University of Tehran Press

Online ISSN: 2345-475X

DESERT

Home page: https://jdesert.ut.ac.ir/

Spatial modeling of a soil fertility index using digital soil mapping (Case study from Honam watershed (Iran))

Fatemeh Ebrahimi Meymand[1*], Hasan Ramezanpour[2], Nafiseh Yaghmaeian[2], Kamran Eftekhari[1]

[1] Department of Soil Identification Research and Land Evaluation, Soil and Water Research Institute, Agricultural Research, Education and Extension Organization (AREEO), Karaj, Iran. E-mail: f.meymand@areeo.ac.ir
[2] Soil Science Department, College of Agriculture, University of Guilan, Rasht, Iran.

Article Info.

Article type:
Research Article

Article history:
Received: 27 Aug. 2023
Received in revised from: 18 Nov. 2023
Accepted: 09 Dec. 2023
Published online: 27 Dec. 2023

Keywords:
Entisols,
Google Earth Engine,
Inceptisols, K-fold cross-validation,
Soil fertility index.

ABSTRACT

Attempt to evaluate soil fertility was and still is one of the most challenging public importance. Soil nutrients are the key factors in soil fertility. For this reason, when constructing soil fertility potential, many researchers prefer to investigate soil nutrient status or use and assessment of qualitative research methods. Quantifying soil fertility is challenging since various factors such as numerous physical and chemical characteristics of soil might affect it. The proper selection of factors that may more accurately describe soil fertility is another issue. So, in this study, we developed a regional soil fertility index (SFI) based on different soil nutrients for quantifying soil fertility. After receiving fertility, a comparative study of machine learning techniques was carried out to construct its distribution map, using digital soil mapping (DSM). The spatial distribution of the SFI map showed that 55% of the studied area had poor fertility, 27.25% had moderately fertile soils, and only a tiny area had fertile soils. The results indicated that heavy soil texture and high calcium carbonate content were the most limiting factor and phosphorus and zinc were the most limiting nutrients across the studied area. Comparing machine learning techniques yielded the finding that the Random forest model has the best performance for predicting SFI (R^2= 0.86) compared with the Decision tree (R^2= 0.53) and Multi-linear regression (R^2= 0.35). Therefore, specific soil fertility management practices and training farmers on the proper use of soil fertility management practices are recommended.

Cite this article: Ebrahimi Meymand, F., Ramezanpour, H., Yaghmaeian, N., Eftekhari, K. (2023). Spatial modeling of a soil fertility index using digital soil mapping (Case study from Honam watershed (Iran)). DESERT, 28 (2), DOI: 10.22059/jdesert.2023.95754

DOI: 10.22059/jdesert.2023.95754

Publisher: University of Tehran Press

1. Introduction

Soil properties vary spatially across landscapes due to intrinsic and extrinsic human activities (Iticha and Takele, 2019). This variability causes differences within the main soil characteristics and along with it, differences in soil fertility. According to Pant et al. (2019), the most basic decision-making tool for a particular land-use system is soil fertility, which helps to establish appropriate management strategies for farmers, extension agents, and policymakers.

Therefore, to plan for sustainability, it is essential to understand the proper way to evaluate soil fertility status and map its spatial distribution across the landscape (Khadka et al., 2019).

Attempts to evaluate soil properties and their fertility were inherent for the primary agricultural civilizations and still are of great public importance; therefore, several generations of scientists focused their efforts thereon. The results of these efforts showed that there are direct and indirect ways of evaluating soil fertility.

Direct evaluation includes several numbers of processes using field and laboratory diagnostics and indirect evaluation consists of developing and applying a variety of mathematical models (Saglam and Dengiz, 2014). So, many studies are applied towards determining soil fertility status and suggested various approaches that concern this main objective:

Evaluation of soil properties that involved assessing soil fertility and mapping their spatial variability using geostatic interpolation techniques (Nguemezi et al., 2020; Iticha and Takele, 2019; Panday et al., 2018; Sirsat et al., 2018). Recently, Panday et al. (2018) used kriging as an ArcGIS geostatistical tool to interpolate measurements of those variables, and multiple digital map layers were developed based on each soil's chemical properties. They demonstrated how these maps help farmers evaluate the current status of their agricultural soils, make decisions about their management more easily and effectively, and ensure sustainable productivity.

Furthermore, using innovative computer-based methods such as GIS and GPS techniques for creating maps that store, analyze, and show spatial soil fertility variability (AlBudeiri and Aloosy, 2019; Babalola et al., 2019; Arif et al., 2017; Prado et al., 2012).

In various studies, SFI determination with the use of feature selection and machine learning techniques to extract the most important soil properties that affect it has been considered (Cao et al., 2021; Haryuni et al., 2020; Merumba et al., 2020; Nguemezi et al., 2020; Mokarram et al., 2018; Awasthi and Bansal, 2017). Similarly, using the ordered weighted averaging (OWA) and therefore the analytical hierarchy process (AHP) method (Fayyaz et al., 2021; Chen et al., 209; Zabihi et al., 2019). Pant et al. (2019) used machine learning algorithms and the Rough Set method to classify SFI levels and determine the accuracy of this method.

It has been highlighted across all research that quantifying soil fertility potential is challenging since a variety of factors might affect its quantification. As, numerous physical and chemical characteristics of the soil. So, the proper selection of factors that may more accurately describe soil fertility is another issue (Moral et al., 2012).

Nowadays, modern computer techniques like digital soil mapping and remote sensing have become the preferred choice for crop monitoring, crop yield forecasting, and soil fertility assessment (Hounkpatin et al., 2022; Subramanian et al., 2022; Weissa et al., 2020; Kumar et al., 2021; Bagheri Bodaghabadi et al., 2019). Additionally, it is a non-destructive, economical, and quick technique for crop growth simulation (Seo et al., 2019). Therefore, this study was conducted with the main objective of mapping the spatial distribution of soil fertility with DSM by using auxiliary variables and the most reliable machine-learning model, within the

region. The study was also conducted to accomplish these objectives: assessing the fertility status by developing an SFI model at the regional level and comparing three machine learning techniques for its spatial prediction.

2. Materials and Methods

2.1. Description of the Study Area

The study area is located in the Honam watershed in the south of Alashtar city, which is one of the most important agricultural areas in the Karkheh River basin of Iran and covers a total area of 4000 hectares. This area consists of two landforms the Piedmont Plain and a valley. Wheat fields are the most commonly used type of land in the study area. The average annual temperature is 8.8 °C, and the average annual rainfall is 554 mm. The soil moisture and temperature regimes are xeric and mesic, respectively, and the soils of the studied area were classified into three main soil great group including Haploxerepts, Calcixerepts and Xerofluvents (Soil Survey Staff, 2014). Figure 1 depicts the location of the study area as well as the locations of the surface soil sampling sites.

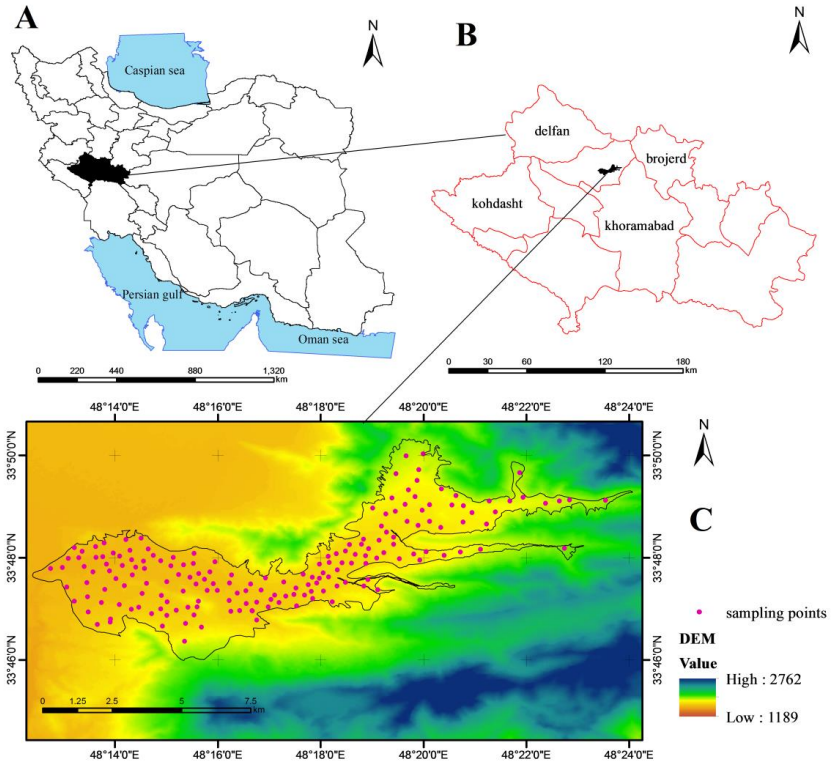

Fig.1. Location of Honam watershed in Lorestan Province, Iran (A), surface soil sampling sites in the study area (B)

2.2. Soil sampling and laboratory analysis

181 surface soil samples (0–30 cm depth) were taken during the field investigation phase utilizing stratified random sampling, which was supported by the expertise of soil scientists and cartographic products (mainly digital elevation models). The latter was used to investigate the interactions between soil and landscape and to encourage sampling in various landforms. A portable GPS was used to map out the locations of each soil sampling site. The soil's nutrient concentrations, including total nitrogen (Nelson & Sommers, 1996), available

phosphorus (Olsen, 1954), available potassium (Jackson, 1973), organic carbon (Walkley and Black, 1934), available copper, iron, manganese, and zinc (DTPA method, Lindsay & Norvell, 1978), were measured after the soil sample was transferred to the lab, air dried, and the fine components were passed through a 2 mm sieve. Similar to that, titrimetric and hydrometer techniques were used to measure the soil's CaCO3 content and particle size distribution (Soil Survey Staff, 1992). Using a pH and EC measuring glass electrode, the samples' soil pH and electrical conductivity were tested at a soil/water ratio of 1:2.5 (Jackson, 1973).

2.3. Statistical analysis and determination of soil fertility status

Using SPSS software, a descriptive study of the minimum, maximum, mean, and standard deviation of the soil nutrient content was carried out. In this work, we used the SFI model to assess the soil's fertility state (Tunçay et al., 2021). This model, which uses a parametric method for each soil type to determine soil fertility using SFI classes, is one of the best fits. Each soil is evaluated based on factor ratings ranging between 10 and 100 using a rating value for each soil parameter (Table 1). The factor rank's least favorable value is 10 and its most beneficial value is 100 (Tunçay et al., 2021). The threshold levels of some soil nutrients have been modified according to the state of this nutrient in the Iranian agricultural soils, extracted and published from 315,000 data related to 50,000 soil types, which were carried out to develop a soil fertility database and assess the fertility of the nation's agricultural land by Shahbazi and Besharati (2013).

SFI is calculated using the value of each factor rating as follows:

$$SFI = [Rmax * \sqrt{A} /100 * B /100 * \ldots\ldots.] \quad (1)$$

Where SFI: soil fertility index, Rmax: maximum ratio: (A + B+….)/n, A, B …: rating value for each indicator, and n: number of indicators (Tunçay et al, 2021).

Table 1. Factor rating of each soil parameter
(some of the soil parameter's factor rating was taken from Tunçay et al, 2021)

Indicators		**Indicator rating**				
soil physical and chemical characteristics*	Unit	100	80	50	20	10
A-Ntotal	%	>0.32	0.32–0.17	0.09–0.17	0.09–0.045	<0.045
B-Pav	$mgkg^{-1}$	>20	15-20	10-15	5-10	5
C-Kav	$mgkg^{-1}$	110-288	288-975	51-110	>975	<51
D-Mn av	$mgkg^{-1}$	14–50	4–14	50–170	>170	<4
E-Zn av	$mgkg^{-1}$	>1	0.75-1	0.5-0.75	0.25-0.5	<0.25
F-Fe av	$mgkg^{-1}$	>7.5	5-7.5	2.5-5	……..	……
G-Cu av	$mgkg^{-1}$	>1	0.25-1	<0.25	……	……
H-Caco3	%	5-15	15-25	25-40	40-60	>60
I-Salt or EC	% or dSm^{-1}	0–0.15 or 0–2	0.15–0.30 or 2–4	0.30–0.50or 4–6	0.50–0.65 or 6–8	>0.65 or >8
J-pH	-	6.5–7.5	7.5–8.5	5.5–6.5	4.5–5.5	<4.5–>8.5
K-SOM	%	>3	2–3	1–2	0.5–1	0–0.5
L-Texture	-	CL, SCL, SiCL,	vfSL, L, SiL, Si, <%50 C	>%50 C, SC, SiC	SL, fSL	S, LS

* Ntotal, Pav, Kav, Mnav, Znav, Feav, and Cuav are total nitrogen, available phosphorus, available potassium, available manganese, available zinc, available iron, and available copper, respectively. In addition, EC and SOM are the conductivity and organic matter of the soil, and the soil texture symbols are CL: clay loam, SCL: sandy clay loam, vfSL: very fine sandy loam, L: loam, C: clay, SL: sandy loam, fSL: fine sandy loam, S: sand, LS: loamy sand, SiCL: silty clay loam, SiL: silty loam, Si: silty, SC: sandy clay, SiC: silty clay.

An example of how to calculate this index for the characteristics of one of the study points is given in Table 2.

Table 2. An example of SFI index calculation for one of the study points

Indicators	A	B	C	D	E	F	G	H	I	J	K	L
Soil characteristics	Ntotal	Pav	Kav	Mn av	Zn av	Fe av	Cu av	Caco3	Salt or EC	pH	SOM	Texture
Actual values	0.111	5.4	390	21.6	0.48	13.8	2.12	34	0.9	7.4	1.11	SIC
Calculated factor rating (According to table 1)	27.87	15.27	80	100	50	100	100	50	91	100	47.07	50

Rmax=(27.87+15.27+80+100+50+100+100+50+91+100+47.07+50)/12=67.60
SFI = [67.60 *√27.87 /100* 15.27 /100*…….]=67.60*0.047=3.22

2.4. Acquisition of Environmental covariates

A stack of environmental variables database for predictive modeling was created using a variety of geospatial datasets associated with soil formation. Several digital terrain attributes, such as elevation, valley depth, curvature, slope, aspect, SAGA wetness index, topographic wetness index, plan curvature, and local curvature, were extracted and computed through a 30*30 m grid cell resolution of a digital elevation model (DEM) derived from NASA's Shuttle Radar Topography Mission (SRTM DEM), which is freely available on the USGS Earth Explorer (https://earthexplorer.usgs).

The spectral bands of Sentinel-2 images were used to develop a set of spectrum indices related to yield, parent material, and soil. To avoid losing important local variations of the study area by using coarse spatial resolution products, we have focused our research on these image data, providing a fantastic opportunity for incorporating extensive datasets with high spatial resolution data. But, due to the large data volumes, deriving spectral indices requires large amounts of data storage, high computational power, and the ability to distribute algorithms. Since its inception, the functionality of Google Earth Engine (GEE) has been freely and publicly available to everyone (https://earthengine.google.com). GEE is a cloud-based platform that offers high-performance computing services and a lot of multi-source satellite data, making satellite imagery computing a relatively fast and flexible process (Xie et al., 2019; Gorelick et al., 2017).

Satellite data were obtained from February to July 2019 (following the trend of one-year NDVI) to better identify the growing patterns of the study area over a year (due to different agricultural management practices), and the mean pixel value of spectral bands with less than 20% cloud coverage was used for remote sensing-based covariance. Together with these topographic variables and remote sensing images, to evaluate parent soil materials and other soil-forming factors, we use geomorphological maps built based on a hierarchical geomorphological approach defined by Zink (1989). According to this approach, geomorphological units are divided into four levels: landscape, landform, lithology, and geomorphological surface (Zink, 1989). There are 19 geomorphological units in this research area (Ebrahimi et al., 2022). More information on the most crucial auxiliary variables collected from sentinel spectral data of the study region can be found in Table 3.

Table 3. Important auxiliary variables derived from sentinel spectral data of the study area

Index	Formula	Description	References
		Vegetation indices	
NDVI	NIR− Red /NIR+ Red	Crop monitoring and empirical studies	Rouse et al., 1973
SAVI	(1+L) (NIR− Red / NIR+ Red +L)	Improving the sensitivity of NDVI to soil backgrounds	Huete, 1994
OSAVI	NIR- Red/NIR+ Red+ X	Calculation of the aboveground biomass, leaf nitrogen content, and chlorophyll content	Rahim et al., 2016
MSAVI	$0.5\{2\cdot N\,IR + 1 - \sqrt{(2\cdot NIR + 1)}\ 2 - 8\,(NIR - Red)$	Reduction of bare soil influence on SAVI	Qi et al.,1994
		Parent material and soil indices	
Carbonate index	R/G	Carbonate response	Amen and Blaszczynski, 2001
Clay index	SWIR 1/SWIR 2	Clay response	Hengl, 2007
Coloration Index	(R-G)/(R+G)	Soil color	Ray et al., 2004

2.5. Geospatial data set preparation

Preparing a dataset of environmental covariates combined with soil data was done in this way: 1-converting polygon maps to raster, 2-resampling all auxiliary variables to the same raster grid of 30 × 30 m, 3-filtering out missing pixels, 4-building a stack of all environmental covariates, 5- intersecting the raster stack with the soil point observation to create a final geospatial dataset for predictive modeling.

2.6. Models used

Three machine learning algorithms (i.e., Multi-linear Regression (MLR), Decision Trees (DT), and Random Forests (RF)) and a set of auxiliary variables (i.e., geomorphology, terrain attributes, and remote sensing data) are used to find a more appropriate model for predicting SFI. These machine-learning models are used because of their successful application in earlier studies and their relatively good accuracy, robustness, and ease of use (Hounkpatin et al., 2022; Hu et al., 2020; Zeraatpisheh et al., 2019; Hengl et al., 2017; Bagheri Bodaghabadi et al., 2015). The entire model was developed using R software.

Prediction of the SFI spatial distribution based on environmental variables as independent variables is generated by a linear regression equation in the MLR model. In this study, we used the lm function of the R software to analyze the MLR. In addition, to identify the most important independent variables in this regression model, we used stepwise regression, which is a popular data mining tool, which uses statistical significance to explain the causal effect of independent variables on the dependent variables. The R software step function is used for this purpose. The second applied model in this study was DT one of the most important and oldest machine learning algorithms that are easy to interpret and quick to learn (Quinlan, 1986). In this study, we used the rpart function of the R software to analyze DT. The RF model as an extension of the regression tree model was the last model that is based on creating multiple classifications, or regression trees, and uses two levels of randomization for each tree in the forest. RF improves the accuracy of predictions and reduces model overfitting (Breiman, 2001). It can handle large amounts of both numerical and categorical data and is unaffected by missing data (Subramanian et al., 2022). The Random Forest package in the R environment is used to predict this model. To identify the most important environmental variables, the varimp function of this package was used.

2.7. SFI spatial prediction and validation

In this study, the performance of different models was assessed using K-fold cross-validation (Kohavi, 1995). The data set is separated into K roughly equal sets using this procedure. These sets include one set specifically for validation. The target variable was then predicted using the model for the reserved data points after it had been calibrated using information from the K-1 sets. Then, using this forecast, the prediction error is calculated. This procedure is carried out K times, with a different set being reserved for validation each time. In this manner, K estimates of the prediction error are obtained, one for each validation sample set (Yigini et al., 2018). With this method, data set calibration and validation depend less on a single random partition. All observations were used for calibration and validation by performing the training procedure k times, with each observation only being used once for validation.

2.8. Evaluation of model performances

The mean square error (RMSE), coefficient of determination (R2), and relative RMSE (%RMSE) were used to assess the efficacy of various estimation models. According to Park and Vlek (2002), the dimensionless % RMSE enables the comparison of the accuracy of variables of various types and volatility ranges.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} [O_i - P_i]^2}{n}} \quad (2)$$

$$R^2 = \left[\frac{\sum_{i=1}^{n} (O_i - O_{avg})(P_i - P_{avg})}{\sqrt{\sum_{i=1}^{n} (O_i - O_{avg})^2 (P_i - P_{avg})^2}} \right]^2 \quad (3)$$

$$\%RMSE = (RMSE/ O_{avg})*100 \quad (4)$$

where Pi, Oi, Oavg, Pavg, n, and p are the observed and predicted values of the soil properties at the i-th point, the mean of the observed and estimated values at that point, the number of observations, and the total number of explanatory variables in the model, respectively (Rahman et al., 2020).

To map the SFI in the study area, a more efficient model was finally used. Jenks (1967) developed the natural break method, which was used to assign different soil fertility classes more accurately.

3. Results and Discussion

Summary statistics of soil characteristics and also SFI index are presented in Table 4. In general, the results showed that 89% of samples have a pH of more than 7.5, 87% have an electrical conductivity of less than 1 dS m-1, 87% have organic carbon of more than 1%, 80% have available phosphorus less than 15 mg kg-1, and 74% have available potassium of more than 300 mg kg-1 soil. Also, 26, 14.3, and 84 percent of samples have available iron, manganese, and zinc, respectively, less than 7, 7, and 1 mg kg-1.

Calcium carbonate levels exceeding 10% are present in 97 percent of soil samples, which shows that it has a significant influence on the soil properties of the study area. Additionally, the research area has a reasonably high clay percentage, which indicates a heavy to relatively heavy soil texture. With a range of 0.75 to 52.33, the SFI levels show a significant variety of fertility throughout the research area.

Table 4. Descriptive statistical summary of the soil characteristics and calculated SFI

	Range	Minimum	Maximum	Mean	Std. Deviation
pH	1.04	6.94	7.98	7.61	0.15
EC	1.96	0.36	2.32	0.77	0.35
TNV	68.10	0.00	68.10	27.48	8.43
Om	12.46	0.86	13.33	2.85	1.70
N_{total}	0.72	0.05	0.77	0.16	0.09
P_{ava}	65.40	0.60	66.00	11.64	10.54
K_{ava}	1043.00	123.00	1166.00	370.28	127.12
Cu_{ava}	1.80	0.90	2.70	1.79	0.35
Zn_{ava}	7.00	0.24	7.24	0.76	0.71
Mn_{ava}	45.90	3.10	49.00	14.60	7.06
Fe_{ava}	29.70	2.10	31.80	10.78	5.14
Clay	27.00	28.00	55.00	42.82	5.57
SFI index	51.58	0.75	52.33	10.03	9.45

3.1. Auxiliary data used in predictive models

The results show that in machine learning models, some auxiliary data affect SFI predictions. In the MLR model based on the step function in the R software, the most important variable was the minimum curvature, lithology, clay index and MSAVI. The most important variables used in the DT model based on *rpart* function in R software were NDVI, OSAVI, MSAVI, clay index, elevation, longitudinal curvature, and carbonate index, respectively. MSAVI, NDVI, SAVI, carbonate index, color index, clay index, aspect and maximum curvature were the most important variables based on *varimp* function in the Random Forest package of R software. Figure 2 shows the result of *varimp* function for important environmental data of Random Forest model (Fig.2).

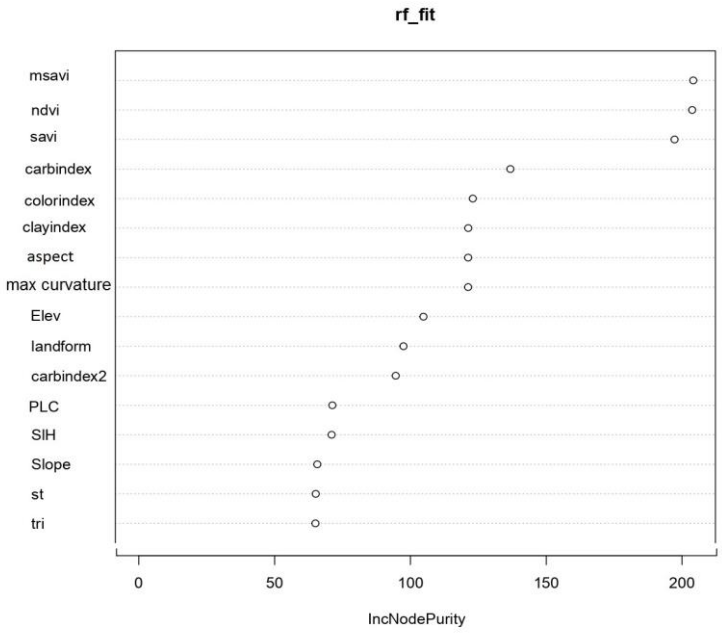

Fig.2. Important environmental data of Random Forest model

Our results demonstrated that spectral vegetation indices were useful predictors for mapping SFI in all used machine-learning models. This result is in line with other research that shows a significant correlation between these kinds of indices and crop yields (Dedeoğlu et al., 2020; Zhao et al., 2020; Thapa et al., 2019; Ferna et al., 2018; Xie et al., 2018).

The findings also indicated that parent material and soil indices (carbonate and clay indices) are additional helpful predictors of SFI in the study area. This finding may be related to the study area's surface soils' high levels of calcium carbonate and heavy soil texture and their effects on the decline in soil fertility. Long-term relationships between soil fertility and soil texture exist because soil porosity influences water holding capacity and water flow, which therefore have an impact on soil fertility (Upadhyay and Raghubanshi, 2020). Elevation and curvature were other important variables that influenced soil fertility. The relationship between terrain attributes and soil properties and thus soil fertility has been described in many studies (Ayele et al., 2019, Kokulan et al., 2018).

3.2. Accuracy of the prediction models

The results of comparing applied machine learning algorithms based on the mentioned accuracy measures for choosing a better SFI predictive model are shown in Table 5.

Table 5. Result of comparing applied machine learning algorithms

Machine learning model	RMSE	R^2	RMSE%
RF	4.31	0.86	0.43
DT	6.51	0.53	0.64
MLR	7.57	0.35	0.75
(SFI=15.2 - 218 min curvature - 0.95 lithology - 11.1 clay index +26.1 MSAVI)			

Notably, MLR performed poorly compared to the non-linear machine learning algorithm. Compared to RF and DT, the MLR model displays marginally worse results. It is important to note that, as compared to the DT model, the RF model performed the best at predicting SFI. This is due to solving multivariable matching problems because RF has combined multiple trees to form a vote allocation mechanism (Subramanian et al., 2022).

The good performance of RF is coherent with the previous studies (Hounkpatin et al., 2022; Hu et al., 2020; Zeraatpisheh et al., 2019; Hengl et al., 2017). RF is advised by Kampichler et al. (2010) due to its performance, modeling simplicity, and interoperability. Although the DT model is less adaptable than the RF model, it performs better predictions than the MLR model. Numerous scholars have noted the MLR's shortcomings in dealing with non-linear correlations between the target and predictive factors, particularly in heterogeneous landscapes (Subramanian et al., 2022; Forkuor et al., 2017; Selige et al., 2006). Scatter plots of observed versus predicted SFI by using different machine learning techniques are shown in Figure 3.

3.3. Spatial distribution of soil fertility

The digital soil map of surface SFI distribution by using a random forest model of the study area is shown in Figure 4. According to Jenks's (1967) natural break method, only a very small region (17.76 percent) had fertile soil, 55 percent of the overall area had poor fertility and 27.25 percent had moderate fertility soil (Table 6).

The fertility map reveals that the study area's northern and southernmost regions have the greatest concentrations of infertile soils. This is probably caused by the area's dense texture

and high calcium carbonate concentration. Additionally, the inadequate availability of plant nutrients like phosphorus and zinc in soils in low-fertility locations limits plant growth.

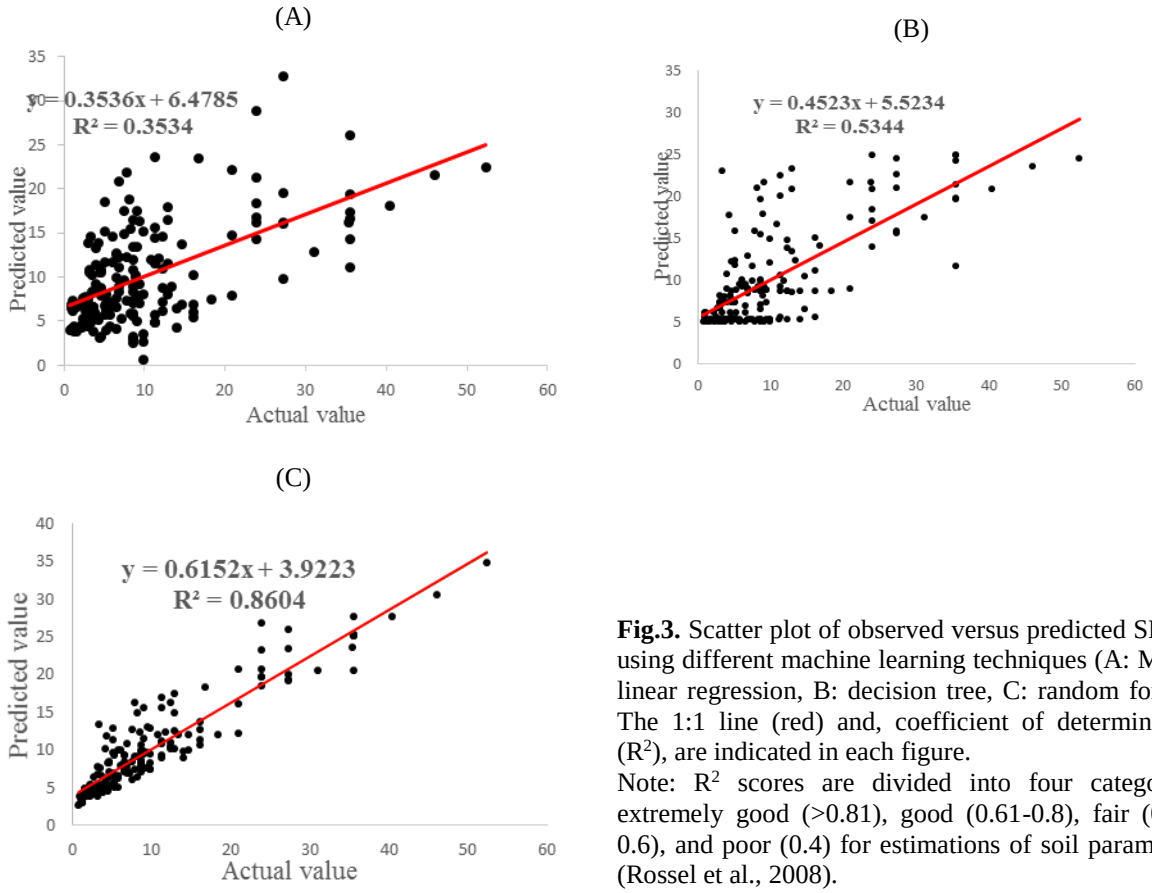

Fig.3. Scatter plot of observed versus predicted SFI by using different machine learning techniques (A: Multi-linear regression, B: decision tree, C: random forest). The 1:1 line (red) and, coefficient of determination (R^2), are indicated in each figure.
Note: R^2 scores are divided into four categories: extremely good (>0.81), good (0.61-0.8), fair (0.41-0.6), and poor (0.4) for estimations of soil parameters (Rossel et al., 2008).

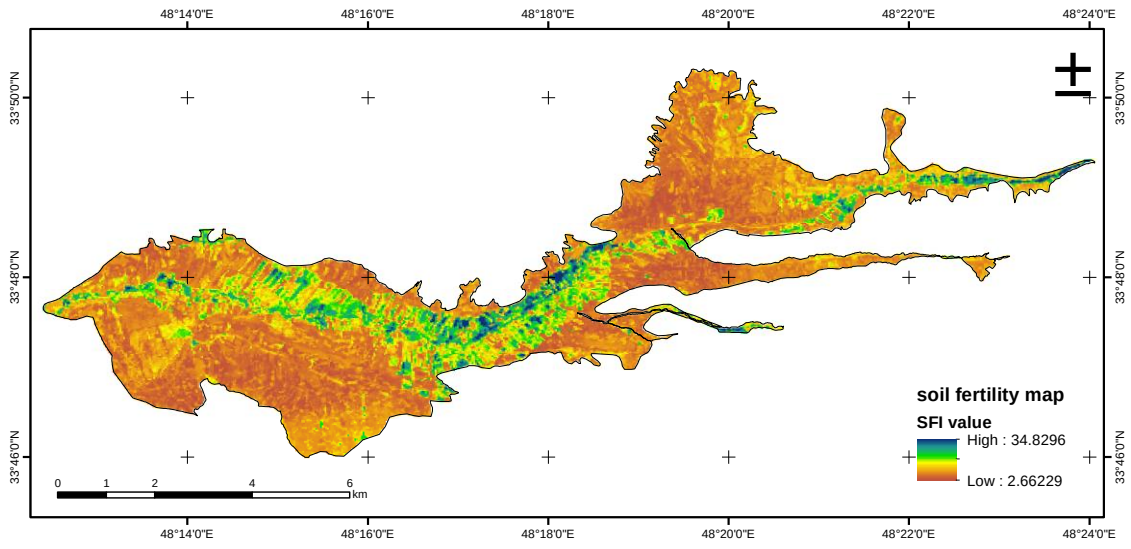

Fig. 4. Distribution map of surface SFI based on random forest model

Table 6. Classification of surface SFI at five levels, according to the natural break method developed by Jenks (1967)

SFI range	Fertility class	Area(hectare)	Percentage
21-34.8	I	95.23	2.51
16-20	II	249.04	6.57
12-15	III	329.06	8.68
8-11	IV	1033.65	27.25
2.66-7	V	2086.13	55.00

4. Conclusions

To map the spatial distribution of soil fertility, this study investigated the use of geographic information systems, digital soil mapping methods, spectral capabilities of Sentinel-2A imagery, and machine learning models, as well as laboratory-analyzed soil samples. Multiple linear regression, random forest, and decision tree were the three machine learning prediction models that were tested and compared. K-fold cross-validation was used to assess the effectiveness of each model.

In all employed machine learning models, the key factors affecting the spatial distribution of soil fertility were identified, and it was found that spectral vegetation indices, which reference yield as well as parent material indices, were effective predictors for mapping SFI.

The result of comparing applied machine learning algorithms shows the good performance of the RF model, due to the combination of multiple trees to form a vote allocation mechanism.

The soil fertility map categorization results show that the research area's fertility level was generally low. The heavy soil texture, high calcium carbonate concentration, and low availability of plant nutrients like phosphorus and zinc were the most important contributors to lowering soil fertility status. Therefore, specific soil fertility management practices are recommended based on limiting nutrients in those fields having inadequate levels of plant nutrients together with training farmers on the proper use of the appropriate soil fertility management practice.

It is advised that more studies be done to confirm the value of SFI in decision-making and implementation for various contexts and regions. Future research should also take into account how human activities, in addition to other physical and environmental aspects including soil parent material, and climatic, hydrological, and ecological elements, affect soil fertility status.

Declarations

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.

Conflict of Interest

The authors have no conflicts of interest to declare. All authors have seen and agree with the contents of the manuscript. We certify that the submission is original work and is not under review at any other publication.

Data Availability

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

References

Al Budeiri, M. H., AL-Aloosy, A. 2019. Development soil fertility map by geographic information system technology for al-souera. Iraqi J Agric Sci. 50:192-203.

Amen, A. and Blaszczynski, J., 2001. Integrated landscape analysis. US Department of the Interior, Bureau of Land Management, National Science and Technology Center.

Arif Özyazici, M., Dengiz, O., Sağlam, M., Erkoçak, A. and Türkmen, F., 2017. Mapping and assessment-based modeling of soil fertility differences in the central and eastern parts of the Black Sea region using GIS and geostatistical approaches. Arab J Geosci. 10:1-9.

Awasthi, N. and Bansal, A., 2017. Application of data mining classification techniques on soil data using R. Int. J. Comput. Sci. electron. 4(1): 33-37.

Ayele, G.T., Demissie, S.S., Jemberrie, M.A., Jeong, J. and Hamilton, D.P., 2019. Terrain effects on the spatial variability of soil physical and chemical propertie. Soil Sys. 4:1-1. https://doi.org/10.3390/soilsystems4010001.

Babalola, T. S., Fasina, A.S., Kediri, W., 2019. Fertility capability classification of soils in two agro ecological zones in the basement complex zone of Nigeria. Int J Res Stud Agric Sci. 5 (4):37-43.

Bagheri Bodaghabadi, M., Amini, A., Saleh, M. H., 2019. Suitability analysis and evaluation of pistachio orchard farming, using canonical multivariate analysis. Sci Hortic. 246: 528-534.

Bagheri Bodaghabadi M, José A, Martínez-Casasnovas M et al (2015) Digital soil mapping using artificial neural network (ANN) and digital elevation model (DEM) attributes. Pedosphere 25:580-591

Breiman L (2001) Random forests. Machine learning 45(1): 5-32.

Cao, H., Jia, M., Song, J., 2021. Rice-straw mat mulching improves the soil integrated fertility index of apple orchards on cinnamon soil and fluvo-aquic soil. Sci Hortic. 278, pp 109837.

DOI:10.1016/j.scienta.2020.109837.

Chen, G., Cao, L., Wang, G., 2009. Application of weighted spatially fuzzy dynamic clustering algorithm in evaluation of soil fertility. Sci Agric Sin. 42(10) pp 3559-3563.

Chen, W., Li, H., Hou, E., 2018. GIS-based groundwater potential analysis using novel ensemble weights-of-evidence with logistic regression and functional tree models. Sci Total Environ. 634:853-867. DOI: 10.1016/j.scitotenv.2018.04.055.

Dedeoğlu, M., Başayiğit, L., Yüksel, M. and Kaya, F., 2020. Assessment of the vegetation indices on Sentinel-2A images for predicting the soil productivity potential in Bursa, Turkey. Environ Monit Assess. 192(1):1-16. DOI:10.1007/s10661-019-7989-8.

Dharumarajan, S., Lalitha, M., Niranjana, K.V. and Hegde, R., 2022. Evaluation of digital soil mapping approach for predicting soil fertility parameters—a case study from Karnataka Plateau, India. Arab J Geosci15 (5): 1-21.

Ebrahimi Meymand, F., Ramezanpour, H., Yaghmaeian Mahabadi, N., & Eftekhari, K. 2022. Increasing the Homogeneity of Soil Map Units Using the Level of Landform Phase in the Geopedologic Approach. Water and Soil, 35(6), 890-873. doi: 10.22067/jsw.2021.72597.1094

Fayyaz, H., Yaghmaeian, N., Sabouri, A. and Shirinfekr, A., 2021. Assessing soil fertility index using Fuzzy-AHP and parametric methods for tea cultivation with different productivities. J of Agricultural Engineering Soil Sci and Agricultural Mechanization. J Agric Sci 44(3): 275-294. DOI: 10.22055/AGEN.2021.38284.1613.

Fern, R.R., Foxley, E.A., Bruno, A. and Morrison, M.L., 2018. Suitability of NDVI and OSAVI as estimators of green biomass and coverage in a semi-arid rangeland. Ecol Indic 94(1):16-21. https://doi.org/10.1016/j.ecolind.2018.06.029.

Forkuor, G., Hounkpatin, O.K., Welp, G. and Thiel, M., 2017. High resolution mapping of soil properties using remote sensing variables in south-western Burkina Faso: a comparison of machine learning and multiple linear regression models. PloS one 12(1) pp e0170478. https://doi.org/10.1371/journal.pone.0170478.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D. and Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens Environ 202: 18-27. https://doi.org/10.1016/j.rse.2017.06.031.

Haryuni, K., Widijanto, A., Supriyadi, H., 2020. Soil fertility index on various rice field management systems in central java, indonesia. Am J Agric Biol Sci. 15: 75-82. DOI: https://doi.org/10.3844/ajabssp.2020.75.82.

Hengl, T., 2007. A Practical Guide to Geostatistical Mapping of Environmental Variables. EUR 22904 EN. Luxembourg (Luxembourg): Office for Official Publications of the European Communities. JRC38153

Hengl, T., Leenaars, J.G., Shepherd, K.D., Walsh, M.G., Heuvelink, G.B., Mamo, T., Tilahun, H., Berkhout, E., Cooper, M., Fegraus, E. and Wheeler, I., 2017. Soil nutrient maps of Sub-Saharan Africa: assessment of soil nutrient content at 250 m spatial resolution using machine learning. Nutr Cycl Agroecosystems 109(1): 77-102.

Hounkpatin, K.O., Bossa, A.Y., Yira, Y., Igue, M.A. and Sinsin, B.A., 2022. Assessment of the soil fertility status in Benin (West Africa)–Digital soil mapping using machine learning. Geoderma Reg, 28, p.e00444. https://doi.org/10.1016/j.geodrs.2021.e00444.

https://earthengine.google.com.

https://earthexplorer.usgs.gov.

Hu, Y., Xu, X., Wu, F., Sun, Z., Xia, H., Meng, Q., Huang, W., Zhou, H., Gao, J., Li, W. and Peng, D., 2020. Estimating forest stock volume in Hunan Province, China, by integrating in situ plot data, Sentinel-2 images, and linear and machine learning regression models. Remote Sens, 12(1), p.186. https://doi.org/10.3390/rs12010186.

Huete, A. R., 1988. A soil-adjusted vegetation index (SAVI). Remote Sen Environ. 25(3): 295-309. https://doi.org/10.1016/0034-4257(88)90106-X.

Iticha, B., Takele, C., 2019. Digital soil mapping for site-specific management of soils. Geoderma. 351: 85-91. https://doi.org/10.1016/j.geoderma.2019.05.026.

Jackson, M. L., 1973. Soil chemical analysis. New Delhi: Prentice Hall of India, Pvt. Ltd.

Jenks, G. F., 1977. Optimal data classification for choropleth maps. Department of Geographiy, University of Kansas Occasional Paper.

Kampichler, C., Wieland, R., Calmé, S., Weissenberger, H. and Arriaga-Weiss, S., 2010. Classification in conservation biology: a comparison of five machine-learning methods. Ecolo Inform, 5(6), pp.441-450. https://doi.org/10.1016/j.ecoinf.2010.06.003.

Khadka, D., Lamichhane, S., Amgain, R., Joshi, S., Shree, P., Kamal, S.A.H. and Ghimire, N.H., 2019. Soil fertility assessment and mapping spatial distribution of Agricultural Research Station, Bijayanagar, Jumla, Nepal. Eurasian J Soil Sci, 8(3): 237-248. DOI:10.18393/ejss.566551.

Kohavi, R. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. 14th Int Joint Conference on Artificial Intelligence Montreal QC Canada. 14(2). pp 1137–1145.

Kokulan, V., Akinremi, O., Moulin, A.P. and Kumaragamage, D., 2018. Importance of terrain attributes in relation to the spatial distribution of soil properties at the micro scale: a case study. Can J Soil Sci 98(2): 292-305. https://doi.org/10.1139/cjss-2017-0128.

Kumar, R. M., Yamanur, A., Basavaraj, P., 2021. NDVI derived LAI model: A novel tool for crop monitoring. J Crop Weed. 17(1): 76-85. DOI:10.22271/09746315.2021.v17.i1.1408.

Lindsay, W. L., Norvell, W., 1978. Development of a DTPA soil test for zinc, iron, manganese, and copper. Soil Sci Soc Am J. 42(3): 421-428.

Merumba, M.S., Semu, E., Semoka, J.M. and Msanya, B.M., 2020. Soil Fertility Status in Bukoba, Missenyi and Biharamulo Districts in Kagera Region, Tanzania. Int J Appl Agric Sci 2020b 6(5): 96-117. DOI:10.11648/j.ijaas.20200605.12.

Mokarram, M., Shaygan, M. and Miliaresis, G.C., 2018. Balancing soil parameters and farmers budget by feature selection and ordered weighted averaging. Geogr Tech. 13(1): 73-84. DOI: 10.21163/GT_2018.131.08.

Moral, F.J., Rebollo, F.J. and Terrón, J.M., 2012. Analysis of soil fertility and its anomalies using an objective model. J Soil Sci Plant Nutr. 175(6): 912-919. https://doi.org/10.1002/jpln.201100361.

Nelson, D.A. and Sommers, L., 1983. Total carbon, organic carbon, and organic matter. Methods of soil analysis: Part 2 chemical and microbiological properties, 9, pp.539-579. https://doi.org/10.2134/agronmonogr9.2.2ed.c29.

Nguemezi, C., Tematio, P., Yemefack, M., Tsozue, D. and Silatsa, T.B.F., 2020. Soil quality and soil fertility status in major soil groups at the Tombel area, South-West Cameroon. Heliyon, 6(2), p.e03432. https://doi.org/10.1016/j.heliyon.2020.e03432.

Olsen, S. R., 1954. Estimation of available phosphorus in soils by extraction with sodium bicarbonate (No. 939). US Department of Agriculture.

Panday, D., Maharjan, B., Chalise, D., Shrestha, R.K. and Twanabasu, B., 2018. Digital soil mapping in the Bara district of Nepal using kriging tool in ArcGIS. PloS one, 13(10), p.e0206350. https://doi.org/10.1371/journal.pone.0206350.

Pant, J., Pant, P., Bhatt, A., Pant, H. and Pandey, N., 2019. Feature Selection towards Soil Classification in the context of Fertility classes using Machine Learning. Int J Eng Innov Technol. 8(12): 4000-4004. DOI: 10.35940/ijitee. L3487.1081219.

Park, S. J., Vlek, P. L., 2002. Environmental correlation of three-dimensional soil spatial variability: a comparison of three adaptive techniques. Geoderma. 109(1-2): 117-140. https://doi.org/10.1016/S0016-7061(02)00146-5.

Prado, R.B., Benites, V.D.M., Polidoro, J.C. and Gonçalves, C.E., 2012. Mapping soil fertility at different scales to support sustainable brazilian agriculture. Int J Agric Biol Eng .6(9): 137-145.

Qi, J., Chehbouni, A., Huete, A.R., Kerr, Y.H. and Sorooshian, S., 1994. A modified soil adjusted vegetation index. Remote Sens Environ 48(2): 119-126. https://doi.org/10.1016/0034-4257(94)90134-1

Quinlan, J. R., 1986. Induction of decision trees. Mach Learn. 1(1): 81-106.

Bin Abdul Rahim, H.R., Bin Lokman, M.Q., Harun, S.W., Hornyak, G.L., Sterckx, K., Mohammed, W.S. and Dutta, J., 2016. Applied light-side coupling with optimized spiral-patterned zinc oxide nanorod coatings for multiple optical channel alcohol vapor sensing. J of Nanophotonics. 10(3): 036009. DOI:10.1117/1.JNP.10.036009.

Rahman, M.M., Zhang, X., Ahmed, I., Iqbal, Z., Zeraatpisheh, M., Kanzaki, M. and Xu, M., 2020. Remote sensing-based mapping of senescent leaf C: N ratio in the sundarbans reserved forest using machine learning techniques. Remote Sens. 12(9): 1375. https://doi.org/10.3390/rs12091375.

Ray, S.S., Singh, J.P., Das, G. and Panigrahy, S., 2004. Use of high-resolution remote sensing data for generating site-specific soil management plan. Int. Arch. Photogramm. Remote Sens Spatial Inf Syst. 35(7): 127-131.

Rossel, R.V., Jeon, Y.S., Odeh, I.O.A. and McBratney, A.B., 2008. Using a legacy soil sample to develop a mid-IR spectral library. Soil Res. 46(1): 1-16. DOI:10.1071/SR07099.

Rouse, J.W., Haas, R.H., Schell, J.A. and Deering, D.W., 1974. Monitoring vegetation systems in the Great Plains with ERTS. NASA Spec. Publ, 351(1), p.309.

Sağlam, M. and Dengiz, O., 2014. Distribution and evaluation of soil fertility based on geostatistical approach in Bafra Deltaic Plain. Türkiye Tarımsal Araştırmalar Dergisi. 1(2): 186-195. DOI:10.19159/tutad.27089.

Selige, T., Böhner, J., Schmidhalter, U., 2006. High resolution topsoil mapping using hyperspectral image and field data in multivariate regression modeling procedures. Geoderma. 136(1-2):235-244.doi.org/10.1016/j.geoderma.2006.03.050.

Seo, B., Lee, J., Lee, K.D., Hong, S. and Kang, S., 2019. Improving remotely-sensed crop monitoring by NDVI-based crop phenology estimators for corn and soybeans in Iowa and Illinois, USA. Field crops Res. 238: 113-128. https://doi.org/10.1016/j.fcr.2019.03.015.

Shahbazi, K., Besharati, H., 2013. An overview of the fertility status of Iranian agricultural soils. Land Management J. 1(1): 1-15.

Sirsat, M.S., Cernadas, E., Fernández-Delgado, M. and Barro, S., 2018. Automatic prediction of village-wise soil fertility for several nutrients in India using a wide range of regression methods. Comput Electron Agric. 154: 120-133. https://doi.org/10.1016/j.compag.2018.08.003.

United States. Division of Soil Survey, 1993. Soil survey manual (No. 18). US Department of Agriculture.

Soil Survey Staff, 2014. Keys to Soil Taxonomy. 12th Edition, USDA-Natural Resources Conservation Service, Washington DC.

Taghizadeh-Mehrjardi, R., Minasny, B., Sarmadian, F. and Malone, B.P., 2014. Digital mapping of soil salinity in Ardakan region, central Iran. Geoderma. 213: 15-28. https://doi.org/10.1016/j.geoderma.2013.07.020.

Thapa, S., Rudd, J.C., Xue, Q., Bhandari, M., Reddy, S.K., Jessup, K.E., Liu, S., Devkota, R.N., Baker, J. and Baker, S., 2019. Use of NDVI for characterizing winter wheat response to water stress in a semi-arid environment. J Crop Improv. 33(5): 633-648.

Tunçay, T., Kılıç, Ş., Dedeoğlu, M., Dengiz, O., Başkan, O. and Bayramin, I., 2021. Assessing soil fertility index based on remote sensing and GIS techniques with field validation in a semiarid agricultural ecosystem. J Arid Environ. 190: 104525. https://doi.org/10.1016/j.jaridenv.2021.104525.

Walkley, A., Black, I. A., 1934. An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. Soil Sci. 37(1): 29-38.

Weiss, M., Jacob, F., Duveiller, G., 2020. Remote sensing for agricultural applications: A meta-review. Remote Sens Environ. 236: 111402.

Xie, Q., Dash, J., Huang, W., Peng, D., Qin, Q., Mortimer, H., Casa, R., Pignatti, S., Laneve, G., Pascucci, S. and Dong, Y., 2018. Vegetation indices combining the red and red-edge spectral information for leaf area index retrieval. IEEE J Sel Top Appl Earth Obs Remote Sens. 11(5): 1482-1493. DOI: 10.1109/JSTARS.2018.2813281.

Xie, S., Liu, L., Zhang, X., Yang, J., Chen, X. and Gao, Y., 2019. Automatic land-cover mapping using landsat time-series data based on google earth engine. Remote Sens. 11(24): 3023. https://doi.org/10.3390/rs11243023.

Yigini, Y., Olmedo, G.F., Reiter, S., Baritz, R., Viatkin, K. and Vargas, R., 2018. Soil organic carbon mapping: cookbook.

Zabihi, H., Alizadeh, M., Kibet Langat, P., Karami, M., Shahabi, H., Ahmad, A., Nor Said, M. and Lee, S., 2019. GIS multi-criteria analysis by ordered weighted averaging (OWA): toward an integrated citrus management strategy. Sustainability. 11(4): 1009. https://doi.org/10.3390/su11041009.

Zeraatpisheh, M., Ayoubi, S., Jafari, A., Tajik, S. and Finke, P., 2019. Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran. Geoderma, 338, pp.445-452. https://doi.org/10.1016/j.geoderma.2018.09.006.

Zhao, Y., Potgieter, A.B., Zhang, M., Wu, B. and Hammer, G.L., 2020. Predicting wheat yield at the field scale by combining high-resolution Sentinel-2 satellite imagery and crop modelling. Remote Sens. 12(6): 1024. https://doi.org/10.3390/rs12061024.

Zinck, J. A., 1989. Physiography and soils. Lecture notes for soil students. Soil Sci Division. Soil survey courses subject matter: K6 ITC Enschede, the Netherlands