

## Assessing agricultural salt-affected land using digital soil mapping and hybridized random forests

Kamal Nabiollahi <sup>a,\*<sup>1</sup></sup>, Ruhollah Taghizadeh-Mehrjardi <sup>b,c,1</sup>, Aram Shahabi <sup>a</sup>, Brandon Heung <sup>d</sup>, Alireza Amirian-Chakan <sup>e</sup>, Masoud Davari <sup>a</sup>, Thomas Scholten <sup>c,f,g</sup>

<sup>a</sup> Department of Soil Science and Engineering, Faculty of Agriculture, University of Kurdistan, Sanandaj, Iran

<sup>b</sup> Faculty of Agriculture and Natural Resources, Ardakan University, Iran

<sup>c</sup> Department of Geosciences, Soil Science and Geomorphology, University of Tübingen, Tübingen, Germany

<sup>d</sup> Department of Plant, Food, and Environmental Sciences, Faculty of Agriculture, Dalhousie University, Canada

<sup>e</sup> Department of Soil Science, Lorestan University, Khorramabad 6815144316, Iran

<sup>f</sup> DFG Cluster of Excellence "Machine Learning", University of Tübingen, Germany

<sup>g</sup> CRC 1070 Resource Cultures, University of Tübingen, Gartenstr. 29, Tübingen, Germany

### ARTICLE INFO

Handling Editor: Budiman Minasny

**Keywords:**

Digital soil mapping

Random forest

Optimization algorithms

Soil salinity

Soil sodicity

### ABSTRACT

Salinization and alkalization are predominant environmental problem world-wide which their accurate assessment is essential for determining appropriate ways to deal with land degradation, for better soil and crop management. In the current research, a combination of random forests and covariate data were used to assess spatial variability of soil salinity and sodicity in 436 km<sup>2</sup> agricultural salt-affected land in Kurdistan Province, Iran. Using the conditioned Latin hypercube sampling method, 295 soil samples were sampled across the study area, and then soil reaction (pH), electrical conductivity (EC), and sodium adsorption ratio (SAR) were measured. Covariate data including terrain attributes, remotely-sensed data, groundwater table, and categorical maps were acquired. Random forest (RF) models were used to predict the spatial distribution of pH, EC, and SAR by making a relationship between soil data and covariates. Furthermore, three optimization algorithms (particle swarm optimization-PSO, genetic algorithm-GA, and bat algorithm-BAT) were used to explore if the hybridized RF works better than the standard RF. Results of 10-fold cross-validation with 100 replications indicated that the accuracy of RF + PSO was higher for predicting pH (RMSE = 0.52 and R<sup>2</sup> = 0.67), EC (RMSE = 2.32 dSm<sup>-1</sup> and R<sup>2</sup> = 0.57), and SAR (RMSE = 8.98 and R<sup>2</sup> = 0.54, respectively) in comparison to the other implemented models. Furthermore, the results disclosed that the most important covariates to predict pH, EC, and SAR were groundwater table, categorical maps, salinity index, and multi-resolution ridge top flatness. Besides, the results indicated that the mean values for pH, EC, and SAR in lowland and bare land were significantly different from the other physiographic units and land uses, respectively. Importantly, the classified map of salt-affected soils highlighted areas with a high risk of exceeding critical threshold values of pH, EC, and SAR, which is located in the center of the study area, and showed that 6.30%, 3.1%, and 4.6% of the study area are saline-sodic soil, saline soil, and sodic soil, respectively. These up to date spatial soil information on severity of soil salinity and sodicity is crucial for agricultural management of affected areas and the proposed method can be used to the other similar regions.

### 1. Introduction

Approximately, 90 percent of Iran is arid and semi-arid (Qadir et al., 2008) and processes of soil salinity and sodicity are influencing the land

degradation in arid and semi-arid area of the world (Abuelgasim and Ammad, 2019; Kumar et al., 2018; Wang et al., 2019a, 2019b), which have a negative effect on agricultural production. In these areas, evapotranspiration exceeds precipitation, which results in the salts

\* Corresponding author.

E-mail addresses: [k.nabiollahi@uok.ac.ir](mailto:k.nabiollahi@uok.ac.ir) (K. Nabiollahi), [taghizadeh-mehrjardi@mnf.uni-tuebingen.de](mailto>taghizadeh-mehrjardi@mnf.uni-tuebingen.de) (R. Taghizadeh-Mehrjardi), [brandon.heung@dal.ca](mailto:brandon.heung@dal.ca) (B. Heung), [amirian.ar@lu.ac.ir](mailto:amirian.ar@lu.ac.ir) (A. Amirian-Chakan), [thomas.scholten@uni-tuebingen.de](mailto:thomas.scholten@uni-tuebingen.de) (T. Scholten).

<sup>1</sup> The first two authors contributed equally to this work.

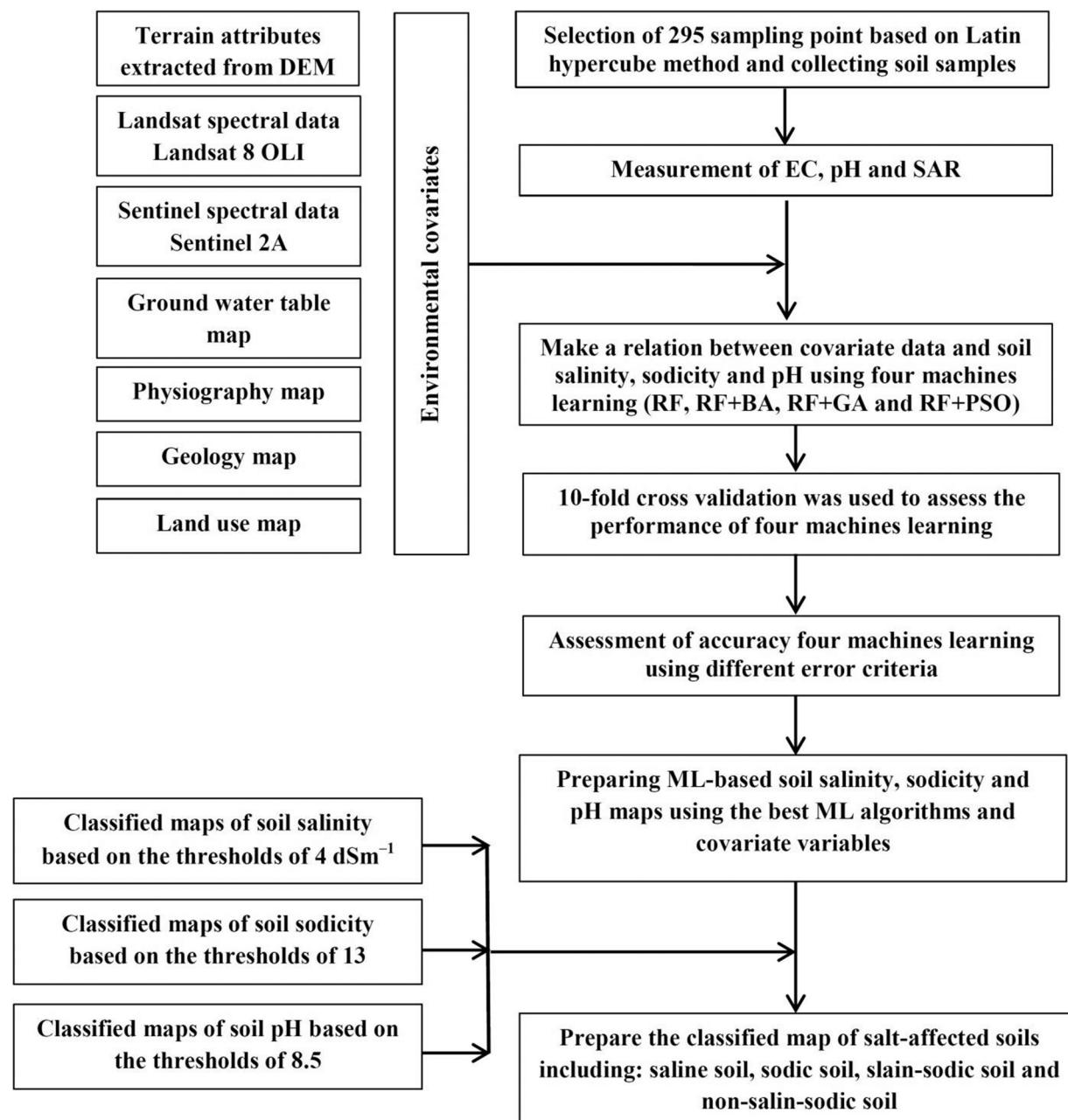


Fig. 1. Flowchart of methodology used in this study.

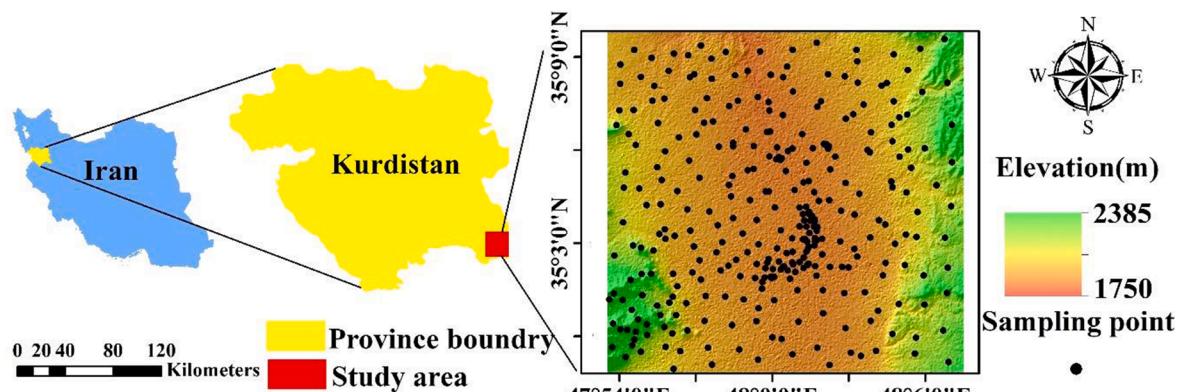
deposition in soil and therefore leading to increased salinity and decreased soil productivity. Moreover, the high content of sodium ion in the soil causes the dispersion of soil components and acceleration of soil aggregates degradation and erosion (Farifteh et al., 2006). Saline soils are specified as having high electrical conductivity (EC;  $>4 \text{ dSm}^{-1}$ ) and low sodium adsorption ratio (SAR;  $<13$ ) values (Richards, 1954). In contrast, alkaline soils are characterized as having low EC ( $<4 \text{ dSm}^{-1}$ ) and high SAR ( $>13$ ) values (Richards, 1954).

The mapping and modeling of soil characteristics is vital to facilitating the mitigation of soil degradation and awareness the processes and factors influencing soil properties (Bogunovic et al., 2018; Pereira et al., 2017; Villarino et al., 2017). Therefore, the surveying and mapping of salt-affected soils also can help improve their management of these soils; however, there is a lack in spatial soil information and associated information in Iran because of the availability of national surveys remains limited (Jafari et al., 2014; Nabiollahi et al., 2018a; Taghizadeh-Mehrjardi et al., 2016b; Yang, 2013). Determination of soil

characteristics through sampling and laboratory analyses is both time consuming and costly (McBratney et al., 2003).

Digital soil mapping (DSM) techniques include a suite of useful tools that facilitate soil mapping at the large-scale in data-poor regions. In DSM, machine learning (ML) algorithms have been used to predict various soil properties using creation of the relationships between soil observations and environmental covariates (Malone et al., 2016; McBratney et al., 2003; Minasny and McBratney, 2016; Pahlavan-Rad and Akbarimoghaddam, 2018; Teng et al., 2018). These covariates can be obtained from digital elevation model, remote sensing data, and other geo-spatial data sources (McBratney et al., 2003; Nabiollahi et al., 2018b, 2019; Taghizadeh-Mehrjardi et al., 2015; Zeraatpisheh et al., 2020).

Different machine learning (ML) models have been used within a DSM approach to predict and map soil characteristics of salt-affected areas around the world (Fathizad et al., 2020; Peng et al., 2018; Taghizadeh-Mehrjardi et al., 2016a; Wang et al., 2019c, 2020a). These



**Fig. 2.** Location map of the study area within the Kurdistan province and Iran.

ML models have included artificial neural networks (Shahabi et al., 2017), partial least square regression (Peng et al., 2018), support vector regression (Wu et al., 2018), stochastic gradient tree-boost (Wang et al., 2019c), regression tree models (Taghizadeh-Mehrjardi et al., 2014), genetic programming (Taghizadeh-Mehrjardi et al., 2016a), and random forests (Fathizad et al., 2020). Among those ML models, random forests (RF) or random decision forests, first developed by Breiman (2001), is the most popular method to predict and map soil characteristics and have successfully applied by many researchers in the last decade (Fathizad et al., 2020; Taghizadeh-Mehrjardi et al., 2020; Rentschler et al., 2019).

One challenge in training the machine learning models and identifying the optimal hyperparameter values (e.g. number of covariates tested at each splitting node, *mtry*, in RF), which may affect the accuracy when predicting complex problems (e.g., soil-landscape relationships). To find the right combination of hyperparameters, an optimization algorithm such as a genetic algorithm (GA), particle swarm optimization algorithm (PSO), artificial bee colony algorithm (ABC), and bat algorithm (BA) may be used to facilitate computation, functionality, robustness, and accuracy aspects of standalone ML models. For instance, Pham et al. (2020) used a hybrid model to predict shear strength of soil and reported that RF + PSO was superior to the single RF model without optimization. Similarly, Yang et al. (2017) obtained the best performance of ML models for predicting soil moisture when they are combined by the PSO optimization algorithm. Wang et al. (2017) introduced the RF + GA as a promising approach in predicting spatial distribution of SOC stocks from available continuous remotely sensed covariates. Similar to the study done by Taghizadeh-Mehrjardi et al. (2015), however, Wang et al. (2017) used GA for covariate selection rather than for optimizing the hyperparameters.

Though different ML models have been used in DSM, to best of our knowledge, there is no study that has carried out the mapping of soil salinity and soil sodicity using RF that has been hybridized with an optimization algorithm for arid and semi-arid regions. Specifically, in this research the objective was to assess, based on a case study, the performance of DSM methods using RF models hybridized with three optimization algorithms for the spatial prediction of three key soil properties (pH, EC, and SAR) of salt affected area. Furthermore, we explored the contribution of each covariate on the prediction to increase our awareness about the processes and factors influencing the soil properties. In addition, the relative influence of land uses, physiographic units, and geological surfaces on the spatial variability of soil salinity and sodicity are explored.

## 2. Materials and methods

The methodological framework for this study is summarized in Fig. 1. Sampling was carried out using the conditioned Latin hypercube

sampling approach (cLHS; Minasny et al., 2006) where measurements of pH, EC, and SAR were made (Section 2.2). A suite of environmental covariates was derived at a common spatial resolution (Section 2.3) and a training dataset was created by spatially intersecting the sampling locations with the environmental covariates and used to train the RF model (Section 2.4). Here, we coupled the RF model with various hyperparameter optimization techniques (Section 2.5) and validate the models using cross-validation (Section 2.6). Lastly, predictions of soil pH, EC, and SAR were produced and the results were reclassified to identify salt-affected areas (Section 2.7).

### 2.1. Study area

The study area (approximately 43674 ha) is located between 47° 53' 30.52" and 48° 07' 29.37" E longitudes and 34° 58' 51.41" and 35° 09' 52.10" N latitudes of the Ghorveh region, Kurdistan Province, west of Iran (Fig. 2). The climate is semi-arid and the mean annual temperature, precipitation, and elevation are 10.8 °C, 369.8 mm, and 2250 m above mean sea level, respectively. The soils are described as Xeric in terms of soil moisture and are located in a Mesic temperature regime (Soil Survey Staff, 2014). Soils in the region are described as predominantly Inceptisols and Entisols based on Soil Taxonomy, and Cambisols and Leptosols based on the World Reference Base for Soil Resources (IUSS, 2015).

### 2.2. Soil sampling and analysis

To capture the high spatial variability of the soil properties over a large spatial extent, a sampling strategy must be carefully considered where sampling sites were selected based on the variability of soil-forming factors that were expected to represent the landscape heterogeneity of the study area. The cLHS was used where 295 soil samples were collected at the 0–30 cm depth increment during the month of September in 2017 (Fig. 2). The samples were air-dried at room temperature (~25 °C), crushed, and passed through a 2-mm sieve prior to chemical analyses. The soil pH and EC were measured in a saturated paste using a pH electrode (McLean, 1982) and conductivity meter (Rhoades, 1982), respectively. Soluble calcium, magnesium and sodium were measured using the EDTA complex metric titration and flame photometric methods (Page, 1992; Jayachandran et al., 2012) and the sodium adsorption ratio (SAR) was calculated using the sodium, calcium, and magnesium measurements.

### 2.3. Environmental covariates

The soil properties of salt affected area are closely associated with topographic properties, ground water quality, vegetation distribution, and lithology (Cui et al., 2019; Fathizad et al., 2020; Nabiollahi et al., 2017; Qadir et al., 2008; Taghizadeh-Mehrjardi et al., 2016a), where the

**Table 1**  
Covariate sets used for spatial modeling.

Covariate Data Source	Symbol	Attribute
Digital Elevation Model	AS CA CC CI CNBL CND EL LS factor MCA MrRTF MrVBF PLC PRC RSP SL TWI VD WE	Aspect Catchment area Cross curvature Convergence index Catchment network base level Catchment network distance Elevation Slope-length factor Modified catchment area Multi-resolution ridge top flatness Multi-resolution valley bottom flatness Plan curvature Profile curvature Relative slope position Slope Topographic wetness index Valley depth Wind effect
Landsat 8 (La) & Sentinel 2A (Se)	La.B1(0.440) & Se.B1(0.443) La.B2(0.482) & Se.B2(0.490) La.B3(0.561) & Se.B3(0.560) La.B4(0.654) & Se.B4(0.665) La.B5(0.864) & Se.B5(0.705) La.B6(1.608) & Se.B6(0.740) La.B7(2.200) & Se.B7(0.783) Se.B8(0.842) & Se.B8a(0.865) Se.B9(0.945) & Se.B10(1.375) Se.B11(1.610) & Se.B12(2.190) La.SI & Se.SI La.CAI & Se.CAI La.CI & Se.CI La.BI & Se.BI La.GI & Se.GI La.NDVI & Se.NDVI La.EVI & Se.EVI	Coastal aerosol Blue band Green band Red band Near infrared & Vegetation Red Edge Shortwave IR-1 & Vegetation Red Edge Shortwave IR-2 & Vegetation Red Edge Near infrared & Vegetation Red Edge Water vapour & SWIR-Cirrus Shortwave IR-1 & Shortwave IR-2 Salinity index: (NIR-RED)/(NIR + RED) Carbonate index: (RED/GREEN) Clay index: (SWIR-1/SWIR-2) Brightness index: ((RED)2 + (NIR)2)/0.5 Gypsum index: (SWIR-1-NIR)/(SWIR-1 + 1NIR) Normalized difference vegetation index: (NIR-RED)/(NIR + RED) Enhanced vegetation index: (NIR-RED)/(NIR + C1 × RED-C2 × BLUE + L)
Ground water table map	GWT map	Ground water table depth
Land use map Geology map Physiography map	Landu map Geo map Physi map	Land use unit Geology unit Physiographic unit

literature shows a wide array of environmental covariates have previously been tested for mapping soil salinity (Fathizad et al., 2020; Shababi et al., 2017; Taghizadeh-Mehrjardi et al., 2014, 2016a; Vermeulen and Van Niekerk, 2017; Wang et al., 2019c; Wu et al., 2018). Here, the covariates derived from digital elevation data, remote sensing data, and other geospatial data are described (Table 1). All covariates were resampled to a common 30 m spatial resolution.

### 2.3.1. Digital elevation data

A suite of 18 terrain attributes was calculated from a 10 m spatial resolution digital elevation model (DEM; National Cartographic Center of Iran, 2014) using the SAGA GIS software (System for Automated

Geoscientific Analysis; Olaya, 2004). Terrain attributes included: multi-resolution ridge top flatness (MrRTF), catchment network base level (CNBL), catchment network distance (CND), elevation (EL), catchment area (CA), slope (SL), convergence index (CI), wind effect (WE), aspect (AS), length-slope factor (LS factor), plan curvature (PLC), profile curvature (PRC), cross curvature (CC), valley depth (VD), topographic wetness index (TWI), relative slope position (RSP), modified catchment area (MCA), and the multi-resolution index of valley bottom flatness (MrVBF) (Table 1).

### 2.3.2. Remote sensing data

The median values of spectral bands of Landsat 8 OLI (operational land imager) and Sentinel-2A acquired on September in 2017 were used. These spectral bands are [(La.B1 (0.440 μm), La.B2 (0.482 μm), La.B3 (0.561 μm), La.B4 (0.654 μm), La.B5 (0.864 μm), La.B6 (1.608 μm), La.B7 (2.200 μm), (Se.B1 (0.443 μm), Se.B2 (0.490 μm), Se.B3 (0.560 μm), Se.B4 (0.665 μm), Se.B5 (0.705 μm), Se.B6 (0.740 μm), Se.B7 (0.783 μm), Se.B8 (0.842 μm), Se.B8a (0.865 μm), Se.B9 (0.945 μm), Se.B10 (1.375 μm) Se.B11 (1.610 μm), Se.B12 (2.190 μm)]. Moreover, the following spectral indices were derived from both satellite images: salinity index (SI – Metternicht and Zinck 2003), gypsum index (GI – Boettinger et al., 2008), carbonate index (CAI – Boettinger et al., 2008), brightness index (BI – 32), normalized difference vegetation index (NDVI – Rouse et al., 1973), clay index (CI – Boettinger et al., 2008), and enhanced vegetation index (EVI – Huete et al., 2002) (Table 1).

### 2.3.3. Other geospatial data

Ground water table map and categorical maps of physiography, geology and land use were also used as covariate. The major physiographic units of this area include mountains, hills, fans, piedmonts, plateaus, and lowlands (Fig. 3). Schist, silty clay flat, marl, gabbro, granite, marble and alluvium constitute the major geological units. The main land uses in the study included cropland (wheat and barley), rangeland, and bare land (Fig. 3). These datasets originated in the polygon format and were rasterized to a 30 m spatial resolution.

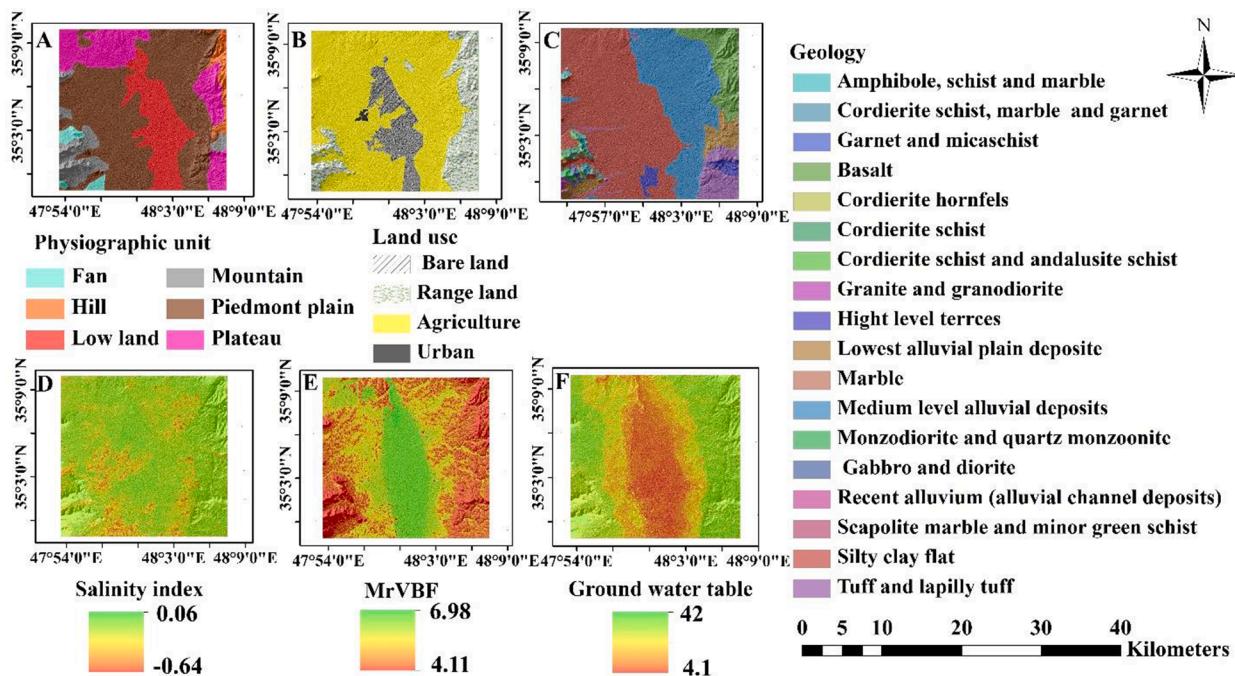
## 2.4. Random forest

Random forest (RF) is a tree-based ensemble models used for both regression and classification tasks (Breiman, 2001). The motivation behind ensemble methods is that by combining multiple weaker learners, a stronger learner is created (Lantz, 2015). A RF is an ensemble of decision trees, where each tree is built and trained using a bootstrap sample from the original dataset. After the forest is generated, each tree in the forest makes a prediction of the target variable. In regression problems, predictions of all trees are averaged to give a single prediction, while in classification tasks, the final prediction is the majority vote of all trees in the ensemble. In RF, averaging the predictions of many uncorrelated trees reduces the sensitivity to noise in training data and thus decrease overall variance without increasing bias when compared to a single decision tree (Lewis, 2015).

### 2.5. Optimization techniques

The genetic algorithm (GA) is a robust, metaheuristic optimization and search technique, has its roots in evolutionary science, and is inspired by the process of genetic and natural selection (Haupt and Haupt, 2004). GAs work based on concepts related to biologic gene operations, which including mutation, crossover, and selection. GA uses specific selection rules to evolve a population of many candidate solutions and, thereafter, the evolved candidate solutions is searched to identify the best solution (Meyer-Bäse and Schmid, 2014). The best solution is identified using an optimizing function, which encodes a potential solution based on chromosome-like data structures.

Metaheuristic nature-based algorithms are becoming popular tools because of their ability to deal with complex optimization problems. The



**Fig. 3.** Examples of covariates used for predicting soil pH, EC, and SAR which included: (A) physiographic map, (B) land use map, (C) geology map, (D) salinity index, (E) Multi-resolution valley bottom flatness index (MrVBF), and (F) ground water table.

idea behind these approaches is inspired from the behavior of biological and/or physical systems in nature. The bat algorithm (BA) is an example of such algorithm—first introduced in Yang (2010) based on the echolocation characteristic of bats. Bats produce very loud sound pulses and use the echoic sound to identify the location and the type of their prey. The BA uses some echolocation behavior of bats such as the variation in frequency, the rate of pulse emission, and loudness efficient strategy during the search process (Yang and He, 2013). Accordingly, several algorithms have formulated the echolocation behavior of microbats and the objective function for optimizing and solving complex problems (Yang, 2013). To develop BA, Yang (2010) idealized and simplified some echolocation characteristics of bats. Accordingly, bats search randomly for prey with velocity  $v_i$  at position  $x_i$  with a fixed frequency  $f_{min}$ , varying wavelength  $\lambda$  and loudness  $A_0$ . Each bat updates  $v_i$ ,  $x_i$  in a  $d$ -dimensional solution space to obtain new solutions for  $v_i$ ,  $x_i$  at iteration  $t$  ( $x_i^t$  and  $v_i^t$ , respectively). Therefore, idealized rules can be formulated as following equations for updating  $x_i^t$  and  $v_i^t$ :

$$f_i = f_{min} + (f_{max} - f_{min})\beta \quad (1)$$

$$v_i^t = v_i^{t-1} + (x_i^t - x^*)\beta \quad (2)$$

$$x_i^t = x_i^{t-1} + v_i^t \quad (3)$$

where  $\beta \in [0, 1]$  is a random vector obtained from a uniform distribution and  $x^*$  is the current best solution among all the bats.

Particle swarm optimization (PSO) a search algorithm developed by Eberhart and Kennedy (1995) based on the swarming behavior of birds and fish. In recent years, the PSO algorithm has received a lot of attention because it provides efficient solutions with minimal implementation effort. It uses specified rules, such as nearest-neighbor velocity matching and acceleration by distance, to reveal swarming behavior in groups of simple agents. The PSO algorithm involves the initialization of a population of random potential solutions (i.e. particles), where each particle inside the swarm is represented by a vector in multidimensional search space of the objective function. This particle space is searched in order to identify optimal solutions by updating generations. Mathematically, suppose  $A \subset \mathbb{R}^n$  is the search space and  $f :$

$A \rightarrow Y \subseteq \mathbb{R}$  is the objective function. The swarm can be defined as a set,  $S = \{x_1, x_2, \dots, x_N\}$  which for  $N$  solution is written as:

$$x_i = (x_{i1}, x_{i2}, \dots, x_{in})^T \in A, \quad i = 1, 2, \dots, N \quad (4)$$

Moving within the search space is adjusted by velocity ( $v_i$ ) which gives the candidate solutions (particles) the ability to shift their position in  $A$ :

$$v_i = (v_{i1}, v_{i2}, \dots, v_{in})^T, \quad i = 1, 2, \dots, N \quad (5)$$

Velocity is modified based on the best location visited by each particle during its search. As these position are stored in a memory, besides  $S$  set, PSO uses a memory set  $P = \{p_1, p_2, \dots, p_N\}$  containing the best positions:

$$p_i = (p_{i1}, p_{i2}, \dots, p_{in})^T \in A, \quad i = 1, 2, \dots, N \quad (6)$$

These position at iteration  $t$  are defines as follow:

$$p_i(t) = \text{argmin}_i(f_i(t)) \quad (7)$$

Considering  $g$  as the best position index with the minimum value for the object function in  $P$  at a given  $t$ , it is possible to describe the best positions as follows:

$$p_g(t) = \text{argmin}(p_i(t)) \quad (8)$$

Finally, PSO is defined as follow:

$$v_{ij}(t+1) = v_{ij}(t) + c_1 R_1 (p_{ij}(t) - x_{ij}(t)) + c_2 R_2 (p_{gi}(t) - x_{ij}(t)) \quad (9)$$

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1) \quad (10)$$

$i = 1, 2, \dots, N$  and  $j = 1, 2, \dots, n$  where  $R_1$  and  $R_2$  are random variables obtained from a uniform distribution within  $[0, 1]$  and  $c_1$  and  $c_2$  are weighting factors.

## 2.6. Validation

Ten-fold cross validation with 100 replications was used to assess the ML algorithms. Root mean square error (RMSE), correlation coefficient, Lin's concordance correlation coefficient (CCC) and mean absolute error

**Table 2**  
Descriptive statistics for soil properties.

Parameters	pH	EC ( $\text{dSm}^{-1}$ )	SAR
Number	293	293	293
Mean	7.97	2.50	8.36
Minimum	6.50	0.23	0.16
Maximum	10.93	22.80	204.00
Standard deviation	0.82	4.09	18.00
Skewness	1.00	3.15	6.51
Kurtosis	1.59	9.70	56.87

(MAE) were used as the validation criteria. Furthermore, we used mean  $\pm 1.65 \times$  standard deviation to prepare the upper and lower limit of the prediction maps.

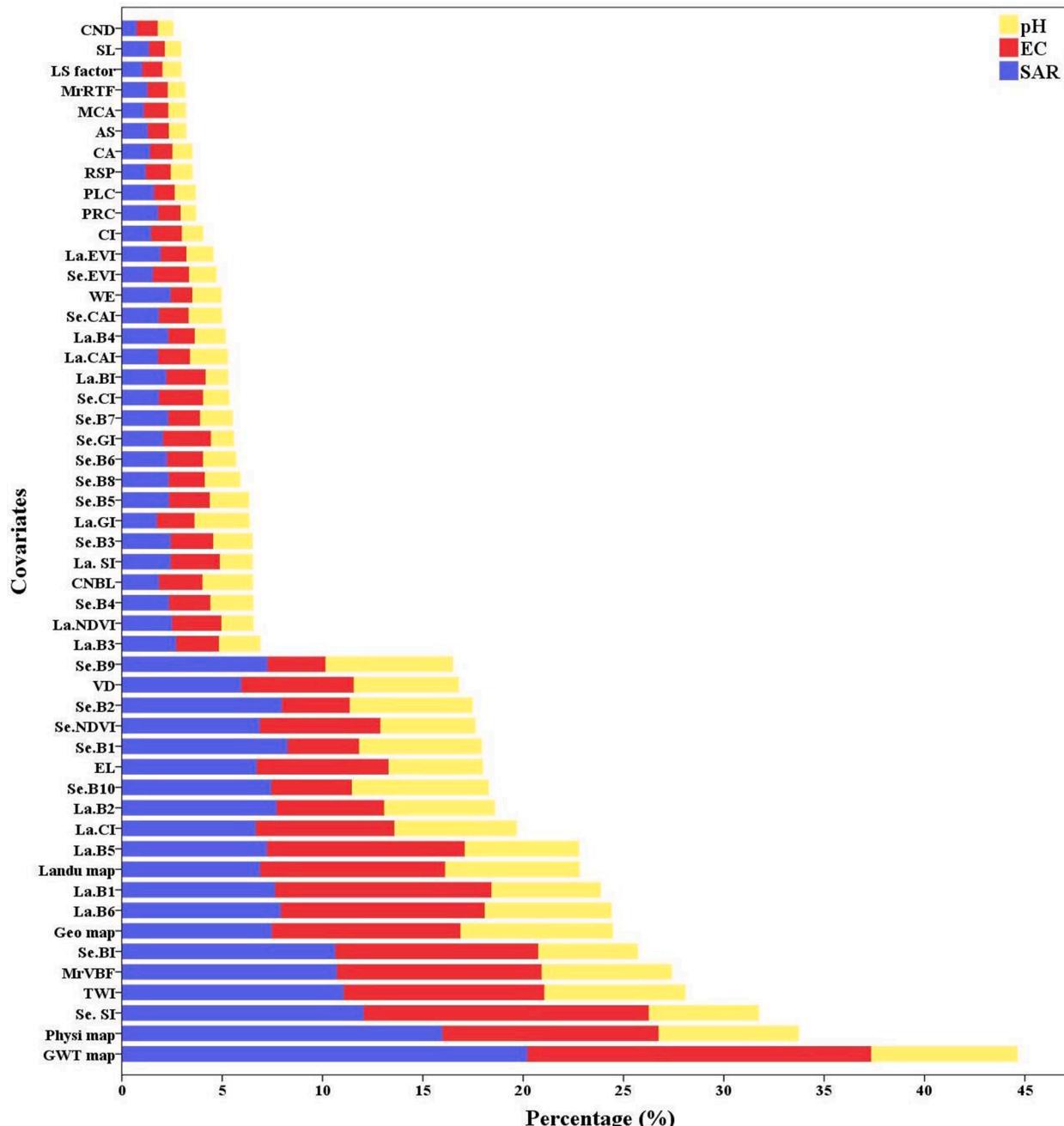
## 2.7. Classified map of soil salinity and sodicity

To highlight soil salinity and sodicity risk across the study area, the resulting maps were classified for soil salinity (based on thresholds EC = 4  $\text{dSm}^{-1}$ ), sodicity (based on a threshold of SAR = 13), and pH (based on a threshold of 8.5; Richards, 1954). Then, classes of salt-affected soils of the study area were prepared based on classified maps of pH, EC, and SAR.

## 3. Results and discussion

### 3.1. Summary statistics of data

**Table 2** shows the descriptive statistics of soil properties in the study area. The average EC was 2.16  $\text{dSm}^{-1}$  and ranged between 0.23 and



**Fig. 4.** Relative influence of covariates on the prediction of soil pH, EC, and SAR.

22.80 dSm<sup>-1</sup>, which indicated that the EC was moderately high. The average pH was 7.36 and ranged between 6.35 and 10.93, which indicated the presence of neutral to alkaline soil. SAR values ranged from 0.16 to 204%, with an average is 7.97%, which indicates SAR levels were low to high. The coefficients of variation for SAR and EC were high (>35%) and showed a high variability across the study area (Wilding, 1985); however, the coefficient of variation for pH was considerably lower (<15%). Approximately 15 to 20% of soil samples had EC, SAR and pH above 4 dSm<sup>-1</sup>, 13, and 8.5, respectively. The samples with the highest salinity, sodicity, and pH level were located in the central region of the study, where the soils are the most severely affected by salt.

### 3.2. Covariates importance

**Fig. 4** shows the significance of each covariate used in the modeling and prediction of salt-affected soils. Here, the relative importance of covariates are reported using the RF + PSO results as it yielded the best model. In summary, for predicting pH, geology map (7.59%), ground water table (7.3%), topographic wetness index (7.03%), physiographic unit (6.98%), land use map (6.7%), and MrVBF (6.49%) were the most important covariates. In terms of predicting EC, ground water table (17.14%), salinity index (14.22%), physiographic unit (10.78%), MrVBF (10.21%), brightness index (10.12%), and topographic wetness index (10.01%) were the most important covariates. For predicting SAR, ground water table (20.18%), physiographic unit (15.95%), salinity index (12.02%), topographic wetness index (12.02%), MrVBF (10.69%), and brightness index (10.61%) were the most important covariates.

The successful use of remote sensing, terrain attribute, groundwater, and physiographic maps as effective covariates in DSM has been emphasized in previous research for Iran (Jafari et al., 2012; Taghizadeh-Mehrjardi et al., 2014; Zeraatpisheh et al., 2018). Elsewhere, Wang et al., (2020a) mapped soil salinity in China using the Cubist model and had also reported that remote sensing indices (e.g., tasseled cap transformation-derived wetness, and salinity index) were the most important covariates. Subsequently, Wang et al., (2020b) mapped soil salinity in China using different ML models and showed similar results where remote sensing data (e.g., canopy response salinity index and NDVI), and terrain attributes (e.g., valley depth, and elevation) were, again, the most important. Taghizadeh-Mehrjardi et al., (2014) predicted the spatial variability of soil salinity in central Iran and showed that the most important covariates included a geomorphological map, and remote sensing data.

Parent material is an important soil-forming factor and plays an important role in both traditional and digital soil mapping (Bockheim et al., 2014) because the chemical, mineralogical, and physical composition of the parent material has a profound effect on soils; however, the importance diminishes with time (Schaetzl and Anderson, 2005). The importance of geology as a control of pH variability in this study area can be attributed to the diverse lithology of the area (**Fig. 4**), where it varied from felsic (e.g. granite and schist) to mafic materials (e.g. basalt, gabbro, and marble) (**Fig. 3**). It was also identified that land use was another controlling factor on pH variability (**Fig. 4**), where the dominant land use type in the study area was agricultural and followed by range land. The relationship between land use and pH was likely attributed to the presence of aboveground and belowground crop residuals as well as agricultural soil management practices such as fertilizer application. For example, oxidation of sulfur and nitrogen from fertilizers and accumulation of organic matter can reduce soil pH (Brady and Weil, 1999). Balstrøm et al. (2013) investigated the spatial distribution of soil pH in Denmark and showed a significant effect of land use on soil pH, especially at the topsoil where they attributed the higher pH of agricultural soils compared to forest due to lime application.

The results indicated that the groundwater table was an important control on pH, EC, and SAR variation in the study area (**Fig. 4**). The groundwater table was lower in area with high EC and SAR, located in center of the study area (**Fig. 3**), compared to the areas with low EC and

SAR. The groundwater, especially in areas with a shallow water table, may affect soil in several ways; for example, the upward movement of water from the groundwater by capillary rise and its evaporation results in the precipitation and accumulation of salts within and on the surface layers. Although rainfall events, which predominantly occur during the winter, facilitate the leaching process, the warm and dry summers associated with the xeric soil moisture regime outweighs the effects of rainfall and therefore favor salt accumulation in the soil. In addition, the long-term use of groundwater as the main source for irrigation and improper irrigation management has resulted in salt accumulation in central part of the area. In line with this study, Fathizad et al. (2020) reported a high contribution of groundwater to soil salinity in central Iran.

Results also indicated importance of RS-based data (e.g. salinity index and brightness index) and terrain attributes (e.g. TWI and MrVBF) in predicting EC and SAR (**Fig. 4**). The high contribution of RS data to the models used to predict EC and SAR can be attributed to the surface features, common in salt-affected soils that affect the reflective behavior of the soil (Farifteh et al., 2006; Zinck and Metternicht, 2009). These surface features include both direct (e.g. surface salts crusts, puffy surface) and indirect (i.e. type and intensity of vegetation cover) indicators of soil salinity (Metternicht & Zinck, 2003; Farifteh et al., 2006). Spectral bands and salinity indices are widely used data to directly detect salinity features, while vegetation indices such as NDVI are often used to reveal the indirect effects of soil salinity on spectral behavior. In salt affected landscapes, the vegetation is influenced by high salt levels, and thereby increasing its visible reflectance and reducing its near-infrared reflectance (Allbed et al., 2014; Bouaziz et al., 2011; Dehni and Louinis, 2012; Sheng et al., 2010). Abdullah et al. (2019), for instance, reported that predictive models that combined both direct and indirect indicators of soil salinity were more effective in soil salinity modeling.

Terrain attributes (e.g. TWI, MrVBF) were also shown to be effective predictors of pH, EC, and SAR across the study area (**Fig. 4**) due to their ability to capture the relief parameters that control several soil-forming process, such as leaching, accumulation, redistribution as well as erosional and depositional processes. For example, MrVBF represents flat valley bottom, which is an indicator of depositional areas (Gallant and Dowling, 2003), where water and salts tend to accumulate and whereas TWI also corresponds with areas of water accumulation (Gallant and Dowling, 2003; Wang and Laffan, 2009) and has also shown to be an important predictor for other soil properties (Abuelgasim and Ammad, 2019; Taghizadeh-Mehrjardi et al., 2014; Wang et al., 2020b; Zare et al., 2019).

Physiographic map unit was also shown as important environmental covariate in predicting pH, EC, and SAR (**Fig. 4**). Widely-used in Iran, physiographic maps are derived from a hierarchical system and are largely based on slope, topography, geology, soil properties (Zeraatpisheh et al., 2020) and other geomorphometric attributes. Therefore, a relationship between physiographic units and soil variability was expected. Specifically, in the central region of study area, there is a lowland physiographic unit with a concave shape, which leads to a shallow groundwater table. As a result, soluble salts, which were washed from the upland areas, accumulated in the soil of this physiography unit and led to increased soil EC. Similar results were reported in the literature and showed that physiographic/landscape parameters and groundwater table are important predictors in arid and semiarid ecosystem (Afify et al., 2010; Fathizad et al., 2020; Jafari et al., 2012; Scull et al., 2005). For instance, Taghizadeh-Mehrjardi et al. (2014) assessed the spatial variability of soil EC in central Iran and indicated that most of the saline soils are located in the playa landform, which also receives more soluble salts from upland areas.

### 3.3. Comparison of different machine learning (ML) algorithms

The ability of four different ML algorithms (RF, RF + BA, RF + GA, and RF + PSO) to predict soil pH, EC, and SAR in Ghorveh region were

**Table 3**

Results of machine learning algorithm for soil pH, EC, and SAR prediction.

		RF	RF + GA	RF + BAT	RF + PSO
pH	RMSE	0.60 ± 0.02	0.52 ± 0.02	0.54 ± 0.03	0.51 ± 0.03
	MAE	0.45 ± 0.03	0.41 ± 0.02	0.41 ± 0.02	0.40 ± 0.02
	R <sup>2</sup>	0.53 ± 0.02	0.62 ± 0.03	0.61 ± 0.05	0.63 ± 0.04
	CCC	0.59 ± 0.02	0.66 ± 0.03	0.66 ± 0.03	0.67 ± 0.03
EC	RMSE	3.21 ± 0.12	2.66 ± 0.07	2.50 ± 0.08	2.32 ± 0.12
	MAE	1.63 ± 0.05	1.57 ± 0.05	1.53 ± 0.04	1.50 ± 0.04
	R <sup>2</sup>	0.50 ± 0.03	0.52 ± 0.04	0.55 ± 0.04	0.57 ± 0.05
	CCC	0.53 ± 0.01	0.57 ± 0.03	0.59 ± 0.03	0.62 ± 0.04
SAR	RMSE	12.22 ± 0.23	9.75 ± 0.16	8.95 ± 0.15	8.90 ± 0.11
	MAE	6.77 ± 0.07	5.86 ± 0.05	5.04 ± 0.06	5.00 ± 0.06
	R <sup>2</sup>	0.45 ± 0.02	0.53 ± 0.06	0.54 ± 0.05	0.55 ± 0.05
	CCC	0.51 ± 0.02	0.55 ± 0.04	0.58 ± 0.03	0.60 ± 0.04

tested using 10-fold cross-validation with 100 replications. Results for root mean square error (RMSE), correlation coefficient ( $R^2$ ), mean absolute error (MAE) and, Lin's concordance correlation coefficient (CCC) for four ML algorithms are summarized in Table 3. Across all accuracy metrics, RF + PSO was the most effective in for mapping all soil attributes comparison to the other three methods. Here, the RMSE and MAE for the soil pH (RMSE = 0.51; MAE = 0.41), EC (RMSE = 2.32; MAE = 1.50), and SAR (RMSE = 8.90; MAE = 5.00) predictions were consistently lower; whereas, the  $R^2$  and CCC for predicting soil pH ( $R^2$  = 0.63; CCC = 0.67), EC ( $R^2$  = 0.57; CCC = 0.62), and SAR ( $R^2$  = 0.55; CCC = 0.60) were higher when using RF + PFO. In particular, the proposed RF + PSO method resulted in improve accuracies (decrease in RMSE) when predicting soil properties in comparison to the other two hybrid models.

RF + PFO had a decrease in RMSE by 2%, 12.8%, and 8.7% for soil pH, EC, and SAR, respectively, in comparison to RF + GA while there was a decrease of 5.5%, 7.2%, and 0.5%, respectively, in comparison to RF + BAT. When compared to the standalone RF model, there was a 15%, 27.7%, and 27.1% decrease for soil pH, EC, and SAR, respectively. The success in using RF + PSO was also reported in studies predicting soil shear strength (Pham et al., 2020) and soil moisture (Yang et al., 2017)).

Although the hybrid methods showed fairly similar values for all accuracy metrics, their performances were significantly different from those obtained by the standalone RF model. For instance, RF + GA, RF + BAT, RF + PSO showed a decrease in RMSE when predicting pH by 13.3%, 10%, and 15%, respectively; predicting EC by 20.6%, 22.1%, and 27.7%, respectively; and predicting SAR by 20.2%, 26.7%, and 27.1%, respectively. The success in using hybrid methods was also reported by Pham et al. (2020), Yang et al. (2017), Wang et al. (2017), and Taghizadeh-Mehrjardi et al. (2015) for predicting other soil properties and therefore, it can be suggested that these hybrid methods could show promise in subsequent DSM studies.

### 3.4. Mapping of soil pH, EC, and SAR

Fig. 5 shows the digital mapping of upper, lower and mean values of soil EC, SAR, and pH produced using the best ML method (RF + PSO). The maps showing the mean predicted EC, SAR and pH from the best ML approach appears in the center of Fig. 5. It is evident from the mean map of soil EC, SAR, and pH that less than one fifth of the area (14%) is characterized by high EC, SAR, and pH. We further classified the three predicted maps (Fig. 6) according to the based on the thresholds of (4

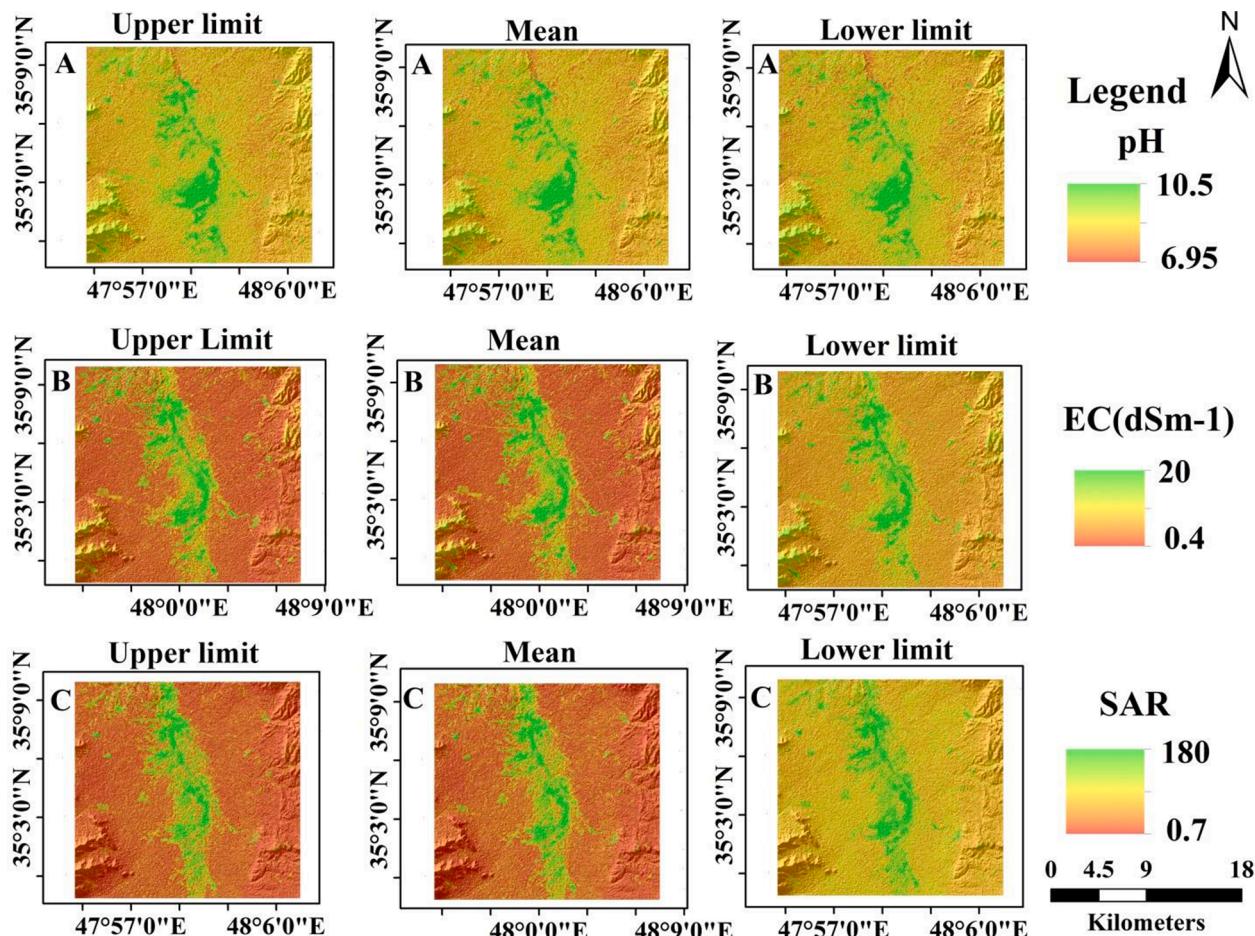


Fig. 5. The spatial distribution of soil (A) pH, (B) EC (dSm<sup>-1</sup>) and (C) SAR predicted (including mean, upper prediction limit, and lower prediction limit) using the RF + PSO model.

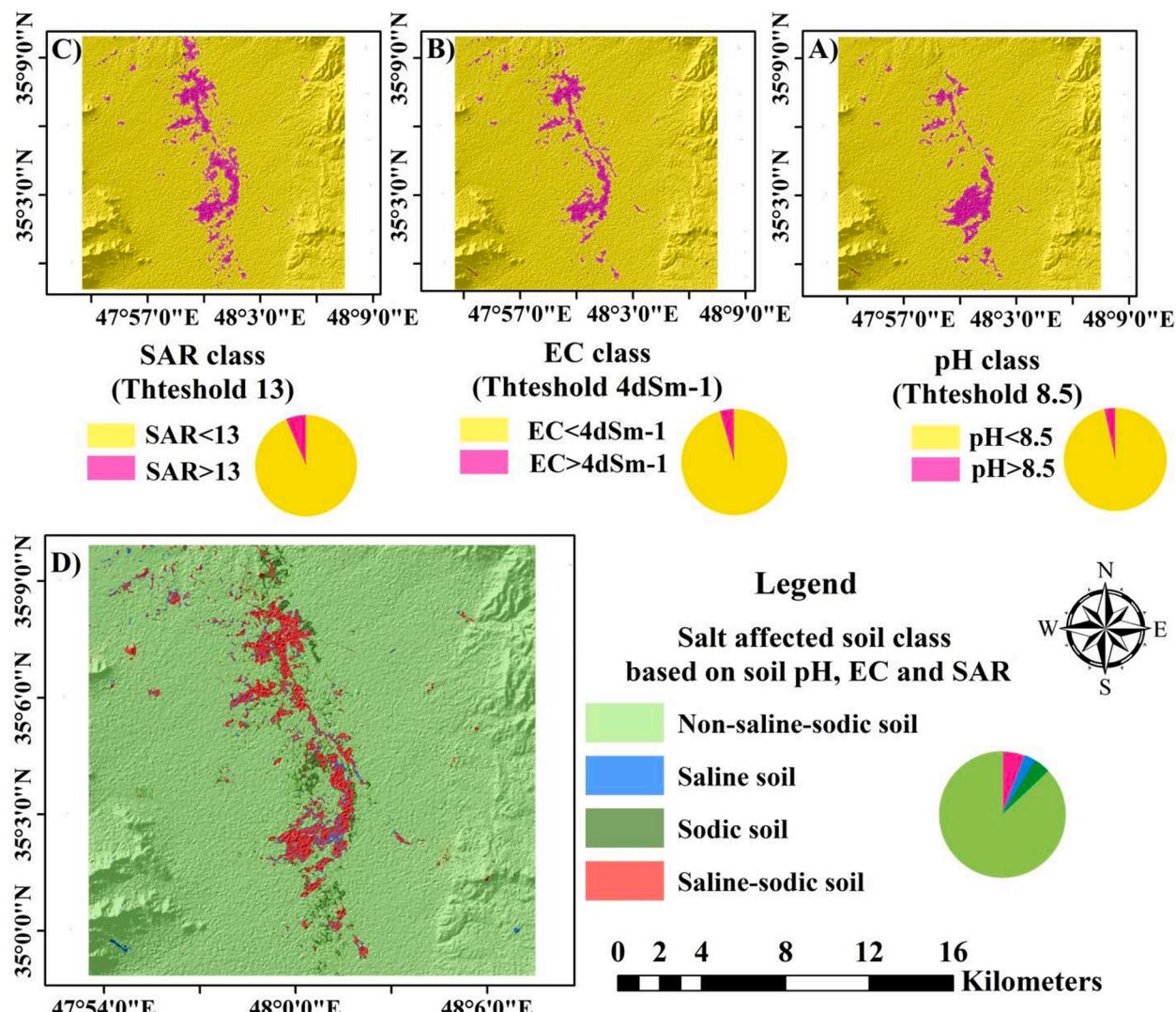


Fig. 6. Classified maps of soil (A) pH, (B) EC ( $\text{dSm}^{-1}$ ), (C) SAR, and (D) salt affected soil class of the study area (D).

$\text{dSm}^{-1}$  for EC; 13 for SAR; and 8.5 for pH) as defined in Richards (1954). These maps show that the central region of the study area has high EC, SAR and pH values. The lower right panel of each map in Fig. 6, indicates the proportion of the study area that are either below or exceed the EC, SAR, and pH thresholds. Based on the EC, SAR, and pH maps, 4.5%, 6.5%, and 3.5% of the study area has EC, SAR and pH values that exceed the thresholds, respectively. Moreover, based on these classified maps of EC, SAR, and pH the salt-affected soils of the study area were further classified as non-saline-sodic ( $\text{pH} < 8.5\%$ ,  $\text{SAR} < 13$ , and  $4 \text{ dSm}^{-1}$ ), saline-sodic soil ( $\text{pH} > 8.5\%$ ,  $\text{SAR} > 13$ , and  $\text{EC} > 4 \text{ dSm}^{-1}$ ), saline soil ( $\text{pH} < 8.5\%$ ,  $\text{SAR} < 13$ , and  $\text{EC} > 4 \text{ dSm}^{-1}$ ), and sodic soil ( $\text{pH} > 8.5\%$ ,  $\text{SAR} > 13$ , and  $\text{EC} < 4 \text{ dSm}^{-1}$ ) (Richards, 1954). Within the study area saline-sodic soil, saline soil, and sodic soil constituted 6.30%, 3.1%, and 4.6% of the study area, respectively; whereas, the remaining 86% of the study area was classified as non-salt-affected (Fig. 6).

### 3.5. pH, EC, and SAR maps according to lithology, physiographic unit and land use

An analysis of variance was used to compare the mean values of EC, SAR, and pH with respect to lithology, physiographic unit, and land use. Here, it was observed that the soil EC ( $4.90 \text{ dSm}^{-1}$ ), SAR (17.12), and pH (8.82) values for the lowland physiographic unit were significantly higher in comparison to other physiographic units (Table 4). Similarly,

the soil EC ( $4.61 \text{ dSm}^{-1}$ ), SAR (15.82), and pH (8.21) were also significantly higher for bare lands in comparison to other land uses (Table 4). Although the soil EC ( $3.10 \text{ dSm}^{-1}$ ), SAR (10.33), and pH (8.82) values were higher for geological units classified as a ‘silty clay flat’, the differences were not significant (Table 4).

In the study area, the ‘lowland’ physiographic unit is located in the central region, which has a low slope, a concave shape, classified as a ‘silty clay flat’ geological unit and bare land. Therefore, more water, along with its soluble salts, are washed out from the upslope areas and accumulated here; hence, leading to a raised ground water (near to the soil surface) in this area and thus increasing the EC (Figs. 4–6). Fathizad et al. (2020) also assessed the spatial and temporal variation of soil salinity using RF in the central desert of Iran and showed that the highest degree of salinity was observed in the lower elevated parts of region.

In the study area the most important control of salinization and alkalization is related to the presence of a groundwater table at the soil surface, which has high salinity and very high sodicity. Here, the salinization and alkalization processes are promoted in areas that are identified as a lowland physiographic unit and in regions with low permeability (i.e. silty clay clay geology unit). Kilic and Kilic (2007) investigated the spatial variability of salinization and alkalization in Turkey and showed that a low quality groundwater table near soil surface was also the most important cause of salinization and alkalization; hence, the high salt contents render these landscapes unsuitable for

**Table 4**

Mean value of soil salinity, sodicity and pH in different land uses, lithology units and physiography units. Means that do not share a letter are significantly different at the 0.01 level ( $p < 0.05$ ) according to Tukey's test.

		pH Mean	EC	SAR
Land use	Poor range land	7.62 <sup>b</sup>	0.74 <sup>b</sup>	1.19 <sup>b</sup>
	Agriculture	7.66 <sup>b</sup>	1.05 <sup>b</sup>	2.50 <sup>b</sup>
	Moderate range land	8.58 <sup>a</sup>	4.61 <sup>a</sup>	15.82 <sup>a</sup>
	P = (Tukey's test)	<0.05**	<0.05**	<0.05**
Physiography unit	Mountain	7.77 <sup>ab</sup>	0.64 <sup>b</sup>	1.60 <sup>b</sup>
	Hill	7.55 <sup>b</sup>	0.73 <sup>b</sup>	3.03 <sup>b</sup>
	Plateau	7.55 <sup>b</sup>	1.20 <sup>b</sup>	1.62 <sup>b</sup>
	Piedmont plain	7.80 <sup>b</sup>	1.12 <sup>b</sup>	3.78 <sup>b</sup>
	Fan	7.83 <sup>b</sup>	0.61 <sup>b</sup>	0.18 <sup>cb</sup>
	Lowland	8.82 <sup>a</sup>	4.90 <sup>a</sup>	17.12 <sup>a</sup>
	p = (Tukey's test)	<0.05**	<0.05**	<0.05**
Lithology unit	Amphibole, schist and marble	7.25 <sup>a</sup>	0.82 <sup>a</sup>	1.85 <sup>a</sup>
	Cordierite hornfels	7.54 <sup>a</sup>	0.32 <sup>a</sup>	0.18 <sup>a</sup>
	Garnet and micaschist,	7.58 <sup>a</sup>	0.48 <sup>a</sup>	0.47 <sup>a</sup>
	Cordierite schist	7.64 <sup>a</sup>	1.11 <sup>a</sup>	1.58 <sup>a</sup>
	Monzodiorite and quartz monzonite	7.64 <sup>a</sup>	0.67 <sup>a</sup>	0.60 <sup>a</sup>
	Basalt,	7.70 <sup>a</sup>	0.61 <sup>a</sup>	
	Silty clay flat	8.21 <sup>a</sup>	3.10 <sup>a</sup>	10.33 <sup>a</sup>
	Gabro and diorite	8.30 <sup>a</sup>	0.59 <sup>a</sup>	0.91 <sup>a</sup>
	Lowest alluvial plain deposit	7.63 <sup>a</sup>	0.86 <sup>a</sup>	0.94 <sup>a</sup>
	Tuff and lapilli tuff	7.71 <sup>a</sup>	0.55 <sup>a</sup>	0.83 <sup>a</sup>
	High level terraces	7.80 <sup>a</sup>	0.53 <sup>a</sup>	1.16 <sup>a</sup>
	Granite and granodiorite,	7.69	0.50 <sup>a</sup>	1.19 <sup>a</sup>
	Marble and phlogopite	7.71 <sup>a</sup>	0.50 <sup>a</sup>	1.22 <sup>a</sup>
	Medium level alluvial deposits	7.83 <sup>a</sup>	0.54 <sup>a</sup>	3.03 <sup>a</sup>
	Recent alluvium	8.24 <sup>a</sup>	1.96 <sup>a</sup>	4.4 <sup>a</sup>
	Cordierite schist, marble and garnet	7.87 <sup>a</sup>	1.12 <sup>a</sup>	1.54 <sup>a</sup>
	Cordierite schist and andalusite schist	7.91 <sup>a</sup>	1.10 <sup>a</sup>	1.85 <sup>a</sup>
	Scapolite marble and minor green schist	7.81 <sup>a</sup>	1.00 <sup>a</sup>	1.75 <sup>a</sup>
	P = (Tukey's test)	ns	ns	ns

\*\*and ns, are significant and non-significant at the 0.01 level, respectively.

many plant communities and certain ecosystem services. Fathizad et al., (2020) showed that the main causes of salinization in the central desert of Iran were related to limited precipitation, high evaporation from the soil, and geological unit; furthermore, anthropogenic factors related to poor drainage of the lands, increased number of wells, and expansion of agricultural lands into areas with poor groundwater quality have also led to increased salinity.

#### 4. Conclusion

Digital soil maps that identify areas of high risk to the salinization and alkalization processes can be used to facilitate better land management and soil management practices. This study assessed spatial variability of soil salinity and sodicity on agriculturally intensive regions of the Kurdistan province, Iran. Specifically, four various machine learning models (i.e., RF, RF + BAT, RF + GA, and RF + PSO) and a set of covariate data were used to digitally map soil pH, EC, and SAR. The results indicated that data related to the groundwater table, physiography, geology, land use, salinity index, multi-resolution ridge top flatness and topographic wetness index were the most important covariates for predicting EC, SAR, and pH. Based on RMSE, MAE, R<sup>2</sup>, and CCC results, the hybridized ML model (i.e., RF + PSO) was the best ML model for predicting soil pH, EC, and SAR compared to the other ML models. Furthermore, the digital soil maps show that the central region of the study area, which covers 14% of the area, has the highest pH, EC, and SAR values. Here, the most predominant controls of salinization and

alkalization were related to the presence of a poor quality groundwater table near the surface of the soil on low slope and concaved topography. Future monitoring of salinity and sodicity in this and other areas of Iran, where salinization and alkalization are a major environmental challenge, we suggest the use of inexpensive and readily available covariate data and hybridized ML models to help identify the areas that are likely to be affected by salts.

#### Declaration of Competing Interest

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

#### Acknowledgements

Ruhollah Taghizadeh-Mehrjardi has been supported by the Alexander von Humboldt Foundation, Germany under the grant number: Ref 3.4-1164573-IRN-GFHERMES-P. Thomas Scholten thanks the German Research Foundation (DFG) for supporting this research through the Collaborative Research Center (SFB 1070) 'Resource Cultures' (sub-projects Z, S and B02) and the DFG Cluster of Excellence 'Machine Learning - New Perspectives for Science', EXC 2064/1, project number 390727645.

#### References

- Abdullah, A.Y.M., Biswas, R.K., Chowdhury, A.I., Billah, S.M., 2019. Modeling soil salinity using direct and indirect measurement techniques: a comparative analysis. Environ. Dev. 29, 67–80.
- Abuelgasim, A., Ammad, R., 2019. Mapping soil salinity in arid and semi-arid regions using Landsat 8 OLI satellite data. Remote Sens. Appl.: Soc. Environ. 13, 415–425.
- Afify, A.A., Arafat, S.S., Ghar, M.A., Khader, M.H., 2010. Physiographic soil map delineation for the Nile alluvium and desert outskirts in middle Egypt using remote sensing data of EgyptSat-1. Egypt. J. Remote Sens. Space Sci. 13 (2), 129–135.
- Allbed, A., Kumar, L., Aldakheel, Y.Y., 2014. Assessing soil salinity using soil salinity and vegetation indices derived from IKONOS high-spatial resolution imagaries: applications in a date palm dominated region. Geoderma 230, 1–8.
- Balström, T., Breuning-Madsen, H., Krüger, J., Jensen, N.H., Greve, M.H., 2013. A statistically based mapping of the influence of geology and land use on soil pH: a case study from Denmark. Geoderma 192, 453–462.
- Bockheim, J.G., Gennadiyev, A.N., Hartemink, A.E., Brevik, E.C., 2014. Soil-forming factors and Soil Taxonomy. Geoderma 226, 231–237.
- Boettinger, J.L., Ramsey, R.D., Bodily, J.M., Cole, N.J., Kienast-Brown, S., Nield, S.J., Saunders, A.M., Stum, A.K., 2008. Landsat spectral data for digital soil mapping. In: Hartemink, A.E., McBratney, A.B., Mendonca-Santos, M.L. (Eds.), Digital Soil Mapping with Limited Data. Springer Science, Australia, pp. 193–203.
- Bogunovic, I., Trevisani, S., Pereira, P., Vukadinovic, V., 2018. Mapping soil organic matter in the Baranja region (Croatia): geological and anthropic forcing parameters. Sci. Total Environ. 643, 335–345.
- Bouaziz, M., Matschullat, J., Gloaguen, R., 2011. Improved remote sensing detection of soil salinity from a semi-arid climate in Northeast Brazil. C.R. Geosci. 343, 795–803.
- Brady, N.C., Weil, R.R., 1999. In: Soil organic matter. The nature and properties of soils. Prentice Hall, Upper Saddle River, New Jersey, pp. 446–490.
- Breiman, L., 2001. Random Forests. Mach. Learn. 45, 12–20.
- Cui, G., Lu, Y., Zheng, C., Liu, Z., Sai, J., 2019. Relationship between soil salinization and groundwater hydration in Yaoba Oasis, Northwest China. Water 11, 175.
- Dehni, A., Lounis, M., 2012. Remote sensing techniques for salt affected soil mapping: application to the Oran region of Algeria. Procedia Eng. 33, 188–198.
- Eberhart, R., Kennedy, J., 1995. A new optimizer using particles swarm theory. In: Proc. Sixth International Symposium on Micro Machine and Human Science (Nagoya, Japan). IEEE Service Center, Piscataway, NJ, pp. 39–43.
- Farifteh, J., Farshad, A., George, R.J., 2006. Assessing salt – affected soils using remote sensing, solute modeling, and geophysics. Geoderma 130, 191–206.
- Fathizad, H., Hakimzadeh Ardakanl, M.A., Sodaiezadeh, H., Kerry, R., Taghizadeh-Mehrjardi, R., 2020. Investigation of the spatial and temporal variation of soil salinity using random forests in the central desert of Iran. Geoderma 365.
- Gallant, J.C., Dowling, T.I., 2003. A multiresolution index of valley bottom flatness for mapping depositional areas. Water Resour. Res. 39 (12).
- Haupt, R.L., Haput, S.E., 2004. Practical Genetic Algorithms, second ed. John Wiley & Sons Inc.
- Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ. 83, 195–213.
- IUSS Working Group WrB., 2015. World reference base for soil resources 2014, update 2015: International soil classification system for naming soils and creating legends for soil maps. World Soil Resources Reports No. 106, p.192.

- Jafari, A., Khademi, H., Finke, P., Wauw, J.V.D., Ayoubi, S., 2014. Spatial prediction of soil great groups by boosted regression trees using a limited point dataset in an arid region, southeastern Iran. *Geoderma* 232–234, 148–163.
- Jafari, A., Finke, P.A., de Wauw, J.V., 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. *Eur. J. Soil Sci.* 63, 284–298.
- Jayachandran, K., Gamare, J.S., Nair, P.R., Xavier, M., Aggarwal, S.K., 2012. A novel biamperometric methodology for thorium determination by EDTA complexometric titration. *Radiochim. Acta* 100, 311–314.
- Kilic, K., Kilic, S., 2007. Spatial variability of salinity and alkalinity of a field having salinization risk in semi-arid climate in northern Turkey. *Environ. Monit. Assess.* 127, 55–65.
- Kumar, N., Singh, S.K., Pandey, H.K., 2018. *Appl. Geomath.* <https://doi.org/10.1007/s12518-018-0218-2>.
- Lantz, B., 2015. *Machine Learning with R*, second ed. Packt Publishing Ltd., Birmingham.
- Lewis, N.D., 2015. 92 applied predictive modeling techniques in R. CreateSpace Independent Publishing Platform.
- Malone, B.P., Jha, S.K., Minasny, B., McBratney, A.B., 2016. Comparing regression-based digital soil mapping and multiple-point geostatistics for the spatial extrapolation of soil data. *Geoderma* 262, 243–253.
- McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. *Geoderma* 117 (1–2), 3–52.
- McLean, E.O., 1982. Soil pH and lime requirement. In: Page, A.L., Miller, R.H., Keeney, D.R. (Eds.), *Methods of Soil Analysis, Part 2. Chemical and Microbiological Properties*, second ed. 9. ASA-SSSA, Madison, WI, pp. 199–224.
- Metternicht, G., Zinck, J.A., 2003. Remote sensing of soil salinity: potentials and constraints. *Remote Sens. Environ.* 85, 1–20.
- Meyer-Bäse, A., Schmid, V., 2014. *Pattern Recognition and Signal Analysis in Medical Imaging*, second ed. Elsevier Press.
- Minasny, B., McBratney, A.B., 2016. Digital soil mapping: a brief history and some lessons. *Geoderma* 264, 301–311.
- Minasny, B., McBratney, A.B., Mendonça-Santos, M.L., Odeh, I.O.A., Guyon, B., 2006. Prediction and digital mapping of soil carbon storage in the lower Namoi Valley. *Aust. J. Soil Res.* 44 (3), 233–244.
- Nabiollahi, K., Eskandari, Sh., Taghizadeh-Mehrjardi, R., Kerry, R., Triantafilis, J., 2019. Assessing soil organic carbon stocks under land use change scenarios using random forest models. *Carbon Manag.* 10 (1), 63–77.
- Nabiollahi, K., Golmohammadi, F., Taghizadeh-Mehrjardi, M., Kerry, R., 2018a. Assessing the effects of slope gradient and land use change on soil quality degradation through digital mapping of soil quality indices and soil loss rate. *Geoderma* 318, 482–494.
- Nabiollahi, K., Taghizadeh-Mehrjardi, M., Eskandari, Sh., 2018b. Assessing and monitoring the soil quality of forested and agricultural areas using soil-quality indices and digital soil-mapping in a semi-arid environment. *Arch. Agron. Soil Sci.* 64 (5), 482–494.
- Nabiollahi, K., Taghizadeh-Mehrjardi, M., Kerry, R., Moradian, Sh., 2017. Assessment of soil quality indices for salt-affected agricultural land in Kurdistan Province, Iran. *Ecol. Ind.* 83, 482–494.
- National Cartographic Center of Iran, 2014. Research Institute of National Cartographic Center, Tehran, Iran. <<http://www.ncc.org.ir/>> (as of 14 February 2014).
- Olaya, V., 2004. A Gentle Introduction to SAGA GIS. p. 216.
- Page, A.L., 1992. In: *Methods of Soil Analysis. ASA and SSSA publishers*, Madison, Wisconsin, USA, p. 321.
- Pahlavani-Rad, M.R., Akbarimoghaddam, A., 2018. Spatial variability of soil texture fractions and pH in a flood plain (case study from eastern Iran). *Catena* 160, 275–281.
- Peng, J., Biswas, A., Jiang, Q., Zhao, R., Hu, J., Hu, B., Shi, Z., 2018. Estimating soil salinity from remote sensing and terrain data in southern Xinjiang Province, China. *Geoderma* 337, 1309–1319.
- Pereira, P., Brevik, E., Munoz-Rojas, M., Miller, B., Smetanova, A., Depellegrin, D., Misiti, I., Novara, A., Cerdá, A., 2017. Soilmapping and processmodelling for sustainable land management. In: Pereira, P., Brevik, E., Munoz-Rojas, M., Miller, B. (Eds.), *Soil Mapping and Process Modelling for Sustainable Land Use Management*. Elsevier, Amsterdam, Netherlands, pp. 29–60.
- Pham, B.T., Qi, C., Ho, L.S., Nguyen-Thoi, T., Al-Ansari, N., Nguyen, M.D., Nguyen, H.D., Ly, H.B., Le, H.V., Prakash, J., 2020. A novel hybrid soft computing model using random forest and particle swarm optimization for estimation of undrained shear strength of soil. *Sustainability* 12 (6), 2218.
- Qadir, M., Qureshi, A.S., Cheraghi, A.M., 2008. Extent and characterization of salt-affected soils in Iran and strategies for their amelioration and management. *Land Degrad. Dev.* 19, 214–224.
- Rentschler, T., Gries, P., Behrens, T., Bruehlheide, H., Kühn, P., Seitz, S., Shi, X., Trogisch, S., Scholten, T., Schmidt, K., 2019. Comparison of catchment scale 3D and 2.5 D modelling of soil organic carbon stocks in Jiangxi Province, PR China. *PLoS ONE* 14.
- Rhoades, J.D., 1982. Soluble salts. In: Page, A.L. (Ed.), *Methods of Soil Analysis, Part II, 2nd ed.*, ASA, Monograph No. 9, Madison, WI, pp. 167–179.
- Richards, L.A., 1954. Diagnosis and improvement of saline and alkali soils. *Agricultural Handbook No. 60. U.S. Salinity Laboratory Riverside, California*.
- Rouse, J.W., Hass, R.H., Schell, J.A., Deering, D.W., 1973. Monitoring vegetation systems in the Great Plains with ERTS. In: Freden, S.C., Mercanti, E.P., Becker, M.A. (Eds.), *Technical Presentations Section A Proceedings of the NASA SP-351: Third Earth Resources Technology Satellite-Symposium*, Washington, DC, USA, 10–14 December 1973, vol. 1. NASA Science and Technology Information Office, Washington, DC, USA, pp. 309–317.
- Schaetzl, R.J., Anderson, S., 2005. Soils: genesis and geomorphology. In: Scull, P., Franklin, J., Chadwick, O.A. (Eds.), *The application of classification of tree analysis to soil type prediction in a desert landscape*. Ecol Model. Cambridge Univ. Press, Cambridge, pp. 1–15.
- Scull, P., Franklin, J., Chadwick, O.A., 2005. The application of classification of tree analysis to soil type prediction in a desert landscape. *Ecol. Model.* 181, 1–15.
- Shahabi, M., Jafarzadeh, A.A., Neyshabouri, M.R., Ghorbani, M.A., Valizadeh Kamran, K., 2017. Spatial modeling of soil salinity using multiple linear regression, ordinary kriging and artificial neural network methods. *Arch. Agron. Soil Sci.* 63 (2), 151–160.
- Sheng, J., Ma, L., Jiang, P., Li, B., Huang, F., Wu, H., 2010. Digital soil mapping to enable classification of the salt-affected soils in desert agro-ecological zones. *Agric. Water Manag.* 97, 1944–1951.
- Soil Survey Staff, 2014. *Keys to Soil Taxonomy*, 12th ed. United States Department of Agriculture, Washington.
- Taghizadeh-Mehrjardi, R., Nabiollahi, K., Rasoli, L., Kerry, R., Scholten, T., 2020. Land suitability assessment and agricultural production sustainability using machine learning models. *Agronomy* 10 (573), 1–20.
- Taghizadeh-Mehrjardi, R., Ayoubi, S., Namazi, Z., Malone, B., Zolfaghari, A., Sadrabadi-Roustaie, F., 2016a. Prediction of soil surface salinity in arid region of central Iran using auxiliary variables and genetic programming. *Arid Land Res. Manag.* 30 (1), 49–64.
- Taghizadeh-Mehrjardi, M., Nabiollahi, K., Kerry, R., 2016b. Digital mapping of soil organic carbon at multiple depths using different data mining techniques in Baneh region, Iran. *Geoderma* 253–254, 67–77.
- Taghizadeh-Mehrjardi, M., Nabiollahi, K., Minasny, B., Triantafilis, J., 2015. Comparing data mining classifiers to predict spatial distribution of USDA-family soil groups in Baneh region, Iran. *Geoderma* 253–254, 67–77.
- Taghizadeh-Mehrjardi, R., Minasny, B., Sarmadian, F., Malone, B.P., 2014. Digital mapping of soil salinity in Ardakan region, central Iran. *Geoderma* 213, 15–28.
- Teng, T., Viscarra Rossel, R.A., Shi, Z., Behrens, T., 2018. Updating a national soil classification with spectroscopic predictions and digital soil mapping. *Catena* 164, 125–134.
- Vermeulen, D., Van Niekerk, A., 2017. Machine learning performance for predicting soil salinity using different combinations of geomorphometric covariates. *Geoderma* 299, 1–12.
- Villarino, S.H., Studdert, G.A., Baldassini, P., Cendoya, M.G., Ciuffoli, L., Mastrandolo, M., Pineiro, G., 2017. Deforestation impacts on soil organic carbon stocks in the Semiariid Chaco Region, Argentina. *Sci Total Environ.* 575, 1056–1065.
- Wang, D., Laffan, S.W., 2009. Characterisation of valleys from DEMs. In: *Proceedings of 18th World IMACS/MODSIM Congress*. IMACS, MSSANZ, Cairns, pp. 2014–2020.
- Wang, B., Waters, C., Orgill, S., Clark, A., Liu, D.L., Simpson, M., Cowie, A., McGowen, I., Sides, T., 2017. Estimating soil organic carbon stocks using machine learning methods in the semi-arid rangelands of New South Wales. *22nd International Congress on Modelling and Simulation*: Hobart, Tasmania, Australia.
- Wang, F., Shi, Z., Biswas, A., Yang, S., Ding, J., 2020a. Multi-algorithm comparison for predicting soil salinity. *Geoderma* 365.
- Wang, F., Yang, S., Yang, W., Yang, X., Jianli, D., 2019c. Comparison of machine learning algorithms for soil salinity predictions in three dryland oases located in Xinjiang Uyghur Autonomous Region (XJUAR) of China. *Eur. J. Remote. Sens.* 52 (1), 256–276.
- Wang, J., Ding, D., Yu, D., Teng, D., He, B., Chen, X., Ge, X., Zhang, Z., Wang, Y., Yang, X., Shi, T., Su, F., 2020b. Machine learning-based detection of soil salinity in an arid desert region, Northwest China: a comparison between Landsat-8 OLI and Sentinel-2 MSI. *Sci. Total Environ.* 707.
- Wang, J., Ding, J., Li, G., Liang, J., Yu, D., Aishan, T., 2019a. Dynamic detection of water surface area of Ebinur Lake using multi-source satellite data (Landsat and Sentinel-1A) and its responses to changing environment. *Catena* 177, 189–201.
- Wang, J., Ding, J., Yu, D., Ma, X., Zhang, Z., Ge, X., 2019b. Capability of Sentinel-2MSI data for monitoring and mapping of soil salinity in dry and wet seasons in the Ebinur Lake region, Xinjiang, China. *Geoderma* 353, 172–187.
- Wilding, L.P., 1985. Spatial variability: it's documentation, accommodation and implication to soil surveys. In: Nielsen, D.R., Bouma, J. (Eds.), *Soil Spatial Variability*. Pudoc, Wageningen, The Netherlands, pp. 166–194.
- Wu, W., Zucca, C., Muhammed, A.S., Al-Shafie, W.M., Fadhil Al-Quraishi, A.M., Nangia, V., Zhu, M., Liu, G., 2018. Soil salinity prediction and mapping by machine learning regression in Central Mesopotamia, Iraq. *Land Degrad. Dev.* 29 (11), 4005–4014.
- Yang, X., Liu, Q., Luo, X., Zheng, Z., 2017. Spatial regression and prediction of water quality in a watershed with complex pollution sources. *Sci. Rep.* 7 (1), 1–11.
- Yang, S.X., 2013. Bat algorithm and cuckoo search: a tutorial. In: Yang, S.X. (Ed.), *Artificial Intelligence, Evolutionary Computing and Metaheuristics*. Springer-Verlag, Berlin Heidelberg.
- Yang, S.X., He, X., 2013. Bat algorithm: literature review and applications. *Int. J. Bio-Inspir. Comput.* 5, 141–149.
- Yang, X.S., 2010. A new meta-heuristic bat-inspired algorithm. In: Gonzalez, J.R. (Ed.), *Nature Inspired Cooperative Strategies for Optimization. Studies in Computational Intelligence*. Springer Berlin, pp. 65–74.
- Zare, S., Fallah Shamsi, S.R., Abtahi, S.A., 2019. Weakly-coupled geo-statistical mapping of soil salinity to Stepwise Multiple Linear Regression of MODIS spectral image products. *J. Afr. Earth Sci.* 152, 101–114.

- Zeraatpisheh, M., Bakhshandeh, E., Hosseini, Alavi, S.M., 2020. Assessing the effects of deforestation and intensive agriculture on the soil quality through digital soil mapping. *Geoderma* 363.
- Zeraatpisheh, M., Ayoubi, Sh., Jafari, A., Tajik, S., Finke, P., 2018. Digital mapping of soil properties using multiple machine learning in a semiarid region, central Iran. *Geoderma* 338, 445–452.
- Zinck, J.A., Metternicht, G., 2009. Soil salinity and salinization hazard. In: *Remote Sensing of Soil Salinization: Impact on Land Management*. CRS Press, Boca Raton, pp. 3–18.