17
K (%)	<0.72	0.72-0.78	0.78-0.91	0.91-0.97

Raman Spectra Pre-processing

All the raw Raman spectra from sample 1 to sample 10 underwent signal processing to remove unwanted noise and enhance the quality of the signal. A total of 20 raw Raman spectra were used in this study, obtained from both the top and bottom readings of each leaf sample. Due to limited resources, only 20 samples were able to be collected and examined for this study as the testing cost for each sample is quite costly. More samples will be collected in the future to increase the accuracy of the classification model. Furthermore, due to the limited number of samples, pre-processing of the Raman spectra were performed on each sample. Pre-processing of the Raman spectra was crucial before extracting features through curve fitting, as it significantly increased the accuracy of the results. Figure 3 shows the sample of raw Raman spectrum (a) and the Raman spectrum after applying the Savitzky-Golay filter (b).

After applying the Savitzky-Golay filter, the Raman spectrum still contained offsets and background noise. Therefore, a baseline correction technique was applied to further clean the spectrum, as depicted in Figure 4(a). Subsequently, the spectrum was segmented to focus specifically on the Raman peak at 1046 cm^{-1} , which is a critical area for extracting features related to the concentration of organic compounds. The result after pre-processing was a clear and smooth signal, as shown in Figure 4(b). These pre-processing steps were performed using the Orange data mining software.

Next, the signal processing continued with curve fitting using the OriginPro software to extract all features encompassed in the 1046 cm^{-1} area. The curve fitting process involved the application of the Lorentz distribution to identify three hidden peaks present in that specific area, as illustrated in Figure 5. Among these hidden peaks, the one closest to 1046 cm^{-1} was selected, and all relevant features such as intensity, FWHM, etc., were extracted. A total of 6 features were successfully extracted through this process.

The accurate extraction of these features from the Raman spectra is crucial for subsequent classification analysis and the development of a robust and accurate machine learning-based health level classification system for oil palm trees.

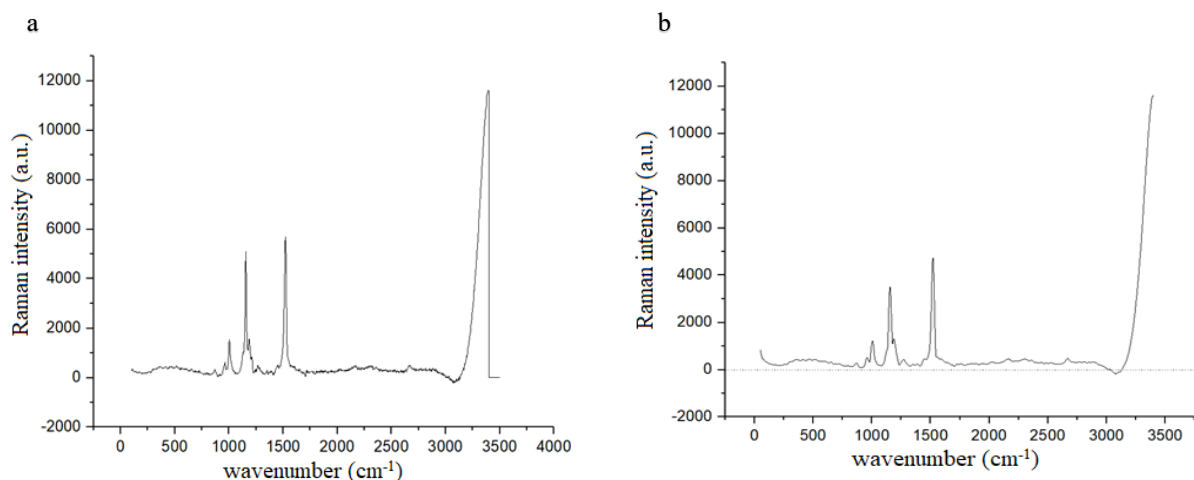


Figure 3. (a) Sample of raw Raman spectrum; (b) Raman spectrum after applying Savitzky-Golay filter

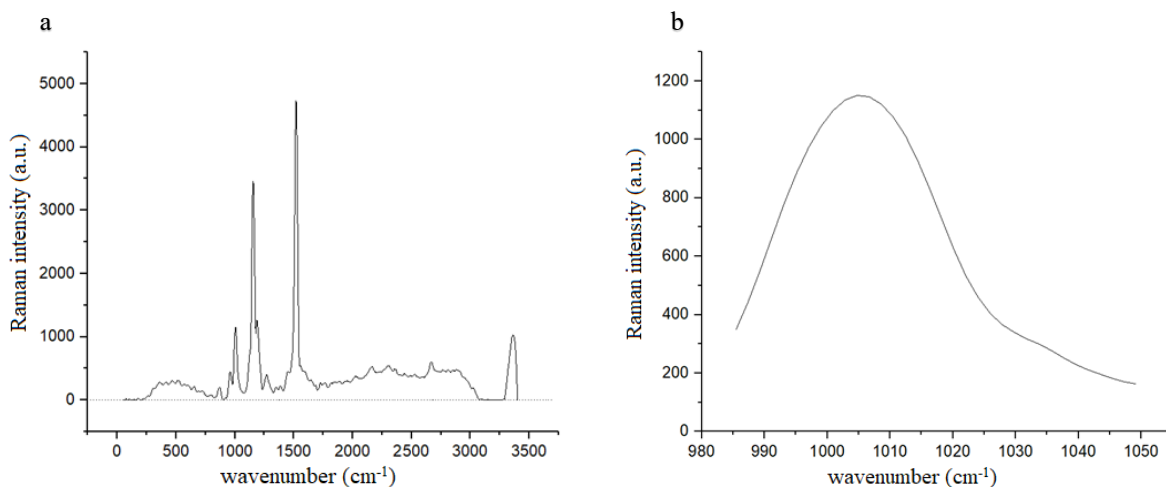


Figure 4. (a) Raman spectrum after applying Baseline Correction; (b) Raman spectrum after segmentation

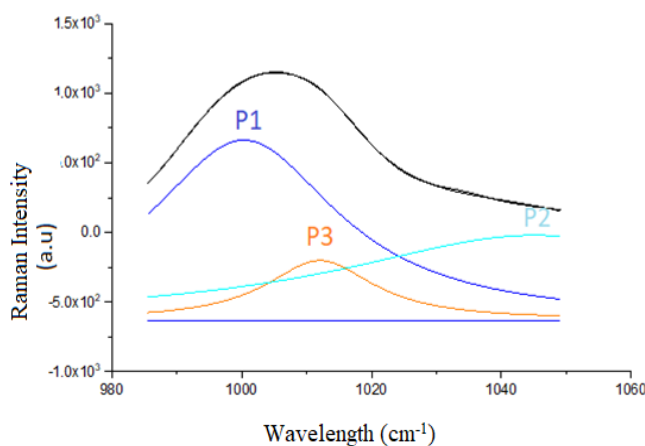


Figure 5. Hidden peaks at 1046 cm^{-1}

Statistical Analysis

The primary objective of the statistical analysis is to identify the significant features among all the extracted features, which will be instrumental in developing the classification system through the machine learning algorithm. Out of the 6 features that were extracted, the statistical analysis revealed that only 3 features are significant for the classification task. These significant features include the peak's intensity, full width at half maximum (FWHM), and peak's area.

The significance of a feature is determined based on the results of the homogeneity test and one-way ANOVA test. A feature is considered significant if it scores a significant value higher than 5% in the homogeneity test and lower than 5% in the one-way ANOVA test. Figure 6 presents the results of the statistical analysis on Raman intensity, FWHM, and area.

For the three significant features, the results of the homogeneity test yielded significant values of 21%, 38%, and 51%, respectively. Additionally, the one-way ANOVA test confirmed the significance of these three features, with significant values of 0.2% and 0.1%.

The identification of these significant features is crucial as they play a vital role in the subsequent classification analysis using machine learning models. These features will serve as key input parameters to develop a robust and accurate classification system for determining the health level of oil palm trees based on Raman spectroscopy data.

Tests of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
intensity	Based on Mean	2.389	1	8	.161
	Based on Median	.655	1	8	.442
	Based on Median and with adjusted df	.655	1	4.594	.458
	Based on trimmed mean	1.848	1	8	.211
fwhm	Based on Mean	1.068	1	8	.332
	Based on Median	.565	1	8	.474
	Based on Median and with adjusted df	.565	1	6.020	.481
	Based on trimmed mean	.862	1	8	.380
area	Based on Mean	.660	1	8	.440
	Based on Median	.119	1	8	.739
	Based on Median and with adjusted df	.119	1	6.111	.741
	Based on trimmed mean	.471	1	8	.512

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
intensity	Between Groups	64000.000	1	64000.000	19.062	.002
	Within Groups	26859.600	8	3357.450		
	Total	90859.600	9			
fwhm	Between Groups	1718.197	1	1718.197	27.143	<.001
	Within Groups	506.407	8	63.301		
	Total	2224.604	9			
area	Between Groups	310.918	1	310.918	38.939	<.001
	Within Groups	63.879	8	7.985		
	Total	374.796	9			

Figure 6. Statistical Analysis results on the Raman intensity, FWHM and area

Classification Analysis

The health level classification system for oil palm leaves was implemented using various machine learning models in MATLAB software with the classification learner toolbox. The sample data imported into MATLAB was trained with different machine learning models, including ANN, CNN, KNN, Tree, and SVM. The accuracy results of these models are presented in Table 3.

Table 3. Machine learning classification models with accuracy

Classification Model	Accuracy (%)
Fine Tree	95.0
Medium Tree	95.0
Coarse Tree	95.0
Fine KNN	95.0
Medium KNN	95.0
Cosine KNN	95.0
Cubic KNN	95.0
Weighted KNN	95.0
Efficient Logistic Regression	95.0
Efficient Linear SVM	90.0

The results indicate that both the Tree and KNN machine learning models achieved a high accuracy of 95%, with only one error in data classification. Further analysis of the classification performance is provided through the confusion matrix and the scattering plot.

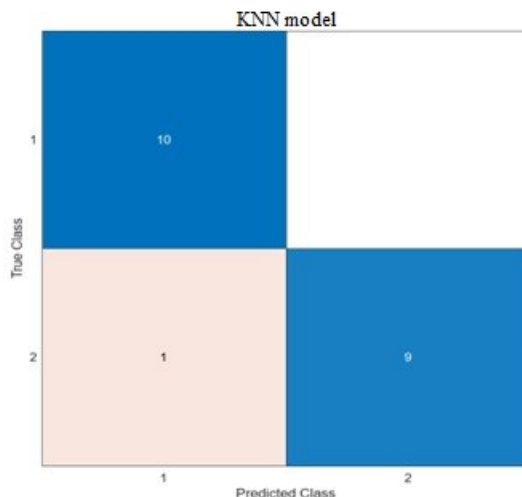


Figure 7 Confusion matrix

The confusion matrix shown in Figure 7 displays the comparison between the predicted class and the true class. Class 1 and Class 2 refer to the healthy and not healthy classes of the samples, respectively. The confusion matrix shows that 9 out of 10 healthy samples were successfully differentiated, while all the unhealthy samples were accurately classified. The scattering plot in Figure 8, shows the distinction between the two groups, with blue dots representing the healthy class and orange dots representing the unhealthy class.

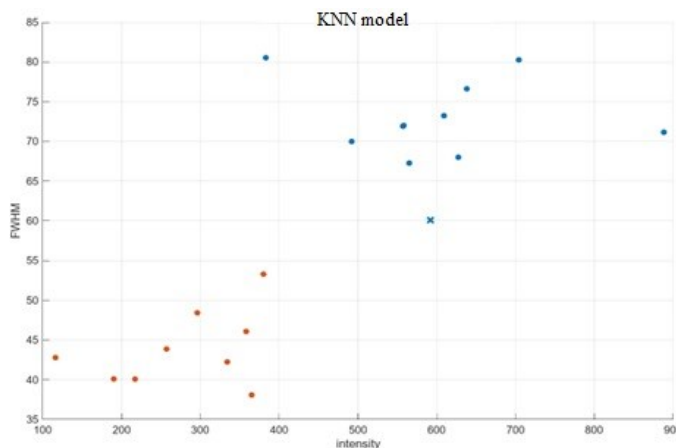


Figure 8 Scattering plot

Table 4. Accuracy from different researches

Researches	Method	Accuracy	Classification Model
[4]	Image Processing	99.67	ANN
[12]	Image Processing	97.00	SVM
[7]	Image Processing	96.00	AlexNet-CNN
This study	Raman Spectra	95.00	Fine KNN

By utilizing 20 study sample data, this study was able to record a very good performance in classifying the health level of oil palm leaves. There are previous studies that also apply a classification system to

classify the health level of oil palm trees. Table 4 shows the performance results carried out by past researchers in applying the classification system of the health level of oil palm trees. The involvement of image processing methods in building a classification system for the health level of oil palm trees is very high and also records a high percentage of accuracy. However, a major drawback of image processing methods is the high dependency on light where the light source must be consistent at all times.

Conclusions

In this study, a classification system for determining the health level of oil palm trees was developed using a machine learning approach and Raman spectroscopy. Traditional methods for classifying the health level of oil palm trees, such as chemical tests, are complex and time-consuming, often requiring invasive procedures to obtain readings of N, P, and K content in oil palm leaves. This study focused specifically on the 17th frond of leaves as the location for taking readings of N, P, and K content.

The developed automatic and non-invasive classification system based on Raman spectroscopy demonstrated high accuracy and provided results in a short period of time. This system can be valuable for farmers in maintaining the health level of their oil palm trees, leading to increased productivity and higher returns. The Raman peak at a wavelength of 1046 cm^{-1} was identified as a biomarker for assessing the organic content of oil palm leaves and classifying the health level of oil palm trees.

The main processes involved in this study included collecting oil palm leaf samples, performing chemical tests to analyse the content of organic compounds (N, P, and K), obtaining Raman spectra, and then applying pre-processing, feature extraction, and analysis of the Raman spectra for classification. The significant characteristics extracted from the Raman peak at 1046 cm^{-1} served as predictors for classifying the health level of oil palm trees into healthy and unhealthy. The Tree and Fine KNN machine learning models achieved the highest accuracy of 95%.

In conclusion, the combination of machine learning and Raman spectroscopy presents a promising and effective approach for accurately assessing the health level of oil palm trees. The developed classification system can contribute to better plant management and increased yields, benefiting both farmers and the palm oil industry. By leveraging the Raman peak at 1046 cm^{-1} as a valuable biomarker, this non-invasive method enables quick and reliable prediction of the health level of oil palm trees.

Conflicts of Interest

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

Acknowledgment

This study was funded by the University Research Grant (GUP-2022-025) and the Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia.

References

- [1] Sabri, M. S. A., Endut, R., Rashidi, C. B. M., Laili, A. R., Aljunid, S., A. & Ali, N. (2019). Analysis of near-infrared (NIR) spectroscopy for chlorophyll prediction in oil palm leaves. *Bulletin of Electrical Engineering and Informatics*, 8(2), 506-513.
- [2] Corley, R. H. V., & Tinker, P. B. (2003). Mineral nutrition of oil palms. In *The oil palm* (4th ed., pp. 327-389). Oxford, U.K.
- [3] Prabowo, N. E., & Foster, H. L. (1998). Variation in oil and kernel extraction rates of oil palms in North Sumatra due to nutritional and climatic factors. *Proc. Int. Oil Palm Conf. Puslit Kelapa Sawit*.
- [4] Hamdani, H., Septiarini, A., Sunyoto, A., Suyanto, S., & Utaminigrum, F. (2021). Detection of oil palm leaf disease based on color histogram and supervised classifier. *Optik*, 245, 167753.
- [5] MPOB. (n.d.). Welcome to Malaysian Palm Oil Board // Malaysian Palm Oil Industry // Washington, DC // 1-202-572-9798.
- [6] Baratto, C., Ambrosio, G., Faglia, G., & Turina, M. (2022). Early detection of esca disease in asymptomatic vines by Raman spectroscopy. *IEEE Sensors Journal*, 22(23), 23286-23292. <https://doi.org/10.1109/jsen.2022.3211616>
- [7] Ong, J. H., Ong, P., & Woon, K. L. (2022). Image-based oil palm leaves disease detection using convolutional neural network. *Journal of Information and Communication Technology*, 21(3), 383-410.
- [8] Hashim, F. H., Mohd Zulkifli, Z., Raj, T. & Huddin, A. B. (2018). A rapid and nondestructive technique in

- determining the ripeness of oil palm fresh fruit bunch (FFB). *Jurnal Kejuruteraan*, 30(1), 93-101.
- [9] Huang, C. H., Singh, D. P., Park, S. H., Chua, N., Ram, R. J. & Park, B. S. (2020). Early diagnosis and management deficiency in plants utilizing raman spectroscopy. *Frontiers in Plant Science*, 11(1), 1-13.
- [10] Gupta, S., Huang, C. H., Singh, G. P., Park, B. S., Chua, N. H. & Ram, R. J. (2020). Portable Raman leaf-clip sensor for rapid detection of plant stress. *Scientific Reports*, 10(1), 1-11.
- [11] Behera, S. K., Rao, B. N., Suresh, K., & Manoja, K. (2015). Soil nutrient status and leaf nutrient norms in oil palm (*elaeis guineensis* Jacq.) plantations grown on Southern Plateau of India. *Proceedings of the National Academy of Sciences, India Section B: Biological Sciences*, 86(3), 691-697. <https://doi.org/10.1007/s40011-015-0508-y>.
- [12] Masazhar, A. N., & Kamal, M. M. (2017). Digital image processing technique for palm oil leaf disease detection using multiclass SVM classifier. *2017 IEEE 4th International Conference on Smart Instrumentation, Measurement and Application (ICSIMA)*.