The Egyptian Journal of Remote Sensing and Space Sciences 26 (2023) 901–918

Contents lists available at ScienceDirect

The Egyptian Journal of Remote Sensing and Space Sciences

journal homepage: www.sciencedirect.com

Research Paper

Prediction of soil nutrients through PLSR and SVMR models by VIs-NIR reflectance spectroscopy

Chiranjit Singha [a], Kishore Chandra Swain [a], Satiprasad Sahoo [b,*], Ajit Govind [b]

[a] Department of Agricultural Engineering, Institute of Agriculture, Visva-Bharati, Sriniketan, Birbhum, West Bengal 731236, India
[b] International Center for Agricultural Research in the Dry Areas (ICARDA), 2 Port Said, Victoria Sq, Ismail El-Shaaer Building, Maadi, 15A, Cairo 11728, Egypt

ARTICLE INFO

Keywords:
VIs-NIR spectroscopy
PLSR
SVMR
Soil nutrient prediction
Soil suitability mapping
Sentinel 2

ABSTRACT

Though soil nutrients play important roles in maintaining soil fertility and crop growth, their estimation requires direct soil sampling followed by laboratory analysis incurring huge cost and time. In this research work, soil nutrients were predicted using VIs-NIR reflectance spectroscopy (range 350–2500 nm) with Partial Least Squares Regression (PLSR) and Support Vector Machine Regression Model (SVMR) model through principal component analysis. Two hundred soil samples were collected from Tarekswar, Hooghly, West Bengal, India to predict eight selected soil nutrients, such as soil organic carbon (OC), pH, available nitrogen (N), available phosphorus (P), available potassium(K), electric conductivity (EC), zinc (Zn) and soil texture (sand, silt, and clay) levels. The OC content was predicted with sound accuracy (R^2: 0.82, RPD: 2.28, RMSE: 0.13, RPIQ: 4.15 FD-SG), followed by P (R^2: 0.71, RPD: 1.83, RMSE: 4575, RPIQ: 3.44 1st derivative). The soil parameters sensitive to the particular band of visible spectrum were also identified viz. wavelengths of 409, 444, 591 and 592 nm for OC, 430 and 505 nm for P, 464 nm for K; 580 nm for Zn, 492,511,596 and 698 nm for N; 493, 569 and 665 nm for EC; 492,567 and 652 nm for pH; 457 nm for sand and 515 nm for clay.

The soil nutrient levels were predicted by PLSR and SVMR models through PCA and Sentinel 2 imagery and soil suitability map were also generated for seven soil parameters such as OC, pH, EC, N, P, K and clay content. Through map query tool in ArcGIS software environment the PLSR and SVMR model successfully identify the suitability class with level of accuracy of 87.2% and 88.9%, respectively, against the direct soil analysis based suitability mapping.

The machine learning technique based soil nutrient and soil suitability prediction can be easily adopted in different regions. This will reduce the cost of laboratory soil analysis and optimize the total time requirement.

1. Introduction

Digital soil mapping (DSM) through precision agriculture adoption is one of the best management practices for enhancing the crop yield from limited land resources (Grunwald and Vasques, 2015). The quality of soil mapping can be enhanced using the Geographical information system (GIS) and remote sensing spectral technologies. The remote sensing technique is the most promising non-destructive alternative approach for estimating soil properties both for small and large scale with the high efficiency and rapid data acquisition mode (Waiser et al., 2007). Remote sensing images collected through satellites are analysed using GIS software environment to identify various soil nutrient levels being verified through groundtruthing data. Various soil samples need to be collected and analysed in laboratory during groundtruthing. In agriculture, the scale of soil status study for better land and crop management is so far limited. In direct estimation processes, soil physio-chemical properties are more accurately determined compared to indirect technique such as remote sensing techniques; however, there is need for intensive ground surveys and unlimited human resources for direct measurement (Viscarra Rossel et al., 2006). Soil spectroscopy (with 350–2500 nm wavelength) is one of the most promising techniques for quick judgments of soil nutrient status. This technique also reduces the cost and time of soil analysis compare to traditional laboratory sampling methods (Soriano-Disla et al., 2014). Diffuse reflectance spectroscopy (DRS) is generally used for real-time basis computational storage in measuring spectra in terms of databases called spectral libraries (Viscarra Rossel et al., 2006). The DRS technique has been preferred over the visible spectrum, near-infrared, and shortwave infrared (VNIR–SWIR) region with 350–2500

* Corresponding author.
E-mail address: satispss@gmail.com (S. Sahoo).

https://doi.org/10.1016/j.ejrs.2023.10.005
Received 29 April 2023; Received in revised form 6 August 2023; Accepted 27 October 2023
Available online 10 November 2023

nm, for the non-invasive technique in near real-time soil properties measurements (Ben-Dor et al., 2009). Since last few decades, DRS has been used to estimate soil physio-chemical properties such as sand, silt, and clay content (Reeves, 2010), soil organic matter (Morgan, 2009), and soil pH (Tekin et al., 2013), soil electrical conductivity (Shrestha, 2006), available nitrogen (Kuang and Mouazen, 2011), available phosphorous (Rakotonindrina et al. 2020), Available potassium (Liu et al., 2020a) etc. Effective use of spectral radiation is vibrant to enhance the model accuracy in soil constraint estimation for emerging indicators of soil health status. The proximal soil sensing is generally carried out for information gathering at higher sampling resolution (Kuang et al., 2012). The primary advantage of Visible to Near-infrared (VIs-NIR) spectroscopy is to facilitate quick and accurate assessment of soil parameters at low cost. This also helps in continuous mapping and monitoring of soil parameters in the agricultural sensor ecosystem. In precision agriculture study, the interest is often limited to characterize the level of variability in a single or several fields within a relatively small region.

Chemometric modeling technique has been used for analyzing complex spectral patterns. This modelling is capable of dealing with a large number of variables including range spectral bands. The whole spectrum of soil spectroscopy can be used to predict a soil nutrient parameter. The most intuitive multivariate techniques in soil spectroscopy are, the partial least squares regression (PLSR) technique, elastic net regression (ENR) and support vector machine regression (SVMR) technique (Savitzky and Golay, 1964) etc. In PLSR, the total number of variables is reduced to a smaller unrelated set of components and least square regression was performed. During smoothing of the spectra estimates, the average of several sample points of the spectrum were carried out to reduce the effect of random noise. Even the derivatives of the spectra (taking either first or second order) is taken to optimize the peaks. The pre-treatment technique also carries out for image corrections of multiplicative effects on the spectral estimates, such as multiplicative scatter correction (MSC) and the standard normal variate (SNV) techniques etc. (Geladi et al., 1985).

Sentinel 2 images with 10 m accuracy can represent the top soil variability in terms of various nutrients. Twelve out of thirteen bands are used for mapping soil and crop parameters (Baroudy et al., 2020). As the images are available free of cost and can be accessed through google cloud, these make them images more viable for prediction soil parameters.

The machine learning models enhanced the model accuracy for predicting land suitability through soil nutrient analysis (El-Sayed et al., 2023). These models can be directly used for prediction of soil nutrients through training and testing of the available soil nutrients levels. Suitable models for predicting the soil nutrient variation may be recommended for future application. A no. of studies has been carried out for estimating soil suitability by measuring soil nutrient through direct method and mapping them in GIS software environment. However, the process involves a huge amount of cost and time, which sometimes nearly impossible for larger areas. This gap can be somewhat fulfilled by using machine learning based soil nutrient prediction through groundtruthing soil sampling.

The study is intended to estimate major soil nutrients, such as available nitrogen (N), available phosphorus (P), available potassium (K), soil pH, organic carbon (OC), electrical conductivity (EC), available zinc (Zn) and soil texture being composed of sand, silt and clay, through in-situ VIs-NIR spectroscopy (with 350–2500 nm wavelength) and also carry out prediction through two machine learning models namely PLSR and SVMR. The study will validate the machine learning models for carrying out land suitability analysis using seven major soil constituents.

2. Materials and methods

2.1. Study site and sampling

Overall, 200 soil samples were directly collected from the study area at 0–30 cm depth, namely, *Tarakeswar* Block, Hooghly district, West Bengal, India covering approximately 300 ha. The study area is located at 2,528,500 to 2,530,600 Northing and 604,500 to 606,500 Easting in UTM projection system for Zone 45 under WGS 1984 Datum system at 40 m from Mean Sea Level (Fig. 1). As a part of Gangetic alluvial plains, the predominant group of soil is sandy loam and loamy soils covering nearly 32 % and 48 %, respectively, of the total cultivated area (Singha et al., 2022).

2.2. Soil sampling and laboratory analysis

The soil survey was conducted during the year 2019–2020 in the study area by collecting soil samples from each plot as per the procedures mentioned in the Soil Survey Manuals (Soil Survey Staff, 1993). The soil samples were collected for a depth of 0–30 cm through the special soil auger system. Each sampling positions was marked and recorded by a handheld GPS receiver (GPS III Plus, Garmin, Olathe, Kansas, USA) with 10–30 m accuracy. All two hundred soil samples were screened through a 2.0 mm sieve and supplied for future analysis in the laboratory. Soil physio-chemical properties were measured through standard techniques, such as pH through water/soil solution ratio of 1:2.5; available nitrogen (N) through Kjehldahl apparatus (Subbaiah and Asija, 1965); available phosphorus (P) following Olsen procedure (Olsen et al., 1954), and organic carbon (OC) determined through Walkely and Black procedure, Available zinc (Zn) by DTPA method (with estimation by Perkin Elmer Atomic Absorption Spectrophotometer). Percentages of sand (>50 μm), silt (2 to 50 μm), and clay (<2 μm) were determined through Hydrometer, and the textural class was identified using the USDA triangle (Piper, 1942). The readings were estimated on the parts per million (PPM) scale. The proposed methodology of the current study summarized (Fig. 2) before the data collection.

2.3. Spectral acquisition using VIs–NIR spectroscopy

For optimized irradiance level, the estimation of spectral reflectance for the soil samples were carried out in a dark experimental setup. ASD FieldSpec® 3 portable spectrometer (Analytical Spectral Devices Inc., Boulder, CO, USA) with a spectral wavelength in the range of 350–2500 nm was used for measuring the reflectance spectrum of each soil sample. The sampling interval of the spectrometer is 1.4 nm for the range of 350–1000 nm and 2 nm for the range 1000–2500 nm, with the resampling interval of 1.0 nm. The specification of the circular containers used to keep the soil sample during the experiment was of 12.0 cm diameter X 1.8 cm depth (though 1.5 cm is considered optically enough thick for soil). About fifty grams of soil was taken in an aluminium soil moisture can of 5 cm diameter and the surface was levelled with a rubber cork being used as a mallet (Liu et al., 2009). Four reflectance spectra were taken for each soil sample from different quadrants over the central part of the circular container. Spectral reflectance from a Spectra on a standard white reference panel (99 % reflectance, Labsphere) was also acquired to optimize the measurement before soil sample analysis. The scan number for the experiment was set at 30 for obtaining an average of thirty scans for each spectrum readings. In this fashion, the reflectance spectra of each soil sample was gathered and recorded with the help of RS3 software application associated with the radiometer.

2.4. VIs–NIR spectral processing

Spectral transformation processing can be carried out by many spectral methods. Wang et al. (2019) reported that spectral processing is necessary to apply a variety of mathematical transformation methods to

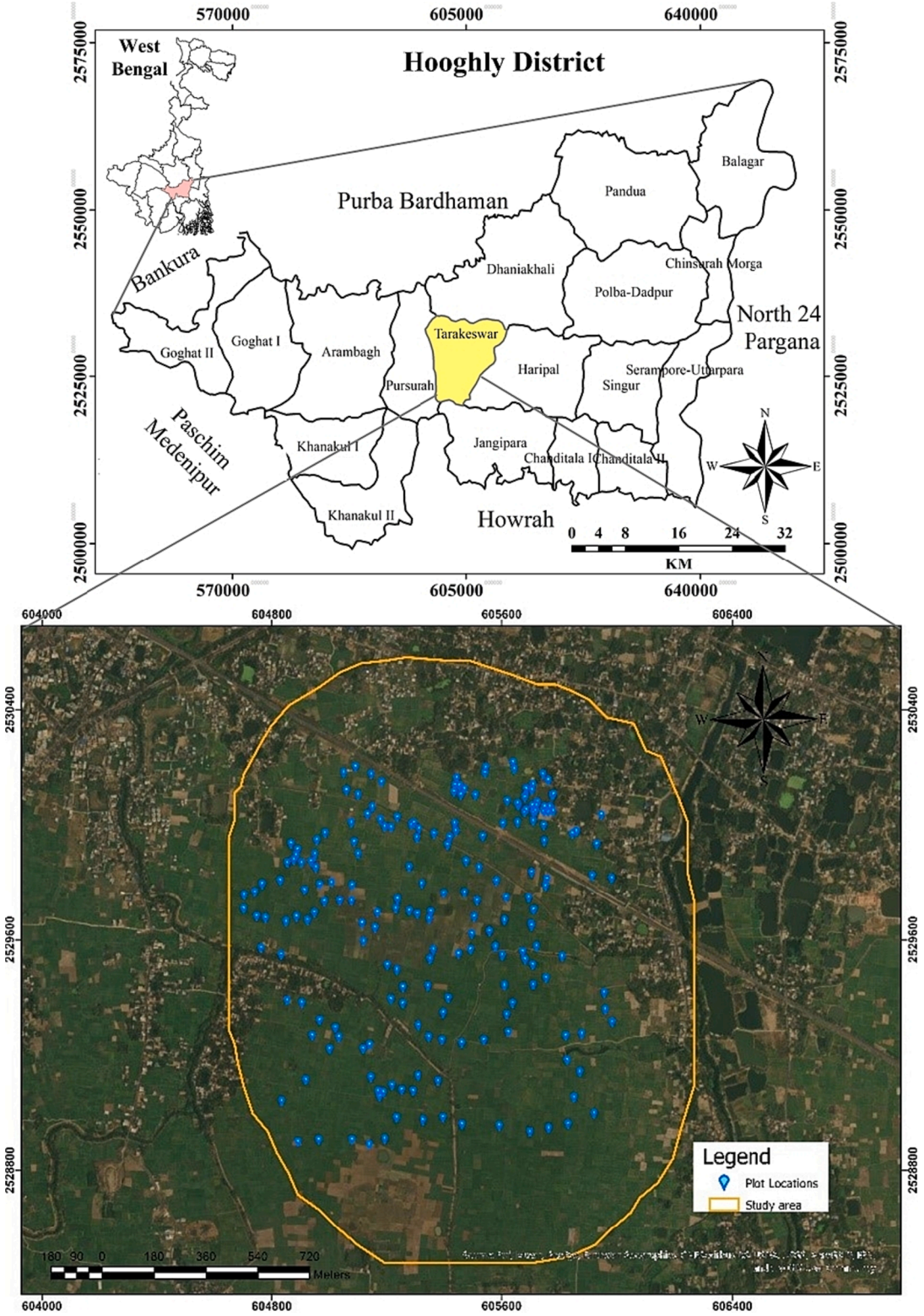

Fig. 1. Study area map.

identify best possible solution. To reduce noise and enhance the signal quality, first-derivative spectra with SG smoothing (FD-SG) and second-derivative spectra with SG smoothing (SD-SG) reflectance using a Savitzky–Golay smoothing filter were used with second order derivative (Savitzky and Golay, 1964; Ludwig et al., 2019). The effect of noise and soil particle size should be optimized for sound readings. The scatter correction with a standard normal variate transform (SNV) technique was used for each spectrum of the whole spectral length. Furthermore, the soil spectral range was optimized to 400–2490 nm range for better accuracy. The overlapped samples were separated to extract the spectral information, as the readings had little variation.

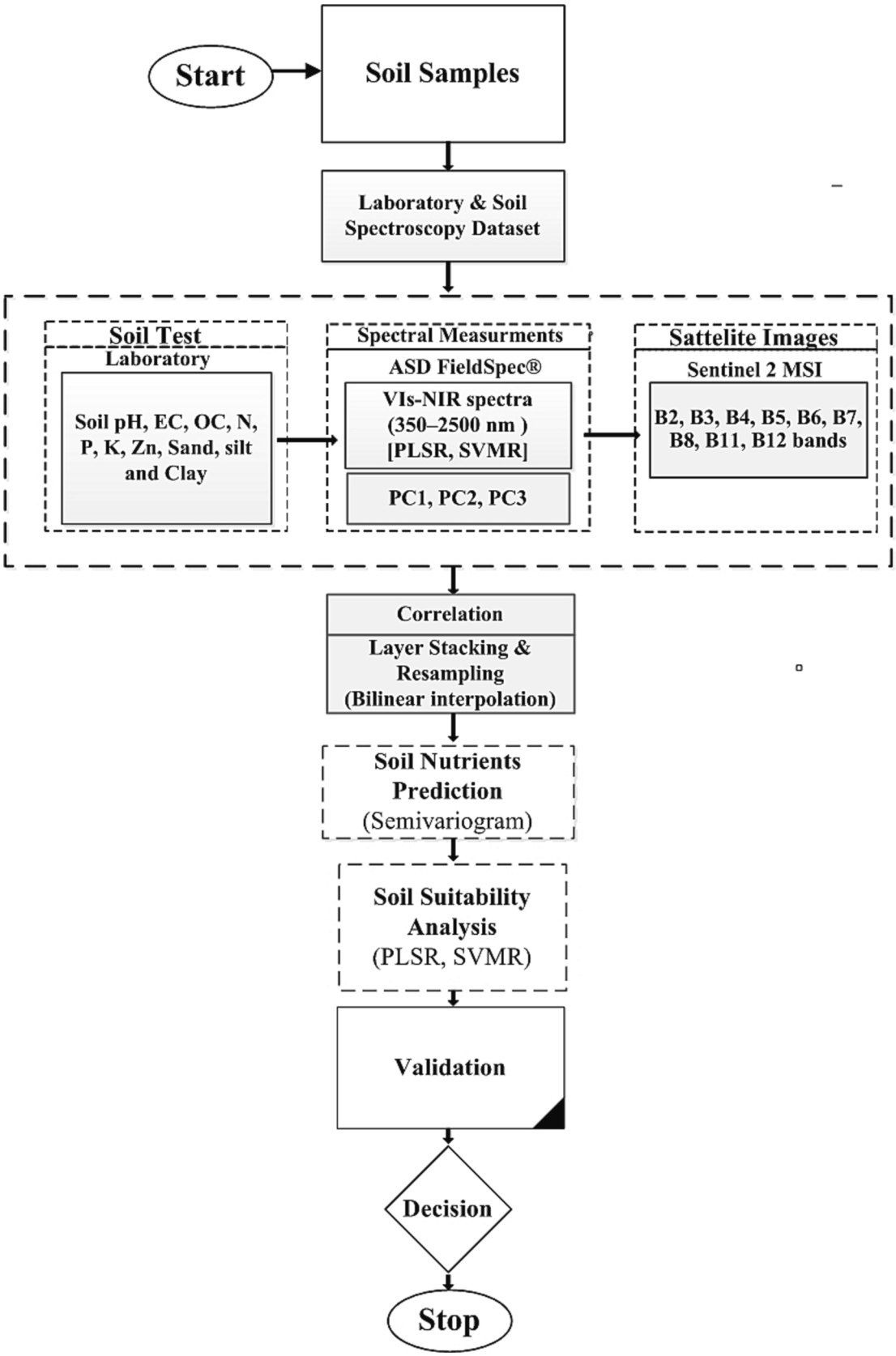

Fig. 2. Systematic methodology for the soil nutrients prediction at Tarakeswar region.

2.5. Chemometric calibration

The calibration technique, used for spectral data analysis, is highly influenced by the selected multivariate statistical approaches (Geladi, 2003) and its filters. For the spectral calibration, VNIR spectra and their matching laboratory data were supplied to the Unscrambler X10.5 software (CAMO, Norway). The dataset was divided into a training set (nearly 70 % of the data) and a validation set (nearly30% of the data). The training set was then analyzed to choose the best pre-treatment options for the prediction spectral wavelength with PLSR and SVMR models. The validation and calibration sets were separated using the sample set partitioning method with x-y distance procedure (Galvão et al. 2005). The entire dataset was sorted based on the nutrients content, and each stratum was separated into a set of validated data for the model validation. A total of 31 samples were selected from the 31 strata for the validation. The smoothed spectra generated using the parameters for CV exhibited the lowest RMSE with the optimal latent variables. Different combinations of pre-treatments including trimming, smoothing (Savitzky-Golay smoothing with FD-SG) and SNV were applied to the spectra which were then used for PLSR and SVMR model run. From these runs, the model performance parameters, such as R^2, Root mean square error (RMSE), Residual prediction deviation (RPD), and Ratio of performance to interquartile distance (RPIQ) were estimated. The optimal number of LVs needed was used as a selection criterion in pre-treatment process. For a successful model, R^2, RPD, RPIQ should be maximized followed by minimized RMSE for minimized number of LVs. Theoretically, the R^2 can be upgraded to touch the value of 1 with additional LVs, which may increase the RMSE and sets the model for overfitting.

2.6. PLSR model

Partial least square regression (PLSR) model is a statistical model which develops regression models by projecting the targeted variables and observable variables into a new space. This bilinear factor model generally works by defining covariance pair by pair. PLSR is a well-known multivariate predictive approach for the chemometric calibration through pre-processing of VIs-NIR spectra for soil properties (Wold et al., 2001). The repeated double cross-validation analysis (Filzmoser and Varmuza, 2017) was employed to discover the potential of VIs-NIR data to quantify their predictive performance of soil properties. The PLSR model integrated the algorithms that extract a minimum number of latent factors to act as independent variables relating to reflectance spectra. These factors are used in regression analysis with the soil data as the dependent variables. The locally-weighted PLSR approach was then used to determine the distance metric that could be used to calculate the similarity between the test and the calibration sample (Gupta et al., 2018). Another factor that can affect the calibration of the model is the sample size. Since soil reflectance spectroscopy is a cost-effective method for soil spectroscopy modeling, it is important to choose the appropriate number of samples. The different parameters used in the PLSR model calibration and validation phases were compute the test samples leverage, F test validation, PCA score, Q residuals generated values etc Lastly in the model calibration and validation stages, the different prediction diagnosis were used, namely: root mean square error of prediction (RMSEP), standard error of prediction (SEP), RMS error of calibration, RMS error of validation, bias, slope, offset, deviance score, and correlation.

According to Abdi et al. (2012), the RPD is the ratio of standard deviation (SD) of the estimated data to the standard error (S.E.) of evaluation. The model with the higher R^2, RPD, RPIQ, and the lesser RMSE is considered as one of the best models to predict soil properties. In agricultural applications, RPD value greater than 2 indicates an excellent judgment, values between 1.5 and 2 good predictive abilities; values between 1 and 1.5 acceptable, and values less than 1 considered not satisfactory, which may be improved further for the better analysis (Barthès et al., 2019; Clingensmith et al., 2019).

2.7. Support vector machine regression (SVMR) model

Radial basis kernel function enables the SVM approach for distributing the complex nonlinear patterns to a new hyperspace for pattern classification and regression analysis. It is capable of self-learning in higher-dimensional feature space with less number of training samples (Xuemei and Jianshe, 2013). It enables putting new data set the correct category. The linear kernel with the SVM has only the cost function as an adaptable constraint. SVM showed better performance compared to other models due to its tolerance against spectral noise and outliers (Viscarra et al. 2006). Wenjun et al. (2014) had successfully applied the SVMR model through in-situ VIs-NIR spectroscopic estimations in Zhejiang, China. SVMR prediction accuracy ($R^2 = 0.75$, RPD 1.90) in situ VIs-NIR spectra outperformed the PLSR results for rice crop in China. VIs-NIR based prediction of anodic properties provide better and reliable output with SVM compared to PLSR technique. However, Di Iorio et al. (2019) recommended both SVM ($1.7 < RPD < 1.9$) and PLSR (RPD 1.9) analytical models for Fe and Al dithionite extraction in European countries.

Before PLSR and SVMR analysis, the 200 soil samples were randomly splitted into 70 % for model calibration and 30 % for validation of data sets for leave-one-out cross validation analysis (Wold et al. 2021; Miloš and Bensa, 2018). The SVMR calibration model was able to provide an average RMSE of 5.10 and a good performance with 120 samples.

Regression models were generated with the calibration data set, which were independently validated with the later data set. Models output results evaluated by the R^2, RMSE, RPD, and RPIQ (Chakraborty et al., 2017), (Eq.1–4).

$$R^2 = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 / \sum_{i=1}^{n} (y_i - \bar{y})^2 \quad (1)$$

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}} \quad (2)$$

$$RPD_{val} = sd_{val} / RMSE_{val} \quad (3)$$

$$RPIQ_{val} = Q3 - Q1 / RMSE_{val} \quad (4)$$

Where,

$\hat{y}_i$: Predicted value of *i th* observation,
$\bar{y}$: Mean observed value,
y_i: Observed value of *i th* observation,
n: Number of samples,
sd_{val}: Standard deviation of the validation set,
$RMSE_{val}$: RMSE of validation, and.
Q3 and *Q1*: 3rd and 1st quartiles of the validation set.

2.8. Sentinel-2 image analysis

The Sentinel 2 MSI (S2) on board the ESA's Copernicus mission was designed to provide high-quality images of the land. The cloud free S2 images were acquired on post-harvest March 15, 2019, and featured 10 m of multi-spectral resolution. A total of nine (i.e., B2, B3, B4, B5, B6, B7, B8, B11, B12 bands) spectral bands were employed for various soil nutrient prediction (ST.1). All the images were geo-rectified using the standard methods (atmospheric and geometric correction) and pre-processed using the SNAP software environment. The thermal bands were not included in the analysis. Finally, the layer stacking all the bands with ground overlay analysis for extracting the ground truth values for soil nutrients prediction mapping. The 10 cross validation method applied (70 %:30 % ratio of sample) through the SVMR and PLSR model validation purpose. Here we used the total twelve input parameters (VIs–NIR spectral of PCs: PC1, PC2, PC3 and nine spectrums of S2 (blue to short-wave infrared): soil pH, EC, OC, N, P, K, Zn, sand, silt

and clay for soil nutrients prediction mapping. Meanwhile the target variables were the laboratory measured of the assigned soil nutrients.

2.9. Semivariogram analysis

Semivariogram is a statistical tool which assesses the average decrease in similarity level between two random variables as the distance between the variables increases. It depicts the spatial autocorrelation of the measured sample points (Singha et al., 2020). The SVMR and PLSR models' predicted values were interpolated using ordinary kriging (OK). The OK geostatistics framework utilized a set of analytical tools to determine the value of the soil properties at an unsampled location. It also checked the normal distribution pattern of collected data.

Semivariograms of the selected soil nutrient parameters were then estimated by averaging the differences between the predicted values of each pair of values (Lark, 2000) (Eq.5)

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [K(y_i) - K(y_i + h)]^2 \quad (5)$$

Where,

γ(h): Semivariance for the interval distance class h, N(h): Number of pairs of the lag interval, K(yi): Observed score at point i, and K (yi + h): Observed score at the location (i + h).

The different semi-variogram model packages (i.e. linear, spherical, circular, gaussian and exponential,) used for the analysis were validated and tested suitability by using different fitted parameters, such as range (A_0), nugget (C_0), sill ($C_0 + C$) and degree of spatial dependence (DSD) are then applied in the kriging technique (Zhang et al., 2015). The suitability of the models was checked through the root mean square errors, RMSSE, and MSE score in ArcGIS software environment.

2.10. Suitability analysis and validation

The area suitable for crops were estimated using seven major soil nutrient parameters, such as pH, EC, OC, Clay content, available nitrogen, available phosphorous and available potassium. The level of sand and silt along with available zinc were not included in suitability study. The levels of each nutrient were equally distributed for four classes of suitability, such as highly suitable, suitable, moderately suitable and not suitable class. Based on laboratory soil nutrient analysis, suitability distribution map was developed for the study area.

Two machine learning techniques, such as PLSR and SVMR were used to predict the ten soil nutrients following 70:30 training to testing sample ratio. Crop suitability estimation was then carried out for both the ML models using seven major soil nutrient parameters.

3. Results

The descriptive statistics of levels of the 10 selected soil variables were estimated (Table 1) in the "R" software environment. Values of average soil pH was 4.93, indicating the presence of acidic soil reaction in the district. The zinc (Zn) content for the dataset varied from 0.04 to 2.99 ppm with an average value of 1.5 ppm. The clay content in the samples ranged from 22.32 % to 63.60 %, whereas the percentage of sand in the samples were ranged from 11.68 % to 66.40 %. The average EC value in the region was 0.44, with a range of 0 to 3 %. The average organic carbon in the study area was 0.48 % (Table 1). The coefficient of variation (CV) of > 55 %, showed higher variability, may be due to variation in the strength of soil parent materials, such as N, P, K, Zn and land use pattern in the entire area. Wider soil variability may be beneficial in improving the predictive accuracy of the calibration model (Kuang et al., 2012).

The soil nutrients generally have unique migration pattern and enrichment trends, leading to different level of correlation for each element. The pH is negatively correlated to sand values, whereas EC, OC,

Table 1
Descriptive statistics of soil parameters.

		Mean	Max	Min	Range	SD	CV	Skewness	Kurtosis	Median	Q1	Q3
pH	All	4.93	6.61	4.11	2.50	0.50	0.10	1.33	1.56	4.82	4.62	5.07
	Calibration	4.94	6.61	4.11	2.50	0.54	0.11	1.22	1.25	4.82	4.59	5.13
	Validation	4.93	6.52	4.26	2.26	0.47	0.09	1.49	2.03	4.82	4.65	5.04
EC (dS/m)	All	0.44	3.00	0.01	2.99	0.43	1.00	1.84	6.18	0.39	0.03	0.62
	Calibration	0.43	1.73	0.01	1.72	0.32	0.76	0.99	2.00	0.43	0.20	0.60
	Validation	0.45	3.00	0.01	2.99	0.52	1.18	1.86	5.05	0.32	0.02	0.65
OC (%)	All	0.48	1.10	0.07	1.03	0.26	0.55	0.08	−1.01	0.52	0.20	0.67
	Calibration	0.48	1.04	0.07	0.97	0.23	0.48	0.02	−0.72	0.49	0.30	0.64
	Validation	0.48	1.10	0.07	1.03	0.30	0.61	0.10	−1.29	0.55	0.17	0.71
N(ppm)	All	24.56	90.54	4.00	86.54	16.70	0.68	1.26	2.48	25.09	10.04	30.11
	Calibration	27.27	84.30	4.94	79.36	15.70	0.58	1.39	3.00	25.09	20.07	30.11
	Validation	21.86	90.54	4.00	86.54	17.30	0.79	1.35	2.64	20.98	5.02	30.28
P(ppm)	All	123.45	320.32	0.88	319.44	90.72	0.73	−0.08	−1.53	130.50	16.68	227.94
	Calibration	148.84	231.63	0.88	230.75	90.75	0.61	−0.54	−1.40	211.86	65.17	228.07
	Validation	98.05	320.32	1.41	318.91	83.69	0.85	0.31	−1.14	100.43	12.58	170.19
K(ppm)	All	88.22	298.80	21.58	277.22	52.02	0.59	1.04	0.72	72.38	46.98	124.75
	Calibration	69.88	151.36	25.90	125.46	30.54	0.44	1.01	0.37	63.25	47.89	79.35
	Validation	106.55	298.80	21.58	277.22	61.87	0.58	0.45	−0.51	111.22	46.48	150.13
Zn (ppm)	All	1.50	4.43	0.01	4.42	1.16	0.77	0.27	−0.82	1.50	0.15	2.30
	Calibration	1.65	3.88	0.01	3.87	0.97	0.59	−0.26	−0.76	1.84	0.97	2.31
	Validation	1.35	4.43	0.01	4.42	1.32	0.97	0.64	−0.75	1.01	0.09	2.28
Sand (%)	All	44.67	66.40	11.68	54.72	12.65	0.28	−0.76	−0.04	47.68	38.96	53.68
	Calibration	43.23	66.40	11.68	54.72	13.19	0.31	−0.64	−0.41	45.68	36.40	53.68
	Validation	46.10	66.40	11.68	54.72	11.98	0.26	−0.90	0.54	48.04	39.68	54.00
Silt (%)	All	17.83	36.00	10.00	26.00	5.70	0.32	0.48	−0.48	16.72	12.72	22.72
	Calibration	18.38	36.00	10.00	26.00	5.95	0.32	0.60	−0.06	17.36	14.00	22.72
	Validation	17.29	28.00	10.00	18.00	5.42	0.31	0.29	−1.33	16.00	12.00	22.72
Clay (%)	All	37.50	63.60	22.32	41.28	8.32	0.22	1.03	0.87	36.32	32.32	41.18
	Calibration	38.39	63.60	22.32	41.28	8.53	0.22	0.81	0.28	36.32	32.32	42.32
	Validation	36.61	63.60	22.32	41.28	8.05	0.22	1.30	1.91	35.60	32.32	39.36

Note: EC-electrical conductivity; OC-organic carbon; N-available nitrogen; P-available phosphorus; K-available potassium; ZN-zinc level; SD-standard deviation; CV-coefficient of variation; SD-Standard deviation; Q1-1st quartiles and Q3- 3rd quartiles.

N, P, K, Zn, silt and clay values were positively correlated (Fig. 3). The organic carbon level also positively correlated with pH, EC, N, P, K, Zn, silt and clay, and negatively correlated to sand level. Through correlation analysis, we also founded strong correlations between OC with Zn ($r^2 = 0.61$), which was mainly caused by Zn deposition in soil with high humus presence. The zinc content of soil may change with variation in humus and OC content (Wang et al., 2014). The phosphorus level was also positively associated with pH, OC, EC, N, Zn, silt, and clay, whereas, sand is negatively correlated. Similarly, silt and clay are negatively correlated to sand. These types of relationships will help in assigning weightage for various land suitability studies through AHP, fuzzy logic software based analysis.

3.1. Qualitative description of the soil spectra

Spectral reflectance characteristics supports in identifying the major soil parameters, such as moisture content, soil color, organic matter, clay content, influence of soil minerals, soil surface condition, soil salinity, and alkalinity etc. Mouazen et al. (2005) established that variation of NIR spectral reflectance was directly dependent on the soil textural classes. The quartz present in sand enhances the level and intensity of spectral reflectance and light scattering parameters (Conforti et al., 2018). The sample outlier detection identified with the Leverage/ Hotelling's T^2, Boxplot and F residuals methods. It defines the distance to the model center. The reduction of the spectral information was performed through the use of the PCA process for various applications such as data exploration and outlier detection. Spectral absorption associated with soil elements (metal and water O-H) rose with higher clay content (Stevens et al., 2013). However, the peak level at 2200 nm was used for the absorption of Al-OH lattice soil texture and level of clay mineral (Viscarra Rossel and Behrens 2010). The level of absorption at 1415 nm and 1915 nm wavelengths of the spectrum is very strong and mostly due to the absorption of moisture from the atmosphere (Nocita et al., 2014). The whole VIs-NIR spectral reflectance profile of the study site through the collected soil samples was estimated (Fig. 4). Soil pretreatment under different spectral indices, FD-SG and SD-SG with Standard normal variable transformation SNV method was obtained (Fu et al., 2018). The SNV transformation was applied to optimize the influence of particle size and curvilinear trend of the measured spectrum. The treatments included SG smoothing, in terms of first and second derivatives for second order polynomial with a range of 10 wavelengths, were applied in unification with SNV (SF.1) (Liu et al., 2020b).

3.2. PCA analysis of soil spectra

The first three principal components covered nearly 93 % of data variation (Fig. 5). PC1 and PC2 accounted for 84 % of the observed variance values, whereas the variation of PC3 was around 9 %. PC1 explained 66 % of the variance, having major influence in the VIs region, whereas OC and clay minerals gave better understanding of the soil properties (Schoell et al., 2005). PC2 explained nearly 18 % of the total variance of level of soil properties. Similarly, VIs range had major influence with positive loadings. PC1 showed a positive correlation with pH, EC, OC, N, P, Zn, and silt, whereas the negative correlation to P and sand content only. PC2 showed a positive correlation with EC, OC, N, P, Zn, silt, and clay, while the negative correlation to pH content only. PC3 showed negative correlation with soil texture analysis especially for silt and clay content (Santra et al., 2015). As the three major PCs explained about 93 % variation in the dataset, an original spectral dataset of 2151 dimensions may be reduced to three prominent dimensions namely PC1, PC2 and PC3 as the principal components (Fig. 5). Based on the PCA multivariate calibration models analysis, the three component of PCs parameter were selected for the model evaluation. The critical F value (0.1 % to 25 %), beta coefficient and Q residuals limit estimated for the model performance.

3.3. Prediction of soil properties using PLSR and SVMR

The values of regression coefficients (R^2), root mean square error (RMSE), residual prediction deviation (RPD), and the ratio of

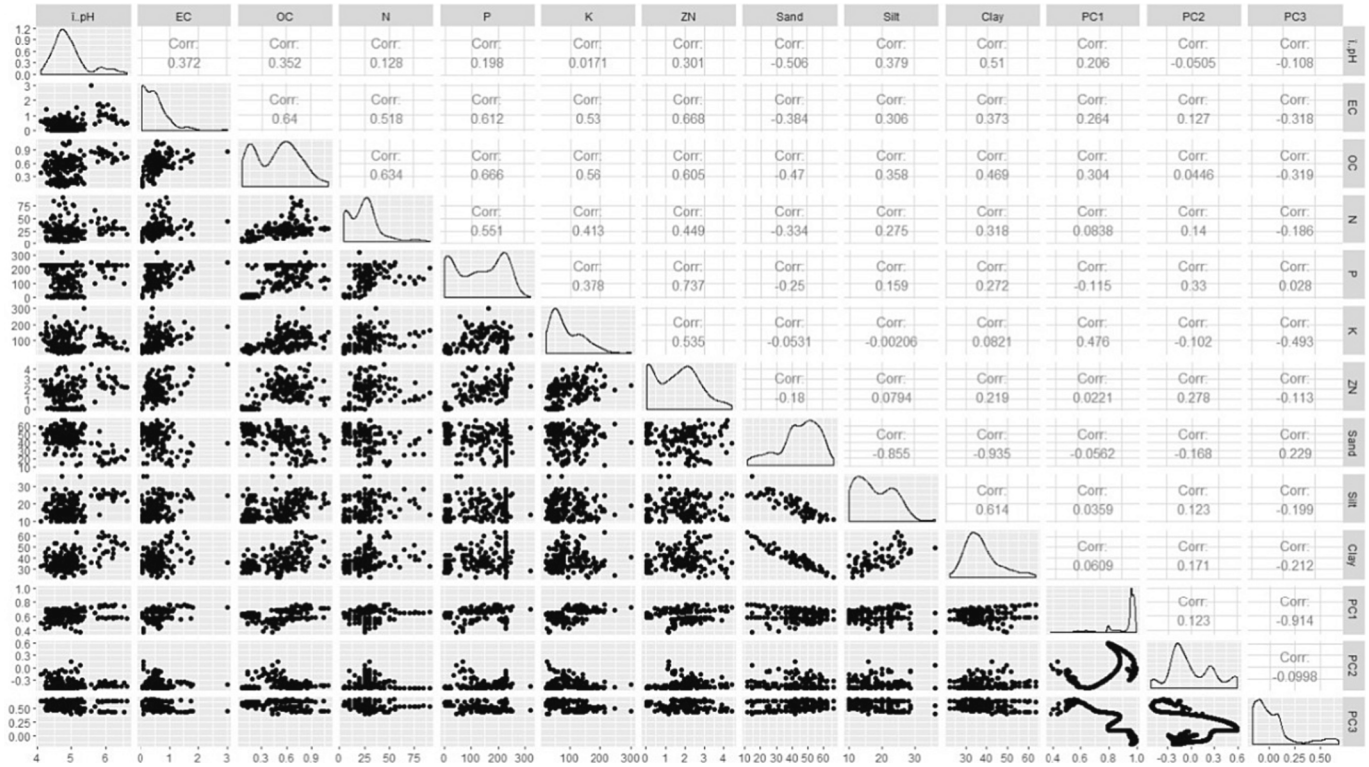

Fig. 3. Correlation measurement distribution of soil properties.

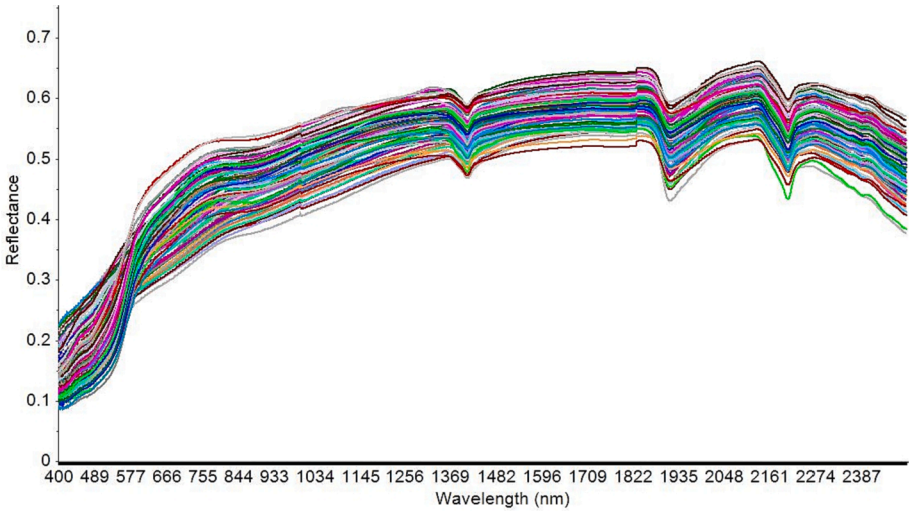

Fig. 4. Spectra measured by an ASD FieldSpec® 3 portable spectrometer USA.

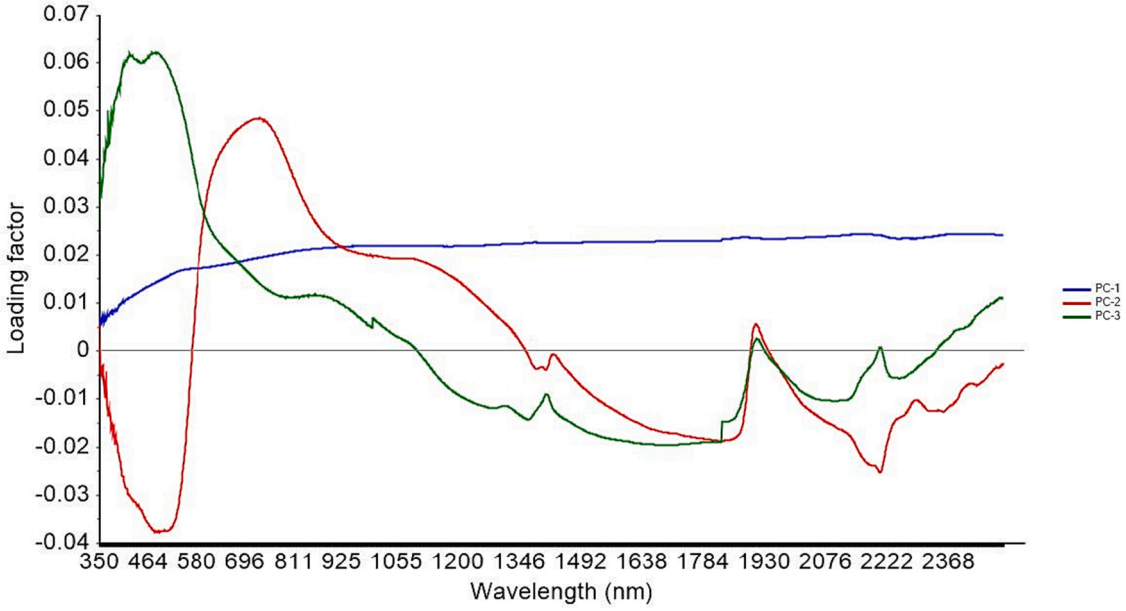

Fig. 5. Loading factor of three major principal components derived from raw reflectance spectra in VIs–NIR.

performance to inter-quartile range (RPIQ) were obtained from each partial least square analysis. The 20 cross validation segments and correlation loadings factor along with scattering plots of measured vs. predicted soil properties were also estimated though PLRS and SVMR models (Tables 2 and 3; Figs. 6 and 7). The maximum number of 20 Latent variables was set to execute the model calibration. In the optimal

Table 2
Calibration and validation results of the PLSR method.

	Calibration 1st Derivative		Validation 1st Derivative				Calibration 2nd Derivative		Validation 2nd Derivative			
Soil Properties	R^2	RMSE	R^2	RMSE	RPD	RPIQ	R^2	RMSE	R^2	RMSE	RPD	RPIQ
pH	0.59	0.34	0.34	0.38	1.23	1.03	0.61	0.34	0.40	0.36	1.30	1.08
EC (dS/m)	0.62	0.20	0.41	0.40	1.31	1.58	0.63	0.19	0.44	0.39	1.34	1.62
OC (%)	0.66	0.13	0.82	0.13	2.28	4.15	0.63	0.14	0.81	0.13	2.28	4.15
N (ppm)	0.42	11.93	0.49	12.34	1.40	2.05	0.36	12.55	0.41	13.28	1.30	1.90
P (ppm)	0.71	48.95	0.71	45.75	1.83	3.44	0.72	48.09	0.68	47.59	1.76	3.31
K (ppm)	0.19	27.57	0.70	34.02	1.82	3.05	0.13	28.55	0.67	35.35	1.75	2.93
Zn (ppm)	0.82	0.41	0.63	0.80	1.65	2.74	0.79	0.44	0.61	0.82	1.62	2.67
Sand (%)	0.61	8.27	0.24	10.47	1.14	1.37	0.60	8.36	0.20	10.69	1.12	1.34
Silt (%)	0.38	4.69	0.16	4.97	1.09	2.16	0.41	4.60	0.09	5.19	1.04	2.07
Clay (%)	0.63	5.19	0.19	7.28	1.11	0.97	0.66	5.02	0.17	7.38	1.09	0.95

Note: EC-Electrical conductivity; OC-Organic carbon; N-Available nitrogen; P-Available phosphorus; K-Available potassium; Zn-Zinc level.

Table 3
Calibration and validation of results through SVMR technique.

SVMR	Calibration 1st Derivative		Validation 1st Derivative				Calibration 2nd Derivative		Validation 2nd Derivative			
	R^2	RMSE	R^2	RMSE	RPD	RPIQ	R^2	RMSE	R^2	RMSE	RPD	RPIQ
pH	0.51	0.37	0.29	0.39	1.20	1.00	0.58	0.36	0.36	0.39	1.20	1.00
EC (dS/m)	0.59	0.21	0.49	0.38	1.38	1.66	0.59	0.21	0.46	0.39	1.34	1.62
OC (%)	0.70	0.12	0.84	0.12	2.47	4.50	0.66	0.13	0.82	0.12	2.47	4.50
N (ppm)	0.37	12.74	0.47	12.89	1.34	1.96	0.40	12.36	0.45	13.04	1.33	1.94
P(ppm)	0.71	48.99	0.70	45.38	1.84	3.47	0.71	49.52	0.68	47.25	1.77	3.34
K (ppm)	0.23	26.87	0.67	35.67	1.73	2.91	0.24	26.87	0.68	34.77	1.78	2.98
Zn (ppm)	0.83	0.39	0.66	0.77	1.71	2.85	0.84	0.39	0.67	0.75	1.76	2.92
Sand (%)	0.69	7.37	0.24	10.44	1.15	1.37	0.66	7.75	0.19	10.78	1.11	1.33
Silt (%)	0.51	4.18	0.17	4.97	1.09	2.16	0.54	4.07	0.10	5.22	1.04	2.05
Clay (%)	0.73	4.44	0.18	7.29	1.10	0.97	0.68	4.85	0.15	7.48	1.08	0.94

Note: EC-electrical conductivity; OC-organic carbon; N-available nitrogen; P-available phosphorus; K-available potassium; Zn-zinc level;

PLRS loadings acquired through 7 Latent variables. These loadings give useful information on the relative contributions of varying wavelengths to the performance of prediction models. FD-SG and SD-SG with SNV datasets were used to test the prediction accuracy for soil attributes. The most suitable PLSR SVMR models were selected through 5 or 6 factors (maximize R^2, RPD, RPIQ, and lowest RMSE). The OC content was predicted with the best accuracy in terms of R^2 of 0.82 (RPD: 2.28, RMSE: 0.13, RPIQ: 4.15 FD-SG), followed by P with R^2 of 0.71 (RPD: 1.83, RMSE: 4575, RPIQ: 3.44 1st derivative), (Table 2). K, Zn, and N predictions were moderately sound from VIs-NIR spectra measurement with R^2 of 0.50–0.70 and RPD of 1.4–1.8, respectively for 1st order derivative. A no. of researchers had already reported successful prediction of N, P, and K with PLSR model using VIs-NIR spectroscopy (Udelhoven et al., 2003; Shao and He, 2011). Zn level was not good enough compared to others, manly as the heavy metal in soil is spectrally featureless and scarcely represented in VIs–NIR spectra (Wang et al., 2014).

EC and pH value had an R^2 of 0.44 and 0.40, RPD of 1.34 and 1.30, respectively, with 2nd order derivative. Soil pH is related to wavelengths of minerals in the VIs-NIR region. The estimation of soil pH was well predicted and more effective compared to P and K (Cohen et al., 2005). VIs-NIR also specifies restrictions to predict particular soil parameters, such as sand, silt, and clay content, for 1st order derivative with RPD between 1.0 and 1.15 (Tables 2 and 3). The measured versus predicted physio-chemical properties levels of all two hundred soil samples were also mapped (Figs. 6 and 7).

The SVMR models predicted OC content well in 1st derivatives, in terms of an RPD of 2.47 and R^2 of 0.84, and provided vital feedback on the important spectra used for partitioning the data. Nawar et al. (2016) also reported different levels of prediction results that were influenced by both on the multivariate and the preprocessing technique comparing PLSR, and SVMR in saline soils. SVMR predictive results were also found moderately accurate for P, K and Zn content, withRPD between 1.7 and 1.8 and R^2 between 0.5 and 0.7). EC, N and pH wwere also poorly predicted with RPD between 1.2 and 1.4 and R^2 between 0.3 and 0.5).

3.4. Model optimization scenario

Predictions within spectra using PLSR and SVMR was estimated (Table 3). Both PLSR and SVMR model output results were compared within soil spectra of each soil properties. Soil OC and P can only be notable with high accuracy for SVMR compared to PLSR (SVMR: R^2 = 0.84; RMSE = 0.12; RPD = 2.47 and RPIQ = 4.50 > PLSR: R^2 = 0.82; RMSE = 0.13; RPD = 2.28; RPIQ = 4.15), followed by P (SVMR- R^2: 0.70, RMSE: 45.38, RPD: 1.84, RPIQ:3.47, PLSR- R^2: 0.71, RMSE: 45.75, RPD: 1.83, RPIQ: 3.44). According to Wenjun et al. (2014) SVMR method, OC can be quantitatively evaluated with better prediction accuracies than PLSR model. From cross-validation result of pH content, the best results were obtained using PLSR based on SD-SG (R^2 = 0.40, RMSE = 0.36, RPD = 1.30 and RPIQ = 1.08; Table 2 and Fig. 6) compared to SVMR. Soil properties portentous acidic (m, H, Al, and H + Al) well performed, being predicted from the VIs-NIR spectra, with RPD value ranging between 1.0 and 1.4 (Terra et al., 2015). The optimal PLSR model prediction accuracy was estimated for N and K value in terms of R^2 and RPD, found better than SVMR model (R^2 = 0.49; RPD = 1.40; RPIQ = 2.05; and R^2 = 0.70; RPD = 1.82, RPIQ = 3.05, respectively), (Raj et al., 2018). In addition, SVMR showed better performance for modeling EC and Zn. All the cross-validation models had RPD and RPIQ above 1.35 and 1.65 over the PLSR model (Fig. 7). The worst results for sand, silt, and clay content in terms of RPD for both the PLSR and SVMR, were in the range of 1.0 and 1.15, respectively (Tables 2 and 3).

3.5. Correlation coefficients of wavelengths

PLSR model successfully predicted the major soil properties from each VIs-NIR spectral range with the *b* coefficient regression value (Table 4). The *b* coefficient summarized the co-relation between all predictors and validated the importance of the X-variables (Lazaar et al., 2020). Reyna et al. (2017) illustrated the importance of the regression coefficient between individual spectral curves and organic carbon. The regression coefficient was validated by the PLSR analysis for selected soil properties SF.2. Particularly, the regression coefficient curve found significant absorption at wavelengths 1432, 1924, and 2245 nm. The spectra value at 1432 nm and 1924 nm wavelength were related to the strong influence of soil moisture (Bo et al., 2010), and the spectra value at 2245 nm wavelength due to the lattice OH in clay minerals, fundamental frequency vibration of O–H, N–H, and C–H, and parameters such as frequency-doubling vibration absorption (Nocita et al., 2014).

For the SG SNV pretreatment spectral setup, maximum peaks used to predict OC, were classified under the visible range with the wavelengths of 427, 436, and 518 nm and classified under NIR range with the wavelengths of 950, 1003, 12340, 1390, 1730, 1878, 2185, 2335 and 2430 nm. The highest correlation in terms of *b* coefficients was identified for the spectral band of 950 and 2450 nm for organic carbon content under NIR range.

The sensitive bands were further identified for P at spectral wavelengths of 430, 505, 780, 1004, 1306, 1387,1490,2162, 2299 and 2433 nm; for K at wavelengths 464, 795, 1004, 1292, 1452, 1940, 1996, 2135 and 2312 nm; and for Zn at wavelengths 580, 775, 1208, 1334,1452, 1930, 1992, 2113 and 2260 nm. The bands at 457, 511,596, 698, 967, 1036, 1108, 1888, 1959, 2165, 2221, 2299 and 2443 nm showed sound relationship with available nitrogen values, whereas bands at 493, 569, 665, 818, 1366, 1840,1953, 2134, 2246, 2407 and 2434 nm for electrical conductivity levels. The soil pH sensitive bands were identified at spectral wavelengths of 492, 567, 652, 818, 984, 1368, 1963, 2138, and 2393 nm in 2nd derivative pretreatment. The highest correlation *b* coefficients were observed for sand, silt, and clay content at 457, 2180, and 515 nm, respectively. Many researchers reported the selected wavebands in the visible NIR region (409, 430, 431, 443, 444, 591, 592 and

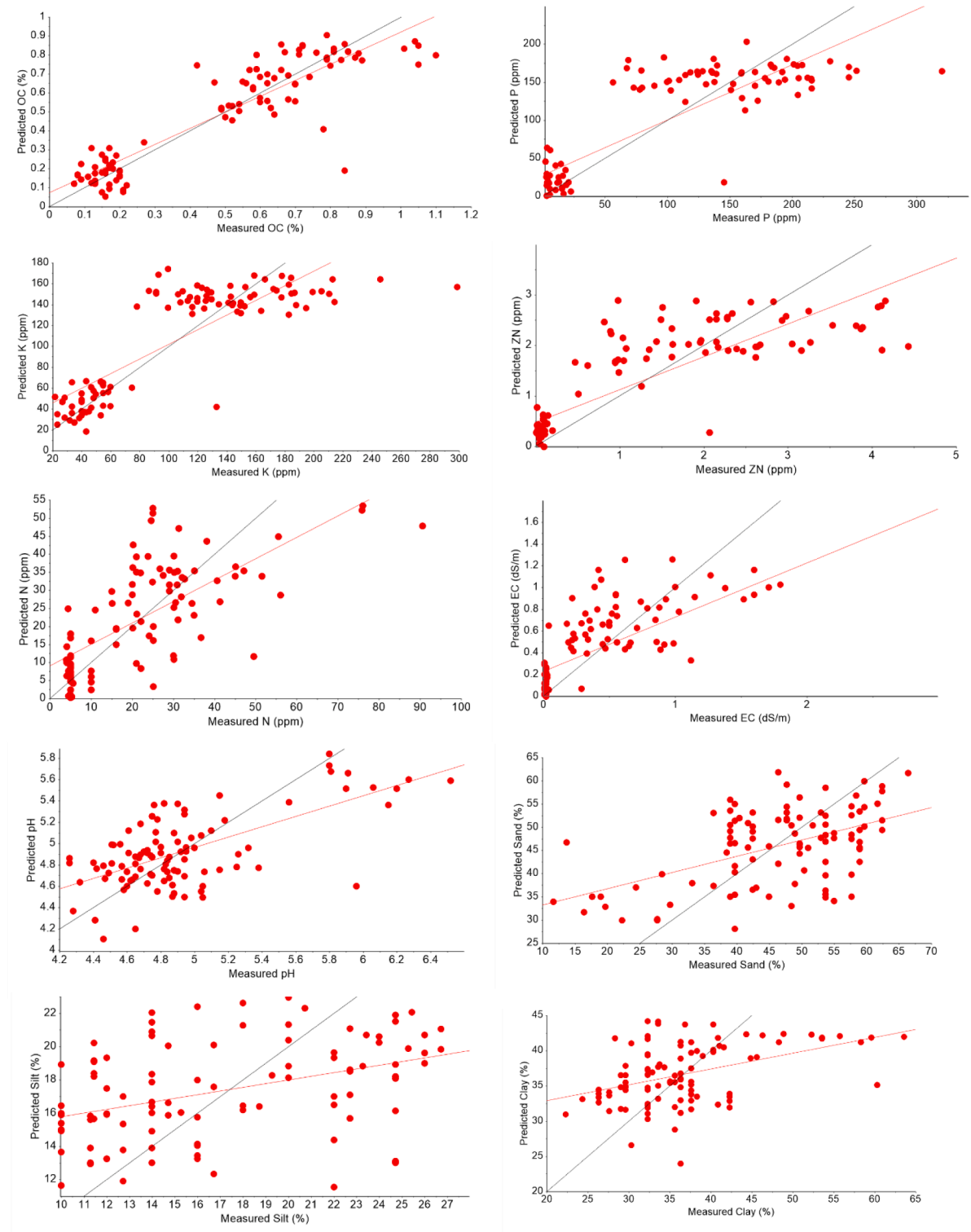

Fig. 6. PLSR validation results for measured vs. predicted values for soil physio-chemical properties. Note: EC-electrical conductivity; pH- pH level, OC-organic carbon; N-available nitrogen; P-available phosphorus; K-available potassium; Zn-zinc level; The red line indicates the lines of regression and the black lines indicate the 1:1 line.

840 and 830 nm) being associated with deep-colored organic matter in the soil, chromophores constituents iron oxides in goethite and hematite minerals (Mortimore et al., 2004).

3.6. Soil nutrients mapping based on semivariogram analysis

The selected soil characteristics' spectral impressions from sentinel 2 imagery were then used to retrieve the predicted values of the nutrients. The prediction models were then used in kriging statistical analyses of other known soil nutrients parameters, such as soil pH, EC, OC, N, P, K,

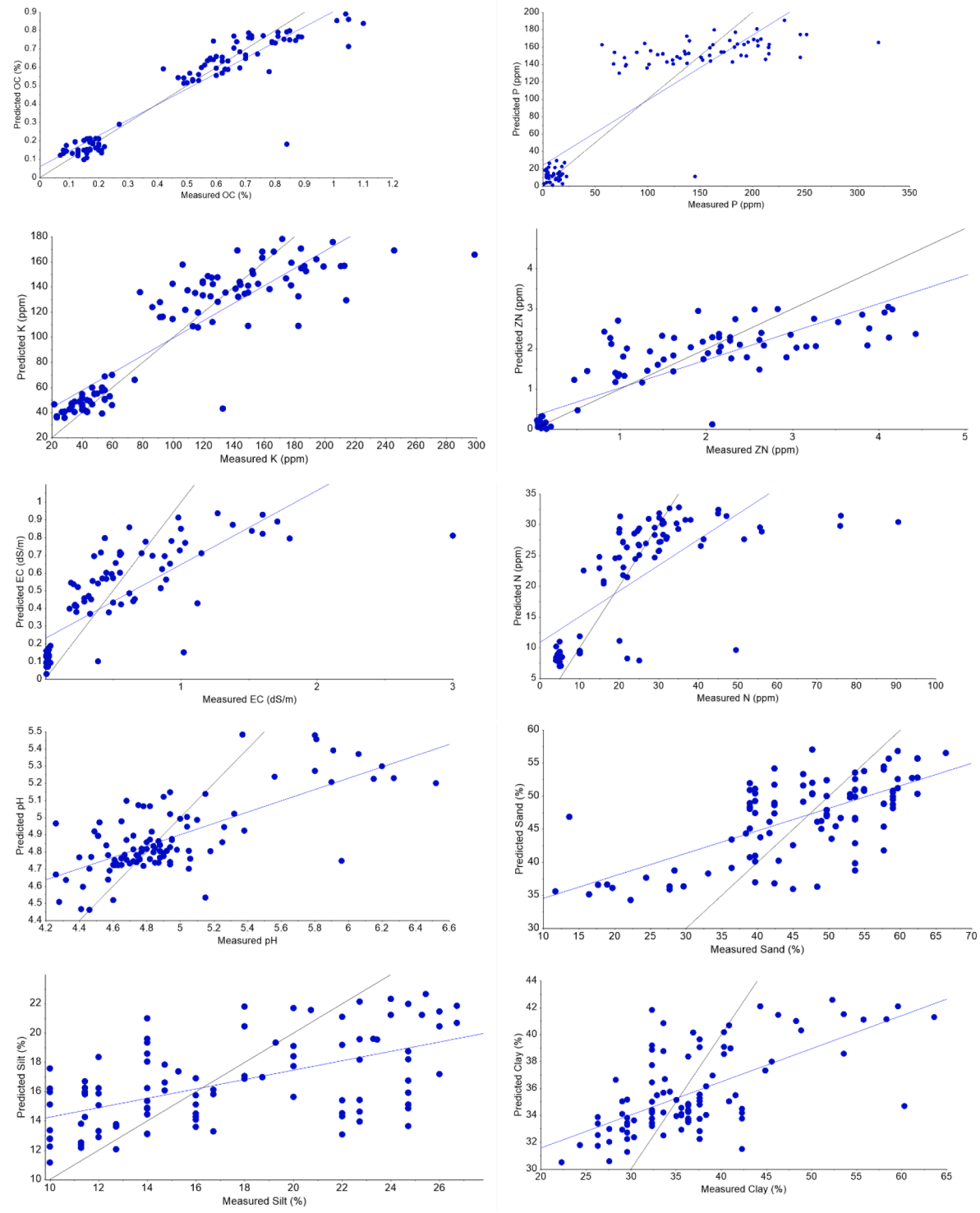

Fig. 7. SVMR validation results for measured vs. predicted values for soil physio-chemical properties *Note:* EC-electrical conductivity; pH- pH level, OC-organic carbon; N-available nitrogen; P-available phosphorus; K-available potassium; Zn-zinc level; The blue lines indicate the line of regression and the black lines indicate the 1:1 line.

Zn, sand, silt and clay (SF.3)The efficiency and performance of the geostatistics semivariogram model in the analysis of the different soil nutrients parameters were then checked by using different matrices namely: RMSSE, MSE, and RMSE. The results of the semivariogram analysis revealed that the Gaussian models were best fitted for available N and available P parameters, where the exponential model best fitted with the soil pH, OC, EC, K, Zn, sand, silt and clay. The different best-fitted tuning parameters were selected for the semivariance geostatistical analysis (Table 5).

3.7. Suitability area estimation

Major seven soil nutrient parameters, such as Soil pH, OC, EC, Clay content, available N, available P and available K were chosen for soil suitability analysis through the outcomes of the SVMR and PLSR model (Figs. 8 & 9). The presence of sand and silt is in exponential relationship

Table 4
Coefficient and factor values for PLSR model.

Soil properties	Pre-treatment	PLSR factor	*b* coefficient
pH	SD-SG	7	0.56
EC(dS/m)	SD-SG	6	7.75
OC (%)	FD-SG	5	2.08
N(ppm)	FD-SG	7	59.42
P(ppm)	FD-SG	3	198.60
K(ppm)	FD-SG	1	99.32
Zn (ppm)	FD-SG	2	1.71
Sand (%)	FD-SG	5	91.54
Silt (%)	FD-SG	3	31.79
Clay (%)	FD-SG	3	32.69

Note: EC-electrical conductivity; OC-organic carbon; N-available nitrogen; P-available phosphorus; K-available potassium; Zn-zinc level; FD-SG- first-derivative spectra with Savitzky–Golay smoothing; SD-SG -first-derivative spectra with Savitzky–Golay smoothing.

with the clay content. Based on laboratory soil nutrient levels, suitability distribution map was developed for the study area (10a). The area of distribution under highly suitable, suitable, moderately suitable and unsuitable classes were 556195 m^2 (19.1 %), 577372 m^2 (19.8 %), 1262091 m^2 (43.5 %)and 511878 m^2(17.6 %), respectively (Fig. 10d & 10e). So, nearly 39 % of study area can be classified as suitable (under highly suitable and suitable class) for growing crops.

Soil suitability estimation was also carried out for both the ML models using the seven major soil nutrient parameters. The final suitability map for both PLSR and SVMR models showed similar pattern as that of laboratory based soil nutrient analysis suitability (Fig. 10a to10e). The SVMR models showed nearly 447500 m^2 (15.3 %), 758200 m^2 (26.1 %), 1286300 m^2 (44.2 %) and 418109 m^2 (14.4 %) were classified under highly suitable, suitable, moderate suitable and unsuitable class respectively (10e). Similarly, the PLSR models distributed nearly 420400 m^2 (14.5 %), 2104800 m^2 (72.3), 354400 m^2 (12.2 %) and 28800 m^2 (1 %) area classified as highly suitable, suitable, moderately suitable and unsuitable class, respectively (10d).

3.8. Validation of suitability distribution

The final suitability maps developed through the machine learning models were validated against the soil analysis based suitability map. The area classified as suitable (both highly suitable and suitable class) under soil analysis based suitability map was validated against SVMR and PLSR suitability class through map query options. Map query tool in ArcGIS software environment was used to identify the variation in suitability area for estimated and predicted distribution (Fig. 10e & 10e). The SVMR model found to be somewhat better (88.9 %) in identifying suitable area compared to PLSR model (Table 6). However, both the models were successful in predicting suitability zones in the study area with accuracy level of nearly 90 %. So, machine learning models can be used to predict soil nutrient and can also guide in estimating crop suitability for a study area.

4. Discussion

Soil parameter modelling mostly targets soil moisture and soil organic carbon content etc. through remote sensing. Though the soil moisture model can facilitate irrigation scheduling but misses many important crop growth parameters. But, a comprehensive study to model major soil nutrient had not been carried out due to a lack of technology and high-cost involvement. On-the-go soil nutrient analysis is still at large for farmers with the least accuracy and high-cost involvement. The availability of wide-range spectral wave length spectroscopy can be used to identify soil parameters through spectral signature analysis. Specific spectral wave length can be identified for a particular soil nutrient, which will be used to identify and quantify the parameter. Regression models have been developed based on the pattern of spectral signature and corresponding wavelength for a particular soil nutrient.

The partial least squares regression (PLSR) and support vector machine regression (SVMR) are found to be most intuitive and commonly applied as multivariate techniques for soil spectroscopy analysis. The pre-treatment VIs-NIR spectrum is imperative for developing PLSR and SVMR models to detect soil attributes (Farwa et al., 2020). The soil nutrients were analyzed by laboratory VIs- spectroscopy measurements for indirectly monitoring the crop and soil quality condition (Mohamed et al., 2016). Principal component analysis (PCA) was undertaken to optimize the no. of components. According to Pasquini (2018), the opportunity of SVMR machine learning techniques with large soil spectral libraries (SSLs) has been performing well, probably due to high level of tolerance shown with outliers and spectral noise which in most cases performed better than PLSR technique. The wavelengths at 830, 840, 2430, 234 in VIs and 2357 nm in NIR region, respectively can be associated with Al-OH bend and O-H stretch combinations, which showed diagnostic absorption features during identification of clay mineral (Knadel et al., 2013; Blaschek et al., 2019; Liu et al., 2019). Residual goethite and hematite minerals have a consequence on the organic matter absorption of phosphorus nutrients (Ramaroson et al., 2018). In the spectra, the slight shoulder at 2340 nm wavelength might displayed illite or mixtures of muscovite minerals (Post and Noble, 1993). Viscarra Rossel and McBratney (1998) also identified significant wavelengths for accurate evaluation of clay content as 1600, 1800, 2000, and 2100 nm classified under the NIR range.

The OC nutrient was estimated successfully using the PLSR and SVMR models. Suitable wavelengths were also identified for available N, P, and K soil nutrients. As observed, soil parameters can also be detected using visible spectrum. The soil parameter sensitive to the visible spectrum with wavelengths of 409, 444, 591 and 592 nm for OC, 430 and 505 nm for P, 464 nm for K; 580 nm for Zn, 492,511,596 and 698 nm for N; 493, 569 and 665 nm for EC; 492,567 and 652 nm for pH; 457 nm for sand and 515 nm for clay. Soil spectral library need to be developed based on VIs-NIR spectra for local soil series. Based on soil spectral libraries, soil nutrient levels may be estimated and used in selection of better crop rotation and better crop production decisions.

Table 5
Summary results of the geostatistical analysis based semi-variogram parameters.

Soil properties	Model Type	Nugget C_0	Partial sill (C_1)	Sill ($C_0 + C_1$)	Range (A_0)	Degree of spatial dependence DSD (%)	Spatial dependence (SD)	Estimated error			
								MSE	ASE	RMSE	RMSSE
pH	Exponential	0.064	0.086	0.149	459	42.54	M	−0.01	0.32	0.34	1.05
EC (ds/m)	Exponential	0.043	0.065	0.107	1088.9	39.68	M	-0.001	0.24	0.27	1.12
OC (%)	Exponential	0.006	0.018	0.024	542.41	24.99	S	-0.005	0.11	0.12	1.09
N (ppm)	Gaussian	66.147	16.636	82.783	804.03	79.90	W	-0.002	9.08	10.07	1.10
P (ppm)	Gaussian	4.939	4939.8	4944.739	53.589	0.09	S	−0.10	106.90	79.57	0.92
K (ppm)	Exponential	442.15	770.18	1212.33	578.56	36.47	M	−0.05	30.79	29.35	1.02
Zn (ppm)	Exponential	0.118	0.396	0.514	734.58	22.98	S	0.04	0.48	0.47	0.98
Sand (%)	Exponential	20.793	170.76	191.553	846.62	10.85	S	−0.01	7.65	7.80	1.02
Silt (%)	Exponential	7.405	24.832	32.238	633.26	22.97	S	−0.003	4.23	3.99	0.98
Clay (%)	Exponential	16.537	72.876	89.413	931.48	18.49	S	−0.02	6.07	5.71	1.08

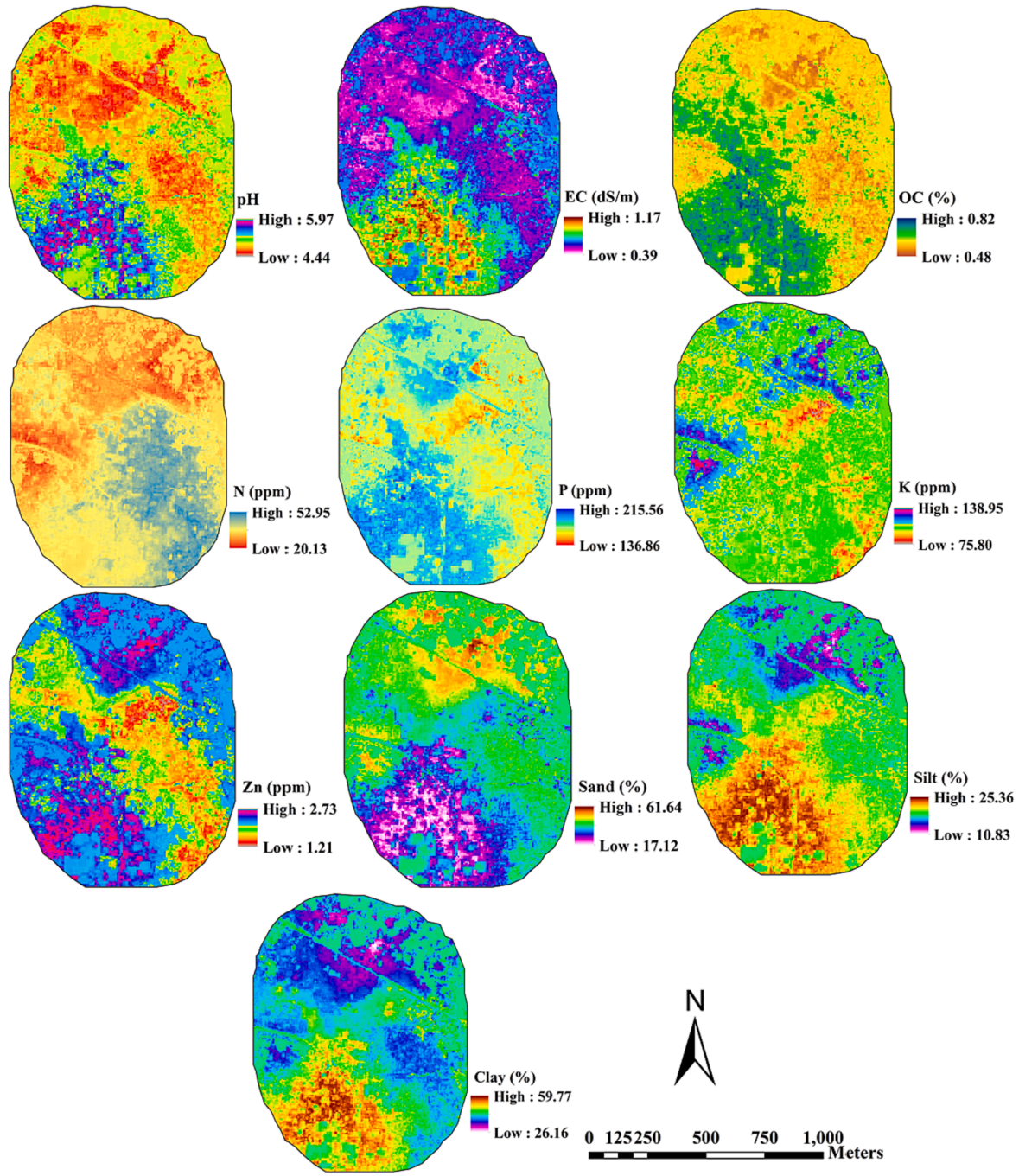

Fig. 8. Spatial variability of predicted soil nutrients by SVMR model (in clockwise, soil pH, electric conductivity (EC), organic carbon (OC), available nitrogen (N), available phosphorus (P), available potassium(K), available zinc (Zn), sand, silt, and clay.

Available soil nutrient-based crop input application will optimize input levels, reduce adverse impact on soil and environment. Somewhat better crop yield and profit will be ensured to the farmers in a sustainable environment (Singha and Swain, 2022; Singha et al., 2022; Suruliandi et al., 2021; Shahare and Gautam, 2022). Nawar et al. (2016) compared PLSR and support vector machines regression (SVMR), to assess organic carbon (OC) and clay content in salt-affected soils, using Savitzky–Golay (SG) smoothing standard normal variate and detrending (SNV-DT), 1st and 2nd derivative. He found different prediction accuracies mostly dependent, both on the multivariate and the preprocessing technique for the study area located in Northern Sinai, Egypt. The prediction results may vary based on different statistical processing procedures and derivatives of spectra (Reeves et al., 2002).

Machine learning models has been successfully utilized in various sectors including medical science, logistic and supply chain, and various industries (Tirkolaee et al., 2021). Even, there is huge scope for machine learning application in agricultural sectors (Singha et al., 2023). These models can help in predicting soil nutrient levels without going for frequent soil sampling and laboratory analysis. Sentinel 2 based 10 m resolution optimal images can be very beneficial in mapping and predicting distribution of soil nutrients through reflectance level of different spectral bands. The 10 m resolution images can support in taking effective production decision. Combining remote sensing images with machine learning models can effectively reduce the cost of direct soil analysis (Mohamed,et al., 2020) and support in taking better production decisions.

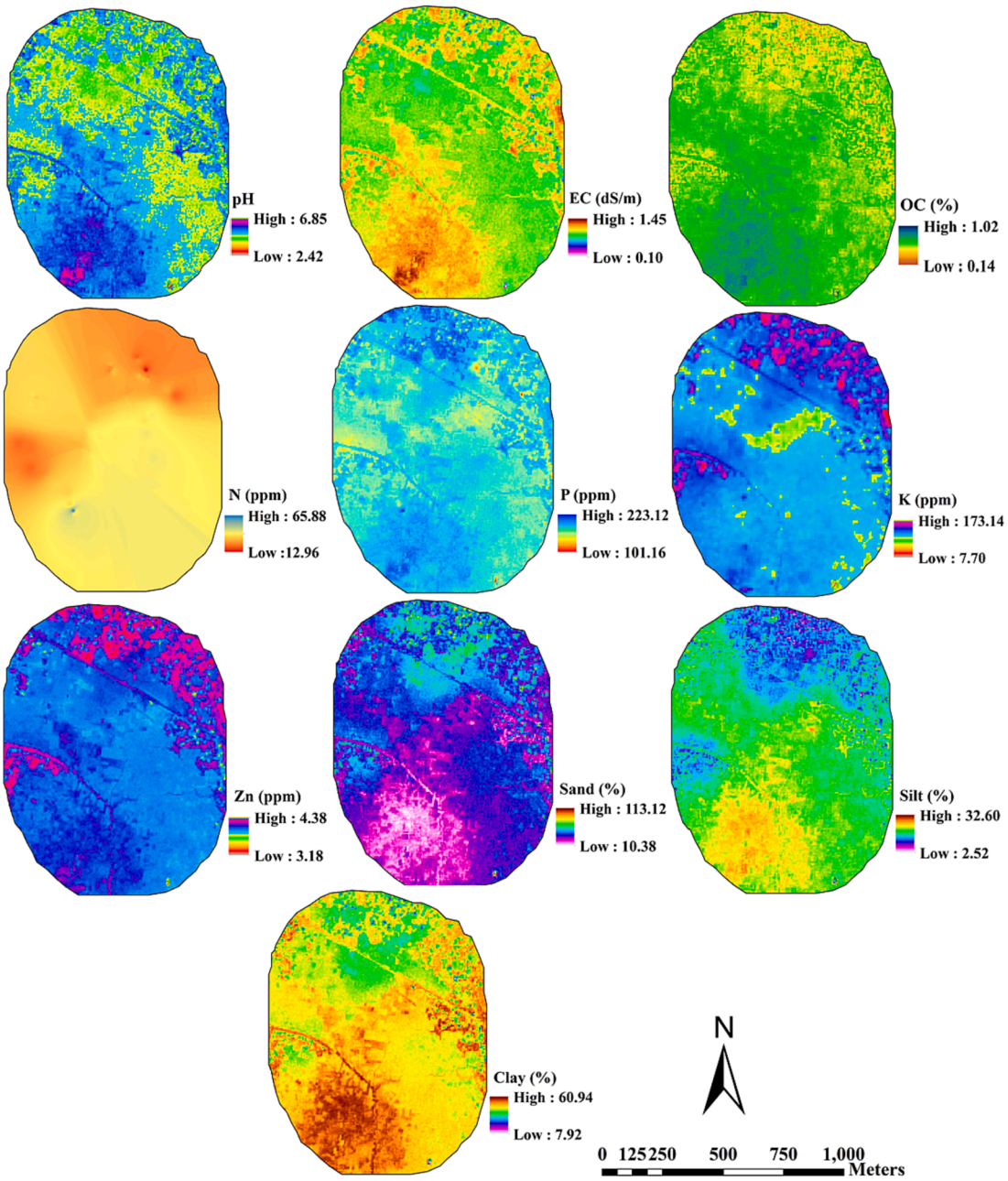

Fig. 9. Spatial variability of predicted soil nutrients by PLSR model (clockwise soil pH, electric conductivity (EC), organic carbon (OC), available nitrogen (N), available phosphorus (P), available potassium(K), available zinc (Zn), sand, silt, and clay.

Seven major crop monitoring soil nutrients namely, OC, pH, EC, available N, available P, available K and clay content were selected to predict the soil nutrient distribution in the study area through PLSR and SVMR machine learning models using eight spectral band data from sentinel-2 imagery. The soil suitability analysis was carried out using the predicted soil nutrient level as well as nutrient level analysed in laboratory in ArcGIS environment.

For wider applicability of machine learning models, the soil suitability maps developed through PLSR and SVMR model were compared with traditional suitability map developed using nutrient levels measured through direct technique. The SVMR model performed slightly better to PLSR model in identifying suitable area in terms of soil nutrients. Similar positive result was also observed by Sarkar et al. (2020) through SVMR model against PLSR model in predicting soluble solid content for food materials.

The accuracy of the distribution maps of the soil traits exhibited by the Ordinary Kriging (OK) validation was authenticated by the results of the semivarigram analysis. The values of the RMSSE for the different types of soil properties were as follows: pH (1.05), EC (1.12), OC (1.09), N (1.10), P (0.92), K (1.02), Zn (0.98), Sand (1.02), silt (0.98) and Clay (1.08), in-line with the outcome validated with Aldabaa and Yousif (2020). The integration of the data from the Sentinel 2 MSI satellite with the in-situ spectral measurements provided a good insight into the spatial variability of soil nutrient distribution in Tarakeswar region, West Bengal, India. The findings of this study demonstrated the utility of satellite images in predicting different properties of soil (Mohamed et al., 2020). The R^2 and RMSE confirmed the expected results of retrieved soil nutrients through satellite images. The validation through R^2 and the RMSSE revealed the expected values of the retrieved soil nutrients by spaceborne remote sensing approach where the R^2 for OC and available N were around >0.80 and >0.7 by PLSR and SVMR model, respectively.

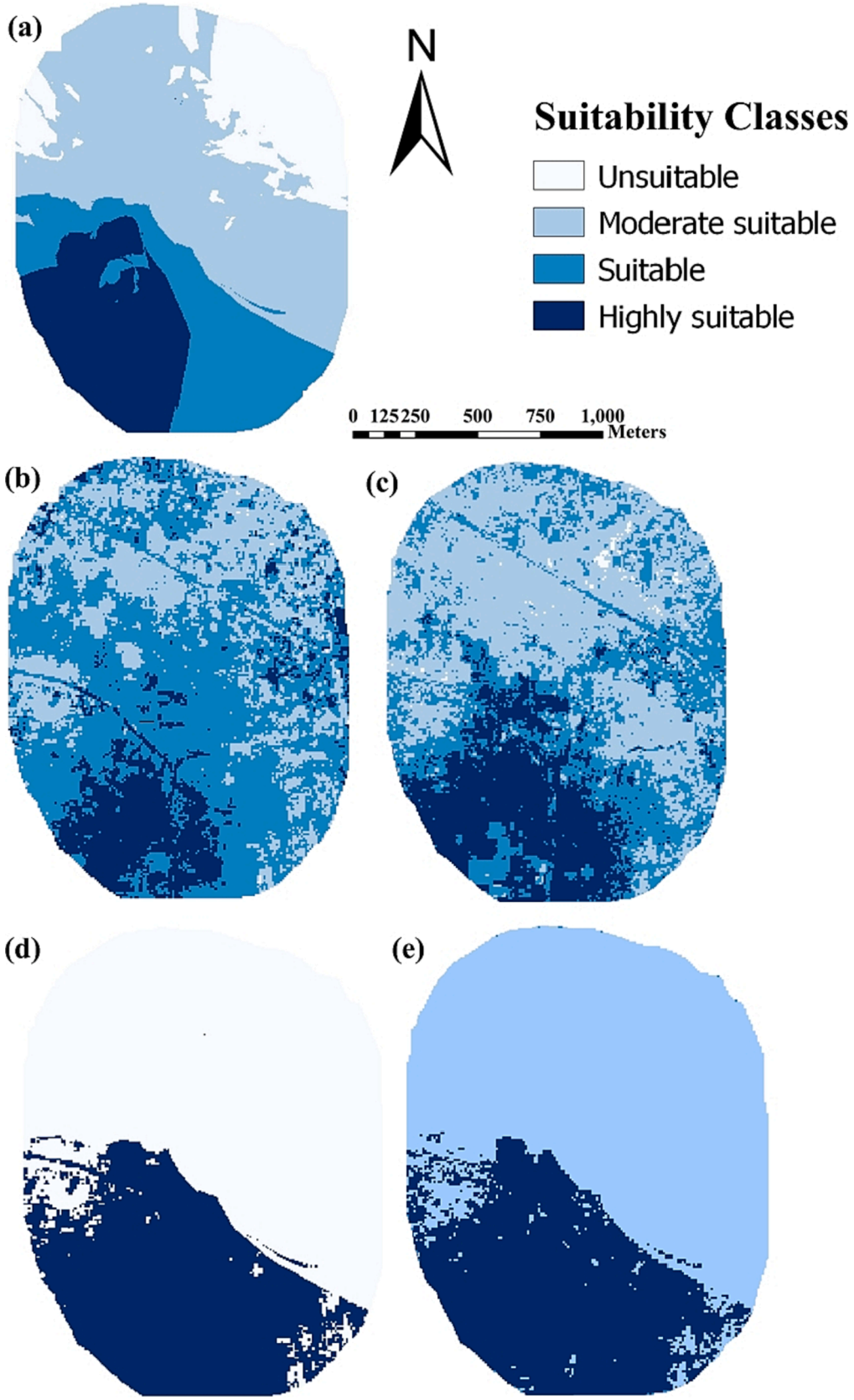

Fig. 10. Comparison of machine learning models suitability distribution with laboratory based analysis (a) Laboratory analysis based soil suitability map, (b), PLSR prediction based soil sutiability map (c) SVMR prediction based soil suitability map, (d) PLSR model based suitability area comparison, (e) SVMR model suitability area comparison.

Table 6
Validation of machine learning model for soil suitability analysis.

Sl. No.	Analysis technique	Suitable area (m^2)	Total area (m^2)	% Accuracy
1	Laboratory analysis based suitability	1,133,567	2,903,536	
2	PLSR	989,200	2,903,536	87.2 %
3	SVMR	1,007,885	2,903,536	88.9 %

Through these machine learning technique, multi years soil nutrient database may be developed for the region to be used in future research and field work. Only upgrading of the soil nutrient level may be carried out time to time.

However, wider applicability of PLSR and SVMR-based soil attribute analysis may require rigorous analysis and extension to various regions for multiple crop seasons. The impact of model values may also be analyzed for mixed cropping, intercropping as well as integrated farming systems (IFS) approaches. The identified spectral bands may be used to develop equipment such as green seeker for quick identification of soil nutrient levels and patterns in the soil.

Limitation

Soil analysis through remote sensing faces issues such as weed cover, variation in moisture level etc. which may affect the final suitability map. Microwave spectrum-based Sentinel 1 and RADARSAT images may perform better under overcast conditions. More machine -learning models may be tested to select best model for soil nutrient estimation in the region. The soil moisture content, particle size, root mix soil is affecting the acquisition the VIs-NIR signature quality as well as the model accuracy.

Advantages and recommendation

VIs-NIR spectroscopy is an inexpensive, non-destructive and time-efficient technique for analyzing the characteristics of soil. It could be an alternative to traditional methods for agricultural monitoring. The development of portable NIR spectrometers has made it easier to carry out the analysis of soil health status. This has also reduced the costs of transportation and handling. The integration of the VIs-NIR spectroscopy and satellite images can help the farmers to apply proper crop inputs such as fertilizer at recommended rate in the field. The farmers easily identified the deficiency of the soil nutrients for better agricultural management. The researchers may utilized VIs-NIR spectroscopy to quickly identify the quality of organic fertilizers. The application of nutrients and other soil amendments in agroecosystems requires precise management. With the help of spectroscopy, it is now possible to assess the chemical and physical attributes of soil in just a couple of minutes. It will have a significant impact on how farmers manage their soil, as it allows them to gain quick access to vital information. VIs-NIR spectroscopy approach will increase the efficiency of precision agriculture operation, digital soil mapping and site-specific nutrient managing. Soil spectroscopic technique for analyzing soil offers numerous advantages over their conventional counterparts. Some of these include: their faster and more accurate technique, agriculture transformation, low cost, minimal environmental impact, portability, non-destructive approach, and availability of large databases. Soil analysis without the use of utensils or chemicals is also feasible by developing specific portable spectrometer.

5. Conclusions

Soil spectral analysis has a great potential for quick estimation of soil nutrients to help farmers in choosing the most suitable or alternative cropping practices.. The pre-treatment of VIs-NIR spectrum is necessary for forming PLSR and SVMR models to identify soil attributes by laboratory VIs- NIR spectroscopy measurements. The PLSR measurement provided the best calibration outcomes (R^2:0.82, 1st derivative) for the Zn content, compared to the validation outcomes for the (R^2: 0.81, 2nd derivative) OC content. Soil OC and P were estimated with high accuracy through SVMR model with $R^2 = 0.84$; RMSE = 0.12; RPD = 2.47 and RPIQ = 4.50. The specific spectral bands in the range of VIs-NIR spectra were identified for the ten soil nutrients, which may be used to develop portable nutrient specific spectrometer.

The PLSR and SVMR machine learning models successfully predicted the ten soil nutrient levels with inputs from sentinel 2 imagery and PCA for the study area. The suitability zone mapping of SVMR found better suited (88.9 %) compared to PLSR as compared against the direct soil nutrient analysis based suitability zoning.

This effective technique has sound applicability in precision agriculture, digital soil mapping, soil security, soil water management, fertility management, soil survey planning, and climate change monitoring.Future studies should evaluate the feasibility of spatial variability mapping of soil available nutrient combining DRS-based model predictions through variable indicators in deep learning, AI-based geostatistical interpolation technologies IoT, Radar-based, UAV, and airborne spectroscopy on a local to global scale.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We acknowledge the project “Integration of Digital Augmentation for sustainable Agroecosystem in Western Lateritic Zone under National Hydrology Project, West Bengal” under which this work is mapped. The author also convey special thanks the International Centre for Agricultural Research in the Dry Areas (ICARDA) for supporting necessary logistics for this research work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejrs.2023.10.005.

References

Abdi, D., Tremblay, G.F., Ziadi, N., Bélanger, G., Parent, L.É., 2012. Predicting soil phosphorus-related properties using reflectance spectroscopy. Soil Sci. Soc Am. J. 76, 2318–2326.

Aldabaa, A., Yousif, I.A.H., 2020. Geostatistical approach for land suitability assessment of some desert soils. Egypt. J. Soil Sci. 60, 195–205.

Baroudy, A.A.E., Ali A.M. Mohamed, E.S., Moghanm, F.S., Shokr, M.S., Savin, I., Poddubsky, A., Ding, Z., Kheir, A.M.S., Aldosari, A.A., et al., 2020. Modeling Land Suitability for Rice Crop Using Remote Sensing and Soil Quality Indicators: The Case Study of the Nile Delta. Sustainability 12, 9653. https://doi.org/10.3390/su12229653.

Barthès, B.G., Kouakoua, E., Clairotte, M., Lallemand, J., Chapuis-Lardy, L., Rabenarivo, M., Roussel, S., 2019. Performance comparison between a miniaturized and a conventional near infrared reflectance (NIR) spectrometer for characterizing soil carbon and nitrogen. Geoderma 338, 422–429.

Ben-Dor, E., Chabrillat, S., Dematte, J.A.M., Taylor, G.R., Hill, J., Whiting, M.L., 2009. Using imaging spectroscopy to study soil properties. Remote Sens. Environ. 13 (1), S38–S55.

Blaschek, M., Roudier, P., Poggio, M., Hedley, C.B., 2019. Prediction of soil available water-holding capacity fromvisible near-infrared reflectance spectra. Sci Rep. 9, 12833.

Bo, S., Rossel, R.A.V., Mouazen, A.M., Wetterlind, J., 2010. Chapter five –visible and near infrared spectroscopy in soil science. Adv Agron. 107 (107), 163–215.

Chakraborty, S., Weindorf, D.C., Deb, S., Li, B., Paul, S., Choudhury, A., Ray, D.P., 2017. Rapid assessment of regional soil arsenic pollution risk via diffuse reflectance spectroscopy. Geoderma 289, 72–81. https://doi.org/10.1016/j.geoderma.2016.11.024.

Clingensmith, C.M., Grunwald, S., Wani, S.P., 2019. Evaluation of calibration subsetting and new chemometricmethods on the spectral prediction of key soil properties in a data-limited environment. Eur. J. Soil Sci. 70, 107–126.

Cohen, M.J., Prenger, J.P., DeBusk, W.F., 2005. Visible-near infrared reflectance spectroscopy for rapid, nondestructive assessment of wetland soil quality. J. Environ Qual. 34, 1422–1434.

Conforti, M., Matteucci, G., Buttafuoco, G., 2018. Using laboratory Vis-NIR spectroscopy for monitoring someforest soil properties. J. Soils Sediments 18, 1009–1019.
Di Iorio, E., Circelli, L., Lorenzetti, R., Costantini, E.A.C., Egendorf, S.P., Colombo, C., 2019. Estimation of andic properties from Vis-NIR diffuse reflectance spectroscopy for volcanic soil classification. Catena 182, 104109. https://doi.org/10.1016/j.catena.2019.104109.
El-Sayed, M.A., Abd-Elazem, A.H., Moursy, A.R.A., Mohamed, E.S., Kucher, D.E., Fadl, M.E., 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, 935. https://doi.org/10.3390/agronomy13030935.
Farwa, U.E., Rehman, A.U., Khan, S.Q., Khurram, M., 2020. Prediction of Soil Macronutrients Using Machine Learning Algorithm. Retrieved from In J Compu. (IJC) 38 (1), 1–14. https://ijcjournal.org/index.php/InternationalJournalOfComputer/article/view/1527.
Filzmoser P, Varmuza K (2017) Chemometrics: Multivariate Statistical Analysis in Chemometrics. R package version 1.4.2. https://CRAN.R-project.org/package=chemometrics.
Fu, C., Gan, S., Yuan, X., Xiong, H., Tian, A., 2018. Determination of soil salt content using a probability neural network model based on particles warm optimization in areas affected and non-affected by human activities. Remote Sens. 10, 1387.
Geladi, P., 2003. Chemometrics in spectroscopy. Part 1. Classical chemometrics. Spectrochim Acta Part B at Spectrosc. 58, 767–782.
Geladi P, MacDougall D, Martens H (1985) Linearization and scatter-correction for near-infrared reflectance spectra of meat. Appl. Spectrosc. 39: 491–500. https://doi. org/10.1366/0003702854248656.
Grunwald, S., Vasques, G.M., 2015. Fusion of soil and remote sensing data to model soil properties. In Advances in Agronomy; Sparks, D.L., Ed.; Advances in Agronomy: Amsterdam. The Netherland 131, 1–109.
Gupta, A., Vasava, H.B., Das, B.S., Choubey, A.K., 2018. Local modeling approaches for estimating soil properties in selected Indian soils using diffuse reflectance data over visible to near-infrared region. Geoderma 325, 59–71.
Knadel, M., Viscarra Rossel, R.A., Deng, F., Thomsen, A., Greve, M.H., 2013. Visible–near infrared spectra as aproxy for topsoil texture and Glacial boundaries. Soil Sci. Soc Am J. 77, 568.
Kuang, B., Mahmood, H.S., Quraishi, M.Z., Hoogmoed, W.B., Mouazen, A.M., van Henten, E.J., 2012. Sensing Soil Properties in the Laboratory, In Situ, and On-Line. Advances in Agron. 155–223 https://doi.org/10.1016/b978-0-12-394275-3.00003-1.
Kuang, B., Mouazen, A.M., 2011. Calibration of visible and near infrared spectroscopy for soil analysis at the fieldscale on three European farms. Eur J. Soil Biol. 62, 629–636.
Lark, R.M., 2000. Estimating variograms of soil properties by the method-of-moments and maximum likelihood. Eur. J. Soil Sci. 51, 717–728.
Lazaar, A., Mouazen, A.M., Hammouti, K.E.L., Fullen, M., Pradhan, B., Memon, M.S., Monir, A., 2020. The application of proximal visible and near-infrared spectroscopy to estimate soil organic matter on the Triffa Plain of Morocco. Int. Soil Wat. Conser. Res. https://doi.org/10.1016/j.iswcr.2020.04.005.
Liu Y, Boss E, Chase AP, Xi H, Zhang X, Röttgers R, Pan Y, Bracher A (2019) Spectral particulateabsorption coefficients and their standard deviation derived from underway AC-S measurements during POLARSTERN cruise PS99.2. PANGAEA.,doi: 10.1594/PANGAEA.898121.
Liu, H., Kawamura, K., Tsujimoto, Y., Nishigaki, T., Razakamanarivo, H., Andrianary, B. H., Andriamananjara, A., 2020b. Prediction of Soil Oxalate Phosphorus using Visible and Near-Infrared Spectroscopy in Natural and Cultivated System Soils of Madagascar. Agriculture 10 (5), 177. https://doi.org/10.3390/agriculture10050177.
Liu, J., Xie, J., Han, J., Wang, H., Sun, J., Li, R., Li, S., 2020a. Visible and near-infrared spectroscopy with chemometrics are able to predict soil physical and chemical properties. J. Soils Sedi. https://doi.org/10.1007/s11368-020-02623-1.
Liu, H.J., Zhang, Y.Z., Zhang, X.L., Zhang, B., Song, K.S., Wang, Z.M., Tang, N., 2009. Quantitative Analysis of Moisture Effect on Black Soil Reflectance. Pedosphere 19 (4), 532–540. https://doi.org/10.1016/s1002-0160(09)60146-6.
Ludwig, B., Murugan, R., Parama, V.R.R., Vohland, M., 2019. Accuracy of Estimating Soil Properties with Mid-Infrared Spectroscopy: Implications of Different Chemometric Approaches and Software Packages Related to Calibration Sample Size. Soil Sci. Soc Am J. https://doi.org/10.2136/sssaj2018.11.0413.
Miloš, B., Bensa, A., 2018. Estimation of SOC Content in Anthropogenic Soils from Flysch Deposits Using Vis-NIR Spectroscopy. Poljoprivreda 24 (1), 45–51. https://doi.org/10.18047/poljo.24.1.6.
Mohamed, E.S., Ali, A.M., El Shirbeny, M.A., Abd El Razek, A.A., Savin, I.Y., 2016. Near infrared spectroscopy techniques for soil contamination assessment in the Nile Delta. Eura Soil Sci. 49, 632–639.
Mohamed, E.S., Baroudy, A.A.E., El-beshbeshy, T., Emam, M., Belal, A.A., Elfadaly, A., Aldosari, A.A., Ali, A.M., Lasaponara, R., 2020. Vis-NIR Spectroscopy and Satellite Landsat-8 OLI Data to Map Soil Nutrients in Arid Conditions: A Case Study of the Northwest Coast of Egypt. Remote Sens. 12, 3716. https://doi.org/10.3390/rs12223716.
Morgan, R.P.C., 2009. Soil erosion and conservation. John Wiley & Sons.
Mortimore, J.L., Marshall, L.J.R., Almond, M.J., Hollins, P., Matthews, W., 2004. Analysis of red and yellow ochresamples from Clearwell Caves and Catalhoyuk by vibrational spectroscopy and other techniques. Spectrochim. Acta A. Mol Biomol. Spectrosc. 60, 1179–1188.
Mouazen, A.M., Karoui, R., De Baerdemaeker, J., Ramon, H., 2005. Classification of soil texture classes by usingsoil visual near infrared spectroscopy and factorial discriminant analysis techniques. J near Infra. Spectrosc. 13, 231–240.
Nawar, S., Buddenbaum, H., Hill, J., Kozak, J., Mouazen, A.M., 2016. Estimating the soil clay content and organic matter by means of different calibration methods of vis-NIR diffuse reflectance spectroscopy. Soil Til Res. 155, 510–522.
Nocita, M., Stevens, A., Toth, G., Panagos, P., van Wesemael, B., Montanarella, L., 2014. Prediction of soil organic carbon content by diffuse reflectance spectroscopy using a local partial least square regression approach. Soil Biol. Biochem. 68, 337e347. https://doi.org/10.1016/j.soilbio.2013.10.022.
Olsen SR, Cole CV, Watanabe FS, Dean LA (1954) Estimation of Available Phosphorus in Soils by Extraction with Sodium Bicarbonate. USDA Circular 939. Washington D.C.
Pasquini, C., 2018. Near infrared spectroscopy: A mature analytical technique with new perspectives—A review. Anal. Chim Acta. 1026, 8–36.
Piper, C.S., 1942. Soils and Plant analysis. Yhe University of Adelaide, Adelaide.
Post, J.L., Noble, P.N., 1993. The near-infrared combination band frequencies of dioctahedral smectites, Micas, and Illites. Clays Clay Min. 41, 639–644. 12 S.
Raj, A., Chakraborty, S., Duda, B.M., Weindorf, D.C., Li, B., Roy, S., Paulette, L., 2018. Soil mapping via diffuse reflectance spectroscopy based on variable indicators: An ordered predictor selection approach. Geoderma 314, 146–159. https://doi.org/10.1016/j.geoderma.2017.10.043.
Ramaroson, V.H., Becquer, T., Sá, S.O., Razafimahatratra, H., Delarivière, J.L., Blavet, D., Vendrame, P.R.S., Rabeharisoa, L., Rakotondrazafy, A.F.M., 2018. Mineralogical analysis of ferralitic soils in Madagascar using NIR spectroscopy. Catena 168, 102–109.
Reeves, J.B., 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? Geoderma 158, 3–14.
Reeves, J., McCarty, G., Mimmo, T., 2002. The potential of diffuse reflectance spectroscopy for the determination of carbon inventories in soils. Environ. Pollut. 116, S277–S284.
Reyna, L., Dube, F., Barrera, J.A., Zagal, E., 2017. Potential Model Overfitting in Predicting Soil Carbon Content by Visible and Near-Infrared Spectroscopy. Applied Sci. 7 (7), 708. https://doi.org/10.3390/app7070708.
Santra, P., Singh, R., Sarathjith, M.C., Panwar, N.R., Varghese, P., Das, B.S., 2015. Reflectance spectroscopic approach for estimation of soil properties in hot arid western Rajasthan, India. Environ. Earth Sci. 74 (5), 4233–4245. https://doi.org/10.1007/s12665-015-4383-x.
Sarkar, S., Basak, J.K., Moon, B.E., Kim, H.T.A., 2020. Comparative Study of PLSR and SVM-R with Various Preprocessing Techniques for the Quantitative Determination of Soluble Solids Content of Hardy Kiwi Fruit by a Portable Vis/NIR Spectrometer. Foods 9, 1078. https://doi.org/10.3390/foods9081078.
Savitzky, A., Golay, M.J.E., 1964. Smoothing and differentiation of data by simplified least squares procedures. Anal Chem. 36, 1627–1639. https://doi.org/10.1021/ac60214a047.
Schoell, A., Zou, Y., Huebner, D., Urquhart, S.G., Schmidt, T., Fink, R., et al., 2005. A comparison of fine structures in high-resolution x-rayabsorption spectra of various condensed organic molecules. J. Chem Phys. 123 (4), 45.
Shahare, Y., Gautam, V., 2022. Soil Nutrient Assessment and Crop Estimation with Machine Learning Method: A Survey. In: Tavares, J.M.R.S., Dutta, P., Dutta, S., Samanta, D. (Eds.), Cyber Intelligence and Information Retrieval. Lecture Notes in Networks and Systems, vol 291. Springer, Singapore. https://doi.org/10.1007/978-981-16-4284-5_22.
Shao, Y., He, Y., 2011. Nitrogen, phosphorus, and potassium prediction in soils, using infrared spectroscopy. Soil Res. 49, 166–172.
Shrestha, R.P., 2006. Relating soil electrical conductivity to remote sensing and other soil properties for assessing soil salinity in northeast Thailand. Land Degrad Dev. 17 (6), 677–689.
Singha C, Swain KC (2022) Rice and Potato Yield Prediction Using Artificial Intelligence Techniques. In: Pattnaik, P.K., Kumar, R., Pal, S. (eds) Internet of Things and Analytics for Agriculture, Volume 3. Studies in Big Data, vol 99. Springer, Singapore. doi: 10.1007/978-981-16-6210-2_9.
Singha, C., Swain, K.C., Swain, S.K., 2020. Best Crop Rotation Selection with GIS-AHP Technique Using Soil Nutrient Variability. Agriculture 10, 213. https://doi.org/10.3390/agriculture10060213.
Singha, C., Gulzar, S., Swain, K.C., Pradhan, D., 2023. Apple yield prediction mapping using machine learning techniques through the Google Earth Engine cloud in Kashmir Valley, India. J. Appl. Remote Sens. 17 (1), 014505 https://doi.org/10.1117/1.JRS.17.014505.
Soriano-Disla, J.M., Janik, L.J., Viscarra Rossel, R.A., Macdonald, L.M., McLaughlin, M. J., 2014. The Performance of Visible, Near-, and Mid-Infrared Reflectance Spectroscopy for Prediction of Soil Physical, Chemical, and Biological Properties. Appl. Spectrosc Rev. 49, 139–186.
Soil Survey Staff (1993) Soil survey manual. Agricultural handbook. No. 18, US Department of Agriculture, Soil Conservation Service, Washington, DC.
Stevens, A., Nocita, M., Toth, G., Montanarella, L., van Wesemael, B., 2013. Prediction of soil organic carbon atthe European scale by visible and near infrared reflectance spectroscopy. PLoS ONE 8, e66409.
Subbaiah, B.V., Asija, G.L., 1965. A rapid procedure for determination of available nitrogen in soil. Current Sci. 25, 259–260.
Suruliandi, A., Mariammal, G., Raja, S.P., 2021. Crop prediction based on soil and environmental characteristics using feature selection techniques. Math. Comput. Model Dyn. Syst. 27 (1), 117–140. https://doi.org/10.1080/13873954.2021.1882505.
Tekin, Y., Kuang, B., Mouazen, A.M., 2013. Potential of On-Line Visible and Near Infrared Spectroscopy for Measurement of pH for Deriving Variable Rate Lime Recommendations. Sensors 13 (8), 10177–10190.

Terra, F.S., Demattê, J.A.M., Viscarra-Rossel, R.A., 2015. Spectral libraries for quantitative analysis of tropical Brazillian soils: Comparing VIS-NIR and MIR reflecytance data. Geoderma 255–256, 81–93.

Tirkolaee, E.B., Sadeghi, S., Mooseloo, F.M., Vandchali, H.R., Aeini, S., 2021. Application of Machine Learning in Supply Chain Management: A Comprehensive Overview of the Main Areas. Article ID 1476043 Math Prob. Eng. 14. https://doi.org/10.1155/2021/1476043.

Udelhoven, T., Emmerling, C., Jarmer, T., 2003. Quantitative analysis of soilchemical properties with diffuse reflectance spectrometry and partial least-squareregression: A feasibility study. Plant Soil. 251, 319–329.

Viscarra Rossel, R.A., Behrens, T., 2010. Using data mining to model and interpret soil diffuse reflectance spectra. Geoderma 158 (1e2), 46e54. https://doi.org/10.1016/j.geoderma.2009.12.025.

Viscarra Rossel, R.A., McBratney, A.B., 1998. Laboratory evaluation of a proximal sensing technique for simultaneous measurement of soil clay and water content. Geoderma 85 (1), 19–39. https://doi.org/10.1016/s0016-7061(98)00023-8.

Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., Skjemstad, J.O., 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma 131 (1–2), 59–75.

Waiser, T.H., Morgan, C.L.S., Brown, D.J., Hallmark, C.T., 2007. In Situ Characterization of Soil Clay Content with Visible Near-Infrared Diffuse Reflectance Spectroscopy. Soil Sci. Soc Ame J. 71 (2), 389.

Wang, S., Chen, Y., Wang, M., Zhao, Y., Li, J., 2019. SPA-Based Methods for the Quantitative Estimation of theSoil Salt Content in Saline-Alkali Land from Field Spectroscopy Data: A Case Study from the Yellow River Irrigation Regions. Remote Sens. 11, 967.

Wang, J.J., Cui, L.J., Gao, W.X., Shi, T.Z., Chen, Y.Y., Gao, Y., 2014. Prediction of low heavy metal concentrations in agricultural soils using visible and near-infrared reflectance spectroscopy. Geoderma 216 (4), 1–9.

Wenjun, J., Zhou, S., Jingyi, H., Shuo, L., 2014. In Situ Measurement of Some Soil Properties in Paddy Soil Using Visible and Near-Infrared Spectroscopy. PLoS ONE 9 (8), e105708.

Wold, S., Sjostrom, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemom. Intell. Lab. Syst. 58, 109–130. https://doi.org/10.1016/S0169-7439 (01):00155-1.

Xuemei, L., Jianshe, L., 2013. Measurement of soil properties using visible and short wave-near infraredspectroscopy and multivariate calibration. Meas. J. Int. Meas Confed. 46, 3808–3814.