University of Tehran Press

DESERT

Online ISSN: 2345-475X

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

Modeling the Spatial Distribution of Sand, Silt, and Clay Particles Based on Global Soil Map and Limited Data

Leila Lotfollahi[1], Mohammad Amir Delavar[1]*, Asim Biswas[2], Mohammad Jamshidi[3], Ruhollah Taghizadeh Mehrjardi[4], Shahrokh Fatehi[5]

[1] Department of Soil Science, Faculty of Agriculture, University of Zanjan, Zanjan, Iran.
E-mail: amir-delavar@znu.ac.ir
[2] School of Environmental Sciences, University of Guelph, Ontario N1G2W1, Canada
[3] Scientific Staff of Soil and Water Research Institute (SWRI), Agricultural Research, Education and Extension Organization (AREEO), Karaj, Iran
[4] Department of Geosciences, University of Tübingen, Rümelinstr. 19-23, Tübingen, Germany
[5] Soil and Water Research Department, Kermanshah Agricultural and Natural Resources Research and Education Center, AREEO, Kermanshah, Iran

Article Info.

Article type:
Research Article

Article history:
Received: 25 Mar. 2023
Received in revised from: 27 Aug. 2023
Accepted: 21 Oct. 2023
Published online: 27 Dec. 2023

Keywords:
Arid and semiarid,
Digital Soil Mapping,
Random Forest,
Decision Tree.

ABSTRACT

Many regions of Iran lack digital map of soil properties. The Chahardowli plain in western Iran is one of these areas. Due to the importance of sand, silt, and clay components, having quantitative and continuous data on abrupt changes in these two properties in this area is very critical. Therefore, to study sand, silt, and clay, samples were taken at depths of 0–5, 5–15, 15–30, 30–60, and 60–100 cm, according to GlobalSoilMap. Finally, 145 samples were collected from 30 soil profiles. The significant covariates were selected by Random Forest Recursive Feature Elimination (RF-RFE). Relationships between these characteristics and environmental predictors were modeled using random forest (RF), decision tree (DT), and multiple linear regression (MLR) models. The accuracy and precision of the models used for all three particles showed that the RF model had the most accurate prediction with R^2 and RMSE of 0.82 and 2.34 for clay, 0.80 and 3.87 for sand, and 0.85 and 2.89 for silt, respectively. In this study, terrain-based variables had a greater impact on improving accuracy than remote-sensing variables. The current study showed that even with limited information, digital mapping of sand, silt, and clay particles under GlobalSoilMap and the use of environmental factors can provide acceptable results.

Cite this article: Lotfollahi, L., Delavar, M.A., Biswas, A., Jamshidi, M., Taghizadeh-Mehrjardi, R., Fatehi, SH. (2023). Modeling the Spatial Distribution of Sand, Silt, and Clay Particles Based on GlobalSoilMap and Limited Data. DESERT, 28 (2), DOI: 10.22059/jdesert.2023.95533

DOI: 10.22059/jdesert.2023.95533

Publisher: University of Tehran Press

1. Introduction

Having precise information on the spatial variability of agricultural soil properties is vital (Dharumarajan *et al.*, 2022; Kaya *et al.*, 2022). Therefore, it is needed to depict the spatial changes of soils in such a way as to provide acceptable maps of the soil. We still lack soil mapping information for many parts of the country. Due to the importance of sand, silt, and clay components, having quantitative and continuous data on vertical changes is very critical (Mirzaeitalarposhti *et al.*, 2022). Sand, silt, and clay components are some of the most essential physical features of soil, expressed as a relative percentage of sand, silt, and clay (Wu *et al.*, 2018; Jeihouni *et al.*, 2020; Keshavarzi *et al.*, 2022). The relative percentage of soil particles affects maintaining soil moisture, cation exchange capacity, soil water movement, the distribution of plant species, dissolved chemicals in soil, crop yield, and nutrient equilibrium in the rhizosphere (Zhang and Shi, 2019; Dharumarajan and Hegde, 2020; Taghizadeh-Mehrjardi *et al.*, 2020; Chen *et al.*, 2021).

Clear or sudden changes in soil particle size percentage affect soil permeability and drainage, and thus ecosystem performance (Mehrabi-Gohari *et al.*, 2019). Sand, silt, and clay components directly affects organic matter accumulation and water movement in the soil (Hook and Burke, 2000). Sand, silt, and clay components, moisture content, and the amount of water available to evaporate from the earth surface, resulting in crop growth and production, were negatively impacted. So, acceptable information from sand, silt, and clay components data is a crucial tool for predicting crop yield (Kaya *et al.*, 2022). The Chahardoli plain in western Iran is one of these areas. There is no information available on sand, silt, and clay components area and spatial distribution. Decreases in water resources and drought have led to limited use of water in agricultural regions. As a result, hydrological studies, crop cultivation spatial planning, and the design of appropriate management methods, as well as engineering work such as land consolidation, drainage management, erosion management, and irrigation systems in the region, necessitate a thorough examination of the spatial diversity of sand, silt, and clay components (Dharumarajan and Hegde, 2020; Keshavarzi *et al.*, 2022).

Sand, silt, and clay components cannot be changed quickly or by agriculture. Furthermore, traditional procedures such as pipette or hydrometer methods require a significant number of samples to determine the proportion of sand, silt, and clay components particle size. On the other hand, sand, silt, and clay components measurement costs are high. Therefore, digital soil mapping methods have overcome this problem to obtain the spatial distribution of soil particles (Dharumarajan and Hegde, 2020). In this approach, soil maps are predicted by a set of available auxiliary variables such as altitude derivatives, climate derivatives, vegetation, and parent materials (Mehrabi-Gohari *et al.*, 2019; Dharumarajan and Hegde, 2020). Different soils have different proportions of clay, silt, and sand (Chen *et al.*, 2021). Plants' water-holding capacity varies according to sand, silt, and clay components. Clay soils typically have a higher percentage of smaller pores and a higher water-holding capacity at low potentials, and they are often associated with poor drainage conditions or limited aeration for plant growth. In dry climates, sandy soils, on the other hand, have comparatively wide pores with a lower water retention capacity (Keshavarzi *et al.*, 2022).

Hengl *et al.* used computerized soil mapping methods to estimate sand, silt, and clay components in their research (Hengl *et al.*, 2015). In a study to predict soil properties, the RF model was better than the multiple linear regression model (Zhang and Shi, 2019). In a report using digital soil mapping methods, African sand amounts were measured (Vagen *et al.*, 2016). In Africa, sand, silt, and clay were estimated through digital mapping with a precision of 1 km (Hengl *et al.*, 2014). Many models were examined to predict sand, silt, and clay components. Research results show that the models perform differently under different conditions (Brungard

et al., 2015; Taghizadeh-Mehrjardi *et al.*, 2016). A RF, which eliminates the recursive feature, is the most accurate or best technique, according to Brungard *et al.* (2015). Jafari *et al.* (2012) predicted sand, silt, and clay component classes using tree regression in an arid region.

In one investigation, soil surface sand, silt, and clay components were determined by multiple linear regression, conventional kriging, basic kriging, and global kriging (Mondejar and Tongco, 2019). The surface horizon sand, silt, and clay components were predicted through MLR (Keshavarzi *et al.*, 2022; Schulz *et al.*, 2023). The RF was used to create a map of sand, silt, and clay component classes for India on a scale of 1:250000 (Dharamarajan and Hegde, 2020). This study examines the limited data on the spatial prediction of three particles of sand, silt, and clay using the samples taken according to GlobalSoilMap. This study is done with RF machine learning algorithms, DT, and MLR. The following are the specific goals of this study:

A) Investigate the relationship between sand, silt, and clay components with environmental variables.

B) Validation of sand, silt, and clay component predictions with 10- fold cross validation.

C) Choosing the best model to predict the spatial distribution of sand, silt, and clay components in the region.

2. Material and methods

The study steps are shown in Fig. 1. In the first step, sampling was performed at the five studied depths according to GlobalSoilMap. The sand, silt, and clay components were measured. Auxiliary data was gathered using digital elevation model (30 m) derivatives and Landsat images (June 2019) in the second step. The significance of variables was examined with Random Forest Recursive Feature Elimination (RF-RFE) (Darst *et al.*, 2018). The fourth phase involved predictions from the RF, DT, and MLR models for the components of sand, silt, and clay. The 5-fold cross-validation approach was used in the fifth phase to assess the prediction models' accuracy, and the 6-step process produced a prediction map of the components of clay, silt, and sand based on the smallest error.

2.1. Site description

The study area is located in the foothills of the Zagros Mountains in western Iran. This region is part of the Chahardowli plain in Kurdistan Province. The lowlands include terraces, newly formed low-level piedmont fans, and valley terrace deposits. Based on the Lambert coordinate system, the plain is located at longitudes 1442000E to 1460000E and latitudes 1275000N to 1290000N, and the area is about 104 km^2 (Fig. 2). The slope of the area decreases from west to east by 268 m (slope average 7%). The highest point is 2051 m, and the lowest point is 1783 m above sea level. The mean annual rainfall is 332 mm. The highest and lowest precipitation observed in November and August, respectively. The average annual temperature is 23 °C. The mean monthly temperature for the study range for the warmest month of the year (August) is 36.4 °C, and for the coldest month of the year (January), it is -3 °C. The highest relative humidity is 77% in February, and the lowest is 35% in August. The area has a semi-arid climate (based on Köppen climate classification method). Soil moisture and temperature regimes in this area are xeric and mesic, respectively.

One of the prominent features of the study area is the presence of temporary and perennial river channels. Surface water flow generally flows from the western heights to the center and from the center to the north. Water flows along the Shurchai River in the north and flows into one of the Sefid-Rud river branches. The studied area is relatively flat, and geomorphological changes are not sharp. According to Mahler's classification (1970), the region's physiographic units include lowlands and sloping plains. The region's main uses include agriculture and

rangeland. In this area are cultivated crops such as wheat, barley, potatoes, chickpeas, and summer crops. Mollisols, Entisols, and Inceptisols are the soil order of the region. The Mollisols order (mostly Typic Calcixerolls) was observed in the southern part of the plain because this part is less affected by river activities and has not been cultivated. The Inceptisols order (mainly Typic Calcixerepts) was observed in the central part of the plain, where river activity is dynamic and the area is under cultivation. The Entisols (Lithic Xerorthents) were observed in the highlands (Soil and Water Research Institute, 1995).

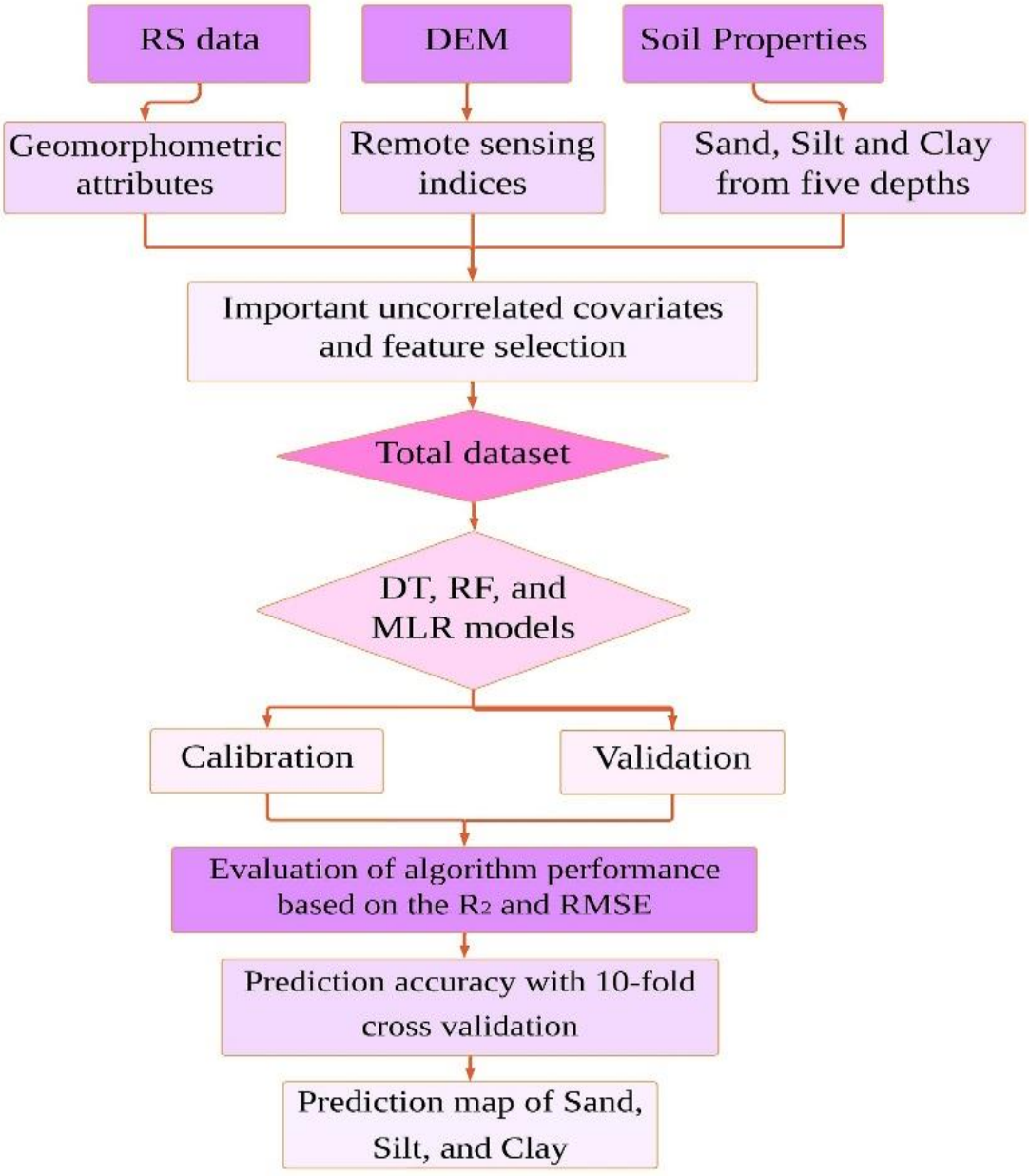

Fig. 1. The Methodological framework of the study.
(RS: remote sensing; DEM: digital elevation model; R^2: correlation coefficient; RMSE: root mean square error; RF: random forest; MLR: Multiple Linear Regression; DT: decision tree).

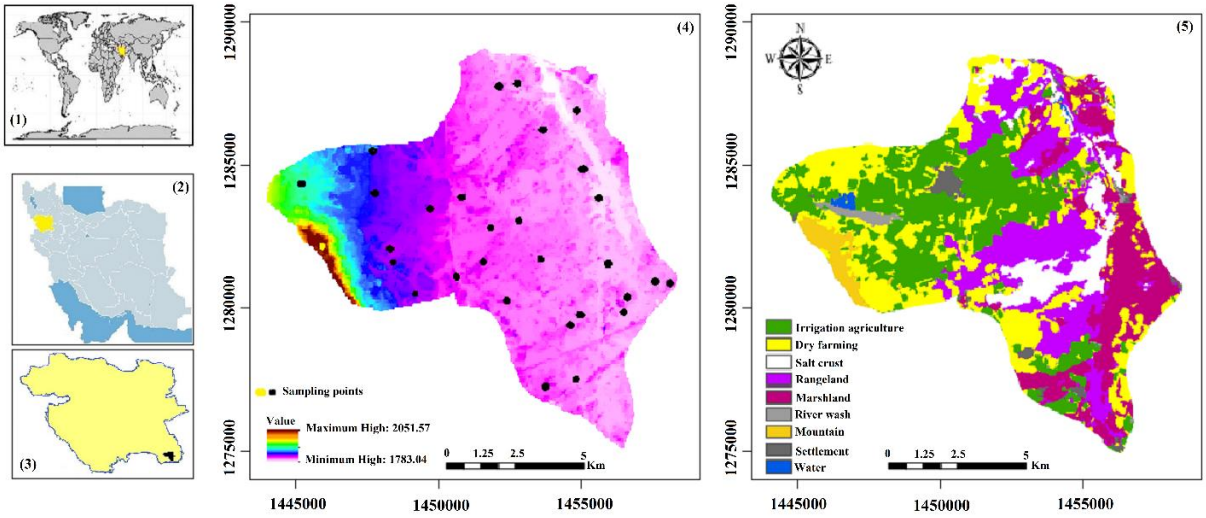

Fig. 2. The location of the study area in Iran and the spatial distribution of soil profiles and the land cover map.

2.2. Soil samples and analysis

In the second step, sampling of points and laboratory analysis was performed. Predominant suborder groups in the plain are Typic Calcixerepts, Typic Calcixerolls, and Aquic Haploxerolls, and in the highland are Lithic Xerorthents (Soil and Water Research Institute, 1995). Soil data quality depends on sampling and laboratory analysis methods. The sampling method involves collecting representative samples from the study area. It provides valid input for creating a prediction model with environmental variables (Biswas and Zhang, 2018). Study points were obtained by conditioned Latin hypercube sampling (cLHS) and the cLHS package in R software (Minasny and McBratney, 2006). CLHS is a random sampling method with the most classifications that ensures complete coverage of variable distribution. This has been developed to select an efficient set of values for the input variables in the development of computer models (Biswas and Zhang, 2018; Jamshidi *et al.*, 2019).

The first step, the selection of the number of points (15, 45, 75, 105, and 135 points), was done in 15,000 replications. The second step was to more accurately determine the number of points taken (15, 30, 45, 60, 75, 90, 105, 120, and 135) with a total of 15000 replications to ensure the optimal solution to minimize the objective function. Since 30% of these points were training points, training points were added to testing points. This resulted in 32 optimal points for the region. According to GlobalSoilMap, the profiles were sampled at depths of 0-5, 5-15, 15-30, 30-60, and 60-100 cm (Hartemink *et al.*, 2010). A total of 145 soil samples were collected from the 30 available profiles. After transfer, all samples were air-dried at 22 °C, the soil was passed through a 2 mm sieve, and then soil sand, silt, and clay components were measured (Sparks *et al.*, 2020).

2.3. Environmental data

Parent material, time, climate, relief, and vegetation indices are among the input variables based on the SCORPAN model (McBratney *et al.*, 2003). Because the area is small and it is not possible to show the effect of temperature and precipitation changes on soil variety, the climate index was not applied. The geological map for the area was large scale (1:50000), and the study area had a small scale, and the geological changes were not significant according to the existing map. Given that more than 95% of the study area has similar parent material and little effect on

soil variability, the parent material index was not considered. Satellite images were used to compute the vegetation index and soil-related indices. Landsat 8 images with cloud cover of less than 10% were selected, including bands 1 (0.433–0.45345 m), band 2 (0.450–0.515.5 m), band 3 (0.525–0.600 m), band 4 (0.630–0.680 m), band 5 (0.845–0.885 m), band 6 (1.560–1.660 m), and band 7 (2.1–2.3 m) (with a pixel size of 30 m). Landsat 8 images are used to provide a land cover map.

The date of sampling (June 2019) was followed while downloading images. Atmospheric and radiometric corrections were made in ENVI software (Version, 5.6). Land cover units were determined using supervised classification (maximum likelihood with an overall accuracy of 90.35 and a Kappa coefficient of 0.87). Finally, nine land cover classes were extracted: irrigation agriculture, dry farming, salt crust, rangeland, marshland, river wash, mountain, settlement, and water (Fig. 2). The DEM (30 m) was downloaded from the ASTER database (ASTER GDEM V2, 30 m), and the necessary preprocessing (Planchon and Darboux, 2002) was performed in SAGA-GIS (Version: 5.0.0). Relief derivatives were selected based on relief analysis methods (Wilson and Gallant, 2000) using SAGA software (Brenning, 2008; Olaya and Conrad, 2009). The DEM derivatives (SAGA Development Team, 2011), vegetation indices (Silleos *et al.*, 2006), and soil-related indices (Taghizadeh-Mehrjardi *et al.*, 2014) are presented below. Ten vegetation indices and 24 DEM indices were computed in SAGA-GIS software (version 5.0.0). Six soil-related indexes were extracted in ArcGIS (Ver. 10.5).

Ten vegetation indices were extracted using SAGA-GIS software (Version: 5.0.0), including corrected transformed vegetation index (CTVI), difference vegetation index (DVI), normalized difference vegetation index (NDVI), normalized ratio vegetation index (NRVI), perpendicular vegetation index (PVI), ratio vegetation index (RVI), soil-adjusted vegetation index (SAVI), transformed soil-adjusted vegetation index (TSAVI), Thiam's transformed vegetation index (TTVI), and transformed vegetation index (TVI). The 23 indices of DEM were obtained using SAGA-GIS software, such as analytical hill shading (AH), channel network base level (CNBL), convergence index (CVI), flow accumulation (FA), longitudinal curvature (LC), aspect (A), slope (S), topographic position index (TPI), topographic wetness index (TWI), vertical distance to channel networks (VDCN), slope length (SL), multiresolution index of valley bottom (MIVB), LS factor (LSF), brightness index (BI), valley depth (VD), longitudinal curvature (LC), catchment area (CA), relative slope position (RSP), slope height (SH), general curvature (GC), stream power index (SPI), and profile curvature (PC). Six soil-related indexes were extracted with ArcGIS (Ver. 10.5), including carbonate index (CI), clay index (CLI), global vegetation moisture index (GVMI), gypsum index (GI), salinity index (SI), and salinity ratio index (SRI). The land cover map was made with Landsat 8 images.

2.4. Correlation analysis and variable reduction

There are many variables in modeling, but only a few have the required information. Variable selection to obtain accurate predictions is essential. In order to provide accurate results, a significant analysis of variables is performed (Brungard *et al.*, 2015; Rumao, 2019). In this study, the significance of variables was examined using Random-Forest-Recursive Feature Elimination (RF-RFE) based on a random forest algorithm. RF-RFE is a wrapper algorithm. This algorithm examines all variables and ranks them. It predicts their importance based on performance (Development Core Team, R, 2016; Campos *et al.*, 2018). The presence of correlated variables affects the model ability to identify the correct variables. The RF-RFE algorithm reduces problems into smaller data sets (Darst *et al.*, 2018). The mean square error (%IncMSE) and node purity are the two most critical parameters in this technique (IncNodePurity). IncMSE is a report that shows the decrease in accuracy or how the prediction

worsens when changing the value variable. In other words, it represents an increase in mean squared error (MSE). That is calculated from the replacement of test data. MSE reports prediction error analysis for each tree. The final importance value is obtained by normalization with a standard error and a formula. That σδbj is the standard deviation. A variable with a higher IncMSE is more critical (Dewi and Chen, 2019), which is the average overall tree (B) of the forest where variable J has been used.

$$\%IncMSE = \frac{\bar{\delta}_{bj}}{\sigma_{\delta_{bj}}/\sqrt{B}} \quad (1)$$

Δbj is obtained based on the following equation.

$$\bar{\delta}_{bj} = \frac{1}{B}\sum_{b=1}^{B}\left(MSE - MSE_{permuted_j}\right) = \frac{1}{B}\sum_{b=1}^{B}\delta_{bj} \quad (2)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (3)$$

The second critical criterion is IncNodePurity, which is related to lost performance and is chosen by the most effective divisions. More useful variables further increase the purity of the node by finding a division that has a significant variance between nodes and a slight variance within the node (Dewi and Chen, 2019).

2.5. Machine Learning Techniques

In this step, the modeling process was performed with selected variables (step 4). We used the variables selected in the above method as a framework for all modeling methods. For modeling, the relationship between soil properties and environmental variables was done through three machine learning models, including RF, DT, and MLR. The following is a brief description of each model: The RF model includes a set of classification and regression (CART) trees that utilize these trees to predict response variables. RF use binary reverse-division trees (Breiman, 2001). The Ranger package was used to adjust the model parameters (Wright and Ziegler, 2015). The optimal number of trees was analyzed from 100 to 1000 trees, with an increase in 100 tree size. The optimal value was found with the lowest error rate in 700 trees. Decision tree models are used for describing, classifying, and generalizing data. The decision tree model consists of zero or more internal nodes and one or more leaf nodes. All of these nodes have branches that test feature expression (Murthy, 1998). Linear models, such as MLR, have been proven to generate a stable model with low variance and high bias and thus run the risk of under-fitting the model (Heung *et al.*, 2016).

2.6. Evaluation of Algorithm Performance

Evaluation of model performance is an essential part of every experiment and crucial when comparing various models. Root mean square error (RMSE) is an objective criterion for comparing model performance and variations. RMSE measures the error rate between two data sets using the following equation: A model with a smaller root mean square error value is more accurate at estimating soil properties. RMSE is defined as Obs is the observed soil property, n is the number of observations, and pred is the projected soil attribute from a given model (Rumao, 2019).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(obs_i - pred_i)^2} \quad (4)$$

Obsi=Actual data point (observation).
Predi=data point predicted.
n=Total number of data points in the test data set.
Bias is called the average prediction error and is defined as follows.

$$Bias = \frac{1}{n}\sum_{i=1}^{n}(obs_i - pred_i) \quad (5)$$

R^2 is the square of the sample correlation coefficient (Pearson) and measures the precision of the relationship between the observations and the relevant predictions. When applied to observed and expected values, the Pearson correlation coefficient r is defined as follows:

$$r = \frac{\sum_{i=1}^{n}(obs_i - \overline{obs}) - (pred_i - \overline{pred})}{\sqrt{\sum_{i=1}^{n}(obs_i - \overline{obs})^2}\sqrt{\sum_{i=1}^{n}(pred_i - \overline{pred})^2}} \quad (6)$$

Concordance, or more technically, the Lin concordance correlation coefficient (Lawrence and Lin, 1989), assesses the relationship between correctness and precision. Along a 45-degree line, it is called fit quality. As a result, it is likely to be a more helpful statistic than R^2 alone. The term "concordance" is defined as follows:

$$\rho_c = \frac{2\rho\sigma_{pred}\sigma_{obs}}{\sigma_{pred}^2 + \sigma_{obs}^2 + (\mu_{pred} - \mu_{obs})^2} \quad (7)$$

Where μpred and μobs are the means of predicted and observed values, respectively. The σpred and σobs are the corresponding variances. The ρ is the correlation coefficient between predictions and observations.

2.7. Prediction accuracy

To evaluate the performance of the ML models, we applied the 10-fold cross-validation method. Ten-fold cross-validation is a resampling technique in which all data are randomly divided into ten equal folds; one fold is put aside for validation at each run, while the remaining 10-1 folds are utilized for calibration. The final accuracy is then calculated using the average accuracy of all folds (Mirzaeitalarposhti *et al.*, 2022). To quantify the error, common indices were calculated, including the root mean square error (RMSE), mean error (ME), coefficient of determination (R^2), and line concordance coefficient (Pc).

3. Results and discussion

3.1. The statistical description of sand, silt, and clay

Table 1 shows sand, silt, and clay contents. Sand amounts vary from 64 to 91%. The highest amount of sand is in layer 0–5 in rangeland and highlands and at the exit of a water channel. The lowest amount is in layer 30–60 cm in marshland. The coefficient of change in the sand does not change substantially with increasing depth. In the study of sand results, it was observed that by increasing the depth to a 30–60 cm layer, the average amount of sand decreases. After this layer, the average amount of sand increases.

Silt results showed that with increasing depth, silt average amounts decreased. The smallest and greatest mean silt values are 3 and 21%, respectively. The highest and lowest average silt values were at 0-5 cm and 60–100 cm, respectively. The highest and lowest surface layer silt were observed in wetlands and areas with severe surface erosion, respectively. The silt coefficient of change is enhanced with increasing depth. With increasing depth, silt is different from numerous land cover. Surface silt differs in agricultural regions.

The clay results showed that the average amount of clay increased up to a 30–60 cm layer, and after this layer, the average amount of clay decreased. Overall, the highest and lowest average amounts of clay ranged from 30 to 60 cm, and the lowest average amounts varied from 0 to 5 cm, respectively. The amount of clay varied from 2 to 23%. The lowest coefficient of variation in the area is related to the percentage of sand, and the highest coefficient of variation is related to the percentage of clay. The Anderson-Darling normality test was performed to test the normality of the data, which showed a normal distribution. Table 1 shows skewness and kurtosis results that confirm normality accuracy. 50 percent of observations in the middle quartile are less than or equal, and 50 percent are significant.

Table 1. Descriptive statistics of sand, silt, and clay components.

Soil properties (%)	Depth (cm)	Min	1st Qu	Median	Mean	3rdQu	Max	Std. D	CV (%)	SKEW	KURT
Sand	0-5	65.0	72.2	75.0	75.4	79.8	85.0	5.84	0.07	0.26	1.38
	5-15	66.0	71.0	75.0	74.9	78.8	85.0	5.56	0.07	-0.03	1.37
	15-30	64.0	68.0	73.0	73.0	78.0	85.0	5.74	0.07	0.0	1.15
	30-60	64.0	68.0	71.5	72.9	76.2	91.0	6.49	0.08	0.15	0.93
	60-100	65.0	71.0	73.0	75.0	80.0	86.0	6.47	0.08	0.55	1.40
Silt	0-5	8.0	12.0	14.5	14.3	16.0	20.0	3.17	0.22	-0.25	1.09
	5-15	6.0	13.0	13.5	13.8	15.0	20.0	3.13	0.22	0.5	2.31
	15-30	7.0	12.0	13.5	13.6	16.0	21.0	2.86	0.21	0.25	1.12
	30-60	5.0	11.0	13.0	13.0	14.6	19.0	3.06	0.23	-0.10	1.36
	60-100	5.5	9.4	11.0	11.4	13.0	17.0	2.80	0.24	0.1	1.00
Clay	0-5	3.0	6.2	10.0	10.3	12.0	19.0	4.41	0.43	-0.30	1.26
	5-15	2.0	7.5	12.0	12.0	15.8	22.0	5.11	0.43	-0.09	1.0
	15-30	4.0	10.5	13.0	13.5	17.0	22.0	4.53	0.34	0.23	1.07
	30-60	4.0	11.8	15.0	14.1	17.2	22.0	4.93	0.35	-0.18	1.60
	60-100	5.0	10.2	15.0	13.8	17.5	23.0	4.94	0.36	-0.31	1.05

Minimum; Min, Lower quartile; 1st Qu, Middle quartile; Median, Mean, Upper quartile; 3rd Qu, Maximum; Max, Standard deviation; Std. D, Coefficient of variation; CV, Skewness; SKEW, kurtosis; KURT.

3.2. Abundance distribution of clay, sand, and silt value

The frequency distribution diagram of all three particles is shown in the Fig 3. The amount of sand particles compared to the other two particles is the highest in five depths and four studied land cover. The amount of clay and silt is almost equal to each other (Fig. 3).

3.3. Selected auxiliary data and relationship between selected variables and sand, silt, and clay

The variables of each layer for each component were selected by the RF-RFE method. Using a specific variable in different layers, the percentage importance of that variable compared to all other variables has been calculated. Fig. 4 shows the importance of auxiliary variables. The variables used for each layer are shown in Table 3. TTVI, CI, A, and CNBL were important variables for identifying sand. The A was a crucial variable in identifying silt, and for clay, CNBL, RVI, SRI, and LSF were the critical variables, respectively (Table. 2). The critical variables used in the sand study have a larger number than silt and clay particles (Fig. 4).

3.4. Effective variables in the spatial prediction of sand, silt, and clay

Clay variables include five terrain-based variables and six remote-sensing variables. Based on the spatial distribution of soil clay, the CNBL, LSF, RVI, and SRI variables had the highest percentage. The CNBL variable shows the effect of height and moisture accumulation on the

amount of clay in the soil (Keshavarzi *et al.*, 2022; Schulz *et al.*, 2023). Several studies have shown that variations in sand, silt, and clay components are closely related to topography (Dharumarajan and Hegde, 2020; Keshavarzi *et al.*, 2022). In one study, the auxiliary data used to predict clay were primarily terrain-based factors (Mahmoudabadi *et al.*, 2017). In southeastern Nigeria, they found that smaller particle sizes (clay and silt) correlated with slope and topographic index, while larger particles (sand) correlated better with area and flow strength index (Wu *et al.*, 2018). According to a study, channel networks play a crucial role in sand, silt, and clay components variations in the region (Pahlavan-Rad and Akbarimoghaddam, 2018). Second place is A and VDCN. These variables are responsible for 9.52% of the variance in clay spatial distribution in the region.

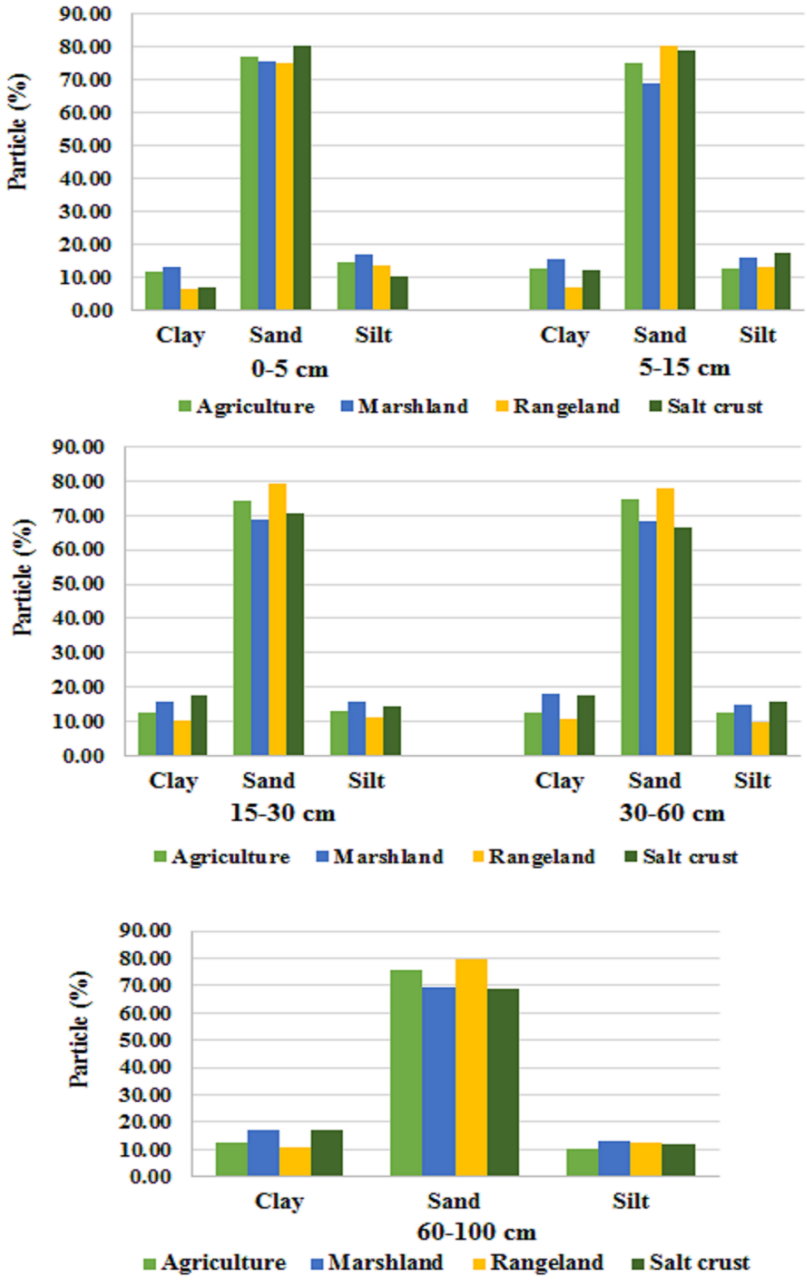

Fig.3. Chart of abundance distribution of clay, sand, and silt value at standard depths and different land cover.

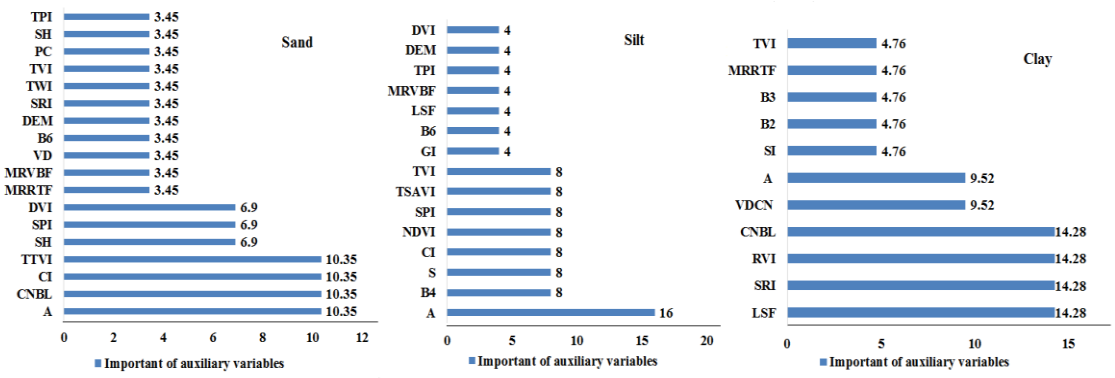

Fig. 4. Important auxiliary variables for each property.

Table 2. Variables used in each layer.

Layer	Sand	Silt	Clay
0-5	A, CNBL, CI, SH, SPI, TTVI	B4,S,A,GI	VDCN, A, RVI, SI
5-15	CNBL, CI, MRRTF, MRVBF, VD	B6, A, LSF, MRVBF, TPI, B4, CI	B2, B3, A, CNBL, MRRTF
15-30	A, CNBL, CI, SH, SPI, TTVI	A, NDVI, SPI, TSAVI, TVI	TVI, CNBL, LSF, SRI
30-60	B6, A, DEM, DVI, SRI, TWI	A, NDVI, SPI, TSAVI, TVI	VDCN, RVI, SRI, LSF, CNBL
60-100	TVI, DVI, PC, SH, TTVI, TPI	S, CI, DEM, DVI	RVI, SRI, LSF, CNBL

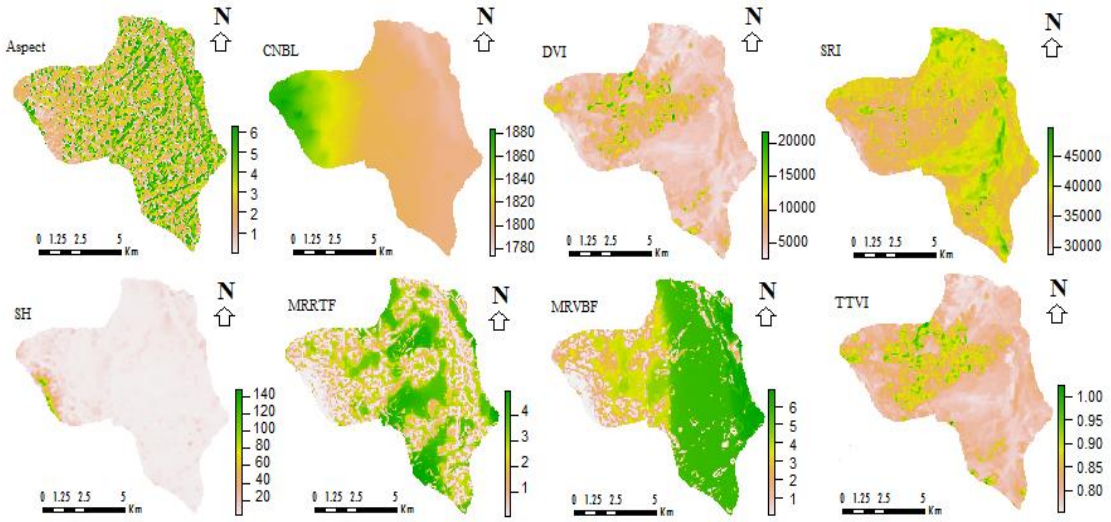

Fig. 5. Channel Network Base Level (CNBL), Difference Vegetation Index (DVI), Slope Height (SH), Ratio Vegetation Index (RVI), Thiams Transformed Vegetation Index (TTVI), Longitudinal Curvature (LC).

In third place, the importance of the variables TVI, MRRTF, B2, B3, and SI is 6.7%. According to the results, the lowlands contained significantly more clay than the highlands (Hook and Burke, 2000). The least amount of clay was observed in rangeland with severe erosion (salt crusts) and highlands. Clay spatial distribution can be determined by soil erosion (Mahmoudabadi *et al.*, 2017). Clay was found in abundance in rangeland with dense vegetation and cultivated fields. It should be emphasized that clay spatial distribution was not uniform

across cultivated fields and quite low in some areas. The link between regional variations in sand, silt, and clay components and topography has been demonstrated in studies (Keshavarzi *et al.*, 2022; Schulz *et al.*, 2023). One of the main processes of catena formation is the redistribution of soil particles by runoff, which generally increases silt, clay, and organic matter in lowlands (Hook and Burke, 2000). To estimate sand, silt, and clay fractions in floodplains, we used height, distance from the river, and groundwater floor distance (Wälder *et al.*, 2008). The depth of the valley and the vertical distance to the channel network were other key variables in their study.

In the study of silt, seven terrain-based variables and eight remote-sensing variables were the most significant. The aspect variable was the most critical variable, with 16% importance. The variables TVI, TSAVI, SPI, NDVI, S, B4, and CI are in the second order of importance, each with a share of 8%. The other variables with an importance of 4% are in the third order of importance. The highest amount of surface layer silt was observed in marshland, and in the salt crust, the lowest amount was observed. Increased silt in lowland areas is due to erosion in highland areas. In the study of sand, there were twelve terrain-based and six remote-sensing variables. The critical variables used are CNBL and CI to a depth of 30 cm. (Wälder *et al*., 2008) identified altitude, distance from the river, and distance to the water channel as the most significant variables for predicting soil properties in floodplains (Wälder *et al.*, 2008). (Rudiyanto *et al*., 2016) used distance to the river as an environmental variable and introduced it as the second critical variable in soil mapping in Indonesia. SH, SPI, and DVI variables are in a second order.

Each of these variables has a 6.9% accuracy rate in accurately predicting the spatial distribution of sand in the area. Sand and clay distribution is significantly linked to landscape location (Karaca *et al.*, 2018). The sand was found in the greatest concentrations in the salt crusts and highlands. Marshlands had the least sand quantity. The amount of sand in cultivated lands increased with increasing depth, due to tillage operations (Nieto *et al.*, 2013). Higher sand content soils have less vegetation, making them more prone to erosion and soil organic matter destruction. Therefore, the percentage of soil particles has a significant impact on organic matter and soil vegetation. In the lowlands, soil organic matter distribution is highly linked to topography and soil particle fraction. In one study, the sand-clay ratio of soil increased with increasing height and slope (Ließ *et al.*, 2012). Topographic impacts are less effective than mechanisms that influence soil particle ratios (Hook and Burke, 2000).

3.5. Machine Learning Performances

Tables 3 illustrate the statistical findings of the model. The prediction performance and accuracy were evaluated using the correlation coefficient (R^2) and the root mean square error (RMSE). According to the validation results of the models, the R^2 value of the RF model was better than the other two models. The accuracy of the model included RF, DT, and MLR, respectively. The difference in prediction accuracy between the RF and DT models was less than the difference in prediction accuracy between the RF and MLR models (Fig. 6). The highest R^2 validations in the RF model were for silt, clay, and sand, respectively. Clay validation results showed that the most accurate prediction was for the RF model. The RF model average R^2 value was 82 percent, with the 30–60 cm layer having the lowest accuracy and an RMSE value ranging from 0.82–3.71. The average R^2 in the DT model was 0.73 percent, while the RMSE for the various layers ranged from 1.7 to 4.33 percent. The R^2 in the MLR model was 0.55 percent on average, with the RMSE ranging from 4.56 to 7.22.

The overall results of the clay study show that the prediction accuracy of the models includes RF, DT, and MLR, respectively. RF is one of the most popular digital mapping techniques for

estimating particle size fractions (Hengl *et al.*, 2018); (Nussbaum *et al.*, 2018); (Zeraatpisheh *et al.*, 2019) tested and evaluated three algorithms for predicting clay content spatial distribution in a region in central Iran. Their results showed that the RF model with an R^2 of 0.15 had the highest performance compared to the RT and MLR models. (Pahlavan-Rad and Akbarimoghadam 2018) also reported the high potential of the RF model for predicting particle size fractions in a floodplain in eastern Iran. Similarly, (Da Silva Chagas *et al.*, 2016) used the RF model and MLR to estimate clay, sand, and silt content. They discovered that the RF model generates more accurate maps of sand, silt, and clay fractions. They obtained R^2 values for clay, sand, and silt of 0.57, 0.63, and 0.25, respectively.

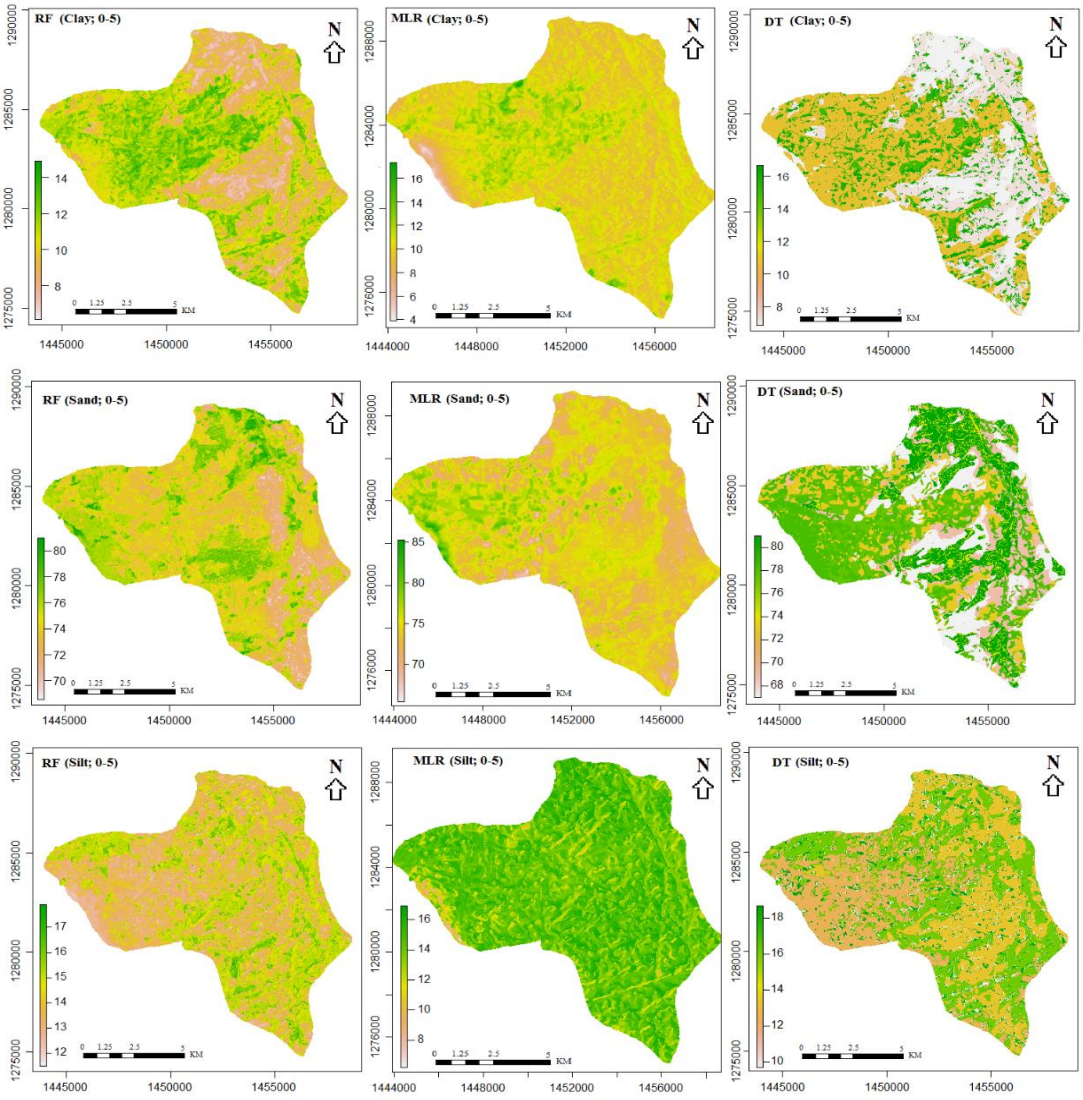

Fig. 6. The result of RF, MLR, and DT models for 0-5 cm soil depth for clay, sand, and silt properties, respectively.

In the one study using the RF model, the R^2 values for clay, sand, and silt were 0.51, 0.54, and 0.56, respectively (Taghizadeh-Mehrjardi *et al.*, 2020). In a study, a tree-based model such as RF and classification tree (CT) performed better than polynomial logistic regression, support vector machines (SVM), and artificial neural networks (ANN) for sand, silt, and clay

classification (Camera *et al.*, 2017; Wu *et al.*, 2018). In one report, the RF model was presented as the most effective strategy among the five machine-learning models for accurately assessing particle size distribution and sand, silt, and clay components classification (Zhang and Shi, 2019). Wadoux (2019) showed that RF is better than CNN (larger correlation coefficient) for predicting soil properties, including clay, sand, and silt.

Sand validation results showed that the RF model had the maximum prediction accuracy. In the RF model, the mean R^2 value was 80%, and the minimum accuracy was for layer 5–15. The RMSE value was between 2.37 and 6.73. In the DT model, the average R^2 value was 0.73%. The RMSE value for the different layers was between 1.05 and 8.22. In the MLR model, the mean R^2 value was 0.72%. The RMSE value was between 2.46 and 12.82. The overall results of this study show that the models are accurate in RF, DT, and MLR, respectively. (Hengl *et al*., 2015) compared linear regression and RF to predict soil properties in Africa and found that RF performed better than the linear regression model with an average reduction of 15–75% in RMSE. (Da Silva Chagas *et al.*, 2016) estimated sand, silt, and clay fractions using RF and multiple linear regression and discovered that RF produces more accurate change maps for sand, silt, and clay fractions than any other method.

The RF model was used to map sand, silt, and clay fractions in eastern Iran (Pahlavan-Rad and Akbarimoghaddam, 2018). Three different methods, including multiple linear regression, stepwise regression, and regression trees, were applied to predict sand, silt, and clay components. The results showed that stepwise regression was the most effective modeling method in this study (Bakker, 2012). Using remote sensing data and an RF algorithm, they studied sand, silt, and clay components at a depth of 0–20 cm with a spatial resolution of 30 m in a part of the Middle East region, and the predicted maps showed that sandy and loamy soils are widespread in surface soils in the Middle East (Poppiel *et al.*, 2021).

Table 3. Results of validation of clay, sand, and silt.

Model	Validation												
	Clay					Sand				Silt			
	Depth	R^2	CCC	RMSE	Bias	R^2	CCC	RMSE	Bias	R^2	CCC	RMSE	Bias
RF	0-5	0.85	0.74	0.96	0.86	0.77	0.79	6.73	-2.23	0.84	0.65	2.24	-0.14
	5-15	0.86	0.72	0.82	0.48	0.50	0.43	3.53	1.98	0.91	0.61	1.32	0.45
	15-30	0.87	0.76	3.67	0.46	0.91	0.78	4.29	0.50	0.80	0.66	4.11	-1.59
	30-60	0.74	0.72	2.54	-1.34	0.87	0.74	2.37	1.25	0.91	0.69	3.26	-2.73
	60-100	0.80	0.71	3.71	0.02	0.96	0.72	2.43	-0.38	0.80	0.67	3.54	-3.21
DT	0-5	0.67	0.46	4.33	3.80	0.88	0.79	1.56	-0.18	0.86	0.72	0.917	0.22
	5-15	0.94	0.73	2.30	-1.56	0.72	0.36	5.88	-4.23	0.64	0.50	3.12	2.64
	15-30	0.72	0.63	1.72	-1.04	0.93	0.79	1.05	0.46	0.80	0.44	2.55	-1.52
	30-60	0.82	0.65	2.98	2.43	0.51	0.41	8.22	-7.16	0.51	0.20	3.78	2.95
	60-100	0.51	0.61	3.19	1.34	0.61	0.53	6.8	3.15	0.75	0.51	3.11	-2.84
MLR	0-5	0.65	0.24	7.22	-1.26	0.76	0.49	4.56	-2.25	0.64	0.36	1.95	1.08
	5-15	0.78	0.37	4.72	-3.94	0.50	0.37	6.75	6.0	0.80	0.61	2.32	1.86
	15-30	0.58	0.23	4.56	-4.09	0.94	0.73	2.46	2.13	0.76	0.12	2.32	-0.55
	30-60	0.52	0.21	6.67	5.34	0.83	0.39	6.05	-2.83	0.82	0.20	3.80	-3.12
	60-100	0.23	0.13	6.55	-5.9	0.59	0.21	12.82	8.69	0.76	0.66	1.13	0.21

The silt validation results showed that the accurate prediction was for the RF model. In the RF model, the mean R^2 was 0.85%. The RMSE value was between 1.32 and 4.11. In the DT

model, the average R^2 value was 0.71%. The RMSE value for the different layers was between 0.9 and 3.78. In the MLR model, the average R^2 value was 0.76%, and the average RMSE was 2.3%. The overall results of the silt study show that the prediction accuracy of the models is RF, DT, and then MLR, respectively. Different multiple regression models were used to correlate the parameters derived from the digital elevation model with soil properties and reported high success rates (Keshavarzi *et al.*, 2022). The variables included in the model showed that topographic difficulties and climatic data are critical variables for sand, silt, and clay components modeling (Bakker, 2012). An estimation of the watershed in Canada based on a set of hydrographic parameters derived from the DEM and the artificial neural network model of sand, silt, and clay components in Canada (Zhao *et al.*, 2009).

3.6. Prediction accuracy

Table 4 shows 10-fold cross-validation results Accuracy varies between measured and predicted values at different depths, and these results are acceptable for different layers. This accuracy is higher for sand than for silt and clay and almost the same for silt and clay. The error in predicting sand is also more significant than that in predicting clay, and clay is heavier than sand. Due to temporary and permanent water channels, sand has less spatial heterogeneity, and the area has less landscape complexity. It is probably because this particle estimation is more accurate than the other two particles.

Table 4. Results of predictions accuracy with 10-fold cross-validation method.

Layers (cm)	Silt		Sand		Clay	
	R^2	RMSE	R^2	RMSE	R^2	RMSE
0-5	0.37	2.56	0.28	5.29	0.21	4.13
5-15	0.37	2.91	0.44	5.82	0.32	5.16
15-30	0.39	2.39	0.39	4.29	0.42	4.35
30-60	0.26	2.85	0.35	5.16	0.20	4.34
60-100	0.30	3.02	0.42	5.73	0.36	5.07

4. Conclusion

The results showed that using variables such as channel network base level and vertical distance to channel networks increased the accuracy of prediction maps. The channel network plays a principal role in the soil distribution of the study area by affecting the rate of discharge that leads to erosion and sedimentation. The accuracy and precision of the models used for all three particles showed that the RF model had the most accurate prediction with R^2 and RMSE of 0.82 and 2.34 for clay, 0.80 and 3.87 for sand, and 0.85 and 2.89 for silt, respectively. In this study, terrain-based variables and remote sensing increased the accuracy of predictions, but in predicting sand, silt, and clay components, terrain-based variables had a greater ratio of predictions and increased their accuracy. The ratio of terrain-based variables used for both surface horizons and subsurface horizons was higher than remote sensing variables. Therefore, the use of geomorphometric features along with the RF model affects the prediction of the spatial distribution of sand, silt, and clay components in the region, and it causes increases in accuracy prediction. The present study showed that digital mapping of sand, silt, and clay particles according to GlobalSoilMap and using environmental variables with a small number of soil samples can provide acceptable results.

References

Bakker, A., 2012. Soil texture mapping on a regional scale with remote sensing data. MSC Thesis, Wageningen University, Centre for Geo-Information, Netherlands, 71 p.

Biswas, A., Y. Zhang, 2018. Sampling designs for validating digital soil maps: a review. Pedosphere, 28(1), 1-15. https://doi.org/10.1016/S1002-0160(18)60001-3.

Breiman, L., 2001. Random forests. Machine learning. 45(1), 5-32. https://doi.org/10.1023/A:1010933404324.

Brenning, A., 2008. Statistical geocomputing combining R and SAGA: The example of landslide susceptibility analysis with generalized additive models. Hamburger Beiträge zur Physischen Geographie und Landschaftsökologie, 19 (23-32), 410.

Brungard, W., J.L. Boettinger, M.C. Duniway, S.A. Wills, Jr.T.C. Edwards, 2015. Machine learning for predicting soil classes in three semi-arid landscapes. Geoderma, 239, 68-83. https://doi.org/10.1016/j.geoderma.2014.09.019.

Camera, Z., J.S. Zomeni, A.M. Noller, I.C. Zissimos, A. Christoforou Bruggeman, 2017. A high resolution map of soil types and physical properties for Cyprus: A digital soil mapping optimization. Geoderma, 285, 35-49. https://doi.org/10.1016/j.geoderma.2016.09.019.

Campos, R., E. Giasson, J.J.F. Costa, I.R. Machado, E.B.D. Silva, B.R. Bonfatti, 2018. Selection of Environmental Covariates for Classifier Training Applied in Digital Soil Mapping. Revista Brasileira de Ciência do Solo, 42. https://doi.org/10.1590/18069657rbcs20170414.

Chen, T. L., Z.L. Shi, A.B. Wen, D.C. Yan, J. Guo, J.C. Chen, Y. Liu, R.Y. Chen, 2021. Multifractal characteristics and spatial variability of soil particle-size distribution in different land use patterns in a small catchment of the Three Gorges Reservoir Region. China. Journal of Mountain Science, 18(1), pp.111-125. https://doi.org/10.1007/s11629-020-6112-5.

Da Silva Chagas, C., W. Carvalho Junior, S.B. de Bhering, B. Calderano Filho, 2016. Spatial prediction of soil surface texture in a semiarid region using random forest and multiple linear regressions. Catena 139, 232–240. https://doi.org/10.1016/j.catena.2016.01.001.

Darst, F., K.C. Malecki, C.D. Engelman, 2018. Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. BMC genetics, 19(1), 65. https://doi.org/10.1186/s12863-018-0633-8.

Development Core Team, R., 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna Austria.

Dewi, C., R.C. Chen, 2019. Random Forest and Support Vector Machine on Features Selection for Regression Analysis. Int. J. Innov. Comput. Inf. Control, 15(6), 2027-2037. DOI: 10.24507/ijicic.15.06.2027.

Dharumarajan, S., R. Hegde, 2022, Digital Mapping of Soil Texture Classes Using Random Forest Classification Algorithm. Soil Use Manag, 38, 135–149. https://doi.org/10.1111/sum.12668.

Dharumarajan, S., R. Hegd, M. Lalitha, B. Kalaiselvi, S.K. Singh, 2019, Pedotransfer Functions for Predicting Soil Hydraulic Properties in Semi-Arid Regions of Karnataka Plateau, India. Curr. Sci., 116, 1237. https://doi: 10.18520/cs/v116/i7/1237-1246.

Hartemink, A. E., J. Hempel, P. Lagacherie, A. McBratney, N. McKenzie, R.A. MacMillan, B. Minasny, L. Montanarella, M.L. de Mendonça Santos, P. Sanchez, M. Walsh, 2010. GlobalSoilMap. Net –a new digital soil map of the world. In Digital soil mapping (pp. 423-428). Springer, Dordrecht. https://doi.org/10.1007/978-90-481-8863-5_33.

Hengl, T., G.B. Heuvelink, B. Kempen, J. G. Leenaars, M.G. Walsh, K.D. Shepherd, A. Sila, R.A. MacMillan, J.M. de Jesus, L. Tamene, J. E. Tondoh, 2015. Mapping soil properties of Africa at 250 m resolution: random forests significantly improve current predictions. PLoS One 10, 1–26. https://doi.org/10.1371/journal.pone.0125814.

Hengl, T., J.M. de Jesus, R.A. MacMillan, N.H. Batjes, G.B. Heuvelink, E. Ribeiro, A. Samuel-Rosa, B. Kempen, J.G. Leenaars, M.G. Walsh, M.R. Gonzalez, 2014. SoilGrids1km-global soil information based on automated mapping. PloS one, 9(8), p.e105992. https://doi.org/10.1371/journal.pone.0105992.

Hengl, T., M. Nussbaum, M.N. Wright, G.B. Heuvelink, B. Gräler, 2018. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ 6, e5518.

Heung, B., H.C. Ho, J. Zhang, A. Knudby, C.E. Bulmer, M.G. Schmidt, 2016. An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. Geoderma 265. 62-77. https://doi.org/10.1016/j.geoderma.2015.11.014.

Hook, P.B., I.C. Burke, 2000. Biogeochemistry in a shortgrass landscape: control by topography, soil texture, and microclimate. Ecology, 81(10), 2686-2703. https://doi.org/10.1890/0012-9658(2000)081[2686:BIASLC]2.0.CO;2.

Jafari, A., P.A. Finke, Vande Wauw, J., Ayoubi, S., Khademi, H., 2012. Spatial prediction of USDA-great soil groups in the arid Zarand region, Iran: comparing logistic regression approaches to predict diagnostic horizons and soil types. European Journal of Soil Science, 63(2), 284-298. https://doi.org/10.1111/j.1365-2389.2012.01425.x.

Jamshidi, M., M.A. Delavar, R. Taghizadehe-Mehrjerdi, C. Brungard, 2019. Evaluating Digital Soil Mapping Approaches for 3D Mapping of Soil Organic Carbon. Iranian Journal of Soil Research, 33(2), 227-239. 10.22092/IJSR.2019.119764.

Jeihouni, M., S.K. Alavipanah, A. Toomanian, A.A. Jafarzadeh, 2020. Soil texture fractions modeling and mapping using LS-SVR algorithm. Desert, 25(2), pp.147-154. 10.22059/JDESERT.2020.79252.

Karaca, S., F. Gülser, R. Selçuk, 2018. Relationships between soil properties, topography and land use in the Van Lake Basin, Turkey. Eurasian Journal of Soil Science, 7(2), 115-120. https://doi.org/10.18393/ejss.348412.

Kaya, F., L. Başayiğit, A. Keshavarzi, R. Francaviglia, 2022. Digital mapping for soil texture class prediction in northwestern Türkiye by different machine learning algorithms. Geoderma Regional, 31, p.e00584.

Keshavarzi, A., M.Á.S. Del Árbol, F. Kaya, Y. Gyasi-Agyei, J. Rodrigo-Comino, 2022. Digital mapping of soil texture classes for efficient land management in the Piedmont plain of Iran. Soil Use and Management, 38(4), pp.1705-1735.

Lawrence, I., K. Lin, 1989. A concordance correlation coefficient to evaluate reproducibility. Biometrics: 45; 255-268. https://doi.org/10.2307/2532051.

Li, Q., F. Gu, Y. Zhou, T. Xu, L. Wang, Q. Zuo, L. Xiao, J. Liu, Y. Tian, 2021. Changes in the Impacts of Topographic Factors, Soil Texture, and Cropping Systems on Topsoil Chemical Properties in the Mountainous Areas of the Subtropical Monsoon Region from 2007 to 2017: A Case Study in Hefeng, China. International Journal of Environmental Research and Public Health, 18(2), p.832. https://doi.org/10.3390/ijerph18020832.

Ließ, M., B. Glaser, B. Huwe, 2012. Making use of the World Reference base diagnostic horizons for the systematic description of the soil continuum-Application to the tropical mountain soil-landscape of southern Ecuador. Catena, 97, 20–30. https://doi.org/10.1016/j.catena.2012.05.002.

Mahler PJ. (Ed.). 1970. Manual of land classification for irrigation. Ministry of Agriculture.

Mahmoudabadi, E., A. Karimi, G.H. Haghnia, A. Sepehr, 2017. Digital soil mapping using remote sensing indices, terrain attributes, and vegetation features in the rangelands of northeastern Iran. Environmental monitoring and assessment, 189(10), 500. https://doi.org/10.1007/s10661-017-6197-7.

Malone, P., B. Minasny, A.B. McBratney, 2017. Using R for digital soil mapping. Basel, Switzerland: Springer International Publishing. Sydney Institute of Agriculture the University of Sydney Eveleigh Australia. 262p. https://doi.org/10.1007/978-3-319-44327-0.

McBratney, A.B., M.M. Santos, B. Minasny, 2003. On digital soil mapping. Geoderma, 117(1-2), pp.3-52. https://doi.org/10.1016/S0016-7061(03)00223-4.

Mehrabi-Gohari, E., H.R. Matinfar, A. Jafari, R. Taghizadeh-Mehrjardi, J. Triantafilis, 2019. The spatial prediction of soil texture fractions in arid regions of Iran. Soil Systems, 3(4), p.65. https://doi.org/10.3390/soilsystems3040065.

Minasny, B., A.B. McBratney, 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. Comput. Geosci. 32:1378–1388. https://doi.org/10.1016/j.cageo.2005.12.009.

Mirzaeitalarposhti, R., H. Shafizadeh-Moghadam, R. Taghizadeh-Mehrjardi, M.S. Demyan, 2022. Digital soil texture mapping and spatial transferability of machine learning models using sentinel-1, sentinel-2, and terrain-derived covariates. Remote Sensing, 14(23), p.5909.

Mondejar, J. P., A.F. Tongco, 2019. Estimating topsoil texture fractions by digital soil mapping-a response to the long outdated soil map in the Philippines. Sustainable Environment Research, 29(1), pp.1-20. https://doi.org/10.1186/s42834-019-0032-5.

Murthy, S. K., 1998. Automatic construction of decision trees from data: A multi-disciplinary survey. Data mining and knowledge discovery, 2(4), 345-389. https://doi.org/10.1023/A:1009744630224.

Nieto, O. M., J. Castro, E. Fernández-Ondoño, 2013. Conventional tillage versus cover crops in relation to carbon fixation in Mediterranean olive cultivation. Plant and Soil, 365(1-2), 321-335. https://doi.org/10.1007/s11104-012-1395-0.

Nussbaum, M., K. Spiess, A. Baltensweiler, U. Grob, A. Keller, L. Greiner, M.E. Schaepman, A. Papritz, 2018. Evaluation of digital soil mapping approaches with large sets of environmental covariates. Soil 4, 1–22. https://doi.org/10.5194/soil-4-1-2018.

Olaya, V., O. Conrad, 2009. Geomorphometry in SAGA. Developments in soil science. 33, 293-308. https://doi.org/10.1016/S0166-2481(08)00012-3.

Pahlavan-Rad, M. R., A. Akbarimoghaddam, 2018. Spatial variability of soil texture fractions and pH in a floodplain (case study from eastern Iran). Catena, 160, 275-281. https://doi.org/10.1016/j.catena.2017.10.002.

Planchon, O., F. Darboux, A. Fast, 2002. Simple and versatile algorithm to fill the depressions of digital elevation models. Catena, 46 (2-3), 159-176. https://doi.org/10.1016/S0341-8162(01)00164-3.

Poppiel, R. R., J.A.M. Demattê, N.A. Rosin, L.R. Campos, M. Tayebi, B.R. Bonfatti, S. Ayoubi, S. Tajik, F.A. Afshar, A. Jafari, N. Hamzehpour, 2021. High resolution middle-eastern soil attributes mapping via open data and cloud computing. Geoderma, 385, p.114890. https://doi.org/10.1016/j.geoderma.2020.114890.

Rudiyanto, B., B. Minasny, I. Setiawan, C. Arif, S.K. Saptomo, Y. Chadirin, 2016. Digital mapping for cost- effective and accurate prediction of the depth and carbon stocks in Indonesian peatlands. Geoderma 117 (1–2), 3–52. https://doi.org/10.1016/j.geoderma.2016.02.026.

Rumao, S., 2019. Exploration of Variable Importance and Variable selection techniques in presence of correlated variables. MSC Thesis. Rochester Institute of Technology College of Science Department of Mathematical Sciences. U.S. states, 71 p.

SAGA DEVELOPMENT TEAM. 2011. System for Automated Geoscientific Analyses [Online]. Available from: http://www.saga-gis.org. [Last accessed: 12 September, 2010].

Schulz, G.A., D.M. Rodriguez, M. Angelini, L.M. Moretti, G.F. Olmedo, L.M. Tenti Vuegen, J.C. Colazo, M. Guevara, 2023. Digital Soil Texture Maps of Argentina and Their Relationship to Soil-Forming Factors and Processes. In Geopedology: An Integration of Geomorphology and Pedology for Soil and Landscape Studies (pp. 263-281). Cham: Springer International Publishing.

Silleos, N.G., T.K. Alexandridis, I.Z. Gitas, K. Perakis, 2006. Vegetation indices: advances made in biomass estimation and vegetation monitoring in the last 30 years. Geocarto International, 21(4), pp.21-28. https://doi.org/10.1080/10106040608542399.

Soil and Water Research Institute. 1995. Semi-detailed studies of geology and land classification of Chahardowli area of Kurdistan province. Journal No. 985, 20 pages. Karaj. Iran.

Sparks, D.L., A.L. Page, P.A. Helmke, R.H. Loeppert, 2020. (Eds.) Methods of soil analysis, part 3: Chemical methods (Vol. 14). John Wiley & Sons.

Taghizadeh-Mehrjardi, R., F. Sarmadian, B. Minasny, J. Triantafilis, M. Omid, 2014. Digital mapping of soil classes using decision tree and auxiliary data in the Ardakan region, Iran. Arid Land Research and Management, 28(2), 147-168. https://doi.org/10.1080/15324982.2013.828801.

Taghizadeh-Mehrjardi, R., M. Mahdianpari, F. Mohammadimanesh, T. Behrens, N. Toomanian, T. Scholten, K. Schmidt, 2020. Multi-task convolutional neural networks outperformed random forest for mapping soil particle size fractions in central Iran. Geoderma, 376, p.114552. https://doi.org/10.1016/j.geoderma.2020.114552.

Taghizadeh-Mehrjardi, R., N. Toomanian, A.R. Khavaninzadeh, A. Jafari, J. Triantafilis, 2016. Predicting and mapping of soil particle-size fractions with adaptive neuro-fuzzy inference and ant colony optimization in central Iran. European Journal of Soil Science 67, no. 6: 707-725. https://doi.org/10.1111/ejss.12382.

Vagen, T.G., L.A. Winowiecki, J.E. Tondoh, L.T. Desta, T. Gumbricht, 2016. Mapping of soil properties and land degradation risk in Africa using MODIS reflectance. Geoderma, 263, pp.216-225. https://doi.org/10.1016/j.geoderma.2015.06.023.

Wadoux, A.M.C., 2019. Using deep learning for multivariate mapping of soil with quantified uncertainty. Geoderma 351, 59–70. https://doi.org/10.1016/j.geoderma.2019.05.012.

Wälder, K., O. Wälder, J. Rinklebe, J. Menz, 2008. Estimation of soil properties with geostatiscal methods in floodplains. Archives of Agronomy and Soil Science 54 (3), 275–295. https://doi.org/10.1080/03650340701488485.

Wilson, J.P., J.C. Gallant, 2000. Digital terrain analysis. Terrain analysis: Principles and applications. 6(12), 1-27.

Wright, M.N., A. Ziegler, 2015. Ranger: A Fast Implementation of Random Forests for High Dimensional Data in C^{++} and R. Journal of Statistical Software. 77, 1–17. https://doi.org/10.48550/arXiv.1508.04409.

Wu, W., A.D. Li, X.H. He, R. Ma, H.B. Liu, J.K. Lv, 2018, A comparison of support vector machines, artificial neural network and classification tree for identifying soil texture classes in southwest China. Computers and Electronics in Agriculture, 144, pp.86-93. https://doi.org/10.1016/j.compag.2017.11.037.

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

Zhang, M., W. Shi, 2019. Systematic comparison of five machine-learning methods in classification and interpolation of soil particle size fractions using different transformed data. Hydrology and Earth System Sciences Discussions, 1-39. https://doi.org/10.5194/hess-2018-584.

Zhao, Z., T.L. Chow, H.W. Rees, Q. Yang, Z. Xing, F.R. Meng, 2009. Predict soil texture distributions using an artificial neural network model. Computers and electronics in agriculture, 65(1), pp.36-48. https://doi.org/10.1016/j.compag.2008.07.008.