

Research Article

Use of Machine Learning Models for Prediction of Organic Carbon and Nitrogen in Soil from Hyperspectral Imagery in Laboratory

Manuela Ortega Monsalve D, Mario Cerón-Muñoz D, Luis Galeano-Vasco D, and Marisol Medina-Sierra D

GAMMA Research Group, Faculty of Agricultural Sciences, Antioquia University, Medellin, Antioquia, Colombia

Correspondence should be addressed to Manuela Ortega Monsalve; manuela.ortegam@udea.edu.co

Received 1 October 2023; Revised 1 December 2023; Accepted 11 December 2023; Published 20 December 2023

Academic Editor: Naveen Mishra

Copyright © 2023 Manuela Ortega Monsalve 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.

Organic carbon and total nitrogen are essential nutrients for plant growth. The presence of these nutrients at acceptable levels can create an optimal environment for the development of crops of interest. The application of spectroscopic techniques and the use of machine learning algorithms have made it possible to calibrate models capable of predicting the number of elements present in the soil. One of these techniques is hyperspectral imaging, which captures portions of the electromagnetic spectrum where the materials present in the soil can be differentiated due to the vibrations of chemical bonds. The objective of this research is to use statistical models to predict OC and N in soils from hyperspectral images. Transformations were applied to spectral and chemical data and the models used were Random Forest (RF) and Support Vector Machine (SVM). To select the best model, the values of the coefficient of determination (R^2), root mean square error of prediction (RMSEP), and the ratio of performance to deviation (RPD) were considered. For OC, the values found for the RF model were an R^2 of 0.87, an RMSEP of 0.10, and an RPD of 6.74; the SVM model presented an R^2 of 0.92, an RMSEP of 0.20, and an RPD of 3.56. For the variable N, the values found for the RF model were an R^2 of 0.79, an RMSEP of 0.03, and an RPD of 5.44; for the SVM model, they were an R^2 of 0.87, an RMSEP of 0.08, and an RPD of 2.76. The RF model showed a better fit for both variables. The SVM model also produced acceptable results. The results show that machine learning models are a good alternative for analysing soil-related variables.

1. Introduction

Precision agriculture is the science that develops techniques that facilitate the process of obtaining results in the field. These techniques include technologies such as Remote Sensing (RS) using satellites, in-field sensors using unmanned aerial vehicles (UAVs), and laboratory sensors [1]. Sensors can use hyperspectral imaging (HSI), which provides information through image pixels to identify the materials that make up the soil [2]. HSI images capture the portion of the electromagnetic spectrum corresponding to the visible region (400–800 nm) and a portion of the near infrared (NIR) and mid-infrared (MWIR) regions (800–2500 nm).

Radiation absorbed by chemical bonds containing carbon or other nonmetals (C=H, N=H, S=H, C=O, and O=H)

is concentrated in the NIR spectral region; therefore, HSI image data corresponding to this region provide information about the chemical composition of the sample [3]. When HSI images are acquired, they provide information in the form of a three-dimensional hypercube, usually with a large amount of data and multicollinearity between them [4]. However, the processing and extraction of such information is complex and requires the application of algorithms and multivariate transformations that are not widely used in general statistics. Nevertheless, HSI imaging offers several advantages, such as high speed and ease of data acquisition, and several machine learning algorithms are available to calibrate this technique [5].

Organic carbon (OC) and nitrogen (N) play a key role in plant nutrition, and the levels of these nutrients are synonymous with soil fertility [6]; however, many farmers are unaware of the OC and N content of their agricultural soils and consequently overapply amendments and fertilizers or, on the contrary, do not supply the soil with the nutrients it needs.

Recently, many authors have attempted to develop equations using sensors and machine learning to calibrate OC [7–9] and N [10–12] in soil. The information provided by such techniques has enabled researchers to determine and interpret various soil properties at both field and regional scales; detailed information can be obtained that allows quantitative analyses of soil constituents [1].

Machine learning models are components of a branch of artificial intelligence and can learn routines on their own. Supervised learning is the process of training a machine learning algorithm on questions and answers to make a prediction. These machine learning algorithms can be classified as classification or regression algorithms. One of the most used models is the Random Forest (RF), which is a series of decision trees that act as a set of classifiers; it can be used to solve both regression and classification problems [13]. Each of the decision trees in the RF model is constructed using different orders of data. One set of data is used for calibration and another for testing. At the end of the RF analysis, the regression prediction is calculated by averaging the individual trees, and a majority vote for the correct classification performs the model ranking. Others are Support Vector Machine (SVM) models, which are based on solving a convex quadratic optimization to obtain a globally optimal solution that overcomes the extreme dilemma of other machine learning techniques; SVM is a nonparametric model and is considered a classification model capable of dealing with high-dimensional data [14]. The training and evaluation of multivariate models allows the evaluation of variables with high dimensionality, where the variables have been masked and subjected to different transformations; however, these processes must be performed iteratively due to the large number of models that can be generated.

Transformations are mathematical equations or formulas that are applied to spectral data to reduce noise. Transformations can improve assumptions when applying statistical models and allow for easier comparisons among the data being analysed. Examples of transformations include Absorbance, Savitzky–Golay (SG), Detrending, Standard Normal Variance (SNV), and Multiplicative Scatter Correction (MSC). Model fit or performance factors are the mathematical criteria evaluated after the statistical model is applied to determine its acceptability. One of the fit factors for machine learning models is the coefficient of determination or R^2 (equation (1)) which relates the sum of error squares to the sum of total squares and indicates the proportion of variance in the response variable that is explained by the predictor variables:

$$R^{2} = 1 - \frac{\sum_{i=1}^{n} (\operatorname{Obs}_{i} - \operatorname{Pred}_{i})^{2}}{\sum_{i=1}^{n} (\operatorname{Obs}_{i} - \overline{\operatorname{obs}})^{2}}.$$
 (1)

Another factor is the root mean square error of prediction or RMSEP (equation (2)) and indicates the difference between the predicted and observed values:

$$\text{RMSEP} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\text{Obs}_i - \text{Pred}_i)^2}.$$
 (2)

The ratio of performance to deviation or RPD (equation (3)) is given by the standard deviation (*sd*) of the observed data over the RMSEP. RPD values greater than 3 are considered excellent in agricultural applications; values greater than 2 indicate good model performance [15]. Authors such as Wadoux et al. [16] consider an RPD >2 to be a good model in soil applications.

$$RPD = \frac{sd}{RMSEP}.$$
 (3)

2. Materials and Methods

2.1. Study Area. Soil sampling was carried out in the department of Antioquia, Colombia, specifically in different subregions and on farms growing flowers, cacao, and pastures for beef and dairy cattle (Figure 1). Within the department, there are different thermal soils classified as high, medium, and low tropic, resulting in soils with highly variable physical and chemical characteristics. Sample processing and image acquisition were carried out at the Faculty of Agricultural Sciences of the University of Antioquia. 1998 soil samples were collected at a depth of 15 cm. The samples were collected between the years 2020 and 2023.

2.2. Chemistry Data. Each of the samples collected contained two bags of soil. This material was mixed and homogenized to ensure sample uniformity. Half of each sample was processed (dried, sieved to 2 mm, and stored) in the laboratory. Drying was performed in a forced air oven at a temperature of 40°C for 48 h. The other half of each soil sample was sent to a wet chemistry laboratory where all soil nutrients were analysed by the conventional method. Results were obtained for the soil chemical variables OC and N, which were analysed using the Walkley–Black and Kjeldahl techniques, respectively. These analyses were used to calibrate the HSI cameras to the data.

2.3. Hyperspectral Image Data Acquisition. Dry soil samples with a particle size of 2 mm were placed in a 10 and 20 cm³ tray. Reflectance values were corrected using a Zenith Lite TM 50% R SG31XX diffuse reflectance target. This target was placed at the front of the dish so that the cameras captured it at the beginning of the procedure. Two cameras were used to capture the images: a Hyspex [®] Baldur V-1024 N (VNIR) with a spectral resolution of 5.4 nm, a spatial resolution of 3289-1024 pixels, and coverage of the spectral range from 485 to 955 nm, and a Hyspex Baldur S-384 N (SWIR) with a spectral resolution of 5.45 nm, a spatial resolution of 1216-384 pixels, and coverage of the spectral range from 951 to 2517 nm.

2.4. Data Preprocessing. Image preprocessing was performed using the Python 3.8.2 programming language [17] and the *SpectralPy*, *Spectral*, and *NumPy* libraries. The region of interest



FIGURE 1: Map points where the soil samples were collected in the department of Antioquia, Colombia.

(ROI) was selected by coordinates within the image. A region was selected in the centre of the image where the edges of the dish were not included and where the sample was homogeneous. An average of the pixels of each band included in the ROI was calculated. Pixels with reflectance less than 0.10 and greater than 0.90 were masked to eliminate shadows and saturated pixels. The overlapping bands were determined, and the transition zone corresponding to band 951 was eliminated. The change between bands 955 and 957 was analysed, and spectra with a change greater than 0.097 were eliminated.

3. Training and Test of Statistical Models

The raw data of the spectra were reflectance values. The spectral signature of the soil samples is shown in Figure 2.

The OC and N variables were transformed by \sqrt{OC} and \sqrt{N} , which is a transformation that has a moderate effect and is weaker than other transformations; it is used to reduce the asymmetry to the right. The spectral data were transformed into absorbance values; other transformations were then applied, including SNV, MSC, first derivative of SG, and detrend. The Mahalanobis distance was applied to the spectral data to detect outliers. No outliers were found, so all data were retained. For the RF model, 500 and 800 trees were used; for the SVM model, radial and linear methods were used. 75% of the data was used for training and 25% for testing the statistical models.

The models were run using the statistical software R-Project [4.2.2]. The *randomForest* and *caret* libraries [3] were used to run the RF model, and the *e1071* library [18] was used for the SVM model. The performance of the models was evaluated based on the R^2 , RMSEP, and RPD metrics and the absence of overuse between training and test data. Figure 3 shows the methodology applied to the soil samples and the spectral information.



FIGURE 2: Spectral signature of the collected soil samples at reflectance values.

4. Results and Discussion

4.1. Characterization of the Variables Used in the Study. According to the descriptive statistics applied to the data, the mean was found to be $2.92\% \pm 2.72$ and $0.31\% \pm 0.23$ for OC and N, respectively. For the transformed data, the mean and median values are more similar, where the standard deviation of the data is significantly reduced. The square root transformation of the soil variables is expected to significantly improve the performance of the statistical models. The results are shown in Table 1.

Neither variable's data show a normal distribution. Most of the data are on the far left of the histogram. The variables represent nonsymmetric data (Figure 4).

The average of the OC variable can refer to soils in warm climates with ideal values or, on the contrary, to soils in cold climates with low values. This research included soils belonging to all thermal soils; therefore, the OC values must be analysed according to the area studied to determine whether they are high, medium, or low. The maximum OC values found are associated with the high tropical zones of the department, since the rate of mineralization of organic matter is inversely proportional to temperature. The average value of N corresponds to overfertilized soil, since the normal range for this nutrient is 0.1-0.2. The analysis of the data by subregions showed that the N values are high in some areas of the high tropics of the department. This result may be related to the high use of nitrogenous fertilizers in dairy cattle production.

5. Statistical Models

In total, 96 statistical models were obtained: 48 models by RF and 48 models by SVM for the two soil variables. The value of 96 was obtained by combining the two types of models and different combinations of transformations and methods for the two variables. For the models by RF, 500 and 800 trees were used; however, in the results, the internal validation method "cv" was used in the results, which is a method to verify the effectiveness of a machine learning model. Its function is to select a part of the dataset that is not used to train the model, to be used later as test data. For the SVM models, the linear and radial methods were used. Only the models with the highest performance for each of the soil variables are shown.



FIGURE 3: Sequential chart of the methodology used in this research.

	FABLE	1:	Descriptive	statistics	of	the	investigated	soil	variable	es
--	--------------	----	-------------	------------	----	-----	--------------	------	----------	----

Component	Mean	Median	Min	Max	sd	CV
OC	2.92	2.92	0.08	15.5	2.72	93.2
\sqrt{OC}	1.55	1.36	0.28	3.93	0.70	45
Ν	0.31	0.23	0.01	1.28	0.23	74.8
\sqrt{N}	0.52	0.48	0.10	1.13	0.19	35.5



FIGURE 4: Histogram of the behaviour of the data of the variables OC and total N.

Table 2 shows the results of the RF and SVM models for the OC variable. In general, high fit values were obtained with all transformations and RF models. In all models, a better fit was obtained when the soil OC variable was transformed. In addition, better performance was obtained for all models and transformations using 800 trees. The model that showed better performance was the application of the absorbance transformation and \sqrt{OC} , where an R^2 of 0.87 was obtained for the test data group, the RMSEP was 0.10, which is one of the lowest values obtained in the present study, and the RPD was 6.74, which was the highest value for the models studied. In addition, the model did not show overfitting as R^2 of the test data was the same as that of the training data. Although the coefficients of determination are lower than those of the SVM models, excellent fits were obtained for RMSEP and RPD.

The RF model performed better than the SVM model for the OC variable. None of the models showed overfitting for the validation data. The best performing RF model was the one that used the transformation of the first derivative of the SG of the spectral data and the transformation for \sqrt{OC} .

Based on a literature review, Vargas et al. [19] concluded that the RF and SVM algorithms are useful for determining the OC in soil. These algorithms have also been studied by other authors. Pouladi et al. [20] used RF models to determine the prediction of soil organic matter, which can be directly related to the OC content through a conversion factor. They found an R^2 of 0.89 and an RMSEP of 4.20. Their relatively large error may be because the study was conducted with relatively few samples. The RMSEP values found by these authors are much higher than those found in the present investigation. Yang et al. [21] have also conducted studies using RF models to determine the OC in harsh climates, where the maximum fit of the model was 0.71 and the RMSEP was 0.48, which are still close to the fit obtained in the present work. Hong et al. [5], who used HSI images in conjunction with RF models to determine OC in soil, obtained an R^2 value of 0.79 and an RMSEP of 0.18, like the values obtained in the present investigation. The research carried out by Nawar and Mouazen [22] shows that RF models are an excellent method for calculating OC and N in soil; these authors found fits as high as 0.97 using crossvalidation of the algorithm, an RPD of 5.58, and an RMSEP of 0.01; these values were found using a set of 528 data points distributed over several European countries.

Table 3 shows the results obtained for the N variables when the RF and SVM models were applied. Although the RF models had a lower R^2 , the RPD obtained was the best among all models and the RMSEP was the lowest among the models. Therefore, the model that showed that the highest performance and its fitting factors are excellent when used with the first derivative SG transformation. The results obtained for this model were an R^2 of 0.79 for the training and test data, an RMSEP of 0.03, and an RPD of 5.44.

For the soil variable *N*, a better performance of the SVM model was obtained using a combination of the first derivative SG transformation and \sqrt{N} with the radial method.

The SVM algorithm gave the best results for the determination of OC and N when combined with different transformations. For this model, Datta et al. [23] obtained a good fit when using the bands with the highest correlation in the spectrum for the OC variable, obtaining an R^2 of 0.90, which is like the R^2 value obtained in the present study. However, Aldana et al. [24] fitted the SVM model and obtained an R^2 of 0.95 and an RMSEP of 0.21 for OC, which confirms our results for the same variable. Meng et al. [25] also applied SVM models and obtained an R^2 of 0.80, an RMSEP of 3.20, and an RPD of 1.71. Although their coefficient of determination is like that found in the present study, the other fit values differ significantly from those in the present work, possibly due to the difference in the number of samples between the studies. Authors such as Vargas et al. [19], through a systematic review, concluded that SVM models are the most suitable machine learning algorithms to determine variables such as organic matter

TABLE 2: Results of adjustment factors for the SVM and RF models for the soil OC variable.

Model	Method	Transformation	R^2 train	R^2 test	RPD
SVM	Radial	$1^{\circ} \text{ der SG} + \sqrt{\text{OC}}$	0.87	0.92	3.56
SVM	Radial	$\text{Det} + \sqrt{\text{OC}}$	0.85	0.90	3.23
SVM	Linear	1° der SG + \sqrt{OC}	0.88	0.90	3.13
RF	800 trees	1° der SG + \sqrt{OC}	0.87	0.87	6.74
RF	800 trees	$MSC + \sqrt{OC}$	0.82	0.82	5.95
RF	800 trees	$\text{Det} + \sqrt{\text{OC}}$	0.83	0.83	5.94

TABLE 3: Results of adjustment factors for the SVM and RF models for the soil total N variable.

Model	Method	Transformation	R^2 train	R^2 test	RPD
SVM	Radial	1° der SG + \sqrt{N}	0.81	0.87	2.76
SVM	Linear	1° der SG + \sqrt{N}	0.82	0.84	2.54
SVM	Linear	$\text{Det} + \sqrt{N}$	0.80	0.83	2.40
RF	800 trees	1° der SG + \sqrt{N}	0.79	0.79	5.44
RF	800 trees	$\text{Det} + \sqrt{N}$	0.76	0.76	5.03
RF	800 trees	$MSC + \sqrt{N}$	0.75	0.75	5.00

and N in soils because they achieve better performance than other multivariate models.

Figure 5 shows all predicted and fitted data obtained using the RF algorithm for the two variables of interest. These plots correspond to the models with the spectral data transformed using the first derivative SG transformation and the square root of the soil variables. That is, it refers to the best RF model observed for each variable.

5.1. Correlation of Spectral Bands. After applying the model, we performed a correlation analysis between the spectral bands and the OC and N variables (Table 4). The correlation analysis was applied to the transformed and untransformed databases, and the bands that gave a better result, with correlations above 0.60 and -0.60, were selected. A small number of bands were found to correlate with nutrients. The detrend transformation resulted in a greater number of band ranges for OC and N. For OC and N, a strong correlation was observed between the band ranges from 500 to 900 nm, which includes portions of the visible and NIR regions, and from 1300 to 1950 nm, which is in the NIR region.

The correlation between the spectral bands and the OC content in the soil is related to the presence of carbon and other elements. In their study carried out to determine the OC in soil using hyperspectral images, Aichi et al. [26] found a high correlation of OC with the range of bands between 400 and 680 nm. In addition, they correlated the concave spectral signature of the soil with a high OC content between the bands at 400 and 950 nm, which was corroborated by the present study because the set of all spectral signatures of the soil resulted in this behaviour. Meng et al. [25] studied the behaviour of soil OC and found that the bands most sensitive to the presence of carbon are in the visible region of the spectrum, which confirms the results of the present investigation, where most of the correlated bands were also found in the visible region. The presence of OC in the visible region of the spectrum can also lead to strong correlations



FIGURE 5: Plots of predicted and observed values of RF for OC and total N.

TABLE 4: Bands with moderately high correlations for the soil OC and total N variables.

Transformation	Bands correlated (nm)
Reflectance	566-879
Absorbance	544-934
SNV	495-506, 733-842, 1410, 1415
Detrend	490-550, 652-836, 1377-1426, 2059-2081
MSC	1410, 1415, 1417
1° der SG	528-593, 815, 932, 1923-1933

because of the relationship between the color of the soil (dark) and its presence in large amounts [20, 27]. Several authors have found a significant relationship between wavelength and the OC of soil. Strong correlations were found in the visible region: in the bands from 550 to 700 nm [28] and between the bands at 526 and 587 nm [29]. These findings support our results, as we found medium and high correlations in the reflectance data in the 566–852 nm spectral range. A high correlation was also observed between OC and reflectance produced near 490 nm [30]. This band showed a high correlation in our research; however, it was detected when the detrend transformation was applied to the spectral data.

Regarding the bands of the spectrum correlated with the variable N, authors such as Patel et al. [31] observed strong absorption peaks near 1400, 1900, 2200, and 2350 nm. In the present study, correlated bands were found between 1412 and 1420 nm, in addition to some bands near 1900 nm. Also, Tahmasbian et al. [32] also found bands highly correlated with the N content, such as the bands between 400 and 900 nm. These bands were also found to be correlated in our research.

6. Conclusions

The results of this study show that the RF and SVM machine learning models can be useful for predicting soil OC and N variables. The SVM model behaves better than the RF model, as indicated by the better R^2 , RMSEP, and RPD values of the

fit to the SVM model. Using the spectral band transformations in this case, the absorbance and the first derivative of SG in combination with the machine learning models can result in a better fit and more accurate prediction of the OC and N data. Few spectral bands with high correlation under the study variables were observed; however, we found certain bands where the correlation is high. These band ranges should allow researchers to work with specific areas of the spectrum in relation to different soil nutrients. The use of HSI can help reduce the use of conventional techniques, which currently have numerous drawbacks.

Data Availability

The spectral and wet chemistry data for the soils used to support the findings of this study have not been made available because they are not directly owned by the authors.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This publication was possible thanks to the database provided by the project "Development and establishment of the Center for Agrobiotechnological Development of Innovation and Territorial Integration El Carmen de Viboral Antioquia Occidente (CEDAIT)" Expert System Component with resources from the General Royalty System and the Government of Antioquia. This publication and the financial support of the MSc students were possible thanks to the project "Design and validation of predictive models to determine Cation Exchange Capacity (CEC), Organic Matter (OM) and Nitrogen (N) in soils from hyperspectral images" through the agreement 2022-7204, financed by the University of Antioquia Foundation.

References

- Y. Ge, J. A. Thomasson, and R. Sui, "Remote sensing of soil properties in precision agriculture: a review," *Frontiers of Earth Science*, vol. 5, no. 3, pp. 229–238, 2011.
- [2] D. Malonakis, Detection Algorithms for Hyperspectral Imaging Applications, Lincon Laboratory, Lexington, MA USA, 2002.
- [3] M. Kuhn, "Building predictive models in R using the caret package," *Journal of Statistical Software*, vol. 28, no. 5, pp. 1–26, 2008.
- [4] C. Guerrero, R. Zornoza, A. Pérez, J. Mataix, I. Gómez, and F. García, "Uso de la espectroscopía en el infrarrojo cercano (NIR) para la estimación rápida del carbono orgánico y la respiración basal en suelos forestales," *Cuaderno de la Sociedad Española de Ciencias Forestales*, vol. 25, pp. 209–214, 2008.
- [5] Y. Hong, S. Chen, Y. Chen et al., "Comparing laboratory and airborne hyperspectral data for the estimation and mapping of topsoil organic carbon: feature selection coupled with random forest," *Soil and Tillage Research*, vol. 199, Article ID 104589, 2020.
- [6] J. Roberts, A. Power, J. Chapman, S. Chandra, and D. Cozzolino, "A Short update on the advantages, applications and limitations of hyperspectral and chemical imaging in food authentication," *Applied Sciences*, vol. 8, no. 4, p. 505, 2018.
- [7] C. Lefevre, F. Rekik, V. Alcantara, and L. Wiese, Soil Organic Carbon: The Hidden Potential, Food and Agriculture Organization of the United Nations (FAO), Rome, Italy, 2017.
- [8] S. Xu, M. Wang, and X. Shi, "Hyperspectral imaging for highresolution mapping of soil carbon fractions in intact paddy soil profiles with multivariate techniques and variable selection," *Geoderma*, vol. 370, Article ID 114358, 2020.
- [9] L. Guo, X. Sun, P. Fu et al., "Mapping soil organic carbon stock by hyperspectral and time-series multispectral remote sensing images in low-relief agricultural areas," *Geoderma*, vol. 398, Article ID 115118, 2021.
- [10] S. Wang, K. Guan, C. Zhang et al., "Using soil library hyperspectral reflectance and machine learning to predict soil organic car- bon: assessing potential of airborne and spaceborne optical soil sensing," *Remote Sensing of Environment*, vol. 271, Article ID 112914, 2022.
- [11] L. Lin and X. Liu, "Mixture-based weight learning improves the random forest method for hyperspectral estimation of soil total nitrogen," *Computers and Electronics in Agriculture*, vol. 192, Article ID 106634, 2022.
- [12] J. Ma, J. Cheng, J. Wang et al., "Rapid detection of total nitrogen content in soil based on hyper- spectral technology," *Information Processing in Agriculture*, vol. 9, no. 4, pp. 566– 574, 2022.
- [13] L. Breinam, "Random forest," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [14] M. Sheykhmousa, M. Mahdianpari, H. Ghanbari, F. Mohammadimanesh, P. Ghamisi, and S. Homayouni, "Support vector machine versus random forest for remote sensing image classification: a meta-analysis and systematic review," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 6308–6325, 2020.
- [15] W. Saeys, A. M. Mouazen, and H. Ramon, "Potential for onsite and online analysis of pig manure using visible and

near infrared reflectance spectroscopy," Biosystems Engineering, vol. 91, no. 4, pp. 393-402, 2005.

- [16] A. Wadoux, B. Malone, B. Minasny, M. Fajardo, and A. McBratney, *Soil Spectral Inference with R*, Springer, Berlin/ Heidelberg, Germany, 2021.
- [17] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual*, CreateSpace, Scotts Valley. CA, USA, 2020.
- [18] D. Meyer, E. Dimitriadou, K. Hornik et al., "Functions of the department of statistics, probability theory group (Formerly: e1071)," *R package version*, vol. 1, no. 2, 2023.
- [19] M. Vargas-Zapata, M. Medina-Sierra, L. F. Galeano-Vasco, and M. F. Cerón-Muñoz, "Algoritmos de aprendizaje de máquina para la predicción de propiedades fisicoquímicas del suelo mediante información es pectral: una revisión sistemática," *Revista de Investigación, Desarrollo e Innovación*, vol. 12, no. 1, pp. 107–120, 2022.
- [20] N. Pouladi, A. Moller, S. Tabatabai, and M. H. Greve, "Mapping soil organic matter contents at field level with Cubist, Random Forest and kriging," *Geoderma*, vol. 342, pp. 85–92, 2019.
- [21] R. M. Yang, G. L. Zhang, F. Liu et al., "Comparison of boosted regression tree and random forest models for mapping topsoil organic carbon concentration in an alpine ecosystem," *Ecological Indicators*, vol. 60, pp. 870–878, 2016.
- [22] S. Nawar and A. Mouazen, "Comparison between random forests, artificial neural networks and gradient boosted machines methods of on-line vis-NIR spectroscopy measurements of soil total nitrogen and total carbon," *Sensors*, vol. 17, no. 10, p. 2428, 2017.
- [23] D. Datta, M. Paul, M. Murshed, S. W. Teng, and L. Schmidtke, "Soil moisture, organic carbon, and nitrogen content prediction with hyperspectral data using regression models," *Sensors*, vol. 22, no. 20, p. 7998, 2022.
- [24] E. Aldana-Jague, G. Heckrath, A. Macdonald, B. van Wesemael, and K. van Oost, "UAS-based soil carbon mapping using VIS-NIR (480-1000 nm) multi-spectral imaging: potential and limitations," *Geoderma*, vol. 275, pp. 55–66, 2016.
- [25] X. Meng, Y. Bao, J. Liu et al., "Regional soil organic carbon prediction model based on a discrete wavelet analysis of hyperspectral satellite data," *International Journal of Applied Earth Observation and Geoinformation*, vol. 89, pp. 102111– 111, 2020.
- [26] H. Aichi, Y. Fouad, D. Causeur, and C. Walter, "Organic carbon and total iron effect on soil Vis-SWNIR spectra and quantification of their contents using PLSR models," *Communications in Soil Science and Plant Analysis*, vol. 51, no. 9, pp. 1253–1267, 2020.
- [27] H. M. Bartholomeus, M. E. Schaepman, L. Kooistra, A. Stevens, W. B. Hoogmoed, and O. S. P. Spaargaren, "Spectral reflectance based indices for soil organic carbon quantification," *Geoderma*, vol. 145, no. 1-2, pp. 28–36, 2008.
- [28] L. S. Galvao and I. Vitorello, "Role of organic matter in obliterating the effects of iron on spectral reflectance and colour of Brazilian tropical soils," *International Journal of Remote Sensing*, vol. 19, no. 10, pp. 1969–1979, 2010.
- [29] I. Melendez-Pastor, J. Navarro-Pedreño, I. Gómez, and M. Koch, "Identifying op timal spectral bands to assess soil properties with VNIR radiometry in semi-arid soils," *Geoderma*, vol. 147, no. 3-4, pp. 126–132, 2008.

- [30] A. M. Mouazen, M. R. Maleki, J. De Baerdemaeker, and H. Ramon, "On-line measurement of some selected soil properties using a VIS-NIR sensor," *Soil and Tillage Research*, vol. 93, no. 1, pp. 13–27, 2007.
- [31] A. K. Patel, J. K. Ghosh, S. Pande, and S. U. Sayyad, "Deeplearning based approach for estimation of fractional abundance of nitrogen in soil from hyperspectral data," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 6495–6511, 2020.
- [32] I. Tahmasbian, Z. Xu, S. Boyd et al., "Laboratory-based hyperspectral image analysis for predicting soil carbon, nitrogen, and their isotopic compositions," *Geoderma*, vol. 330, pp. 254–263, 2018.