



OPEN A PCA-based standardized spectral index from Sentinel-2 for modeling soil macronutrients in the Miandoab region, Iran

Ali Dianati, Mir Hassan Rasouli-Sadaghiani✉, Farrokh Asadzadeh✉, Ebrahim Sepehr & Kamal Khosravi Aqdam

Spatial monitoring of soil macronutrients—especially total nitrogen (TN), available phosphorus (AP), and available potassium (AK)—using remote sensing technologies is a promising approach to promote precision agriculture. The aim of this study is to evaluate the effectiveness of Sentinel-2 satellite data and selected spectral indices in modeling the concentrations of these important soil macronutrients on agricultural land. A total of 181 soil samples were collected from a depth of 0–30 cm and analyzed in the laboratory using standard methods to determine nitrogen, phosphorus, and potassium content. Twelve common spectral indices and a newly proposed index—the standardized spectral reflectance index (SSRI), derived from the first principal component of a PCA—were extracted from the Sentinel-2 data. Linear regression modeling revealed that TN provided the most accurate predictions ($R^2 = 0.77$, RMSE = 0.04%, MSE = 0.01, RPIQ = 2.43 (good predictive performance)), followed by AK ($R^2 = 0.72$, RMSE = 166.49 ppm, MSE = 0.27, RPIQ = 2.21 (good predictive performance)), while AP showed relatively weak model performance (no significant regression and no predictive RPIQ), probably due to its limited spectral expression in remote sensing data. The newly introduced SSRI outperformed the conventional indices in nitrogen modeling, which is an important novelty of this study. Overall, the results indicate that the integration of Sentinel-2 data with optimized spectral indices provides a feasible and effective approach for the indirect estimation of TN and, to a lesser extent, AK. The application of this method has the potential to reduce reliance on costly field sampling, improve fertilizer management, and contribute to the sustainability of agricultural systems. It is recommended that multi-temporal Sentinel-2 imagery be used in future studies to refine SSRI extraction and provide a rational and innovative method for estimating these critical soil macronutrients.

Keywords Sentinel-2, Soil macronutrients, Standardized spectral reflectance index, Total nitrogen

The availability of nutrients in the soil is a key factor influencing agricultural productivity, long-term sustainability, and environmental quality. Macronutrients, including total nitrogen (TN), available phosphorus (AP), and available potassium (AK), are essential elements that directly control fundamental plant physiological processes such as cell division, photosynthesis, nutrient uptake, and energy transfer. These nutrients have a significant impact on biomass accumulation and yield formation by supporting the metabolic functions and structural development of plants^{1,2}. Due to the high requirements of plants, maintaining optimal levels of these macronutrients through continuous monitoring, accurate soil testing, and appropriate fertilization strategies is crucial for maintaining soil fertility, avoiding nutrient depletion, and improving plant performance. In addition, balanced nutrient management also mitigates negative environmental impacts such as nutrient leaching, greenhouse gas emissions, and soil degradation³.

Accurate determination of soil macronutrient content is fundamental to precision agriculture as it enables site-specific fertilization and minimize the overuse of agrochemicals that can affect soil and water quality⁴. Conventional soil testing methods, such as field sampling followed by chemical analysis in laboratory, are widely used but have several limitations, including high cost, high labor intensity, and limited spatial coverage^{5,6}. These limitations are particularly important in large-scale or heterogeneous agricultural regions where nutrient variability is high and timely decision-making is essential^{7,8}.

Department of Soil Science, Faculty of Agriculture, Urmia University, Urmia, Iran. ✉email: m.rsadaghiani@urmia.ac.ir; f.asadzadeh@urmia.ac.ir

With the advancement of Earth observation technologies, remote sensing has become a powerful tool for characterizing the spatial variability of soil properties. Satellite-based multispectral data, especially from the Sentinel-2 platform, has proven to be an effective tool for mapping soil properties due to its high temporal frequency, spectral richness, and spatial resolution (10–20 m)^{2,9–11}. Various soil properties, such as texture, moisture, organic matter, and nutrient content, can influence the reflectance captured in specific spectral bands¹². Therefore, remote sensing offers the opportunity to indirectly estimate soil macronutrients on a broad scale, with lower costs and improved temporal monitoring¹³.

Previous studies have applied various machine learning and statistical techniques to model soil nutrient distribution using satellite imagery^{7,14}. However, most studies focused on individual nutrients or used simple vegetation or soil indices, often neglecting the interdependencies among soil attributes or the full spectral potential of the Sentinel-2 data. In addition, the presence of noise, redundancy, and multicollinearity in raw spectral bands can affect model performance and reduce generalizability⁴. To overcome these challenges, dimensionality reduction techniques such as principal component analysis (PCA) have become increasingly important in soil and environmental modeling⁸. PCA enables the transformation of correlated spectral bands into uncorrelated principal components, preserving the most important spectral variance and reducing noise and redundancy. In combination with appropriate standardization techniques, spectral indices derived from PCA can increase the robustness and interpretability of remote sensing models and improve their predictive performance in complex landscapes¹.

In this study, a standardized spectral index called Principal Component Analysis–Standardized Spectral index (PCA-SSI) was developed based on PCA and standardization to accurately estimate the spatial variability of soil macronutrients—TN, AP, and AK—using Sentinel-2 imagery. The Miandoab plain in northwestern Iran was selected as the study area because it is an important agricultural region characterized by high soil heterogeneity, variable fertility levels, and diverse cropping systems. Challenges such as nutrient deficiency and soil salinization in this region highlight the need for accurate, data-driven methods to monitor macronutrient status, as improved soil management here can significantly increase productivity and enhance food security. Therefore, the main objectives of this research are: (1) extract the main spectral components from Sentinel-2 images using PCA, (2) develop a PCA-SSI that integrates the most informative bands for macronutrient prediction, (3) model and map the spatial distribution of TN, AP, and AK using this index, and (4) evaluate the performance of the proposed index in capturing soil fertility variability across the study area. By combining multivariate spectral transformation, statistical standardization, and high-resolution remote sensing data, this work provides a robust and practical framework for precise soil nutrients that can be used for digital soil mapping and optimized nutrient management programs in semi-arid agricultural systems.

Materials and methods

Geographical, climatic, and soil characteristics of the study region

The study area encompasses approximately 8.3×10^4 km² of agricultural land in the northwest of Iran, located between 45° 50' 00" and 46° 20' 00" east longitude and 37° 10' 00" and 38° 50' 00" north latitude (Fig. 1). The region is characterized by a semi-arid climate according to the Köppen climate classification, with an average annual temperature of around 9 °C. Rainfall patterns are bimodal, with a wet season that lasts from early autumn to mid-spring and a dry season that lasts from mid-spring to early autumn. The average annual rainfall during the wet period is about 289 mm. The lithological composition of the region is diverse and consists of igneous, sedimentary, and metamorphic formations such as diorite, marble, andesitic volcanic rocks, marl, limestone, sandstone, shale, conglomerate, and Precambrian granite to granodiorite. These parent materials have contributed to the development of different soil types, which are mainly classified as Entisols and Inceptisols and show a low to moderate levels of pedogenic development. The main land use is agriculture with the main crops being winter wheat, barley, peas, sugar beets, alfalfa, oilseeds, corn, and various fruit crops. Although furrow irrigation was the predominant irrigation practice in the past, increasing water scarcity in recent years has led to a gradual transition to more water-efficient systems, in particular center pivot irrigation (<https://www.swri.ir>).

Soil sampling and macronutrient analysis

After delimiting and defining the boundaries of the study area, soil sampling was conducted randomly at 181 precise geographical locations distributed in the region (Fig. 1). The main objective of sampling strategy was to obtain a comprehensive representation of the chemical properties of the surface soils in the plain. At each sampling point, composite soil samples were collected from the 0 to 30 cm depth interval. More specifically, five subsamples (Fig. 2) were taken at a specific radius around each point and thoroughly mixed into a single composite sample Faramarzi et al.¹⁵. This depth range was chosen because it encompasses the zone of highest biological activity, nutrient exchange, and root development for cultivated plants. After collection, the samples were transported to the laboratory where they were air-dried at room temperature, gently ground, and passed through a 2 mm mesh sieve to prepare them for chemical analyzes. The concentrations of the key macronutrients were then determined according to established standard protocols. Total nitrogen (TN) was quantified using the classical Kjeldahl digestion method¹⁶. The available phosphorus (AP) was determined using the Olsen procedure¹⁷, and soil available potassium (AK) concentration was measured using a flame photometric technique¹⁸. These methods were selected due to their proven suitability for the analysis of calcareous soils, which are prevalent in the study region.

Remote sensing data preparation and preprocessing

In this study, Sentinel-2 satellite images—including Band 2 (Blue, 0.458–0.523 µm), Band 3 (Green, 0.53–0.59 µm), Band 4 (Red, 0.64–0.67 µm), Band 8 (NIR, 0.85–0.88 µm), Band 11 (SWIR1, 1.57–1.65 µm) and Band 12 (SWIR2, 2.11–2.25 µm)—were used to extract the relevant indices. The data was obtained via the Google

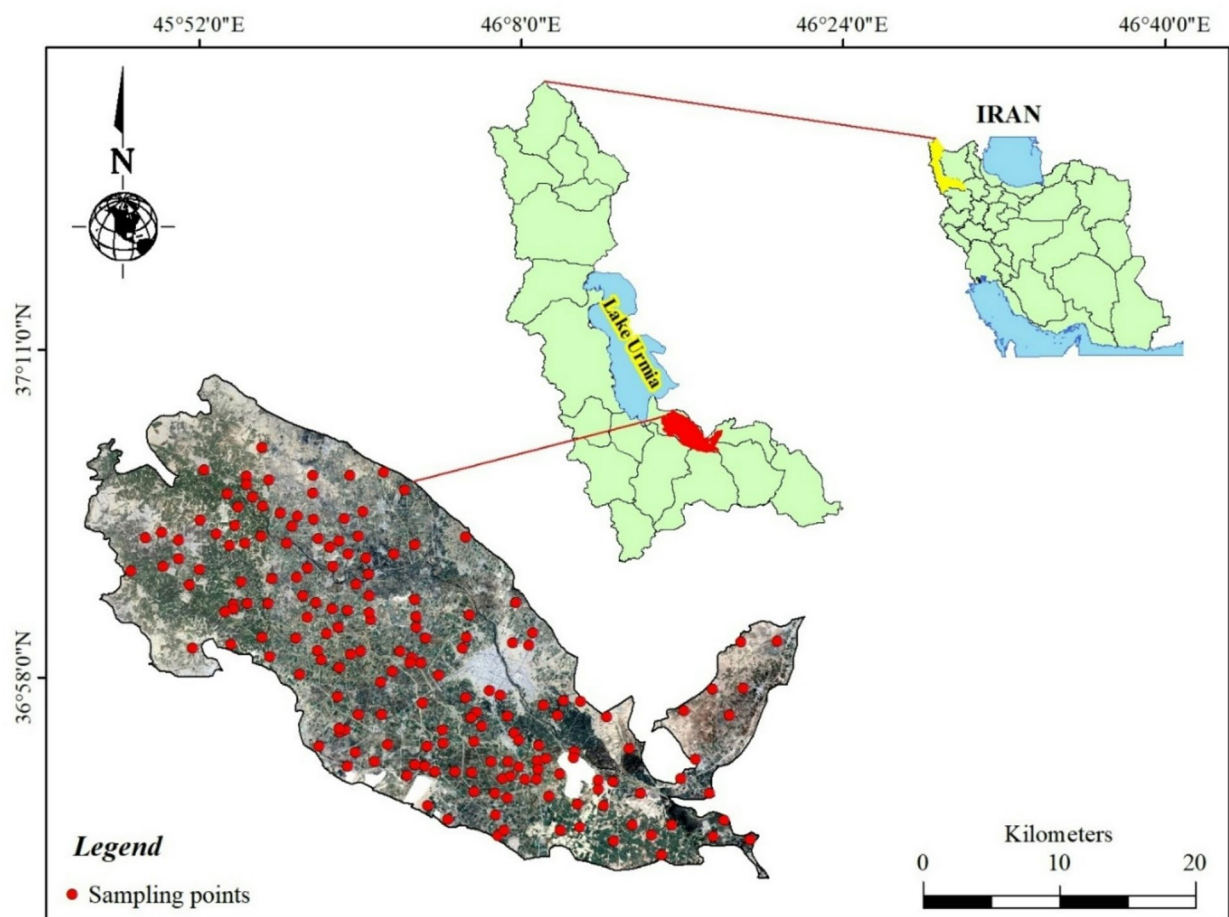


Fig. 1. Study area boundary and sampling locations in Miandoab plain, West Azerbaijan Province.

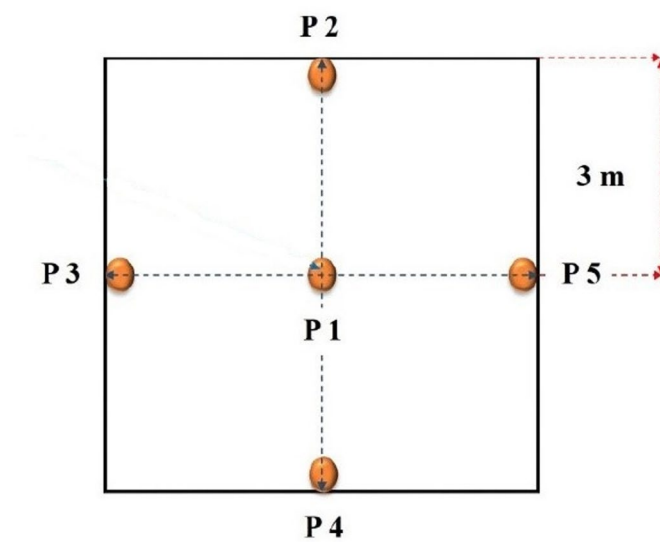


Fig. 2. Soil sampling locations using the 5-point sampling method.

Earth Engine platform, a robust cloud-based computing environment designed for large-scale remote sensing and geospatial data processing. A specific time window was selected to minimize cloud cover, restricting it to less than 10% to ensure data quality. To enhance spectral fidelity and reduce atmospheric interference, an atmospheric correction was applied to the images. This process used the physically based SEN2COR algorithm, which was developed specifically for Sentinel-2 datasets. SEN2COR uses atmospheric parameters such as water vapor, aerosols, and surface pressure to remove scattering and absorption effects, and convert the data from top-of-atmosphere (TOA) reflectance to bottom-of-atmosphere (BOA) reflectance. For validation purposes, complementary atmospheric correction methods based on empirical models, including dark object subtraction (DOS), were also implemented and cross-compared with the SEN2COR outputs. After atmospheric correction, all images were subjected to geometric co-registration, and the spectral bands were resampled to a uniform spatial resolution of 20 m using bilinear interpolation. This resampling technique was chosen to avoid spatial discontinuities and distortions. It is noteworthy that the Sentinel-2 bands (e.g. blue, green, red, and near-infrared bands) originally recorded at a resolution of 10 m and those recorded at a resolution of 60 m (typically in conjunction with atmospheric correction bands) were rescaled accordingly to the targeted spatial resolution of 20 m. This harmonization was essential for the subsequent integrated data analysis¹⁹.

Standardized spectral reflectance index calculation

To determine the standardized spectral reflectance index (SSRI), all relevant reflectance bands from the Sentinel-2 images were first atmospherically corrected and spatially resampled to a common resolution. Each band was then normalized to a range of [0,1] to eliminate scale disparities among them, according to the Eq. 1^{20,21}.

$$R_t = \frac{R_i - R_{i-min}}{R_{i-max} - R_{i-min}} \quad (1)$$

where, R_i denotes the reflectance value of band i and R_t is the normalized reflectance.

PCA was then applied to the matrix of normalized reflectance values. This dimensionality reduction technique enabled the extraction of principal spectral patterns by transforming the correlated spectral bands into a series of uncorrelated components. The first principal component (PC1), which has the highest variance in the spectral data, was selected as the core component of the SSRI. To standardize PC1, a Z-score normalization was performed as follows (Eq. 2).

$$SSRI = \frac{(PC1 - \mu)}{\sigma} \quad (2)$$

where, μ and σ represent the mean and standard deviation of PC1 values across the entire image, respectively.

The SSRI enhances spectral contrast and enables effective discrimination of surface features such as vegetation, bare soil, water bodies, and urban areas. By integrating spectral information from all relevant bands into a single index, the SSRI provides a robust metric for analyzing land cover and assessing spectral variability. Importantly, this ensures that the SSRI is not just another spectral index but rather a generalized and standardized framework that can be consistently applied across different regions and time periods.

Development of predictive models for spatial distribution of soil macronutrients using the SSRI

In this study, prediction models for macronutrient concentrations in soil were developed based on the SSRI. First, the point data obtained from chemical soil analyzes were divided into two different subsets. To ensure the robustness of the model, a random sampling approach was employed, which 70% of the data used for model train and the remaining 30% for test and performance evaluation. After calculating the SSRI values for the entire study area, simple linear regression analyzes were performed to investigate the relationships between spectral reflectance (expressed by SSRI) and soil chemical properties. Separate regression models were established for each macronutrient element, with SSRI as the independent variable and measured nutrient concentration as the dependent variable. Using these regression equations and the spatial SSRI map, predictions of nutrient concentration were made for each image pixel, resulting in spatial distribution maps of soil macronutrients²¹.

Model evaluation

To assess the spatial accuracy and reliability of the predictive maps, the estimated nutrient concentrations were extracted at the geographic locations corresponding to the points in the test dataset and then compared with the actual measured values. The quality of the predictions was assessed using various statistical metrics, including the coefficient of determination (R^2), root mean square error (RMSE), mean squared error (MSE), and the ratio of performance to interquartile distance (RPIQ). While R^2 , RMSE, and MSE provide information on the overall fit and error magnitude, RPIQ offers a scale-independent measure of predictive performance by relating the model error to the interquartile range of observed values. This combination of metrics provided a more comprehensive understanding of the model's performance in capturing the spatial variability of soil macronutrients^{22,23}. The statistical metrics were calculated as follows (Eqs. 3–6).

$$R^2 = 1 - \frac{\sum_{n=1}^n (y_i - y_j)^n}{\sum_{n=1}^n (\bar{y} - y_j)^n} \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2} \quad (4)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \quad (5)$$

$$RPIQ = \frac{Q_3 - Q_1}{RMSE} = \frac{IQR(y)}{RMSE} \quad (6)$$

where, y_i and y_j are the observed and predicted values, respectively; \bar{y} is the mean of the observed values; Q_1 and Q_3 represent the first and third quartiles of the observed values; and $IQR(y)$ denotes the interquartile range, which is defined as the difference between the third and first quartiles. A higher RPIQ value indicates better model performance. Typically, RPIQ values are interpreted as follows: $RPIQ > 2$ indicates a good model; $1.5 \leq RPIQ \leq 2$ indicates an acceptable model; and $RPIQ < 1.5$ indicates a weak model.

Results

Descriptive statistical analysis of soil macronutrients

The mean TN content in the soil was found to be 0.092%. The minimum and maximum values were recorded at 0.010% and 0.80%, respectively. A coefficient of variation (CV) of 75.2% indicates a relatively high variability of nitrogen concentrations across the studied soils (Table 1). This heterogeneity may be due to differences in agricultural practices, vegetation cover, and the rate of organic matter decomposition at different sites²⁴. Among the analyzed macronutrients, AP showed the greatest variability, with a CV of 164.9%, indicating considerable variation in its concentration. AP levels ranged from 1.8 to 361.0 ppm, with a mean value of 20.78 ppm (Table 1). This pronounced variability can be attributed to the dynamic behavior of AP in soils, its particular chemical properties, and anthropogenic influences. In contrast to other macronutrients, P is relatively immobile in soil, as it frequently becomes fixed in insoluble forms such as calcium, iron, and aluminum phosphates. As a result, despite repeated applications of phosphate fertilizers by farmers in each growing season, AP tends to accumulate in localized and non-uniform patterns within the soil profile. In addition, many farmers routinely apply phosphorus fertilizers each year without conducting soil tests. This practice often leads to excessive phosphorus loading in certain areas, and exacerbates spatial heterogeneity. The problem is particularly evident in agricultural lands with a long history of chemical fertilizer usage²⁵. The mean AK content was 377.42 ppm, with values ranging from 52 to 1321 ppm. The CV for K was 49.3%, indicating moderate variability compared to nitrogen and phosphorus (Table 1). This variability can be influenced by factors such as the mineralogical composition of the parent material, cropping pattern, irrigation management, and site-specific potassium fertilization practices^{26,27}.

Correlation between spectral bands and soil macronutrients

The Pearson correlation matrix between selected Sentinel-2 spectral bands and three soil macronutrients showed that TN had the strongest correlation with spectral reflectance. The correlation coefficients for TN were consistently high across all bands and statistically significant at a confidence level of 99% ($\alpha = 0.01$). The highest correlation was observed in the green band (Band 3) with a coefficient of 0.878, closely followed by the red band (Band 4) with a value of 0.876 (Table 2). These strong associations are likely due to the direct influence of nitrogen on chlorophyll content in plants, which significantly affects reflectance in the visible and near-infrared (NIR) regions of the spectrum. In contrast, AP showed the weakest correlation with the spectral bands. Its correlation coefficients were generally low, ranging from 0.22 to 0.24, and none reached statistical significance (Table 2). This weak correlation may reflect relatively low mobility of phosphorus in soil and its indirect influence on plant spectral characteristics. AK showed moderate correlations with several spectral bands, with the highest correlation observed at Band 11 (SWIR 1), at 0.591 (Table 2). Considering the role of potassium in stomatal regulation and plant water balance mechanisms, this relationship is biologically plausible, although it is significantly weaker than that observed for nitrogen. Overall, the results indicate that nitrogen is more strongly associated with spectral reflectance in the visible, NIR, and SWIR regions, while phosphorus exerts a comparatively smaller influence on spectral responses. These results indicate that the nitrogen content in soil can be estimated more accurately and reliably by remote sensing data than the other two macronutrients investigated.

Statistic	TN (%)	AP (ppm)	AK (ppm)
Minimum	0.01	1.80	52.00
Maximum	0.80	361.00	1321.00
Mean	0.09	20.78	377.42
Standard deviation	0.07	34.28	185.97
CV	75.23	164.94	49.27

Table 1. Summary of descriptive statistics for soil macronutrient concentrations.

Variables	N	AP	AK	Band 2	Band 3	Band 4	Band 8	Band 11	Band 12
TN	1								
AP	0.112	1							
AK	0.161	0.319*	1						
Band 2	0.861**	0.240*	0.561**	1					
Band 3	0.878**	0.224	0.576**	0.959**	1				
Band 4	0.876**	0.227	0.574**	0.959**	0.969**	1			
Band 8	0.846**	0.241*	0.578**	0.947**	0.953**	0.954**	1		
Band 11	0.850**	0.236	0.591**	0.953**	0.953**	0.963**	0.962**	1	
Band 12	0.865**	0.224	0.542**	0.961**	0.946**	0.954**	0.944**	0.959**	1

Table 2. Pearson correlation coefficients between Sentinel-2 spectral bands and soil macronutrients.

**Correlation is significant at the 0.01 level. *Correlation is significant at the 0.05 level.

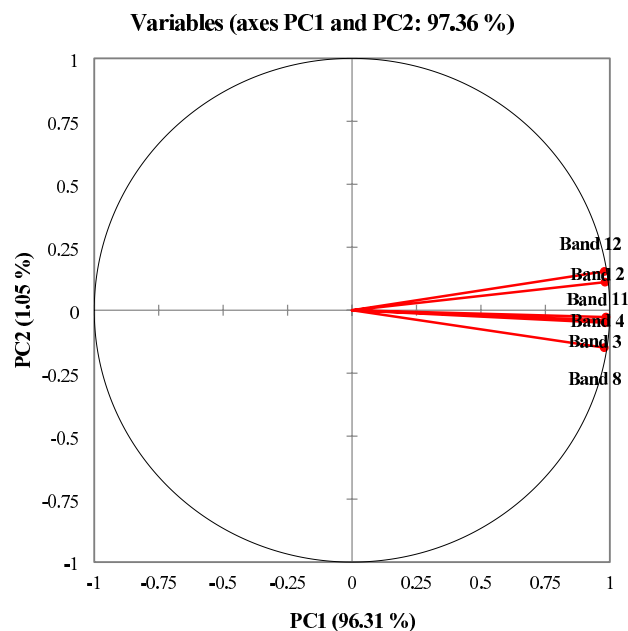


Fig. 3. PCA of various spectral bands from Sentinel-2 satellite imagery.

Correlation analysis between the SSRI and soil macronutrients

PCA was performed to reduce the dimensionality of the spectral data and to extract the most important features from selected Sentinel-2 bands. The results showed that the first principal component alone explained 96.31% of the total variance in spectral reflectance values (Fig. 3). This high percentage indicates that most of the useful information contained in the spectral bands is actually concentrated in the first component. Consequently, this component was standardized and referred to as SSRI to be used in the subsequent correlation analyses with soil macronutrients.

Pearson correlation analysis revealed a strong and statistically significant positive correlation between SSRI and TN in soil ($r=0.875$), significant at the 99% confidence level (Fig. 4). This finding indicates that nitrogen, which directly influences chlorophyll concentration and plant greenness, plays a dominant role in controlling spectral reflectance variations. A moderate and statistically significant correlation was also observed between SSRI and AK ($r=0.560$) (Fig. 4). This relationship may reflect the involvement of potassium in important physiological processes such as water regulation, photosynthesis efficiency, and cellular structural stability. In contrast, the correlation between SSRI and AP was relatively weak ($r=0.230$) and not statistically significant at a 99% confidence level (Fig. 4). This result is consistent with the limited mobility of phosphorus in the soil and its indirect influence on the reflectance properties of plants.

Regression analysis of SSRI for estimating soil macronutrient concentrations

To evaluate the potential of the SSRI, derived from Sentinel-2 satellite data, in estimating soil macronutrient concentrations, regression analyses were performed for TN, AK, and AP. The results revealed that SSRI had a

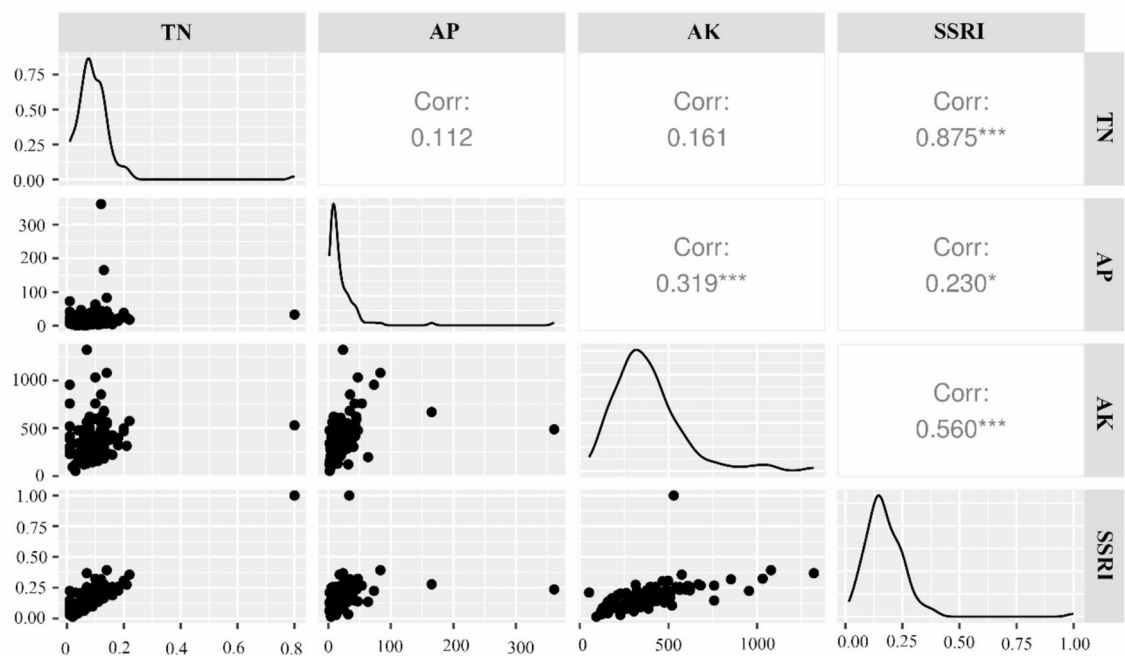


Fig. 4. Correlation between spectral SSRI and soil macronutrient concentrations.

Dataset (n = 181)	Train (n = 126)	Test (n = 55)
RMSE-TN (%)	0.04	0.02
RMSE-AK (ppm)	166.49	110.37
MSE-TN (%)	0.01	0.001
MSE-AK (ppm)	0.27	0.12
RPIQ-TN (%)	2.43	1.25
RPIQ-AK (ppm)	2.21	1.66

Table 3. Model validation results for soil macronutrient prediction.

statistically significant linear relationship with TN and AK concentrations, whereas no such relationship was observed for AP; consequently, no regression model was developed for phosphorus.

For TN (%), the regression equation (Eq. 7) showed high predictive performance during the training phase, with an R^2 of 0.77, RMSE of 0.04%, and MSE of 0.01. The corresponding RPIQ of 2.43 indicates that the model is good in terms of predictive capability during training (Table 3 and Fig. 5a). When applied to the testing dataset (55 samples), the R^2 decreased to 0.34, indicating lower explanatory power. Nevertheless, RMSE and MSE remained low at 0.02% and 0.001%, respectively, suggesting limited numerical error. The RPIQ for testing was 1.25, reflecting weak predictive performance, which indicates that while the model maintains reasonable accuracy, its generalization ability is somewhat limited (Table 3 and Fig. 5c).

For AK, the regression model (Eq. 8) also performed satisfactorily. During training, R^2 reached 0.72, with RMSE = 166.49 (ppm) and MSE = 0.27, and an RPIQ of 2.21, indicating good predictive performance (Table 3 and Fig. 5b). In the testing phase, numerical accuracy improved, with RMSE decreasing to 110.37 (ppm) and MSE to 0.12, although R^2 decreased to 0.46 (Table 3 and Fig. 5d). The RPIQ of 1.66 suggests acceptable predictive performance, demonstrating that the AK model retains relatively stable predictive ability across both datasets despite a slight decline in explanatory power. Also, regression analysis for AP did not show a statistically significant relationship with SSRI, and thus no predictive model was developed for this nutrient.

$$\text{TN} = -1.73\text{E} - 02 + 0.65 \times \text{SSRI} (0 - 1) \quad (7)$$

$$\text{AK} = 194.75 + 1090.86 \times \text{SSRI} (0 - 1) \quad (8)$$

Spatial analysis of soil macronutrient levels across the study area

Figure 6a illustrates the spatial distribution of TN in the soils of the study area, which varies between 0.01 and 0.80%. It is evident that nitrogen concentrations are predominantly in a medium to high range in most parts of the region, especially in the western and central zones where the colors change from yellow to pink, indicating elevated levels. This pattern can be attributed to agricultural activities, the use of nitrogen-based fertilizers, and

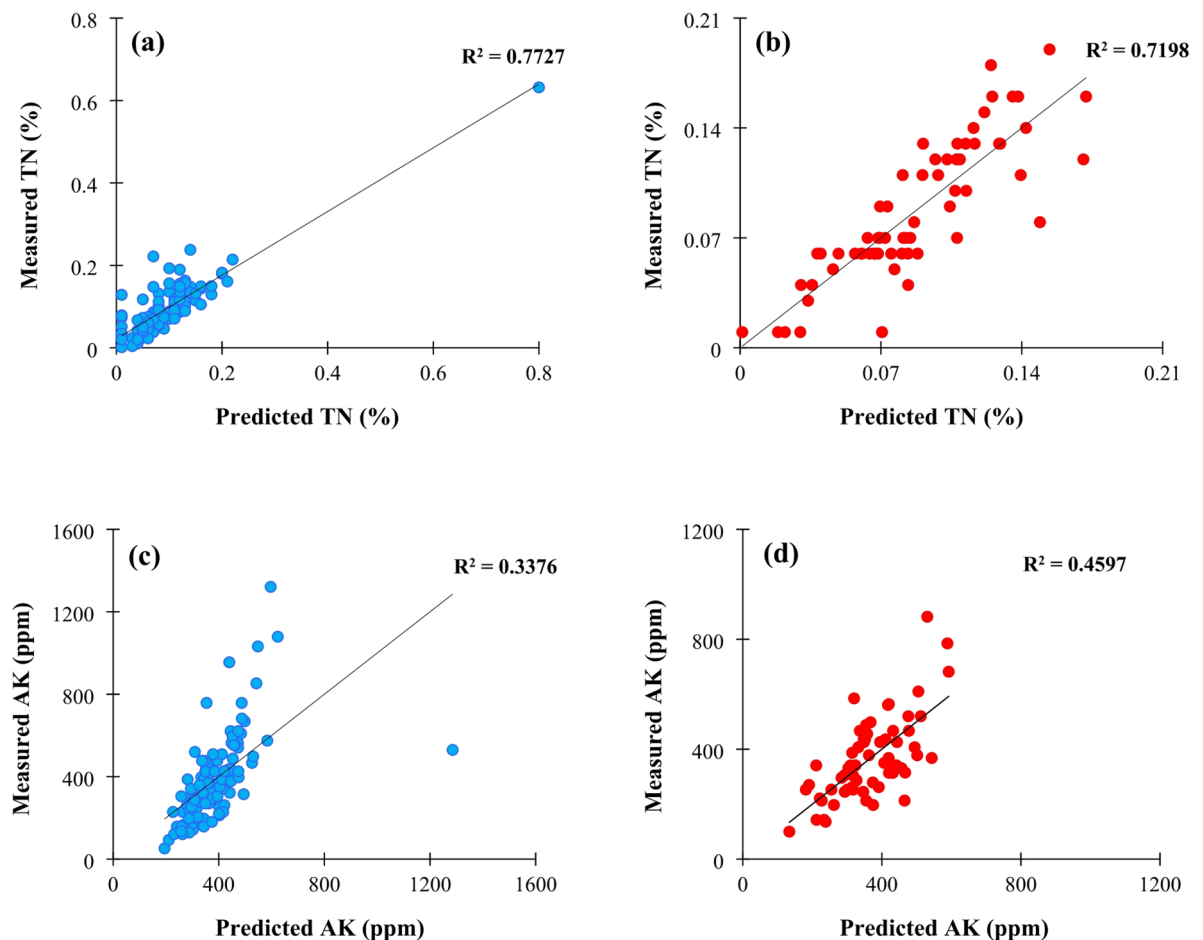


Fig. 5. Measured versus predicted values of soil macronutrient concentrations for: (a) TN (%) training data, (b) TN (%) test data, (c) AK (ppm) training data, and (d) AK (ppm) test data.

favorable climatic conditions that enhance nitrogen retention in the soil. Conversely, Fig. 6b shows the spatial distribution of AK throughout the area, with values ranging from 52 to 1321 ppm. The color gradients show that large parts of the study region—especially the southern, southeastern, and central areas—marked by yellow to orange hues, corresponding to low to moderate potassium availability. Only limited sections in the northern and eastern parts show higher potassium concentrations, represented by purple and pink shades. This uneven distribution indicates a widespread potassium deficiency throughout the study area. Given the important role of AK in regulating osmotic pressure, enzymatic activities, and nutrient transport in plants, such deficiency can significantly affect crop yields and soil health²⁸. Therefore, it is crucial to include potassium supplementation as a priority in soil management and fertilization strategies in these affected zones. Overall, these results highlight the need for site-specific nutrient management to optimize the distribution and application of essential elements across the region.

Discussion

The results of the present study show that the SSRI models the spatial variability of certain macronutrients in soil with satisfactory accuracy. The descriptive statistical analysis (Table 1) revealed that AP had the highest spatial variability within the study area, with a CV of 164.94%. In contrast, TN and AK showed moderate to high variability, with CV values of 75.23% and 49.27%, respectively. The pronounced heterogeneity observed, particularly for AP, may be attributed to the improper and untested application of phosphate fertilizers, the low stability of phosphorus in soil, and the formation of insoluble compounds^{29–31}. Conversely, the relatively lower variability of AK could be related to the mineralogical origin of the soil, cropping patterns, and irrigation practices^{32–34}.

Correlation analysis between the Sentinel-2 spectral bands and the soil nutrient elements (Table 2) showed that TN had the strongest significant correlations in all bands, especially with the green band (Band 3, $r = 0.878$) and the red band (Band 4, $r = 0.876$). These results are in good agreement with the direct involvement of nitrogen in chlorophyll synthesis and its influence on the spectral reflectance properties of vegetation³⁵. In contrast, AP showed weak correlations with the spectral bands ($r \approx 0.22$ – 0.24), none of which reached statistical significance. This probably reflects the limited mobility of phosphorus in the soil and its indirect effects on plant cover. AK

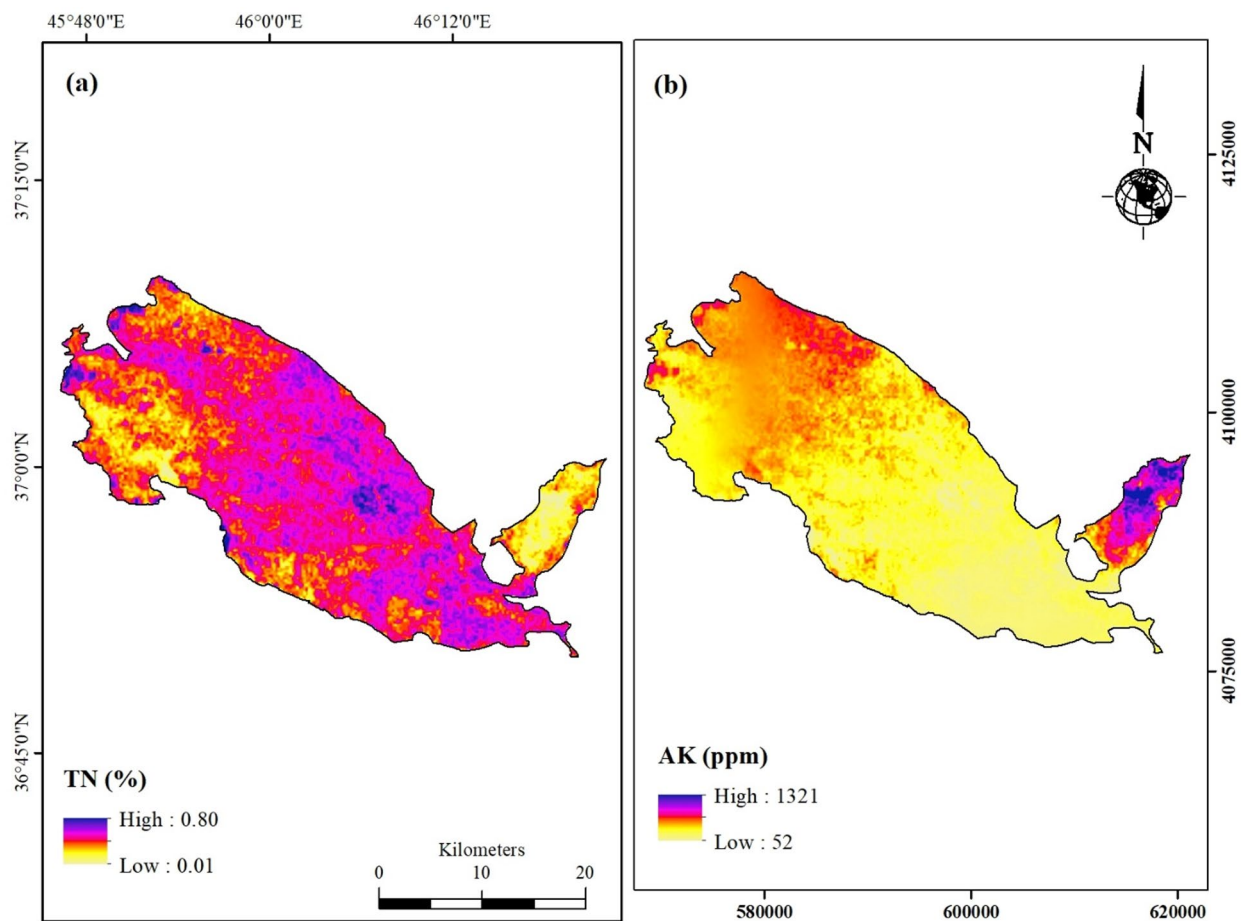


Fig. 6. Spatial patterns of TN (a) and AK (b) across the study area.

showed moderate correlations, especially with the SWIR1 band (Band 11, $r=0.591$), which may be attributed to role of potassium in regulating water balance, stomatal function, and stabilizing cell structure²¹.

PCA revealed that PC1 accounted for more than 96.31% of the total variance in the spectral data (Fig. 3). This substantial proportion indicates that the most important and dominant information contained in the selected bands is encapsulated in this component. After standardization of PC1 using the Z-score method, the SSRI was derived and subsequently used as a composite spectral reflectance index in further analyzes. The relationship between the SSRI and the macronutrient elements (Fig. 4) showed significant positive correlations with TN ($r=0.875$) and AK ($r=0.560$), while AP continued to show a weak and statistically insignificant correlation ($r=0.230$). This pattern confirms the superior ability of SSRI in modeling nitrogen content compared to other nutrients, especially considering that nitrogen is more strongly associated with plant growth and spectral reflectance Faramarzi et al.¹⁵, Cheng³⁶.

The regression analysis between the SSRI and the macronutrient elements (Fig. 5 and Table 3) confirmed the previous results. The TN model showed high accuracy in the training phase with a R^2 of 0.77, RMSE of 0.04%, MSE of 0.01, and RPIQ = 2.43 (good predictive performance); however, its performance decreased in the test phase ($R^2=0.34$, RMSE = 0.02%, MSE = 0.001, RPIQ = 1.25 (weak predictive performance)). The pronounced decrease in R^2 during validation, particularly for TN, is likely attributable to the high mobility and dynamic behavior of nitrogen in soils. Nitrogen is among the most labile soil nutrients, subject to continuous transformations through mineralization, immobilization, leaching, denitrification, and rapid plant uptake. These pronounced spatiotemporal variations result in considerable heterogeneity in nitrogen distribution, thereby constraining the predictive capacity of the model during validation. Consequently, the observed decline in model performance for TN reflects the inherent variability and complexity of the soil nitrogen cycle rather than limitations of the modeling approach. Notably, despite this decline, the low RMSE during testing indicates that the model retains strong practical applicability and reliability for field-scale predictions^{28,37}.

The AK model exhibited a similar trend, with R^2 values of 0.72 and 0.46, RMSE values of 166.49 ppm and 110.37 ppm, MSE values of 0.27 and 0.12, and RPIQ values of 2.21 (good predictive performance) and 1.66 (acceptable predictive performance) in the training and testing phases, respectively (Fig. 5 and Table 3). Compared with the TN model, the AK model showed relatively greater stability during the testing phase. This behavior can be explained by the chemical characteristics of potassium in soils. AK is predominantly present in ionic form (K^+) in the soil solution and, unlike phosphorus, it is not readily immobilized through fixation

reactions. Consequently, it is more mobile within the soil profile and less likely to be bound in insoluble forms such as calcium precipitates or iron and aluminum oxides. This higher mobility and availability in the soil solution means that variations in AK are more strongly reflected in plant nutritional status and, indirectly, in canopy spectral reflectance, thereby allowing more accurate estimation from remote sensing data^{38–41}.

In contrast, the results showed no significant relationship between AP and the SSRI index, and thus no predictive model for AP could be developed. This limitation is closely tied to the inherent characteristics of phosphorus in soils. Phosphorus is a nutrient with extremely low mobility; it is often concentrated in subsurface soil layers and is rapidly fixed through reactions with Ca^{2+} in alkaline soils or Fe^{3+} and Al^{3+} in acidic soils, resulting in the formation of insoluble compounds. As a result, the fraction of phosphorus that remains available in the soil solution is minimal and its direct influence on plant reflectance signals is weak^{21,42,43}.

From a remote sensing perspective, spectral indices primarily capture signals from the soil surface and vegetation canopy. Consequently, nutrients with low solubility and limited translocation, such as phosphorus, exert little to no direct effect on spectral reflectance. This suggests that the lack of predictability for AP is not only a limitation of the SSRI approach but also reflects the broader constraints of spectral indices for estimating low-mobility nutrients. In other words, while SSRI proved effective for more mobile nutrients such as TN and AK, the prediction of phosphorus content remains challenging due to its low mobility, subsurface distribution, and weak spectral activity^{44–46}.

The spatial analysis (Fig. 6a and b) further supports the statistical findings. TN concentrations predominantly range from moderate to high across most parts of the study area, especially in the western and central regions, probably due to repeated applications of nitrogenous fertilizers as well as more favorable climatic conditions for nitrogen retention. Conversely, the spatial distribution of AK shows a widespread deficiency in the southern, southeastern, and central parts of the region. This heterogeneous pattern emphasizes the need for site-specific nutrient management, especially for potassium, to increase crop productivity. In addition, the use of high-resolution maps generated through this approach can play a crucial role in optimizing the use of nutrients^{47,48}. In summary, the results of this study show that the SSRI-based approach in combination with simple statistical models offers significant potential for predicting the spatial variability of nitrogen and potassium in soil. However, elements with lower mobility, such as AP, still require direct measurement methods or more sophisticated modeling approaches that integrate remote sensing data, soil properties, and land management history.

Conclusion

The results of this study demonstrated that the Sentinel-2 satellite data and the spectral indices derived from them—in particular the novel SSRI based on principal component analysis—have considerable capabilities in the spatial modeling of certain soil macronutrients, especially TN and AP. The observed significant correlations between the spectral indices and TN content reflect the indirect but reliable influence of vegetation cover on spectral reflectance, allowing inference of plant nutritional status from remote sensing data. The model performance for AK was also acceptable, although with lower accuracy compared to nitrogen. Conversely, AP modeling faced limitations mainly due to its limited mobility, subsurface concealment in the soil layers, and lack of direct impact on the spectral indices. These results suggest that the integration of soil science knowledge with remote sensing technologies can provide an effective, non-destructive, rapid, and cost-efficient approach for monitoring soil nutrient status at agricultural and regional scales. Based on these results, it is recommended that future research aiming to improve the modeling accuracy of low-mobility or spectrally inactive elements include a combination of remote sensing data with ground-truth information, soil physicochemical properties, and topographic parameters. In addition, the use of multi-temporal Sentinel-2 imagery throughout the growing season and the derivation of spectral reflectance indices from these time-series images can better capture nutrient dynamics, enabling dynamic monitoring and timely management recommendations. The expansion of such integrated approaches in the context of precision agriculture has the potential to optimize fertilizer use, increase productivity, and reduce environmental impacts in agricultural systems. Moreover, the SSRI index could serve as a valuable environmental variable for integration into nonlinear and machine learning models, potentially enhancing predictive accuracy and providing deeper insights in future studies.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request. The data of satellite images and digital elevation model are available in the Google Earth Engine platform.

Received: 14 September 2025; Accepted: 16 October 2025

Published online: 20 November 2025

References

1. Dutta, A., Banerjee, M. & Ray, R. Land capability assessment of Sali watershed for agricultural suitability using a multi-criteria-based decision-making approach. *Environ. Monit. Assess.* **196**, 237. <https://doi.org/10.1007/s10661-024-12393-9> (2024).
2. Zhang, M. et al. Evaluation of soil fertility using combination of Landsat 8 and Sentinel-2 data in agricultural lands. *Environ. Monit. Assess.* **196**, 131. <https://doi.org/10.1007/s10661-024-12301-1> (2024).
3. Sharma, S., Kaur, G., Singh, P., Boparai, A. & Dhaliwal, S. S. Development of artificial neural networks for predicting soil micro-nutrients availability under rice-based cropping systems of north-western India. *J. Soil Sci. Plant Nutr.* **24**, 884–904. <https://doi.org/10.1007/s42729-023-01593-w> (2024).
4. Sishodia, R. P., Ray, R. L. & Singh, S. K. Applications of remote sensing in precision agriculture: A review. *Remote Sens.* **12**, 3136. <https://doi.org/10.3390/rs12193136> (2020).
5. Dimkpa, C. et al. In *Sustainable Agriculture Reviews* (ed Eric Lichtfouse) 1–43 (Springer International Publishing, 2017).

6. Mikula, K. et al. Controlled release micronutrient fertilizers for precision agriculture—a review. *Sci. Total Environ.* **712**, 136365. <https://doi.org/10.1016/j.scitotenv.2019.136365> (2020).
7. Mazur, P., Gozdowski, D. & Wnuk, A. Relationships between soil electrical conductivity and Sentinel-2-derived NDVI with pH and content of selected nutrients. *Agronomy* **12**, 354. <https://doi.org/10.3390/agronomy12020354> (2022).
8. Ren, J., Zhao, Y. & Yu, K. LIBS in agriculture: A review focusing on revealing nutritional and toxic elements in soil, water, and crops. *Comput. Electron. Agric.* **197**, 106986. <https://doi.org/10.1016/j.compag.2022.106986> (2022).
9. Nandal, A. et al. Soil quality under different tree species in an urban university campus: A multidimensional study. *Environ. Earth Sci.* **83**, 613. <https://doi.org/10.1007/s12665-024-11902-w> (2024).
10. Ramut, R. et al. The effect of broadcast struvite fertilization on element soil content and microbial activity changes in winter wheat cultivation in southwest Poland. *Sci. Rep.* **15**, 6451. <https://doi.org/10.1038/s41598-025-90384-7> (2025).
11. Shankar, K. & Moorthi, M. An efficient IoT-based crop yield prediction framework using optimal ensemble learning and hybridized optimization model. *Earth Sci. Inf.* **18**, 156. <https://doi.org/10.1007/s12145-024-01525-x> (2025).
12. Hu, B. et al. Fine-resolution baseline maps of soil nutrients in farmland of Jiangxi Province using digital soil mapping and interpretable machine learning. *CATENA* **249**, 108635. <https://doi.org/10.1016/j.catena.2024.108635> (2025).
13. Khanal, S., Kc, K., Fulton, J. P., Shearer, S. & Ozkan, E. Remote sensing in agriculture—accomplishments, limitations, and opportunities. *Remote Sens.* **12**, 3783. <https://doi.org/10.3390/rs12223783> (2020).
14. Pandey, P. C. et al. Geospatial approach to assess the impact of nutrients on rice equivalent yield using MODIS sensors'-based MOD13Q1-NDVI data. *IEEE Sens. J.* **15**, 6108–6115. <https://doi.org/10.1109/JSEN.2015.2451113> (2015).
15. Faramarzi, S. E., Pazira, E., Masihabadi, M. H., Mohammadi Torkashvand, A. & Motamedvaziri, B. Modeling and estimating the spatial distribution of soil organic matter content in irrigated lands. *Int. J. Environ. Sci. Technol.* **19**, 7399–7410. <https://doi.org/10.1007/s13762-022-03909-2> (2022).
16. Bremner, J. M. Methods of Soil Analysis. *SSSA Book Series*, 1085–1121 (1996).
17. Olsen, S. & Sommers, E. Phosphorus soluble in sodium bicarbonate. In *Methods of Soil Analysis*, Part 2, 404–430 (1982).
18. Helmke, P. A. & Sparks, D. L. In *Methods of Soil Analysis SSSA Book Series* 551–574 (1996).
19. Abbaszad, P., Asadzadeh, F., Rezapour, S., Khosravi Aqdam, K. & Shabani, F. Evaluation of Landsat 8 and Sentinel-2 vegetation indices to predict soil organic carbon using machine learning models. *Model. Earth Syst. Environ.* **10**, 2581–2592. <https://doi.org/10.1007/s40808-023-01916-x> (2024).
20. Khosravi Aqdam, K. et al. Predicting the spatial distribution of soil mineral particles using OLI sensor in northwest of Iran. *Environ. Monit. Assess.* **193**, 377. <https://doi.org/10.1007/s10661-021-09163-2> (2021).
21. Miran, N. et al. Predicting soil nutrient contents using Landsat OLI satellite images in rain-fed agricultural lands, northwest of Iran. *Environ. Monit. Assess.* **193**, 607. <https://doi.org/10.1007/s10661-021-09397-0> (2021).
22. Malone, B. P., Minasny, B. & McBratney, A. B. Using R for Digital Soil Mapping (eds Brendan P. Malone, Budiman Minasny, & Alex B. McBratney) 1–5 (Springer International Publishing, 2017).
23. Xu, X., Wang, Z., Song, X., Zhan, W. & Yang, S. A remote sensing-based strategy for mapping potentially toxic elements of soils: Temporal-spatial-spectral covariates combined with random forest. *Environ. Res.* **240**, 117570. <https://doi.org/10.1016/j.envres.2023.117570> (2024).
24. Viaud, V., Angers, D. A. & Walter, C. Toward landscape-scale modeling of soil organic matter dynamics in agroecosystems. *Soil Sci. Soc. Am. J.* **74**, 1847–1860. <https://doi.org/10.2136/sssaj2009.0412> (2010).
25. Li, C. et al. Response of agronomic traits and phosphorus uptake to soil P deficiency during rice cultivars improvement. *Agronomy* **15**, 983 (2025).
26. Haider, M. W. et al. Assessment of salicylic acid and potassium nitrate to mitigate frost stress in autumn-sown potato crop cv Sutlej. *Sci. Rep.* **15**, 1942. <https://doi.org/10.1038/s41598-025-85769-7> (2025).
27. Souza, V. S. et al. Cover crops enhance soil health, crop yield and resilience of tropical agroecosystem. *Field Crop Res.* **322**, 109755. <https://doi.org/10.1016/j.fcr.2025.109755> (2025).
28. Sardans, J. & Peñuelas, J. Potassium control of plant functions: Ecological and agricultural implications. *Plants* **10**, 419. <https://doi.org/10.3390/plants10020419> (2021).
29. Chien, S. H., Prochnow, L. I., Tu, S. & Snyder, C. S. Agronomic and environmental aspects of phosphate fertilizers varying in source and solubility: An update review. *Nutr. Cycl. Agroecosyst.* **89**, 229–255. <https://doi.org/10.1007/s10705-010-9390-4> (2011).
30. Zhu, J., Li, M. & Whelan, M. Phosphorus activators contribute to legacy phosphorus availability in agricultural soils: A review. *Sci. Total Environ.* **612**, 522–537. <https://doi.org/10.1016/j.scitotenv.2017.08.095> (2018).
31. Ibrahim, M. et al. Phosphorus mobilization in plant-soil environments and inspired strategies for managing phosphorus: A review. *Agronomy* **12**, 2539. <https://doi.org/10.3390/agronomy12102539> (2022).
32. Adhikari, G. & Bhattacharyya, K. G. Correlation of soil organic carbon and nutrients (NPK) to soil mineralogy, texture, aggregation, and land use pattern. *Environ. Monit. Assess.* **187**, 735. <https://doi.org/10.1007/s10661-015-4932-5> (2015).
33. Bakker, E., Lanson, B., Findling, N., Wander, M. M. & Hubert, F. Mineralogical differences in a temperate cultivated soil arising from different agronomic processes and plant K-uptake. *Geoderma* **347**, 210–219. <https://doi.org/10.1016/j.geoderma.2019.04.010> (2019).
34. Cheng, J. et al. Divergent responses of cropland soil available potassium to fertilization across the Sichuan Basin of China. *Soil Till. Res.* **253**, 106654. <https://doi.org/10.1016/j.still.2025.106654> (2025).
35. Serret, M. D. et al. Vegetation indices derived from digital images and stable carbon and nitrogen isotope signatures as indicators of date palm performance under salinity. *Agric. Water Manag.* **230**, 105949. <https://doi.org/10.1016/j.agwat.2019.105949> (2020).
36. Cheng, L. Comparing sentinel-2 and Landsat 8 spectral reflectance indices for predicting soil organic carbon. *Environ. Earth Sci.* **84**, 216. <https://doi.org/10.1007/s12665-025-12235-y> (2025).
37. Chen, B., Liu, E., Tian, Q., Yan, C. & Zhang, Y. Soil nitrogen dynamics and crop residues: A review. *Agron. Sustain. Dev.* **34**, 429–442. <https://doi.org/10.1007/s13593-014-0207-8> (2014).
38. Pätzold, S. et al. Predicting plant available phosphorus using infrared spectroscopy with consideration for future mobile sensing applications in precision farming. *Precis. Agric.* **21**, 737–761. <https://doi.org/10.1007/s11119-019-09693-3> (2020).
39. Jain, S., Sethia, D. & Tiwari, K. C. A critical systematic review on spectral-based soil nutrient prediction using machine learning. *Environ. Monit. Assess.* **196**, 699. <https://doi.org/10.1007/s10661-024-12817-6> (2024).
40. Lachgar, A., Mulla, D. J. & Adamchuk, V. Implementation of proximal and remote soil sensing, data fusion and machine learning to improve phosphorus spatial prediction for farms in Ontario Canada. *Agronomy* **14**, 693. <https://doi.org/10.3390/agronomy14040693> (2024).
41. Saidi, S. et al. Use of Vis-NIR reflectance spectroscopy for estimating soil phosphorus sorption parameters at the watershed scale. *Soil Till. Res.* **248**, 106460. <https://doi.org/10.1016/j.still.2025.106460> (2025).
42. Guo, J., Jin, S. & Wang, K. Machine learning and genetic algorithm for mapping soil available phosphorus in coastal provinces in Southeast China. *Ecol. Ind.* **166**, 112294. <https://doi.org/10.1016/j.ecolind.2024.112294> (2024).
43. Lou, H. et al. Detecting and analyzing soil phosphorus loss associated with critical source areas using a remote sensing approach. *Sci. Total Environ.* **573**, 397–408. <https://doi.org/10.1016/j.scitotenv.2016.08.048> (2016).
44. Odehri, O. et al. Remote sensing of depth-induced variations in soil organic carbon stocks distribution within different vegetated landscapes. *CATENA* **243**, 108216. <https://doi.org/10.1016/j.catena.2024.108216> (2024).
45. Kok, M., Sarjant, S., Verweij, S., Vaessen, S. F. C. & Ros, G. H. On-site soil analysis: A novel approach combining NIR spectroscopy, remote sensing and deep learning. *Geoderma* **446**, 116903. <https://doi.org/10.1016/j.geoderma.2024.116903> (2024).

46. Yang, J., Weisberg, P. J. & Bristow, N. A. Landsat remote sensing approaches for monitoring long-term tree cover dynamics in semi-arid woodlands: Comparison of vegetation indices and spectral mixture analysis. *Remote Sens. Environ.* **119**, 62–71. <https://doi.org/10.1016/j.rse.2011.12.004> (2012).
47. Jin, J. & Jiang, C. Spatial variability of soil nutrients and site-specific nutrient management in the P.R. China. *Comput. Electron. Agric.* **36**, 165–172. [https://doi.org/10.1016/S0168-1699\(02\)00099-6](https://doi.org/10.1016/S0168-1699(02)00099-6) (2002).
48. Briat, J.-F., Gojon, A., Plassard, C., Rouached, H. & Lemaire, G. Reappraisal of the central role of soil nutrient availability in nutrient management in light of recent advances in plant nutrition at crop and molecular levels. *Eur. J. Agron.* **116**, 126069. <https://doi.org/10.1016/j.eja.2020.126069> (2020).

Author contributions

Ali Dianati: Writing—original draft, data curation, software, methodology, investigation, formal analysis, conceptualization. Mir Hassan Rasouli-Sadaghiani: Review & editing, methodology, conceptualization, supervisor. Farrokh Asadzadeh: Review & editing, methodology, data curation, conceptualization, supervisor. Ebrahim Sepehr: Review & editing, methodology, conceptualization, supervisor. Kamal Khosravi Aqdam: Review & editing, methodology, conceptualization, supervisor.

Funding

This study was supported by the Urmia University, Agricultural Jihad Organization of West Azarbaijan province, and the soil laboratory of AZAR KHAK. The authors of this article would like to thank from them for accepting all expenses of this study.

Declarations

Competing interests

The authors declare no competing interests.

Ethical approval

All authors have read, understood, and have complied as applicable with the statement on "Ethical responsibilities of Authors" as found in the Instructions for Authors and are aware that with minor exceptions, no changes can be made to authorship once the paper is submitted.

Additional information

Correspondence and requests for materials should be addressed to M.H.R.-S. or F.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025