JTEP Jurnal Keteknikan Pertanian,
Vol. 14 No. 1, p 108-125, 2026
P-ISSN 2407-0475 E-ISSN 2338-8439

Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.014.1.108-125

Research Article

OPEN ACCESS Check for updates

Rapid NIR-Based Prediction of Free Fatty Acid and Moisture Content in Intact Oil Palm Fruits Using PLSR and Hybrid PLS-ANN Models

Annisy Syahida Aulia[1], I Wayan Budiastra[2,3*], Y Aris Purwanto[2], Yunisa Trisuci[4], Agus Arip Munawar[5], Daniel Mörlein[6]

[1]Postharvest Study Program, Faculty of Engineering and Technology, IPB University, Jalan Lingkar Akademik, Kampus IPB Dramaga, Babakan, Kec. Dramaga, Kabupaten Bogor, Jawa Barat 16002, Indonesia.
[2]Division of Biosystem Engineering, Faculty of Engineering and Technology, IPB University, Jalan Lingkar Akademik, Kampus IPB Dramaga, Babakan, Kec. Dramaga, Kabupaten Bogor, Jawa Barat 16002, Indonesia.
[3]Center for Research on Engineering Application in Tropical Agriculture (CREATA) IPB University, Jalan Lingkar Akademik, Kampus IPB Dramaga, Babakan, Kec. Dramaga, Kabupaten Bogor, Jawa Barat 16002, Indonesia.
[4]Agricultural Engineering Study Program, Faculty of Engineering and Technology, IPB University, Jalan Lingkar Akademik, Kampus IPB Dramaga, Babakan, Kec. Dramaga, Kabupaten Bogor, Jawa Barat 16002, Indonesia.
[5]Department of Agricultural Engineering, Pusmeptan / PR-ITP / ARC research center, University Syiah Kuala, Banda Aceh, Indonesia.
[6]Department für Nutztierwissenschaften, Georg-August-Universität Göttingen, Germany.
*Corresponding author, email: wbudiastra@apps.ipb.ac.id

Article Info

Submitted: 6 February 2026
Revised: 22 March 2026
Accepted: 30 March 2026
Available online: 16 April 2026
Published: March 2026

Keywords:
NIR; oil palm fruits; PLS-ANN; moisture content; free fatty acid.

How to cite:
Aulia, A. S., Budiastra, I. W., Purwanto, Y. A., Trisuci, Y., Munawar, A.A., Morlein, D. (2026). Rapid NIR-Based Prediction of Free Fatty Acid and Moisture Content in Intact Oil Palm Fruits Using PLSR and Hybrid PLS-ANN Models. Jurnal Keteknikan Pertanian, 14(1): 108-125.
https://doi.org/*10.19028/jtep.014.1.108-125.*

Abstract

Rapid, non-destructive, and in-situ methods are essential for predicting the chemical composition of oil palm fruit to improve the harvesting efficiency. This study aimed to evaluate the performance of Partial Least Squares Regression (PLSR) and Partial Least Squares Regression-Artificial Neural Network (PLS-ANN) models with different spectral pre-treatments for predicting free fatty acid (FFA) and moisture content of oil palm fruit using a portable NIR spectrometer (740–1070 nm). A total of 408 oil palm fruits of the Tenera variety (Elaeis guineensis Jacq. var. tenera), representing 10 maturity stages (3–6 months), were used in this study. The reflectance spectra of the samples were acquired using a portable NIR Spectrometer and then transformed into absorbance spectra. The samples were then subjected to FFA and moisture content analysis using the chemical method. Some spectral pretreatments were applied to the NIR absorbance data before calibration. PLSR and a hybrid method integrating PLS and ANN were used to build calibration models for predicting FFA and moisture content. Performance evaluation revealed that the best model for predicting FFA was achieved using a combination of first derivative Savitzky-Golay and smoothing Savitzky-Golay pretreatments through PLS-ANN calibration (R^2 = 0.81, RPD_val = 2.34, and consistency = 87.88%). For moisture content, the

JTEP Jurnal Keteknikan Pertanian,
Vol. 14 No. 1, p 108-125, 2026
P-ISSN 2407-0475 E-ISSN 2338-8439
Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.014.1.108-125

best model was obtained using detrending pre-treatment through PLS-ANN calibration (R^2 = 1, RPD_val = 12.52, and consistency = 86.47%). These results indicate that the FFA prediction model is suitable for rough screening, whereas the moisture prediction model is suitable for various applications. These models demonstrate a strong potential for practical application at both the farmer and industrial levels.

Doi: https://doi.org/10.19028/jtep.014.1.108-125

1. Introduction

Indonesia is the world's largest producer of oil palm (*Elaeis guineensis* Jacq.) producer. The primary product (CPO) contributes 3.5% to the total gross domestic product (Kementerian Pertanian, 2023; USDA, 2025). Smallholder plantations (<25 ha/holding) often experience inconsistent harvest quality. The adoption of innovative technologies is a key strategy for improving quality consistency and maintaining operational efficiency. Key challenges that need to be addressed through technology include determining the optimal harvest time and assessing the internal quality of Fresh Fruit Bunches (FFB).

The internal components, namely water, lipids, and free fatty acids (FFA), play an important role in quality control, indicate the level of deterioration, and directly influence market profitability (Misron et al., 2017; Nanda et al., 2024; Novianty, Gilang et al. 2023a). High FFA levels increase susceptibility to rancidity, flavor deterioration, saponification, and reduced yields. High moisture also affects hydrolysis, increasing the potential for FFA formation, particularly when the fruits are damaged or delayed before processing (Azeman et al., 2015). Mature oil palm fruits typically contain approximately 24.3% moisture (Misron et al., 2017) and 3.725% FFA (Nanda et al., 2024). Refiner association standards state that free fatty acids (expressed as palmitic acid) in crude palm oil (CPO) must not exceed 5% for human consumption (Che Man et al., 1999). This study focused on indicators of quality deterioration in relation to the maturity stage, rather than merely estimating yield potential. Therefore, FFA and moisture content were selected as indicators of quality degradation that require rapid monitoring.

A key limitation of conventional assays is that they can damage the sample, are labor-intensive, time-consuming, and costly, as they require extensive sample preparation. Near-infrared (NIR) spectroscopy has been widely applied as a reliable technique for determining the chemical components of oil palm materials (Adiarifia et al., 2024; Budiastra et al., 2024; Isdhiyanti et al., 2024; Nanda et al., 2025; Novianty et al. 2023a). The most commonly used spectrometer is the benchtop (laboratory-scale) type, which is costly, bulky, heavy, and unsuitable for field applications. A promising alternative is the deployment of compact, highly portable, rapid analysis instruments that are simple to operate, such as the SCiO portable NIR spectrometer.

JTEP Jurnal Keteknikan Pertanian,
Vol. 14 No. 1, p 108-125, 2026
P-ISSN 2407-0475 E-ISSN 2338-8439
Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.014.1.108-125

However, the high dimensionality of NIR spectral datasets can pose a significant challenge for conventional regression modeling because of multicollinearity and redundant information (Silalahi et al., 2020). Therefore, a calibration method based on Partial Least Squares Regression (PLSR) was employed to extract latent variables that capture the most relevant spectral information. The PLSR performance was compared with that of a hybrid method combining PLS and Artificial Neural Networks (ANN). ANN captures complex patterns within the data and models the nonlinear relationships. This study aimed to evaluate the performance of PLSR and PLS-ANN models with spectral pretreatments for predicting free fatty acid and moisture content of oil palm fruit using a portable NIR spectrometer.

2. Material and Methods

2.1 Research Framework

This study was carried out in six main steps: (1) preparing the instruments and materials, (2) NIR spectra acquisition, (3) performing chemical analyses, (4) pretreatment of the spectral data, (5) developing calibration models using PLSR and PLS-ANN, and (6) validating the models. The complete workflow is shown in Figure 1.

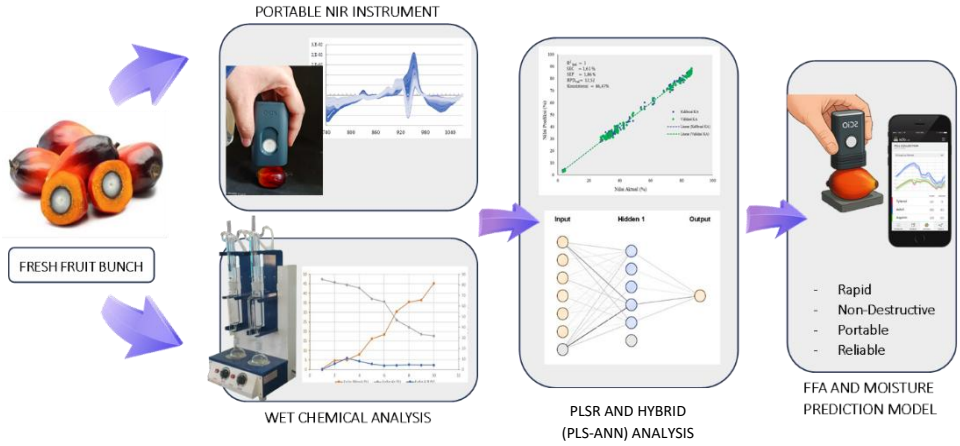


Figure 1. Research framework for FFA and moisture content prediction using PLSR and the hybrid (PLS-ANN) method.

2.2 Sample Preparation

Field sampling was conducted at the Cikabayan Oil Palm Plantation of IPB University, Bogor, Indonesia. Oil palm fruit samples (*Elaeis guineensis Jacq*. var. tenera), representing ten ripeness levels (3M, 4M, 4M1W, 4M2W, 4M3W, 5M, 5M1W, 5M2W, 5M3W, and 6M; M = month; W = week). Samples were collected from four different oil palm trees. To ensure representativeness, fruits were sampled from several positions within each bunch, including the upper, middle, and lower sections, as well as

Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.014.1.108-125

the distal, central, and proximal regions. For ripeness levels 3M to 5M, nine fruits were collected from each tree. For the more advanced ripeness levels (5M1W to 6M), 12 fruits were collected from each tree to account for the larger bunch size. A total of 408 fruit samples were analyzed. The samples were cleaned of contaminants, leaves, bunch residues and dust. The oil palm fruit samples were transported from the plantation to the laboratory using compartmentalized fruit container packaging to prevent mechanical damage. The fruits were harvested in the morning, and spectral acquisition was conducted within 8 h to avoid significant changes in chemical composition (Ruswanto et al., 2020). The samples were then stored under refrigerated conditions until the laboratory chemical analysis stage.

2.3 Spectral Acquisition

The NIR spectrometer used in this study was a handheld SCiO molecular sensor (version 1.2), which was selected because of its compact dimensions. The spectral signal was captured within the range of 740–1070 nm (≈13,514–9,346 cm^{-1}). The reflectance spectra were converted to absorbance using the log(1/R) transformation. Data management was carried out using the SCiO smartphone software (The Lab, version 1.3.8 (189)). The SCiO operates using a cloud-based data management system. Each fruit was scanned three times and averaged to obtain one spectrum (n = 408). To minimize spectral variation among the fruits, each sample was scanned individually with the sensor positioned perpendicularly to the fruit. The dataset was then randomly divided into training and testing subsets in a 7:3 proportion (Nguyen et al., 2021) using the Split Data operator in RapidMiner with the automatic sampling mode before any signal processing was applied.

2.4 Chemical Analysis

FFA measurements were conducted using titration with 0.1 N NaOH. The FFA content was expressed as % palmitic acid and calculated according to AOAC ed.18th 940.28, 2005, as shown in Equation 1:

$$\% \text{ Palmitic acid} = \frac{MW \times N \times V}{W \times 1000} \times 100\% \quad (1)$$

where MW is the molecular weight of palmitic acid (256), N is the normality of NaOH solution, V is the volume of the NaOH titrant (mL), and W is the sample weight (g).

Moisture content was measured using the oven-drying method according to the Indonesian National Standard (SNI 01-3555-1998) and reported on a wet basis:

$$m = \frac{a-b}{a} \times 100\% \quad (2)$$

Where m is the moisture content of the sample on a wet basis (%wb), a is the sample weight before drying (g), and b is the sample weight after drying (g).

JTEP Jurnal Keteknikan Pertanian,
Vol. 14 No. 1, p 108-125, 2026
P-ISSN 2407-0475 E-ISSN 2338-8439
Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.014.1.108-125

2.5 Spectral Pretreatment

Several pretreatment algorithms were applied to eliminate the influence of noise and improve the signal-to-noise ratio of the NIR spectra. The pretreatments used included a combination of first derivative Savitzky-Golay (D1 SG) + smoothing Savitzky-Golay (smoothing SG), D1 SG, Detrending, and a combination of D1 SG + Detrending. Derivative spectra are effective in enhancing the resolution of overlapped peaks, removing baseline variations and background interferences, thereby facilitating the development of a more accurate calibration model. However, increasing the derivative order tends to reduce spectral intensity and the signal-to-noise ratio, which may negatively affect model performance (Tong et al., 2015). Derivative-based preprocessing methods are typically combined with smoothing techniques, such as the Savitzky–Golay smoothing, to eliminate noise amplification that naturally occurs when differentiating noisy signals (Oliveri et al., 2019). In Detrending, the baseline trend is estimated using a least-squares approach by fitting each spectrum with a first- or second-order polynomial, which is then subtracted from the original spectrum to yield the corrected spectrum. Detrending can correct baseline shifts (Pruksapha et al., 2023). The first derivative SG (D1 SG) was calculated with derivative order 1 and polynomial order 2 using a 3-point window. Smoothing SG was performed using a polynomial order of 0 with a 3-point window (1 left, 1 right). The detrending was performed using a polynomial order of 2. These pretreatment parameters yielded the best model performance and were therefore selected for subsequent modeling studies.

2.6 Partial least squares regression (PLSR)

PLSR is applied to calibrate NIR spectra data with chemical data. The Unscrambler X version 10.4 (Camo Software) was used for PLSR analysis. During the PLSR analysis, outliers were identified and removed based on Hotelling's T^2 and Q residuals, resulting in a reduced number of samples compared to the original dataset. PLSR simultaneously decomposes the predictor and response variables into latent components (factor component(s)) ranked according to their correlations. Components with the highest correlations were selected for modeling to improve the model accuracy and efficiency (Yi et al., 2025). The optimal number of components was determined based on the lowest SEP while maintaining a small difference between the SEC and SEP to avoid overfitting. PLSR method decomposes the predictor design matrix X and the response matrix Y. This approach establishes a linear X–Y relationship, as expressed in Equation 3.

$$Y = \alpha_0 + \alpha_1 S_{1\lambda} + \alpha_2 S_{2\lambda} + \ldots + \alpha_n S_{n\lambda} \quad (3)$$

Where α_n denotes the regression coefficient calculated from the linear relationship between $S_n\lambda$ vs. $Y\lambda$ within the calibration iterations process, and $S_n\lambda$ represents the score of the n-th factor component at wavelength λ.

JTEP Jurnal Keteknikan Pertanian,
Vol. 14 No. 1, p 108-125, 2026
P-ISSN 2407-0475 E-ISSN 2338-8439
Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.014.1.108-125

2.7 Artificial neural network (ANN)

An Artificial neural network (ANN) is an information-processing system modeled after the human brain's mechanisms for acquiring knowledge through a network of logically organized processing units known as neurons. Their interconnections carry scalar weights that represent the strength and characteristics of the interactions (Mingione et al., 2022). ANN typically includes an input layer that receives data, one or more hidden layers that learn the empirical patterns, and an output layer that generates the final response. The model was constructed, trained, and implemented using the Backpropagation Neural Network (BPP-NN) toolbox in RapidMiner software. The general formula for an ANN is presented in Equation 4.

$$OA = Tf_0\left(\sum_{h=1}^{L} w_h^o . HL_h^k + bi_o\right) \quad (4)$$

Where: OA: Final output at the output layer, Tf_o: Activation function applied to the k-th output layer, L: Number of neurons in the k-th hidden layer, $W_h{}^o$: Weight value between the h-th in the k-th hidden layer and the neuron in the output layer, $HL_h{}^k$: Output of the h-th neuron in the k-th hidden layer, bi_o: Bias value in the output layer.

2.8 Model validation

The model was validated using test set validation, which involved evaluating the model using a subset of samples that were not included in the model calibration. The best model was selected based on a higher coefficient of determination (R^2) and lower standard error (SE) in both the calibration (C) and prediction (P) (Palou et al., 2023). Model performance was also assessed using the ratio of prediction to deviation (RPD), where values above 2.5 indicate fair predictive capability (Williams, 2014), as well as model consistency within the range of 80–110% (Andasuryani & Ifmalinda, 2024).

$$R^2 = 1 - \left(\frac{\sum_{i=1}^{n}(Y_{NIR}-Y_R)^2}{\sum_{i=1}^{n}(Y_{NIR}-Y_m)^2}\right) \quad (5)$$

$$SE = \left(\frac{1}{N-1} \times \sum(Y_R - Y_{NIR} - BIAS)^2\right)^{1/2} \quad (6)$$

$$RPD = \frac{SD}{SE} \quad (8)$$

$$Consistency = \frac{SEC}{SEP} \times 100\% \quad (9)$$

where N represents the number of samples, and SEC and SEP denote the standard errors of the calibration and validation, respectively. where SD refers to the standard deviation, YNIR is the predicted value, YR is the reference value, and Ym is the mean.

JTEP Jurnal Keteknikan Pertanian,
Vol. 14 No. 1, p 108-125, 2026
P-ISSN 2407-0475 E-ISSN 2338-8439
Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.014.1.108-125

3. Results and Discussion

3.1 Chemical Value

Table 1 presents the descriptive statistics of the chemical compositions of the oil palm fruit samples. The FFA content showed a standard deviation of 1.94, indicating moderate variation among the samples. In contrast, the moisture content exhibited a much higher standard deviation (22.92), suggesting greater variability in moisture levels. The trends in the average chemical content are shown in Figure 2. The average FFA content remained relatively low, generally under 5%, except at 4 months 3 weeks. At an early age, the FFA levels tended to be low. However, during this period, lipase enzymes are already present in the oil palm fruit. These enzymes act as catalysts in the hydrolysis of triglycerides (a type of lipid found in oil palms). Free fatty acids (FFA) are formed through hydrolysis during oil processing and through natural biochemical processes in palm fruits (Nanda et al., 2024). As shown in Figure 2, the moisture content tended to decrease with increasing maturity. The reduction in moisture content is triggered by the accumulation of assimilates from photosynthesis, such as oil, proteins, and other compounds, and the formation of kernel mass within the mesocarp (Krisdiarto et al., 2017).

Table 1. Descriptive statistics of the chemical content.

Chemical content	n	Min (%)	Max (%)	Mean (%)	SD (%)
FFA	345	0.00	7.77	3.62	1.94
Moisture	314	27.75	86.39	58.61	22.92

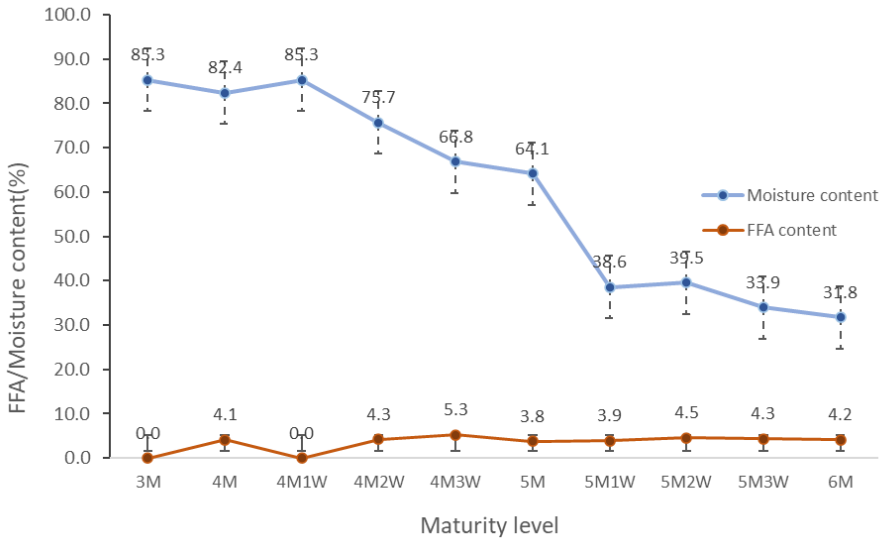


Figure 2. Reference (measured) values of FFA and moisture content across ripeness stages.

JTEP Jurnal Keteknikan Pertanian,
Vol. 14 No. 1, p 108-125, 2026
P-ISSN 2407-0475 E-ISSN 2338-8439
Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.014.1.108-125

The two parameters, FFA and moisture content, used different numbers of samples in this study. FFA analysis was performed on 345 of the 408 samples after removing outliers and samples with extreme values. Similarly, the moisture content analysis used 314 samples after the outlier detection step during PLSR analysis.

3.2 Spectra Characteristics

Figure 3 shows the original spectra and spectra generated using the pretreatment methods. The D1 SG + smoothing SG (Figure 3 (b)) causes the absorbance range to be concentrated on a scale of 10^{-2}. The spectra appear to be dominated by D1 SG (similar to Figure 3 (d)) with smoothing effects. The peak–valley features became more concentrated, particularly between 900 and 950 nm, allowing for more detailed and specific chemical analyses. The similarity in the spectral patterns is mainly due to the dominant effect of the first derivative Savitzky–Golay (D1 SG) transformation, which strongly determines the spectral shape after preprocessing. The smoothing SG step primarily reduces high-frequency noise and has a minimal impact on the overall spectral structure. As a result, the spectra remain visually similar to those obtained using the D1 SG pretreatment alone. Minor differences can still be observed on a finer axis scale.

The spectrum resulting from the detrending pre-treatment (Figure 3 (c)) exhibits peaks and valleys positioned similarly to the original spectra along the X-axis, but with a more centered and specific range, as well as a reduced absorbance range on the Y-axis. Similarly, the combination of D1 SG with detrending produced comparable spectral patterns because detrending mainly removed baseline variations without significantly altering the derivative spectral features. Consequently, although different pretreatments were applied, the resulting spectra appeared visually similar. Differences were observed, particularly at wavelengths above approximately 1040 nm.

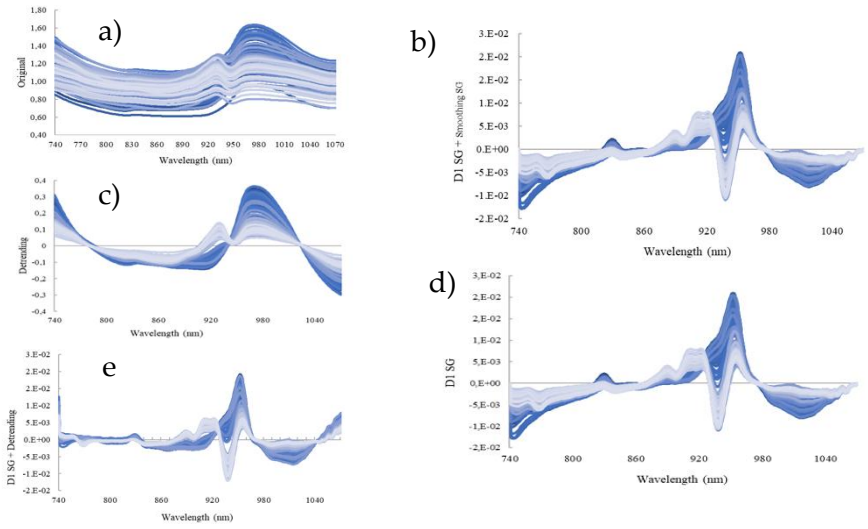


Figure 3. NIR spectral absorbance of oil palm fruits: (a) original; (b) D1 SG + smoothing SG; (c) detrending; (d) first derivative SG (D1 SG); (e) first derivative SG + Detrending.

JTEP Jurnal Keteknikan Pertanian,
Vol. 14 No. 1, p 108-125, 2026
P-ISSN 2407-0475 E-ISSN 2338-8439
Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.014.1.108-125

NIR spectra typically contain absorption bands associated with three main types of chemical bonds: C–H (fats, oils, and hydrocarbons), O–H (water and alcohols), and N–H (proteins) (Novianty, Gilang et al. 2023a; Stuart, 2004) [A38.1]. The peak in the 900–950 nm region that appeared in the spectral output of D1 SG + smoothing SG is linked to the molecular features characteristic of fats and oils (MacArthur et al., 2020). The Detrended spectra exhibited a minor peak at 930 nm and a major peak at 970 nm. These wavelengths are strongly associated with water (~970 nm) and oil (~930 nm) (Clark et al., 2003; Li et al., 2018).

3.3 PLSR and PLS-ANN Implementation

The graph in Figure 4 illustrates the relationship between the number of PLS factor components and the amount of variance explained. The higher the explained variance, the more accurate the model; the closer the calibration and validation lines, the more consistent the model. The factors obtained from the FFA prediction model resulting from the D1 SG + smoothing SG pretreatment are shown in Figure 4 (a). The optimal number of factors was 11, which showed a high explained variance. The factors obtained from the moisture content prediction model through detrending pretreatment are shown in Figure 4 (b); the optimal number of factors is 15, which shows the highest explained variance. The development of the PLSR model involves the application of regression, whereas in PLS-ANN, the process ends with the generation of factor components (FC) from PLS analysis, which are subsequently used as inputs for the ANN.

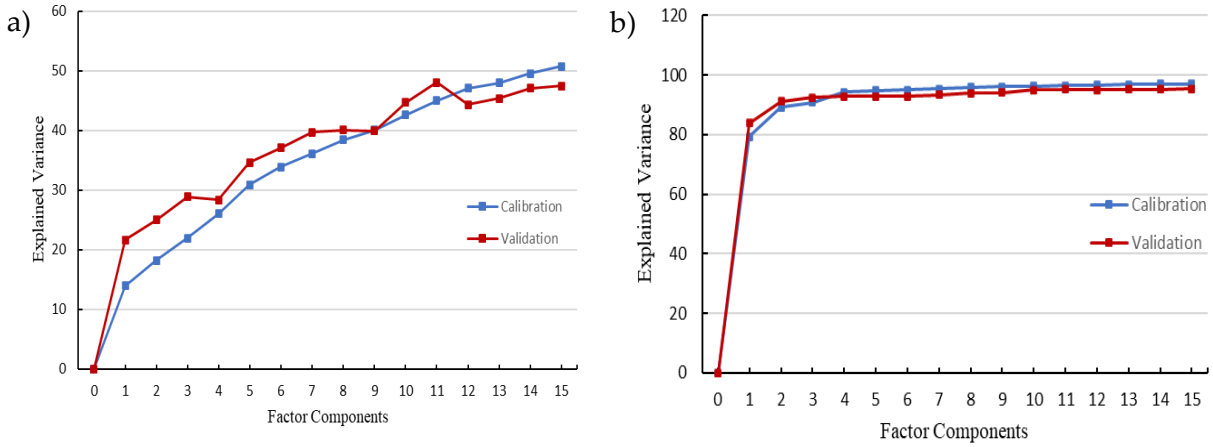


Figure 4. Graphs of explained variance for the number of FCs (a) FFA (b) moisture.

The structure of the ANN algorithm is an MLP (multi-layer perceptron) composed of input, hidden, and output layers. The optimal combination of neurons and activation functions yielded a structure with 11 neurons for FFA prediction and 30 neurons for moisture content prediction, arranged in two hidden layers, and employing a logistic sigmoid activation function. An appropriate range for the

JTEP Jurnal Keteknikan Pertanian,
Vol. 14 No. 1, p 108-125, 2026
P-ISSN 2407-0475 E-ISSN 2338-8439
Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.014.1.108-125

number of neurons lies between N and 2N + 1, where N represents the number of input factors (Walczak & Cerpa, 2003). This aligns with our results, as the optimal configuration was within the suggested range. Table 2 summarizes the ANN architecture.

Table 2. ANN architecture.

ANN Parameters	Information of Prediction Models	
	FFA Content	Moisture Content
Number of input factor	11	15
Number of hidden layers	1	2
Neuron number	11	30
Learning rate	0,01	0,02
Momentum	0,9	0,9
Training cycle	500	500
Maximum training cycle	1500	1500
Activation Function	Logistic sigmoid	Logistic sigmoid

3.4 Model Evaluation

3.4.1 Free Fatty Acid

The performance of the prediction model for the FFA content is presented in Table 3. The model developed using the PLSR method demonstrated the best performance, with R^2cal = 0.44, SEC = 1.38%, SEP = 1.54%, RPD_val = 1.38, and consistency of 89.54%. The applied spectral pretreatment was a combination of D1 SG + Detrending, utilizing 10 factor components. The model was not recommended for application according to the RPD degree of merit (Williams, 2014). Despite the low accuracy, the model consistency of 89.54% indicates a stable predictive performance between the calibration and validation. Therefore, further optimization of the modeling process is required to improve its accuracy and reliability.

Table 3. Evaluation of the FFA content prediction model performance.

Pretreatment	Method	FC	Calibration (n=241)			Validation (n=104)		Consistency (%)
			R^2	SEC (%)	RPD_cal	SEP (%)	RPD_val	
Ori	PLSR	10	0.36	1.48	1.25	1.67	1.27	88.62
		6	0.28	1.57	1.18	1.74	1.23	90.47
		2	0.21	1.64	1.13	1.77	1.21	92.91
	PLS-ANN	10	0.62	0.97	1.61	1.20	2.08	80.84
		6	0.57	1.21	1.53	1.38	1.55	88.02
		2	0.35	1.50	1.24	1.54	1.38	96.80
D1 SG + smoothing SG	PLSR	15	0.51	1.30	1.42	1.55	1.38	83.87
		11	0.45	1.37	1.35	1.54	1.38	89.14
		10	0.43	1.40	1.32	1.59	1.34	88.23

JTEP Jurnal Keteknikan Pertanian,
Vol. 14 No. 1, p 108-125, 2026
P-ISSN 2407-0475 E-ISSN 2338-8439
Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.014.1.108-125

Pretreatment	Method	FC	Calibration (n=241)			Validation (n=104)		Consistency (%)
			R^2	SEC (%)	RPD_cal	SEP (%)	RPD_val	
	PLS-ANN	15	0.88	0.65	2.85	0.96	2.22	67.50
		11	0.81	0.80	2.31	0.91	2.34	87.88
		10	0.70	1.02	1.82	1.09	1.95	92.94
D1 SG + Detrending	PLSR	15	0.51	1.30	1.42	1.56	1.37	83.43
		10	0.44	1.38	1.34	1.54	1.38	89.54
		7	0.38	1.38	1.27	1.66	1.28	87.74
	PLS-ANN	15	0.81	0.81	2.29	1.23	1.74	65.79
		10	0.77	0.89	2.29	1.23	1.74	88.35
		7	0.60	1.18	1.57	1.43	1.49	82.20
Detrending	PLSR	10	0.38	1.46	1.27	1.67	1.28	87.62
		7	0.34	1.50	1.23	1.67	1.28	89.96
		6	0.33	1.51	1.23	1.66	1.29	91.02
	PLS-ANN	10	0.68	1.05	1.77	1.41	1.51	74.12
		7	0.53	1.08	1.44	1.41	1.52	76.86
		6	0.58	1.20	1.54	1.25	1.71	96.34

The performance comparison between the PLSR and PLS-ANN models is visually illustrated in the scatter plot in Figure 5, where the sample points appear more concentrated and closer to the regression line, shown a significant improvement in performance through PLS-ANN. The obtained values were R^2cal = 0.81, SEC = 0.80%, SEP = 0.91%, RPD_val = 2.34, and consistency = 87.88%. The model was constructed using D1 SG and smoothing SG pretreatments with 11 factor components. Overall, the parameters indicate that the model is suitable for rough screening purposes, such as coarse classification of samples (Williams, 2014).

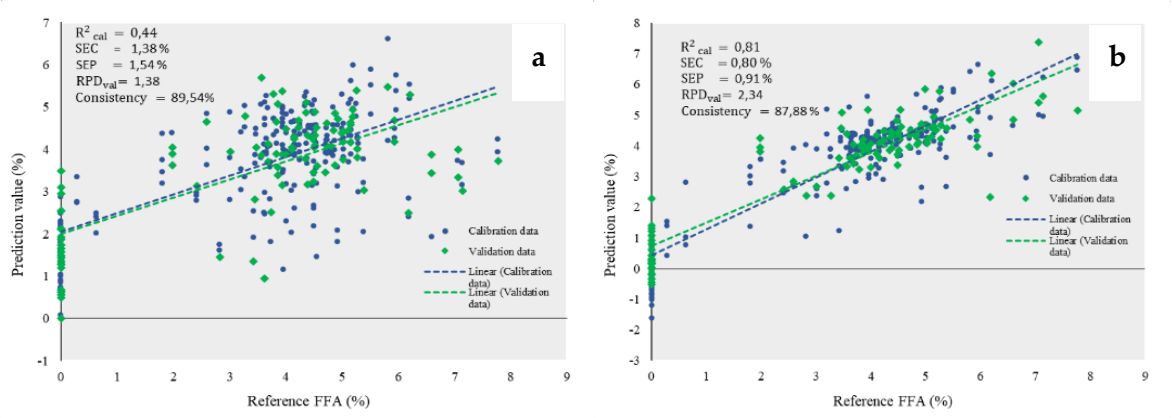


Figure 5. Predicted vs. actual value scatter plot: (a) PLSR; b) PLS-ANN in FFA content prediction models.

Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.014.1.108-125

The proposed model demonstrated competitive performance compared with that of previous studies. For example, a study using a portable NIR instrument, the Si-Ware NeoSpectra Micro SWS62231 (1300–2500 nm), with PLSR calibration reported an R^2 value of only 0.70 (Hermawan, 2025). Similarly, Kaufman et al. (2019) used a portable NIR spectrometer (TIDA-00554 DLP; Texas Instruments, Dallas, TX, USA) with PLSR calibration and reported relatively low predictive performance, with an R^2 of 0.37 and RPD of 1.20. Studies that used laboratory-scale instruments with a broader spectral range, such as the NIRFlex N-500 (1000–2500 nm), with PLS-ANN models achieved higher predictive performance, with R^2 and RPD values of 0.99 and 11.47 by Adiarifia et al. (2024) and 0.96 and 4.24 by Budiastra et al. (2024). Even more advanced methods using NIRFlex N-500, such as deep learning (LSTM) (Nanda et al., 2024), produced high accuracy (R^2 0.959). In contrast, this study used an SCiO Portable NIR spectrometer (740–1070 nm), which has a much narrower wavelength range, resulting in more limited chemical information being captured.

Fatty acids are a type of carboxylic acid (an organic acid that possesses a -COOH functional group) characterized by a long hydrocarbon side chain, and are commonly present in fats, oils, and biological membranes. The low accuracy in FFA prediction may stem from the difficulty in measuring acids using NIR, since the functional group of acid (-COOH) contains covalent bonding between two large atoms, carbon and oxygen, resulting in a smaller dipole moment and low absorptivity compared to other chemical bonds between small and large atoms (C-H or O-H) (Arruda de Brito et al., 2022; Li et al., 2018). Additionally, the low prediction performance may be influenced by the low concentration of FFA in the oil palm and the narrow FFA distribution range (0.0–7.77%).

In addition, water is a strong NIR absorber, and its presence can dominate the spectra and interfere with the detection of other components, including the functional groups of acids, particularly at high moisture content. A study by Li et al. (2018) using a portable SCiO NIR sensor to predict acid functional groups in feijoa fruit reported similar limitations, with relatively low accuracy (R^2 = 0.44 and RMSE = 0.4). This limitation was presumed to be associated with interference from the high moisture content (80-82%). In contrast, in this present study, oil palm fruits with a lower moisture content (27.75–86.39%) exhibited the same general challenge; however, a higher prediction accuracy was achieved.

3.4.2 Moisture Content

As shown in Table 4, the best PLSR model achieved R^2cal = 0.97, SEC = 3.99%, SEP = 4.74%, RPD_val = 4.92, and consistency = 84.30%. The model was generated using D1 SG pre-treatment with 10 FC. The RPD_val value of 4.92, which is greater than 4, reflects excellent predictive performance, indicating that the model is suitable for various applications to the same type of material (Williams, 2014). Overall, the PLSR model demonstrated robust reliability across nearly all parameters. Nonetheless, there is still potential to maximize its overall performance in the future.

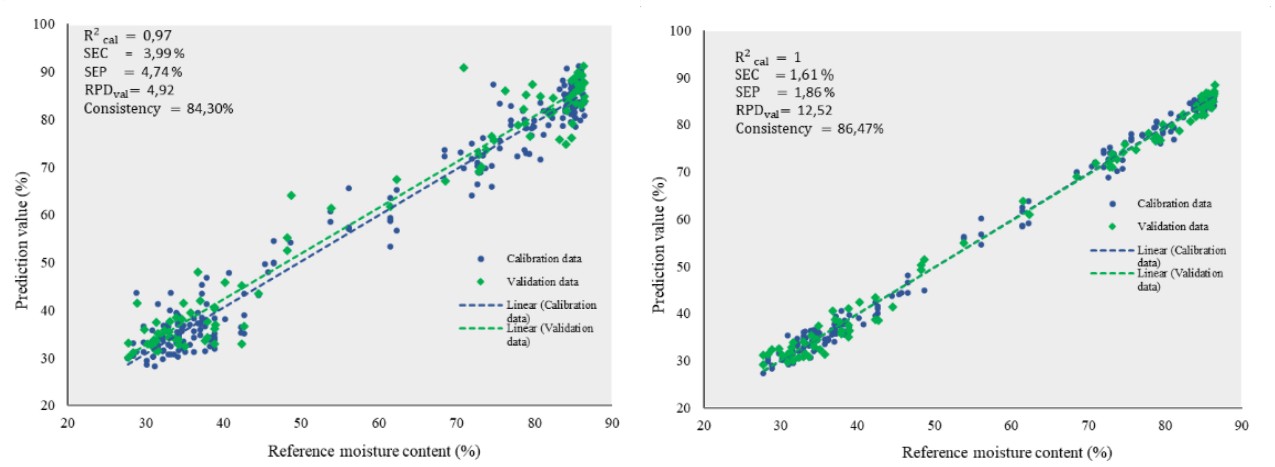
JTEP Jurnal Keteknikan Pertanian,
Vol. 14 No. 1, p 108-125, 2026
P-ISSN 2407-0475 E-ISSN 2338-8439
Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.014.1.108-125

Table 4. Evaluation of the performance of the moisture content prediction model.

Pretreatment	Method	FC	Calibration (n=220)			Validation (n=94)		Consistency (%)
			R^2	SEC (%)	RPD_cal	SEP (%)	RPD_val	
Ori	PLSR	10	0.96	4.30	5.30	5.12	4.55	83.98
		7	0.95	5.09	4.48	6.19	3.76	82.21
		4	0.93	5.99	3.81	6.08	3.84	98.51
	PLS-ANN	10	0.98	2.81	8.12	3.69	6.32	76.06
		7	0.98	3.07	7.42	3.78	6.16	81.22
		4	0.98	3.54	6.45	4.10	5.69	86.22
D1 SG+ smoothing SG	PLSR	15	0.97	3.71	6.15	4.77	4.89	77.76
		10	0.97	4.00	4.91	4.75	4.91	84.29
		7	0.96	4.55	5.01	5.60	4.17	81.32
	PLS-ANN	15	0.99	1.69	13.51	3.23	7.22	52.20
		10	0.99	1.90	11.97	2.05	11.38	92.99
		7	0.98	2.87	7.93	3.26	7.16	88.29
Detrending	PLSR	15	0.97	3.95	5.77	4.82	4.83	81.98
		10	0.96	4.39	5.19	5.00	4.67	87.88
		7	0.95	4.93	4.62	5.80	4.02	84.97
	PLS-ANN	15	1.00	1.61	14.16	1.86	12.52	86.47
		10	0.99	2.13	10.72	2.66	8.75	79.82
		7	0.98	2.82	8.09	3.00	7.78	94.05
D1 SG	PLSR	10	0.97	3.99	5.71	4.74	4.92	84.30
		7	0.96	4.54	5.02	5.58	4.18	81.32
		2	0.90	7.06	3.23	6.71	3.47	105.13
	PLS-ANN	10	0.99	1.95	11.67	2.31	10.08	84.45
		7	0.99	2.68	8.52	3.12	7.48	85.85
		2	0.97	3.85	5.92	4.08	5.72	94.48

The hybrid PLS-ANN model demonstrated superior performance. The comparison between PLSR and the hybrid PLS-ANN model performance is visually presented in the scatter plot in Figure 6. The data points of the PLS-ANN method were closer to the regression line, indicating that the predicted values were more consistent with the actual values than those of the PLSR method. The best model achieved R^2cal = 1, SEC = 1.61%, SEP = 1.86%, RPD_val = 12.52, and a consistency of 86.47%. The model was developed using detrending pre-treatment with 15 factor components. An RPD_val value of 12.52 is generally classified as excellent and suitable for various applications, including high-accuracy quantitative analysis (Williams, 2014). Water strongly absorbs NIR radiation, making NIR highly sensitive to moisture measurements (ABVista, 2018).

JTEP Jurnal Keteknikan Pertanian,
Vol. 14 No. 1, p 108-125, 2026
P-ISSN 2407-0475 E-ISSN 2338-8439
Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.014.1.108-125



(a) (b)

Figure 6. Predicted vs. actual value scatter plot (a) PLS; b) PLS-ANN, in the moisture prediction models.

The performance of the moisture prediction model was higher than that reported by Teye et al. (2023), who used a similar instrument and wavelength range but obtained R^2 values of only 0.361 (PLS) and 0.943 (SiPLS). Compared with studies employing the NIRFlex N-500 (1000–2500 nm), which reported R^2 values of 0.975 with an RMSE of 2,487 (Novianty et al., 2023b) and R2 of 0.99 with an RPD of 15.51 (Adiarifia et al., 2024), the performance achieved in this study was competitive.

4. Conclusion

The Hybrid PLS-ANN model resulted in higher accuracy compared to the PLS method for both FFA and moisture content predictions. Based on the performance evaluation, the proposed model can predict the FFA content with R^2cal = 0.81, SEC = 0.80%, SEP = 0.91%, RPD_val = 2.34, and consistency = 87.88%. The FFA prediction model is suitable for coarse preliminary sorting at the early stage of evaluation. The prediction model for moisture content can be applied to various quantitative analyses with high accuracy, with R^2cal = 1, SEC = 1.61%, SEP = 1.86%, RPD_val = 12.52, and a consistency of 86.47%.

5. AI Writing Statement

During the preparation of this work, the authors used [ChatGPT and Grammarly] in order to [do a grammar check]. The authors have reviewed and edited the content as needed. The authors will take full responsibility for the content of the published article.

Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.014.1.108-125

6. Acknowledgements

This research was supported by the BIMA Program organized by the Directorate General of Research and Development, Ministry of Higher Education, Science, and Technology of the Republic of Indonesia, under the Research Program Implementation Contract for Fiscal Year 2025 (Number: 006/c3/DT.05.00/PL/2025).

7. References

ABVista. (2018). A Guide to NIR: Understanding NIR spectra. https://www.abvista.com/news/a-guide-to-nir-understanding-nir-spectra

Adiarifia, N., Budiastra, I., & Mardjan, S. S. (2024). Non-Destructive prediction of chemical content in palm oil fruit using near-infrared spectroscopy and artificial neural network. Jurnal Keteknikan Pertanian, 12(1), 128–139.

Andasuryani, & Ifmalinda. (2024). Rapid prediction of moisture and ash content in sungkai leaves herbal tea (Peronema canescens Jack.) using NIR Spectroscopy. 12(3), 301–313. https://doi.org/10.19028/jtep.012.3.301-313

Arruda de Brito, A., Campos, F., dos Reis Nascimento, A., Damiani, C., Alves da Silva, F., de Almeida Teixeira, G. H., & Cunha Júnior, L. C. (2022). Non-destructive determination of color, titratable acidity, and dry matter in intact tomatoes using a portable Vis-NIR spectrometer. Journal of Food Composition and Analysis, 107, 104288. https://doi.org/https://doi.org/10.1016/j.jfca.2021.104288

Azeman, N. H., Yusof, N. A., & Othman, A. I. (2015). Detection of free fatty acid in crude palm oil. Asian Journal of Chemistry, 27(5), 1569–1573. https://doi.org/10.14233/ajchem.2015.17810

Budiastra, W., Marjan, S., Adiarifia, N., Novianty, I., & Suci, Y. T. (2024). Non-destructive prediction of oil and free fatty acid of oil palm fruitlets using near-infrared spectroscopy and hybrid calibration method. INMATEH - Agricultural Engineering, 73(2), 461–472. https://doi.org/10.35633/inmateh-73-39

Che Man, Y. B., Moh, M. H., & van de Voort, F. R. (1999). Determination of free fatty acids in crude palm oil and refined-bleached-deodorized palm olein using fourier transform infrared spectroscopy. Journal of the American Oil Chemists' Society, 76(4), 485–490. https://doi.org/10.1007/s11746-999-0029-z

Clark, C. J., McGlone, V. A., Requejo, C., White, A., & Woolf, A. B. (2003). Dry matter determination in 'Hass' avocado by NIR spectroscopy. Postharvest Biology and Technology, 29(3), 301–308. https://doi.org/https://doi.org/10.1016/S0925-5214(03)00046-2

Hermawan, H. F. (2025). Prediksi kandungan kimia buah kelapa sawit secara non-destruktif menggunakan [IPB University]. https://repository.ipb.ac.id/handle/123456789/164889

Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.014.1.108-125

Isdhiyanti, M. I. T., Budiastra, I. W., Mala, D. M., Mardjan, S., Hasrini, R. F., & Supriatna, D. (2024). Prediction of chemical contents of cooking oil using near-infrared spectroscopy. Food Research, 8(6), 147–152. https://doi.org/https://doi.org/10.26656/fr.2017.8(6).520

Kaufman, K., Favero, F., Vasconcelos, M., Godoy, H., Sampaio, K., & Barbin, D. (2019). Portable nir spectrometer for prediction of palm oil acidity. Journal of Food Science, 84(3), 406–411. https://doi.org/doi.org/10.1111/1750-3841.14467

Kementerian Pertanian. (2023). Kontribusi minyak kelapa sawit indonesia mengatasi krisis pangan global. https://ditjenbun.pertanian.go.id/kontribusi-minyak-kelapa-sawit-indonesia-mengatasi-krisis-pangan-global/

Krisdiarto, A. W., Sutiarso, L., & Widodo, K. H. (2017). Optimasi kualitas tandan buah segar kelapa sawit dalam proses panen-angkut menggunakan model dinamis. Agritech, 37(1), 102. https://doi.org/10.22146/agritech.17015

Li, M., Qian, Z., Shi, B., Medlicott, J., & East, A. (2018). Evaluating the performance of a consumer scale SCiOTM molecular sensor to predict quality of horticultural products. Postharvest Biology and Technology, 145(March), 183–192. https://doi.org/10.1016/j.postharvbio.2018.07.009

MacArthur, R. L., Teye, E., & Darkwa, S. (2020). Predicting adulteration of Palm oil with Sudan IV dye using shortwave handheld spectroscopy and comparative analysis of models. Vibrational Spectroscopy, 110, 103129. https://doi.org/https://doi.org/10.1016/j.vibspec.2020.103129

Mingione, E., Leone, C., Almonti, D., Menna, E., Baiocco, G., & Ucciardello, N. (2022). Artificial neural networks application for analysis and control of grapes fermentation process. Procedia CIRP, 112(March), 22–27. https://doi.org/10.1016/j.procir.2022.09.018

Misron, N., Aliteh, N. A., Harun, N. H., Tashiro, K., Sato, T., & Wakiwaka, H. (2017). Relative estimation of water content for flat-type inductive-based oil palm fruit maturity sensor. In Sensors (Vol. 17, Issue 1). https://doi.org/10.3390/s17010052

Nanda, M. A., Amaru, K., Rosalinda, S., Novianty, I., & Park, T. (2025). Multi-parameter prediction of oil palm fruit quality through near infrared spectroscopy combined with chemometric analysis. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 343, 126505. https://doi.org/https://doi.org/10.1016/j.saa.2025.126505

Nanda, M. A., Amaru, K., Rosalinda, S., Novianty, I., Sholihah, W., Mindara, G. P., Faricha, A., & Park, T. (2024). Higuchi fractal dimension and deep learning on near-infrared spectroscopy for determination of free fatty acid (FFA) content in oil palm fruit. Journal of Agriculture and Food Research, 18(February), 101437. https://doi.org/10.1016/j.jafr.2024.101437

Nguyen, Q. H., Ly, H.-B., Ho, L. S., Al-Ansari, N., Le, H. Van, Tran, V. Q., Prakash, I., & Pham, B. T. (2021). Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil. Mathematical Problems in Engineering, 2021(1), 4832864. https://doi.org/https://doi.org/10.1155/2021/4832864

Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.014.1.108-125

Novianty, I., Nanda, M.A., Gilang, R., & Iqbal, M. (2023a). Empirical mode decomposition of near-infrared spectroscopy signals for predicting oil content in palm fruits. Information Processing in Agriculture, 10(3), 289–300. https://doi.org/10.1016/j.inpa.2022.02.004

Novianty, I., Sholihah, W., Mindara, G. P., Nurulhaq, M. I., Faricha, A.,Sinambela, R., Purwandoko, P.B., Nanda, M.A. (2023b). Shannon entropy on near-infrared spectroscopy for nondestructively determining water content in oil palm. Int. J. of Electrical and Computer Engineering 13(5), 5397–5405. https://doi.org/10.11591/ijece.v13i5.pp5397-5405

Oliveri, P., Malegori, C., Simonetti, R., & Casale, M. (2019). The impact of signal pre-processing on the final interpretation of analytical outcomes – A tutorial. Analytica Chimica Acta, 1058, 9–17. https://doi.org/https://doi.org/10.1016/j.aca.2018.10.055

Palou, A., Jiménez, P., Casals, J., & Masaló, I. (2023). Evaluation of the Near Infrared Spectroscopy (NIRS) to predict chemical composition in Ulva ohnoi. Journal of Applied Phycology, 35(5), 2007–2015. https://doi.org/10.1007/s10811-023-02939-8

Pruksapha, P., Khongkaew, P., Suwanvecho, C., Nuchtavorn, N., Phechkrajang, C., & Suntornsuk, L. (2023). Chemometrics-assisted spectroscopic methods for rapid analysis of combined anti-malarial tablets. Journal of Food and Drug Analysis, 31(2), 338–357. https://doi.org/10.38212/2224-6614.3449

Silalahi, D. D., Midi, H., Arasan, J., Mustafa, M. S., & Caliman, J.-P. (2020). Robust wavelength selection using filter-wrapper method and input scaling on near infrared spectral data. In Sensors (Vol. 20, Issue 17). https://doi.org/10.3390/s20175001

Stuart, B. (2004). Infrared spectroscopy: fundamentals and applications. John Wiley & Sons, Ltd. https://doi.org/10.1002/0470011149

Teye, E., Amuah, C. L. Y., Yeh, T., & Nyorkeh, R. (2023). Nondestructive detection of moisture content in palm oil by using portable vibrational spectroscopy and optimal prediction algorithms. Journal of Analytical Methods in Chemistry, 2023(1). https://doi.org/10.1155/2023/3364720

Tong, P., Du, Y., Zheng, K., Wu, T., & Wang, J. (2015). Chemometrics and intelligent laboratory systems improvement of NIR model by fractional order savitzky – golay derivation (FOSGD) coupled with wavelength selection. Chemometrics and Intelligent Laboratory Systems, 143, 40–48. https://doi.org/10.1016/j.chemolab.2015.02.017

USDA. (2025). United States Department of Agriculture 2025: Production-Palm Oil. https://www.fas.usda.gov/data/production/commodity/4243000

Walczak, S., & Cerpa, N. (2003). Artificial Neural Networks (R. A. B. T.-E. of P. S. and T. (Third E. Meyers (ed.); pp. 631–645). Academic Press. https://doi.org/https://doi.org/10.1016/B0-12-227410-5/00837-1

Williams, P. (2014). The RPD statistic : a tutorial note. NIR News, 25(1), 22–26. https://doi.org/https://doi.org/10.1255/nirn.1419

JTEP Jurnal Keteknikan Pertanian,
Vol. 14 No. 1, p 108-125, 2026
P-ISSN 2407-0475 E-ISSN 2338-8439
Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.014.1.108-125

Yi, M., Chen, W., & Liu, J. (2025). Total partial least square regression and its application in infrared spectra quantitative analysis. Measurement, 247, 116794. https://doi.org/https://doi.org/10.1016/j.measurement.2025.116794.