

Estimating Macronutrient Content of Paddy Soil Based on Near-Infrared Spectroscopy Technology Using Multiple Linear Regression

Jonni Firdaus^{1,2,*}, Usman Ahmad³, I Wayan Budiastra³, I Dewa Made Subrata³

¹Agricultural Engineering Science Study Program, IPB University, Bogor, Indonesia

²National Research and Innovation Agency Republic of Indonesia, Jakarta, Indonesia

³Department of Mechanical and Biosystem Engineering, IPB University, Bogor, Indonesia

Received 04 August 2023; received in revised form 10 November 2023; accepted 11 November 2023

DOI: <https://doi.org/10.46604/aiti.2023.12683>

Abstract

This study investigates the feasibility of employing near-infrared (NIR) spectroscopy with multiple linear regression (MLR) to estimate macronutrients in paddy soil compared with partial least squares (PLS) and principal component regression (PCR). Seventy-nine soil samples from West Java Province, Indonesia, are subject to conventional nutrient analysis and NIR spectroscopy (1000-2500 nm). The reflectance data undergoes various pretreatment techniques, and MLR models are calibrated using the forward method to achieve correlations exceeding 0.90. The best model calibrations are selected based on high correlation coefficients, determination coefficients, RPD, and low RMSE values. Meanwhile, the comparison of performance MLR is made with the PLS and PCR models. Results indicate that simple MLR models perform less than PLS for all nutrients, better than PCR for nitrogen, and below PCR for phosphorus and potassium. However, MLR reliably estimates soil nitrogen, phosphorus, and potassium content with ratio of performance to deviation (RPD) exceeding 2.0. This study demonstrates the potential of MLR for precise macronutrient estimation in paddy soil.

Keywords: fertility, near-infrared spectroscopy, nitrogen, phosphorus, potassium

1. Introduction

Soil plays a crucial role in agriculture by providing the necessary nutrients for plants to grow and ultimately supporting human food availability. Due to the diverse soil conditions, the practices of agricultural land management must be adapted accordingly. Maintaining a consistent nutrient supply during the crop-growing phase maximizes crop productivity [1]. Hence, soil health is the continuation of the soil's capacity to function as a vital living ecosystem that supports plants and ensures all essential soil functions. One of the critical aspects of soil health is soil fertility. Soil fertility is primarily determined by the presence of macronutrients such as nitrogen (N), phosphorus (P), and potassium (K), which are required in large quantities. Furthermore, soil fertility parameters reflect the soil's ability to supply plant nutrients.

Regular monitoring of soil nutrient levels is crucial to ensure optimal soil nutrient availability. Traditional laboratory methods of soil analysis are time-consuming, costly, and risky. It often requires heedful operation for hazardous chemicals harmful to the environment, resulting in impracticability for farmers. Therefore, developing accurate, environmentally friendly, time-efficient, and cost-effective soil analysis methods is imperative [2]. However, it is essential to identify various near-infrared (NIR) wavelengths that impact nutrients through straightforward multivariate models such as multiple linear regression (MLR) equations in the initial phase of creating a device that is lightweight, portable, and user-friendly.

* Corresponding author. E-mail address: jonni_firdaus@yahoo.com

West Java is one of Indonesia's provinces with the most extensive rice fields, requiring fast soil fertility monitoring to ensure practical support for plant growth. Earlier research has indicated significant variations in the accuracy of soil nutrient prediction models across different regions and soil types. Consequently, studying and exploring precision and practical application effects are continuously needed, particularly concerning diverse origins and types of croplands [3].

NIR spectroscopy is a promising alternative to traditional soil analysis methods. Specifically, NIR spectroscopy is a non-destructive chemical content detection technology that is fast and simple in sample preparation, while chemicals are not needed, conducting to an environmentally friendly solution. Numerous studies have demonstrated the efficacy of NIR spectroscopy in soil analysis, with its ability to rapidly and accurately predict soil nutrient contents and other soil properties. Therefore, adopting NIR spectroscopy as a soil analysis tool can significantly benefit farmers and facilitate sustainable agriculture practices. Many studies have been conducted on NIR to determine property characteristics and soil fertility [4-8]. The accuracy of deploying NIR to estimate soil nitrogen was relatively high, with an R^2 value of 0.75-0.95 [9-15]. Research on phosphorus estimation showed a low R^2 value [16], but several other studies obtained high R^2 values of 0.63-0.91 [17]. Meanwhile, potassium prediction obtained 0.47-0.59 for R^2 value [2], and other researchers got higher R^2 values standing at 0.92-0.99 [3, 18].

Nitrogen prediction has been studied using Foss NIRSystems 5000 (FOSS NIRSystems, Inc, Laurel, USA) (1100-2498 nm) and partial least squares (PLS) with reflectance (R) and first derivative (D1) pretreatment in 360 samples, resulting in an R^2 value of 0.77 and ratio of performance to deviation (RPD) of 2.10 [9]. The study used a FieldSpec® Pro sensor (Analytical Spectral Devices Inc., Colorado, USA), visible-near infrared (Vis-NIR) range (350-2500 nm), and support vector machine (SVM) with absorbance pretreatment in a total of 210 samples, resulting in an R^2 value of 0.75 [11]. Munawar et al. [12] employed the benchtop NIR instrument Thermo Nicolet Antaris II (Thermo Fisher Scientific Inc, Waltham, USA) range (1000-2500 nm), principal component regression (PCR), and PLS in 40 samples, resulting in R^2 values of 0.85 and 0.87, respectively.

Regarding RPD values, they stood at 2.00 and 3.50, respectively. Pudełko and Chodak [13] used the Antaris II FT-NIR analyzer (Thermo Fisher Scientific Inc, Waltham, USA) range (1000-2500 nm) and PCR, PLS, artificial neural network (ANN), principal component analysis-artificial neural network (PCA-ANN), PLS-ANN with absorbance, baseline offset (BO), D1, and D1+BO pretreatments in 90 samples, resulting in R^2 values of 0.90, 0.89, 0.91, 0.93, and 0.91, respectively. Reda et al. [14] used the NIR portable spectrometer (1100-2500 nm) and PLS with absorbance pretreatment in 400 samples, resulting in an R^2 value of 0.80 and RPD of 2.77. They also employed back propagation neural network (BPNN), backward variable elimination-back propagation neural network (BVE-BPNN), and ensemble learning modeling (ELM) with multi scatter correction (MSC) pretreatment, resulting in R^2 values of 0.93, 0.90, and 0.94, while RPD values are 3.84, 3.03, and 4.91, respectively.

Ng et al. [16] used the NeoSpectra Module SWS62221 (Si-Ware Systems, Cairo, Egypt) range (1300-2600 nm) and Cubist Model with Savitzky-Golay (SG) second order polynomial and standard normal variate (SNV) pretreatments in 1601 samples, resulting in R^2 values of 0.52. Several studies have successfully predicted soil nitrogen content using different types of NIR equipment and multivariate methods, achieving R^2 values ranging from 0.75 to 0.94.

Munawar et al. [12] used a benchtop NIR instrument Thermo Nicolet Antaris II (Thermo Fisher Scientific Inc, Waltham, USA) for phosphorus and applied PCR and PLS, achieving R^2 values of 0.93 and 0.99, respectively, with RPD values of 3.86 and 5.41. The NeoSpectra Module SWS62221 (Si-Ware Systems, Cairo, Egypt) was used with a Cubist model, SG and second order polynomial, and SNV pretreatment, but the R^2 value was merely 0.47. It is noteworthy that phosphorus is considered one of the most challenging nutrients to predict using NIR spectroscopy, as it presents in soil at low concentrations and is highly reactive with other soil minerals. Thus, more research is needed to improve the accuracy of phosphorus estimation using NIR spectroscopy [16].

For potassium, Munawar et al. [12] also used a benchtop NIR instrument Thermo Nicolet Antaris II (Thermo Fisher Scientific Inc, Waltham, USA) with PCR and PLS multivariate methods to achieve R^2 values of 0.88 and 0.90, respectively, with RPD values of 2.04 and 2.68. Tang et al. [19] employed the ASD AgriSpec spectrometer (Malvern Panalytical Inc. Boulder, Colorado, USA) range (350-2500 nm) with a Cubist model, SG, and second-order polynomial with 392 samples, but the R^2 value merely reached 0.36. Tang et al. [19] also tried using other spectrometers, such as the Malvern Panalytical Inc. Boulder (Spectral Evolution Inc. Lawrence, MA, USA) range (350-3500 nm), the NeoSpectra module SWS62221 (Si-Ware Systems, Cairo) range (1250-2500 nm), and the NIRVascan ASP-NIR-350M-Reflect (Allied Scientific Pro., Quebec, Canada) range (900-1700 nm), but the R^2 values were all lower than 0.40.

Most of these studies use various mathematical models such as PLS, artificial neural networks, and machine learning, where the wavelengths used have a wide range between 400 to 2498 nm [20], 900-1700 nm [21], 350-2500 nm [22], and 350-2500 nm [23], for instance. MLR is rarely used in NIR spectroscopic prediction models, whereas MLR has some virtues. One of the typical virtues is providing simple interpretation where regression coefficients in MLR models represent the relative contribution of each independent variable to the dependent variable [24]. MLR is a relatively computationally simple model compared to more complex models like neural networks or deep learning, enhancing the efficiency to implement in limited computational resources. Therefore, in this research, the MLR method was assessed to estimate macronutrients based on NIR spectroscopy and compared with commonly used models such as PLS and PCR.

2. Methodology

The implementation of the research commences with soil sampling, followed by the collection of NIR data in the form of reflectance spectra and conventional macro-nutrient soil data. Subsequently, data pretreatment was conducted on the reflectance spectra. The following steps involve the development of MLR, PLS, and PCR models through calibration between NIR spectra and conventional soil macro-nutrient data. Once the models were constructed, performance testing was carried out to verify the effectiveness and reliability of the developed models in soil analysis.

2.1. Soil sample collection

Table 1 Soil type and parent material at the sampling location in West Java Province

Districts	Sub-districts	Number of samples	Dominant soil type	Parent material
Bogor	Jasinga	7	Gleisol district, cambisol gleik	Clay deposits
	Ciampea	6	Regosol district, latosol haplik	Andesit, basalt
	Leuwiliang	7	Gleisol distrik, cambisol distrik	Clay deposits
	Tenjolaya	4	District regosol, latosol haplik	Andesite, basalt
	Cihoe	6	Latosol haplik, podsolik haplik	Clay deposits
	Ciseeng	7	Gleisol district, cambisol district	Clay deposits
	Rumpin	7	Latosol haplik, podsolik haplik	Clay deposits
Sukabumi	Cikembar	7	Cambisol district, andosol district	Andesit
	Warungkiara	7	Latosol haplik, district andosol	Andesit, basalt
Indramayu	Patrol	7	Gleisol eutrik	Clay deposits
	Anjatan	7	Gleisol eutrik, gleisol vertik	Clay deposits
Subang	Pamanukan	7	Gleisol eutrik, gleisol fluvik	Clay deposits

Source: Soil type based on the National Soil Classification of Indonesia [25-28].

A total of 79 samples of paddy soil (500 g each) at a depth of 0-20 cm were employed in this study. The samples in this study came from 4 districts in West Java Province, Indonesia. When the soil samples were adopted, the weather varied between sunny and cloudy. Soil sampling was accomplished by randomized purposive sampling based on P and K's soil nutrient availability map [25-28]. The soil sample was situated in a sealed plastic bag and brought to the laboratory. The type of soil

and soil parent material at the sampling locations are presented in Table 1. The sampling locations have dominant soil types and parent material. In general, the dominant soil types at the sampling location were Gleisol, Cambisol, Regosol, Latosol, Podzolic, and Andosol (based on The National Soil Classification of Indonesia), and the main material types consisted of clay, andesite, and basalt deposits.

2.2. Retrieval and pretreatment of NIR spectra data

The samples were dried (average moisture content 8.31%, standard deviation 1.14), cleaned of foreign material, and then sieved with a size of 0.75 mm. After sieving, the soil sample was placed in a plastic bag, and then it was mixed and shaken for 30 seconds to uniform the nutrient distribution in the soil sample. Subsequently, the soil sample was inserted into the petri dish 9 cm in diameter and a height of 2 cm. Before collecting the NIR spectra, the soil in the petri dish was compacted by tapping until the soil surface level did not decrease any further to ensure the soil density in each petri dish was the same. Next, the soil was leveled using a glass spatula to level the soil surface in a petri dish. Eventually, the reflectance data was taken using the NIR Spectrometer NIRflex N500 Buchi in the 1000-2500 nm wavelength range with 1501 wavelength data. Reflectance data collection for each sample was carried out in 3 replications by rotating the petri dish position by 1/3 turn to obtain 237 reflectance data. The reflectance data were transformed into absorbance ($A = \log(1/R)$). Moreover, the data were pretreated in the form of the D1, second derivative (D2), normalized, and SNV using NIRCal software with the following formula:

D1 by Savitzky-Golay (dg1)

$$f'(X_i) = \frac{-86f(X_{i+4}) + 142f(X_{i+3}) + 193f(X_{i+2}) + 126f(X_{i+1}) - 126f(X_{i-1}) - 193f(X_{i-2}) - 142f(X_{i-3}) + 86f(X_{i-4})}{188} \quad (1)$$

D2 by Savitzky-Golay (dg2)

$$f''(X_i) = \frac{28f(X_{i+4}) + 7f(X_{i+3}) - 8f(X_{i+2}) - 17f(X_{i+1}) - 20f(X_i) - 17f(X_{i-1}) - 8f(X_{i-2}) + 7f(X_{i-3}) + 28f(X_{i-4})}{462} \quad (2)$$

Standard normal variate (SNV)

$$X_{snv} = \frac{X - \text{mean}(X)}{\text{St.dev}(X)} \quad (3)$$

Normalization 0-1 (norm)

$$X_{norm} = \frac{X_i - \min(X)}{\max(X) - \min(X)} \quad (4)$$

2.3. Conventional soil nutrient data collection

Soil samples, for which spectral data had been collected, were analyzed for their macronutrient content using conventional methods. Total nitrogen was determined using the Kjeldahl method, where nitrogen compounds were oxidized in a concentrated sulfuric acid environment with a selenium mixture catalyst, forming $(\text{NH}_4)_2\text{SO}_4$. The subsequent ammonium content in the extract was determined using spectrophotometry with the indophenol blue color generator. Total phosphorus and potassium in the soil were extracted using a wet digestion method with a mixture of concentrated HNO_3 and HClO_4 . The total phosphorus in the extract was measured using a spectrophotometer, while the total potassium in the extract was measured using atomic absorption spectrophotometry. The water content was determined by the gravimetric method.

2.4. Data calibration

Data calibration aims to adjust the NIR spectrum response to the values of macronutrient measurements, such as total nitrogen, total phosphorus, and total potassium, to provide accurate measurement results. This process involves measuring the

NIR spectrum response from samples with varying concentrations of macronutrients, and this data is evaluated to create mathematical models such as MLR, PLS, and PCR,

(1) Multiple linear regression (MLR)

NIR spectral data was calibrated with conventional nitrogen, phosphorus, and potassium (NPK) value reading data and used the MLR model with the step-forward method in IBM SPSS Statistics software. 150 spectra data were used as calibration data, and 87 spectra were used as validation data. In the step-forward method, the variable wavelength selection was carried out in stages where the wavelengths were entered sequentially into the model. The first wavelength variable in the equation contains the most significant partial correlation to the N, P, and K nutrient content variables. The following wavelength variable is re-selected, which gives the most significant total correlation to the NPK nutrient content to be included in the equation model. The selection and addition of the wavelength variable will be stopped once the total correlation of the MLR equation model has reached > 0.90.

$$Y = a + b_1(X_1) + b_k(X_k) \quad (5)$$

where Y is the estimated value of total nutrients (nitrogen, phosphorus, and potassium), a is the constant, b_1 is the first wavelength variable coefficient, b_k is the variable coefficient of the nth wavelength, X_1 is the 1st wavelength NIR spectral intensity value, and X_k is the nth wavelength NIR spectral intensity value.

In the calibration process, the pretreatment of spectral data was considered as input data (X) for constructing an MLR model using the following models:

- a. Input data is raw data as R
- b. Input data is the D1 of the reflectance spectra (Rdg1)
- c. Input data is the D2 of the reflectance spectra (Rdg2)
- d. Input data is SNV of reflectance spectra (Rsnv)
- e. Input data is a normalization of reflectance spectra (Rnorm)
- f. Input data is absorbance (A), which is the transformation of $\log(1/\text{reflectance})$
- g. Input data D1 of the absorbance spectra (Adg1)
- h. Input data is the D2 of the absorbance spectra (Adg2)
- i. Input data is SNV of absorbance spectra (Asnv)
- j. Input data was normalized from absorbance spectra (Anorm)

(2) Partial least squares (PLS) and principal component regression (PCR)

The performance of MLR in predicting soil nutrients will be compared with other commonly used methods in NIR spectroscopy, i.e., PLS and PCR. Besides, PLS and PCR are multivariate calibration techniques. This algorithm uses both spectral data and reference matrices in multivariate regression. It is achieved by projecting spectral data into a reduced-dimensional space. They were initially designed for high-dimensional and collinear multivariate scenarios. Cross-validation was employed in both PLS and PCR to determine the best number of terms latent variables (LV) for PLS and principal component (PC) for PCR in the calibration model to prevent overfitting [13]. Cross-validation method using 237 random samples with 20 segments. Calibration process using The Unscrambler software.

(3) Model performance assessment

The best model calibrations were selected based on high correlation coefficients r, determination coefficients R^2 , and the ratio of performance deviation (RPD) values while minimizing root mean square error (RMSE) values. The stability of the model was determined by the RPD

$$RPD = \frac{St.dev}{RMSE} \quad (6)$$

St.dev is the standard deviation of reference, and RMSE is the root mean square error between prediction and measurement. $RPD > 2.0$ indicates accurate predictions, $RPD 1.4-2.0$ indicates less accurate predictions and $RPD < 1.4$ indicates unreliable predictions, and the model cannot be used to predict soil properties [29].

3. Results and Discussion

Observations exhibit high variation coefficients in the diversity of macronutrient content in the soil. This variability ensures that the model built in the calibration process can represent nutrient content over a wide range to generate a high generalization value for the model. Specifically, observations of soil content at sampling locations and calibration models to predict macronutrient content using NIR waves are explicated further.

3.1. Soil conditions at the soil sampling location

The heterogeneity is desperately needed in developing soil nutrient estimation models to elicit the obtained equations having a wide and varied range of uses for soil types, parent material, and soil nutrient status as found at the sampling locations. The total nitrogen, phosphorus, and potassium of the soil samples are described in Table 2.

Table 2 Descriptive statistics of soil nutrient content in 79 samples

Soil nutrient content	Mean (g/100 g)	Minimum (g/100 g)	Maximum (g/100 g)	SD (g/100 g)	CV (%)
Total nitrogen	0.16	0.09	0.24	0.04	22.93
Total phosphorus	0.07	0.01	0.36	0.07	96.34
Total potassium	0.11	0.01	0.39	0.12	112.41

SD is the standard deviation, and CV is the coefficient variation.

Table 2 is drawn regarding the soil nutrient content in 79 samples. First, total nitrogen exhibits a relatively low variation, with a CV of 22.93%. The range between maximum and minimum is about 0.15 g/100 g. Second, conversely, total phosphorus evinces a larger variation with a CV of 96.34%, signifying significant differences in phosphorus content among the soil samples with a range between maximum and minimum of about 0.35 g/100 g. Eventually, the total potassium content demonstrates a high level of variation, reflected in a CV of 112.41%, with a range between a maximum and a minimum of about 0.38 g/100 g. Heterogeneity in soil nutrient content values with a wide range of values is needed in building a soil nutrient estimation model using NIR waves so that the resulting model can be used in a wide range of soil nutrients according to calibration data.

3.2. Data calibration and validation

The calibration and validation process aims to construct a model for estimating the total content of soil macronutrients based on the NIR wavelength spectrum. The results of the calibration models MLR, PLS, and PCR for estimating total nitrogen, total phosphorus, and total potassium using NIR wavelengths will be further elaborated.

(1) Total nitrogen

The selected MLR models for estimating total nitrogen are presented in Table 3. Table 3 illustrates that applying data pretreatments such as SNV, normalization, D1, and SNV of $\log(1/R)$ can enhance the performance of total nitrogen estimation in the soil compared to using only raw reflectance data. Based on the calibration results, the best performance is observed in Model 5, which employs normalization of the reflectance data, indicated by the highest values of r , R^2 , and RPD, followed by the smallest RMSE values of 0.93, 0.86, 2.6, and 0.014 g/100 g, respectively. Normalization helps eliminate scale differences among spectra affecting the MLR model. With normalization, each spectrum was transformed to have a similar standard

deviation, helping to overcome the non-linearity effects caused by large-scale variations among spectra. Additionally, normalization can improve the consistency and stability of the spectrum in MLR analysis, thereby enhancing the model's ability to capture the linear relationship between spectral data and total nitrogen.

Table 3 MLR model calibration and validation result for estimating total nitrogen with several pretreatment spectra data

Model	Spectra pretreatment data	Number of wavelengths	Calibration				Validation	
			r	R ²	RMSE (g/100 g)	RPD	RMSE (g/100 g)	RPD
1	R (raw data)	9	0.91	0.82	0.015	2.4	0.019	2.0
2	Rdg1	5	0.90	0.81	0.016	2.3	0.019	1.9
3	Rdg2	5	0.90	0.81	0.016	2.3	0.021	1.8
4	Rsnv	7	0.92	0.84	0.014	2.5	0.017	2.1
5	Rnorm	6	0.93	0.86	0.014	2.6	0.018	2.1
6	A	12	0.89	0.79	0.017	2.2	0.021	1.7
7	Adg1	7	0.91	0.83	0.015	2.4	0.019	2.0
8	Adg2	5	0.91	0.82	0.015	2.4	0.026	1.4
9	Asnv	5	0.91	0.83	0.015	2.5	0.018	2.1
10	Anorm	7	0.91	0.82	0.015	2.4	0.022	2.1

The best model is marked in bold, r: coefficient correlation, R²: coefficient determination, RMSE: root mean square error, RPD: ratio of performance deviation.

A comparison of the MLR model with PLS and PCR models in estimating total nitrogen is presented in Table 4. MLR demonstrates good performance on the training data with an R² of around 0.86 but experiences a decline in performance on the validation data with an R² of around 0.77. This decline may indicate overfitting or the absence of the model's generalization ability. PLS consistently performs well on calibration and validation data sets, with R² values above 0.80. Hence, PLS might be a favorable choice, with a note to consider the optimality of the number of LV. On the other hand, PCR exhibits the lowest performance, with Rdg1 providing an r-value of 0.87, R² of 0.76, and RMSE of 0.018. PCR in the validation stage shows consistent performance with r, R², and RMSE values relatively similar to those in the calibration stage. The analysis results specify that PLS with Rsnv outperforms the other two models with r of 0.93, R² of 0.87, and RMSE of 0.013 g/100 g. Although MLR with Rnorm performs well with r of 0.93 and R² of 0.86, PLS is significantly better in more accurate predictions. PCR with Rdg1 shows moderate performance with r of 0.87, R² of 0.76, and RMSE of 0.018 g/100 g.

Table 4 Comparison of MLR model with PLS and PCR in predicting total nitrogen

Model	Pretreatment Data			Calibration				Validation			
				r	R ²	RMSE (g/100 g)	RPD	r	R ²	RMSE (g/100 g)	RPD
MLR	Rnorm	WL	6	0.93	0.86	0.014	2.6	0.87	0.77	0.018	2.1
PLS	Rsnv	LV	11	0.93	0.87	0.013	2.8	0.92	0.84	0.014	2.5
PCR	Rdg1	PC	8	0.87	0.76	0.018	2.0	0.85	0.73	0.019	1.9

WL: number of wavelengths, LV: number of latent variables, PC: number of principal components, r: coefficient correlation, R²: coefficient determination, RMSE: root mean square error, RPD: ratio of performance deviation.

RPD is an indicator to understand the reliability of the model in predicting total nitrogen content. PLS with Rsnv manifests the highest RPD values, namely 2.8 in the calibration stage and 2.5 in the validation stage. A relatively high RPD indicates that the PLS model's predictions are accurate and can provide significant predictive value compared to the natural variation in the data. Specifically, the high RPD in PLS may be due to its ability to extract relevant information from independent variables, thus improving prediction accuracy. MLR with Rnorm also shows reasonably good RPD, with values of 2.6 in the calibration and 2.1 in the validation. Although not as optimal as PLS, MLR still provides good prediction accuracy, especially considering its simplicity in interpretation. PCR with Rdg1, while having a decent RPD value, shows lower performance compared to PLS and MLR, with values of 2.0 in the calibration and 1.9 in the validation. This lower performance denotes that PCR may not be able to provide prediction accuracy as good as PLS and MLR.

The research conducted by Munawar et al. [12] deployed the Benchtop NIR instrument Thermo Nicolet Antaris II (Thermo Fisher Scientific Inc, Waltham, USA) at 1000-2500 nm and used PCR and PLS in 40 samples, resulting in R^2 values of 0.85 and 0.87, respectively exhibited relatively similar performance to this study, achieving an R^2 value of 0.86 in the MLR and 0.85 in PCR models, but lower than the PLS model from this study. PLS, through the partial component formation approach, can capture complex patterns in data and improve prediction accuracy. Given such results, PLS may be the primary choice for accurate predictions, while MLR remains relevant for simple model interpretation. Moreover, PCR may be an alternative in significant multicollinearity issues despite the disposition of lower performance. The PLS and PCR models are obtained by projecting spectral data into smaller dimensions, i.e., LV for PLS and PC for PCR. However, in practical terms, it is challenging to implement this for building portable measuring devices. On the other hand, the MLR model enables the direct use of spectral data as input for the portable model to be constructed.

The specified wavelengths in Model 5 correlating highly with total nitrogen, are 1871, 2059, 1873, 1929, 2013, and 2082 nm (Fig. 1), with a correlation coefficient of 0.93 (Table 3). Fig. 1 shows the normalized reflectance spectra from all soil samples in 1800-2100 nm and the contribution of each wavelength to the correlation coefficient that has been achieved. The wavelength of 1871 nm provides the highest contribution to the correlation coefficient of the MLR model (Model 5), amounting to 78.2%, following 8.6% in 1929 nm. Fig. 2 illustrates the calibration and validation results between actual total nitrogen and predicted total nitrogen, yielding calibration determination coefficients (R^2) of 0.86, validation root mean square error (RMSEv) of 0.018 g/100 g, and validation ratio of performance to deviation (RPDv) of 2.1.

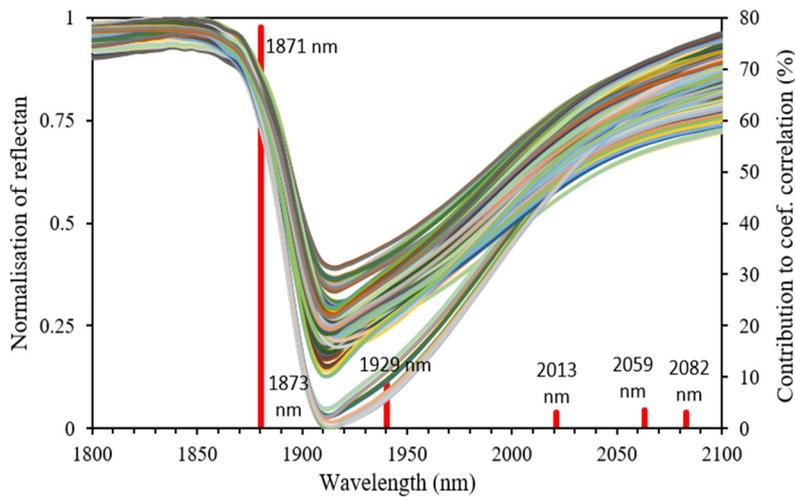


Fig. 1 Wavelength contribution to MLR model (Model 5) for total nitrogen prediction

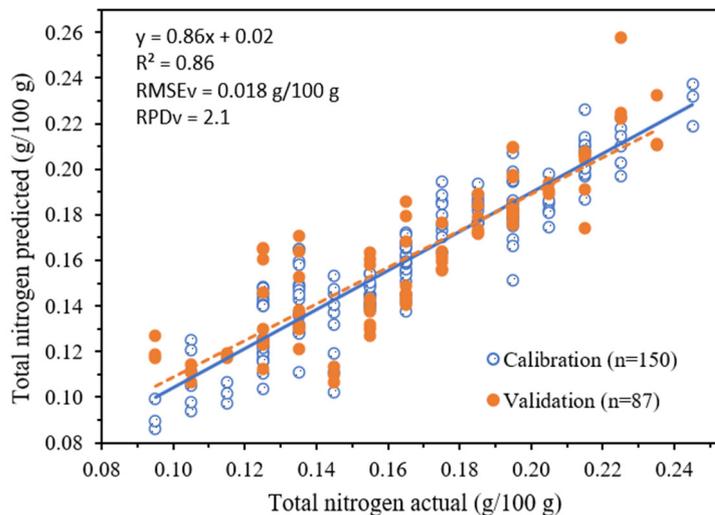


Fig. 2 MLR Model (Model 5) calibration and validation for predicted total nitrogen

(2) Total phosphorus

The selected MLR model for estimating total phosphorus is presented in Table 5. All data pretreatment can improve the performance of phosphorus estimation compared to employing only raw reflectance data. Based on the calibration results, the best performance is attained by model 5 with data pretreatment of normalization of reflectance, indicated by the highest values of r , R^2 , and RPD, followed by the smallest RMSE values of 0.92, 0.85, 2.6, and 0.028 g/100 g, respectively in the calibration data set. Similar to total nitrogen estimation, normalization in phosphorus estimation can also assist in addressing the non-linearity effects caused by large-scale spectra variations. Normalization can improve the consistency and stability of the spectrum in MLR analysis, thereby enhancing the model's ability to capture the linear relationship between spectral data and total phosphorus.

Table 5 MLR model calibration and validation result for estimating total phosphorus with several pretreatment spectra data

Model	Spectra pretreatment data	Number of wavelengths	Calibration				Validation	
			r	R^2	RMSE (g/100 g)	RPD	RMSE (g/100 g)	RPD
1	R (raw data)	7	0.78	0.60	0.045	1.6	0.056	1.2
2	Rdg1	13	0.91	0.82	0.030	2.3	0.056	1.2
3	Rdg2	10	0.90	0.81	0.031	2.3	0.043	1.5
4	Rsnv	10	0.90	0.81	0.031	2.3	0.039	1.7
5	Rnorm	13	0.92	0.85	0.028	2.6	0.036	1.9
6	A	7	0.79	0.63	0.043	1.6	0.050	1.3
7	Adg1	12	0.90	0.82	0.030	2.3	0.052	1.3
8	Adg2	11	0.91	0.82	0.030	2.4	0.055	1.2
9	Asnv	12	0.91	0.82	0.030	2.3	0.036	1.8
10	Anorm	10	0.91	0.83	0.030	2.4	0.034	1.9

The best model is marked in bold, r : coefficient correlation, R^2 : coefficient determination, RMSE: root mean square error, RPD: ratio of performance deviation.

A comparison of the MLR model with PLS and PCR models in estimating total phosphorus is presented in Table 6. MLR used normalized data (Rnorm) as a preprocessing method, resulting in a strong correlation between prediction and observation in the calibration ($r = 0.92$). However, in the validation, this model exhibited a significant decrease in R^2 and RPD, reflecting the potential for overfitting or the absence of generalization to new data. Meanwhile, PLS demonstrated consistent performance improvement using the same pretreatment data as Rnorm in MLR. With a higher calibration correlation ($r = 0.93$) and an increased R^2 of 0.86, PLS remained consistent and even enhanced predictions in the validation phase with a high R^2 (0.83) and a relatively high RPD (2.5). These results indicate the ability of PLS to provide stable and accurate predictions. Employing normalization pretreatment data (Anorm), PCR showed calibration results similar to MLR. However, this model experienced a slight performance decrease in the validation with an RPD of 2.4. Consequently, the use of principal components as predictors may be less optimal in terms of generalization to new data.

Table 6 Comparison MLR model with PLS and PCR in predicting total phosphorus

Model	Pretreatment data			Calibration				Validation			
				r	R^2	RMSE (g/100 g)	RPD	r	R^2	RMSE (g/100 g)	RPD
MLR	Rnorm	WL	13	0.92	0.85	0.028	2.6	0.85	0.72	0.036	1.9
PLS	Rnorm	LV	11	0.93	0.86	0.026	2.6	0.91	0.83	0.028	2.5
PCR	Anorm	PC	12	0.92	0.85	0.027	2.6	0.91	0.82	0.029	2.4

WL: number of wavelengths, LV: number of latent variables, PC: number of principal components, r : coefficient correlation, R^2 : coefficient determination, RMSE: root mean square error, RPD: ratio of performance deviation.

PLS is the most effective model in predicting total phosphorus content, especially in the calibration due to the production of the highest r value (0.93), high R^2 (0.86), and low RMSE (0.026 g/100 g). Additionally, the high RPD (2.6) indicates the superiority of PLS in providing accurate and consistent predictions. Despite the decent results MLR and PCR yielded, PLS remains superior in most evaluation parameters. MLR and PCR perform similarly in the calibration, with r values of 0.92 and

R² of 0.85. PLS maintains higher performance in the validation, while MLR and PCR experience a decline. These results indicate that PLS is a better choice for addressing the challenges of predicting total phosphorus content, especially in dealing with the complexity of relationships between variables. PLS's ability to handle multicollinearity and produce more stable predictions signifies the superiority of MLR and PCR models in the context of this research.

Despite the evident superiority of PLS in predicting total phosphorus content, particularly during the calibration, it is noteworthy that MLR still yields commendable accuracy. Both MLR and PCR demonstrate comparable performance in the calibration, with r values of 0.92 and R² of 0.85. These metrics indicate that MLR is still a viable option concerning accuracy. Moreover, it is crucial to acknowledge the practicality of MLR, especially its simplicity of use. MLR's simplicity enables users to implement straightforwardly, being regarded as a convenient choice for applications where simplicity and quick application are prioritized. The ability to easily select specific wavelengths enhances the adaptability of MLR to portable devices, which is a factor that might be crucial in specific practical scenarios. While PLS excels in addressing the complexity of relationships between variables and handling multicollinearity, MLR's accuracy and simplicity rationalize itself in situations where computational simplicity and interpretability are paramount.

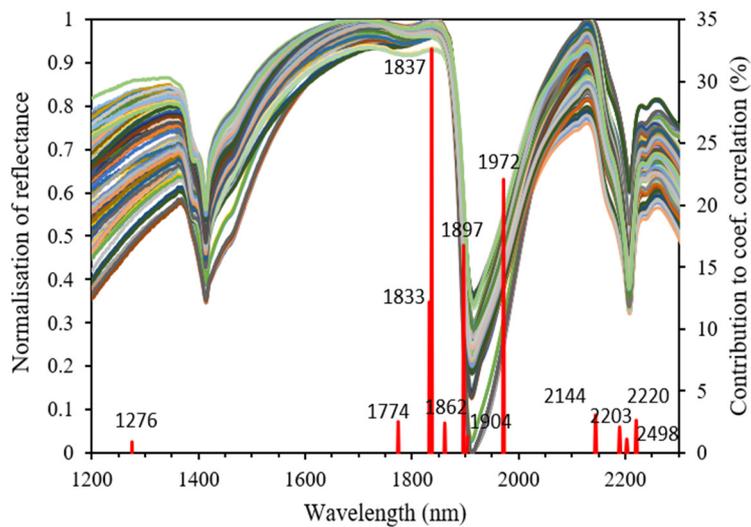


Fig. 3 Wavelength contribution to MLR model (Model 5) for total phosphorus prediction

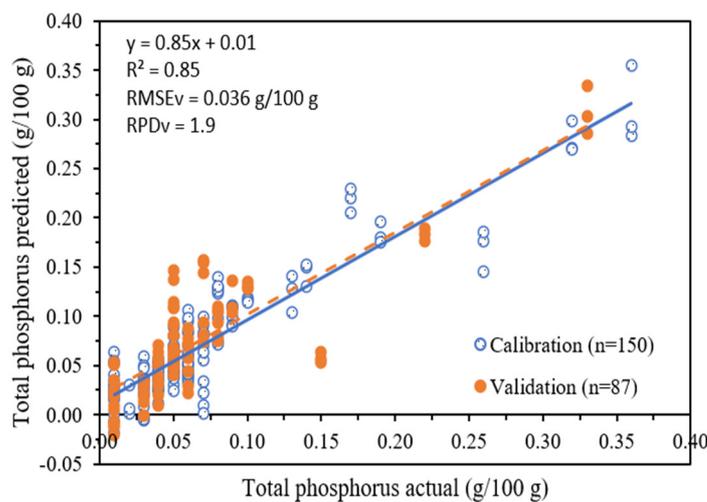


Fig. 4 MLR model (Model 5) calibration and validation for predicted total phosphorus

The selected wavelengths in Model 5, which have a high correlation with total phosphorus, are 1837, 1897, 1972, 1833, 1904, 1862, 2220, 2189, 2498, 1276, 1774, 2203, and 2144 nm (Fig 3), with a correlation coefficient of 0.92 (Table 5). Fig. 3 shows the normalized reflectance from all soil samples in 1200–2200 nm and the contribution of each wavelength to the

correlation coefficient having been achieved. The wavelength of 1837 nm provides the highest contribution to the correlation coefficient of the MLR model (Model 5), amounting to 32.6%, following 22.1% in 1972 nm and 16.7% in 1897 nm. Fig. 4 illustrates the calibration and validation results between actual total phosphorus and predicted total phosphorus, yielding calibration determination coefficients (R^2) of 0.85, RMSEv of 0.036 g/100 g, and RPDv of 1.9.

(3) Total potassium

The selected MLR model for estimating total potassium is depicted in Table 7. All pretreatment methods applied can improve the performance of MLR in predicting total potassium compared to using raw data (reflectance) alone. Based on calibration results, the D1 treatment applied to the raw data (reflectance) yielded the highest model performance regarding r , R^2 , and RPD, with values of 0.97, 0.94, and 4.2 g/100 g, respectively.

NIR spectra often contain nonlinear variations that can affect the relationship between input and output variables. The D1 can help “flatten” nonlinear changes in the spectrum, thus creating a more linear relationship between NIR spectra and total potassium. The D1 in potassium estimation using MLR can enhance the model’s ability to capture linear patterns in the NIR spectrum. The effect of D2 is also similar to that of D1 in reducing nonlinear effects.

However, in this study, the performance of the D2 is lower than it is in the D1. This lower performance could be attributed to second-order derivatives being more sensitive to noise in the data. Furthermore, noise can be accentuated during the calculation of the D2, potentially leading to overfitting, especially if the model captures noise in the training data rather than the true underlying patterns. Additionally, the model may struggle to generalize to new data. Second-order derivatives, while capable of highlighting peaks and valleys, might also result in the loss of certain information in the original spectra.

Table 7 MLR model calibration and validation result for estimating total potassium with several pretreatment spectra data

Model	Spectra pretreatment data	Number of wavelengths	Calibration				Validation	
			r	R^2	RMSE (g/100 g)	RPD	RMSE (g/100 g)	RPD
1	R (raw data)	6	0.90	0.81	0.051	2.3	0.057	2.2
2	Rdg1	4	0.97	0.94	0.029	4.2	0.036	3.6
3	Rdg2	4	0.97	0.93	0.031	3.8	0.037	3.4
4	Rsnv	4	0.95	0.91	0.037	3.2	0.034	3.7
5	Rnorm	4	0.94	0.88	0.042	2.9	0.055	2.3
6	A	4	0.90	0.82	0.051	2.3	0.055	2.3
7	Adg1	4	0.96	0.93	0.032	3.7	0.035	3.8
8	Adg2	4	0.96	0.92	0.034	3.5	0.050	2.6
9	Asnv	4	0.95	0.91	0.036	3.3	0.034	3.6
10	Anorm	4	0.93	0.87	0.044	2.7	0.063	2.0

The best model is written in bold, r : coefficient correlation, R^2 : coefficient determination, RMSE: root mean square error, RPD: ratio of performance deviation.

A comparison of the MLR model with PLS and PCR models in estimating total phosphorus is depicted in Table 8. The MLR model, with the D1, demonstrates excellent performance in the calibration with an R^2 value of 0.94 and a r of 0.97. Besides, the RMSE value of 0.029 g/100 g and RPD value of 4.2 indicate significant prediction accuracy. However, in the validation, this model experiences a slight decrease in performance with R^2 at 0.92 and r at 0.96. Meanwhile, with absorbance ($\log 1/R$), the PLS model exhibits higher performance in calibration and validation. This model achieves high accuracy with R^2 values of 0.95 and 0.97 and r at 0.98 in the calibration. The low RMSE value (0.026 g/100 g) and high RPD (4.7) affirm the reliability of the PLS model’s predictions. On the other hand, the PCR model, with the normalization of absorbance, demonstrates performance comparable to MLR. With R^2 at 0.94 in the calibration and 0.93 in the validation, this model provides good predictions. Although slightly lower than PLS, PCR yields a low RMSE value (0.029 g/100 g) and a sufficiently high RPD (3.9).

Table 8 Comparison MLR model with PLS and PCR in predicting total potassium

Model	Pretreatment data	-	Calibration					Validation			
			r	R ²	RMSE (g/100 g)	RPD	r	R ²	RMSE (g/100 g)	RPD	
MLR	Rdg1	WL	0.97	0.94	0.029	4.2	0.96	0.92	0.036	3.6	1.9
PLS	A	LV	0.98	0.95	0.026	4.7	0.97	0.95	0.028	4.3	2.5
PCR	Anorm	PC	0.97	0.94	0.029	4.1	0.97	0.93	0.031	3.9	2.4

WL: number of wavelengths, LV: number of latent variables, PC: number of principal components, r: coefficient correlation, R²: coefficient determination, RMSE: root mean square error, RPD: ratio of performance deviation.

In the calibration, the PLS model r value of 0.98 is higher than MLR (0.97) and PCR (0.97). This result denotes that PLS can better elaborate the training data’s relationship between input variables and total potassium. Regarding the R² calibration, PLS also excels with a score of 0.95, while MLR and PCR have values of 0.94 and 0.941, respectively. Therefore, PLS can better capture the variation in the calibration data. Furthermore, the Root Mean Square Error (RMSE) during calibration, PLS, has the lowest RMSE value (0.026 g/100 g), and therein lies the lower prediction error of this model compared to MLR (0.029 g/100 g) and PCR (0.029 g/100 g). In the validation stage, the results are also noteworthy. PLS maintains good performance with an R² value of 0.95, whereas MLR and PCR show a slight decrease in performance with values of 0.92 and 0.933, respectively.

Regarding the RPD values in validation, both PLS (4.3) and MLR (3.6) experience a decline from the calibration RPD values, but PLS remains superior. Overall, the PLS model stands out in its predictive ability, both in the calibration and validation, with high correlation, good R², and low prediction error. While MLR and PCR also provide good results, PLS might be more optimal for predicting total potassium thereon.

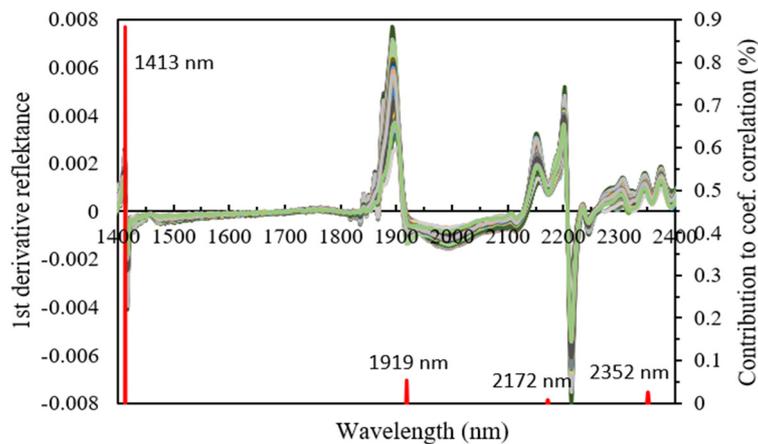


Fig. 5 Wavelength contribution to MLR model (Model 2) for total potassium prediction from D1 reflectance spectra

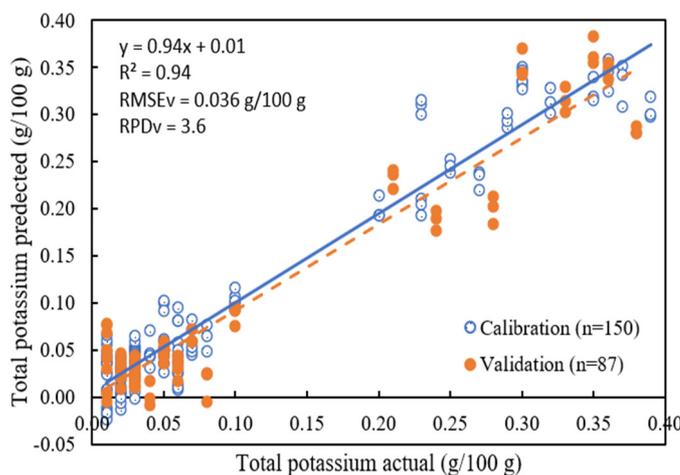


Fig. 6 MLR model (Model 2) calibration and validation for predicted total potassium

The selected wavelengths in Model 2, which correlate highly with total potassium, are 1413, 1919, 2352, and 2172 nm (Fig. 5), with a correlation coefficient of 0.97 (Table 7). Fig. 5 describes the D1 of reflectance from all soil samples in 1400–2400 nm and the contribution of each wavelength to the correlation coefficient that has been achieved. The wavelength of 1413 nm provides the highest contribution to the correlation coefficient of the MLR model (Model 5), amounting to 88.0%. Fig. 6 illustrates the calibration and validation results between actual total potassium and predicted total potassium, yielding calibration R^2 of 0.94, RMSE_v of 0.036 g/100 g, and RPD_v of 3.6.

4. Conclusion

Estimating nitrogen, phosphorus, and potassium content in paddy soil using NIR at 1000-2500 nm and the MLR model evinced lower performance than PLS for all nutrients but better than PCR for nitrogen, and it was below PCR for phosphorus and potassium. However, MLR is well-performed owing to achieving a high RPD > 2.0 in the best model for each nutrient. The study using simple MLR induces different pretreatment data and the number of wavelengths in soil nitrogen, phosphorus, and potassium.

- (1) Nitrogen: the highest performance was achieved by normalization of reflectance in pretreatment data with six wavelengths as variables, giving results with r , R^2 , RMSE, and RPD values of 0.93, 0.86, 0.014 g/100 g, and 2.6, respectively.
- (2) Phosphorus: the highest performance was achieved by normalization of reflectance in pretreatment data with thirteen wavelengths as variables, giving results with r , R^2 , RMSE, and RPD values of about 0.92, 0.85, 0.028 g/100 g, and 2.6, respectively.
- (3) Potassium: the highest performance achieved by the D1 of reflectance in pretreatment data with four wavelengths as a variable, giving results with r , R^2 , RMSE, and RPD values is about 0.97, 0.94, 0.029 g/100 g, and 4.6, respectively.

This research shows that all selected models provide statistically excellent correlation values ($r > 0.80$) for all soil nutrients. Likewise, the selected model provides an accurate level of prediction ($RPD \geq 2.0$) for all soil nutrient estimation models using NIR spectra, except the validation of the PCR model for nitrogen estimation and the MLR model for phosphorus estimation indicating less accurate prediction ($RPD < 2.0$).

Acknowledgments

This research was funded by the Ministry of Education and Culture of the Republic of Indonesia through a doctoral dissertation research program (Contract No. 082/E5/PG.02.00.PT/2022, year 2022).

Conflicts of Interest

The authors declare no conflict of interest.

References

- [1] M. A. Munaf and A. M. Mouazen, "Development of a Soil Fertility Index Using On-Line Vis-NIR Spectroscopy," *Computers and Electronics in Agriculture*, vol. 188, article no. 106341, September 2021.
- [2] J. M. Johnson, E. Vandamme, K. Senthilkumar, A. Sila, K. D. Shepherd, and K. Saito, "Near-Infrared, Mid-Infrared or Combined Diffuse Reflectance Spectroscopy for Assessing Soil Fertility in Rice Fields in Sub-Saharan Africa," *Geoderma*, vol. 354, article no. 113840, November 2019.
- [3] Devianti, Sufardi, R. Bulan, and A. Sitorus, "Vis-NIR Spectra Combined with Machine Learning for Predicting Soil Nutrients in Cropland from Aceh Province, Indonesia," *Case Studies in Chemical and Environmental Engineering*, vol. 6, article no. 100268, December 2022.

- [4] S. Alomar, S. A. Mireei, A. Hemmat, A. A. Masoumi, and H. Khademi, "Comparison of Vis/SWNIR and NIR Spectrometers Combined with Different Multivariate Techniques for Estimating Soil Fertility Parameters of Calcareous Topsoil in an Arid Climate," *Biosystems Engineering*, vol. 201, pp. 50-66, January 2021.
- [5] M. Pusch, A. Samuel-Rosa, P. S. G. Magalhães, and L. R. do Amaral, "Covariates in Sample Planning Optimization for Digital Soil Fertility Mapping in Agricultural Areas," *Geoderma*, vol. 429, article no. 116252, January 2023.
- [6] S. Srisomkiew, M. Kawahigashi, P. Limtong, and O. Yuttum, "Digital Soil Assessment of Soil Fertility for Thai Jasmine Rice in the Thung Kula Ronghai Region, Thailand," *Geoderma*, vol. 409, article no. 115597, March 2022.
- [7] Y. Wang, H. Huang, and X. Chen, "Predicting Organic Matter Content, Total Nitrogen and pH Value of Lime Concretion Black Soil Based on Visible and Near Infrared Spectroscopy," *Eurasian Soil Science*, vol. 54, no. 11, pp. 1681-1688, November 2021.
- [8] G. Zheng, A. Wang, C. Zhao, M. Xu, C. Jiao, and R. Zeng, "Evolution of Paddy Soil Fertility in a Millennium Chronosequence Based on Imaging Spectroscopy," *Geoderma*, vol. 429, article no. 116258, January 2023.
- [9] B. G. Barthès, E. Kouakoua, M. Clairotte, J. Lallemant, L. Chapuis-Lardy, M. Rabenarivo, et al., "Performance Comparison Between a Miniaturized and a Conventional Near Infrared Reflectance (NIR) Spectrometer for Characterizing Soil Carbon and Nitrogen," *Geoderma*, vol. 338, pp. 422-429, March 2019.
- [10] F. B. de Santana and K. Daly, "A Comparative Study of MIR and NIR Spectral Models Using Ball-Milled and Sieved Soil for the Prediction of a Range Soil Physical and Chemical Parameters," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 279, article no. 121441, October 2022.
- [11] U. J. dos Santos, J. A. de Melo Dematte, R. S. C. Menezes, A. C. Dotto, C. C. B. Guimarães, B. J. R. Alves, et al., "Predicting Carbon and Nitrogen by Visible Near-Infrared (Vis-NIR) and Mid-Infrared (MIR) Spectroscopy in Soils of Northeast Brazil," *Geoderma Regional*, vol. 23, article no. e00333, December 2020.
- [12] A. A. Munawar, Y. Yunus, Devianti, and P. Satriyo, "Calibration Models Database of Near Infrared Spectroscopy to Predict Agricultural Soil Fertility Properties," *Data in Brief*, vol. 30, article no. 105469, June 2020.
- [13] A. Pudełko and M. Chodak, "Estimation of Total Nitrogen and Organic Carbon Contents in Mine Soils with NIR Reflectance Spectroscopy and Various Chemometric Methods," *Geoderma*, vol. 368, article no. 114306, June 2020.
- [14] R. Reda, T. Saffaj, B. Ilham, O. Saidi, K. Issam, L. Brahim, et al., "A Comparative Study Between a New Method and Other Machine Learning Algorithms for Soil Organic Carbon and Total Nitrogen Prediction Using Near Infrared Spectroscopy," *Chemometrics and Intelligent Laboratory Systems*, vol. 195, article no. 103873, December 2019.
- [15] P. Zhou, Y. Zhang, W. Yang, M. Li, Z. Liu, and X. Liu, "Development and Performance Test of an In-Situ Soil Total Nitrogen-Soil Moisture Detector Based on Near-Infrared Spectroscopy," *Computers and Electronics in Agriculture*, vol. 160, pp. 51-58, May 2019.
- [16] W. Ng, Husnain, L. Anggria, A. F. Siregar, W. Hartatik, Y. Sulaeman, et al., "Developing a Soil Spectral Library Using a Low-Cost NIR Spectrometer for Precision Fertilization in Indonesia," *Geoderma Regional*, vol. 22, article no. e00319, September 2020.
- [17] R. Reda, T. Saffaj, S. E. Itqiq, I. Bouzida, O. Saidi, K. Yaakoubi, et al., "Predicting Soil Phosphorus and Studying the Effect of Texture on the Prediction Accuracy Using Machine Learning Combined with Near-Infrared Spectroscopy," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 242, article no. 118736, December 2020.
- [18] H. T. Cai, J. Liu, J. Y. Chen, K. H. Zhou, J. Pi, and L. R. Xia, "Soil Nutrient Information Extraction Model Based on Transfer Learning and Near Infrared Spectroscopy," *Alexandria Engineering Journal*, vol. 60, no. 3, pp. 2741-2746, June 2021.
- [19] Y. Tang, E. Jones, and B. Minasny, "Evaluating Low-Cost Portable Near Infrared Sensors for Rapid Analysis of Soils from South Eastern Australia," *Geoderma Regional*, vol. 20, article no. e00240, March 2020.
- [20] P. Berzaghi, J. H. Cherney, and M. D. Casler, "Prediction Performance of Portable Near Infrared Reflectance Instruments Using Preprocessed Dried, Ground Forage Samples," *Computers and Electronics in Agriculture*, vol. 182, article no. 106013, March 2021.
- [21] Z. Li, J. Song, Y. Ma, Y. Yu, X. He, Y. Guo, et al., "Identification of Aged-Rice Adulteration Based on Near-Infrared Spectroscopy Combined with Partial Least Squares Regression and Characteristic Wavelength Variables," *Food Chemistry: X*, vol. 17, article no. 100539, March 2023.
- [22] Q. Wang, H. Zhang, F. Li, C. Gu, Y. Qiao, and S. Huang, "Assessment of Calibration Methods for Nitrogen Estimation in Wet and Dry Soil Samples with Different Wavelength Ranges Using Near-Infrared Spectroscopy," *Computers and Electronics in Agriculture*, vol. 186, article no. 106181, July 2021.

- [23] A. Cambou, B. G. Barthès, P. Moulin, L. Chauvin, E. Faye, D. Masse, et al., "Prediction of Soil Carbon and Nitrogen Contents Using Visible and Near Infrared Diffuse Reflectance Spectroscopy in Varying Salt-Affected Soils in Sine Saloum (Senegal)," *Catena*, vol. 212, article no. 106075, May 2022.
- [24] Z. Patonai, R. Kicsiny, and G. Géczi, "Multiple Linear Regression Based Model for the Indoor Temperature of Mobile Containers," *Heliyon*, vol. 8, no. 12, article no. e12098, December 2022.
- [25] BBSDLP, Semi-Detailed Soil Map of Bogor Regency, West Java Province. Bogor, Indonesia: Center for Agricultural Land Resources, Agency for Agricultural Research and Development, 2016.
- [26] BBSDLP, Semi-Detailed Soil Map of Sukabumi Regency, West Java Province. Bogor, Indonesia: Center for Agricultural Land Resources, Agency for Agricultural Research and Development, 2016.
- [27] BBSDLP, Semi-Detailed Soil Map of Indramayu Regency, West Java Province. Bogor, Indonesia: Center for Agricultural Land Resources, Agency for Agricultural Research and Development, 2017.
- [28] BBSDLP, Semi-Detailed Soil Map of Subang Regency, West Java Province. Bogor, Indonesia: Center for Agricultural Land Resources, Agency for Agricultural Research and Development, 2017.
- [29] B. Miloš, A. Bensa, and B. Japundžić-Palenkić, "Evaluation of Vis-NIR Preprocessing Combined with PLS Regression for Estimation Soil Organic Carbon, Cation Exchange Capacity and Clay from Eastern Croatia," *Geoderma Regional*, vol. 30, article no. e00558, September 2022.



Copyright© by the authors. Licensee TAETI, Taiwan. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC) license (<https://creativecommons.org/licenses/by-nc/4.0/>).