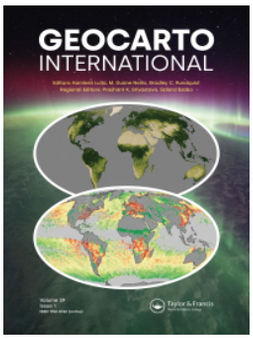

Geocarto International

ISSN: 1010-6049 (Print) 1752-0762 (Online) Journal homepage: www.tandfonline.com/journals/tgei20

Advancing soil property prediction: integration of Vis-NIR sensors' data and machine learning models for rapid and accurate analysis

G. S. Tagore, Devid Kumar Sahu, Bhabani Prasad Mondal, Suman Dutta, Sumanta Das, Arghya Chattopadhyay, Rahul Sadhukhan, R. K. Nema, Mohamed E. Fadl, Costanza Fiorentino, Paola D'Antonio & Ali R. A. Moursy

To cite this article: G. S. Tagore, Devid Kumar Sahu, Bhabani Prasad Mondal, Suman Dutta, Sumanta Das, Arghya Chattopadhyay, Rahul Sadhukhan, R. K. Nema, Mohamed E. Fadl, Costanza Fiorentino, Paola D'Antonio & Ali R. A. Moursy (2026) Advancing soil property prediction: integration of Vis-NIR sensors' data and machine learning models for rapid and accurate analysis, Geocarto International, 41:1, 2656536, DOI: 10.1080/10106049.2026.2656536

To link to this article: https://doi.org/10.1080/10106049.2026.2656536

GEOCARTO INTERNATIONAL
2026, VOL. 41, NO. 1, 2656536
https://doi.org/10.1080/10106049.2026.2656536

Taylor & Francis
by informa

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Advancing soil property prediction: integration of Vis-NIR sensors' data and machine learning models for rapid and accurate analysis

G. S. Tagore[a,1], Devid Kumar Sahu[a], Bhabani Prasad Mondal[b,1], Suman Dutta[c], Sumanta Das[c], Arghya Chattopadhyay[d], Rahul Sadhukhan[e], R. K. Nema[f], Mohamed E. Fadl[g], Costanza Fiorentino[h], Paola D'Antonio[h] and Ali R. A. Moursy[i]

[a]Department of Soil Science and Agricultural Chemistry, College of Agriculture, Jawaharlal Nehru Krishi Vishwa Vidyalaya (JNKVV), Jabalpur, Madhya Pradesh, India; [b]Department of Soil Science, Bihar Agricultural University, Sabour, Bhagalpur, Bihar, India; [c]Ramkrishna Mission, Vivekananda Educational and Research Institute, Kolkata, West Bengal, India; [d]Department of Soil Science and Agricultural Chemistry, Institute of Agricultural Science, Banaras Hindu University, Varanasi, Uttar Pradesh, India; [e]Department of Agronomy, Multi-Technology Testing Centre and Vocational Training Centre and College of Horticulture, Imphal, Mizoram, India; [f]PI-NAHEP-CAAST, ICAR, College of Agricultural Engineering, Jawaharlal Nehru Krishi Vishwavidyalaya, Madhya Pradesh, India; [g]Division of Scientific Training and Continuous Studies, National Authority for Remote Sensing and Space Sciences (NARSS), Cairo, Egypt; [h]School of Agricultural, Forestry and Environmental Sciences (SAFE), University of Basilicata, Potenza, Italy; [i]Soil and water Department, Faculty of Agriculture, Sohag University, Sohag, Egypt

ABSTRACT

In Madhya Pradesh, agricultural soils face constraints, such as nutrient depletion, low fertility in Alfisol, drought and drainage issues in Vertisol, salinity, waterlogging and erosion-driven degradation. Sustainable soil management requires timely soil testing and spatial variability assessment, yet conventional methods are slow, costly and poorly accessible. This study examines visible–near infrared spectroscopy (Vis–NIR) with machine learning to estimate soil pH, electrical conductivity (EC) and soil organic carbon (SOC) from 2216 georeferenced surface samples. Laboratory results showed pH 4.55–8.39, EC 0.03–0.97 dS m^{-1}, and SOC 1.05–11.25 g kg^{-1}, with duplicates confirming high precision. Algorithms tested included partial least squares regression (PLSR), SVMR, ANN, RF, CatBoost, ELM, XGBoost and LightGBM. PLSR delivered the most accurate predictions for pH and SOC, while EC showed moderate predictability. Future research should combine mid-infrared spectroscopy, soil covariates, calibration and global spectral libraries to improve model transferability and support scalable soil health monitoring.

ARTICLE HISTORY
Received 4 June 2025
Accepted 1 April 2026

KEYWORDS
Soil spectroscopy; machine learning; soil properties; prediction accuracy; Vis–NIR.

1. Introduction

Soil is considered a complex, structured natural system that is difficult to fully understand. Its components are called 'aggregates', which are dynamic and have complex matrices compared to other chemical, synthetic or architectural forms. Besides, soil consists of the solid phase, which includes inorganic and organic materials; the solution phase and the soil atmosphere. Moreover, in its surface or in solution, physico-chemical and biological interactions occur in non-linear complex ways (Yudina and Kuzyakov 2023). Still, researchers cannot completely understand how to deal with such material. Furthermore, soil formation is caused by several interactions among time, climate, topography, organisms and parent materials (Amundson 2021). When talking about soil complexity, the main reason is the special processes, such as soil carbon sequestration, which is particularly affected by molecular diversity, environmental conditions and agricultural activities (Lehmann et al. 2020). Therefore, understanding these intricate relations is necessary to achieve effective land management and promote sustainable agriculture (Hou et al. 2020; Moursy et al. 2022).

CONTACT Bhabani Prasad Mondal bpmondal1@gmail.com; Costanza Fiorentino costanza.fiorentino@unibas.it
[1] Authors have made equal contributions.
Supplemental data for this article can be accessed online at https://doi.org/10.1080/10106049.2026.2656536.

Soil physico-chemical parameters, which are directly related to crop yield, are important for soil health evaluation. Soil pH is considered one of the most important environmental indicators that can be used to determine the availability and mineralization of nutrients (Maurya et al. 2020). The soil organic carbon (SOC) availability enhances the soil physico-chemical and biological characteristics and enhances soil, water and nutrient retention (Mesfin et al. 2021; Thabit et al. 2023). High soil salinity negatively affects root respiration, soil organic matter (SOM) decomposition and nitrification processes (Singh 2016; Abd-Elazem et al. 2024). Moreover, high salinity affects the stability of soil aggregates and increases the possibility of wind and water erosion. Thus, the variability in pH, electrical conductivity (EC) and soil organic carbon (SOC) in different soils is greatly influenced (Dhaliwal et al. 2021).

Soil analysis plays several crucial roles in various contexts, such as agriculture, where analyzing the soil pH, EC and SOC can help in fertilizer recommendations and adjusting the crop requirements (Nadporozhskaya et al. 2022). Additionally, regular soil analysis aids in identifying potential deficiencies and adjusting management practices to enhance soil quality and sustainability, support the objective of feeding a growing population as well as reducing ecological impacts (Lehmann et al. 2020).

Although traditional methods of soil analysis remain valuable tools for sustainable land use, effective agriculture and the most reliable (Piccini et al. 2024), they have several drawbacks, such as cost and time, specifically when extensive sampling is required across large areas. The second limitation is soil spatial variability, whereas a single soil sample may not fully represent the entire area. Moreover, soil sampling, transportation or analysis errors affect the results as proper protocols need to be followed. The frequent analysis is needed because of the continuous change in the soil properties. The soil analysis using a lot of chemical solutions is hazardous to humans and the environment (Ng et al. 2022). Moreover, conventional soil analysis is a labour-intensive process, often requires specialized equipment and expertise. Additionally, the diagnostic value of total soil analysis is limited, as it does not account for differences in nutrient availability across various soil fractions. The same soil sample can be different in its reports if analyzed in different soil testing laboratories (Nadporozhskaya et al. 2022). The use of soil extractants is empirical and only as reliable as the calibration curve, which must be specifically developed for the soils being tested. Furthermore, predicting plant nutrient uptake and responses to fertilization based solely on soil analysis can be challenging due to the complex interaction of soil properties, environmental factors as well as plant varieties (Potdar et al. 2021).

Therefore, there is an essential need for a rapid, cost-effective, accurate, non-laborious and eco-friendly alternative to the traditional protocols of soil analysis. Soil spectroscopy technology was found to be a potential technique to estimate soil parameters. This technique offers several advantages over traditional soil analysis methods, such as rapid assessment, whereas soil spectroscopy provides near-instant results, reducing the time required for analysis compared to traditional laboratory-based methods (Thabit and Moursy 2023). Another advantage is the high-spatial resolution, whereas spectroscopic techniques allow for detailed mapping of soil properties across a field, capturing spatial variability more effectively (Liu et al. 2023). These techniques are cost-effective, with initial setup costs; long-term savings come from reduced sampling efforts and faster decision-making (Viscarra Rossel et al. 2022). Diffuse reflectance spectroscopy (DRS) covers a range of soil parameters simultaneously, including SOM, minerals, moisture, etc. Visible–near infrared (Vis–NIR) spectroscopy became a suitable substitution for wet-chemistry analysis because it is rapid and non-destructive and allows multiple soil properties to be estimated simultaneously using a single spectrum. The spectroscopic estimation does not require destructive sampling, preserving soil integrity for ongoing studies. The portability of spectroscopic equipment also enables onsite analysis, making it a practical tool for precision farming and site-specific nutrient management, ultimately supporting sustainable agricultural practices and better decision-making for land management (Piccini et al. 2024). However, the hyperspectral remote sensing (HRS) technique plays vital functions in estimating soil properties as well as creating soil maps using different types of sensors (Shi et al. 2023). There are popular sensors that are used for estimating various soil properties, such as optical sensors, which use visible, near infrared, and mid-infrared (vis-NIR-MIR) light to analyze soil reflectance. These devices include an analytical spectral device (ASD), which is used for estimating various soil parameters in the Vis–NIR spectral range between 350 and 2500 nm (Ahmadi et al. 2021; Ai et al. 2022; George et al. 2024; Francos et al. 2024; Moursy 2025).

Machine learning (ML) techniques are popularly employed for predicting soil parameters utilizing the Vis–NIR spectroscopy information (Tavakoli et al. 2023). Several algorithms have been explored, including support vector machine regression (SVMR), partial least squares regression (PLSR), artificial neural networks (ANNs), random forest (RF), gradient boosted regression trees (GBRTs) and cubist regression (Bulan and Sitorus 2022). These models leverage the spectral information encoded in Vis–NIR data for determining many soil characteristics like SOC, total nitrogen (N), phosphorus (P), potassium (K) and pH, with high accuracy (Chang et al. 2023). So, the applicability of spectroscopy combined with ML models for predicting soil properties has seen significant growth in recent years.

However, the predictability of soil properties on the basis of spectroscopy varies significantly depending on the specific soil properties being analyzed and ML models employed (Tavakoli et al. 2023). Research indicates that SOC, N, P and K can be predicted with high accuracy (R^2 values over 0.9) using ML algorithms such as Cubist, while properties, such as cation exchange capacity (CEC) and pH, show moderate predictability, and textural components, such as clay and sand, exhibit lower performance due to their complex relationships with spectral data (Clingensmith and Grunwald 2022). Therefore, the integration of spectroscopic data with advanced ML approaches enhances the efficiency and accuracy of soil property estimation, yet the variability in predictability underscores the need for tailored models that consider the unique characteristics of the soil being studied.

Previous efforts in estimating soil properties have utilized a variety of methods, including traditional laboratory analyses, remote sensing (RS) and advanced ML techniques. For example, studies have employed Vis-NIR spectroscopy combined with chemometric methods, such as partial PLSR and RF, to predict key soil attributes, such as SOC, N and CEC, achieving high accuracy in predictions (Clingensmith and Grunwald 2022; Ramírez et al. 2023). Additionally, ML algorithms, including SVMR and ANN, have been applied to airborne sensors' imagery data to estimate soil texture and EC, demonstrating effective spatial distribution modelling (El-Sayed et al. 2023). Recently, one study compared various multivariate ML models, such as PLSR, SVMR, RF and MARS, for the prediction of major soil nutrients, such as SOC, available N, P and K using combined laboratory-based hyperspectral spectra (spectroradiometer) and airborne spectra of AVIRIS-NG (Mondal et al. 2025a). Apart from soil nutrients and chemical properties, ML models, such as RF and ANN models, have also utilized for the prediction of soil physical properties, such as saturated hydraulic conductivity [saturated permeability, saturated and unsaturated shear strength parameters and soil water characteristics curve (SWCC)] (Li et al. 2022). The multivariate statistical models were also applied to determine the threshold suction limit in predicting the permeability function of unsaturated soils beyond residual suction (Tang et al. 2022). Bello and associates' researchers recently developed and validated a new mathematical model for accurate estimation of the bimodal distribution of SWCC under wetting, and it showed high reliability for engineering applications (Bello et al. 2025). Li et al. (2024) proposed a new method using the wetting SWCC and loess soil properties for accurately estimating wetting-induced settlements, which outperformed conventional approaches. (Li et al. 2024). Notable approaches included the integral methodologies for estimating soil hydraulic properties, which enhance the understanding of soil water dynamics. Recently, hybrid methods that combine ML models with interpolation techniques, such as regression kriging, have become popular for predicting and digitally mapping soil properties (Tziachris et al. 2019). These diverse approaches highlight the ongoing advancements in soil property estimation, aiming to improve agricultural practices and land management strategies through a better understanding of soil health (Dharumarajan et al. 2022). By integrating soil laboratory and ground-based sensor measurements with ML algorithms, researchers can develop predictive models for soil properties, overcoming the limitations of traditional protocols (Teixeira et al. 2024). The main concern in this regard is the accuracy of prediction. Although various researchers have reported varying levels of accuracy in soil property prediction using ML approaches, there is still a strong need to test and validate different ML models under diverse agronomic conditions and management practices. Hence, this study seeks to identify the best-performing ML model for soil property prediction in the study area by evaluating not only widely used ML models, such as PLSR, RF, SVM and ANN but also comparatively less explored models, such as gradient boosting regression (GBR), extreme gradient boosting (XGBoost) and extreme learning machine (ELM). Apart from this, to increase the accuracy of prediction, the study employed a large number of samples (2216 samples) in the modelling process. Although the primary objective of this study is to identify the most suitable ML models for soil property prediction in the study area, the study recognize that reproducibility and upscaling of such models depend on globally accepted frameworks, such as the Global Soil Laboratory Network

(GLOSOLAN) of Food and Agricultural Organization (FAO) and the FAO Global Soil Spectroscopy Program for standardization (Viscarra Rossel and McBratney 2016; Francos et al. 2023). Hence, this study should be viewed as a step towards developing region-specific calibrations that can later be harmonized with national and global soil spectral libraries through agreed-upon standard operating procedures (SOPs) and calibration transfer protocols (Safanelli et al. 2023). Therefore, the specific objectives of this study are as follows:

(i) To obtain reliable reference data on soil pH, EC and SOC through laboratory analyses of soil samples from the study area.
(ii) To develop predictive models for these properties using ground-based Vis–NIR spectra with both common and less explored ML algorithms.
(iii) To evaluate model performance and identify the most suitable model for accurate soil property estimation and region-specific calibration.

2. Materials and methods

2.1. Study area and soil sampling

The study was conducted in the Kymore Plateau and Satpura hill Zone, which includes Jabalpur, Katni, Seoni, Panna, Rewa, Sidhi, Satna and Singhroli districts (Figure 1). There is a wide variation in physiography, soils, rainfall, irrigation and cropping patterns in various parts of the zone. It comes under the wheat rice zone, having mixed red and medium black soils. The region is characterized by a tropical climate with annual rainfall ranging from 1100 to 1400 mm. Global positioning system (GPS)-based 2216 surface soil samples (0–15 cm) were collected from farmers' fields during 2023. Soil sampling was restricted to the 0–15 cm layer of soil because it represents the plough zone, where nutrient dynamics, root activity and management interventions are more pronounced (Hengl et al. 2015). This soil depth also

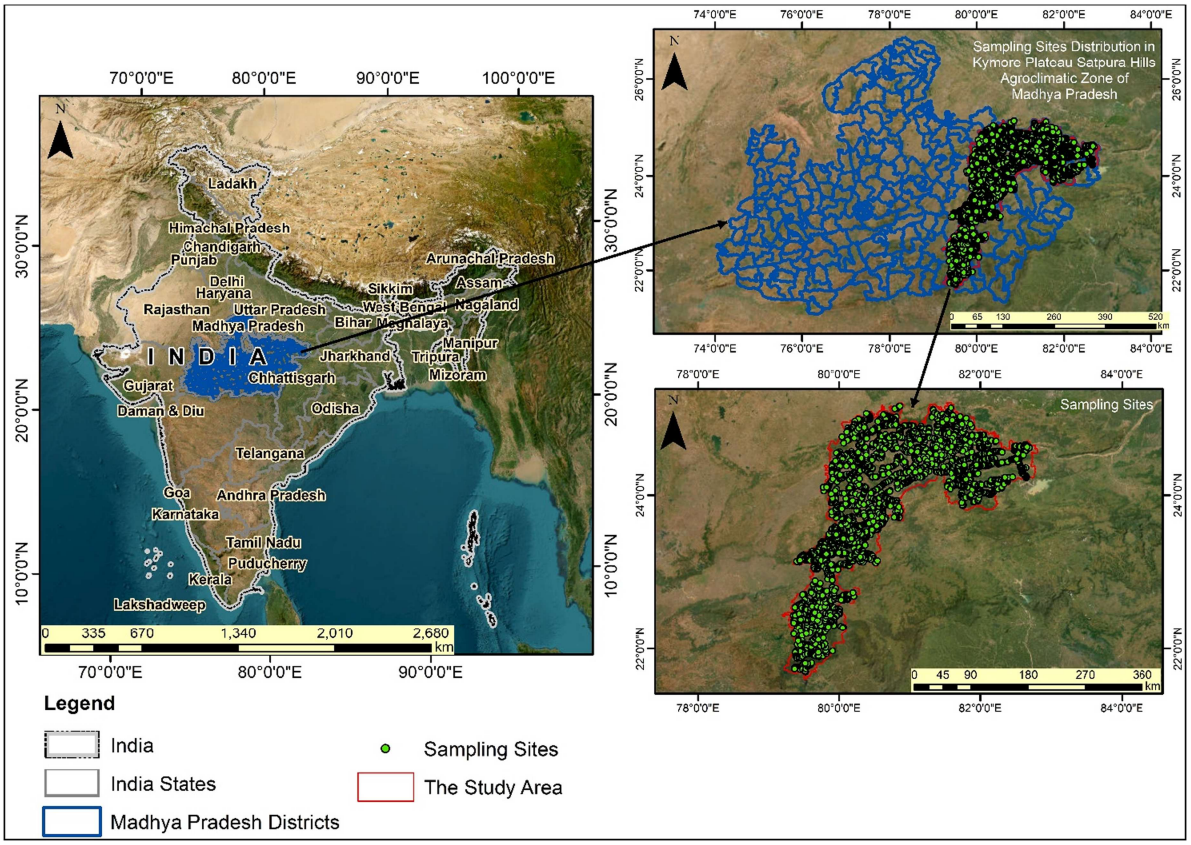

Figure 1. Location map of study area and soil sampling points.

exhibits the strongest soil spectral response due to direct exposure, making it the most appropriate for fertility assessment and remote sensing-based soil studies because the sensitivity of remote sensing is mainly confined to the surface soil layer (Viscarra Viscarra Rossel and Behrens 2010).

2.2. Soil samples preparation and analysis

The collected soil samples were air-dried, crushed and sieved using a 2 mm sieve, whereas stones and plant residues were eliminated. Soil pH was determined in 1:2.5 soil:water suspension using pH meter, while the EC was measured using EC meter in the extract as described in Jackson (1973). The SOC was determined using the wet oxidation protocol as explained by Walkley and Black (1934).

2.3. Spectroscopic measurement

The Vis–NIR spectra were measured through Spectroradiometer (RS-3500) using a high-intensity contact probe with a spectral range of 350 to 2500 nm and a spectral resolution of 1 nm. Each sample was placed on a rectangular black disk 5 cm in diameter and 2 cm in depth by tapping the tray on a table to ensure a smooth surface. The soil-filled rectangular black disk was kept on the dark background of a table, and light reflectance was measured. A spectral range of 400 to 2400 nm was used because the spectra outside this range were very noisy. Reflectance (R) spectra were transformed into absorbance ($A = \log10(1/R)$).

2.4. Machine learning model

Several algorithms, including PLSR, SVM, RF, ANN, XGBoost, CatBoost, LightGBM and ELM, were generated using R Studio software (RCoreTeam 2018), and these models were used to predict soil physicochemical properties. The whole data set ($n = 2217$) was divided into two datasets at a 1:5 ratio (every fifth sample used for the testing dataset) into two datasets viz., training dataset ($n = 1773$) and a testing dataset ($n = 443$) at an 80:20 ratio.

2.4.1. Partial least squares regression (PLSR)

The PLSR provides a constructed prediction model, which is used for conducting quantitative analyses of the spectral data using highly-correlated predictor variables. This algorithm is utilized in selecting the orthogonal components, which help in increasing spectral predictor 'X's variance as well as the 'Y' response variables (measured soil parameters' data). PLSR works on decomposing 'X' and 'Y' variables to factor loadings as 'P' and 'Q' and scores 'T'. The residuals 'E' and 'F' are used in the following equations as removed noises (Martens and Næs 1989).

$$X = TP + E,$$

$$Y = TQ + F.$$

The 'pls' package in the RStudio (RCoreTeam 2018) is used for developing different parameter's 'calibration' and 'validation' prediction models through integrating the Vis–NIR hyperspectral data and measured soil data using the 'data normalization', whereas each single reflectance value ranges between 0 and 1. Moreover, dividing the data into two data divisions are 80% and 20% for calibration and validation datasets, respectively. Afterwards, the two datasets are sorted based on their weights, whereas each dataset receives uniformly distributed values. Using 'leave one out' technique is used for the cross-validation process, whereas the components' evaluation occurs for multiple known numbers of components either in the calibration or validation prediction models using the root mean square error (RMSE) parameter (Efron and Tibshirani 1994). The component with the lowest RMSE is selected to represent the dataset.

2.4.2. Random forest (RF)

One machine learning approach for forecasting continuous output variables is called RF regression. Building several decision trees on various subsets of the training data and grouping their predictions

is how it operates. With random choosing of the features' subset, this approach chooses the best feature at each split, minimizing tree correlation and increasing model accuracy. Large and complicated datasets can be handled using the robust RF regression, which also resists overfitting problems.

2.4.3. Support vector machine regression (SVMR)

It is an ML algorithm that finds a hyperplane maximizing the margins among predicted outputs' values and observations, while being capable of accurately estimating unknown samples. It uses kernel functions for handling the non-linear relations within the inputs and outputs, as well as handling large datasets and high-dimensional input sizes. Moreover, it is sensitive in selecting the hyperparameters and computationally expensive for large datasets.

2.4.4. Artificial neural network (ANN)

For identifying the best data weights in ANN, the 'Levenberg–Marquardt training' approach is employed using the R software version 3.5.1 (RCoreTeam 2018) to ensure the availability of the minimum required number of neurons for simulating the training dataset precisely (Paul et al. 2020). Different internal processes are done, like 'sigmoidal linear activation functions', which are useful in preventing the 'overfitting' while model developing; and 'hidden neurons' were allocated in the model to achieve this process. The whole dataset is divided to three parts: calibration or training (80%), and testing (20%), as a similar methodology has been applied by Prashanth et al. (2020) and Xu et al. (2021). The ANN's main function was described in the following equation.

$$P = f_n\left(b_0 + \sum_{k=1}^{h}\left(w_k f_n\left(b_{hk} + \sum_{i=1}^{m} w_{ik}x_i\right)\right)\right),$$

where: P = predicted data; f_n = function transferring; b_0 = bias of layer output; h = the neurons number of the hidden layer; k = the neuron value; w_k = the value connects k and neuron output; $b_{hk} = k$ and b_0 bias; m = input variables' number; I = input layer; w_{ik} = the value connects i and k and x_i = values of input.

2.4.5. Gradient boosting regression (GBR)

GBR as an ML algorithm, is used in regression tasks, which work through iterative addition of weak decision-tree models to the ensemble, where each new tree corrects the errors of the previous tree. The GBR minimizes a loss function by gradient descent, whereas the gradient is computed with respect to the predictions of the current ensemble. However, the GBR deals with complex and nonlinear relations among the inputs and outputs as well as resists over-fitting.

2.4.6. Extreme gradient boosting (XGBoost)

The XGBoost is a regression ML algorithm that is commonly utilized for performing regression tasks by leveraging an ensemble of weak prediction models, specifically the decision trees, in order to create a strong prediction model. The XGBoost processes by an iterative boosting of an ensemble model with decision trees, whereas each subsequent tree is designed to reduce the residual errors of its predecessor till a predetermined limit of trees is reached.

2.4.7. Extreme learning machine (ELM)

The ELM is a ML algorithm that generalizes a single-hidden-layer feed-forward network (SLFN) with a weight and hidden-layer threshold in the first layer that are randomly assigned and a weight in the output layer that is calculated directly by the least-squares method. The entire learning process is completed in one round, with no iterations required; therefore, this algorithm performs at an extremely fast learning speed. For N distinct samples (x_i, t_i), where $x_i = [x_{i1}, x_{i2}, .., x_{in}]^T \in R^n$, x_i were soil spectra, while t_i were the measured soil data of pH, EC or SOC. Given a hidden node number $\hat{N}$, the activation function is defined as follows.

$$g(x) = \sum_{j=1}^{\hat{N}} \beta_j \delta_j(x_i) = \sum_{j=1}^{\hat{N}} \beta_j g(w_j x_i + b_j) = o_i, \quad i = 1, 2, \ldots, N; j = 1, 2, \ldots, \hat{N},$$

where $w_j \in Rn$ is the weight vector connecting the input nodes to the j^{th} hidden node and $b_j \in R$ is the threshold of the jth hidden node, $\beta j \in R$ represents the weight vector connecting the jth hidden node and the output nodes. To approach the real results of the training data infinitely, the prediction results o_i must be consistent with the real result t_i, in which case $\sum_{i=1}^{N} oi - ti = 0$. Under these conditions, Equation (1) can be expressed as follows: $\sum_{i=1}^{N} \beta_j g(w_j x_i + b_j) = t_i$, which is represented by a matrix:

$$\mathrm{H}\beta = \mathrm{T},$$

where:

$$\mathrm{H} = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & \cdots & g(w_{\hat{N}} \cdot x_1 + b_{\hat{N}}) \\ \vdots & \ddots & \vdots \\ g(w_1 \cdot x_N + b_1) & \cdots & g(w_{\hat{N}} \cdot x_N + b_{\hat{N}}) \end{bmatrix}_{N \times \hat{N}}, \quad \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_{\hat{N}} \end{bmatrix}_{\hat{N} \times 1}, \quad \textit{and } \mathrm{T} = \begin{bmatrix} t_1 \\ \vdots \\ t_N \end{bmatrix}_{N \times 1}^{T},$$

when the input weight $w_j \in Rn$ and bias $b_j \in \mathrm{R}$ are randomly assigned, the output matrix H in the hidden layer can be calculated by the ELM, after which the output weight β is calculated by $\hat{\hat{\beta}} = \mathrm{H}^{\dagger}\mathrm{T}$ where $\mathrm{H}^{\dagger}$ is the Mosse–Penrose generalized inverse of H.

The ELM regression ability varies significantly with the number of initial hidden neurons. To select the number of hidden neurons, we conducted an experiment by varying the number of hidden neurons from 1 to 120 in steps of 1. The optimal number of hidden neurons can be considered the number that is capable of creating the minimum prediction error through the cross-validation stage, either in the calibration or validation datasets. However, this whole process and calculations were conducted in R software version 3.5.1 (RCoreTeam 2018).

2.5. Model evaluation

To evaluate and compare models' performance, some statistical functions, namely, coefficient of determination (R^2), root mean square error (RMSE), relative percentage difference (RPD1) and ratio of performance to deviation (RPD2), were applied as described in the following equations.

$$R^2 = 1 - \left(\frac{\sum (Y_{\text{pred}} - Y_{\text{meas}})^2}{\sum (Y_i - Y_{\text{meas}})^2} \right),$$

$$\text{RMSE} = \sqrt{\frac{1}{n \sum (Y_{\text{pred}} - Y_{\text{meas}})^2}},$$

$$RPD1 = \frac{|X_1 - X_2|}{\left(\frac{X_1 + X_2}{2} \right)} \times 100,$$

$$\text{RPD2} = \frac{\text{SD}}{\text{RMSE}},$$

where Y_{pred} = predicted/estimated values; Y_i = mean of observed/actual/measured values; Y_{meas} = measured values; n = measured or predicted values; X_1 represents the original value; and X_2 represents the duplicate analysis value and SD is the standard deviation. Repeatability was assessed by computing RPD1 between original and duplicate measurements. On the other hand, RPD2 was calculated to assess the prediction accuracy of different multivariate models for soil property prediction using Vis–NIR reflectance spectra.

2.6. Reliability and repeatability of lab data and spectroscopic measurement

To ensure the analytical rigor and reliability of the laboratory-analyzed data, standard quality control procedures were implemented during the analysis of soil physico-chemical properties (pH, EC and SOC). Approximately 10% of the total samples means about 221 samples were reanalyzed (duplicate) to assess repeatability. For these duplicates, RPD and coefficient of variation (CV) were computed to quantify intra-laboratory precision. In addition, standard reference soil samples and internal control samples were included at regular intervals in each analytical batch to monitor instrument stability and potential drift. Calibration standards were re-run periodically to ensure the accuracy of the measurement.

In case of spectroscopic measurement, a total of 20 scans were done for each soil sample, and later they were averaged to make a single spectrum for each soil sample, thereby reducing the noise levels in the spectra. Besides that, calibration of the spectroradiometer sensor was repeated after obtaining spectra of 50 soil samples using white Spectralon to optimize its performance (Mondal et al. 2025a). Overall, these quality control practices or repeatability for both chemical analysis as well as spectroscopic measurements were in line with the internationally accepted recommendations, including the recommendations of FAO-GLOSOLAN and the ISO standards of soil analysis, which emphasizes the importance of duplicates, blanks, reference samples, repeated measurements for quality assurance in large scale soil monitoring programs (Francos et al. 2024).

2.7. Variables' Selection Approach

The competitive adaptive reweighted sampling (CARS) approach is based on Darwin's principle 'the survival of the fittest,' which was utilized to select the soil parameters' Vis–NIR hyperspectral variables (bands/wavelengths) that are strongly sensitive or correlated to each soil parameter; and described in the following equation (Jobson 2012).

$$C_j = \left| \frac{\bar{b_j}}{S(b_j)} \right|.$$

2.8. Multiple linear regressions (MLR)

The general equation for the proposed multiple linear regression (MLR) modelling was formulated as follows,

$$Y = a + bX_1 + cX_2 + dX_3 + \ldots + X_n,$$

where 'Y' = soil parameter content/concentration; 'X_1', 'X_2', 'X_3', …, and 'X_n' are the spectral variables of each soil parameter of soil pH, EC and SOC (Chatterjee and Price 1997; George and Maller 2003).

3. Results and discussion

3.1. Distribution of soil physico-chemical properties

3.1.1. Soil pH

The soil pH ranged from 5.13 to 8.22, 4.55 to 7.92, 5.64 to 8.02, 4.88 to 7.87, 5.24 to 7.89, 4.87 to 8.05, 4.86 to 7.71 and 4.79 to 8.39, with a mean value of 6.99, 6.47, 7.08, 6.58, 6.58, 6.43, 6.35 and 6.93 and CV value of 9.35, 10.34, 7.50, 9.40, 6.62, 10.12, 8.50 and 9.54% in soils of Jabalpur, Katni, Panna, Seoni, Rewa, Sidhi, Singrauli and Satna district, respectively (Table 1). The soil pH in the districts exhibited the order Panna > Jabalpur > Satna > Rewa > Seoni > Katni > Sidhi > Singrauli. Overall, the soil pH in the Kymore Plateau and Satpura hill zones varied between 4.55 and 8.39 (average and CV values = 6.72 and 9.75%, respectively). A low soil pH in the Singrauli district may be due to basic cations leaching and these nutrients being removed during crop harvest. The main reason of problems with soil acidity in the area is the intense cultivation of the rice–wheat cropping system. According to Elias (2017) and Yeneneh et al.

Table 1. Wet-chemistry data of soil pH, EC and SOC for different sampling sites.

	pH				EC				SOC			
Kymore Plateau &Satpura Hills	Min	Max	Mean	CV (%)	Min	Max (dS m^{-1})	Mean	CV (%)	Min	Max (g kg^{-1})	Mean	CV (%)
Jabalpur ($n = 379$)	5.13	8.22	6.99	9.35	0.03	0.66	0.21	49.58	1.50	11.25	5.34	30.73
Katni ($n = 220$)	4.55	7.92	6.47	10.34	0.08	0.56	0.28	36.21	3.15	8.10	5.57	9.75
Panna ($n = 301$)	5.64	8.02	7.08	7.50	0.11	0.97	0.34	48.12	1.05	10.35	5.32	27.33
Seoni ($n = 300$)	4.88	7.87	6.58	9.40	0.05	0.59	0.21	41.97	2.25	10.65	6.06	20.39
Rewa ($n = 284$)	5.24	7.89	6.58	6.62	0.04	0.72	0.25	45.92	2.40	9.30	5.60	17.99
Sidhi ($n = 210$)	4.87	8.05	6.43	10.12	0.05	0.92	0.22	52.62	1.54	8.10	5.48	13.97
Singrauli ($n = 222$)	4.86	7.71	6.35	8.50	0.05	0.73	0.22	46.85	1.65	9.45	5.38	17.48
Satna ($n = 300$)	4.79	8.39	6.93	9.54	0.08	0.89	0.28	52.74	1.65	9.90	5.34	23.02
Overall ($n = 2216$)	**4.55**	**8.39**	**6.72**	**9.75**	**0.03**	**0.97**	**0.25**	**51.31**	**1.05**	**11.25**	**5.51**	**22.35**

Note: Bold values denotes the result of total sample size.

(2022), the extensive leaching of bases and continuous use of acid-forming fertilizers are the main causes of extremely acidic soil reaction in intensively grown cereal crops. The results of this investigation support these findings. Similar findings were recorded in Patel et al. (2023) and Thakur et al. (2021).

3.1.2. Soil EC

The soil EC ranged from 0.03 to 0.66, 0.08 to 0.56, 0.11 to 0.97, 0.05 to 0.59, 0.04 to 0.72, 0.05 to 0.92, 0.05 to 0.73 and 0.08 to 0.89 dS m^{-1}, with mean values of 0.21, 0.28, 0.34, 0.21, 0.25, 0.22, 0.22 and 0.28 dS m^{-1} and CV values of 49.58, 36.21, 48.12, 41.97, 45.92, 52.62, 46.85 and 52.74% in soils of Jabalpur, Katni, Panna, Seoni, Rewa, Sidhi, Singrauli and Satna districts, respectively (Table 1). The soil EC in districts exhibited the order Panna > Katni > Satna > Rewa > Singrauli > Sidhi > Jabalpur > Seoni. Overall, the soil EC in the Kymore Plateau and Satpura hill zones differed among 0.03 and to 0.97 dS m^{-1}, whereas the average and CV values were 0.25 dS m^{-1} and 51.21%, respectively. Lower EC values can be attributed to the process of salts' leaching from the top to the lower horizons due to heavy precipitation. Similar results were found Kumawat and Gehlot (2020); Malo et al. (2023).

3.1.3. SOC

The SOC varied between 1.50 and 11.25, 3.15 and 8.10, 1.05 and 10.35, 2.25 and 10.65, 2.40 and 9.30, 1.54 and 8.10, 1.65 and 9.45 and 1.65 and 9.90 g kg^{-1}, with mean values of 5.34, 5.57, 5.32, 6.06, 5.60, 5.48, 5.32 and 5.34 g kg^{-1}, respectively, and the CV values of 30.73, 9.75, 27.33, 20.39, 17.99, 13.97, 17.48 and 23.02% in the soils of Jabalpur, Katni, Panna, Seoni, Rewa, Sidhi, Singrauli and Satna district, respectively (Table 1). The soil SOC in the districts exhibited the order Seoni > Rewa > Katni > Sidhi > Singrauli > Satna > Jabalpur > Panna. Overall, SOC in the Kymore Plateau and Satpura hill zones ranged among 1.05 and 11.25 g kg^{-1}, whereas the average and CV values were 5.51 g kg^{-1} and 22.35%, respectively. Higher SOC can be attributed to the regions' low agroclimatic temperature that suppresses the overall biological activities on the way of reducing the organic materials' decomposition as well as their accumulation on soil's surface as described by Srinidhi et al. (2020). Shrivastava et al. (2023) reported increased SOC due to increased contribution from the biomass. Our observations are in consistence with those for Santhi et al. (2018); Garnaik et al. (2020); Yadav et al. (2023) and Singh et al. (2023).

3.2. Intra-laboratory repeatability of wet chemistry measurements

The overall datasets ($n = 2216$) showed mean values of 6.72 for pH, 0.25 ds m^{-1} for EC and 5.51 g kg^{-1} for SOC, with CVs of 9.75%, 51.31% and 22.35%, respectively. The intra-laboratory repeatability assessment based on duplicate analysis of 221 randomly selected soil samples showed close agreement, with mean values of 7.22 (pH), 0.24 dS m^{-1} (EC) and 5.06 g kg^{-1} (SOC). This assessment also revealed high precision for most of the physicochemical properties (Table 2). Precision assessment indicated low variability for pH (CV = 7.97%) and SOC (CV = 19.22%), while EC showed higher variability (CV = 41.84%). Strong reproducibility was confirmed by very high R^2 values (more than 0.95) across all the soil properties. Again, RMSE values for all the soil properties were minimal (0.01–0.05), supporting analytical consistency. RPD1 values indicated very high repeatability for pH (0.36) and SOC (2.15), with moderate repeatability for EC (9.31). However, all these RPD1 values were below 10, meaning small RPD1 values, indicating high

Table 2. Intra-laboratory repeatability of soil physico-chemical properties (based on duplicate soil samples analysis, $n = 221$).

Soil property	Mean	CV (%)	R^2	RMSE	RPD1*	Interpretation
Soil pH	7.22	7.97	0.98	0.02	0.36	Very high precision
Soil EC	0.24	41.84	0.95	0.05	9.31	Acceptable precision
SOC	5.06	19.22	0.99	0.01	2.15	Very high precision

***RPD1** = relative percent difference.

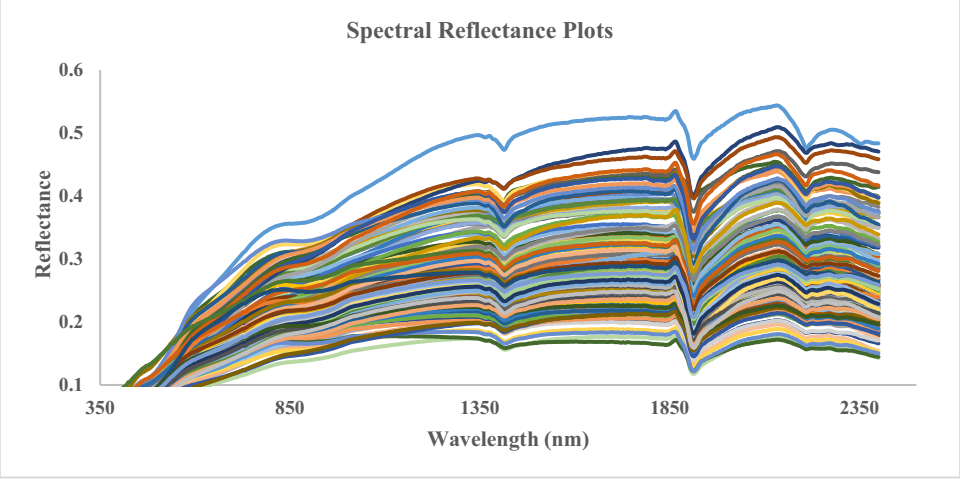

Figure 2. Spectral reflectance plots (reflectance vs wavelength) of studied soil samples.

precision in laboratory-based wet chemistry analysis. Overall, these results confirmed a reliable intra-laboratory performance, particularly for pH and SOC.

3.3. Soil spectral behaviour

Hyperspectral signatures of the 2216 soil samples were acquired in the laboratory conditions using the analytical spectral device (ASD) in the spectral range of the Vis–NIR between 350 and 2500 nm (Figure 2). The collected spectral data were converted to text format with reflectance values between 0 and 1 using the ASD software. However, there are strong absorptions were distinguished in the collected spectral signatures near 600 and 900 nm which are primarily associated with some minerals, such as hematite and goethite (Morris et al. 1985), and other absorption was recorded near the wavelength of 600 nm which is related to the organic compounds (Stenberg et al. 2010). Moreover, the functional spectral groups of the hydroxyl groups were recognized in the spectral locations of the 1400, 1900 and 2200 nm (Stoner and Baumgardner 1981; Clark et al. 1990; Stenberg et al. 2010).

3.4. Prediction of soil physicochemical properties using multiple ML models

The performance of different ML models for predicting soil physicochemical properties (pH, EC and SOC) was evaluated. The prediction performances of various ML models were tabulated (Table 3) and depicted using scatterplots (Figures 3–5). For pH, PLSR achieved the highest accuracy ($R^2 = 0.99$, RMSE = 0.36 and RPD2 = 2.14 in the testing dataset), followed closely by SVMR ($R^2 = 0.97$). ANN and RF also performed well, with testing R^2 values of 0.91 and 0.95, respectively. XGBoost, LightGBM and CatBoost showed moderate performance, while ELM had the lowest predictive accuracy. For EC, PLSR again outperformed all other models ($R^2 = 0.88$, RMSE = 0.06, RPD2 = 1.75 in the testing data set), whereas SVMR, RF and ANN displayed lower accuracies, with testing R^2ranging between 0.28 and 0.68. Tree-based boosting methods (XGBoost, LightGBM and CatBoost) and ELM demonstrated very poor performance for EC prediction, with R^2 below 0.25 in the testing dataset. For SOC, PLSR consistently provided superior results

Table 3. Prediction performances of various multivariate models in predicting pH, EC and SOC.

		Training			Testing		
Property	Model	R^2	RMSE	RPD2	R^2	RMSE	RPD2
pH	PLSR	0.99	0.35	2.15	0.99	0.36	2.14
	SVMR	0.98	0.36	2.09	0.97	0.37	2.06
	ANN	0.95	0.38	2.01	0.91	0.41	2.00
	RF	0.97	0.36	2.05	0.95	0.37	2.02
	XGBoost	0.90	0.41	1.98	0.86	0.45	1.88
	LightGBM	0.84	0.44	1.95	0.80	0.47	1.77
	CatBoost	0.85	0.42	1.99	0.83	0.45	1.86
	ELM	0.79	0.57	1.13	0.74	0.64	1.75
EC ($dS\ m^{-1}$)	PLSR	0.92	0.12	1.97	0.88	0.06	1.75
	SVMR	0.72	0.15	1.65	0.68	0.06	1.45
	ANN	0.37	0.82	1.10	0.28	0.96	1.03
	RF	0.62	0.27	1.79	0.47	0.35	1.41
	XGBoost	0.28	0.95	1.39	0.22	0.98	1.25
	LightGBM	0.18	1.07	1.08	0.13	1.11	1.00
	CatBoost	0.20	1.02	1.10	0.14	1.05	1.09
	ELM	0.11	1.12	1.06	0.10	1.16	1.02
SOC ($g\ kg^{-1}$)	PLSR	0.99	0.41	2.20	0.99	0.56	2.17
	SVMR	0.98	0.55	2.14	0.98	0.60	2.12
	ANN	0.95	0.58	2.09	0.93	0.61	2.03
	RF	0.97	0.56	2.11	0.96	0.60	2.09
	XGBoost	0.92	0.69	2.07	0.89	0.77	2.01
	LightGBM	0.88	0.77	2.05	0.79	0.79	1.97
	CatBoost	0.88	0.78	2.04	0.85	0.79	1.99
	ELM	0.77	0.97	1.97	0.73	1.02	1.87

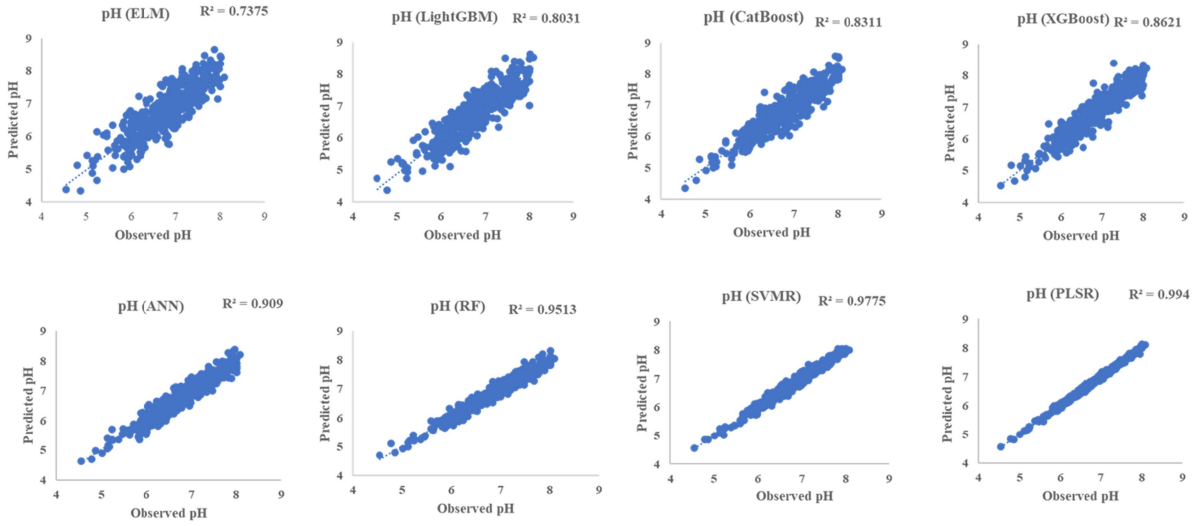

Figure 3. Scatter plots of observed vs predicted pH showing the predictive performance of various multivariate models in the test data set.

($R^2 = 0.99$, RMSE = 0.56 and RPD2 = 2.17 in the testing dataset), followed by SVMR ($R^2 = 0.98$) and RF ($R^2 = 0.96$). ANN gave satisfactory performance, while XGBoost, LightGBM and CatBoost performed moderately well, with R^2 values ranging between 0.79 and 0.89. ELM was found to be the least suitable ML model for SOC prediction. Overall, PLSR demonstrated stable and reliable predictions across all the soil properties.

3.5. Selected sensitive spectral bands

Using the CARS algorithm, sensitive spectral bands were selected for each soil parameter. The MLR algorithm was applied for developing the regression equation for each soil parameter, which can be used for estimating these properties in new soil samples using the same spectral regions. The sensitive bands recorded for soil pH were 400, 401, 422, 423 and 1910 nm; for the soil EC were 401, 402, 403, 404, 423 and

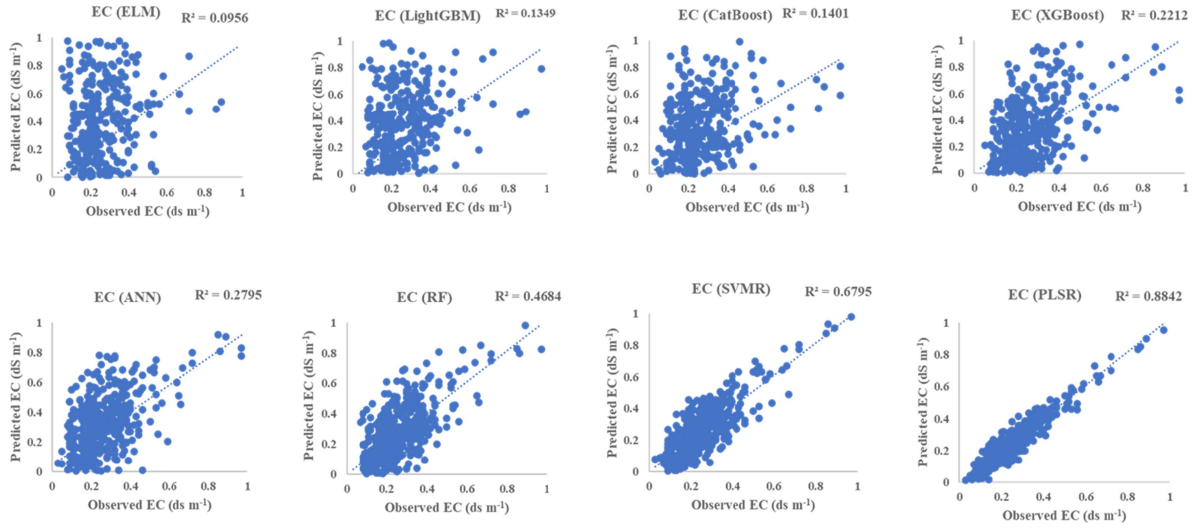

Figure 4. Scatter plots of observed vs predicted EC (dS m^{-1}) showing the predictive performance of various multivariate models in the test data set.

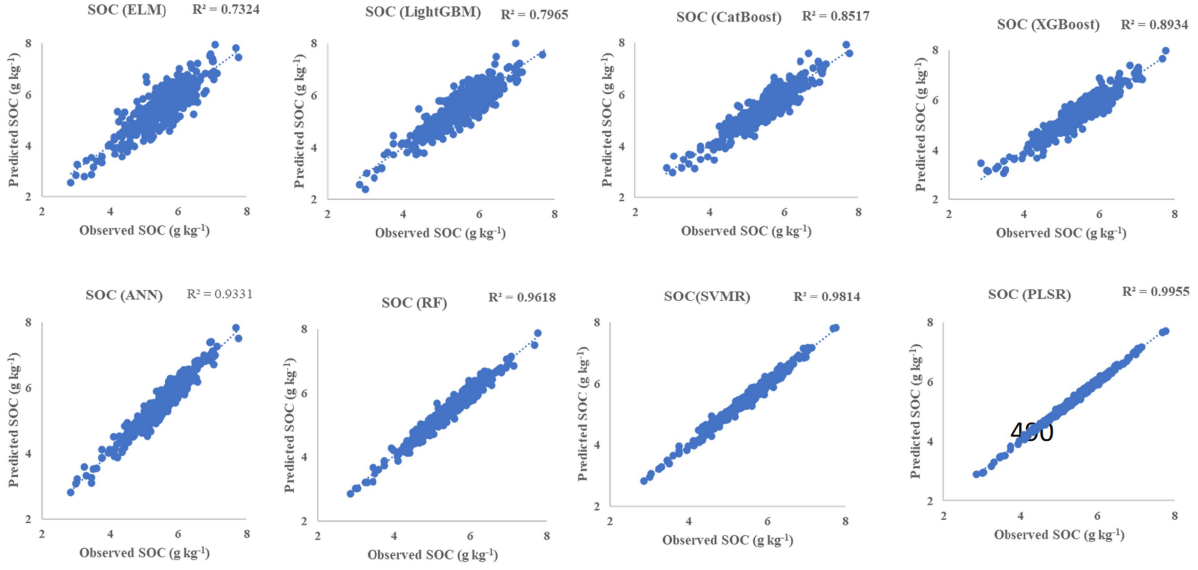

Figure 5. Scatter plots of observed vs predicted SOC (g kg^{-1}) showing the predictive performance of various multivariate models in the test data set.

424 nm; while for SOC were 429, 487, 1908, 1912, 1913, 1914, 1915 and 1916 nm. The developed regression equations to estimate soil pH, EC and SOC are written in the following equations.

Soil pH = $6.619 + 399.767_{\times 400} - 332.688_{\times 401} - 877.673_{\times 422} + 827.189_{\times 423} - 3.022_{\times 1910}$.

Soil EC (dS m^{-1}) =

$0.1416 + 322.5476_{\times 401} - 607.2487_{\times 402} + 474.3602_{\times 403} - 196.1975_{\times 404} + 116.1385_{\times 423} - 107.3770_{\times 424}$SOC (g kg^{-1}).

$= 5.653 + 63.102_{\times 429} - 53.553_{\times 487} - 2416.628_{\times 1909} + 31458.290_{\times 1912} - 69865.063_{\times 1913} + 70801.571_{\times 1914} - 40719.189_{\times 1915} + 10742.932_{\times 1916}$.

3.6. Comparative efficiency of spectroscopy vs wet chemistry analysis

Table 4 represents the comparative evaluation of analytical methods that highlights the trade-offs between conventional wet chemistry and Vis–NIR spectroscopy (Table 4). While wet chemistry is destructive, time-

Table 4. Comparative assessment between wet chemistry and spectroscopic analysis.

Method of analysis	Time per sample	Cost per sample	Nature	Precision
Wet-chemistry	1–2 h	500 (INR)	Destructive	Highly precise and accurate
Spectroscopic (Vis–NIR) analysis	2–3 min	80 (INR)	Non-destructive	Variable accuracy depending on sample size, model selection and spectral relationship (direct/indirect)

*INR = Indian Rupee (1 INR = 0.011 US Dollar).

consuming (1–2 h per sample), and relatively costly (approximately 500 INR per sample), it provides highly precise and accurate measurements. In contrast, Vis–NIR spectroscopy offers rapid (2–3 min per sample analysis), low-cost (approximately 80 INR per sample) and non-destructive analysis, though its accuracy depends on the sample size, calibration models and spectral relationships. These differences emphasize the complementary roles of both approaches in soil analysis, with spectroscopy being well suited for large-scale, high-throughput applications and wet chemistry serving as the reference standard.

4. Discussion

4.1. Addressing reproducibility concerns in large-scale soil monitoring

Comparison of the full dataset with duplicate analysis highlights consistent laboratory performance, with means and CVs showing strong alignments across both datasets. The low CV, RMSE and RPD1 values, especially for pH and SOC, confirm the high precision and repeatability of wet chemistry methods, which is in agreement with the previous reports on soil laboratory quality assurance (Reeves III 2010; Nocita et al. 2015). The higher variability in EC reflects known challenges in its measurements due to temporal and handling effects (Viscarra Rossel and Behrens 2010). Importantly, the duplicate results demonstrate that intra-laboratory variability is minimal compared to the natural variation captured in the full dataset. This strengthens confidence in using the dataset for validating ML-based spectroscopic models. Moreover, the reproducibility achieved here aligns with FAO-GLOSOLAN recommendations for soil monitoring networks (Safanelli et al. 2023). Overall, the findings assure that both wet chemistry and spectroscopy can be integrated under standardized protocols for scalable soil property predictions.

4.2. Comparisons of various multivariate models in soil property prediction

The predictability of three important soil properties using various multivariate ML models was evaluated using the thresholds proposed by Chang et al. (2001). The thresholds include three categories of soil property prediction based on RPD2 values, i.e. RPD2 > 2.0 indicates 'excellent' prediction, whereas RPD2 values between 1.4 and 2.0 indicates 'fair' prediction, and <1.4 RPD2 value indicates 'poor' prediction of a particular soil property by a particular model (Chang et al. 2001). Applying these thresholds to the present results placed pH and SOC in the high predictability class for the top performing models (PLSR and SVMR), whereas EC showed low to moderate predictability by various multivariate models. PLSR produced exceptionally high testing statistics for pH and SOC ($R^2 = 0.99$ and RPD2 > 2.0), indicating robust calibration and validation performances for those attributes in the datasets. Another probable reason for getting such high accuracy might be the usage of a huge number of samples (2216 soil samples) for training the spectral model for soil property prediction. However, in spite of this, the predictability of EC by any model was low. The present finding is very similar to the previous findings that reported difficulty in the reliable prediction of EC from Vis–NIR spectra (Stenberg et al. 2010; Soriano-Disla et al. 2014). Mustafa and Moursy (2020) utilized the PLSR model for predicting soil pH and EC in some Indian soils based on the Vis–NIR spectra and obtained R^2 of 0.69 and 0.31, respectively. Weak predictability of EC can be attributed to the absence of a direct spectral relationship in the Vis–NIR region, confounding by texture and moisture, and landscape or climatic influences that vary spatially (Minasny et al. 2009; Weindorf et al. 2016). Another group of researchers also mentioned that dissolved slats were not well represented in the Vis–NIR spectral regions (Soriano-Disla et al. 2014). The dominance of PLSR for pH and SOC here is consistent with many previously reported studies, where the computed latent variables (LVs) effectively explained high-dimensional, collinear spectral predictors for soil property predictions

(Zornoza et al. 2008; Mondal and Sekhon 2019). The PLSR focuses on covarying spectral information into a few predictive variables, which stabilizes the model across concentration ranges and sample heterogeneity (Stenberg et al. 2010). Non-linear models (SVMR, RF, ANN and some boosting algorithms) showed mixed predictive performance. After PLSR, SVMR and RF performed almost equally well in soil property prediction, particularly for pH and SOC. A recent study also reported superior performance of the SVMR model in SOC prediction, both from laboratory-based Vis–NIR spectra and airborne (AVIRIS-NG) spectra (Mondal et al. 2025a). SVMR and RF were competitive for SOC prediction in several studies, where some boosting models gave only moderate predictability in soil property prediction (Yang et al. 2019; Zayani et al. 2023). SVMR and RF excel when spectral relationships are non-linear or where interactions among predictor variables are complex (Mondal et al. 2025a). These two non-linear models are highly suitable for predicting and mapping soil nutrients, including macronutrients and micronutrients (Mondal et al. 2025a, 2025b). Underperforming models in this study (ELM and LightGBM) likely reflect model sensitivity to tuning, overfitting risks or insufficient representation of the target's spectral signatures. Although PLSR demonstrated superior performance in this study, comparative literature demonstrated mixed outcomes: some studies reported that non-linear methods outperformed the linear PLSR model (Yang et al. 2019; Mondal et al. 2025a), while others documented PLSR's continued competitiveness depending on the dataset composition (Mondal and Sekhon 2019; Nawar and Mouazen 2022). Performances of various ML-based models also depend on some methodological factors, such as spectral preprocessing techniques (smoothening, noise removal, etc.), band selection and hyperparameter tuning, which help to improve the capacity to detect subtle spectral features (Lagacherie et al. 2008; Mondal et al. 2025b). Based on the performances of various multivariate models, the present study highlighted the usage of the PLSR model as a robust baseline model for soil pH and SOC prediction; while systematically testing the non-linear SVMR or RF model could be used to predict those properties having a non-linear relationship with spectra. Overall, this study reinforces the view that spectroscopy coupled with robust modelling techniques remains a powerful tool for rapid soil property assessment, but its success is property specific and depends on aligning the spectral domain, sampling design and algorithm choice with the target or specific soil property.

4.3. Sensitive bands for soil property predictions

4.3.1. Soil pH

Regarding the sensitive bands to the soil pH, (Gozukara et al. 2022) found that the wavelengths (560, 610, 600, 1450, 1510, 1760, 1840 and 2310 nm) are significant for estimating soil pH; while Yang et al. (2019) found spectral ranges of (500:700), (800:900), (1150:1250), (1300:1450), (1550:1750), (1900:2000) and (2100:2200) nm are sensitive to predict the soil pH. Moreover, Xu et al. (2018) selected the wavelengths of 480, 780, 1120, 1910, 2200, and 2390 nm as the best spectral variables for predicting the soil pH. Viscarra Rossel et al. (2006) stated that soil pH, which is calibrated in calcium chloride, has negative spectral peaks in the range from 470 to 670 nm, and at 600 nm. They added that the soil pH is related to the soil moisture, whereas through using the NIR spectral range, the O–H bonds have strong vibrations at 1400, 1900 and 2200 nm.

4.3.2. Soil electrical conductivity (EC)

Gozukara et al. (2022) selected the significant spectral variables that are related to the soil EC, and those that are 500, 510, 710, 2130, 1470, 1650, 1880, 2030 and 2080; while Levi et al. (2020) selected the wavelengths of 1363, 1898 and 1982 nm as responsible for predicting EC. Furthermore, Yang et al. (2020) mentioned that the wavelengths range between 600 and 700, 800 and 850, 1000 and 1800 and 1900 and 2100 nm are essential to predict soil EC. Levi et al. (2020) found the 1898 and 1982 nm wavelengths are strongly related to the soil EC; while Yang et al. (2020) found the spectral ranges 1000–1800 nm and 1900–2100 nm are significantly correlated to soil EC. Furthermore, Gozukara et al. (2022) stated that 590, 2360 and 800 nm for a 1:1 soil-to-water ratio; 2150, 940 and 960 nm for a 1:2.5 soil-to-water ratio; and 2280, 770 and 1430 nm for a 1:5 soil-to-water ratio were helpful in estimating EC. Therefore, the ratio of extraction is changing the spectral behaviour and controlling the relation between EC and Vis–NIR reflectance (Grunwald et al. 2015).

4.3.3. Soil organic carbon (SOC)

Several studies have concluded that the sensitive bands for predicting the SOC are 1100, 1600, 1700–1800, 2000 and 2200–2400 nm (Stenberg and Viscarra-Rossel 2010). Viscarra Rossel et al. (2006) found that 410, 570 and 660 nm wavebands are sensitive to predict SOC. Furthermore, Clark (1999) selected wavelengths of 2300, 1700 and 1100 nm as significant spectra for detecting C–H bonds; however, the SOC bands represent other organic and inorganic molecules.

4.4. Limitations

Although the multivariate models demonstrated high accuracy for soil pH and SOC predictions, several limitations of this study should be acknowledged for designing better, futuristic work on spectroscopy-based soil property predictions. First, reliance solely on Vis–NIR spectra to predict a particular soil property restricted the capability to capture attributes like EC that exhibit weak spectral responses in this domain (Soriano-Disla et al. 2014). Second, boosting models and ELM underperformed partly due to their sensitivity to hyperparameter tuning and the possibility of overfitting, suggesting that more systematic optimization may be required. Third, although the sample size is large, it may not be sufficient to represent the total heterogeneity or spatial variability of soil properties across broader landscapes, which can limit the transferability of calibrations to new environments (Wetterlind et al. 2008). Fourth, the study did not integrate auxiliary variables (soil moisture, texture, other terrain attributes, etc.), which could have strengthened EC prediction and potentially improved robustness across soil properties. Fifth, calibration transfer was not performed because only one instrument (spectroradiometer) was available at the institute. Finally, the absence of uncertainty quantification (prediction intervals, standard error mapping) limits the operational reliability of the models for precision agriculture and soil fertility management.

4.5. Forward-looking perspectives and recommendations

Future research should focus on integrating mid-infrared (MIR) and shortwave infrared (SWIR) spectroscopy, which can capture ionic and vibrational features more relevant for soil property prediction, like EC prediction (Omran 2016; Nawar and Mouazen 2022). Hybrid modelling approaches that combine PLSR for dimensionality reduction with non-linear learners such as SVMR or RF on residuals may improve robustness by exploiting both linear and non-linear features (Yang et al. 2019). Incorporating auxiliary soil information such as soil texture, soil moisture and environmental covariates derived from a digital elevation model (SRTM-DEM) and using sensor fusion techniques, such as combining multispectral (Senyinel-2, Landsat-8/9 data) and hyperspectral (AVIRIS-NG, HyMap) datasets for soil nutrient prediction, can enhance transferability and field-scale applications (Grunwald et al. 2015; Naimi et al. 2022; Mondal et al. 2025b). Future work should include inter-instrument transfer and integration of national as well as global soil spectral libraries, such as FAO-GLOSOLAN, to standardize the work in a globally accepted framework. Rigorous spectral preprocessing (standard normal variate, continuum removal, noise removal, different orders of derivatives) and targeted band selection using various techniques, such as VIP and CARS, in the SWIR domain should be systematically explored to improve model performances (Wang and Wang 2022; Mondal et al. 2025a). Moreover, the inclusion of uncertainty quantification methods, such as the bootstrap technique and spatially explicit error mapping, is strongly recommended to inform end users of prediction reliability (Mondal et al. 2025b). From an operational perspective, PLSR remains a reliable baseline for routine mapping of soil pH and SOC, while non-linear learners should be tested in complex and heterogeneous soils. For soil EC, direct field measurements or MIR-based calibrations may be necessary to improve prediction accuracy. Overall, advancing soil spectroscopy requires aligning spectral domain, sampling strategy and modelling approaches with specific soil attributes while also ensuring transparent reporting, reproducibility and field-scale applicability.

5. Conclusions

The study demonstrated that Vis–NIR spectroscopy, combined with multivariate models, can reliably predict soil pH and SOC with high accuracy, while EC prediction remained challenging due to its weak

spectral response. Among the multivariate models, the linear model PLSR outperformed all other models in predicting soil properties. Among non-linear models, SVMR and RF demonstrated their potential in spectral-based soil property prediction after PLSR. Boosting-based models showed moderate accuracy in soil property predictions. The intra-laboratory repeatability assessment based on duplicate analysis confirmed the robustness of wet chemistry as a benchmark method, while also providing evidence that spectroscopic approaches, though faster and more cost-effective, require careful calibration and validation to ensure reproducibility and reliability. Compared with conventional wet chemistry, spectroscopy significantly reduced analysis time and cost per sample, highlighting its potential for large-scale soil fertility assessment and precision agriculture applications. A unique contribution of this work lies in explicitly quantifying repeatability, reproducibility and cost–time trade-offs, thereby addressing a critical gap in many spectroscopic studies, where such aspects remain underexplored or only a little bit explored. Nonetheless, the findings underscore the need for domain-specific modelling strategies, the integration of auxiliary variables and calibration transfer across instruments to increase operational reliability. Overall, this study establishes a framework that bridges laboratory precision with field-scale applicability, paving the way for scalable, transparent and reproducible soil spectroscopy applications in sustainable soil health management.

Acknowledgements

This research was supported by the National Agriculture Higher Education Project, Centre for Advanced Agriculture Science and Technology (NAHEP-CAAST) on 'Skill Development to use Data for Natural Resources Management in Agriculture', College of Agricultural Engineering, Jawaharlal Nehru KrishiVishwaVidyalaya, Jabalpur, for providing financial support. The third author (Bhabani Prasad Mondal), having equal contribution like the first author, also conveyed thanks to the Directorate of Research, Bihar Agricultural University, Sabour, for providing the BAU Communication Number for submitting this article to this internationally reputed journal.

Author contributions

Conceptualization, G.S.T, D.K.S., B.P.M., S.D., S.Das, A.C., R.S., R.K.N., M.E.F., C.F., P.D. and A.R.A.M.; methodology, G.S.T., D.K.S., B.P.M., S.D., S.Das, A.C., R.S., R.K.N., M.E.F., C.F., P.D. and A.R.A.M.; software, G.S.T., D.K.S., B.P.M., S.D., S.Das, A.C., R.S., R.K.N. and A.R.A.M.; validation, G.S.T., D.K.S., B.P.M., S.D., S.Das, A.C., R.S., R.K.N., M.E.F., C.F., P.D. and A.R.A.M.; formal analysis, G.S.T., D.K.S., B.P.M., S.D., S.Das, A.C., R.S., R.K.N. and A.R.A.M.; investigation, G.S.T., D.K.S., B.P.M., S.D., S.Das, A.C., R.S., R.K.N. and A.R.A.M.; resources, G.S.T., D.K.S., B.P.M., S.D., S.Das, A.C., R.S., R.K.N., M.E.F., C.F., P.D. and A.R.A.M.; data curation, G.S.T., D.K.S., B.P.M., S.D., S.Das, A.C., R.S., R.K.N., M.E.F., C.F., P.D. and A.R.A.M.; writing—original draft preparation, G.S.T., D.K.S., B.P.M., S.D., S.Das, A.C., R.S., R.K.N. and A.R.A.M.; writing—review and editing, G.S.T., D.K.S., B.P.M., S.D., S.Das, A.C., R.S., R.K.N., M.E.F., C.F., P.D. and A.R.A.M.; visualization, G.S.T., D.K.S., B.P.M., S.D., S.Das, A.C., R.S., R.K.N., M.E.F., C.F., P.D. and A.R.A.M.; supervision, G.S.T., D.K.S., B.P.M., S.D., S.Das, A.C., R.S., R.K.N., M.E.F., C.F., P.D. and A.R.A.M.; project administration, G.S.T., D.K.S., B.P.M., S.D., S.Das, A.C., R.S., R.K.N., M.E.F., C.F., P.D. and A.R.A.M. and funding acquisition, G.S.T., D.K.S., B.P.M., S.D., S.Das., A.C., R.S., R.K.N., M.E.F., C.F., P.D. and A.R.A.M.

All authors have read and agreed with the published version of the manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Abd-Elazem AH et al. 2024. Estimating soil erodible fraction using multivariate regression and proximal sensing data in arid lands, south Egypt. Soil Systems. 8(2):48. https://doi.org/10.3390/soilsystems8020048

Ahmadi A, Emami M, Daccache A, He L. 2021. Soil properties prediction for precision agriculture using visible and near-infrared spectroscopy: a systematic review and meta-analysis. Agronomy. 11(3):433. https://doi.org/10.3390/agronomy11030433

Ai W et al. 2022. Application of hyperspectral imaging technology in the rapid identification of microplastics in farmland soil. Sci Total Environ. 807:151030. https://doi.org/10.1016/j.scitotenv.2021.151030

Amundson R. 2021. Factors of soil formation in the 21st century. Geode. 391:114960. https://doi.org/10.1016/j.geoderma.2021.114960

Bello N, Satyanaga A, Gofar N, Kim J. 2025. Estimation of bimodal soil-water characteristics curve under wetting process. PLoS One. 20(6):e0325646. https://doi.org/10.1371/journal.pone.0325646
Bulan R, Sitorus A. 2022. Vis-NIR spectra combined with machine learning for predicting soil nutrients in cropland from Aceh province, Indonesia. Case Studies in Chemical and Environmental Engineering. 6:100268. https://doi.org/10.1016/j.cscee.2022.100268
Chang CW, Liard DA, Mausbach MJ, Hurburgh CR. 2001. Near-infrared reflectance spectroscopy-principal components regression analysis of soil properties. Soil Science Society of American Journal. 65:480–490. https://doi.org/10.2136/sssaj2001.652480x
Chang N et al. 2023. Soil organic carbon prediction based on different combinations of hyperspectral feature selection and regression algorithms. Agronomy. 13(7):1806. https://doi.org/10.3390/agronomy13071806
Chatterjee S, Price B. 1997. Regression Analysis by Example. 2nd. John Wiley & Sons: New York.
Clark RN. 1999. Spectroscopy of rocks and minerals and principles of spectroscopy. In: Rencz AN, editor. Remote Sensing for the Earth Sciences. John Wiley & Sons: Chichester, UK. p 3–58.
Clark RN, King TV, Klejwa M, Swayze GA, Vergo N. 1990. High spectral resolution reflectance spectroscopy of minerals. Journal of Geophysical Research: Solid Earth. 95(B8):12653–12680. https://doi.org/10.1029/JB095iB08p12653
Clingensmith CM, Grunwald S. 2022. Predicting soil properties and interpreting vis-NIR models from across continental United States. Senso. 22(9):3187. https://doi.org/10.3390/s22093187
Dhaliwal SS et al. 2021. The pedospheric variation of DTPA-extractable Zn, Fe, Mn, Cu and other physicochemical characteristics in major soil orders in existing land use systems of Punjab, India. Sust. 14(1):29. https://doi.org/10.3390/su14010029
Dharumarajan S, Lalitha M, Gomez C, Vasundhara R, Kalaiselvi B, Hegde R. 2022. Prediction of soil hydraulic properties using VIS-NIR spectral data in semi-arid region of Northern Karnataka plateau. Geoderma Regional. 28:e00475. https://doi.org/10.1016/j.geodrs.2021.e00475
Efron B, Tibshirani RJ. 1994. An Introduction to the Bootstrap. New York, NY, USA: CRC Press; p. 456.
Elias E. 2017. Characteristics of nitisol profiles as affected by land use type and slope class in some Ethiopian highlands. Environmental Systems Research. 6:1–15. https://doi.org/10.1186/s40068-017-0097-2
El-Sayed MA, Abd-Elazem AH, Moursy AR, Mohamed ES, Kucher DE, Fadl ME. 2023. Integration vis-NIR spectroscopy and artificial intelligence to predict some soil parameters in arid region: a case study of wadi Elkobaneyya, south Egypt. Agronomy. 13(3):935. https://doi.org/10.3390/agronomy13030935
Francos N et al. 2023. A spectral transfer function to harmonize existing soil spectral libraries generated by different protocols. Applied and Environmental Soil Science. 2023:14155390–17. https://doi.org/10.1155/2023/4155390
Francos N et al. 2024. Mapping soil organic carbon stock using hyperspectral remote sensing: a case study in the Sele river plain in Southern Italy. Remote Sens. 16(5):897. https://doi.org/10.3390/rs16050897
Garnaik S, Sekhon BS, Sahoo S, Dhaliwal SS. 2020. Comparative assessment of soil fertility status of various agroecological regions under intensive cultivation in northwest India. Environ Monit Assess. 192:1–18. https://doi.org/10.1007/s10661-020-08290-6
George D, Maller P. 2003. SPSS for Windows Step by Step. A Simple Guide and Reference. 11. 4th. Pearson Education, Inc: Boston. p 400.
George EB, Gomez C, Kumar ND. 2021. Adapting prediction models to bare soil fractional cover for extending topsoil clay content mapping based on AVIRIS-NG hyperspectral data. Remote Sens. 16(6):115091.
Gozukara G., Altunbas S., Dengiz O., Adak A.. 2022. Assessing the effect of soil to water ratios and sampling strategies on the prediction of EC and pH using pXRF and Vis-NIR spectra. Comput Electron Agric. 203:107459. https://doi.org/10.1016/j.compag.2022.107459
Grunwald S, Vasques GM, Rivero RG. 2015. Fusion of soil and remote sensing data to model soil properties. In: Adv Agron. 131, .p 1–109. https://doi.org/10.1016/bs.agron.2014.12.004
Hengl T et al. 2015. Mapping soil properties of Africa at 250 m resolution: random forests significantly improve current predictions. PLoS One. 10(6):e0125814. https://doi.org/10.1371/journal.pone.0125814
Hou D, Bolan NS, Tsang DC, Kirkham MB, O'connor D. 2020. Sustainable soil use and management: an interdisciplinary and systematic approach. Sci Total Environ. 729:138961. https://doi.org/10.1016/j.scitotenv.2020.138961
Jackson ML. 1973. Soil chemical analysis-advanced course. In: A manual of methods useful for instruction and research in soil chemistry, physical chemistry, soil fertility and soil genesis. 2ndMadison, US.
Jobson JD. 2012. Applied Multivariate Data Analysis: Regression and Experimental Design. Berlin/Heidelberg, Germany: Springer Science & Business Media
Kumawat R, Gehlot Y. 2020. GIS based mapping of soil fertility status of Tehsil Jobat, district Alirajpur, Madhya Pradesh, India. Int. J. Curr. Microbiol. App. Sci. 9(10):60–69. https://doi.org/10.20546/ijcmas.2020.910.009
Lagacherie P, Baret F, Feret JB, Madeira JN, Robbez-Masson JM. 2008. Estimation of soil clay and calcium carbonate using laboratory, field and airborne hyperspectral measurements. Remote Sens Environ. 112:825–835. https://doi.org/10.1016/j.rse.2007.06.014
Lehmann J, Bossio DA, Kögel-Knabner I, Rillig MC. 2020. The concept and future prospects of soil health. Nature Reviews Earth & Environment. 1(10):544–553. https://doi.org/10.1038/s43017-020-0080-8
Levi N, Karnieli A, Paz-Kagan T. 2020. Using reflectance spectroscopy for detecting land use effects on soil quality in drylands. Soil Tillage Res. 199:104571. https://doi.org/10.1016/j.still.2020.104571

Li C et al. 2024. Estimation of the wetting induced settlement of loess soils from the wetting soil-water characteristic curve. Sci Rep. 14(1):31533. https://doi.org/10.1038/s41598-024-83258-x
Li Y, Rahardjo H, Satyanaga A, Rangarajan S, Lee DTT. 2022. Soil database development with the application of machine learning methods in soil properties prediction. Eng Geol. 306:106769. https://doi.org/10.1016/j.enggeo.2022.106769
Liu S et al. 2023. Prediction of soil organic carbon in soil profiles based on visible–near-infrared hyperspectral imaging spectroscopy. Soil Tillage Res. 232:105736. https://doi.org/10.1016/j.still.2023.105736
Malo SK, Rahaman S, Saha S. 2023. Assessment of soil fertility status using soil nutrient index (SNI)With special reference to pH, SOC, EC, N, P, K and S of Jabalpur block in Jabalpur district, M.P, India. International Journal of Multidisciplinary. 8(12):114–126.
Martens H, Næs T. 1989. Multivariate Calibration. Chichester, UK: John Wiley and Sons; p. 419.
Maurya S., Abraham J. S, Somasundaram S., Toteja R., Gupta R, Makhija S. 2020. Indicators for assessment of soil quality: a mini-review. Environ Monit Assess. 192:1–22.
Mesfin S, Gebresamuel G, Haile M, Zenebe A. 2021. Modelling spatial and temporal soil organic carbon dynamics under climate and land management change scenarios, Northern Ethiopia. Eur J Soil Sci. 72(3):1298–1311. https://doi.org/10.1111/ejss.13060
Minasny B, Tranter G, McBratney AB, Brough DM, Murphy BW. 2009. Regional transferability of mid-infrared diffuse reflectance spectroscopic prediction for soil chemical properties. Geode. 153:155–162. https://doi.org/10.1016/j.geoderma.2009.07.021
Mondal B. P., Sekhon B. S. 2019. Using diffuse reflectance spectroscopy for assessing soil phosphorus status of an intensively cropped region. Agric Res J. 56:657. https://doi.org/10.5958/2395-146x.2019.00102.9
Mondal BP et al. 2024. Spatial variability of soil microbiological properties under different land use systems. African Journal of Agricultural Research. 20(9):825–839. https://doi.org/10.5897/AJAR2024.16720
Mondal BP et al. 2025a. Comparison of multivariate machine learning models for major soil nutrients prediction using laboratory-based and airborne (AVIRIS-NG) visible near-infrared spectroscopy. Eur J Agron. 170:127726. https://doi.org/10.1016/j.eja.2025.127726
Mondal BP et al. 2025b. Digital mapping of DTPA extractable micronutrients using combined AVIRIS-NG hyperspectral and multispectral data of Sentinel-2 and SRTM-DEM. Comput Electron Agric. 239:110905. https://doi.org/10.1016/j.compag.2025.110905
Morris RV, Jr, Lauer HV, Lawson CA, Jr, Gibson EK, Nace GA, Stewart C. 1985. Spectral and other physicochemical properties of submicron powders of hematite (α-Fe_2O_3), maghemite (γ-Fe_2O_3), magnetite (Fe_3O_4), goethite (α-FeOOH), and lepidocrocite (γ-FeOOH). Journal of Geophysical Research: Solid Earth. 90(B4):3126–3144. https://doi.org/10.1029/JB090iB04p03126
Moursy AR. 2025. Hyperspectral remote sensing as an alternative to conventional methods of soil analysis. Smart Technologies in Sustainable Agriculture. Apple Academic Press; p. 221–254.
Moursy AR, Hassan MN, Elhefny TM. 2022. Sampling and analysis of soil and water: a review. Int J Geogr Geol Environ. 4:34–41.
RCoreTeam. 2018. R:ALanguage and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria https://www.R-project.org/.
Mustafa AERA, Moursy AR. 2020. Using a multivariate regression model and hyperspectral reflectance data to predict soil parameters of Agra, India. International Journal of Geography, Geology and Environment. 2(1):4–9. https://doi.org/10.22271/27067483.2020.v2.i1a.12
Nadporozhskaya M, Kovsh N, Paolesse R, Lvova L. 2022. Recent advances in chemical sensors for soil analysis: a review. Chemosensors. 10(1):35. https://doi.org/10.3390/chemosensors10010035
Naimi S, Ayoubi S, Di Raimo LADL, Dematte JAM. 2022. Quantification of some intrinsic soil properties using proximal sensing in arid lands: application of Vis-NIR, MIR, and pXRF spectroscopy. Geoderma Reg. 28:e00484. https://doi.org/10.1016/j.geodrs.2022.e00484
Nawar S, Mouazen AM. 2022. Combining mid infrared spectroscopy with stacked generalisation machine learning for prediction of key soil properties. Eur J Soil Sci. 73(6):e13323. https://doi.org/10.1111/ejss.13323
Ng W, Minasny B, Jeon SH, McBratney A. 2022. Mid-infrared spectroscopy for accurate measurement of an extensive set of soil properties for assessing soil functions. Soil Security. 6:100043. https://doi.org/10.1016/j.soisec.2022.100043
Nocita M et al. 2015. Soil spectroscopy: an alternative to wet chemistry for soil monitoring. Adv Agron. Vol. 132; p. 139–159. https://doi.org/10.1016/bs.agron.2015.02.002
Omran ESE. 2016. Inference model to predict heavy metals of Bahr El-Baqar soils, Egypt using spectroscopy and chemometrics technique. Modeling Earth Systems and Environment. 2:1–17. https://doi.org/10.1007/s40808-016-0259-7
Patel BK, Singh PK, Dubey V, Sharma N, Patel BL, Pandey R. 2023. Soil profiling and physical parameters of leguminous crops grown soils around established industries in Singrauli (M.P.). International Journal of Pharmaceutical Research and Applications. 8(6):2150–2161.
Paul SS, Coops NC, Johnson MS, Krzic M, Chandna A, Smukler SM. 2020. Mapping soil organic carbon and clay using remote sensing to predict soil workability for enhanced climate change adaptation. Geode. 363:114177. https://doi.org/10.1016/j.geoderma.2020.114177
Piccini C et al. 2024. In-field soil spectroscopy in Vis–NIR range for fast and reliable soil analysis: a review. Eur J Soil Sci. 75(2):e13481. https://doi.org/10.1111/ejss.13481

Potdar RP, Shirolkar MM, Verma AJ, More PS, Kulkarni A. 2021. Determination of soil nutrients (NPK) using optical methods: a mini review. J Plant Nutr. 44(12):1826–1839. https://doi.org/10.1080/01904167.2021.1884702
Prashanth DS, Mehta RVK, Sharma N. 2020. Classification of handwritten Devanagari number–an analysis of pattern recognition tool using neural network and CNN. Procedia Comput. Sci. 167:2445–2457. https://doi.org/10.1016/j.procs.2020.03.297
Ramírez PB, Calderón FJ, Jastrow JD, Ping CL, Matamala R. 2023. Applying NIR and MIR spectroscopy for C and soil property prediction in Northern cold-region ecosystems. Which approach works better?. Geoderma Regional. 32:e00617
Reeves Iii JB. 2010. Near-versus mid-infrared diffuse reflectance spectroscopy for soil analysis emphasizing carbon and laboratory versus on-site analysis: where are we and what needs to be done?. Geode. 158(1-2):3–14. https://doi.org/10.1016/j.geoderma.2009.04.005
Safanelli JL et al. 2023. An interlaboratory comparison of mid-infrared spectra acquisition: instruments and procedures matter. Geode. 440:116724. https://doi.org/10.1016/j.geoderma.2023.116724
Santhi R et al. 2018. Soil fertility appraisal for Villupuram district of Tamil Nadu using GPS and GIS techniques. J Indian Soc Soil Sci. 66(2):158–165. https://doi.org/10.5958/0974-0228.2018.00020.8
Shi Z et al. 2023. Comparison of depth-specific prediction of soil properties: MIR vsdot Vis-NIR spectroscopy. Senso. 23(13):5967. https://doi.org/10.3390/s23135967
Shrivastava AK, Thakur RK, Bisen NK, Rai SK, Sarvade S. 2023. Response of integrated nutrient management on crop productivity and soil fertility under rice–wheat cropping system in Chhattisgarh plain agro-climatic zone. Journal of Experimental Agriculture International. 45(12):201–208. https://doi.org/10.9734/jeai/2023/v45i122279
Singh K. 2016. Microbial and enzyme activities of saline and sodic soils. Land Degradation & Development. 27(3):706–718. https://doi.org/10.1002/ldr.2385
Singh K, Sharma S, Saraswat A. 2023. Assessment of soil fertility status of different villages of Depalpur block of Indore district, Madhya Pradesh, India. Environment and Ecology. 41(1A):191–196.
Soriano-Disla JM, Janik LJ, Viscarra Rossel RA, Macdonald LM, McLaughlin MJ. 2014. The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. ApSRv. 49(2):139–186. https://doi.org/10.1080/05704928.2013.811081
Srinidhi P et al. 2020. Physico-chemical analysis of soils in Madanapalle block, Chittor district of Andhra Pradesh. Int J Commun Syst. 8(3):154–158.
Stenberg B, Viscarra-Rossel RA. 2010. Diffuse reflectance spectroscopy for high resolution soil sensing. In proximal soil sensing, progress in soil science 1. In: Viscarra-Rossel RA29–47pp.
Stenberg B, Viscarra Rossel RA, Mouazen AM. 2010. Visible and near infrared spectroscopy in soil science. Adv Agron. Vol. 107; p. 163–215. https://doi.org/10.1016/S0065-2113(10)07005-7 andWetterlind, J.
Stenberg B, Rossel RAV, Mouazen AM, Wetterlind J. 2010. Visible and near infrared spectroscopy in soil science. Adv Agron. Vol. 107; p. 163–215. https://doi.org/10.1016/S0065-2113(10)07005-7
Stoner ER, Baumgardner MF. 1981. Characteristic variations in reflectance of surface soils. SSSAJ. 45(6):1161–1165. https://doi.org/10.2136/sssaj1981.03615995004500060031x
Tang L, Tian G, Dai G, Zhai Q, Rahardjo H, Satyanaga A. 2022. Effect of threshold suction on the prediction of the permeability function by using the statistical method. Results in Engineering. 14:100456. https://doi.org/10.1016/j.rineng.2022.100456
Tavakoli H, Correa J, Sabetizade M, Vogel S. 2023. Predicting key soil properties from Vis-NIR spectra by applying dual-wavelength indices transformations and stacking machine learning approaches. Soil Tillage Res. 229:105684. https://doi.org/10.1016/j.still.2023.105684
Teixeira MPR et al. 2024. Near-infrared spectroscopy as an alternative tool for predicting soil erodibility in a watershed under desertification. Land Degradation & Development. 35(4):1526–1540. https://doi.org/10.1002/ldr.5003
Thabit FN, El-Shater AH, Soliman W. 2023. Role of silt and clay fractions in organic carbon and nitrogen stabilization in soils of some old fruit orchards in the Nile floodplain, sohag governorate, Egypt. Journal of Soil Science and Plant Nutrition. 23(2):2525–2544. https://doi.org/10.1007/s42729-023-01209-3
Thabit FN, Moursy AR. 2023. Sensors efficiency in smart management of the environmental resources. Handbook of Nanosensors: Materials and Technological Applications. Cham: Springer Nature Switzerland; p. 1–40.
Thakur N, Sharma R, Kumar A, Sood K. 2021. Soil fertility appraisal for pea growing regions of himachal pradesh using GPS and GIS techniques. Indian J Agric Res. 55(4):452–457.
Tziachris P, Aschonitis V, Chatzistathis T, Papadopoulou M. 2019. Assessment of spatial hybrid methods for predicting soil organic matter using DEM derivatives and soil parameters. Catena. 174:206–216. https://doi.org/10.1016/j.catena.2018.11.010
Viscarra Rossel RA, Behrens T. 2010. Using data mining to model and interpret soil diffuse reflectance spectra. Geode. 158(1–2):46–54. https://doi.org/10.1016/j.geoderma.2009.12.025
Viscarra Rossel RA, McBratney AB. 1998. Laboratory evaluation of a proximal sensing technique for simultaneous measurement of soil clay and water content. Geode. 85:19–39. https://doi.org/10.1016/S0016-7061(98)00023-8
Viscarra Rossel RA, Walvoort DJJ, McBratney AB, Janik LJ, Skjemstad JO. 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geode. 131:59–75. https://doi.org/10.1016/j.geoderma.2005.03.007

Viscarra Rossel RA et al. 2022. Diffuse reflectance spectroscopy for estimating soil properties: a technology for the 21st century. Eur J Soil Sci. 73(4):e13271. https://doi.org/10.1111/ejss.13271

Walkley A, Black IA. 1934. An examination of the Degtjareff method for determining organic cabon in soil: effect of variation in digestion condition and of inorganic soil constituents. Soil Sci. 63:251–263.

Wang L, Wang R. 2022. Determination of soil pH from Vis-NIR spectroscopy by extreme learning machine and variable selection: a case study in lime concretion black soil. Spectrochim Acta, Part A. 283:121707. https://doi.org/10.1016/j.saa.2022.121707

Weindorf DC, Chakraborty S, Herrero J, Li B, Choudhury C, Casta˜neda CA. 2016. Simultaneous assessment of key properties of arid soil by combined PXRF and Vis–NIR data. Eur J Soil Sci 67(2):173–183. https://doi.org/10.1111/ejss.12320

Wetterlind J, Stenberg B, Jonsson A. 2008. Near infrared reflectance spectroscopy compared with soil clay and organic matter content for estimating within-field variation in N uptake in cereals. Plant Soil. 302:317–327. https://doi.org/10.1007/s11104-007-9489-9

Xu S, Zhao Y, Wang M, Shi X. 2018. Comparison of multivariate methods for estimating selected soil properties from intact soil cores of paddy fields by Vis–NIR spectroscopy. Geode. 310:29–43. https://doi.org/10.1016/j.geoderma.2017.09.013

Xu L, Mei X, Chang J, Wu G, Jin Q, Wang X. 2021. Rapid assessment of quality changes in French fries during deep-frying based on ATR-FTIR spectroscopy combined with artificial neural network. J Oleo Sci. 70:1373–1380. https://doi.org/10.5650/jos.ess21006

Yadav TC, Singh YP, Yadav SS, Singh A, Patle T. 2023. Spatial variability in soil properties, delineation site-specific management division based on soil fertility using fuzzy clustering in Gwalior, Madhya Pradesh, India. Int. J. Plant Soil Sci. 35(6):49–78. https://doi.org/10.9734/ijpss/2023/v35i62840

Yang Y, Tong X, Zhang Y. 2020. Spatial variability of soil properties and portable X-Ray fluorescence-quantified elements of typical golf courses soils. Sci Rep. 10(1):1–14. https://doi.org/10.1038/s41598-020-57430-y

Yang M, Xu D, Chen C, Li H, Shi Z. 2019. Evaluation of machine learning approaches to predict soil organic matter and pH using vis-NIR spectra. Senso. 19:263. https://doi.org/10.3390/s19020263. 1-14.

Yeneneh N, Elias E, Feyisa GL. 2022. Assessment of the spatial variability of selected soil chemical properties using geostatistical analysis in the north-Western highlands of Ethiopia. ActaAgriculturaeScandinavica, Section B—Soil & Plant Science. 72(1):1009–1019. https://doi.org/10.1080/09064710.2022.2142658

Yudina A, Kuzyakov Y. 2023. Dual nature of soil structure: the unity of aggregates and pores. Geode. 434:116478. https://doi.org/10.1016/j.geoderma.2023.116478

Zayani H et al. 2023. Using machine-learning algorithms to predict soil organic carbon content from combined remote sensing imagery and laboratory Vis-NIR spectral datasets. Remote Sens. 15(17):4264. https://doi.org/10.3390/rs15174264

Zornoza R, Guerrero C, Mataix-Solera J, Scow KM, Arcenegui V, Mataix-Beneyto J. 2008. Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils. Soil Biol Biochem. 40:1923–1930.