JTEP Jurnal Keteknikan Pertanian,
Vol. 12 No. 2, p 242-258, 2024
P-ISSN 2407-0475 E-ISSN 2338-8439

Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.012.2.242-258

Macro-Nutrient Prediction of Paddy Field Soil Using Artificial Neural Network and NIR Spectroscopy

Jonni Firdaus[1,4*], Usman Ahmad[2], I Wayan Budiastra[2,3], I Dewa Made Subrata[2]

[1]Agricultural Engineering Science Study Program, IPB University, Jalan Raya Darmaga, School of Graduate Program Building, IPB, Kampus IPB Darmaga Bogor 16680, West Java, Indonesia.

[2]Department of Mechanical and Biosystem Engineering, Faculty of Agricultural Engineering and Technology, IPB University, Jalan Lingkar Akademik, Kampus IPB Dramaga, Babakan, Kec. Dramaga, Kabupaten Bogor, Jawa Barat 16680, Indonesia

[3]Centre for Research on Engineering Application in Tropical Agriculture, IPB University, Jalan Lingkar Akademik, Kampus IPB Dramaga, Babakan, Kec. Dramaga, Kabupaten Bogor, Jawa Barat 16680, Indonesia

[4]National Research and Innovation Agency, Republic of Indonesia, Jakarta

*Corresponding author, email: jonni_firdaus@yahoo.com

Article Info

Submitted: 19 June 2024
Revised: 25 June 2024
Accepted: 5 July 2024
Available online: 7 August 2024
Published: August 2024

Keywords:
Near-infrared; soil; nitrogen; phosphorus; potassium.

***How to cite**:*
Firdaus, J., Ahmad, U., Budiastra, I. W., Subrata, I. D. M., 2024. Macro-Nutrient Prediction of Rice Field Soil Using Artificial Neural Network and NIR Spectroscopy. Jurnal Keteknikan Pertanian, 12(2): 242-258. https://doi.org/ 10.19028/jtep.012.2.242-258.

Abstract

Understanding soil fertility, influenced by macronutrients such as nitrogen, phosphorus, and potassium, is essential for the implementation of adaptive agriculture based on various soil conditions. Near-infrared spectroscopy technology provides non-destructive, rapid soil property measurements without chemicals, and is applicable both in the field and in the laboratory. However, the wide NIR spectrum range and neural network complexities can hinder Artificial Neural Network (ANN) training and inference, leading to time and resource inefficiency, particularly without sophisticated computing devices. This study examined data reduction methods to enhance ANN performance in predicting soil macronutrients using NIR spectra. Multiple Linear Regression (MLR) and Principal Component Analysis (PCA) were applied to select wavelengths from the 1000–2500 nm for ANN input, comparing their performance. Approximately 237 NIR reflectance data points from paddy soils were transformed into absorbance data. MLR used forward selection to identify wavelengths with correlations higher than 0.9, whereas PCA selected wavelengths corresponding to the loading factor peaks for each principal component. These selected wavelengths served as inputs for the ANN model. The ANN performance was assessed using the correlation and determination coefficients, RMSE, RPD, and model consistency. For nitrogen, the PCA+ANN model with reflectance spectra performed better (RPD 2.4-4.8) than the MLR+ANN model (RPD 2.2-2.6) using fewer wavelengths (5-9 for PCA+ANN vs. 9-12 for MLR+ANN). For phosphorus estimation, the PCA+ANN model also excelled (RPD 2.3-7.0 vs. 2.3-2.4) with fewer wavelengths (4-7 vs. 7). The PCA+ANN model showed superior performance (RPD 4.3-9.5 vs. 4.2-4.4), using the same number of wavelengths (4-8 vs. 4-6).

Doi: https://doi.org/10.19028/jtep.012.2.242-258

JTEP Jurnal Keteknikan Pertanian,
Vol. 12 No. 2, p 242-258, 2024
P-ISSN 2407-0475 E-ISSN 2338-8439
Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.012.2.242-258

1. Introduction

The role of soil in agriculture is crucial because it provides essential nutrients for plant growth. The application of adaptive agriculture to diverse soil conditions, knowing soil fertility, one of which is determined by macronutrients such as nitrogen (N), phosphorus (P), and potassium (K), is very important. However, traditional soil analysis is time consuming, costly, and potentially environmentally damaging because of chemical waste. Therefore, there is an urgent need for efficient, environment-friendly, and economical soil analysis methods to ensure sustainable soil nutrient availability for optimal agriculture (Johnson et al., 2019).

Soil is a complex mixture of water, air, and organic and inorganic matter particles. The diverse composition of soil organic matter significantly affects factors such as biological processes, availability of nutrients such as N, P, and K, soil dynamics, structure, aggregation, and water holding capacity. These properties vary depending on the source and type of the soil. The relationship between NIR spectra and soil macronutrients is not always linear but can be nonlinear, influenced by the diversity of soil physical, chemical, and biological properties, which are very complex.

Near-infrared spectroscopy (NIRS) is a promising alternative to conventional soil-analysis techniques. NIRS technology has several advantages, such as being non-destructive, simple sample preparation without the need for chemicals, fast measurement, the ability to predict soil properties through one scan, and the ability to be used both in the field and the laboratory.

Several studies have explored the potential of NIRS to estimate soil properties and fertility (Barthès et al., 2019; Pusch et al., 2023; Srisomkiew et al., 2022; Zheng et al., 2023). In N estimation, high R^2 values between 0.75 - 0.95 were obtained (Bachion de Santana and Daly 2022; Munawar et al., 2020; Ng et al., 2020; Pudełko and Chodak, 2020; Reda et al., 2019; Zhou et al., 2019). Estimation of P by Reda et al., 2020 showed a low R^2 value, but in other studies, a high R^2 value was obtained ranging from 0.63 - 0.91 (Cai et al., 2021). For K prediction, it showed a low R^2 value between 0.47 - 0.59 (Johnson et al., 2019), but in other studies, a high R^2 value was obtained between 0.92 - 0.99 (Tang et al., 2020). This study used a wide wavelength range of 1000-2500 nm. According to Firdaus et al. (2024), the use of Multiple Linear Regression (MLR) as an NIR wave calibration model of nutrients, even though it has been able to reduce the number of wavelength variables from 1501 to 4-12 variables but showed R^2 values still below 0.90 for reflectance and absorbance spectra, so there is still a need for a calibration model that can improve estimation accuracy with the involvement of a minimal number of wavelengths.

In NIRS, the construction of a calibration model aims to relate the spectral data to the soil nutrient values. Machine learning is used to develop algorithms and models that can learn from data, detect patterns, and make predictions and decisions. One of the models used in machine learning is the Artificial Neural Network (ANN). Several studies using ANN have shown a strong correlation

Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.012.2.242-258

between NIR spectra and soil NPK content. However, ANN have disadvantages, particularly in terms of their complexity. The ANN model is difficult to interpret because of the large amount of input data for up to 1501 NIR spectra (1000-2500 nm), which slows the understanding of the contribution of each spectral feature to the prediction results. In addition, ANN training and inference processes are time- and resource-intensive. A large number of parameters and complex calculations can slow down this process, especially in the absence of sophisticated computing devices.

One approach to reducing model complexity and data volume is data reduction techniques. Data dimensionality reduction is often used to simplify or reduce data complexity, increase computational efficiency, and improve the interpretability and model performance. One of the data dimension reduction techniques that can be applied to NIR spectral data is Principal Component Analysis (PCA), which uses the loading score value in PCA and shows the contribution of data to its main component (Ouassila et al., 2023; Razmi et al., 2022). Another technique to reduce wavelengths by selecting wavelengths that contribute to soil nutrients is MLR. MLR is a statistical technique that is used to model the linear relationship between one dependent variable and two or more independent variables.

This study aimed to analyze the performance of MLR and PCA in selecting and reducing wavelengths as input data for ANN models to predict soil macronutrients (NPK) in rice fields using NIRS.

2. Materials and Methods

2.1 Soil Sample

In this research, a total of 79 soil samples were collected from rice fields in Bogor, Subang, Indramayu, and Sukabumi districts of West Java Province. These samples, weighing 500 g each, were obtained using a purposive random sampling approach that was based on soil nutrient availability maps. The soil types that were predominantly found at the sampling sites were Gleisol, Cambisol, Regosol, Latosol, Podzolic, and Andosol. The primary parent material categories encompass clay, andesite, and basalt deposits (BBSDLP, 2016a; BBSDLP, 2016b; BBSDLP, 2017a, 2017b). Soil samples were obtained from a depth of 0-20 cm using a hoe and subsequently placed in plastic bags for transportation to the laboratory for examination.

2.2 NIR Reflectance Retrieval and Data Pretreatment

The soil samples were dried to a mean moisture content of 8.31% with a standard deviation of 1.14%. Afterward, the samples were freed of foreign materials, sifted through a 0.75 mm sieve, and then transferred to a 9 cm diameter and 2 cm high Petri dish. Reflectance data (R) were acquired using a Buchi NIRflex N500 spectrometer (1000-2500 nm), with a total of 1501 data points. For each of the

JTEP Jurnal Keteknikan Pertanian,
Vol. 12 No. 2, p 242-258, 2024
P-ISSN 2407-0475 E-ISSN 2338-8439
Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.012.2.242-258

237 spectral data points, each sample was measured three times by rotating the Petri dish by 1/3 of its diameter. The reflectance data were then transformed into absorbance values (A = log(1/R)).

2.3 Conventional Measurement of Soil Macronutrient Content

Soil samples with spectral data were then conventionally measured for macronutrients: total N using the Kjeldahl method, total P using the HNO_3 extraction method, and total K using the HNO_3 extraction method (Balit Tanah 2009).

2.4 ANN Model Calibration

The calibration process uses an ANN model with two wavelength selection methods as the input data, MLR and PCA (Figure 1). The architecture of the developed ANN model consisted of an input layer, one hidden layer with the number of nodes optimized for the best performance, and one output layer for each nutrient, N, P, or K.

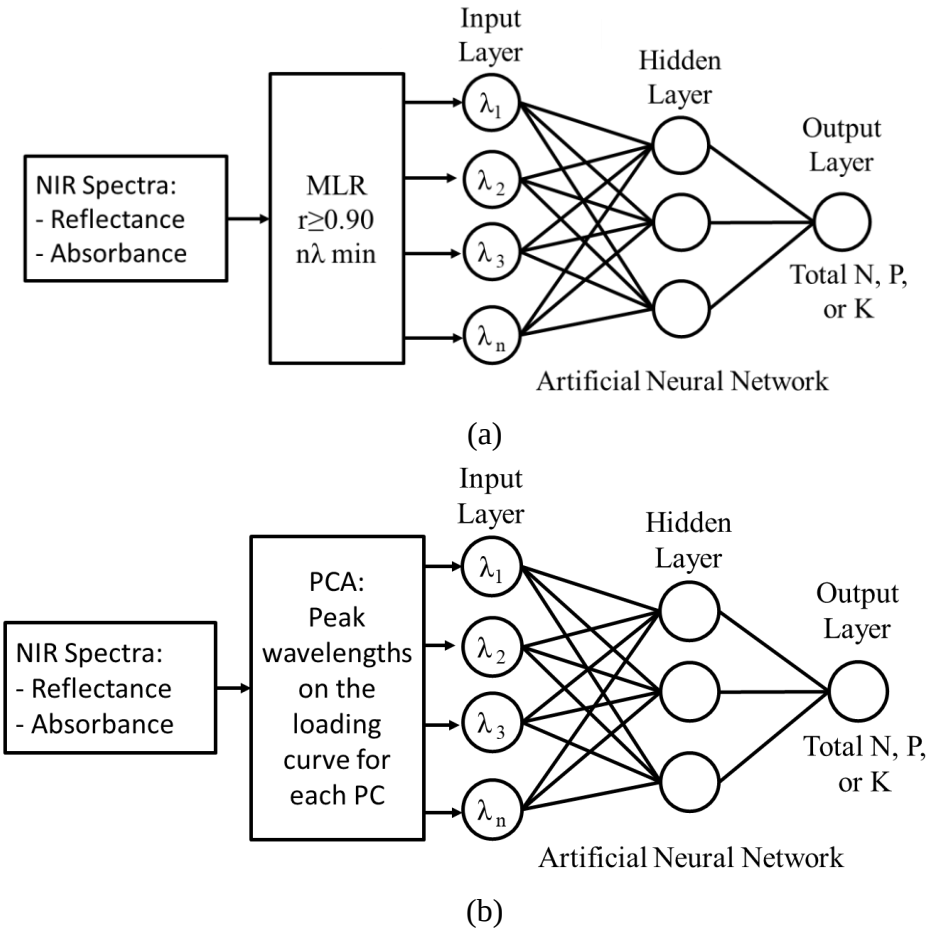

Figure 1. Artificial Neural Network Model with Input Selection: (a) MLR+ANN and (b) PCA+ANN

Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.012.2.242-258

The model uses feedforward backpropagation with the Levenberg-Marquardt training algorithm. The transfer function used in the hidden layer is sigmoid, where the sigmoid function is nonlinear, which can capture and model nonlinear relationships in the data (Dubey et al. 2022), whereas the output layer uses a linear function because the linear function has an input scale value proportional to the output, and the prediction is on the same scale as the original target (Feng and Lu 2019).

The calibration procedure used 237 spectral data points, with 150 data points for calibration and 87 data points for validation. Model calibration was performed by optimizing the number of nodes in the hidden layer for each R and A data spectrum using MATLAB software. Both the R and A spectra were used to determine which spectra had a stronger relationship with nutrient content values. The transformation of R to A may cause a change in the linearity of the spectra concerning nutrient content.

2.4.1 Wavelength selection based on MLR output

The MLR method was the forward method using SPSS software. The selection and addition of wavelength variables were stopped when the total correlation of the MLR model exceeded 0.90. The input data for the MLR equation consisted of R and A data in the range 1000-2500 nm (1501 data). The selected wavelengths were then used as input data for the ANN.

2.4.2 Wavelength selection using PCA

PCA is a technique that reduces the dimensionality of multivariate data into a smaller set of linear combinations but captures most of the variance of the original data. The PCA process used was Unscramble X 10.4 software. Wavelength selection was determined based on the loading value that becomes the peak of the curve in each principal component (PC) in the R and A spectra. The input data for the PCA equation consisted of R and A data in the range–1000-2500 nm (1501 data). The selected wavelengths were then used as input data for the ANN.

2.5 ANN Model Validation Performance

The good model performance was assessed by a high correlation coefficient (r), high coefficient of determination (R^2), low Root Mean Square Error (RMSE), high ratio of performance deviation (RPD), and 80-110% consistency.

$$RPD = (St.dev)/RMSE \quad (1)$$

$$Consistensy = RMSE_{calibration}/RMSE_{validation} \times 100\% \quad (2)$$

dev represents the standard deviation of the reference and RMSE is the root-mean-square error between the predicted and actual values.

JTEP Jurnal Keteknikan Pertanian,
Vol. 12 No. 2, p 242-258, 2024
P-ISSN 2407-0475 E-ISSN 2338-8439
Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.012.2.242-258

3 Result and Discussion

3.1 Calibration and Prediction Result Using PCA-ANN and MLR-ANN

3.1.1 Nitrogen

Nitrogen in the soil exists in the form of inorganic compounds, such as nitrate (NO_3^-), nitrite (NO_2), ammonium (NH_4^+), and organic compounds, such as amino acids (proteins), glucosamine (hexosamine), purines, and pyrimidines. The total N content in mineral soils is very small, ranging from 0.02% to 0.5% (Havlin et al., 2017), and in Indonesia, it varies between 0.01% and 0.59% (Ng et al., 2020). The N content in this study ranged from 0.09% to 0.24%.

The low N content causes the absorbance of ions associated with N to be small, making it difficult to observe wave peaks (Figure 2). Therefore, the selection of NIR wavelengths that were strongly correlated (r>0.9) with N was performed using MLR. PCA was used to determine the main wavelengths with the largest variations in the data.

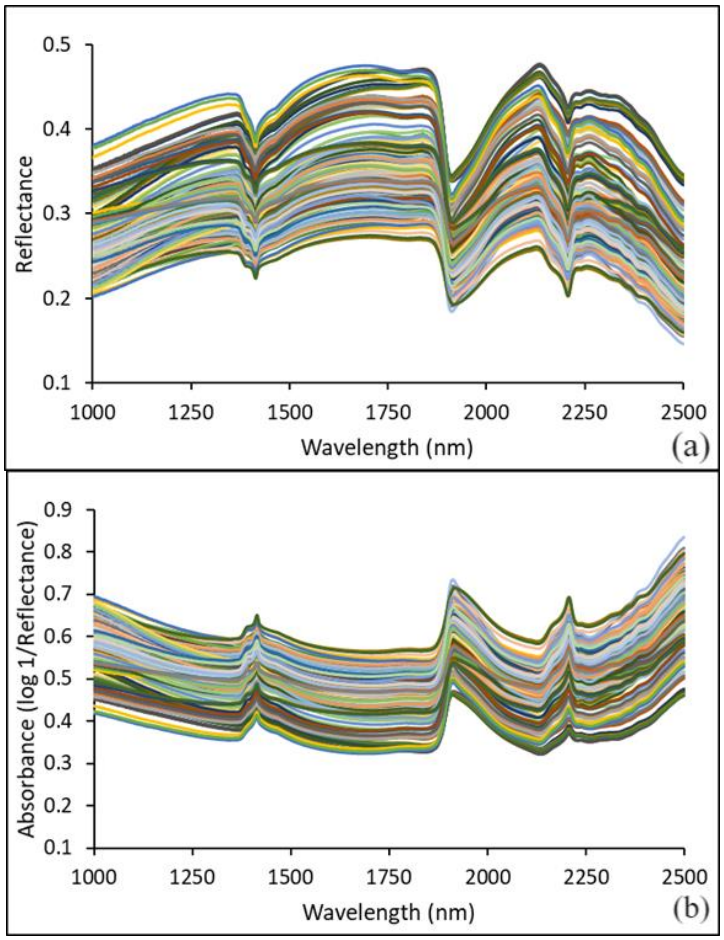

Figure 2. (a) Reflectance spectra and (b) absorbance spectra of NIR wavelengths from soil samples.

JTEP Jurnal Keteknikan Pertanian,
Vol. 12 No. 2, p 242-258, 2024
P-ISSN 2407-0475 E-ISSN 2338-8439
Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.012.2.242-258

Some of the NIR wavelengths with the hight contributions to the MLR model and principal components are listed in Table 1. These wavelengths were used as ANN input data. The MLR model using R spectra obtained 9 wavelengths with an r value of 0.91, while using A spectra produced more wavelengths (12 wavelengths), but had a smaller r of 0.89 (Firdaus et al., 2024).

Table 1. Selected Wavelengths for Nitrogen Nutrients Using MLR and PCA.

Selection Model	Spectra	r	Wavelengths (nm)											
			1	2	3	4	5	6	7	8	9	10	11	12
MLR	R	0.91	1266	1438	2059	1885	1398	2201	2216	2302	1083	-	-	-
	A	0.89	1255	1399	2168	1383	1406	1000	1003	1030	1007	2203	2216	1036
PCA	R	-	1908	2208	1601	2051	1380	2161	1716	2260	1416	-	-	-
			PC3	PC3	PC2	PC2	PC1	PC1	PC1	PC2	PC3			
	A	-	1612	1871	2260	1390	1383	1908	1838	-	-	-	-	-
			PC2	PC4	PC2	PC1	PC4	PC3	PC4					

The PCA model using the R spectra obtained nine wavelengths, consisting of three wavelengths each in PC1, PC2, and PC3. Meanwhile, the PCA model with spectral data A produced fewer than seven wavelengths, consisting of one wavelength in PC1 and PC3, two wavelengths in PC2, and three wavelengths in PC4.

In Table 2, the PCA + ANN model consistently performed better than the MLR + ANN model, both in the calibration and validation stages. This was indicated by the higher r-value, higher R^2, and lower RMSE of the PCA + ANN model. The PCA + ANN model also produced higher RPD values, indicating a higher prediction accuracy for determining N nutrients.

Table 2. Performance of ANN in Determining Nitrogen Nutrients Based on Wavelength Selection Model.

Selection Model	Spectra	Number of Inputs	Nodes in Hidden Layer	Calibration				Validation				Consistency (%)
				r	R^2	RMSE (%)	RPD	r	R^2	RMSE (%)	RPD	
MLR+ANN	R	9	3	0.93	0.87	0.013	2.8	0.93	0.86	0.014	2.6	92.9
	A	12	3	0.92	0.84	0.014	2.5	0.89	0.79	0.017	2.2	82.4
PCA+ANN	R	5	4	0.92	0.84	0.015	2.5	0.91	0.83	0.015	2.4	96.8
		6	6	0.96	0.92	0.011	3.5	0.93	0.86	0.013	2.6	83.0
		7	**8**	**0.97**	**0.94**	**0.009**	**4.1**	**0.96**	**0.92**	**0.011**	**3.6**	**81.2**
		8	6	0.97	0.93	0.009	3.9	0.96	0.92	0.011	3.5	82.4
		9	5	0.96	0.92	0.011	3.5	0.95	0.91	0.011	3.3	97.7
	A	5	7	0.95	0.90	0.012	3.2	0.92	0.84	0.014	2.5	84.4
		6	7	0.96	0.93	0.010	3.7	0.96	0.92	0.011	3.5	90.2
		7	8	0.97	0.93	0.009	3.8	0.96	0.92	0.011	3.5	82.3

The best model is indicated by writing it in bold

Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.012.2.242-258

Specifically, on the R spectra, the PCA + ANN model showed excellent performance with R^2 (0.84-0.94) and RPD (2.5- 4.1) in the calibration stage, and R^2 of approximately 0.83 to 0.92 and RPD of approximately 2.4 to 3.6 in the validation stage. The use of the PCA + ANN model, especially with the R spectra, yielded better results in predicting the N nutrient content compared to the MLR + ANN model.

When viewed from the value of consistency in nitrogen nutrient estimation, all models, both MLR + ANN and PCA + ANN for the R and A spectra, had consistency values above 80%, indicating that all models had relatively similar performance between the calibration and validation datasets. The high consistency of the PCA + ANN model indicates that it performs similarly for calibration and validation data. High consistency close to 100% indicates good generalization ability and reliability in nitrogen nutrient prediction for new data. Consistency was calculated using Equation (2).

Previous studies on predicting soil nitrogen content using NIR spectroscopy have shown varying degrees of success with different wavelength ranges and multivariate methods. Barthès et al. (2019), using the range of 1100-2498 nm with the PLS method, obtained an R^2 value of 0.77 and an RPD of 2.10. Munawar et al., (2020) used the range of 1000-2500 nm with PCR and PLS methods, resulting in R^2 values of 0.85 and 0.87 and RPD of 2.0 and 3.5. Pudełko and Chodak (2020) applied various methods to the 1000-2500 nm range and obtained R^2 values between 0.89 and 0.93. Reda et al., (2019) with a range of 1100-2500 nm and the PLS method, achieved an R^2 value of 0.80 and an RPD of 2.77, and using BPNN, BVE-BPNN, and ELM, getting an R^2 value of up to 0.94 and an RPD of up to 4.91. Ng et al. (2020) using the 1300-2600 nm range with the Cubist model, resulted in an R^2 value of 0.52. In this study, which uses ANN and input R spectra from selection using the loading value in PCA, although it only uses seven wavelength variables as ANN input, the results show improved accuracy with higher r, R^2, RMSE, and RPD values of 0.97, 0.94, 0.009, and 4.1 for calibration, and 0.96, 0.92, 0.011, and 3.6, respectively, for validation.

3.1.2 Phosphorus

Phosphorus in soil is present in inorganic forms, including various phosphate compounds such as calcium (fluorapatite, hydroxyapatite, trisodium phosphate, octacalcium phosphate, dicalcium phosphate, and phosphate of potassium dihydrate), iron (strengite), and aluminum (variscite, potassium taranacite, and berlinite), as well as in the form of $H_2PO_4^-$ and HPO_4^{2-} ions. Organic forms of P in soil include inositol phosphate, nucleic acids, phospholipids, phosphoproteins, and metabolic phosphates (Havlin et al., 2017). According to Ng et al. (2020), the potential P content in Indonesia ranges from 0.28 to 129.58 mg/100 g soil (0.00028-0.1296%), while in this study, it varied between 0.01% and 0.36%.

JTEP Jurnal Keteknikan Pertanian,
Vol. 12 No. 2, p 242-258, 2024
P-ISSN 2407-0475 E-ISSN 2338-8439
Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.012.2.242-258

The P content was also low, leading to minimal absorbance from the ionic bonds associated with P, making it difficult to observe the wavelength peak. The relationship between the NIR wavelength and P content was examined using the MLR equation. PCA was used to determine the main wavelength with the largest variation in data.

Some of the NIR wavelengths with the largest contributions to the MLR model and principal components are listed in Table 3. These wavelengths were used as ANN input data. The MLR model using R spectra produces 7 wavelengths with an r value of 0.78 while using A spectra produces the same wavelengths (7 wavelengths) and an r value of 0.79 (Firdaus et al., 2024).

Table 3. Selected Wavelengths for Phosphorus Nutrients Using MLR and PCA.

Selection Model	Spectra	r	Wavelengths (nm)						
			1	2	3	4	5	6	7
MLR	R	0.78	1838	1411	1417	1821	1900	1393	1416
	A	0.79	1838	1411	1417	1821	1898	1866	1903
PCA	R	-	1716	1908	2051	2208	1416	1380	-
			PC1	PC3	PC2	PC3	PC3	PC1	
	A	-	1908	1838	1871	2076	1612	2404	2386
			PC3	PC4	PC4	PC2	PC2	PC4	PC1

The PCA model using R spectra obtained six wavelengths: two wavelengths in PC1, one wavelength in PC2, and three wavelengths in PC3. Meanwhile, the PCA model with spectra A produced more wavelengths (seven wavelengths), consisting of one wavelength in PC1 and PC3, two wavelengths in PC2, and four wavelengths in PC4.

Based on Table 4, it can be observed that in the calibration stage, the MLR + ANN model shows a high R^2 value above 0.90, indicating that the model can explain the variation in the data well. However, the PCA + ANN model with R-spectra performed better, with an R^2 value close to 0.99. In addition, the low RMSE of the PCA + ANN model indicated the accuracy of the prediction of the model concerning the actual data. Furthermore, in the validation stage, both the MLR + ANN and PCA + ANN models showed good performance with high R^2 values and low RMSE. However, the PCA + ANN model tended to provide more consistent results than the MLR + ANN model, especially when using the R spectra.

In addition, the RPD value can be seen, which is an indicator of the accuracy of the prediction of the actual value. The higher the RPD value, the better the model performance. In this case, the PCA + ANN model with R-spectra generally provided a high RPD value, indicating its prediction accuracy. The RPD is calculated based on equation (1)

Thus, the PCA + ANN model using R spectra performed very well in determining the P nutrient concentrations. Although the MLR + ANN model also provided satisfactory results, the PCA + ANN

Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.012.2.242-258

model showed more advantages, particularly in terms of performance and prediction accuracy for new data.

Table 4. Performance of ANN in Determining Phosphorus Nutrients Based on Wavelength Selection Model.

Selection Model	Spectra	Number of Inputs	Nodes in Hidden Layer	Calibration				Validation				Consistency (%)
				r	R^2	RMSE (%)	RPD	r	R^2	RMSE (%)	RPD	
MLR+ANN	R	7	2	0.95	0.91	0.022	3.20	0.92	0.84	0.028	2.3	78.6
	A	7	2	0.95	0.90	0.023	3.10	0.92	0.85	0.027	2.4	85.2
PCA+ANN	R	4	10	0.99	0.98	0.011	7.0	0.99	0.97	0.011	6.1	100.2
		5	6	0.99	0.98	0.010	6.9	0.99	0.98	0.010	6.6	96.2
		6	**6**	**0.99**	**0.98**	**0.009**	**7.9**	**0.99**	**0.98**	**0.011**	**7.0**	**80.3**
	A	4	5	0.91	0.82	0.028	2.4	0.90	0.82	0.031	2.3	90.8
		5	6	0.98	0.96	0.014	5.0	0.98	0.96	0.016	4.8	82.1
		6	6	0.99	0.97	0.011	6.2	0.98	0.97	0.013	5.6	81.4
		7	6	0.98	0.97	0.013	5.5	0.98	0.96	0.014	5.2	89.1

The best model is indicated by writing it in bold

When viewed from the consistency value in phosphorus nutrient estimation, all models show a consistency value above 80%, which means that the model has a relatively similar performance between the calibration and validation datasets, except for MLR + ANN on reflectance spectra, which has a consistency below 80%. The high consistency in the PCA + ANN model is especially evident in the reflectance with four input wavelengths, which reached 100.2%. This shows that the model has a 100% similar performance between the calibration and validation data.

Previous research on estimating soil phosphorus content using NIR spectroscopy technology has shown varying degrees of success with different methods and wavelength ranges. Munawar et al., 2020 using PCR and PLS methods (1000-2500 nm) produced R^2 values of 0.93 and 0.99 with RPD values of 3.86 and 5.41, respectively. In contrast, Ng et al. (2020) used the cubist model and second-order polynomial SG pretreatment, and SNV (1300-2600) resulted in a lower R^2 value of 0.47. The current study used ANN and input R spectra from selection using loading values in PCA; although using only six wavelength variables as ANN input, this study achieved the best performance with r, R^2, RMSE, and RPD values for calibration of 0.99, 0.98, 0.009, and 7.9, and 0.99, 0.98, 0.011, and 7.0, respectively. These results show higher performance compared to previous studies, mainly because of the use of ANN, which is better able to capture the complex and nonlinear relationship between NIR spectra and soil phosphorus content, as well as the selection of relevant variables through PCA.

JTEP Jurnal Keteknikan Pertanian,
Vol. 12 No. 2, p 242-258, 2024
P-ISSN 2407-0475 E-ISSN 2338-8439
Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.012.2.242-258

3.1.3 Potassium

Potassium in soil exists in the form of ions in soil moisture, exchangeable K bound to soil colloids, and non-exchangeable K, which is slowly available to plants. K can also be found in soil minerals such as feldspar groups (orthoclase, microcline), mica groups (muscovite, biotite), granitic rocks (feldspar, mica, and quartz), and secondary minerals (clay/mica hydrate, vermiculite, and chlorite). Mineral soils generally contain K in the range of 0.05-3% (Havlin et al., 2017). According to Ng et al. (2020), the potential K content in Indonesia ranges from 0.43 to 166.09 mg/100 g soil (0.00043-0.1661%), while in this study, it varied between 0.01% and 0.39%. The K content was also low, resulting in minimal absorbance from the ionic bonds associated with K, making it difficult to observe the wave peaks. The relationship between NIR wavelengths and K content was examined using the MLR equation, in addition to determining the main wavelength that had the largest variation in the data using PCA.

Some of the NIR wavelengths with the largest contributions to the MLR model and the principal components are listed in Table 5. These wavelengths were used as input data for the ANN. The MLR model using R spectra produces 6 wavelengths with an r value against K of 0.90 while using A spectra produces fewer wavelengths (4 wavelengths) and an r value of 0.90 (Firdaus et al., 2024). The PCA model using the R spectra obtained eight wavelengths consisting of two wavelengths in PC1 and three wavelengths each in PC2 and PC3. Meanwhile, the PCA model with spectra A produces fewer wavelengths (six wavelengths), consisting of one wavelength in PC1, PC2, and PC3, and three wavelengths in PC4.

Table 5. Selected Wavelengths for Potassium Nutrients Using MLR and PCA.

Selection Model	Spectra	r	Wavelengths (nm)							
			1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
MLR	R	0.90	1907	1922	1000	1013	1911	1096	**-**	**-**
	A	0.90	1907	1922	1911	1000	-	-	**-**	**-**
PCA	R		1908	2051	1416	1380	1601	2260	2208	2326
			PC3	PC2	PC3	PC1	PC2	PC2	PC3	PC1
	A		1838	2386	1908	2404	1612	1871	-	-
			PC4	PC1	PC3	PC4	PC2	PC4		

Based on the calibration results in Table 6, the PCA + ANN model with the R and A spectra showed superior performance compared to the MLR+ANN model. The PCA + ANN model with R spectra achieved the highest r value of 1, with an R^2 of 0.99, RMSE of 0.012%, and RPD of 10.1. The model also maintained high performance in the validation phase, with r and R^2 values of 0.99, RMSE of 0.015%, and RPD remained high. The model showed high reliability for the prediction of K levels.

For the MLR + ANN model, both the R and A spectra performed well, but not as well as for the PCA + ANN model. and R^2 values reached 0.98 and 0.96 with RMSE of 0.022% and 0.025% and RPD

Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.012.2.242-258

of 5.3 and 4.8 in calibration. This performance decreased slightly in validation, with r and R^2 values of 0.97 and 0.95, and RMSE of 0.028% and 0.03%, respectively

Table 6. Performance of ANN in Determining Potassium Nutrients Based on Wavelength Selection Model.

Selection Model	Spectra	Number of Inputs	Nodes in Hidden Layer	Calibration				Validation				Consistency (%)
				r	R^2	RMSE (%)	RPD	r	R^2	RMSE (%)	RPD	
MLR+ANN	R	6	5	0.98	0.96	0.022	5.3	0.97	0.95	0.028	4.4	78.6
	A	4	3	0.98	0.96	0.025	4.8	0.97	0.95	0.030	4.2	83.3
PCA+ANN	R	5	4	0.97	0.94	0.030	3.9	0.96	0.93	0.035	3.7	85.7
		6	2	0.98	0.95	0.026	4.7	0.97	0.95	0.027	4.5	92.9
		7	8	0.99	0.98	0.015	8.0	0.99	0.98	0.017	7.5	89.2
		8	**6**	**1.00**	**0.99**	**0.012**	**10.1**	**0.99**	**0.98**	**0.015**	**7.2**	**80.7**
	A	4	5	0.97	0.95	0.026	4.5	0.97	0.95	0.031	4.3	85.5
		5	5	0.99	0.98	0.018	6.5	0.99	0.98	0.021	6.3	88.0
		6	8	0.99	0.99	0.012	9.7	0.99	0.99	0.014	9.5	86.7

The best model is indicated by writing it in bold

Model consistency in potassium nutrient estimation showed that all models showed consistency values above 80%, which means that the models have relatively similar performance between the calibration and validation datasets. Except for MLR + ANN on reflectance spectra, which had a consistency below 80% (78.6 %). The highest consistency was found in the PCA + PLS model, reaching 92.9%.

Previous research on estimating soil potassium content using NIR spectroscopy technology has shown a variety of different results. Munawar et al. (2020) used multivariate PCR and PLS methods in the range of 1000 - 2500 nm, resulting in R^2 values of 0.88 and 0.90 respectively with RPD values of 2.04 and 2.68. Tang et al., (2020) used a range of 1250 - 2500 nm, resulting in an R^2 value of 0.29 and an RMSE of 0.53. Tang et al., (2020) also used the 900 - 1700 nm range, resulting in an R^2 value of 0.19 and an RMSE of 0.58. The current study uses an ANN with R spectral input, which is selected using the loading value in PCA. Despite using only eight wavelength variables as ANN inputs, the results showed the best performance, with r, R^2, RMSE, and RPD values for calibration of 1.0, 0.99, 0.012, and 10.1, and for validation of 0.99, 0.98, 0.015, and 7.2. The high performance of this study compared to previous studies can be attributed to the use of ANN, which is more effective in capturing the complex and nonlinear relationship between NIR spectra and soil potassium content, and the selection of relevant variables through PCA.

The best predicted values of the calibration model for N, P, and K can be seen in Figure 3.

JTEP Jurnal Keteknikan Pertanian,
Vol. 12 No. 2, p 242-258, 2024
P-ISSN 2407-0475 E-ISSN 2338-8439
Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.012.2.242-258

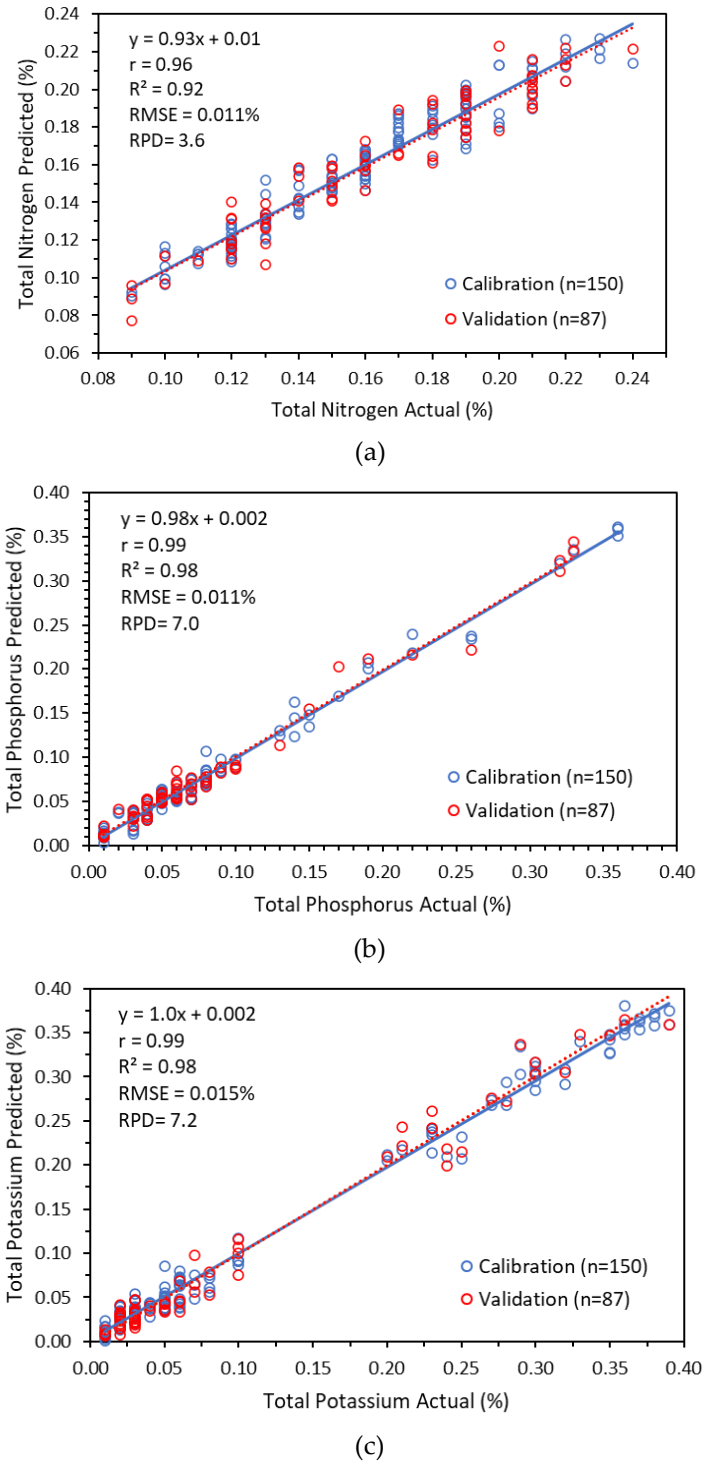

Figure 3. The Best Calibration and Validation for Nutrients: (a) Nitrogen, (b) Phosphorus, (c) Potassium

Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.012.2.242-258

Overall, the PCA + ANN model with R spectra performed best in calibration and validation, with the highest r and R^2 values, as well as the optimal RMSE and RPD values, and showed high accuracy and reliability. PCA can select the wavelength with the greatest variation from each principal component using the loading value that peaks on the curve. These highly variable wavelengths can have a linear or nonlinear relationship with nutrient values, whereas the MLR method only has a linear relationship between the NIR spectra and nutrient values. ANN models with wavelength inputs using the PCA selection method have become stronger in predicting nutrient values because they can accommodate linear and nonlinear relationships with nutrients. The ANN method, which uses the wavelength input from the selection using the MLR method, obtains only linear input data.

The results of the analysis also show that the use of R spectra provides a higher predictive value than the use of A spectra. This is because the R spectra are more linear than the A spectra for soil nutrient values. The log(1/R) transformation in spectra A changes the linearity of spectra R to soil nutrient content in a nonlinear direction in spectra A. Small changes in reflectance values correspond to or are linear with changes in NPK composition. In spectra A, the changes became smaller and less linear with changes in the soil NPK composition.

Strong predictive values based on the results of the analysis were K, P, and N. This was influenced by the variation in nutrient data, as indicated by the coefficient of variation. The largest coefficient of variation of nutrients in the sample was K = 112.4%, P = 96.3%, and N = 22.9%. A larger coefficient of variation indicated that the data variation was over a wider range, thus accommodating more nutrient values in the construction of the calibration model.

4. Conclusion

The PCA + ANN model with R spectra demonstrated superior results for N estimation, exhibiting an RPD range of 2.4-4.8, compared to the MLR + ANN model, which had an RPD range of 2.2-2.6. The PCA + ANN model utilized 5-9 wavelengths, whereas the MLR + ANN model required 9-12 wavelengths. In the P estimation, the PCA + ANN model with R spectra outperformed MLR + ANN with an RPD range of 2.3-7.0 compared to 2.3-2.4. The PCA + ANN model used 4-7 wavelengths, whereas MLR + ANN required seven wavelengths. For K estimation, the PCA + ANN model with R spectra showed better performance, with an RPD range of 4.3-9.5, compared to MLR + ANN, which had an RPD range of 4.2-4.4. The PCA + ANN model utilized 4-8 wavelengths, whereas the MLR + ANN utilized 4-6 wavelengths. The ANN model, which incorporated R spectra and wavelengths from the loading values in the PCA, exhibited the highest performance and employed the fewest wavelengths. Thus, it is recommended as a valuable alternative method for estimating N, P, and K in paddy soils using NIRS waves.

Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.012.2.242-258

Acknowledgement

The researchers express their gratitude to the Ministry of Education, Culture, Research, and Technology for the financial support of this project through the Doctoral Dissertation Research Scheme Year 2022, with contract number 082/E5/PG.02.00.PT/2022.

5. References

Bachion de Santana, F. dan K. Daly. (2022). 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. Spectrochim Acta - Part A Mol Biomol Spectrosc. 279: 121441. Doi: https://doi.org/10.1016/j.saa.2022.121441

Balit Tanah. (2009). Analisis kimia tanah, tanaman, air, dan pupuk. Balai Penelitian Tanah, Badan Litbang Pertanian, Bogor, Indonesia.

Barthès, B.G., E. Kouakoua, M. Clairotte, J. Lallemand, L. Chapuis-Lardy, M. Rabenarivo dan S. Roussel. (2019). Performance comparison between a miniaturized and a conventional near infrared reflectance (NIR) spectrometer for characterizing soil carbon and nitrogen. Geoderma. 338: 422–429. Doi: https://doi.org/10.1016/j.geoderma.2018.12.031

BBSDLP. 2016a. Semi-Detailed Soil Map of Bogor Regency, West Java Province. Center for Agricultural Land Resources, Agency for Agricultural Research and Development, Bogor, Indonesia.

BBSDLP. 2016b. Semi-Detailed Soil Map of Sukabumi Regency, West Java Province. Center for Agricultural Land Resources, Agency for Agricultural Research and Development, Bogor, Indonesia.

BBSDLP. 2017a. Semi-Detailed Soil Map of Indramayu Regency, West Java Province. Center for Agricultural Land Resources, Agency for Agricultural Research and Development, Bogor, Indonesia.

BBSDLP. 2017b. Semi-Detailed Soil Map of Subang Regency, West Java Province. Center for Agricultural Land Resources, Agency for Agricultural Research and Development, Bogor, Indonesia.

Cai, H.T., J. Liu, J.Y. Chen, K.H. Zhou, J. Pi dan L.R. Xia. (2021). Soil nutrient information extraction model based on transfer learning and near infrared spectroscopy. Alexandria Eng J. 60(3): 2741–2746. Doi: https://doi.org/10.1016/j.aej.2021.01.014

Dubey, S.R., S.K. Singh dan B.B. Chaudhuri. (2022). Activation functions in deep learning: A comprehensive survey and benchmark. Neurocomputing. 503: 92-108. Doi: https://doi.org/10.1016/j.neucom.2022. 06.111

Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.012.2.242-258

Feng, J. dan S. Lu. (2019). Performance Analysis of Various Activation Functions in Artificial Neural Networks. Journal of Physics: Conf. Series 1237: 022030. Doi:https://doi.org/10.1088/1742-6596/1237/2/022030

Firdaus, J., U. Ahmad, I.W. Budiastra dan I.D.M Subrata. (2024). Estimating Macronutrient Content of Paddy Soil Based on Near-Infrared Spectroscopy Technology Using Multiple Linear Regression. Adv Technol Innov. 9(1):50–64. Doi: https://doi.org/10.46604/aiti.2023.12683

Havlin, J.L., S.L. Tisdale, W.L. Nelson dan J.D. Beaton. (2017). Soil Fertility and Fertilizers an Introduction to Nutrient Management. 8th ed. Pearson India Education Services Pvt. Ltd., Uttar Pradesh, India.

Johnson, J.M., E. Vandamme, K. Senthilkumar, A. Sila, K.D. Shepherd dan K. Saito. (2019). Near-infrared, mid-infrared or combined diffuse reflectance spectroscopy for assessing soil fertility in rice fields in sub-Saharan Africa. Geoderma. 354: 113840. Doi: : https://doi.org/10.1016/j.geoderma. 2019.06.043

Munawar, A.A., Y. Yunus dan P. Satriyo. (2020). Calibration models database of near infrared spectroscopy to predict agricultural soil fertility properties. Data Br. 30: 105469. Doi: https://doi.org/10.1016/j.dib.2020.105469

Ng, W., L. Anggria, A.F. Siregar, W. Hartatik, Y. Sulaeman, E. Jones and B. Minasny. (2020). Developing a soil spectral library using a low-cost NIR spectrometer for precision fertilization in Indonesia. Geoderma Reg. 22: e00319. Doi: https://doi.org/10.1016/j.geodrs.2020.e00319

Ouassila, B., T.F. Zohra, L. Laid dan B. Hizia. (2023). Neural networks based linear (PCA) and nonlinear (ISOMAP) feature extraction for soil swelling pressure prediction (North East Algeria). Heliyon. 9(8). Doi: https://doi.org/10.1016/j.heliyon.2023.e18673

Pudełko, A. &M. Chodak. (2020). Estimation of total nitrogen and organic carbon contents in mine soils with NIR reflectance spectroscopy and various chemometric methods. Geoderma. 368: 114306. Doi: https://doi.org/10.1016/j.geoderma.2020.114306

Pusch, M., A. Samuel-Rosa, P.S.G. Magalhães dan L.R. do Amaral. (2023). Covariates in sample planning optimization for digital soil fertility mapping in agricultural areas. Geoderma. 429: 116252. Doi: https://doi.org/10.1016/j.geoderma.2022.116252

Razmi, A., M. Rahbar dan M. Bemanian. (2022). PCA-ANN integrated NSGA-III framework for dormitory building design optimization: Energy efficiency, daylight, and thermal comfort. Appl Energy. 305: 117828. Doi: https://doi.org/10.1016/j.apenergy.2021.117828

Reda, R. T. Saffaj, B. Ilham, O. Saidi, K. Issam dan L. Brahim. (2019). A comparative study between a new method and other machine learning algorithms for soil organic carbon and total nitrogen prediction using near infrared spectroscopy. Chemom Intell Lab Syst. 195:103873. Doi: https://doi.org/10.1016/j.chemolab.2019.103873

Available online:
http://journal.ipb.ac.id/index.php/jtep
DOI: 10.19028/jtep.012.2.242-258

Reda, R. T., Saffaj, S.E. Itqiq, I. Bouzida, O. Saidi, K. Yaakoubi, B. Lakssir dan N. El Mernissi. (2020). Predicting soil phosphorus and studying the effect of texture on the prediction accuracy using machine learning combined with near-infrared spectroscopy. Spectrochim Acta Part A Mol Biomol Spectrosc. 242: 118736. Doi: https://doi.org/10.1016/j.saa.2020.118736

Srisomkiew, S., M. Kawahigashi, P. Limtong dan O. Yuttum. (2022). Digital soil assessment of soil fertility for Thai jasmine rice in the Thung Kula Ronghai region, Thailand. Geoderma. 409: 115597. Doi: https://doi.org/10.1016/j.geoderma.2021.115597

Tang, Y., E. Jones dan B. Minasny. (2020). Evaluating low-cost portable near infrared sensors for rapid analysis of soils from Southeastern Australia. Geoderma Reg. 20: e00240. Doi: https://doi.org/10.1016/j.geodrs.2019.e00240

Zheng, G., A. Wang, C. Zhao, M. Xu, C. Jiao dan R. Zeng. (2023). Evolution of paddy soil fertility in a millennium chronosequence based on imaging spectroscopy. Geoderma. 429: 116258. Doi: https://doi.org/10.1016/j.geoderma.2022.116258

Zhou, P., Y. Zhang, W. Yang, M. Li, Z. Liu dan X. Liu. (2019). Development and performance test of an in-situ soil total nitrogen-soil moisture detector based on near-infrared spectroscopy. Comput Electron Agric. 160: 51–58. Doi: https://doi.org/10.1016/j.compag.2019.03.016