❒   506

Analysis of Near-infrared (NIR) spectroscopy for chlorophyll prediction in oil palm leaves

Mohd. Shafiq Amirul Sabri[1], R. Endut[2], C. B. M. Rashidi[3], A. R. Laili[4], S. A. Aljunid[5], N. Ali[6]

[1,2,3,5]School of Computer and Communication Engineering (SCCE), Universiti Malaysia Perlis, 02600, Arau, Perlis, Malaysia
[1,2,3,5,6]Centre of Excellence Advance Communication Engineering (ACE) Optics, SCCE, Universiti Malaysia Perlis, 01000, Kangar, Perlis, Malaysia
[4]Photonics R&D, Department of Industrial Electronics Technology (IET), MIMOS Berhad, Technology Park Malaysia, 43300 Kuala Lumpur, Malaysia
[6]School of Microelectronic Engineering (SCCE), Universiti Malaysia Perlis, 02600, Arau, Perlis, Malaysia

Article Info

ABSTRACT

Oil palm nutrient content is investigated with using chlorophyll as a representative factor correlated with NIR spectroscopy spectral absorbance. NIR spectroscopy method of sampling have been tested to overcome time consuming, complex chemical analysis procedure and invasive sampling method in order to identify chlorophyll content in an oil palm tree. Spectral absorbance data from range 900 nm to 1700 nm and chlorophyll data, then tested through five pre-processing methods which is Savitzky-Golay Smoothing (SGS), Multiplicative Scatter Correction (MSC), Single Normal Variation (SNV), First Derivative (1D) and also Second Derivative (2D) using Partial Least Square (PLS) regression prediction model to evaluate the correlation between both data. The overall results show, SGS has the best performance for preprocessing method with the results, the coefficient of determination (R^2) values of 0.9998 and root mean square error (RMSE) values of 0.0639. In summary, correlation of NIR spectral absorbance data and chlorophyll can be achieved using a PLS regression model with SGS pre-processing technique. Thus, we can conclude that NIR spectroscopy method can be used to identify chlorophyll content in oil palm with using time saving, simple sampling and non-invasive method.

Corresponding Author:

Mohd. Shafiq Amirul Sabri,
Centre of Excellence Advance Communication Engineering (ACE) Optics,
School of Computer and Communication Engineering,
Universiti Malaysia Perlis, 01000 Kangar, Perlis, Malaysia.
Email: shafiq.amirul23@gmail.com

1. INTRODUCTION

Oil palm (Elaeis guineensis Jacq) which originate from West Africa [1] has been one of the largest commodity agricultural industry in Malaysia. Economically, Malaysia and Indonesia has contributed 90% of global oil palm production and 85% as global exporter [2]. More importantly, this tree has produced more product than we can imagine not only in the food industry and cosmetics, but also harness as a source of renewable energy, biofuels [3].

In order to optimize the production yield, a healthy palm can be maintained through proper fertilizer application. In establishing oil palm industry, fertilizer application could cost not less than 50% of the plantation maintenance cost and 25% of overall production cost [4]. Due to an increment of fertilizer cost, management must make sure the supplied fertilizer were optimized in balance composition according to their respective deficiency, by the same time increasing the yields [5].

In contrast, oil palm tree shows physical effect in plant appearance affected by the deficiency of nutrient which shown by plant height, leaf number, leaf area of frond, chlorophyll content and also nitrogen and phosphorus content of the leaves [6]. From all the listed part, leaves part are one of the obvious observation to be considered. As part of the plant biological system, leaves contain a number of nutrients that can be observe such as chlorophyll, nitrogen, phosphorus, potassium and etc [7].

As a matter of fact, chlorophyll were corresponded to nutrient content in a plant, representing the health of the tree itself [8-10]. For this reason, an experiment has been conducted to observe the nutrient content of oil palm tree by selecting chlorophyll as the representative for overall nutrient.

In the meantime, established method of sampling for the leaves are time consuming for large scale of the sample, implicate with complex chemical analysis and also involved destructive method of sampling [11-13]. In brief, to overcome this issue, near-infrared (NIR) spectroscopy has been introduced and are currently being used in recent related paper by researchers [11-18].

NIR spectroscopy is ranged in wavelength spectrum from 780 nm to 2500 nm. Spectral analysis of absorbance in NIR spectrometer will be tested and chlorophyll meter will be the calibrator for this experiment. This study will focus on the chlorophyll content of oil palm tree from different frond named F3, F9, and F17. In essence, spectral data will be treated to increase the correlation between both data, which will be determined in the prediction model analysis.

2.   MATERIALS AND METHODS
2.1. Sampling

For leaf sampling, total of 72 leaf reading is taken from 24 oil palm, from different leaf and frond, left and right side, different level, mixing green leaf with some leaf that are yellowish to make sure samples are covering different types of leaf appearance and different deficiency. 5 years old oil palm tree is chosen because of nutrient variation will decrease after age of 6 years reflected by an established canopy [19]. Moreover, this age is crucial because healthy oil palm will start fruiting at the age of 3 years old [20].

For a common oil palm tree, it usually has 40 to 60 fronds [20, 21]. However, for this analysis, three significance fronds identically named frond 3 (F3), frond 9 (F9) and frond 17 (F17) is preferred as the variables. Figure 1 shows how to identify areas of frond based on section level of F3, F9 and F17. Based on previous studies [22], F17 has been selected as the primary reference for analysis because the frond are at the midpoint position of the tree. Alongside, the F3 and F9 are selected as a control in this study as both were regularly being used as a particular leaf analysis for micronutrient representations of oil palms below 3 years old [23].
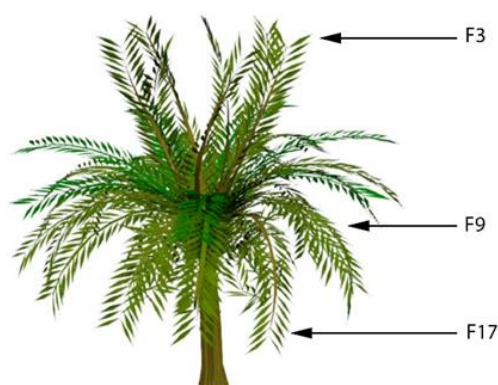

Figure 1. Section level of oil palm frond F3, F9, F17 areas

Besides that, it is also taken from two different compound area to cover different soil properties. All samples were taken from establishing farm in Chuping, Perlis, ensuring that oil palm is previously fruiting, has been periodically fertilized and taken care under plantation management. Because of the weather and environmental condition of the plantation, few pre-processing sampling method has been done as precautionary procedure to avoid data reading error. Every leaf that involves with data reading will be wiped gently with wet tissue containing distilled water to discard contaminant like dust and oil from polluted air and industrial activity [19]. Besides that, the data are taken under low exposure of sunlight to make sure data have no disruption with natural light scattering [24].

2.2. Calibration and data measurement

Minolta SPAD-502 (Minolta, Osaka, Japan) Chlorophyll meter device has been used to identify the Chlorophyll content, giving "greenness" percentage of the leaf based on two wavelengths that are specified from spectral analysis [25]. SPAD-502 need to calibrate before used and data taken by clipping the leaf of oil palm on-site without cutting or destroying the leaf. For the spectral absorbance, data is measured using the DLPNIRscan Nano Evaluation Module (EVM) (Texas Instrument, Texas, United States). NIRscan data are presented in absorbance contrary to reflectance given by the light flashed to the surface of the leaf. Absorbance is detailed in spectrum and measured from the 900 nm to 1700 nm for every data reading [26]. Before starting reading process, NIRScan need to calibrate with Polytetrafluoroethylene (PTFE) compound, which giving the highest reflectance and lowest absorbance in order for the device to give best result with less calibration error [27].

For every leaf, 15 points of reading are taken from different segment of leaf from near-end of the tips to the frond as shown in Figure 2 using both NIRscan and SPAD-502 and then averaged to get the total averaged reading per leaf. NIRscan reading is taken side by side with the SPAD-502 so that the exact same point is taken before being averaged. For the SPAD-502, initial data taken is saving up to 15 reading and average option is ready for the user, but for the NIRscan need to be averaged using spreadsheet software as example Microsoft Excel. So in order to average the data, we need to average by individual absorbance reading from each spectrum row without averaging the overall spectrum. From this overall spectrum plot, we can see the spectral analysis and pattern of the absorbance respond in the oil palm leaf specifically.
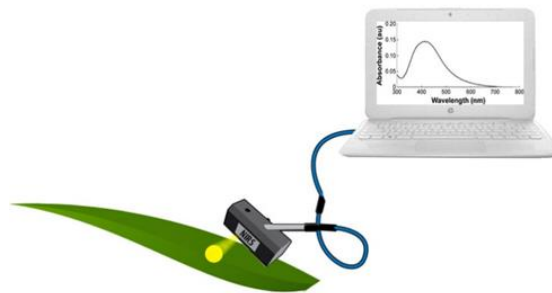

Figure 2. NIRS spectrometer taking the reading of oil palm leaves

2.3. Pre-processing technique

Meanwhile, 50 nm of the first and last data points were removed from original absorbance spectra to avoid data interruption and noise, leaving remaining 950 nm to 1650 nm from overall NIR spectrum wavelength [28]. Besides that, data are also going through the pre-treatment process to remove negative, null and harsh data. From few of the pre-processing techniques tested, about five out of all will be analyzed in this experiment.

Five of them are Savitzky-Golay Smoothing (SGS), Multiplicative Scatter Correction (MSC), Single Normal Variation (SNV), First Derivative (1D) and also Second Derivative (2D). Consequently, all these technique performance is evaluated after testing with partial least squares (PLS) regression prediction model analysis. The comparison between them and the best pre-processing technique will be concluded in the result by evaluating the coefficient of determination (R^2) and root mean square error (RMSE).

2.4. Prediction model

Basically, Partial Least Square (PLS) regression is a method for constructing predictive models when the factors are many and highly collinear [29]. In other words, the model emphasis on predicting the responses and not necessarily on trying to understand the underlying relationship between the variables. The X-and Y-scores are selected with the goal that the connection between progressive sets of scores is as solid as could be expected under the circumstances. Fundamentally, this resembles a robust form of redundancy analysis, looking for direction in the factor space that are related with high variability in the reactions yet biasing them toward a direction that are accurately predicted [29].

PLS regression generalizes and combines features from principle component analysis (PCA) and multiple regression (MR) to predict or analyze a group of dependent variables from a group of variables or predictors. Eventually, prediction is retrieved by extracting it from the predictors set of orthogonal factors

called latent variables (LV) which consisting best predictive power. The orthogonal basis of LV is constructed by PLS regression, aligned along the direction of maximal covariance between the spectra matrix X and the response Y. As for this analysis, maximum number of LV is set at 10 because any number higher than this is resulting non-actual data [19].

For this reason, PLS regression will be preferred as the prediction model to validate the correlation between calibration data of chlorophyll and NIR spectral data. Prior to developing the calibration model, two groups of data consisting 75% for calibration and 25% prediction is randomly distributed from the overall sample data. This ratio is the ideal ratio preferred for the prediction model analysis that are performed in previous studies [30]. Full cross validation is selected to evaluate quality and avoid over-fitting of the graph to achieve the calibration models while prediction set validates the models. PLS regression analysis and pre-processing technique are tested and evaluate using data statistical software named 'The Unscrambler X' (version 10.4, Camo Process AS, Oslo, Norway).

3. RESULT AND DISCUSSION

In Figure 3 shows spectral plot for 72 sample reading from 24 oil palm tree about 5 years old, from overall frond F3, F9 and F17 in NIR spectrum range of 950 nm to 1650 nm. The curve show respond of absorbance varies with wavelength between all of the leaf samples. From this graph, oil palm leaf absorption is intensively high at range 1400 nm to 1600 nm. Henceforth, these absorbance were then analyzed using a prediction model to evaluate the correlation between absorbance and chlorophyll data reading.
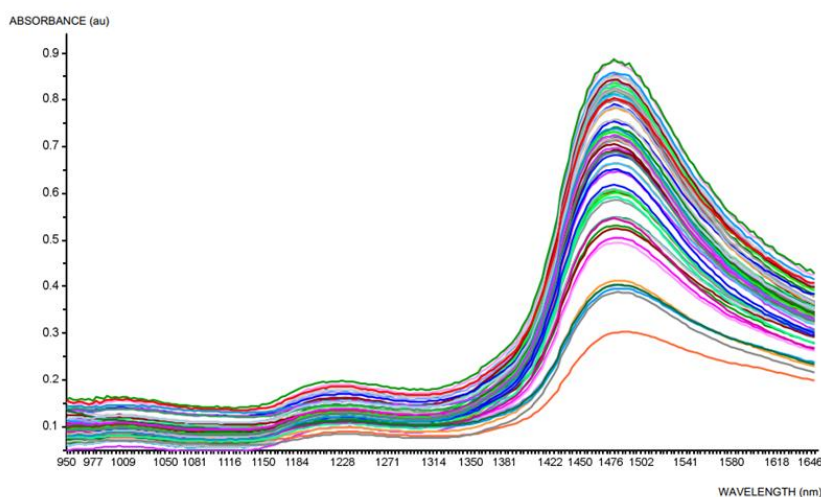

Figure 3. Spectral graph of oil palm leaf absorbance of F3, F9 and F17

As a result, both chlorophyll and spectral absorbance data is then evaluate using PLS regression prediction model with the five selected top performer per-processing technique. There is five pre-processing of interest that is SGS, MSC, SNV, 1D and 2D. In Table 1, the evaluation of chlorophyll data and spectral absorbance data is presented with a different pre-processing technique that have been tested. The data is analyzed in order that overall data included F3, F9 and F17 combined and tested together.

Table 1. PLS regression evaluation with optimum pre-processing technique

Pre-processing technique	Calibration		Prediction	
	R^2	RMSE	R^2	RMSE
Savitzky-Golay Smoothing (SGS)	0.9998	0.0559	0.9998	0.0639
Multiplicative Scatter Correction (MSC)	0.8763	1.8799	0.8028	2.3771
Single Normal Variation (SNV)	0.8757	1.8848	0.8191	2.3310
First Derivative (1D)	0.9359	1.3534	0.8844	1.7844
Second Derivative (2D)	0.8859	1.8059	0.7131	2.8867

The selected absorbance spectral is split into 7 range section that is distributed evenly, and section 6th (1400 nm to 1600 nm) is selected as the observation range to be analyzed because this range show highly anticipated correlation for both data. For this analysis, R^2 and RMSE will be the benchmark of performance for the correlation between chlorophyll data and spectral absorbance data. In spite of R^2, best performance for the correlation to happen is between the ranges of 0.70 to 0.90, or even better if possible, nearer to 1.00 [31]. On the contrary, best RMSE are evaluated in ranges of 0.10 to 0.90 or even better if possible to reach nearer 0.00.

From Table 1, SGS technique shows the most anticipated and intended calibration with R^2 values of 0.9998 and RMSE of 0.0559. More importantly, the prediction result in the best expected range, which are same R^2 as calibration at the values of 0.9998 and RMSE of 0.0639 which only have slightly different from the calibration. This evaluation show that prediction for chlorophyll data is highly correlated to the specified spectral absorbance data using this technique. For graph Figure 4, Figure 5, Figure 6, Figure 7 and Figure 8, blue lines and dot represent calibration data, and red lines and dot represent prediction data. As we can see from Figure 4, the graph is proportionally stack with each other comparing between calibration and prediction graph with slight point on the edge.
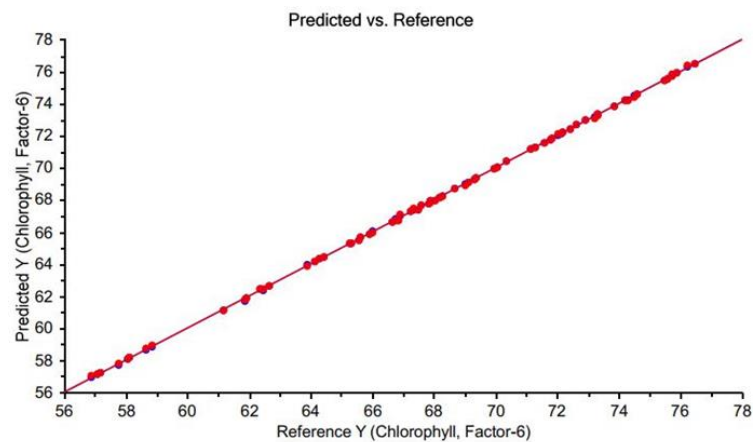

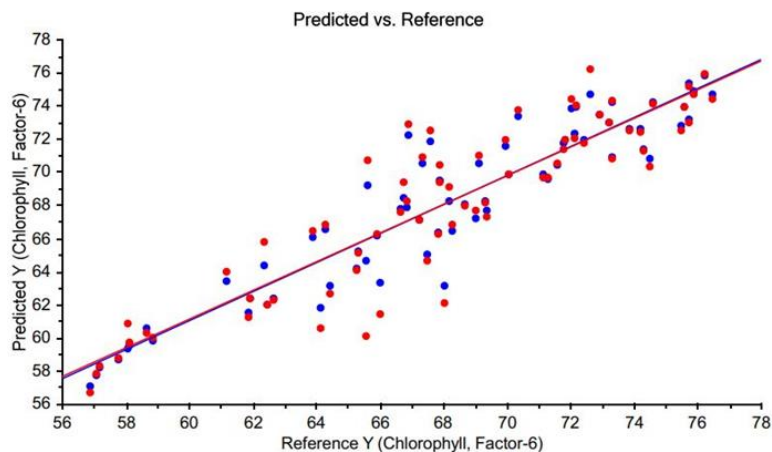Figure 4. PLS regression prediction using SGS technique

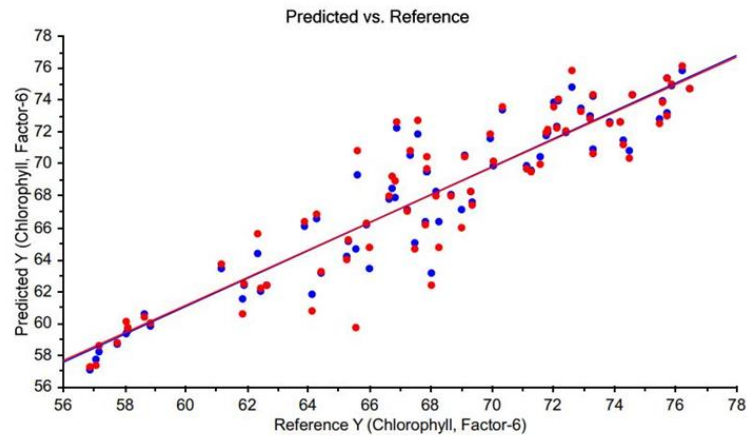Figure 5. PLS regression prediction using MSC technique

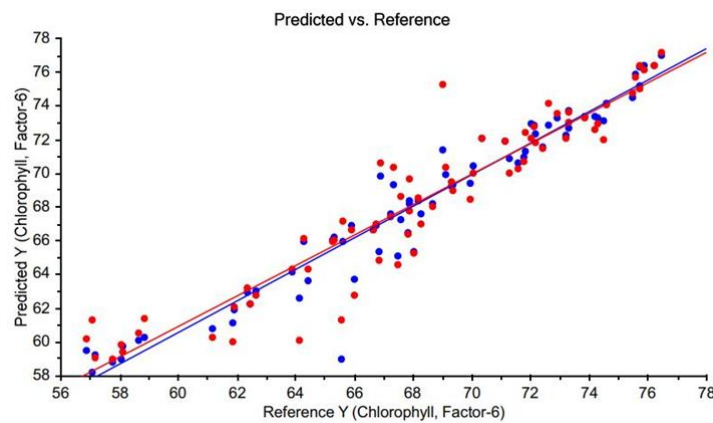Figure 6. PLS regression prediction using SNV technique

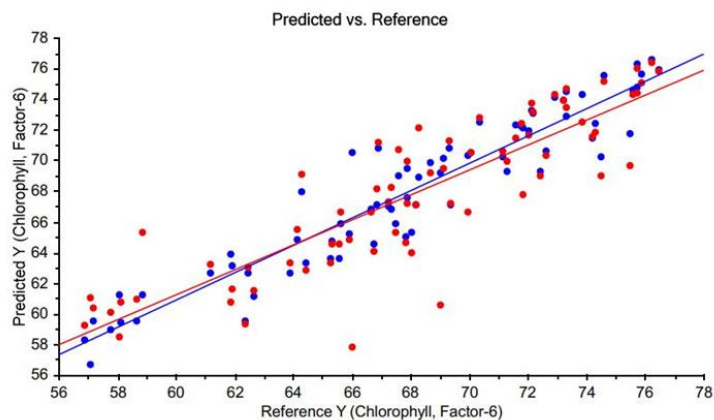Figure 7. PLS regression prediction using 1D technique

Figure 8. PLS regression graph using 2D technique

Comparatively, 1D technique has achieved second highest performer for this analysis with calibration R^2 values of 0.9359 and RMSE at values of 1.3534. Looking at the prediction, the technique has scored 0.8844 for R^2 and 1.7844 for RMSE. Although, the R^2 is quite good for both calibration and prediction, the RMSE is out of the expected benchmark score range which we can say are not balanced. Furthermore, 2D technique has attained the calibration results of 0.8859 for R^2 and 1.8059 for RMSE.

Comparing to the prediction, R^2 has decreased to 0.7131 and RMSE is increasing significantly with the values of 2.8867. For this reason, the technique has scored the highest RMSE for prediction which alternately remark as a low correlation result.

Meanwhile, observing the MSC technique, for calibration has scored 0.8763 for R^2 and 1.8799 for RMSE. Important to realize that the score for prediction is 0.8028 for R^2 and 2.3771 for RMSE. On the other hand, for SNV technique, R^2 has been evaluated at 0.8757 and RMSE at values of 1.8848 for calibration. Correspondingly, SNV has resulted in 0.8191 for R^2 and 2.3310 for RMSE. From this point, we can notice prediction RMSE has quite large for 2D, MSC and SNV technique. Besides, when we observe Figure 5, Figure 6 and Figure 8, the correlation pattern is slightly equivalent to each other. Given this point, we can conclude that the SGS pre-processing technique has scored the best pre-processing technique comparing to other four observed techniques for the correlation of chlorophyll data and spectral absorbance data.

As a suggestion for future research, a lot of other variables must be considered for observation which can also affect the performance of the prediction model such as weather, soil type, temperature and also humidity. Besides that, to increase correlation and accuracy of the prediction model, large number of samples is expected, so that can help prediction model increase its prediction factor.

4.   CONCLUSION

In brief, after going through overall testing and observation, SGS pre-processing has been identified as the best performer for pre-processing technique in the analysis of chlorophyll data correlation with spectral absorbance data of the oil palm leaf. Moreover, we can summarize that using NIR spectral absorbance analysis, chlorophyll content of oil palm tree can be identify more efficiently using time saving, non-invasive, less complex sampling procedure with NIR spectroscopy method of sampling using SGS pre-processing technique and PLS regression model. For future research, numerous of nutrients can be proposed for observation using NIR spectroscopy instrument as an efficient tool.

REFERENCES

[1]   D. Boucher, P. Elias, K. Lininger, M.-T. Calen, S. Roquemore, and E. Saxon, "The root of the problem what's driving tropical deforestation today?," *Union Concerned Sci.*, vol. 61, no. 1, pp. 5-8-28-110, 2011.
[2]   Dr. Ahmad Kushairi Din, "Malaysian Oil Palm Industry Performance 2016 and Prospects for 2017," Malaysia, 2017.
[3]   S. Sumathi, S. P. Chai, and A. R. Mohamed, "Utilization of oil palm as a source of renewable energy in Malaysia," *Renew. Sustain. Energy Rev.*, vol. 12, no. 9, pp. 2404–2421, 2008.
[4]   S. Sudradjat, H. Saputra, and S. Yahya, "Optimization of NPK Compound Fertilizer Package Rate on One Year Old Oil Palm (Elaeis guineensis , Jacq) Trees," *Ijsbar*, vol. 20, no. 1, pp. 365–372, 2015.
[5]   C. B. Teo et al., "Optimising Return From Fertilizer for Oil Palms:," no. 212, pp. 7–10, 1998.
[6]   I. Journal, S. Samsuri, and U. Teknologi, "Effect Of Different Fertilizers On Oil Palm (Elaeis Guineensis) Growth And Performance At Nurs ....," no. March, pp. 0–5, 2017.
[7]   A. Laboratories, "Plant Nutrient Analysis Sampling Guide," vol. 58267, no. 701, pp. 1–8, 2013.
[8]   G. Argenta, P. R. F. da Silva, and L. Sangoi, "Leaf relative chlorophyll content as an indicator parameter to predict nitrogen fertilization in maize," *Ciência Rural,* vol. 34, no. 5, pp. 1379–1387, 2004.
[9]   B. Bojović and A. Marković, "Correlation between nitrogen and chlorophyll content in wheat (Triticum aestivum L.)," *Kragujev. J. Sci.*, vol. 31, no. 5827, pp. 69–74, 2009.
[10]   S. Toan and P. Phu, "Sustainable Living with Environmental Risks," 2014.
[11]   D. Pietro Cavallo, M. Cefola, B. Pace, A. F. Logrieco, and G. Attolico, "Contactless and non-destructive chlorophyll content prediction by random forest regression: A case study on fresh-cut rocket leaves," Comput. Electron. *Agric.*, vol. 140, pp. 303–310, 2017.
[12]   I. R. Lawler, L. Aragones, N. Berding, H. Marsh, and W. Foley, "Near-infrared reflectance spectroscopy is a rapid, cost-effective predictor of seagrass nutrients," *J. Chem. Ecol.*, vol. 32, no. 6, pp. 1353–1365, 2006.
[13]   A. Gitelson, Y. Gritz, and M. Merzlyak, "Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves.," *J. Plant Physiol.*, vol. 160, no. 3, pp. 271–282, 2003.
[14]   N. Shetty, Asmund Rinnan, and R. Gislum, "Selection of representative calibration sample sets for near-infrared reflectance spectroscopy to predict nitrogen concentration in grasses," *Chemom. Intell. Lab. Syst.,* vol. 111, no. 1, pp. 59–65, 2012.

[15] M. Wachendorf, B. Ingwersen, and F. Taube, "Prediction of the clover content of red clover- and white clover-grass mixtures by near-infrared reflectance spectroscopy," *Grass Forage Sci*., vol. 54, no. 1, pp. 87–90, 1999.

[16] C. Petisco, B. García-Criado, B. R. V. De Aldana, I. Zabalgogeazcoa, S. Mediavilla, and A. García-Ciudad, "Use of near-infrared reflectance spectroscopy in predicting nitrogen, phosphorus and calcium contents in heterogeneous woody plant species," *Anal. Bioanal. Chem*., vol. 382, no. 2, pp. 458–465, 2005.

[17] K. Almendingen, H. M. Meltzer, J. I. Pedersen, B. N. Nilsen, and M. Ellekjær, "Near infrared spectroscopy-A potentially useful method for rapid determination of fat and protein content in homogenized diets," *Eur. J. Clin. Nutr*., vol. 54, no. 1, pp. 20–23, 2000.

[18] A. Fassio, E. G. Fernández, E. A. Restaino, A. La Manna, and D. Cozzolino, "Predicting the nutritive value of high moisture grain corn by near infrared reflectance spectroscopy," *Comput. Electron. Agric.,* vol. 67, no. 1–2, pp. 59-63, 2009.

[19] H. Anusia et al., "Application of Spectroscopy for Nutrient Prediction of Oil Palm," *Journal of Experimental Agriculture International*. vol. 15, no. 3, pp. 1–9, 2017.

[20] B. S. Wanasuria, H. Setyobudi, I. B. Mayun, and B. Suprihatno, "Iron Deficiency of Oil Palm in Sumatra," *Better Crops International,* vol. 13, no. 1, pp. 33–35, 1999.

[21] P. Yew Chin, J. A. Varley, and J. B. Ward, Foliar Composition of the Oil Palm in West Malaysia I. Variation in Leaf Nutrient Levels in Relation to Sampling Intensity, 1970; 6(2): 113-121.

[22] E. Pushparajah, Leaf Analysis and Soil Testing for Plantation Tree Crops. ASPAC & FFTC, 1994.

[23] "Oil Palm: Nutrient Disorders and Nutrient Management-Pocket Guide," vol. 89, no. 2, p. 287599, 2005.

[24] Metrohm, "NIR Spectroscopy : A guide to near-infrared spectroscopic analysis of industrial manufacturing processes," *Monograph*, pp. 1–46, 2013.

[25] J. Markwell, J. C. Osterman, and J. L. Mitchell, "Calibration of the Minolta SPAD-502 leaf chlorophyll meter," *Photosynth. Res.,* vol. 46, no. 3, pp. 467–472, 1995.

[26] A. Thangappan and V. Pathangay, "Non-Invasive Blood Glucose Estimation from Near Infrared Spectrum using Gradient Boosted Tree Models," Engineering in Medicine and Biology Society (EMBC), Annual International Conference of the IEEE, At Orlando, Florida, USA, 2016.

[27] M. R. C. Inácio, K. M. G. de Lima, V. G. Lopes, J. D. C. Pessoa, and G. H. de Almeida Teixeira, "Total anthocyanin content determination in intact açaí (Euterpe oleracea Mart.) and palmitero-juçara (Euterpe edulis Mart.) fruit using near infrared spectroscopy (NIR) and multivariate calibration," *Food Chem*., vol. 136, no. 3, pp. 1160–1164, 2013.

[28] A. Hervé, "Partial least squares regression and projection on latent structure regression (PLS Regression)," *Wiley Interdiscip. Rev. Comput. Stat*., vol. 2, no. 1, pp. 97–106, Jan. 2010.

[29] R. D. Tobias, "An introduction to partial least squares regression," *SAS Conf. Proc. SAS Users Gr. Int. 20 (SUGI 20)*, pp. 2–5, 1995.

[30] N. Mat Nawi, G. Chen, T. Jensen, and S. A. Mehdizadeh, "Prediction and classification of sugar content of sugarcane based on skin scanning using visible and shortwave near infrared," *Biosyst. Eng*., vol. 115, no. 2, pp. 154–161, 2013.

[31] Konstantinos G. Kyprianidis and Jan Skvaril, Developments in Near-Infrared Spectroscopy. Croatia: Janeza Trdine 9, 51000 Rijeka, Croatia, 2017.