

Research Article

Quick Determination of Soil Quality Using Portable Spectroscopy and Efficient Multivariate Techniques

Ernest Teye ¹, Charles L. Y. Amuah ², Kofi Atiah ³, Ransford Opoku Darko ¹, Thomas Abindaw ³, Kwadwo Kusi Amoah ⁴, Michael Miyittah ⁵, Emmanuel Afutu ⁴, and Rebecca Owusu ⁶

¹School of Agriculture, Department of Agricultural Engineering, University of Cape Coast, College of Agriculture and Natural Sciences, Cape Coast, Ghana

²School of Physical Sciences, Department of Physics, University of Cape Coast, College of Agriculture and Natural Sciences, LAFOC, Cape Coast, Ghana

³School of Agriculture, Department of Soil Science, University of Cape Coast, College of Agriculture and Natural Sciences, Cape Coast, Ghana

⁴School of Agriculture, Department of Crop Science, University of Cape Coast, College of Agriculture and Natural Sciences, Cape Coast, Ghana

⁵School of Biological Sciences, Department of Environmental Sciences, University of Cape Coast, College of Agriculture and Natural Sciences, Cape Coast, Ghana

⁶School of Agriculture, Department of Agric Economics and Extension, University of Cape Coast, College of Agriculture and Natural Sciences, Cape Coast, Ghana

Correspondence should be addressed to Ernest Teye; ernest.teye@ucc.edu.gh

Received 19 December 2022; Revised 17 April 2023; Accepted 19 April 2023; Published 11 May 2023

Academic Editor: Maria C. Yebra-Biurrun

Copyright © 2023 Ernest Teye et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Rapid and onsite determination of the soil status and quality parameters holds a brighter potential for improving food security, and minimizing waste of the excessive application of soil amendments hence reducing environmental pollution. In this study, a pocket-sized shortwave NIR spectroscopy (740–1070 nm) and multivariate statistics were used to classify soil from different land-use types and simultaneously predict nitrogen (N), phosphorus (P), potassium (K), calcium (Ca^{2+}), magnesium (Mg^{2+}), and pH in Ghana. *Different Algorithms*. Linear discriminant analysis (LDA), support vector machine (SVM), and partial least squares algorithms (full-range partial least square, FrPLS; interval partial least squares, IPLS; synergy interval partial least squares, Si-PLS) were attempted for building a suitable classification and quantification model. The models were assessed by the classification rate, coefficient of determination (R_p^2), and root mean square error of prediction (RMSEP) in the prediction set. A total of 110 soil samples from 0 to 15 cm, 15 to 30 cm, and 30 to 45 cm layers were collected from the field of different land-use cropping systems. The results obtained showed that SVM had a 98.61% classification rate of the soil from the cropping system. While Si-PLS was superior in predicting N, P, K, Mg^{2+} , Ca^{2+} , and pH. The performance of the Si-PLS model for N, P, K, Mg^{2+} , Ca^{2+} , and pH were 0.571, 0.779, 0.910, 0.778, 0.826, and 0.904 for R_p^2 and 0.033%, 0.738 $\text{mg}\cdot\text{kg}^{-1}$, 0.117 $\text{cmol}\cdot\text{kg}^{-1}$, 0.654 $\text{cmol}\cdot\text{kg}^{-1}$, 3.0219 $\text{cmol}\cdot\text{kg}^{-1}$, and 0.4760 pH unit for RMSEP, respectively. The results revealed that the portable NIR spectroscopic technique could be used to measure the soil status and some quality parameters. However, further studies are needed to proof its application. This could lead to improving the yield and saving the cost of fertilizer application.

1. Introduction

Soil quality and soil fertility management play a significant role in agricultural productivity and environmental

pollution control, and therefore, a rapid knowledge of the soil quality status of the soil is a vital step. The status of soils is normally measured by analysing the soil using the traditional laboratory technique known as wet chemistry to

provide useful information. The wet chemistry approach comes with its numerous challenges, such as it is expensive, time consuming, involve chemical usage, often restricted to fewer samples, or samples are bulked from an area to provide representative composites and prompting the use of pedo-transfer functions as a substitute [1, 2]. It is also known to generate unwanted waste and destructive to the original soil samples [3]. Above all, it is limited to the laboratory and cannot be used in the field where it is needed most to provide rapid and accurate results to assist in the promotion of precision agriculture. This shows that an alternative technique is required in the face of promoting in situ determination to encourage precision agriculture. The development of alternative measurement methods that are accurate, rapid, and inexpensive is of great value [4].

NIR spectroscopy is an advanced analytical technique that has gained ground in various fields including agriculture. It provides many useful advantages over the traditional analytical methods. These advantages include the following: it is physical, nondestructive, rapid results, and no chemical usage, hence environmentally friendly and inexpensive [2]. Research conducted by other researchers have revealed the potential usefulness of NIR spectroscopy for soil analysis and notable among them include the measurement of heavy metals in soil [4], soil physical, chemical, and biochemical properties [5], soil carbon and nitrogen [6, 7], discrimination of three major soil types [8] and discrimination of organic matter in soil from grass and forest [9]. All these aforementioned studies have proven that NIR spectroscopy could provide the needed alternative for soil analysis. However, all these studies involve the use of large NIR machine that defeat the purpose of onsite usage. Hence, there is little or no attempt to use a small NIR spectrometer for simultaneous determinations of soil health properties. However, due to the advances in computers and electronics, portable or small NIR spectroscopy has been proposed and developed coupled with chemometric. This could provide an added advantage over the laboratory-based NIR spectroscopy. However, up until now, little or no studies have been done in Ghana on the use of pocket-sized user-friendly NIR spectroscopy for soil analysis on the classification of different land-use types and also for predicting soil health quality parameters.

This research, therefore, seeks to investigate the feasibility of applying pocket-sized NIR spectroscopic techniques coupled with multivariate statistics by employing a variable-wise selection protocol for the simultaneous classification and detection of soil health properties to inform the stepwise precision application of soil amendment. The specific objectives are to predict the identification of soil under different land-use types and determine N, P, K, pH, Ca^{2+} , and Mg^{2+} simultaneously by employing synergy interval variable selection optimum.

2. Materials and Methods

2.1. Sample Collection. A total of 110 soil samples were collected at different depth (0–15, 15–30, and 30–45 cm) from different land-use types such as arable, native, pasture,

and plantation as describe by others [10]. Physically, any rough stones and plant debris were removed before the soil samples were air dried. The soil samples were then individually uniformly ground, sieve through a 2 mm sieved, and then package in a well-labelled polythene bag before analysis.

2.2. Sample Spectral Acquisition. The spectrum of each sample was obtained in the reflectance mode using a pocket-sized spectrometer (SCIO™) in a spectral range of 740 nm–1070 nm in a 1 nm resolution for spectra data recording. To scan the samples, a 60 g sample was poured into a glass container as seen in Figure 1 and scanned four times after rotating it at 45°. The whole process was carried out at 28–31°C and 65% relative humidity. The raw dataset of 110 soil samples stored in the cloud based were downloaded using a research license of SCIO lab and imported into MATLAB version 9.5.0 (Mathworks Inc., USA). The downloaded raw dataset was divided into two subsets called the calibration set (77 samples) for developing the model and the prediction set (33 samples) for evaluating the predictability of the developed model. To avoid bias in the selection of members in each subset, the Kennard–Stone algorithm was used in the partitioning of the dataset.

2.3. Reference Methods. The pH of the soils was measured in a 1:2.5 (w/v) soil: water ratio with a pH meter [11]. Total nitrogen (N) was determined using the micro Kjeldahl digestion method [12]. Available phosphorous in the soil was determined following the Bray-1 acid method [13]. Ca, Mg, and K were determined through extraction using the ammonium acetate method at pH 7 [14]. All the analysis were done in triplicates, and the measured soil chemical properties were statistically processed in terms of the range (maximum to minimum values), mean, and standard deviation (SD) as seen in Table 1.

2.4. Mathematical Signal Treatments. In this study, five mathematical spectral signal pretreatments (MC, mean centring; MSC, multiplicative scatter correction; SNV, standard normal variate; FD, first derivative; and SD, second derivative) were comparatively used to obtain the best model developed. In NIR modelling, it has become very necessary to pretreat the raw data set with the best techniques and the challenge, however, is there are several of them. It has, therefore, become a huge task coupled with the fact that it cannot be left-out. Spectral pretreatment is known to be an effective method to reduce or eliminate the optical scattering from different particles, reduce noises, and thereby improve prediction accuracy and robustness of the developed model [15]. Also, any interferences caused by light scattering, baseline shift, and slope variations caused by the particle size are causing unwanted signals to be removed [16]. MC uses the principle of calculation of average; thus, this average spectrum of the data set is calculated and this average is subtracted from each spectrum of the acquired data [17]. SNV is normally used to remove scatter variation from the



FIGURE 1: Soil spectral data acquisition setup.

TABLE 1: Wet chemistry measurement of soil health properties.

Parameters	No. of samples	Ranges	Mean	STD
N (%)	77	0.010–0.26	0.059	0.040
	33	0.015–0.18	0.063	0.042
P (mg·kg ⁻¹)	77	0.0045–3.99	0.769	1.230
	33	0.0081–3.54	0.766	1.257
K (cmol _c ·kg ⁻¹)	77	0.020–0.90	0.199	0.202
	33	0.020–0.74	0.200	0.228
Mg ²⁺ (cmol _c ·kg ⁻¹)	77	0.16–3.99	0.873	0.788
	33	0.08–3.54	1.043	0.978
Ca ²⁺ (cmol _c ·g ⁻¹)	77	0.64–18.88	3.562	3.948
	33	0.72–21.36	2.744	4.067
pH	77	4.83–7.57	6.233	0.781
	33	4.81–7.39	5.935	0.599

light source in the spectral data by eliminating multiplicative interferences and scatter [16, 18]. MSC is a unique preprocessing technique that is normally used for the correction of scattered light and to remove different inclinations of spectral peak. For more information, refer to [19]. Also, FD and SD derivatives spectra pretreatments are used to separate overlapping peaks and eliminate the baseline shift and it is improved by using the Savitzky–Golay algorithm.

2.5. Quantification Models. The partial least squares (PLSs) algorithm is a well-known linear multivariate algorithm proposed by Herman Wold for modelling complicated data set [20]. It has recently found its use for analysing spectra data with strong collinear, noise, and redundant variables. However, the original PLS works on full spectrum and involves a larger sample matrix which often has both useful and unwanted information. To overcome this bottle neck in the PLS model, other researchers have resorted to the manual selection of different spectral regions to estimate some chemical composition [21, 22]. This approach, however, is slow and cumbersome and requires a prior experienced knowledge about unique spectra selection. To solve the aforementioned challenges associated to the PLS model, the interval partial least squares (IPLSs) and synergy interval partial least squares (Si-PLSs) models were proposed. For IPLS, it works by splitting the spectra into smaller equidistant regions and they develop the model for each subinterval by the original PLS, while Si-PLS also split the data set into a number of intervals and then calculate all possible

PLS models for all possible combinations of more than one interval (two, three, and four intervals). The best interval for IPLS and Si-PLS are selected based on the lowest root mean square error of calibration (RMSEC) for a single selected interval and for a combination of intervals with the for the best outcome is chosen respectively. The results of the model are normally evaluated by using three main parameters, namely, the RMSECV, the root mean square error of prediction (RMSEP), and the coefficient of determination (R^2) [23, 24]. These parameters are calculated by using the following equation:

$$\begin{aligned} \text{RMSEC} &= \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}, \\ \text{RMSEP} &= \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}, \\ R &= \sqrt{1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}}, \end{aligned} \quad (1)$$

where n = the number of samples. y_i = the reference measurement results for sample i , \hat{y}_i = the estimated results of the model for the sample i , and \bar{y} = the mean of the reference measurement results for all samples in the data set.

3. Results and Discussion

3.1. Spectral Data Presentation. Spectral profile obtained contains useful information for modelling. Figure 2(a) presents the raw spectra of soil samples from different land-use types and this revealed several absorptions bands. However, the spectra profile appears to show similarities with no unique differences when looked at with the naked eyes. Furthermore, the spectra profile appears to have no useful information and this, therefore, called for the use of multivariate algorithms to assist in the building of qualitative and quantitative models for predicting useful parameters of interest. Also, the wavelength range (740–1070 nm) used possesses unique functional groups such as C-H stretch, C-H deformation, S-H, N-H, CH₂, and CH₃ that could correspond to various parameters in soil such as N, P, K, pH, and other distinct attributes (as seen in Table 1) that could be useful for differentiating the various soil types, as seen in Figure 2(b). The wet chemistry results obtained in this study showed a wide range of chemical properties as seen in Table 1, and this could be attributed to the wide array of land use types for the study from which the samples were collected. The results obtained also agree with those of other authors [10]. Furthermore, the relationship between the spectral absorption wavelength and soil chemical composition (absorption of C-H, O-H and N-H bonds) made it possible to quantify specific soil health parameter of interest using appropriate selection of the wavelength region [25], and this could be attributed to the clear separation as observed in Figure 2(b). Also, the organic matter present in the samples used have distinct spectral fingerprints in the NIR

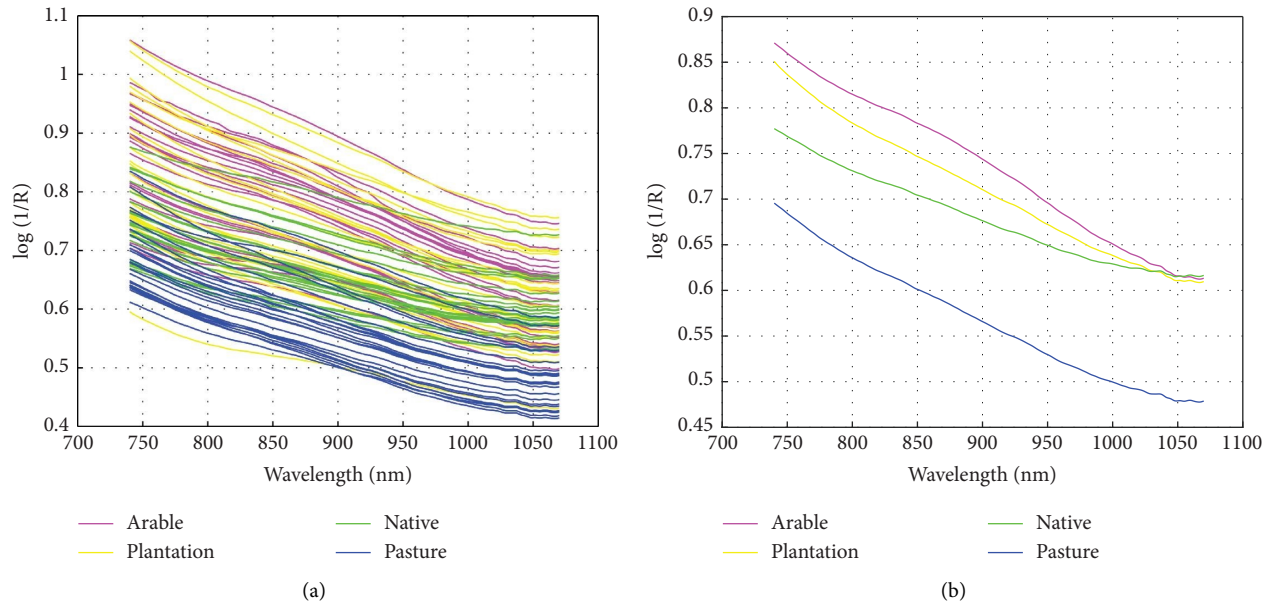


FIGURE 2: Spectra profile (a) raw and (b) mean of soil samples from different land-use types.

region because the relatively strong absorption of overtone and the combination modes relative to several functional groups (CH: aliphatic, CO: carboxyl, NH: amine and amide) are usually present in the organic compounds [26].

3.2. Principal Component Analysis (PCA). Principle component analysis offers an unsupervised pattern recognition tool in a dimensional space for observing any possible cluster trends. It works by reducing the dimension of the data matrix and translating useful information into interpretable variables known as principal components (PCs). Figure 3(a) shows the outcome of PCA and it revealed that there were four distinct soil groups. All the samples clustered well along the two PCS planes where PC1 and PC2 could explain 92.68% and 6.68% of the variance, respectively, giving a total accumulative contribution of 99.37% variance for the 110 samples used in this study. This means the first two principal components (PC1 and PC2) cover the maximum information and provided the chemical compositional information in the NIR region for modelling. Soil samples have considerable unique differences in chemical properties in accordance with their land use type. Since PCA is not a classification tool, LDA and SVM multivariate classification techniques were used for building a classification model.

3.3. Classification Model. There are several classification algorithms and most often the selection of the ones to use is a big challenge. In this experiment, linear discriminant analysis (LDA) and the support vector machine (SVM) were comparatively used. This was because every multivariate classification model has its own strength and weakness. From Table 2, it could be observed that the LDA model had its optimum classification rate at 98.65% and 97.22% in the calibration set and the prediction set, respectively, after the FD preprocess technique was applied on the raw data. This

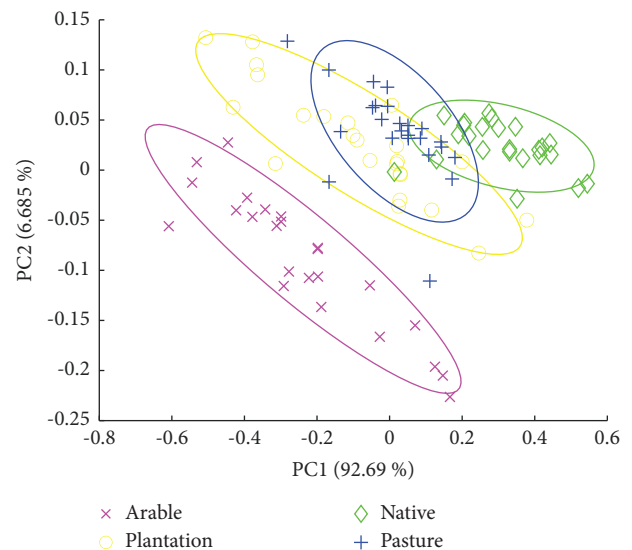


FIGURE 3: PCA cluster trends of soil samples from different land-use types.

finding supports the aforementioned fact that preprocessing methods are known to improve modelling results as it normally eliminates unwanted information, reduce noise, improved accuracy, and enhance robustness of the developed classification model [15].

On the other hand, the SVM obtained the best results comparatively at a classification rate of 99.32% and 98.61% in the calibration and prediction sets, respectively, as seen in Table 2. Also, among the preprocessing techniques used, and MSC and FD improved the raw spectra data set, hence enhanced the final classification rate. It could be explained that MSC is unique in the correction of scattered light and to remove different inclination of the spectral peak while FD enhanced spectra separation. In this research, the MSC-

TABLE 2: Identification rate of different land use classes by LDA and SVM.

Models	Preprocessing	Correct classification rates (%)	
		Calibration set (77)	Prediction set (33)
LDA	RAW	93.20	94.44
	MC	96.62	97.22
	MSC	97.30	98.61
	SNV	96.62	97.22
	FD	98.65	97.22
SVM	RAW	95.83	96.87
	MC	96.62	95.75
	MSC	95.97	98.610
	SNV	96.62	98.83
	FD	99.32	98.61

The bold values represents the best results.

SVM/FD-SVM model was superior in the classification of land use types. A cross-validated analysis was done, and Figure 4 shows cross validation done using randomly selected spectra to test the model. Among the samples used in Figure 4, it was observed that only one sample was misclassified. This sample was the one from the pasture land-use type. It could be explained that the SVM created a hyper-plane that allowed the separation in the higher dimension feature space because the SVM is a transformational tool that converts data from a low dimension input space to a high dimension feature space [17].

Explaining the phenomenon of the accurate classification is vital. Figure 5 reveals the total contribution of the unique wavelengths that contributed to the neat separation and classification of the land use types. At the first component, the major peak was found around 900 nm and this corresponds to CH_3 and CH_2 at the third overtone [26] associated with organic materials, while at the second and third components, the major peaks were found around 800–830 nm, 850–875 nm, 925–950 nm, and 1000–1050 nm. These wavelengths correspond with RNH_2 , ArCH , CH_3 , CH_2 , and RONH_2 [26] that are associated with chemical properties like nitrogen, pH, organic carbons, and among others in the soils used in this research.

3.4. Quantitative Models. The spectral prediction of nitrogen, phosphorus, potassium, pH, calcium, and magnesium were modelled by using different PLS and other wavelength selection techniques (IPLS and Si-PLS). From the results obtained by using the full PLS algorithm, first derivative spectra preprocessing performed better than the others in all the soil quality parameters as seen in Table 3. This performance could be due to first derivatives spectra pre-treatment's ability in greatly defining the presence and locations of hidden absorption bands [27]. Also, from Table 4, the parameters measured did not show any well-defined pattern for the preprocessing models' performance. The parameters measured did not show any well-defined pattern for preprocessing model performance. Thus, mean centring (MC) preprocessing was superior for nitrogen and calcium, while first derivative and SNV outperformed the others for phosphorus, pH, and potassium. Generally, results

Arable	17				100.0%	
Native		16			100.0%	
Pasture			14	1	93.3%	6.7%
Plantation				16	100.0%	
	100.0%	100.0%	100.0%	94.1%		5.9%
	Arable	Native	Pasture	Plantation		
	Predicted Class					

FIGURE 4: Cross-validation identification rate matrix by FD-SVM of different land-use types.

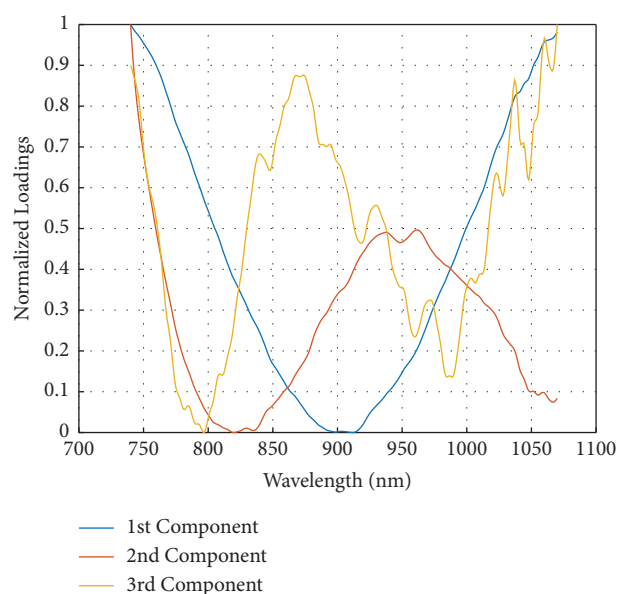


FIGURE 5: Eigenvector values of three principal components of different land-use types.

obtained by using Si-PLS showed an optimal performance for all the parameters studied, as seen in Table 5. Specifically, FD preprocessing spectra treatment also enhanced the results of most quality parameters (N, P, and K), while MC enhanced Calcium results and No preprocessing treatment was needed for pH and magnesium.

Comparatively, as seen from Table 6, IPLS performed the least followed by full PLS, while Si-PLS performed best for all the parameters (N, P, K, Mg^{2+} , Ca^{2+} , and pH) studied. These revelations could be explained by that each PLS type has its unique properties. PLS performed on the full spectral region of the soil samples and contained some irrelevant spectral information which inevitably reduces the performance of the PLS model, while IPLS actually overcome the challenges of PLS by selecting a maximum region of interest to calibrate the PLS model. However, only a single interval selection gives way for the neglect of other useful spectral information.

TABLE 3: Optimal selection of the preprocessing technique using the PLS model.

Parameters	Models	Calibration set (77)			Prediction set (33)		
		R^2	RMSEC	Bias	R^2	RMSEP	Bias
N (%)	RAW	0.5597	0.0335	0.0038	0.5517	0.0349	0.0061
	MSC	0.4886	0.0353	0.0040	0.7505	0.0276	0.0048
	MC	0.5589	0.0335	0.0038	0.5829	0.0340	0.0059
	SNV	0.4982	0.0350	0.0040	0.7510	0.0276	0.0048
	FD	0.5606	0.0335	0.0038	0.7775	0.0263	0.0046
P (mg·kg ⁻¹)	RAW	0.7533	0.7596	0.0866	0.8121	0.7226	0.1258
	MSC	0.7532	0.8039	0.0916	0.8322	0.6866	0.1195
	MC	0.7818	0.7619	0.0868	0.8380	0.6757	0.1176
	SNV	0.7633	0.7894	0.0900	0.8332	0.6847	0.1192
	FD	0.7899	0.7494	0.0854	0.8384	0.6750	0.1175
K (cmol _c ·kg ⁻¹)	RAW	0.7894	0.1234	0.0141	0.6616	0.1690	0.0294
	MSC	0.8445	0.1076	0.0123	0.8826	0.1060	0.0184
	MC	0.8566	0.1037	0.0118	0.8701	0.1111	0.0193
	SNV	0.8454	0.1073	0.0122	0.8872	0.1040	0.0181
	FD	0.8781	0.0962	0.0110	0.8789	0.1075	0.0187
Mg ²⁺ (cmol _c ·kg ⁻¹)	RAW	0.7437	0.5237	0.0597	0.8014	0.5765	0.1003
	MSC	0.7208	0.5430	0.0619	0.7888	0.5924	0.1031
	MC	0.7385	0.5282	0.0602	0.7946	0.5851	0.1019
	SNV	0.7308	0.5347	0.0609	0.7898	0.5911	0.1029
	FD	0.7119	0.5502	0.0627	0.8078	0.5681	0.0989
Ca ²⁺ (cmol _c ·kg ⁻¹)	RAW	0.7629	2.5357	0.2890	0.7982	2.4129	0.4200
	MSC	0.7605	2.5471	0.2903	0.8646	2.0127	0.3504
	MC	0.7629	2.5359	0.2890	0.7973	2.4177	0.4209
	SNV	0.7613	2.5434	0.2899	0.8697	1.9770	0.3442
	FD	0.7830	2.4399	0.2780	0.7871	2.4706	0.4301
pH	RAW	0.8843	0.3624	0.0413	0.5972	0.4738	0.0825
	MSC	0.8912	0.3521	0.0401	0.7667	0.3792	0.0660
	MC	0.8708	0.3817	0.0435	0.5911	0.4765	0.0829
	SNV	0.8914	0.3518	0.0401	0.7855	0.3655	0.0636
	FD	0.8740	0.3773	0.0430	0.5927	0.4758	0.0828

TABLE 4: Optimal selection of the preprocessing technique using the IPLS model.

Parameters	Models	Calibration set (77)			Prediction set (33)		
		R^2	RMSEC	Bias	R^2	RMSEP	Bias
N (%)	RAW	0.4576	0.0361	0.0006	0.4894	0.0390	-0.0137
	MSC	0.4710	0.0358	0.0000	0.4891	0.0389	-0.0135
	MC	0.4991	0.0351	0.0006	0.5138	0.0379	-0.0111
	SNV	0.4477	0.0364	0.0000	0.4477	0.0401	-0.0131
	FD	0.4231	0.0368	0.0012	0.4982	0.0383	-0.0119
P (mg·kg ⁻¹)	RAW	0.6194	0.9627	0.0102	0.6445	0.9611	-0.1017
	MSC	0.5739	1.0074	0.0037	0.5636	1.1048	-0.3496
	MC	0.5533	1.0256	0.0322	0.5973	1.0237	-0.2469
	SNV	0.5157	1.0582	0.0073	0.5090	1.0876	-0.2161
	FD	0.6740	0.9087	0.0127	0.7005	0.9029	0.1131
K (cmol _c ·kg ⁻¹)	RAW	0.6170	0.1586	0.0110	0.6229	0.1802	-0.0364
	MSC	0.7890	0.1253	0.0020	0.7604	0.1700	-0.0864
	MC	0.6977	0.1445	0.0010	0.6746	0.1767	-0.0569
	SNV	0.8104	0.1184	0.0020	0.8268	0.1476	-0.0700
	FD	0.7763	0.1275	0.0030	0.7221	0.1642	-0.0515
Mg ²⁺ (cmol _c ·kg ⁻¹)	RAW	0.7702	0.5029	0.0008	0.7764	0.6087	-0.0252
	MSC	0.5490	0.6597	0.0040	0.5817	0.7861	-0.0578
	MC	0.6285	0.6141	0.0040	0.7510	0.6574	0.1355
	SNV	0.5619	0.6557	0.0108	0.5504	0.8051	-0.0274
	FD	0.5103	0.6760	0.0040	0.5544	0.8040	-0.0526

TABLE 4: Continued.

Parameters	Models	Calibration set (77)			Prediction set (33)		
		R^2	RMSEC	Bias	R^2	RMSEP	Bias
Ca^{2+} ($\text{cmol}_c\cdot\text{kg}^{-1}$)	RAW	0.6073	3.1258	0.0313	0.5450	3.3710	-0.2670
	MSC	0.7369	2.6520	0.0119	0.7346	2.8916	-0.9689
	MC	0.7773	2.4693	0.0749	0.7868	2.5842	-0.7115
	SNV	0.5755	3.2082	0.0242	0.5809	3.2835	-0.2832
	FD	0.6646	2.9410	0.1847	0.6218	3.2138	-0.1248
pH	RAW	0.3762	0.7206	0.0440	0.3209	0.6363	0.2764
	MSC	0.5937	0.6254	0.0048	0.6042	0.5648	0.0948
	MC	0.5815	0.6316	0.0021	0.4628	0.5458	0.1008
	SNV	0.5624	0.6425	0.0217	0.4477	0.5901	0.0398
	FD	0.6887	0.5648	0.0407	0.5005	0.5301	0.1165

TABLE 5: Optimal selection of the preprocessing technique using the Si-PLS model.

Parameters	Preprocessing	Calibration set (77)			Prediction set (33)		
		R^2	RMSEC	Bias	R^2	RMSEP	Bias
N (%)	RAW	0.5540	0.0337	0.0001	0.6571	0.0348	-0.0141
	MSC	0.5158	0.0349	0.0012	0.5307	0.0378	-0.0130
	MC	0.5549	0.0338	-0.001	0.3845	0.0414	-0.0139
	SNV	0.5160	0.0348	0.0000	0.3220	0.0455	-0.0178
	FD	0.5713	0.0333	0.0005	0.6993	0.0335	-0.0139
P ($\text{mg}\cdot\text{kg}^{-1}$)	RAW	0.7714	0.7809	0.0289	0.8382	0.6815	0.0870
	MSC	0.7551	0.8025	0.0310	0.8129	0.7244	0.0432
	MC	0.7779	0.7699	0.0438	0.8360	0.6880	0.0979
	SNV	0.7549	0.8024	0.0304	0.7982	0.7573	-0.0289
	FD	0.7796	0.7659	0.0284	0.8042	0.7381	0.0488
K ($\text{cmol}_c\cdot\text{kg}^{-1}$)	RAW	0.9044	0.0859	0.0005	0.8496	0.1360	-0.0554
	MSC	0.8921	0.0911	0.0032	0.8310	0.1318	-0.0355
	MC	0.9092	0.0838	0.0001	0.8804	0.1254	-0.0501
	SNV	0.8966	0.0892	0.0004	0.8593	0.1229	-0.0319
	FD	0.9108	0.0833	0.0005	0.8980	0.1176	-0.0470
Mg^{2+} ($\text{cmol}_c\cdot\text{kg}^{-1}$)	RAW	0.8010	0.4699	0.0071	0.7586	0.6835	0.0178
	MSC	0.7961	0.4809	0.0116	0.8205	0.5591	0.0532
	MC	0.8066	0.4655	0.0066	0.7976	0.6100	0.0175
	SNV	0.7893	0.4825	0.0028	0.8209	0.5704	0.0293
	FD	0.7786	0.4929	0.0160	0.7613	0.6542	0.1097
Ca^{2+} ($\text{cmol}_c\cdot\text{kg}^{-1}$)	RAW	0.8342	2.1857	0.1060	0.6470	3.0853	-0.1829
	MSC	0.8160	2.3095	0.0142	0.3827	3.9873	-0.4315
	MC	0.8317	2.2040	0.0507	0.6712	3.0206	-0.5359
	SNV	0.8126	2.2983	0.0443	0.5245	3.4507	-0.0758
	FD	0.8269	2.2286	0.0642	0.6662	3.0219	-0.4531
pH	RAW	0.9043	0.3317	0.0004	0.6465	0.4760	-0.0369
	MSC	0.8790	0.3709	0.0052	0.5749	0.5162	0.0465
	MC	0.8927	0.3505	0.0093	0.5445	0.558	-0.0577
	SNV	0.8767	0.3755	0.0050	0.5931	0.5324	-0.0753
	FD	0.8901	0.3543	0.0071	0.5796	0.5105	-0.0040

Hence, it could be seen that IPLS performance declined drastically. On the other hand, its counterpart (Si-PLS model) uses the combination of more than one useful selection of intervals to model the parameter of interest as in the case of this study. Therefore, Si-PLS showed its own superiority over PLS and IPLS because it overcame the demerits showed by both techniques (full PLS and IPLS).

More specifically, for nitrogen prediction, Si-PLS performed best, as seen in Table 6. The optimal spectral interval selected were 770–784, 945–958, and 973–986 nm at 4 PLS

components, as seen in Figure 6(a). These spectra corresponded to various absorption bands for the nitrogen content in soil as these ranges are associated with RNH_2 according to others [26]. These wavelengths are also associated with C-H and N-H third overtones. For phosphorus, the optimum selected wavelengths were 768–781, 894–907, 973–986, and 1058–1070 nm at 3 PLS components as seen in Figure 6(b), which represents the third overtone region and correspond to ArOH, CH_3 , and ArCH. The mobilization of phosphorus plays a vital role in capturing, storing, and

TABLE 6: Modelling results from different variable selection models for soil quality parameters.

Parameters	Models	Var	Wavelengths (nm), selected	Calibration (77)			Prediction (33)		
				R^2	RMSEC	Bias	R^2	RMSEP	Bias
N (%)	PLS	331	740–1070	0.560	0.033	0.0038	0.777	0.026	0.004
	IPLS	17	842–858	0.441	0.036	-0.0002	0.413	0.038	-0.005
	Si-PLS	43	770–784 945–958 973–986	0.571	0.033	-0.0005	0.699	0.033	-0.013
P (mg·kg ⁻¹)	PLS	331	740–1070	0.763	0.789	0.0900	0.803	0.684	0.119
	IPLS	17	1039–1054	0.698	0.882	0.0144	0.758	0.817	0.093
	Si-PLS	45	768–781 894–907 950–963 1058–1070	0.779	0.765	0.0284	0.804	0.738	0.048
K (cmol _c ·kg ⁻¹)	PLS	331	740–1070	0.878	0.096	0.0110	0.878	0.107	0.018
	IPLS	17	842–858	0.798	0.121	-0.0019	0.732	0.134	0.016
	Si-PLS	60	846–860 876–890 921–935 996–1010	0.910	0.083	0.0005	0.898	0.117	-0.047
Mg ²⁺ (cmol _c ·kg ⁻¹)	PLS	331	740–1070	0.711	0.550	0.0627	0.807	0.568	0.098
	IPLS	17	1055–1070	0.770	0.502	0.0008	0.776	0.608	-0.025
	Si-PLS	53	768–781 824–836 967–979 1019–1030	0.778	0.492	0.0160	0.761	0.654	0.109
Ca ²⁺ (cmol _c ·kg ⁻¹)	PLS	331	740–1070	0.783	2.439	0.2780	0.787	2.470	0.430
	IPLS	17	774–790	0.736	2.652	0.0119	0.734	2.8916	-0.9689
	Si-PLS	60	756–770 801–815 936–950 981–995	0.826	2.228	0.0642	0.666	3.0219	-0.4531
pH	PLS	331	740–1070	0.891	0.351	0.0401	0.785	0.3655	0.0636
	IPLS	17	876–890	0.688	0.564	0.0407	0.500	0.5301	0.1165
	Si-PLS	53	810–823 824–837 922–935 1019–1031	0.904	0.331	-0.0004	0.646	0.4760	-0.0369

* Par: quality parameter, Var: selected variable. Bold values represents best options.

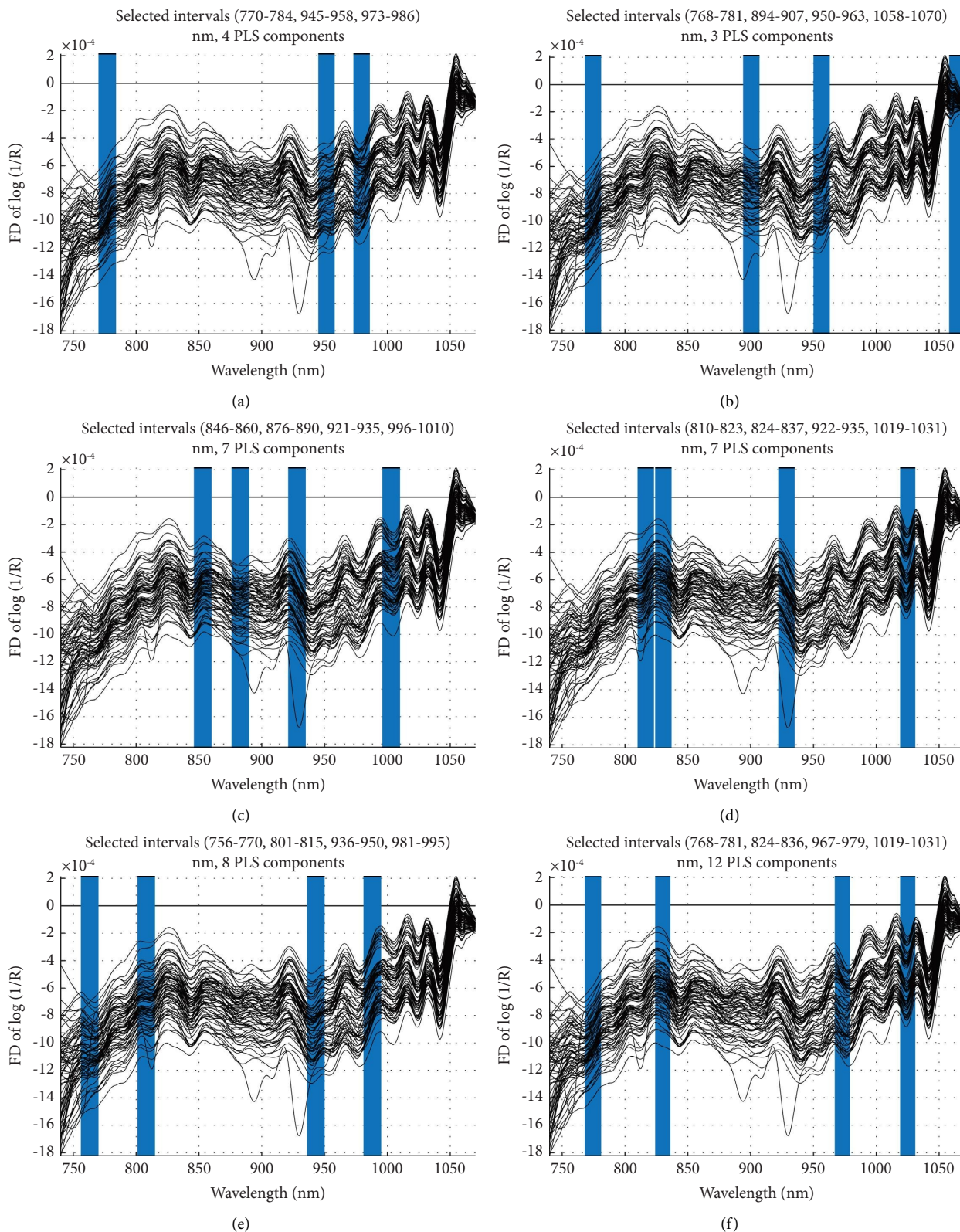


FIGURE 6: Optimal wavelength range selection of Si-PLS prediction for soil properties: (a) nitrogen, (b) phosphorus, (c) potassium, (d) pH, (e) calcium, and (f) magnesium.

converting the sun's energy into biomolecules, such as adenosine triphosphate (ATP) that drives biochemical reaction (photosynthesis). While for potassium, the optimal spectra range was found around 846–860, 876–890, 921–935, and 996–1010 nm with 7 PLS components in the second overtone region, which represents ArCH and CH in the electromagnetic wave as seen in Figure 6(c). Potassium supports transporting and forming sugars and starch through the plant. It is also vital in water regulation in plant. The total pH in soil is very important because it influences several soil factors affecting plant growth such as soil structure, soil bacteria, and nutrient availability among others and it is described as the master soil variable [28]. In this study, the optimum spectra tool selected four unique wavelengths for pH were 810–823, 824–837, 922–935, and 1019–1031 nm at 7 PLS components, as shown in Figure 6(d). These wavelengths represent C-H₃, C-H₂, C-H, and O-H corresponding to acidity [26]. It is particularly important to rapidly determine soil pH onsite as it readily gives a hint of the soil condition and the expected direction of many soil processes and can also be applied for nutrient cycling for plant nutrition and soil remediation [28]. For the optimum modelling of calcium and magnesium (Figures 6(e) and 6(f)), the Si-PLS method selected 756–770, 801–815, 936–950, and 981–995 nm at 8 PLS component and 768–781, 824–836, 967–979, and 1019–1030 nm at 12 PLS component, respectively. Ca²⁺ and Mg²⁺ are micronutrients required by plants for growth though in minute quantities. More specifically, Ca²⁺ is a component of plant cell that maintains cell walls strength and improves the fruit set and quality. Also, it has a positive effect on soil properties by improving the soil structure by enabling nitrogen-fixing bacteria on the roots of leguminous plants to capture atmospheric nitrogen into the soil. Mg²⁺, on the other hand, is an essential component of chlorophyll molecule; therefore, it is essential for photosynthesis in plant. Notably, Ca²⁺ and Mg²⁺ levels and their balances are two important factors affecting the growth of plant [29]. Furthermore, heavy metals do not absorb NIR; however, such constituents which do not absorb NIR radiation can be predicted owing to their correlation with other spectrally active parameters [30, 31]. Also, the findings in this study were similar to those of other researchers [31, 32]. And, as can be seen from Table 6, the results means that the model could be used acceptably for screening and other “approximate” calibration and the range 0.83–0.90 could be usable with caution for most applications, including research [33].

4. Conclusion

For the first time, this work has revealed that pocket-sized NIR spectroscopy in the range of 740–1080 nm could be used onsite to differentiate soils of different land used types and N, P, K, Mg²⁺, Ca²⁺, and pH simultaneously. The systematic comparison of different PLS calibration models for the prediction of soil health parameters revealed that the efficient spectral interval showed its superiority in measuring N, P, K, Mg²⁺, Ca²⁺, and pH in soils with the coefficient of correlation ranging from 0.699 to 0.898 and RMSEP between

0.033 and 3.02 in the prediction set. This means that for the models developed, the nitrogen model could be acceptable for very rough to rough screening, while the other could also be acceptable for screening, other “approximate” calibration, and usable with caution for most applications, including research [33]. These findings mean that portable NIR spectroscopy could be used for the rapid prediction of the soil status and quality parameters simultaneously with caution. However, more studies are needed to proof the robustness of the findings as it has a huge possibility of reducing the use of the time-wasting wet chemistry technique. It could also assist in making precision fertilizer application a reality in resource-poor communities, especially in developing countries. Also, this study only provides a feasibility study of using portable NIRS, and further studies are therefore required at different geographical locations and wide land-use types.

Data Availability

The data used to support the findings of this study are available upon request from the corresponding author.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Authors' Contributions

Ernest Teye developed the concept, provided resources, designed the experiment, wrote the first draft, and provided supervision. Charles L. Y. Amuah and Kwadwo K. Kusi analysed the spectra data and wrote and reviewed the first draft. Ransford Opoku Darko performed the experiment and analysed the soil chemistry data. Michael Miyittah, Rebecca Owusu, and Emmanuel Afutu revised the drafted manuscript and provided supervision. Kofi Atiah supervised the soil analysis and reviewed the first draft. Thomas Abindaw performed the soil analysis and analysed the data.

Acknowledgments

The support provided by the Directorate of Research, the Innovation and Consultancy (RSG/INT/CANS/2021/101) of the University of Cape Coast is highly acknowledged. The authors are also thankful to Mr Steve Adu, Osei Agyemang, and Mr Francis Padi Lamptey for their support during the sample collection and laboratory analysis.

References

- [1] B. Minasny, A. B. McBratney, and K. L. Bristow, “Comparison of different approaches to the development of pedotransfer functions for water-retention curves,” *Geoderma*, vol. 93, no. 3-4, pp. 225–253, 1999.
- [2] J. M. Soriano-Disla, L. J. Janik, R. A. Viscarra Rossel, L. M. Macdonald, and M. J. McLaughlin, “The performance of visible, near-and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties,” *Applied Spectroscopy Reviews*, vol. 49, no. 2, pp. 139–186, 2014.

- [3] M. Mariela Fuentes, "Nir spectroscopy: an alternative for soil analysis," *Communications in Soil Science and Plant Analysis*, vol. 43, 2012.
- [4] M. Todorova, A. M. Mouazen, H. Lange, and S. Atanassova, "Potential of near-infrared spectroscopy for measurement of heavy metals in soil as affected by calibration set size," *Water, Air, and Soil Pollution*, vol. 225, no. 8, pp. 1–19, 2014.
- [5] R. Zornoza, C. Guerrero, J. Mataix-Solera, K. Scow, V. Arcenegui, and J. Mataix-Beneyto, "Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils," *Soil Biology and Biochemistry*, vol. 40, no. 7, pp. 1923–1930, 2008.
- [6] M. Siddique, "Determination of N mineralization, total N and cation exchange capacity of soil through NIR spectroscopy for decision support in rice farming," *International Journal of Business, Management and Social Research*, vol. 1, no. 01, pp. 47–50, 2015.
- [7] B. Esquivel-Valenzuela, J. A. Cueto-Wong, C. O. Cruz-Gaistardo, A. Guerrero-Peña, A. Jarquín-Sánchez, and D. Burgos-Córdova, "Carbono orgánico y nitrógeno total en suelos forestales de México mediante espectroscopia VIS-NIR," *Revista Mexicana de Ciencias Farmacéuticas*, vol. 9, no. 47, pp. 295–313, 2018.
- [8] M. H. Todorova and S. L. Atanassova, "Near infrared spectra and soft independent modelling of class analogy for discrimination of Chernozems, Luvisols and Vertisols," *Journal of Near Infrared Spectroscopy*, vol. 24, no. 3, pp. 271–280, 2016.
- [9] D. Ertlen, D. Schwartz, M. Trautmann, R. Webster, and D. Brunet, "Discriminating between organic matter in soil from grass and forest by near-infrared spectroscopy," *European Journal of Soil Science*, vol. 61, no. 2, pp. 207–216, 2010.
- [10] T. Abindaw, "Influence of land use types on the distribution of selected soil properties in tropical soils of the coastal savanna zone," *Heliyon*, vol. 9, 2023.
- [11] J. Anderson and J. Ingram, *A handbook of methods*, Vol. 221, CAB International, Wallingford, Oxfordshire, 1993.
- [12] S. E. Allen, *Chemical Analysis of Ecological Materials*, Blackwell Scientific Publications, Hoboken, NJ, USA, 1974.
- [13] K. J. Maghanga, *Comparison of Soil Phosphorous Extraction by Olsen and Double Acid Methods in Acid Soils of Western Kenya*, Rongo University Press, Rongo, Kenya, 2015.
- [14] J. Rhoades, "Cation exchange capacity," *Methods of Soil Analysis: Part 2 Chemical and Microbiological Properties*, vol. 9, pp. 149–157, 1983.
- [15] C. An, X. Yan, C. Lu, and X. Zhu, "Effect of spectral pretreatment on qualitative identification of adulterated bovine colostrum by near-infrared spectroscopy," *Infrared Physics and Technology*, vol. 118, Article ID 103869, 2021.
- [16] D. Kusumaningrum, H. Lee, S. Lohumi, C. Mo, M. S. Kim, and B. K. Cho, "Non-destructive technique for determining the viability of soybean (*Glycine max*) seeds using FT-NIR spectroscopy," *Journal of the Science of Food and Agriculture*, vol. 98, no. 5, pp. 1734–1742, 2018.
- [17] E. Teye, X. Huang, H. Dai, and Q. Chen, "Rapid differentiation of Ghana cocoa beans by FT-NIR spectroscopy coupled with multivariate classification," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 114, pp. 183–189, 2013.
- [18] H. Buddenbaum and M. Steffens, "The effects of spectral pretreatments on chemometric analyses of soil profiles using laboratory imaging spectroscopy," *Applied and Environmental Soil Science*, vol. 2012, Article ID 274903, 12 pages, 2012.
- [19] C. L. Amuah, E. Teye, F. P. Lamptey, K. Nyandey, J. Opoku-Ansah, and P. O. W. Adueming, "Feasibility study of the use of handheld NIR spectrometer for simultaneous authentication and quantification of quality parameters in intact pineapple fruits," *Journal of Spectroscopy*, vol. 2019, Article ID 5975461, 9 pages, 2019.
- [20] S. Wold, M. Sjöström, and L. Eriksson, "PLS-regression: a basic tool of chemometrics," *Chemometrics and Intelligent Laboratory Systems*, vol. 58, no. 2, pp. 109–130, 2001.
- [21] E. Teye and X. Huang, "Novel prediction of total fat content in cocoa beans by FT-NIR spectroscopy based on effective spectral selection multivariate regression," *Food Analytical Methods*, vol. 8, no. 4, pp. 945–953, 2015.
- [22] Y. G. Sun, "Determination of the contents of free amino acids, caffeine and tea polyphenols in green tea by Fourier transform nearinfrared spectroscopy," *Chinese Journal of Spectroscopy Laboratory*, vol. 21, pp. 940–943, 2004.
- [23] F. Han, X. Huang, E. Teye, F. Gu, and H. Gu, "Nondestructive detection of fish freshness during its preservation by combining electronic nose and electronic tongue techniques in conjunction with chemometric analysis," *Analytical Methods*, vol. 6, no. 2, pp. 529–536, 2014.
- [24] X. Huang, E. Teye, J. D. Owusu-Sekyere et al., "Simultaneous measurement of titratable acidity and fermentation index in cocoa beans by electronic tongue together with linear and non-linear multivariate technique," *Food Analytical Methods*, vol. 7, no. 10, pp. 2137–2144, 2014.
- [25] M. Kodaira and S. Shibusawa, "Using a mobile real-time soil visible-near infrared sensor for high resolution soil property mapping," *Geoderma*, vol. 199, pp. 64–79, 2013.
- [26] Z. Xiaobo, Z. Jiewen, M. J. Povey, M. Holmes, and M. Hanpin, "Variables selection methods in near-infrared spectroscopy," *Analytica Chimica Acta*, vol. 667, no. 1–2, pp. 14–32, 2010.
- [27] M. Ruiz and P. Chen, "Use of the first derivative of spectral reflectance to detect mold on tomatoes," *Transactions of the ASAE*, vol. 25, no. 3, pp. 759–0762, 1982.
- [28] D. Neina, "The role of soil pH in plant nutrition and soil remediation," *Applied and Environmental Soil Science*, vol. 2019, Article ID 5794869, 9 pages, 2019.
- [29] H. X. Li, T. Zhou, and Y. Liu, "High potassium to magnesium ratio affected the growth and magnesium uptake of three tomato (*Solanum lycopersicum* L.) cultivars," *Journal of Integrative Agriculture*, vol. 17, no. 12, pp. 2813–2821, 2018.
- [30] E. Ben-Dor and A. Banin, "Near infrared analysis (NIRA) as a method to simultaneously evaluate spectral featureless constituents in soils," *Soil Science*, vol. 159, no. 4, pp. 259–270, 1995.
- [31] M. Chodak, "Application of near infrared spectroscopy for analysis of soils, litter and plant materials," *Polish Journal of Environmental Studies*, vol. 17, no. 5, 2008.
- [32] M. Chodak, B. Ludwig, P. Khanna, and F. Beese, "Use of near infrared spectroscopy to determine biological and chemical characteristics of organic layers under spruce and beech stands," *Journal of Plant Nutrition and Soil Science*, vol. 165, no. 1, pp. 27–33, 2002.
- [33] P. Williams, M. Manley, and J. Antoniszyn, *Near Infrared Technology: Getting the Best Out of Light*, African Sun Media, Stellenbosch, South Africa, 2019.