DESERT

Desert
Online at http://desert.ut.ac.ir

Desert 24-1 (2019) 153-169

Application of soil properties, auxiliary parameters, and their combination for prediction of soil classes using decision tree model

M. Shahini Shamsabadi[a], I. Esfandiarpour-Borujeni[a*], H. Shirani[a], M.H. Salehi[b]

[a] *Soil Science Department, Faculty of Agriculture, Vali-e-Asr University of Rafsanjan, Rafsanjan, Iran*
[b] *Soil Science Department, Faculty of Agriculture, Shahrekord University, Shahrekord, Iran*

Received: 3 October 2018; Received in revised form: 12 December 2018; Accepted: 14 December 2018

Abstract

Soil classification systems are very useful for a simple and fast summarization of soil properties. These systems indicate the method for data summarization and facilitate connections among researchers, engineers, and other users. One of the practical systems for soil classification is Soil Taxonomy (ST). As determining soil classes for an entire area is expensive, time-consuming, and almost impossible, this research has tried to predict the soil classes in each level of the ST system (up to family level) by using the data of 120 excavated pedons and some auxiliary parameters (such as derivatives of digital elevation model, i.e., DEM) in Shahrekord plain, central Iran. For this reason, the decision tree model was encoded and implemented in the MATLAB software for three conditions: use of soil properties, auxiliary parameters, and its combination. According to the results, soil class prediction error by using soil properties, auxiliary parameters, and its combination was estimated to be 0, 3.33 and 0% for order and suborder levels; 0.83, 15 and 0.83% for great group level; 3.33, 22.5 and 3.33% for subgroup level and 30, 52.5 and 30% for family level, respectively. In addition, the use of kriging maps of soil properties (instead of 120 observational points) decreased the prediction error of the modeling in all levels of the ST system. It seems that the effect of auxiliary parameters (in comparison to soil properties) is not very significant for predicting soil classes in low-relief areas.

Keywords: Soil classification; Kriging maps; Digital soil mapping; Sensitivity analysis

1. Introduction

Soil classification systems indicate a set of quantitative methods which are used to show similar soils and compare different soils (Das, 2000). One of the practical systems for soil classification is Soil Taxonomy (ST), which is more useful in agriculture because of the consideration of soil particle size distribution (Soil Survey Division Staff, 1993). Today, one of the most important research subjects in geology is the comprehension of estimation accuracy of soil classes based on the limited point data, previous researches, and the correlation between soil and landscape (Goodman and Owen, 2012). McBratney *et al*. (2003) suggested the use of correlation between soil data and auxiliary parameters for estimating soil classes and soil properties. They expressed that the auxiliary data can be chosen based on soil formation factors. According to the viewpoints of these researchers, the factors of soil formation can be introduced in the SCORPAN model, in accordance to the following equation:

$$S = f\ (s, c, o, r, p, a, n) + \varepsilon \quad (1)$$

S shows the soil class under prediction, which depends on factors including soil (s), climate (c), organisms (o), topography (r), parent material

* Corresponding author. Tel.: +98 34 31312019
Fax: +98 34 31312042
E-mail address: esfandiarpoor@vru.ac.ir

(p), time (a) and location (n). Moreover, ε shows the model prediction error.

Massawe *et al*. (2018) used methods of machine learning for the prediction of the soil classes in Tanzania and concluded that the use of derivatives of the digital elevation model (DEM) can be useful in predicting the soil classes. Bagheri Bodaghabadi *et al*. (2011) considered DEM derivatives as the most important source for input of digital soil maps in Boroujen, Chaharmahal and Bakhtiari province, central Iran. These researchers declared that wetness index, radiation duration, slope, and sediment index are more important than other DEM derivatives for predicting soil properties and for determining the pattern of soil distribution in that area. Jafari *et al*. (2012) used geomorphology maps as an input for digital soil maps and expressed that these maps have improved prediction accuracy of soil classes. They concluded that some soils, which were influenced by geomorphology and topography, had been predicted more exactly.

One of the practical models in digital soil mapping is the decision tree model. This model is based on a consecutive division of data into a set of discrete groups and tries to increase the distance between groups in the separation process. Various researchers have used the decision tree model to solve many problems of classification and regression. Taghizadeh-Mehrjardi *et al*. (2015) used six methods of data mining including regression, artificial neural network, support vector machine, the nearest neighbor k, random forest, and decision tree to predict the soil families in Baneh, western Iran. They concluded that the decision tree model and the artificial neural network have been the most accurate methods. Hash *et al*. (2009) used the decision tree model to predict soil class in Nevada. They concluded that this method predicted soil groups more accurately. Saunders and Boettinger (2007) evaluated efficiency of the decision tree in soil classes' prediction in the Wyoming State as optimum. Scull *et al*. (2005) applied the decision tree model to predict soil classes using remote sensing data and DEM derivatives, and expressed a clear correlation between soil and landscape as necessary for experts to be able to use this method. In addition, they suggested that because of the wide range of auxiliary parameters, which are used in predicting the soil classes and the high flexibility of the decision tree, the mentioned model has a high efficiency in the prediction of soil classes. The other researchers such as Lagacherie and Holmes (1997) in France and Russia, Odgers *et al*. (2014) in Australia, Holmes *et al*. (2015) in Australia and Adhikari, and Hartemink (2016) in Denmark, have used the decision tree model to predict soil classes as well.

Although many studies have been done about predicting soil classes during the recent years, these predictions were rarely conducted in different levels of the ST system for low-relief regions (such as the Shahrekord plain). In addition, the separation of the effect of soil properties and auxiliary parameters has not been done for predicting soil classes. Therefore, the main goal of the present research is to use soil properties, auxiliary parameters, its combination in predicting soil classes, and comparing the results at different levels of the ST system (up to family class) using the decision tree model. The use of the kriging maps of soil properties as an input of the decision tree model and comparing the results of these maps with the results obtained from the soil characteristics measured at 120 observation points as inputs of the model is another goal of this research. Iran lacks a correct definition for soil series. Therefore, this research does not consider this level of the ST system.

2. Materials and Methods

2.1. The Study Area

The study area is a part of the Shahrekord plain, located in Chaharmahal-va-Bakhtiari Province, Iran (Figure 1). This region is located between 32˚12' to 32˚23' N, and 50˚45' to 50˚59' E, with a mean altitude of about 2060m a.s.l. The mean annual rainfall and temperature of the region, during a 50-year period (1966-2016), is 329mm and 12.3 ℃, respectively. The approximate area of this region is 10000ha, with 4800ha of the area including cultivated land. The main cultivated crops in this region are wheat, alfalfa, potato, and forage maize. The soil moisture and temperature regimes of the area are xeric and mesic, respectively (Soil Survey Staff, 2014). The soils of this area have been formed on Quaternary shale and foliated clayey limestone deposits (Geological Survey and Mineral Exploration of Iran, 2017).

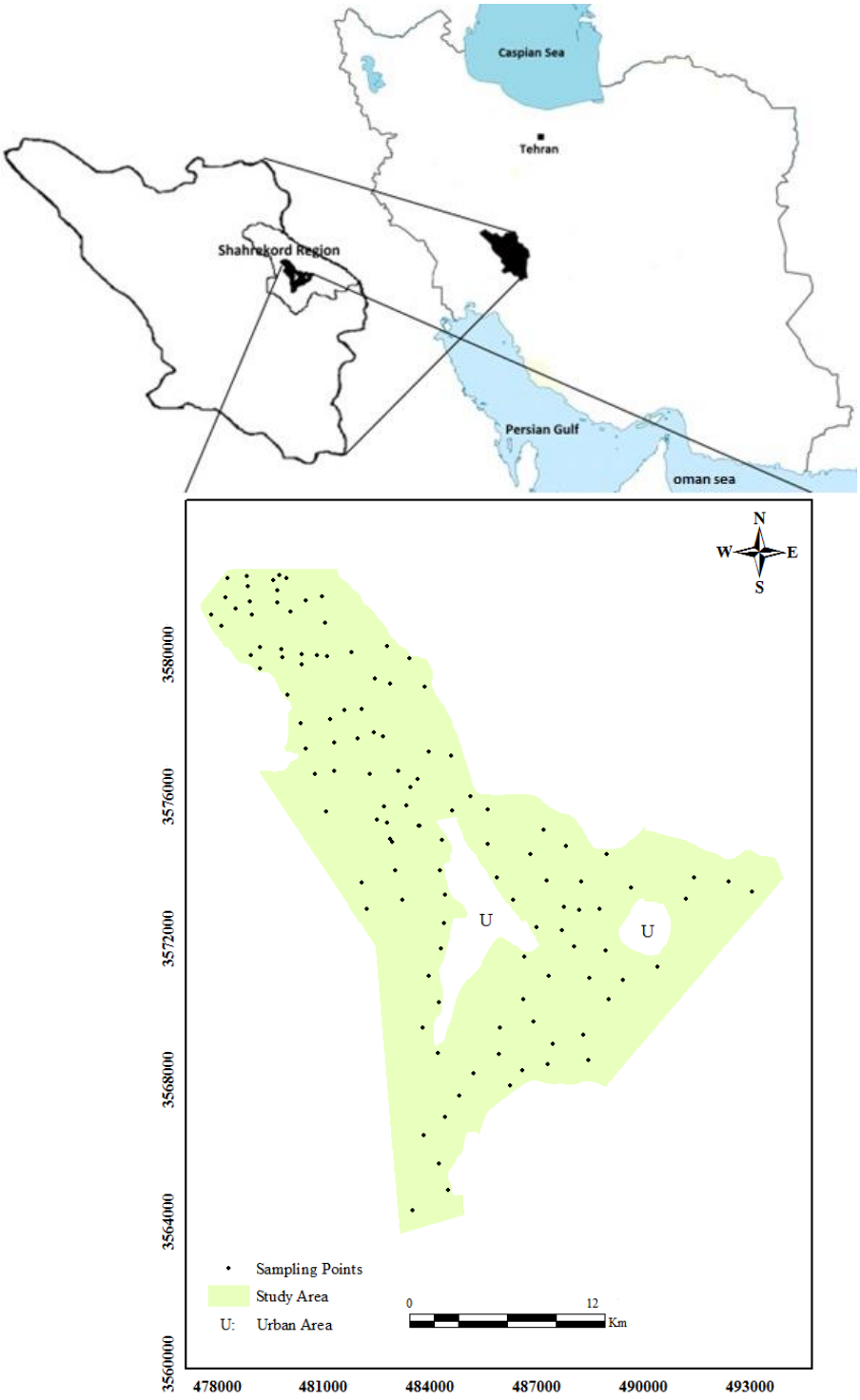

Fig. 1. The location of study area with 120 sampling points

2.2. The Soil Properties Used for Soil Class Prediction

The soil properties used in this research, have been extracted from the data of 120 excavated pedons in the Shahrekord plain (Figure 1) obtained by Mosleh *et al.* (2017) in the form of a random stratified sampling pattern for the preparation of a digital map of soil classes using random forest (RF), boosted regression tree (BRT), artificial neural network (ANN), and multinomial logistic regression (MLR) models. In the present study, only the important properties affecting the classification of soils at different levels of the ST System (Soil Survey Staff, 2014) were considered for modeling soil

classes by the decision tree model. In other words, after final classification of all pedons, only the main discriminant properties of studied soils in each level of ST System were considered, and after modeling, the most effective properties were determined and compared from the model's perspective by sensitivity analysis and tree structure. Therefore, only a property of presence or absence of argillic horizon for order level (separating Inceptisols from Alfisols), and two properties of presence or absence of argillic horizon and/or aquic moisture regime for suborder level (separating Aquepts, Xerepts, and Xeralfs) were used for predicting the soil classes. In addition, some properties including the presence or absence of cambic, argillic, calcic, petrocalcic horizons, aquic moisture regime, secondary carbonates, and chroma of 2 or less, were used for the prediction of soil classes in the great group and subgroup levels.

In order to predict soil classes at the family level, all of the used characteristics in the subgroup level were used to determine the soil families (such as carbonatic mineralogy class, cation exchange activity class, and presence or absence of shallow depth class). It should be noted that in order to define the qualitative properties of the modelling process, zero marked the non-existence mode and one marked the existence mode of a property.

2.3. The Auxiliary Properties Used for Soil Class Prediction

One of the auxiliary parameters used in this research was DEM derivatives. Therefore, DEM of area was prepared with spatial resolution of 30×30m which was downloaded from the Aster GDEM database (US Geology Survey, 2014) and projected to the Universal Transverse Mercator (UTM) projection. Then, DEM derivatives including longitudinal curvature, cross sectional curvature, aspect, elevation, slope, analytical hill shading, convergence index, closed depressions, catchment area, topographic wetness index, LS factor, channel network base level, vertical distance to channel network, valley depth, and relative slope position were prepared. In addition to DEM derivatives, geo-form maps of the area (including the landform map and the landform phase map) were prepared using aerial photographs (1:40000) and based on the hierarchy presented by Zinck (1989). The geologic map of the area with a scale of 1:25000 (Geological Survey and Mineral Exploration of Iran, 2017) and the soil map of the area with a scale of 1:50000 (Mohammadi, 1986) were used as auxiliary parameters. Additionally, normalized difference vegetation index (NDVI) was computed and used as an auxiliary parameter with the following equation:

$$\mathrm{NDVI} = (\mathrm{NIR} - \mathrm{RED}) / (\mathrm{NIR} + \mathrm{RED}) \qquad (2)$$

NIR and RED are the amount of reflection of near infrared and red waves, respectively Rouse *et al*., 1973). It must be mentioned that this index has been calculated by images taken from Landsat 8 satellite in June, 2017.

2.4. Decision Tree Model

This model is a non-parametric method which can use a set of qualitative and quantitative predictive variables for classification. In fact, a decision tree shows the direct and indirect correlations of several independent variables with a target variable (dependent) as a tree structure and by a reversible classification of data (Taghizadeh-Mehrjerdi *et al*., 2014). In the mentioned tree structure above, if a variable separates the classes in higher branches, it will have more influence on the class prediction.

DEM derivatives and a resampled map were first inputed in the MATLAB software (version 2015) for the modelling process. Then, sensitivity analysis was done by the StatSoft method (StatSoft Inc, 2004) to determine the most effective properties in soil classes' estimation. In this method, the sensitivity value for each input property is derived from the division of the network error in the absence of the desired input property on the network error in the presence of all input variables. It means that at first the model was created with all the input variables. The amount of error index (ε') was computed after gaining the best performance or the least amount of errors. Afterwards, a certain input variable was deleted and the model was remade by other input properties. In addition, the amount of error index (ε) was determined in this condition after obtaining the most suitable structure and performance of the model. The amount of output sensitivity compared to the input variable under study was computed by the ratio of the error index in the second condition (deletion of input properties to the first condition (presence of all inputs)). In this method, any property with a sensitivity analysis of more than one value was more valueable in the prediction of soil classes. In other words, the numerical value in this method is as follows:

$$n = \frac{\varepsilon}{\varepsilon\prime} \qquad (3)$$

Where, n is the numerical value of the sensitivity analysis by the StatSoft method for a special property, ε is the rate of the prediction error with all properties except a special property, and ε' is the rate of the prediction error in the presence of all the properties (along with the same special property) (StatSoft Inc, 2004).

In addition, the structure of the decision tree can show the most effective properties for the prediction of soil classes. This means that some properties in higher branches are more effective for this prediction. The estimation error percentage was computed and compared by the following equation, for each level of ST system:

$$e = \frac{n}{N} \times 100 \quad (4)$$

where, e is the percentage of error, n is the number of points which have not been predicted correctly, and N is the total number of the studied points.

2.5. Kriging Maps of Soil Properties

In this research, the measured soil properties in 120 observation points were considered as the model inputs. After that, the kriging maps of these properties were drawn and used as the input model data. In fact, the authors of current research wanted an answer to if the modelling results of the two mentioned conditions were different in terms of accuracy level. For this purpose, the kriging maps of soil properties were prepared by ArcGIS software (version 10.3). These maps include: the presence or absence of argillic, calcic, and petrocalcic horizons, chroma of one or less, chroma of two or less, secondary carbonates in genetic horizons of pedons, and the percentage of calcium carbonate equivalent, sand, clay, clay-size carbonates, particles with diameters of 0.1 to 75mm, coarse fragments (≥ 2mm), and amounts of cation exchange activity in the control section of family level of ST system. It must be mentioned that kriging maps of soil classes were prepared in five levels of the ST system (up to family level). In fact, the total number of region pixels were determined (125000 pixels) in relation to the area of each pixel (0.08ha) and the area of the whole region (10000ha). The amounts of soil properties and soil classes were determined for each of the quintuple levels of the ST system (up to family level) based on geographical coordinates of each used point prepared by kriging maps.

3. Results and Discussion

3.1. Studied Soils Abundance and Their Distribution Pattern

Figure 2 shows the taxonomy results of the studied pedons and its abundance percentage to the subgroup level. As indicated in Figure 2, Typic Calcixerepts subgroup has the most abundant soil in the region and Typic Endoaquepts, Typic Haploxeralfs, and Aquic Haploxerepts subgroups are the least abundant soils in the Shahrekord plain. 9.17% of the studied soils have been considered as the Alfisols order class. Figure 3 shows a distribution of the studied soils in the region at subgroup level. It is observed that most parts of the Alfisols order class were located in the west and northwest of the studied region. Bockheim and Hartemink (2013) suggested that argillic horizon could be formed in pergelic, cryic, frigid, mesic, thermic and hyper-thermic soil temperature regimes, and in aquic, udic, ustic, xeric and aridic soil moisture regimes. However, the formation of argillic horizon can occur in moisture regions more than dry ones (Gunal and Ransom, 2006; Khormali *et al*., 2012). Gunal and Ransom (2006) also stated that the least amount of annual rainfall for the formation of argillic horizon is about 400-500mm. As the mean annual rainfall in the studied region is 329mm, it seems that the formation of argillic horizon in this region is a result of a wetter paleoclimate. There are also alternative periods of wetness and dryness, which causes the subsurface dry soil to absorb moisture when the water and clay transferred to the subsurface dry soil, so the clay illuviates as clay films in the walls of pores. For this reason, it is probable that long-term irrigation in the agricultural lands of the region would form argillic horizon and as a result, the formation of Alfisols.

There are some low lands in the southern or central parts of the study area in which surface water accumulation has occurred due to the high groundwater levels. This has caused for an aquic moisture regime too take place. Additionally, the accumulation of secondary carbonates have occurred in much of the soils, due to the calcareous nature of the parent material of the study area. Petrocalcic horizon is often formed. For this reason, some parts of the studied soils have been located in the great groups of Calcixerepts or petrocalcic subgroups (Figure 3).

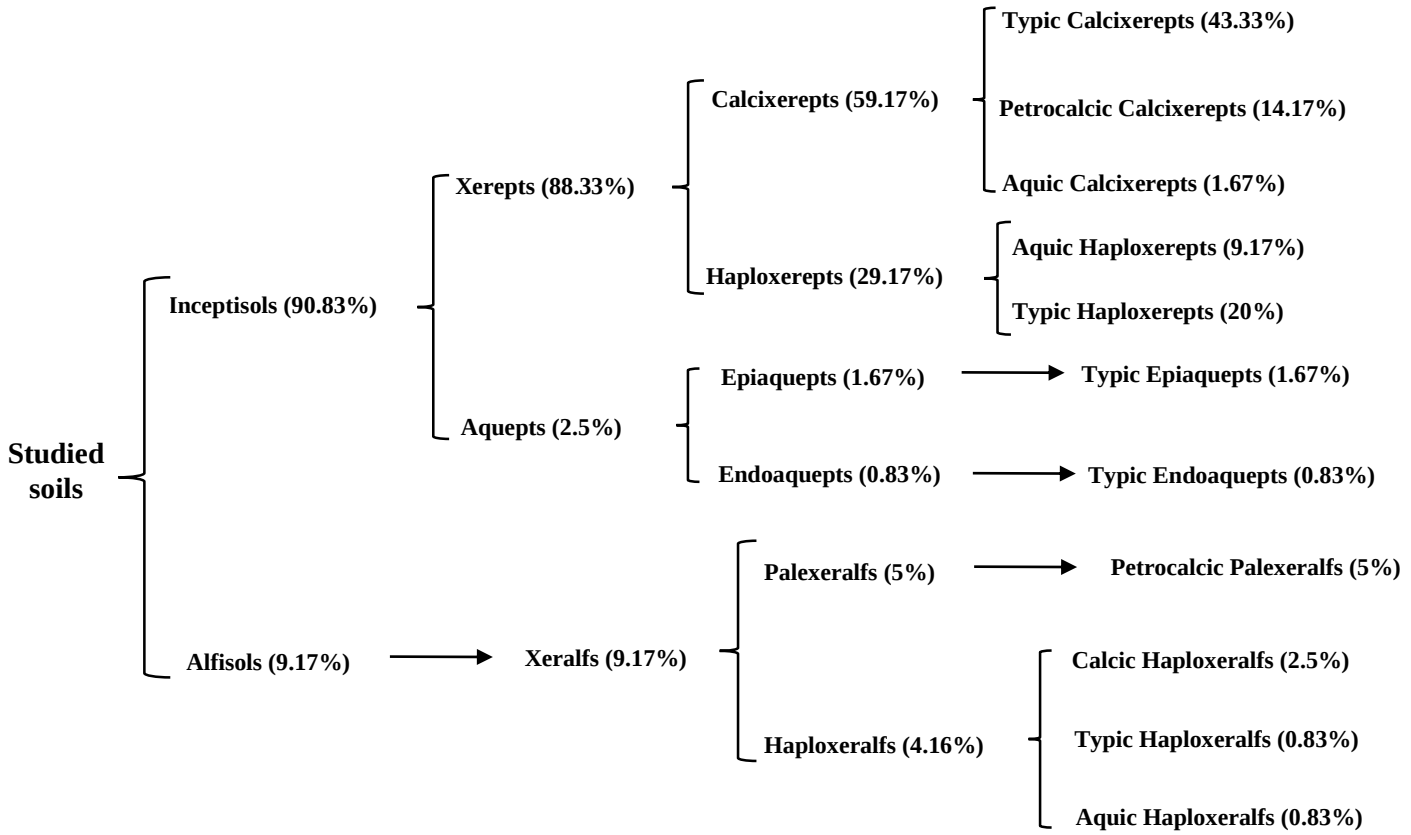

Fig. 2. The abundance of soil classes to subgroup level in the study area

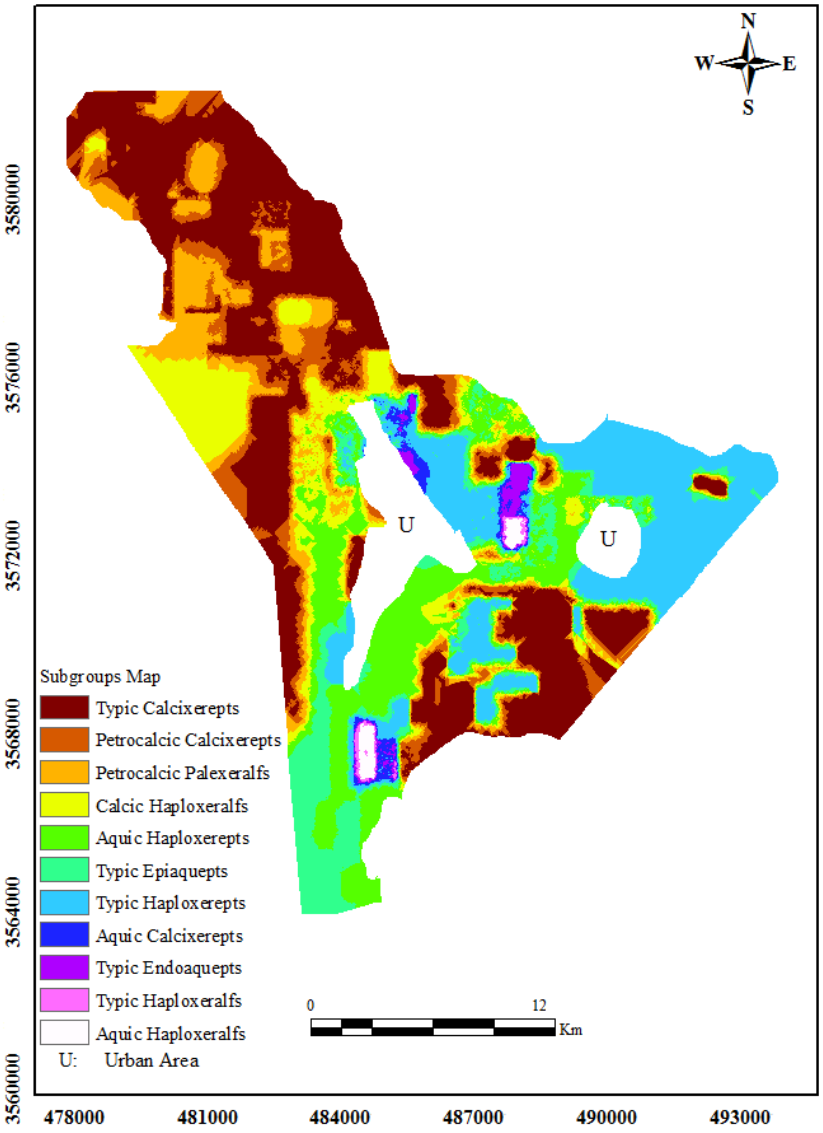

Fig. 3. Soil subgroups distribution map in the study area

3.2. Soil Classes Prediction by Auxiliary Parameters

Table 1 shows the sensitivity analysis results of the prediction of soil classes by auxiliary parameters in different levels (up to family level) of the ST system. As it was mentioned, each property with a sensitivity value of more than 1 is important in predicting soil classes. For example, at the order level, two properties of surface relief, such as aspect and analytical hill shading, are the most important for the prediction of soil classes. In the decision tree structure (its scheme for the decision tree has not been showed), other than the mentioned properties, some properties like catchment area, convergence index, soil map, landform, and elevation have also affected the prediction of soil classes at order level.

In the suborder level which presence or absence of argillic horizon and aquic condition were used for classification of soils at this level, some properties including cross-sectional curvature, slope, analytical hill shading, closed depressions, catchment area, and channel-network base level were the most effective properties in predicting soil classes. On the other hand, in the tree structure of suborder prediction (the scheme of the decision tree has not been shown), some properties like analytical hill shading, channel-network base level, catchment area, cross-section curvature, and slope have been located on the higher branches. Therefore, they are considered as the most effective properties of the suborder prediction. These properties are related to surface relief and can be in accordance with the formation of argillic horizon, and aquic moisture regime in the low land positions of the study area.

Table 1. Sensitivity analysis results for predicting soil classes by auxiliary parameters

Auxiliary parameters	Soil Taxonomy levels				
	Order	Suborder	Great group	Subgroup	Family
Geologic map	1.00	1.00	1.00	1.00	1.00
Landform map	1.00	1.00	1.00	1.00	1.00
Landform phase map	1.00	1.00	1.00	0.86	1.00
Soil map	1.00	1.00	1.00	1.00	1.00
NDVI	1.00	1.00	1.00	0.92	1.02
Longitudinal curvature	1.00	1.00	1.00	1.03	1.00
Cross-sectional curvature	1.00	3.33	1.42	1.00	1.00
Aspect	1.20	1.00	1.17	0.89	0.90
Elevation	1.00	1.00	1.00	1.00	1.02
Slope	1.00	1.67	1.00	1.00	1.00
Analytical hill shading	1.40	3.00	1.08	0.95	1.00
Convergence index	1.00	1.00	1.00	1.00	0.97
Closed depressions	0.40	1.33	1.00	0.81	1.00
Catchment area	1.00	2.66	1.08	1.03	1.00
Topographic wetness index	1.00	1.00	1.08	0.92	1.03
LS factor	1.00	1.00	1.00	1.00	1.00
Channel-network base level	1.00	2.00	0.83	0.84	1.00
Vertical distance to channel network	1.00	1.00	1.00	0.97	0.98
Valley depth	1.00	1.00	1.00	1.00	0.93
Relative slope position	1.00	1.00	1.08	1.00	1.00

Also, it is observed that for the great group and subgroup levels, DEM attributes have been the most effective properties in predicting soil classes (Table 1). Moore *et al*. (1991) expressed that the reason for the effectiveness of DEM and its derivatives in predicting soil classes was its description of soil moisture, soil erosion, and the conditions of its sedimentation by these derivatives. Thompson *et al*. (2001) described the advantages of the soil-landscape models as showing soil properties in the entire landscape continuously, quantifying the effects of environmental factors such as topography on the soil properties, and spatial prediction of soil properties throughout the landscape, even in areas which have not been sampled. Wilson (2012) has recognized the following points: the importance of climate elevation on vegetation and potential energy, the importance of slope on the amount of sediment, the flow velocity of surface and undersurface water, the amount of runoff, the content of soil water, the flow direction, sunny hours, the amount of evaporation, and the abundance of floral and faunal population. He also considered the topographic wetness index as a saturated area with runoff, and regarded it as a function of the transferability of soil and the amount of slope. Maynard and Johnson (2014) defined terrain curvatures like longitudinal curvature and cross-sectional curvature as the amount of slope changes in a special direction and considered it very important in determining surface distribution and undersurface water. Machado *et al*. (2018) expressed that auxiliary parameters including landform, slope and wetness index can be useful in predicting soil classes. Mirakzehi *et al*. (2018) prepared a digital soil map of the region of Sistan using the RF model, and concluded that the channel networks, valley depth, convergence, NDVI, and catchment area were one of the most important covariates. Bagheri Bodaghabadi *et al*. (2015) also predicted the soil classes of Borujen in Chaharmahal and Bakhtiari province, using the ANN model. He stated that the use of relief attributes was sufficient in achieving good prediction results.

At the family level, which mainly uses the soil's inherent characteristics (such as texture and cation exchange capacity) in classification, in addition to the topographic characteristics (i.e., elevation and topographic wetness index), the NDVI feature was also influenced in predicting soil family classes (Table 1). Figure 4 shows the scheme of the decision tree in predicting the soil family classes, which confirms the above conclusion. Mosleh *et al*. (2017) declared NDVI as the most important auxiliary parameter for predicting soil classes by multinomial logistic regression (MLR) and artificial neural network (ANN) models.

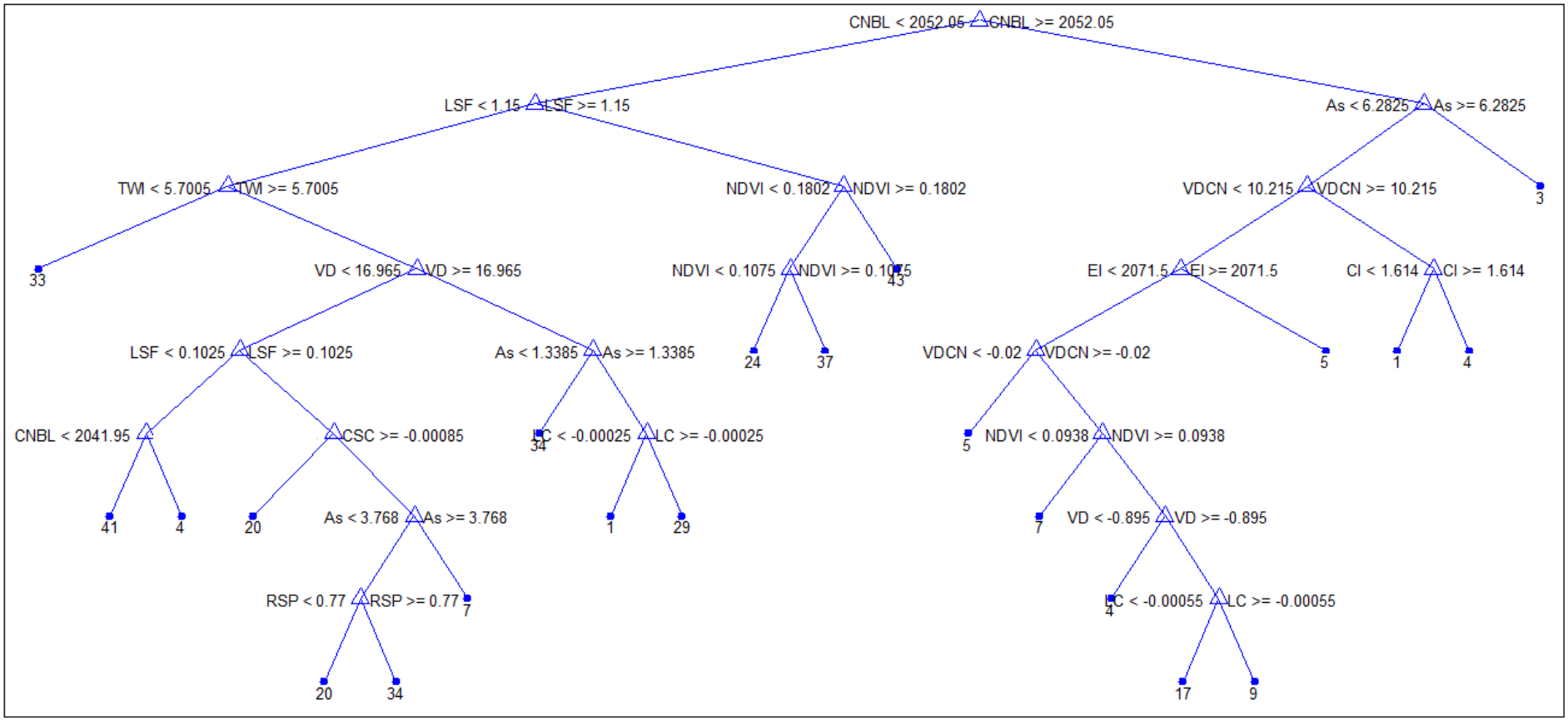

Fig. 4. Scheme of decision tree for predicting soil family classes using auxiliary parameters
CNBL: Channel-network base level; As: Aspect; LSF: LS factor; VDCN: Vertical distance to channel network; CI: Convergence index; El: Elevation; VD: Valley depth; LC: Longitudinal curvature; CSC: Cross-sectional curvature; RSP: Relative slope position; NDVI: Normalized difference vegetation index

3.3. Soil Classes Prediction by Soil Properties

As shown in Table 2, the properties of the presence and absence of argillic, calcic, and petrocalcic horizons have been the most effective properties for predicting the soil great groups. In addition, these properties, with the presence and absence of chroma equal to 1-2, were the most effective for predicting the subgroup level. These results are fully consistent with the principles of Keys to Soil Taxonomy (2014), which were used to classify the studied pedons. For example, in the great group level, about 60% of the pedons were located in the Calcixerepts class (Figure 2) due to the presence of calcic horizon in these pedons. According to the results of the sensitivity analysis (Table 2), the corresponding number of this horizon is 9 for the great group level and it is 7.5 for the subgroup level.

Table 2. Sensitivity analysis results for predicting soil classes using soil properties

Soil properties	Soil Taxonomy levels				
	Order	Suborder	Great group	Subgroup	Family
Calcic horizon	-	-	9	7.05	1.14
Cambic horizon	-	-	1	1.00	1.00
Argillic horizon	Inf	Inf	11	3.50	1.06
Secondary carbonates	-	-	1	1.00	1.00
Xeric	-	-	1	1.00	1.00
Chroma 1-2	-	-	1	2.00	1.00
Chroma <=1	-	Inf	1	1.00	1.00
Petrocalcic horizon	-	-	4	5.70	1.08
Shallow depth	-	-	-	-	1.00
Subactive	-	-	-	-	1.00
Semiactive	-	-	-	-	1.00
Active	-	-	-	-	1.03
Superactive	-	-	-	-	1.00
Carbonatic	-	-	-	-	1.00
Silty	-	-	-	-	0.94
Loamy	-	-	-	-	1.00
Less than 60 percent clay (Fine)	-	-	-	-	1.06
Skeletal	-	-	-	-	1.08
Less than 18 percent clay (Coarse)	-	-	-	-	1.00
18 to less than 35 percent clay (Fine)	-	-	-	-	1.00
Clayey	-	-	-	-	1.00

- : shows nonuse of the considered properties for predicting soil class in the relevant level.
Inf : shows a condition that based on it the rate of prediction error is equal to zero (with special properties).

In addition, at the order level, the presence and absence of the argillic horizon has caused about 9% of the soils to fall into the Alfisols, and the rest (91%) are located into the Inceptisols (Fig. 2). For this reason, the sensitivity analysis value of this horizon is 11 for the great group level and it is 3.5 for the subgroup level. But in the suborder level, the presence and absence of aquic conditions (chroma 1 or less) have caused 2.5% of the studied soils to be included in Aquepts suborder and the rest being located in Xerepts and Xeralfs suborder classes (Figure 2). Therefore, the decision tree model is not able to show the effect of aquic conditions in predicting the great group and subgroup levels. This can be due to the low number of affected points by this feature (only 2.5%), and increasing inputs to predict of great group and subgroup levels. In spite of this matter, the error amount is 0.83% for the prediction of the great group (Table 4). This means that the model can predict those points under the effect of the presence and absence of aquic conditions based on other inputs and the model predicted only one point wrongly. The point was 105, which has been located in the Endoaquepts class, but was incorrectly predicted under the Epiaquepts class.

The amount of error for subgroup prediction is 3.33% (Table 4). This means that 4 out of the 120 points, including points number 93 (Typic Epiaquepts), 97 (Typic Calcixerepts), 105 (Typic Endoaquepts) and 120 (Aquic Haploxeralfs), have been wrongly predicted. It is apparent that for the subgroup level, only the aquic conditions caused errors.

The schematic of the decision tree to predict the great group and subgroup (the schematic of the decision tree for these levels is not presented) is also in accordance with the sensitivity analysis values, and the features of the presence and absence of argillic, calcic, and petrocalcic horizons as well as aquic moisture regime are the most effective properties for predicting the great group and subgroup level. These properties are located in the upper parts of the tree structure. A remarkable point in the tree structure of subgroup prediction is that the chroma (the aquic conditions), which did not show its effect in the sensitivity analysis, was located in the higher branches of the tree structure. Therefore, it has

been one of the most effective properties in predicting the subgroup level.

At the family level, the properties of the presence and absence of argillic, calcic, and petrocalcic horizons are still influential, with sensitivity values of more than one. This is entirely logical because those properties which were effective in the higher levels of the ST system (order, suborder, great group and subgroup levels), will also be effective in the lower levels of the system. In addition to these properties, the cation exchange capacity class (i.e., active) and the particle size distribution class (i.e., fine and skeletal) have been more important than the other properties for predicting soil family classes. This issue is also fully in line with the principles of Keys to ST (Soil Survey Staff, 2014), because most of the studied soils are differentiated at the family level based on the differences in the cation exchange capacity classes and the particle size distribution classes. Figure 5 shows the scheme of the tree structure in the prediction of soil family classes. It is observed that the cation exchange capacity class and particle size distribution class are in the higher branches. As a result, these properties are more effective in predicting soil family classes than the presence and absence of diagnostic horizons.

In regards to Table 4, the error value for predicting the soil family classes is 30%. This means that 36 out of the 120 points were predicted incorrectly. The predicted families differ in one or more properties with the observed families. Ten predicted families in the particle size distribution class, 4 cases in CEC class, 4 cases in mineralogy class, 5 cases in shallow depth class, and 24 cases in the higher levels than family class differ from observed families. 8 of the 24 cases in order level (these 8 cases were Alfisols which had been wrongly predicted as Inceptisols,), 4 cases in suborder level (2 cases were Xerepts but were predicted as Aquepts, and 2 cases were Aquepts which were wrongly predicted as Xerepts), 1 case in great group level (Calcixerept but wrongly predicted as Haploxerept), and 11 cases in subgroup level (11 cases were Aquic Haploxerepts which were wrongly predicted as Typic Haploxerepts) differ from the families of the observation points. This means that adding new inputs for predicting soil family classes disrupts the prediction of higher levels of family. Bagheri Bodaghabadi (2015) predicted soil classes by ANN and concluded that adding a new input variable sometimes disrupts the network and increases the error.

3.4. Soil Classes Prediction by Combining Auxiliary Parameters and Soil Properties

Table 3 shows the results of sensitivity analysis of predicting soil classes in different levels of the ST system by combining auxiliary parameters and soil properties. It was observed that among these properties, the presence and absence of argillic horizon have just been effective in predicting the order class. The error content equaled to zero, and in accordance with the StatSoft sensitivity analysis method, the odd denominator is zero, which is shown in Table 3 with the "Inf" mark. Other auxiliary parameters had little effect in this level. For the suborder level, the soil properties of the presence and absence of argillic horizon and aquic conditions had been effective in predicting soil classes (scheme of decision tree has not been shown).
In great group and subgroup levels, some soil properties such as the presence and absence of argillic, calcic, petrocalcic horizons, and chroma equal to 1-2 (only for subgroup level) are the most effective properties in predicting soil classes. The auxiliary parameters showed a very weak effect. The probable reason for this can be the relatively high effect of soil properties in areas with low relief variation, such as the Shahrekord plain, or an estimation of DEM derivatives, rather than real soil data. In addition to some soil properties, landform phase as an auxiliary parameter influences the lower branches of decision tree structure for prediction of soil classes at great group level (data not shown). Esfandiarpoor Borujeni *et al*. (2010) suggested using landform phase to improve the soil maps prepared by the geopedology method (Zinck, 1989). Landform property which is related to auxiliary parameters of the geo-form map is effective in the subgroup level (data not shown). Machado *et al*. (2018) concluded that auxiliary parameters of landform can be useful for predicting soil classes. For soil family level, the results of sensitivity analysis (Table 3) showed that some soil properties and NDVI are effective, while in addition to these properties in the structure of the decision tree (Figure 6), other parameters, like channel-network base level, landform, and longitudinal curvature, have also been effective.

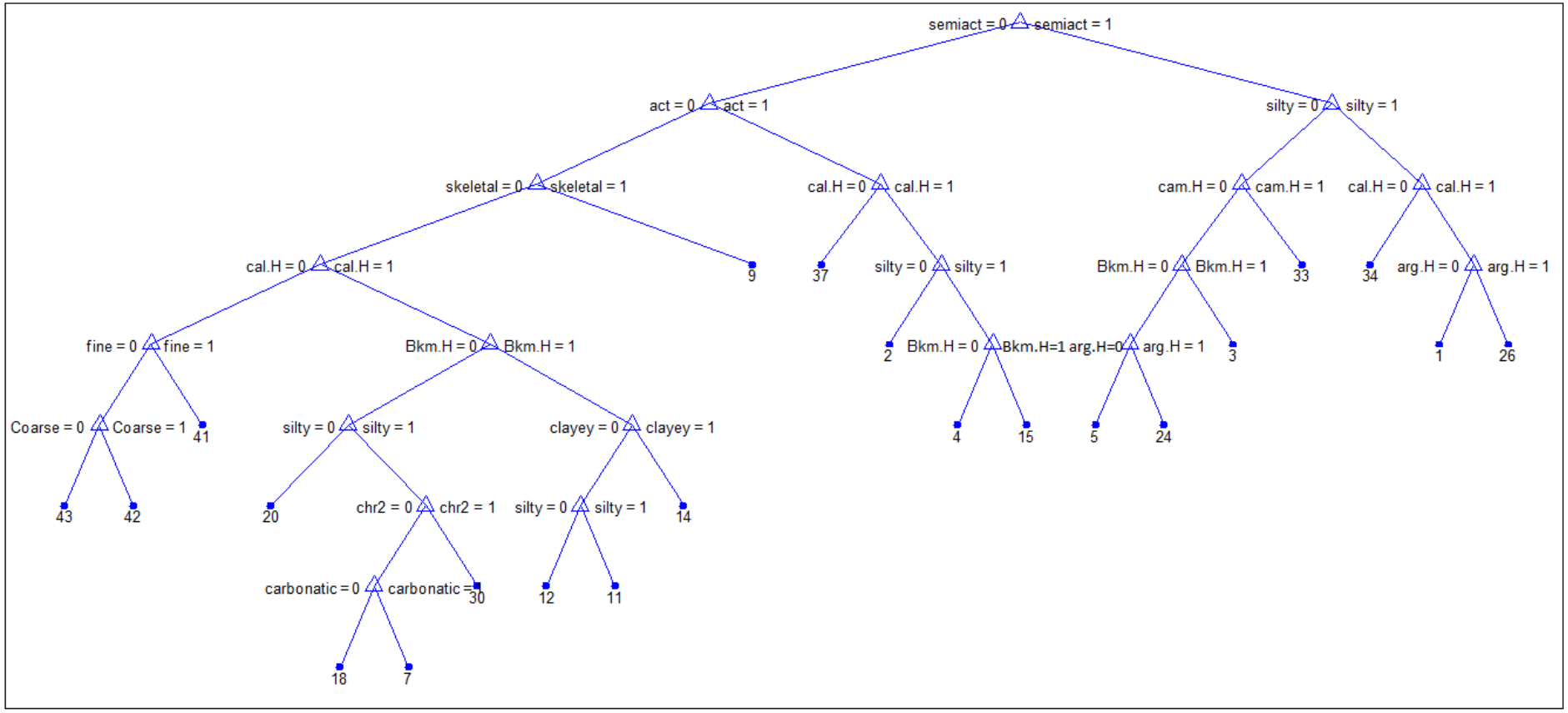

Fig. 5. Scheme of decision tree for predicting soil family classes using soil properties
The statements of semiact, act, silty, skeletal, coarse, fine, clayey and carbonatic show the presence and absence of classes of cation exchange activity of semiactive and active, particle size distribution classes of silty, skeletal, coarse, fine and clayey, and carbonatic mineralogy class, respectively. In addition, some signs such as cam. H, cal. H, Bkm. H and arg. H show presence and absence of cambic, calcic, petrocalcic and argillic horizon, respectively. The sign chr2 shows the presence and absence of chroma 2 or less

Table 3. Sensitivity analysis results for predicting soil classes using combining auxiliary parameters and soil properties

Properties' name	Soil Taxonomy levels				
	Order	Suborder	Great group	Subgroup	Family
Soil properties					
Calcic horizon	-	-	4	1.75	1.05
Cambic horizon	-	-	1	1.00	1.00
Argillic horizon	Inf	Inf	5	1.75	1.00
Secondary carbonates	-	-	1	1.00	1.00
Xeric	-	-	1	1.00	1.00
Chroma 1-2	-	-	1	1.75	1.00
Chroma <=1	-	Inf	1	1.00	1.00
Petrocalcic horizon	-	-	3	3.25	1.03
Shallow depth	-	-	-	-	1.00
Subactive	-	-	-	-	1.00
Semiactive	-	-	-	-	1.03
Active	-	-	-	-	1.05
Superactive	-	-	-	-	1.00
Carbonatic	-	-	-	-	1.00
Silty	-	-	-	-	0.97
Loamy	-	-	-	-	1.00
Less than 60 percent clay (Fine)	-	-	-	-	1.11
Skeletal	-	-	-	-	1.11
Less than 18 percent clay (Coarse)	-	-	-	-	1.00
18 to less than 35 percent clay (Fine)	-	-	-	-	1.00
Clayey	-	-	-	-	1.05
Auxiliary parameters					
Geologic map	0.10	0.43	1	1	1.00
Landform map	0.14	0.34	1	1	1.00
Landform phase map	0.20	0.17	1	1	1.00
Soil map	0.25	0.35	1	1	1.00
NDVI	0.10	0.17	1	1	1.03
Longitudinal curvature	0.15	0.24	1	1	1.00
Cross-sectional curvature	0.30	0.28	1	1	1.00
Aspect	0.11	0.17	1	1	1.00
Elevation	0.42	0.33	1	1	1.00
Slope	0.33	0.41	1	1	1.00
Analytical hill shading	0.24	0.18	1	1	1.00
Convergence index	0.36	0.10	1	1	1.00
Closed depressions	0.18	0.31	1	1	1.00
Catchment area	0.10	0.27	1	1	1.00
Topographic wetness index	0.23	0.26	1	1	1.00
LS factor	0.44	0.37	1	1	0.95
Channel network base level	0.26	0.22	1	1	1.00
Vertical distance to channel network	0.16	0.10	1	1	1.00
Valley depth	0.14	0.16	1	1	1.00
Relative slope position	0.15	0.26	1	1	1.00

- : shows nonuse of the considered properties for predicting the soil class in the relevant level
Inf : shows a condition that based on it the rate of prediction error is equal to zero (with special properties)

Table 4 presents the comparison of prediction errors of different levels of ST system by using soil properties, auxiliary parameters and combination of both features.

Table 4. Amount of prediction error of soil classes at different levels of the ST system

Soil Taxonomy level	Percentage of prediction error		
	Soil properties	Auxiliary parameters	Combination of soil properties and auxiliary parameters
Order	0.00	3.33	0.00
Suborder	0.00	3.33	0.00
Great group	0.83	15.00	0.83
Subgroup	3.33	22.50	3.33
Family	30.00	52.50	30.00

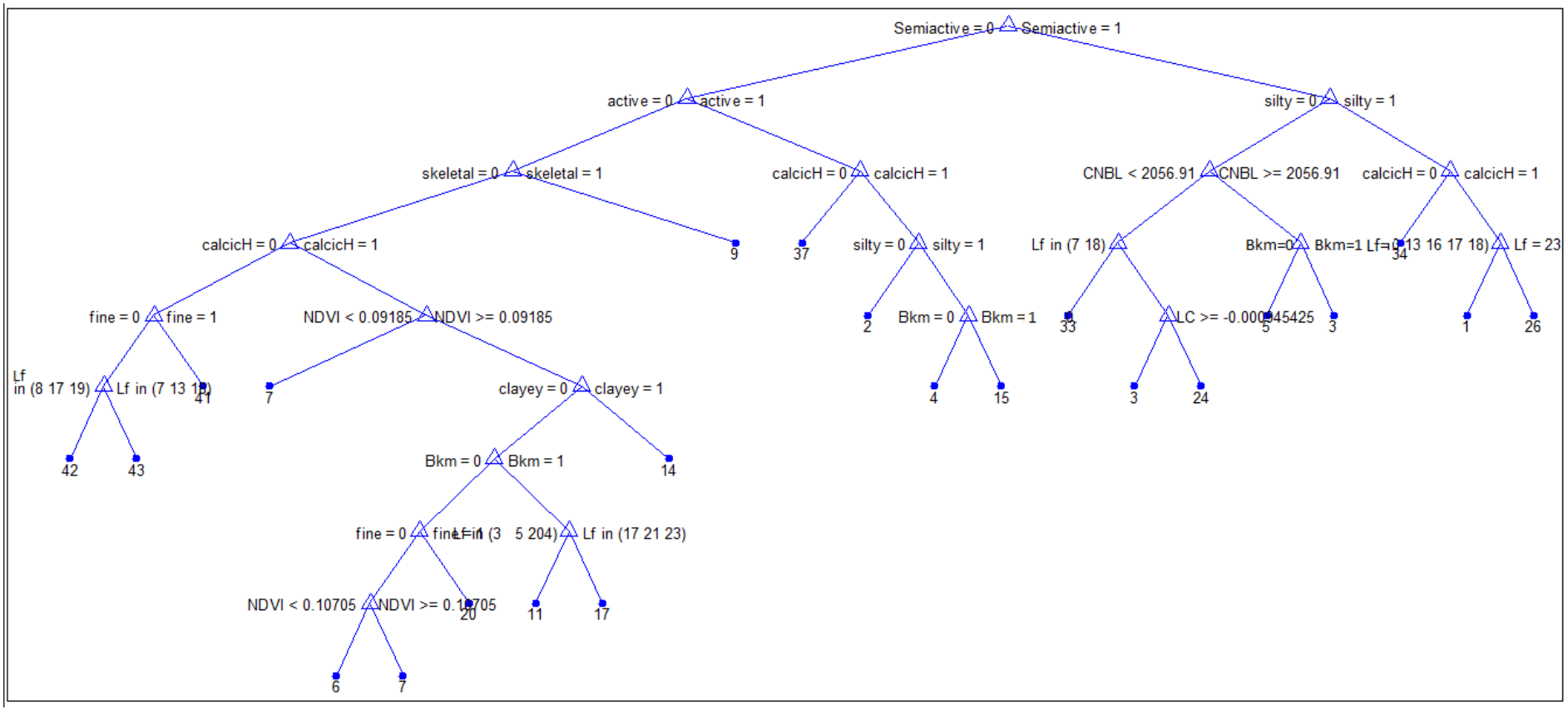

Fig. 6. Scheme of decision tree for soil family level prediction by combining auxiliary parameters and soil properties
CNBL: Channel-network base level; LC: Longitudinal curvature; LF: Landform map
The statements of semiactive, active, silty, skeletal, fine and clayey show the presence and absence of classes of cation exchange activity of semiactive and active and particle size distribution classes of silty, skeletal, fine and clayey, respectively. The signs of cal. H and Bkm. H show presence and absence of calcic and pertrocalcic horizons, respective

It is observed that the prediction error values have increased from upper levels of classification (order) to lower levels (family). The probable cause of this issue can be the entry of more details into the soil classification at lower levels of the ST system, which increase the number of classes. Besides, role of the soil properties which have used at each taxonomic level maybe affect this issue. Heung *et al*. (2016) suggested that the reason for the decrease in the overall accuracy from order to great group level is due to the increasing details in great group rather than order level. They also introduced the decision tree as a suitable method for when the number of predictable classes is low. Brungard *et al*. (2015) also considered the reason of the decrease in the overall accuracy of their maps, prepared by the decision tree model, the large number of soil classes that must be predicted. This issue has also been concluded by Mosleh et al. (2017) and Taghizadeh-Mehrjardi *et al*. (2015). In addition, the calculated errors in Table 4 can implicitly show the high efficiency of the decision tree in applying the qualitative features (such as the presence or absence of a diagnostic horizon or property) to predict soil classes. In other words, a good agreement between the performed decision tree with the rules of Keys to ST (Soil Survey Staff, 2014) is understandable by this way.

The amount of computed prediction error for conditions which have just used soil properties is lower than that of which have used the auxiliary parameters. This is shown in Table 4. Additionally, the amount of prediction error of the soil classes with the combined application of soil properties and auxiliary parameters are similar to those of the soil properties. With simultaneous application of soil properties and auxiliary parameters, the effect of soil qualitative properties in predicting soil classes were so high that the auxiliary parameters failed to show its effect in predicting soil classes.

3.5. Soil Classes Prediction Using Kriging Maps of Soil Properties

To predict soil classes using kriging maps, the previously used properties to estimate soil classes (Table 2) were solely used. As shown in Table 5, the prediction error values had increased rather than the soil properties (Table 4) of the great group and subgroup levels. However, the error rate had decreased only in the family level. The probable cause of this reduction in error is the use of quantitative properties for the prediction of the soil family classes.

Table 5. Prediction error percentage of soil classes using kriging maps of soil properties

Soil Taxonomy level	prediction error by previous inputs (Table 2)	prediction error by new inputs (qualitative and quantitative)
Order	0.000	0.000
Suborder	0.000	0.000
Great group	2.880	0.002
Subgroup	9.530	0.006
Family	0.014	0.014

In order to reduce the prediction error for higher levels of classification, in addition to the qualitative characteristics, quantitative properties including calcium carbonate percentage, sand percentage and clay percentage were considered as inputs. Table 6 shows the sensitivity analysis values of using kriging maps of soil properties for predicting soil classes at different levels of the ST system.

According to Table 6, the presence and absence of argillic horizon and aquic conditions were only effective for order and suborder levels, respectively. The presence and absence of calcic horizon and the percentage of sand and clay were effective for great group level. The presence and absence of argillic, calcic, petrocalcic horizons, aquic conditions, chroma of 1-2, and the percentage of sand, clay, and calcium carbonate equivalents were effective in the subgroup level. To predict the soil family classes based on Keys to ST (Soil Survey Staff, 2014), a number of new properties, such as the percentage of a particle size greater than 1mm, percentage of particle size greater than 2mm, percentage of clay-size carbonates, cation exchange capacity, and the presence and absence of a root-limiting layer at a 50cm depth from the mineral soil surface (as many as 125000 records) were also added. It has been observed that, in addition to these properties, the presence and absence of argillic, calcic, cambic, petrocalcic horizons, aquic conditions and the percentage of clay and calcium carbonate equivalents were effective for predicting soil family classes. These results are entirely consistent with the principles of the ST System. In other words, the same properties used for the classification of pedons were considered as influential in predicting soil classes.

As shown in Table 5, the input of kriging maps of quantitative data to the decision tree model at higher levels of soil family, significantly reduced the amount of prediction error for great group and subgroup levels. Thus, the error value at the great group level dropped from 2.88 to 0.002% and the value in the subgroup level decreased from 9.53 to 0.006%.

Of the 125,000 records in the great group, subgroup, and family levels, only 2, 8, and 17 records were incorrectly predicted, respectively.

Table 6. Sensitivity analysis results for predicting soil classes using kriging maps of soil properties

Properties' name	Soil Taxonomy levels				
	Order	Suborder	Great group	Subgroup	Family
Calcic horizon	0.25	0.34	2	1.30	1.18
Cambic horizon	0.36	0.36	1	1.00	1.06
Argillic horizon	Inf	Inf	1	1.63	1.37
Secondary carbonates	0.29	0.32	0	1.00	0.87
Chroma 1-2	0.18	0.36	1	1.50	1.00
Chroma <=1	0.27	Inf	1	1.63	1.12
Petrocalcic horizon	0.43	0.28	1	3.25	1.19
Shallow depth	-	-	-	-	0.94
Sand	-	-	2	15.50	0.87
Clay	-	-	2	16.75	1.62
$CaCO_3$	-	-	1	12.75	1.31
clay-size carbonates	-	-	-	-	1.81
Particles more than 1 mm	-	-	-	-	1.44
CEC	-	-	-	-	1.37
Rock fragments (≥ 2 mm)	-	-	-	-	1.19

- : shows nonuse of the considered properties for predicting the soil class in the relevant level.
Inf : shows a condition that based on it the rate of prediction error is equal to zero (with special properties).

4. Conclusion

The results of this study showed that the decision tree model had a good performance in predicting soil classes at different levels of the ST system in the Shahrekord plain. The usage of the StatSoft sensitivity analysis method and the decision tree model well illustrated the effects of different inputs in predicting soil classes. It was seen that the prediction error values have increased from upper levels of classification (order) to lower levels (family). The probable cause of this issue can be the entry of more details into the soil classification at lower levels of the ST system, which increase the number of classes. It was also observed that the soil class prediction error along with the simultaneous use of soil properties and auxiliary parameters were similar to the soil class prediction error using soil properties. Auxiliary parameters did not influence the decrease of the soil class prediction error. In other words, the effect of soil qualitative properties in predicting soil classes was considerable and the auxiliary parameters could not properly show their effects in this regard. Therefore, it seemed that the effect of auxiliary parameters, relative to soil properties, in predicting soil classes in low relief variation areas were not significant. The usage of kriging maps for quantitative and qualitative properties of soil in the decision tree model resulted in a significant reduction in the prediction error of soil classes. Due to the fact that low relief areas (plains) are widely used for agricultural purposes and that there is a strong demand for accurate information on its soil and the variability of these regions, the results of this research can respond to users' needs in order to understand these variations.

References

Adhikari, K., A.E. Hartemink, 2016. Linking soils to ecosystem services - A global review. Geoderma, 262; 101–111.

Bagheri Bodaghabadi, M., M.H. Salehi, J.A. Martínez-Casasnovas, J. Mohammadi, N. Toomanian, I. Esfandiarpoor Borujeni, 2011. Using Canonical Correspondence Analysis (CCA) to identify the most important DEM attributes for digital soil mapping applications. Catena, 86; 66–74.

Bagheri Bodaghabadi, M., J.A. Martínez-Casasnovas, M.H. Salehi, J. Mohammadi, I. Esfandiarpoor Borujeni, N. Toomanian, A. Gandomkar, 2015. Digital Soil Mapping using Artificial Neuronal Networks (ANN) and Terrain-Modelling Attributes. Pedosphere, 25; 580-591.

Bockheim, J.G., A.E. Hartemink, 2013. Distribution and classification of soils with clay-enriched horizons in the USA. Geoderma, 209–210; 153–160.

Brungard, C.W., J.L. Boettinger, M.C. Duniway, S.A. Wills, T.C. Edwards, 2015. Machine learning for predicting soil classes in three semi-arid landscapes. Geoderma, 239-240; 68–83.

Das, M.D., 2009. Principles of Geotechnical Engineering (7^{th} ed.), Cengage Learning, Stamford, CT.

Elliott, P.E., P.J. Drohan, 2009. Clay accumulation and argillic-horizon development as influenced by aeolian deposition vs. local parent material on quartzite and limestone-derived alluvial fans. Geoderma, 151; 98–108.

Esfandiarpoor Borujeni I., J. Mohammadi, M.H. Salehi, N. Toomanian, R.M. Poch, 2010. Assessing geopedological soil mapping approach by statistical and geostatistical methods: A case study in the Borujen region, Central Iran. Catena, 82; 1–14.

Geological survey and mineral exploration of Iran. 2017. http://www.gsi.ir.

Goodman, J. M., P. R. Owens, 2012. Predicting soil organic carbon using mixed conceptual and geostatistical models. In: B. Minasny, B. P.Malone, A. B. McBratney (eds), Digital soil assessments and beyond (pp. 155–159). London: CRC Press.

Gunal, H., M.D. Ransom, 2006. Clay illuviation and calcium carbonate accumulation along a precipitation gradient in Kansas. Catena, 68; 59–69.

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.

Holmes, K.W., E.A. Griffin, N.P. Odgers, 2015. Large-area spatial disaggregation of a mosaic of conventional soil maps: evaluation over Western Australia. Soil Research, 53; 865–880.

Jafari, A., P.A. Finke, J. VandeWauw, S. Ayoubi, H. Khademi, 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; 284–298.

Kalavathi, K., P.V. Nimitha Safar, 2015. Performance Comparison between Naive Bayes, Decision Tree and k-Nearest Neighbor. International Journal of Emerging Research in Management and Technology, 4; 152-161.

Khormali, F., S. Ghergherechi, M. Kehl, S. Ayoubi, 2012. Soil formation in loess-derived soils along a subhumid to humid climate gradient, Northeastern Iran. Geoderma, 179–180; 113–122.

Lagacherie, P., S. Holmes, 1997. Addressing geographical data errors in a classification tree for soil unit prediction. International Journal of Geographical Information Science, 11; 183–198.

Machado, I.R., E. Giasson, A.R. Campos, J. Janderson, F. Costa, E.B. Silva, B.R. Bonfatti, 2018. Spatial Disaggregation of Multi-Component Soil Map Units Using Legacy Data and a Tree-Based Algorithm in Southern Brazil. Rev Bras Cienc Solo; 42: e0170193.

Massawe, B.H.J., S.K. Subburayalu, A.K. Kaaya, L. Winowiecki, B.K. Slater, 2018. Mapping numerically classified soil taxa in Kilombero Valley, Tanzania using machine learning. Geoderma, 311; 143–148.

Maynard, J.J., M.G. Johnson, 2014. Scale-dependency of LiDAR derived terrain attributes in quantitative soil-landscape modeling: Effects of grid resolution vs. neighborhood extent. Geoderma, 230–231; 29–40.

McBratney, A. B., M. L. Mendonc, B. Minasny, 2003. On digital soil mapping. Geoderma, 117; 3–52.

Mirakzehi, K., M. Pahlavan-Rad, A. Shahriari, 2018. Digital soil mapping of deltaic soils: A case of study from Hirmand (Helmand) river delta. Geoderma, 313; 233–240.

Mohammadi, M., 1986. Semi-detailed soil studies report Chaharmahal-Va-Bakhtiari province (Shahrekord and Borujen area). Tehran, Iran. Iranian Soil and Water Research Institute.

Moore, ID., R.B. Grayson, A.R. Ladson, 1991. Digital terrain modelling: a review of hydrological. geomorphological and biological applications. Hydrol Process, 5; 3-30.

Mosleh, Z., M.H. Salehi, A. Jafari, I. Esfandiarpoor Borujeni, A. Mehnatkesh, 2017. Identifying sources of soil classes variations with digital soil mapping approaches in the Shahrekord plain, Iran. Environ Earth Sci, 76; 748p.

Odgers, N.P., W. Sun, A.B. McBratney, B. Minasny, D. Clifford, 2014. Disaggregating and harmonising soil map units through resampled classification trees. Geoderma, 214; 91–100.

Olaya, V. F., 2004. A gentle introduction to SAGA GIS. User Manual. Germany, DC; Gottingen.

Rossiter, D. G., 2000. Methodology for soil resource inventories. Lecture notes. 2nd revised version. Enschede, The Netherlands: Soil Science Division, International Institute for Aerospace Survey and Earth Science (ITC).

Rouse, J. W., R. H. Hass, J. A. Schell, D.W. Deering, 1974. Monitoring vegetation systems in the Great Plains with ERTS. Proceedings of 3rd Earth Resource Technology Satellite (ERTS) Symposium, 1; 48-62.

Saunders, A. M., J. L. Boettinger, 2007. Incorporating classification trees into a pedogenic understanding raster classification methodology, Green River Basin, Wyoming, USA. In: P. Lagacherie McBratney A. B., Voltz M. (ed.), Digital Soil Mapping: An introductory perspective. Developments in Soil Science. Elsevier, Amsterdam, 31; 389-399.

Scull, P., J. Franklin, O.A. Chadwick, 2005. The application of classification tree analysis to soil type prediction in a desert landscape. Ecological Modelling, 181; 1–15.

Soil Survey Division Staff, 1993. Soil Survey Manual. Soil Conservation Service, U.S. Department of Agriculture Handbook 18 (Chapter 3).

Soil Survey Staff, 2014. Soil taxonomy: a basic systems of soil classification for making and interpreting soil surveys (12th ed.). USDA; NRCS.

Taghizadeh-Mehrjardi, R., B. Minasny, J. Triantafilis, F. Sarmadian, M. Omid, 2014. Digital mapping of soil classes using decision tree and auxiliary data in theArdakan region, Iran. Arid Land Research and Management, 42; 225-237.

Taghizadeh-Mehrjardi, R., K. Nabiollahi, B. Minasny, J. Triantafilis, 2015. Comparing data mining classifiers to predict spatial distribution of USDA-family soil groups in Baneh region. Iran. Geoderma, 253–254; 67–77.

Thompson, J.A., J.C. Bell, C.A. Butler, 2001. Digital elevation model resolution: effects on terrain attribute calculation and quantitative soil-landscape modeling. Geoderma, 100; 67–89.

US Geology Survey, 2016. Geology.com/news/2010/free-Landsat images- from-USGS-2. (http://glovis.usgs.gov).

Wilson, J.P. 2012. Digital terrain modeling. Geomorphology, 137; 107–121.

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; 86–93.

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