Environmental Science and Pollution Research (2021) 28:10804–10817
https://doi.org/10.1007/s11356-020-11319-5

RESEARCH ARTICLE

Susceptibility mapping of groundwater salinity using machine learning models

Amirhosein Mosavi[1,2] · Farzaneh Sajedi Hosseini[3] · Bahram Choubin[4] · Fereshteh Taromideh[5] · Marzieh Ghodsi[6] · Bijan Nazari[7] · Adrienn A. Dineva[8]

Received: 10 July 2020 / Accepted: 18 October 2020 / Published online: 25 October 2020

Abstract

Increasing groundwater salinity has recently raised severe environmental and health concerns around the world. Advancement of the novel methods for spatial salinity modeling and prediction would be essential for effective management of the resources and planning mitigation policies. The current research presents the application of machine learning (ML) models in groundwater salinity mapping based on the dichotomous predictions. The groundwater salinity is predicted using the essential factors (i.e., identified by the simulated annealing feature selection methodology) through k-fold cross-validation methodology. Six ML models, namely, flexible discriminant analysis (FDA), mixture discriminant analysis (MAD), boosted regression tree (BRT), multivariate adaptive regression spline (MARS), random forest (RF), support vector machine (SVM), were employed to groundwater salinity mapping. The results of the modeling indicated that the SVM model had superior performance than other models. Variables of soil order, groundwater withdrawal, precipitation, land use, and elevation had the most contribute to groundwater salinity mapping. Results highlighted that the southern parts of the region and some parts in the north, northeast, and west have a high groundwater salinity, in which these areas are mostly matched with soil order of Entisols, bareland areas, and low elevations.

Keywords Salinity mapping · Machine learning · Feature selection · Simulated annealing · Dichotomous prediction

Responsible Editor: Marcus Schulz

✉ Bahram Choubin
b.choubin@areeo.ac.ir

✉ Adrienn A. Dineva
adrienndineva@duytan.edu.vn

Amirhosein Mosavi
amirhosein.mosavi@tdtu.edu.vn

Farzaneh Sajedi Hosseini
farzsajedi@ut.ac.ir

Fereshteh Taromideh
f.taromide@stu.Sanru.ac.ir

Marzieh Ghodsi
mghodsi@alumni.ut.ac.ir

Bijan Nazari
b.nazari@eng.ikiu.ac.ir

1 Environmental Quality, Atmospheric Science and Climate Change Research Group, Ton Duc Thang University, Ho Chi Minh City, Vietnam

2 Faculty of Environment and Labour Safety, Ton Duc Thang University, Ho Chi Minh City, Vietnam

3 Reclamation of Arid and Mountainous Regions Department, Faculty of Natural Resources, University of Tehran, Karaj, Iran

4 Soil Conservation and Watershed Management Research Department, West Azarbaijan Agricultural and Natural Resources Research and Education Center, AREEO, Urmia, Iran

5 Department of Irrigation, Sari Agricultural Sciences and Natural Resources University, Sari, Iran

6 Faculty of Geography, University of Tehran, Tehran, Iran

7 Department of Water Sciences and Engineering, Imam Khomeini International University, Qazvin, Iran

8 Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam

Introduction

Importance and sole dependency on the groundwater in an arid oasis of the globe is magnified due to the scarcity of the other freshwater resources. Furthermore, due to the recent severe effects of climate change on the expansion and devastation of the long-term droughts, and the ever-increasing surface water pollutions, this dependency has become even more visible and the need for the groundwater is continuously growing worldwide (Kundzewicz and Doell 2009; Stoll et al. 2011; Stuart et al. 2011; Gallardo 2013; Alberti et al. 2016). Thus, groundwater plays an essential role in reducing vulnerability and enhancing resilience to climate change (Lipczynska-Kochany 2018; Mas-Pla and Menció 2019; Yihdego et al. 2017; Khan et al. 2008).

Groundwater overdraft, salinity, and industrial and agricultural pollutions are among the significant challenges for maintaining groundwater quality (Davoodi et al. 2019; Tavakoli-Kivi et al. 2019; Odeh et al. 2019). The formation and travel time of groundwater in a natural water cycle is very long. This cycle, depending on the depth and characteristics of aquifers, may vary from centuries to millennia. Thus, survival and sustainable development in arid areas directly depend on effective management for maintaining a healthy quality and acceptable level of groundwater (Hosseini et al. 2019). Increasing groundwater salinity has recently raised severe environmental and health concerns around the world such as destroying the soil structure, reducing the biodiversity and agricultural production, water pollution, and problems for human health (Naser et al. 2020; Li et al. 2019; Xiao et al. 2019; Banda et al. 2019; Sang et al. 2018; Sofiyan Abuelaish and Camacho Olmedo 2018). Thus, the development of the novel approaches for salinity modeling and prediction has been seen as an essential approach for effective management of the resources and mitigation policies (Chien and Lautz 2018; Chowdhury et al. 2018; Aydin et al. 2017; Suzuki et al. 2017; Giannoccaro et al. 2017).

The literature includes a different range of statistical, numerical, hydrological, and physically based models for groundwater salinity prediction (Masciopinto et al. 2017; Bourke et al. 2017; Pauw et al. 2017; Levanon et al. 2017; Delsman et al. 2017; Gil-Márquez et al. 2017). However, recently, the availability and the abundance of hydrochemical, hydrogeochemical, and electrochemical datasets have motivated using advanced data-driven methods (Haselbeck et al. 2019; Duque et al. 2019; M'nassri et al. 2019; Delsman et al. 2018). The data-driven methods are based on a model of brain cell interaction which is introduced in 1949 by Donald Hebb (Hebb 1949) and developed by Samuel (1959). The main advantages of these models can be summarized as (i) easily identifies the trends and patterns of variables; (ii) can identify implicit relationships in data-sets; (iii) automatically adaptation with the dynamic system and environment without human intervention; (iv) handling multi-dimensional data; (v) wide applications; and (vi) saving both time and money (Lu 1990; Yu and Liu 2003; Wuest et al. 2016; Alpaydin 2020). Among them, artificial neural networks (ANN) have been particularly popular for modeling groundwater salinity (Alagha et al. 2017; Haselbeck et al. 2019; Nozari and Azadi 2019). Nevertheless, the research for spatial modeling of salinity and susceptibility mapping has been limited (Amiri-Bourkhani et al. 2017).

Although the application of novel machine learning methods has already been established in various aspects of groundwater modeling, salinity has been left behind, despite its importance. Consequently, the contribution of this paper has been set to investigate the application of various machine learning in groundwater salinity mapping for the first time. A comparative analysis has been designed to study the performance of the essential and ensemble machine learning methods. In this study, single machine learning methods, i.e., flexible discriminant analysis (FDA), multivariate adaptive regression spline (MARS), support vector machine (SVM), mixture discriminant analysis (MAD), and also ensemble machine learning methods, i.e., random forest (RF) and boosted regression tree (BRT), have been used to build the models.

Material and methods

Study area

The study case is the Karaj watershed which is placed between 50° 43′ 07″ to 51° 35′ 18″ E and 35° 06′ 19″ to 35° 58′ 31″ N with an area of 2144 km^2 and surroundings of 50 km on the southern section of Alborz Mountain range. The maximum elevation of the watershed is about 3294 m above mean sea level (a.m.s.l) in the northern region, and the lowest is 805 m a.m.s.l in the southern region (Fig. 1). The average daily temperature is about 15.8 °C, and the average yearly precipitation of the region is about 260 mm for the years from 1985 to 2017. Thus, the climate of the region is mostly arid/semi-arid, in which the amount of precipitation is less than one-third of the annual average rainfall of the Earth (Tabari and Aghajanloo 2013). Due to the climate of the region, the level of surface water is not adequate to meet all requirements, so there is a heavy reliance on groundwater as a freshwater resource that increases the concerns on its quantity and quality (Heidarnejad et al. 2006). Cities of Karaj, Shahriar, Eslamshahr, and Robat Karim, with a total population of 2,500,000 people, are located in this watershed that the primary source of drinking water in these cites is groundwater (about 80%) (Statistical Center of Iran 2016).

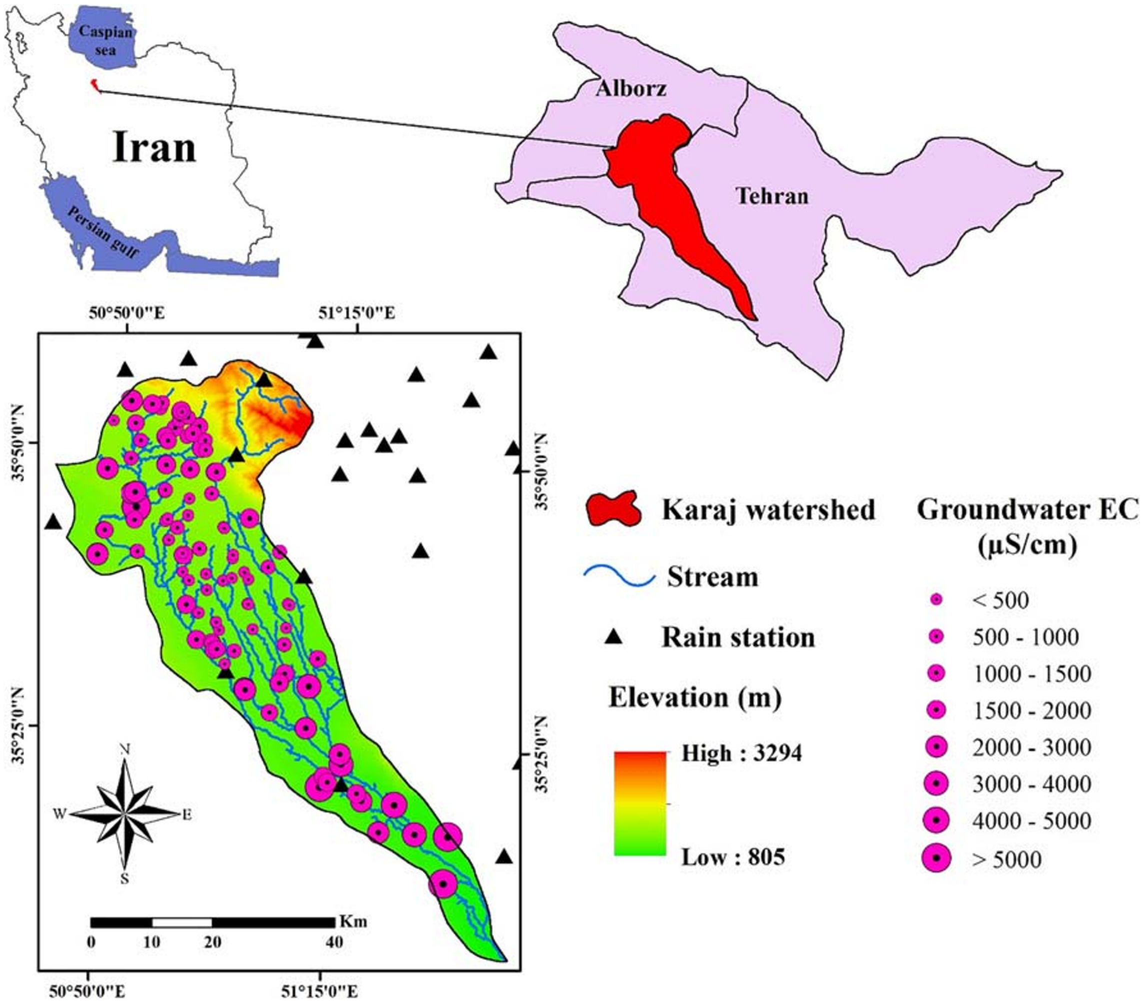


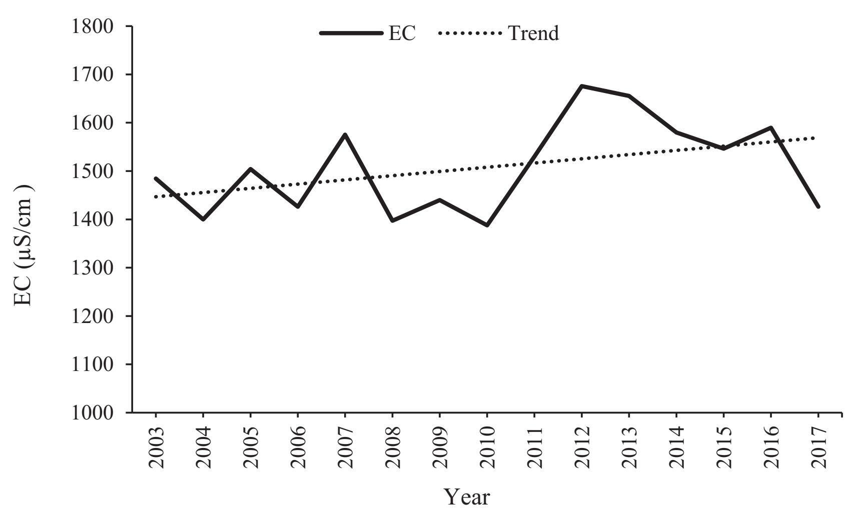
Fig. 1 Location of the Karaj watershed

Also, the temporal variations of the electrical conductivity (EC) for the watershed indicate a gradually increasing trend that increases the concerns of the water-management communities (Fig. 2).

Modeling process

The groundwater salinity modeling in this study is based on the dichotomous (yes/no) predictions by six machine learning

Fig. 2 Temporal variations of the EC values for the Karaj watershed



(ML) models. ML modeling is done based on the dependent and independent variables. The EC data, as the real and best way to express salinity (Rodríguez-Rodríguez et al. 2018; Song et al. 2019; Ali et al. 2017; Burger and Čelková 2003), was considered as the dependent variable, and the groundwater salinity influencing factors (GSIFs) were considered as independent variables. According to the national water standard of Iran (No. 1053) the maximum allowable EC is equal to 1500 μS/cm (ISIRI 2010; Amouei et al. 2012), so this threshold was considered to determine the saline and non-saline wells respectively by assigning values of 1 and 0 (as the dependent variable using the mean yearly EC values in each well). After that, the groundwater salinity was predicted using the important factors (identified by the feature selection). Description of the dataset, feature selection methodology, applied ML models, and performance analysis is presented as follows:

Dataset

Salinity data The EC data were obtained for 114 groundwater monitoring wells (Fig. 1) from 2003 to 2017 from the Iranian Water Resources Management Company (IWRMC). The mean annual values of EC during the period for each well are considered in this study, because data sampling was not regular during the years, and most of the months and some of the seasons not had any samples in a year, and this status was different during the different years. In Fig. 1, the spatially amount of mean EC for each well is presented. The lower EC values are mostly in the central parts, while the higher values are in the southern, north, and northwest parts (Fig. 1). Also, temporal variations of the EC for the watershed are presented in Fig. 2. The minimum and maximum average of EC is equal to 1388 and 1676 μS/cm for 2010 and 2012. Although there is fluctuation during the years, a gradually increasing trend of EC is observable during the period (Fig. 2).

Groundwater salinity influencing factors (GSIFs) The considered GSIFs in this study are categorized into four groups, including topographic, hydro-climate, groundwater, geologic, and land cover factors:

– Topographic factors: The considered topographic factors were including elevation, slope, aspect, curvature, and topographic position index (TPI) (Fig. 3a to e), which are calculated using the ASTER digital elevation model (DEM) 30 × 30 m in ArcGIS software. Factors of elevation, curvature, and TPI have an inverse relationship with groundwater salinity, as the accumulation of salinity in low elevations, lower curvatures (i.e., concave areas), and valleys' floor (i.e., lower TPI) is more than other regions due to groundwater drainage and washing the soil by surface runoff (Madyaka 2008; Shafapour Tehrany et al. 2013; Mosavi et al. 2020a). Different slopes and aspects can create different microclimate conditions (due to different sunlight, precipitation, evaporation, soil humidity, etc.) that affect the expansion of vegetation, generation, and infiltration of runoffs, and finally, other effects on groundwater.
– Hydro-climate factors: The considered hydro-climate factors were including distance from stream (DFS), precipitation (PCP), evaporation (Eva), and topographic wetness index (TWI) (Fig. 3f to i). The area near to streams and rivers can increase the groundwater recharge, but the increasing or decreasing groundwater salinity in this area is related to the dissolved materials during the infiltration. An increase in the amount of precipitation has an inverse impact on groundwater salinity, while an increase in evaporation has a direct impact on it (Newman and Goss 2000; Gholami et al. 2010). The TWI is a factor demonstrating the spatial pattern of soil moisture (Moore and Burch 1986). It shows the relation between the surface slope of the terrain and the amount of moisture at the surface (Pourghasemi and Beheshtirad 2014). However, the TWI influences the groundwater recharge and groundwater level.
– Groundwater factors: Two factors of withdrawal discharge and groundwater level (GWL) are considered (Fig. 3 j and k). There is a direct relationship between groundwater salinity and withdrawal discharge of groundwater, while reduction of groundwater level may lead to leaching the saline water into the aquifer and therefore increasing the salinity of groundwater (Gholami et al. 2010).
– Geologic factors: The geologic factors were including distance from fault (DFF), lithology, and soil order (Fig. 3l to n). The fault and distance to it may affect the quality of groundwater through conjunction with surface water (Mosavi al., 2020a). Geologically, most of the area is covered by the Quaternary formations (i.e., low- and high-level pediment fan and valley terrace deposits; fluvial conglomerate, piedmont conglomerate, and sandstone). However, other formations such as Karaj formation (i.e., well-bedded green tuff and tuffaceous shale), Upper red formation (i.e., red marl, gypsiferous marl, sandstone, and conglomerate), and Oligocene andesitic lava flows (Oav), andesitic volcanics (Eav), basaltic volcanic tuff (Ebvt), calcareous shale with subordinate tuff (Ek.a), and Gypsiferous marl (Murmg) are the main geological formations of the watershed (Fig. 3n). Different lithologies and soils control the amount of penetration, leaching of soil materials, and recharge of groundwater which affects the groundwater quality (Gowd 2004; Yun et al. 2011; Mosvi et al. 2020).

– Land cover factors: The land use types and distance from road (DFR) were the land cover factors in this study (Fig. 3 o and p). Drainage and infiltration of surface water into groundwater in different land uses such as agriculture, bareland, and urban areas have different impacts on groundwater quality. Also, changing the natural landscapes by humans, such as roads, where roads intersect saline geologic formations, can increase dissolving the saline materials into the surface water and groundwater (McRobert and Foley 1999; Liu et al. 2006). Moreover, other human activities such as de-icing of road surfaces using the salt pose serious risks through draining via runoff and seeping into groundwater.

Feature selection

Feature selection (FS) is a procedure for decreasing the number of input variables when the dimension of the input variables is large. It is a promising method to decrease the training times, dimensionality, and overfitting problems during the modeling (Bermingham et al. 2015). Two main categories of the FS methods are filters and wrappers (Lualdi and Fasano 2019). Filters are divided into univariate filters (such as correlation, information gain ratio) and multivariate filters (such as factor analysis, principal component analysis) which select the important variables based on a score or correlation function (Lualdi and Fasano 2019; Azzellino et al. 2019; Busico et al. 2018), but wrappers apply machine learning models using some search strategies such as genetic algorithms and simulated annealing (SA) (Lualdi and Fasano 2019). Since it has been empirically demonstrated that wrappers show better performance (Jović et al. 2015; Lualdi and Fasano 2019), so, in the present research, feature selection was conducted using the SA algorithm. The SA is a metaheuristic global optimization method in a large search space based on the minimum energy configuration theory (Bertsimas and Tsitsiklis 1993; Choubin et al. 2019; Sajedi Hosseini et al. 2020), whereby a solid is gradually cooled such that its structure is frozen (Choubin et al. 2020; Mosavi et al. 2020b). In this study in the feature selection process, the random forest is used as the estimator. The R environment was used to implement the SA method through the Caret package (Kuhn 2015).

Model description

Six machine learning (ML) models, namely, flexible discriminant analysis (FDA), multivariate adaptive regression spline (MARS), boosted regression tree (BRT), random forest (RF), support vector machine (SVM), mixture discriminant analysis (MAD), were employed to groundwater salinity mapping through k-fold cross-validation methodology by the SDM R package (Naimi and Araújo 2016). A brief description of each ML model is presented as follows:

The FDA model is an assortment model that depends on a combination of linear regression models, which utilizes optimal scoring to transform the response variable so that the data are in a better formation for linear separation, and multiple adaptive regression splines to produce the discriminant surface. Applying the FDA with standard linear regression yields Fisher's discriminant vectors (Hastie et al. 1994).

The MARS model is a non-parametric regression way that creates numerous linear regressions between the ranges of predictor amounts. By separating the dataset for feeding the linear regression model on every individual division, MARS provides a unique modeling capability. The MARS creates no suppositions about the connection between the dependent and the independent variables. It forms a set of basic functions (BF), which in this manner, the range of predictor amounts is separated into multiple groups. A separate linear regression is shaped for each group. The relations among the different regression lines are called knots. The MARS procedure attempts for the best spots to put the knots. Each knot has a pair of essential functions. These functions explain the connection between the independent and dependent variables (Leathwick et al. 2006).

The RF model is an ensemble learning manner for regression, classification, and further applications that acts by manufacturing plenty of decision trees at calibrating time and creation the class, which is the mode of the classes (classification) or average forecast (regression) of the exclusive trees (Ho 1995, 1998). The RF amends the problematic habit of decision trees, such as overfitting, and yields a low bias (Hastie et al. 2008). The RF can be utilized to rank the importance of variables in a classification or regression problem in a natural manner.

The BRT model is a robust algorithm and operates very well with the big dataset. BRT model is a composition of two methods: decision trees and boosting methods (Elith et al. 2008). Like the RF model, the BRT frequently fit plenty of decision trees to enhance the model accuracy. The major difference among the approaches is the strategy of choosing the decision trees. Either of the models picks a random subset of the entire dataset for every recently available tree. The RF model utilizes the bagging approach, which assures that every incidence has an equivalent probability of being chosen in future samples. In contrast, the BRT model applies the boosting approach in which the input data for the future trees is weighted. This strategy of managing the weights ensures that data that was poorly modeled by former trees has a higher probability of being chosen in the recent tree (DeAth 2007).

The SVM model is extremely preferred by plenty, as it generates considerable accuracy with less calculation power. SVM can be used for both regression and classification

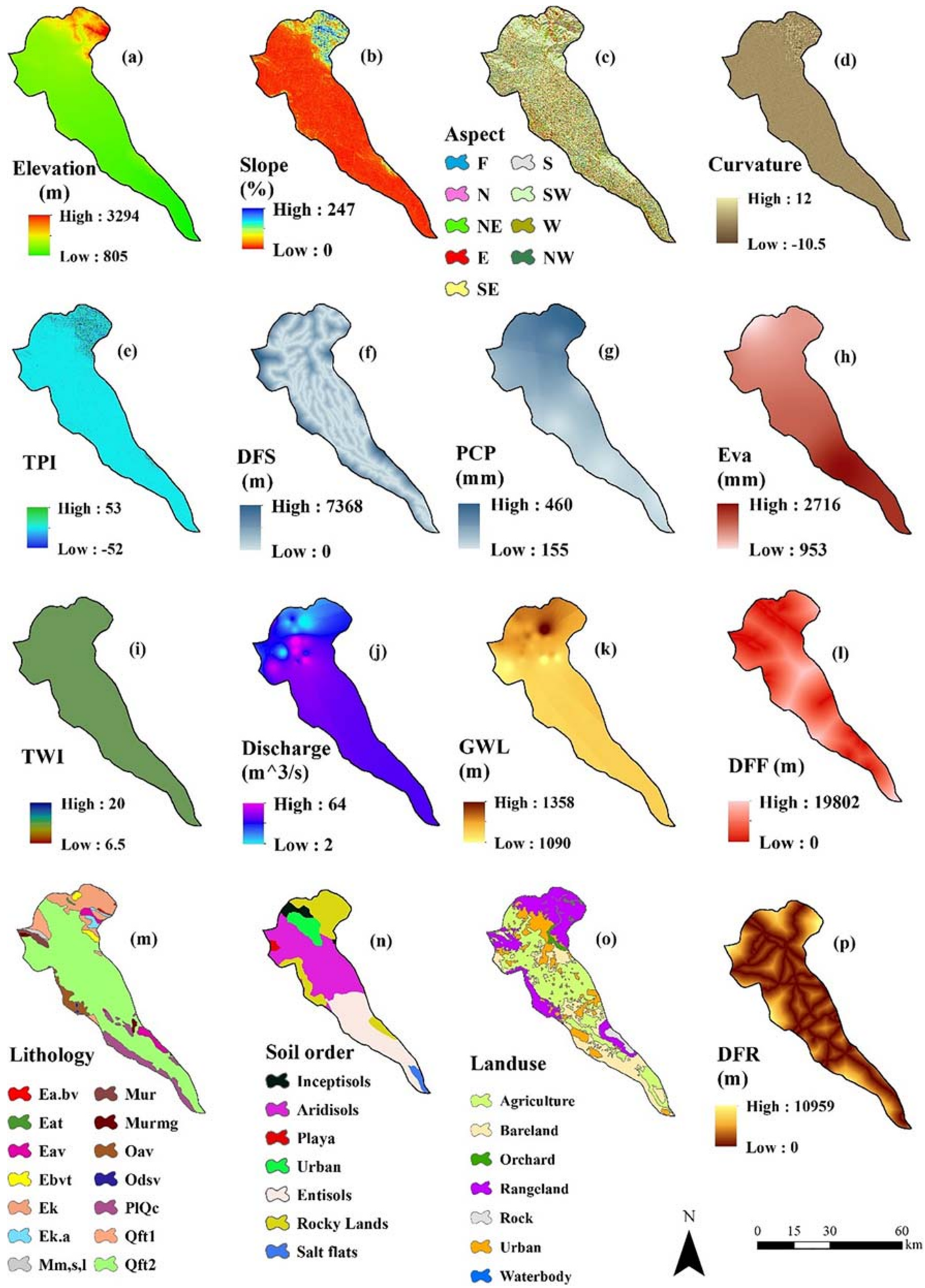


Fig. 3 Groundwater salinity influencing factor **a** elevation, **b** slope, **c** aspect, **d** curvature, **e** topographic position index (TPI), **f** distance from stream (DFS), **g** precipitation (PCP), **h** evaporation (Eva), **i** topographic wetness index (TWI), **j** discharge, **k** groundwater level (GWL), **l** distance from fault (DFF), **m** lithology, **n** soil order, **o** landuse, and **p** distance from road (DFR)

tasks. However, it is extensively used in classification purposes. The purpose of the algorithm is to obtain a hyperplane in N-dimensional space (N—the number of features) that separately assorts the data points. To discrete the two classes of data points, many feasible hyperplanes could be selected. Our purpose is to find a plane that has the maximum border (i.e., the maximum extent between data points of both classes). Maximizing the border extent supplies some amplification so that subsequent data points can be categorized with more reliance (Evgeniou and Pontil 2001).

For MAD, there are classes, and each class is supposed to be a Gaussian mixture of subclasses, where each data point has a possibility of depending on each class. Equalization of a covariance matrix, among classes, is still supposed. The model formulation is generative, and the posterior probability of class membership is applied to categorize an unlabeled observation. Each subclass is supposed to have its mediocre vector, but all subclasses share the same covariance matrix for model parsimony (Bashir and Carter 2005; Hastie and Tibshirani 1996).

Performance evaluation

Evaluation of the models' performance was conducted using some important metrics in dichotomous (yes/no) predictions, including accuracy (Eq. 1), kappa (Eq. 2), probability of detection (POD) (Eq. 4), false alarm ratio (FAR) (Eq. 5), critical success index (CSI) (Eq. 6), and Heidke skill score (HSS) (Eq. 7). Accuracy is the proportion of the number of correct predictions to the collected amount of input samples. The accuracy ranges from 0 to 1 (Pourghasemi et al. 2012). Kappa is between 0 and 1, where 0 shows the value of agreement that can be expected from random chance, and 1 shows complete agreement between the raters (Cohen 1960; Marston 2010). POD is the ratio of the amount of miss data to the total amount of observed occurrences. It confines from 0 to 1. For this statistic 1 is the perfect score (Wilks 1995). FAR is the proportion of the total false alarms (FA) to the total anticipated events (H + FA). The FAR can be improved by consistently forecasting scarce events. It ranges from 0 to 1, and the desired score is 0 (Barnes et al. 2007). CSI supplies no individual confirmation information since it is a function of both FAR and POD; understanding its conduct can aid in identifying which component would be more useful to purpose in an alarming strategy. Its range is 0 to 1, with a value of 1 showing a perfect forecast (JWGFVR 2009). The HSS is a popular metric because it is relatively ordinary to calculate, and probably because of the standard anticipation, chance, is relatively easy to beat. The HSS calculates the fractional improvement of the anticipation over the standard anticipation. The range of the HSS is −1 to 1.

$$\mathrm{Accuracy} = \frac{\mathrm{H+CN}}{\mathrm{H+FA+M+CN}} \tag{1}$$

$$\mathrm{Kappa} = \frac{\mathrm{Accuracy}-\mathrm{P_e}}{1-\mathrm{P_e}} \tag{2}$$

$$P_e = \frac{\mathrm{(H+FA)(H+M)+(M+CN)(FA+CN)}}{\mathrm{(H+FA+M+CN)}^2} \tag{3}$$

$$\mathrm{POD} = \frac{\mathrm{H}}{\mathrm{H+M}} \tag{4}$$

$$\mathrm{FAR} = \frac{\mathrm{FA}}{\mathrm{H+FA}} \tag{5}$$

$$\mathrm{CSI} = \frac{\mathrm{H}}{\mathrm{H+M+FA}} \tag{6}$$

$$\mathrm{HSS} = \frac{2[(\mathrm{H}\times\mathrm{CN})-(\mathrm{FA}\times\mathrm{M})]}{[\mathrm{(H+M)(M+CN)+(H+FA)(FA+CN)}]} \tag{7}$$

where H, FA, M, and CN are cells of the contingency table respectively indicating the number of hits, false alarms, misses, and correct negatives.

Results

Feature selection results

A trial and error (100 runs) were conducted to select the best number of features during each resampling fold (k). Figure 4 shows the variation in accuracy based on the number of features (size) and k-resampling folds and based on 100 iterations and k = 10 folds. For better understanding, Fig. 5 indicates the best number of features in each resampling which is determined based on the higher accuracy during 100 iterations. For example, Fold 5 had the best performance with a number of 10 features as input. As Fig. 5 shows, the best number of variables varies between 7 and 12 variables among 10-fold resamples. Therefore, the allowable minimum and a maximum number of features were respectively 7 and 12 variables. Finally, to select the best number of features, the occurrence frequency of variables in k-fold resample (%) was investigated. We decided eight variables (including GWL, PCP, Landuse, DFS, soil order, withdrawal discharge, elevation, and lithology) with occurrence frequency equal or more than 50% in 10-fold resamples (Table 1). Therefore, these variables were used for groundwater salinity mapping.

Model performance evaluation

The performance of the models is shown in Table 2. Although all of the models have an accuracy greater than 0.82, the SVM

Table 1 Selected features using the SA method

Selected features	Occurrence frequency in k-fold resample (%)
Groundwater level (GWL)	90
Precipitation (PCP)	80
Landuse	70
Distance from stream (DFS)	70
Soil order	70
Withdrawal discharge	60
Elevation	50
Lithology	50
Curvature	30
Evaporation	30
distance from road (DFR)	20
Topographic wetness index (TWI)	20
Topographic position index (TPI)	20
Slope	10
Aspect	10
Distance from fault (DFF)	10

model has higher accuracy (0.88) than others. The Kappa values vary between 0.64 and 0.76, respectively, for the BRT and SVM models, which indicate the good performance for the models (0.55 < kappa <0.85; Monserud and Leemans 1992). POD index is higher respectively for FDA, SVM, MARS, MDA, BRT, and RF (respectively equal to 0.94, 0.89, 0.83, 0.83, 0.78, and 0.78). Given the FAR, the RF and SVM models have lower false alarms. Also, the SVM model had higher CSI (CSI = 0.80) and HSS (HSS = 0.76) rather than other models (Table 2).

Generally, according to the metrics of accuracy, kappa, CSI, and HSS, the SVM model specified that had superior performance than other models. To the best of the authors' knowledge, there are no references for comparison which used machine learning for spatial mapping of groundwater salinity. However, like the results of this study, previous literature such as Alagha et al. (2017) and Isazadeh et al. (2017) have demonstrated the good potential of the SVM model in the time series (not spatial) prediction of the groundwater salinity.

Variable importance

Significance of the variables was analyzed through the Jackknife test (Efron 1982), and the decrease in area under the ROC curve (DAUC) (Choubin et al. 2019) was calculated after excluding each variable from the modeling process. Fig. 6 indicates the outcome of the sensitivity analysis for the best model (i.e., SVM), in which the greater DAUC indicates greater importance. According to the SVM results, the variables of soil order and withdrawal discharge with DAUC 27.7% and 13.4% were the most important variables that had a higher contribution in the groundwater salinity mapping. Also, the precipitation (PCP) (DAUC = 8.5%), land use (DAUC =

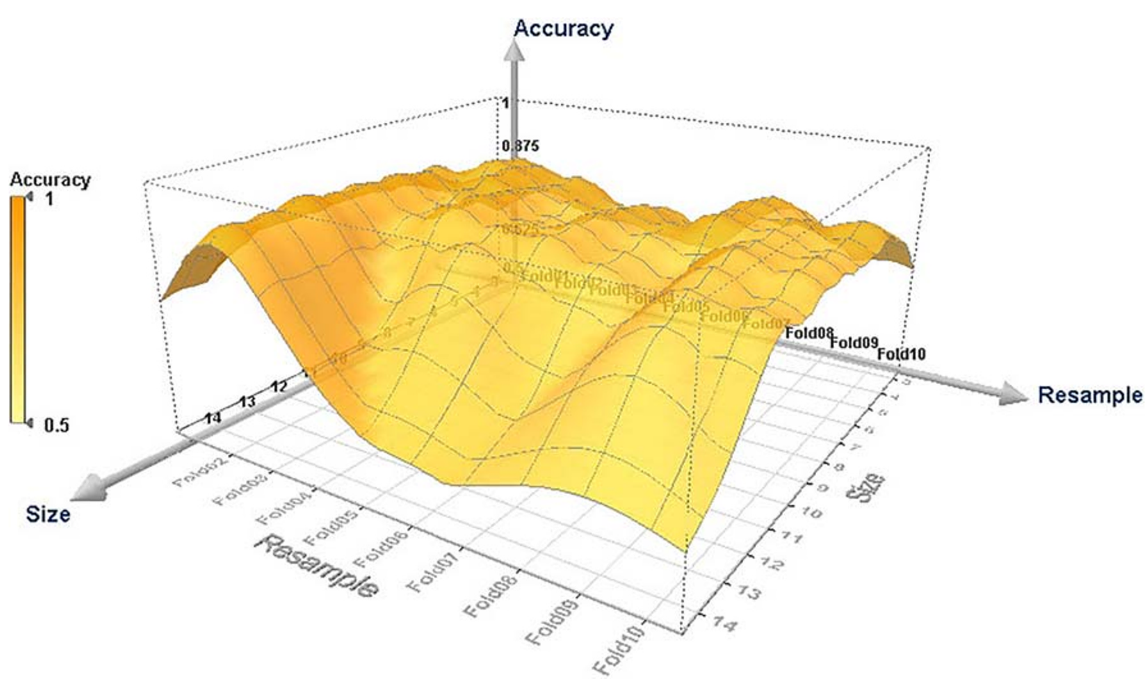


Fig. 4 Variation in the accuracy vs the number of features in resampling folds

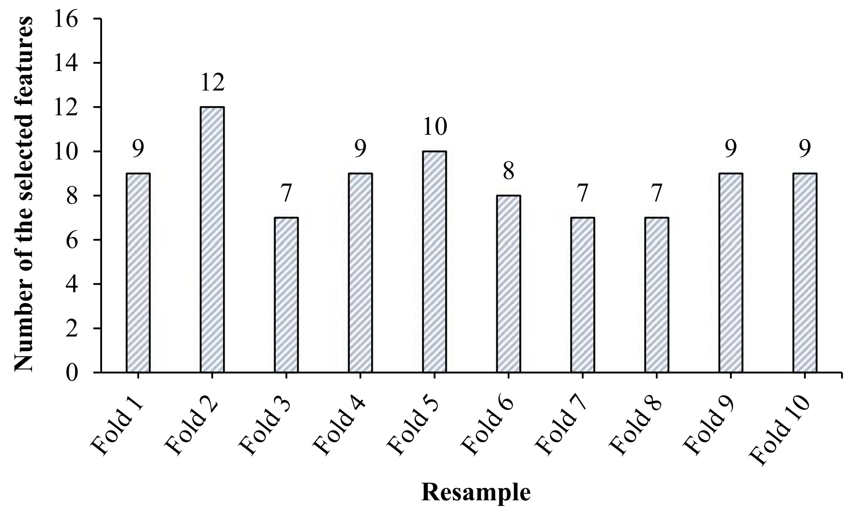


Fig. 5 The best number of the features in each resampling

8.1%), and elevation (DAUC = 7.8%) indicated a moderate importance (Fig. 6; Table 3). The results of the sensitivity analysis for other models are summarized in Table 3.

Groundwater salinity mapping

The pixels' value of predictor variables for the whole region was used as input to spatial mapping the groundwater salinity using the calibrated models. Therefore, a predicated map with cell size 30 × 30 m was produced for each model. Figure 7 shows the groundwater salinity maps predicted by the models, which were classified into three classes of low, moderate, and high regions based on the Natural Breaks method. This classification method is based on the Jenks optimization approach that identifies classes based on the minimum difference within classes and the maximum difference between classes (Brewer and Pickle 2002). Also, in other studies of groundwater quality/potential mapping (e.g., El-Hoz et al. 2014; Guru et al. 2017; Miraki et al. 2019; El-Meselhy et al. 2020; Barilari et al. 2020; Garewal et al. 2020) this method has been used. As can be seen, the high groundwater salinity

Table 2 Performance of the models

Metric	FDA	RF	BRT	MARS	SVM	MDA
Accuracy	0.85	0.85	0.82	0.85	0.88	0.85
Kappa	0.69	0.70	0.64	0.70	0.76	0.70
POD	0.94	0.78	0.78	0.83	0.89	0.83
FAR	0.19	0.07	0.13	0.12	0.11	0.12
CSI	0.77	0.74	0.70	0.75	0.80	0.75
HSS	0.69	0.70	0.64	0.70	0.76	0.70

Table 3 Importance of the variables based on the DAUC (%)

Variable	SVM	FDA	RF	BRT	MARS	MDA
Soil order	27.7	34.2	9.3	5.5	42.9	53.5
Withdrawal discharge	13.4	17.4	24.9	68.3	27.4	27.7
Precipitation (PCP)	8.5	44.4	5.0	1.5	49.1	22.5
Landuse	8.1	10.4	1.9	0.6	2.9	4.0
Elevation	7.8	7.1	6.8	1.9	26.4	12.8
Distance from stream (DFS)	4.3	0.5	0.9	0.7	0.2	4.7
Groundwater level (GWL)	3.2	0.2	1.9	1.6	28.6	1.7
Lithology	3.1	3.8	0.5	0.0	1.0	8.2

Fig. 6 Result of DAUC (%) for the SVM model

regions are mostly located in the southern parts and some parts of the north, northeast, and west regions. However, the middle of the region has a low groundwater salinity (Fig. 7). The high susceptibility regions mainly correspond to soil type of Entisols, bareland areas, and low elevations.

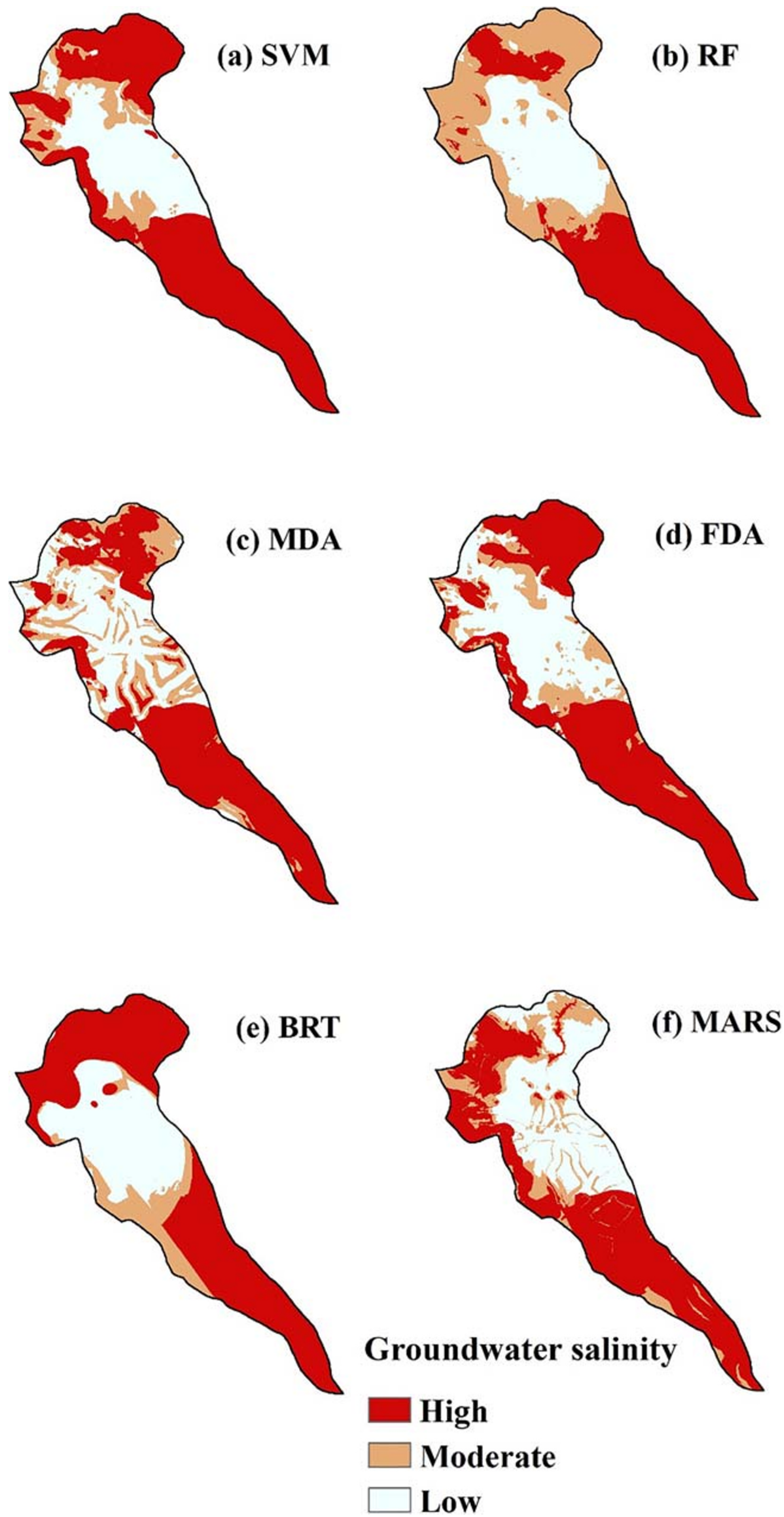


Fig. 7 Groundwater salinity map by **a** SVM, **b** FDA, **c** MDA, **d** FDA, **e** BRT, and **f** MARS

Discussion

Monitoring the groundwater quality and vulnerability assessment are the fundamental practices, especially in the arid and semi-arid areas where the access to the freshwater recourses are limited (Ben Ammar et al. 2020; Nahin et al. 2019). In this study using the potential of machine learning models, the susceptibility of the groundwater salinity in the Karaj watershed was modeled. The predicted high susceptibility regions in this study mainly correspond to bareland areas and low elevations. In the bareland areas, there is a significant salt in the soil that penetrates the groundwater during the recharge periods (He et al. 2014; Thiam et al. 2019). Also, high salinity in low lands is because of washing the soil materials by surface runoff and accumulating them in these areas that penetrate groundwater (Madyaka 2008; Shafapour Tehrany et al. 2013; Mosavi et al. 2020).

Intra-annual and inter-annual changes of the EC data are the main concern of the study that may affect the susceptibility mapping. In this study because of data monitoring conditions (i.e., irregular data sampling and lack of continuous samples for months and seasons), the mean annual values of EC during 2003–2017 for each well are considered, so, considering the inter-annual and intra-annual variations was not possible. To address the intra-annual concerns, susceptibility mapping for dry and wet seasons (i.e., discharge and recharge seasons) is recommended for future studies. Also, for addressing climate variability effects and inter-annual concern in future studies, dividing the period into equal sub-periods and spatial modeling for each sub-period can be a proper solution (Mosavi et al. 2020).

On the other hand, considering the land use and climate changes mostly for a long-term dataset is of utmost importance. Change in climate has been recently contributing in irregularity in temperature and precipitation that affect the groundwater, too; which has substantially increased the uncertainty of the predictive models (Akbari et al. 2020; Wang et al. 2019; Waqas et al. 2019; Geng and Boufadel 2017). Also, land use changes such as changing the rangelands to agricultural areas, and increasing the bare lands result in high salt concentrations in the soil. However, considering the effects of climate and land use changes during the long period dataset (and also for predictions of future salinity) is essential in the susceptibility mapping.

Conclusion

The application of six important machine learning methods for the first time has been evaluated in spatial modeling of the groundwater salinity. The modeling process indicated that the variables of soil order, groundwater withdrawal, precipitation, land use, and elevation were the most important variables that had the most contribute to the groundwater salinity mapping. Although all of the models had an accuracy greater than 0.82, the results highlighted that the SVM model had a superior performance than other models. The superior models of this study have potential value for water resource managers for identifying the vulnerable locations to plan and enforce effective policies to control and reduce the salinity of groundwater. For future research advancing hybrid machine learning methods are strongly proposed to identify an optimal model with a higher level of adaptivity, accuracy, and generalization ability.

Authors' contributions Conceptualization, AM and BC; data preparation, FSH; formal analysis, FSH, BC and AM; investigation, MG, AAD, and BN; methodology, BC, and FSH; project administration, AAD and BC; supervision, AM and BC; validation, FT; visualization, BN and AAD; writing—original draft, MG, BN, and FT; writing—review and editing, AAD.

Data availability Not applicable.

Compliance with ethical standards

Competing interests The authors declare that they have no conflict of interest.

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

References

Akbari M, Najafi Alamdarlo H, Mosavi SH (2020) The effects of climate change and groundwater salinity on farmers' income risk. Ecol Indic 110:105893

Alagha JS, Seyam M, Md Said MA, Mogheir Y (2017) Integrating an artificial intelligence approach with k-means clustering to model groundwater salinity: the case of Gaza coastal aquifer (Palestine). Hydrogeol J 25:2347–2361

Alberti L, Cantone M, Colombo L, Oberto G, La Licata I (2016) Assessment of aquifers groundwater storage for the mitigation of climate change effects. Rend Online Soc Geol Ital 39:89–92

Ali G, Haque A, Basu NB, Badiou P, Wilson H (2017) Groundwater-driven wetland-stream connectivity in the prairie pothole region: inferences based on electrical conductivity data. Wetlands 37:773–785

Alpaydin E (2020) Introduction to machine learning. MIT press

Amiri-Bourkhani M, Khaledian MR, Ashrafzadeh A, Shahnazari A (2017) The temporal and spatial variations in groundwater salinity in mazandaran plain, Iran, during a long-term period of 26 years. Geofizika 34:119–139

Amouei AI, Mahvi AH, Mohammadi AA, Asgharnia HA, Fallah SH, Khafajeh AA (2012) Physical and chemical quality assessment of potable groundwater in rural areas of Khaf, Iran. World Appl Sci J 18(5):693–697

Aydin BE, Rutten M, Abraham E, Oude Essink GHP, Delsman J (2017) Model predictive control of salinity in a polder ditch under high saline groundwater exfiltration conditions: a test case. IFAC-PapersOnLine 50:3160–3164

Azzellino A, Colombo L, Lombi S, Marchesi V, Piana A, Andrea M, Alberti L (2019) Groundwater diffuse pollution in functional urban areas: the need to define anthropogenic diffuse pollution background levels. Sci Total Environ 656:1207–1222

Banda KE, Mwandira W, Jakobsen R, Ogola J, Nyambe I, Larsen F (2019) Mechanism of salinity change and hydrogeochemical evolution of groundwater in the Machile-Zambezi Basin, South-Western Zambia. J Afr Earth Sci 153:72–82

Barnes LR, Gruntfest EC, Hayden MH, Schultz DM, Benight C (2007) False alarms and close calls: a conceptual model of warning accuracy. Wea Forecasting 22:1140–1147

Bashir S, Carter E (2005) High breakdown mixture discriminant analysis. J Multivar Anal 93(1):102–111

Ben Ammar S, Taupin JD, Ben Alaya M, Zouari K, Patris N, Khouatmia M (2020) Using geochemical and isotopic tracers to characterize groundwater dynamics and salinity sources in the Wadi Guenniche coastal plain in northern Tunisia. J Arid Environ 178:104150

Bermingham ML, Pong-Wong R, Spiliopoulou A, Hayward C, Rudan I, Campbell H, Wright AF, Wilson JF, Agakov F, Navarro P, Haley CS (2015) Application of high-dimensional feature selection: evaluation for genomic prediction in man. Scientific Reports 5(1):10312

Bertsimas D, Tsitsiklis J (1993) Simulated annealing. Stat Sci 8:10–15. https://doi.org/10.1214/ss/1177011077

Bourke SA, Hermann KJ, Hendry MJ (2017) High-resolution vertical profiles of groundwater electrical conductivity (EC) and chloride from direct-push EC logs. Hydrogeol J 25:2151–2162

Brewer CA, Pickle L (2002) Evaluation of methods for classifying epidemiological data on choropleth maps in series. Ann Assoc Am Geogr 92(4):662–681

Burger F, Čelková A (2003) Salinity and sodicity hazard in water flow processes in the soil. Plant Soil Environ 49(7):314–320

Busico G, Cuoco E, Kazakis N, Colombani N, Mastrocicco M, Tedesco D, Voudouris K (2018) Multivariate statistical analysis to characterize/discriminate between anthropogenic and geogenic trace elements occurrence in the Campania plain, southern Italy. Environ Pollut 234:260–269

Chien NP, Lautz LK (2018) Discriminant analysis as a decision-making tool for geochemically fingerprinting sources of groundwater salinity. Sci Total Environ 618:379–387

Choubin B, Mosavi A, Alamdarloo EH, Hosseini FS, Shamshirband S, Dashtekian K, Ghamisi P (2019) Earth fissure hazard prediction using machine learning models. Environ Res 179:108770. https://doi.org/10.1016/j.envres.2019.108770

Choubin B, Abdolshahnejad M, Moradi E, Querol X, Mosavi A, Shamshirband S, Ghamisi P (2020) Spatial hazard assessment of the PM10 using machine learning models in Barcelona, Spain. Sci Total Environ 701:134474

Chowdhury AH, Scanlon BR, Reedy RC, Young S (2018) Fingerprinting groundwater salinity sources in the Gulf coast aquifer system, USA. Hydrogeol J 26:197–213

Cohen J (1960) A coefficient of agreement for nominal scales. Educ Psychol Meas 20:37–46

Davoodi K, Darzi-Naftchali A, Aghajani-Mazandarani G (2019) Evaluating Drainmod-s to predict drainage water salinity and groundwater table depth during winter cropping in heavy-textured paddy soils. Irrig Drain 68:559–572

DeAth G (2007) Boosted trees for ecological modeling and prediction. Ecology 88(1):243–251

Delsman JR, de Louw PGB, de Lange WJ, Oude Essink GHP (2017) Fast calculation of groundwater exfiltration salinity in a lowland catchment using a lumped celerity/velocity approach. Environ Model Softw 96:323–334

Delsman JR, Van Baaren ES, Siemon B, Dabekaussen W, Karaoulis MC, Pauw PS, Vermaas T, Bootsma H, De Louw PGB, Gunnink JL, Wim Dubelaar C, Menkovic A, Steuer A, Meyer U, Revil A, Oude Essink GHP (2018) Large-scale, probabilistic salinity mapping using airborne electromagnetics for groundwater management in Zeeland, the Netherlands. Environ Res Lett 13

Duque C, Olsen JT, Sánchez-Úbeda JP, Calvache ML (2019) Groundwater salinity during 500 years of anthropogenic-driven coastline changes in the Motril-Salobreña aquifer (south East Spain). Environ Earth Sci 78

Efron B (1982). The Jackknife, the Bootstrap, and Other Resampling Plans 38 SIAM

El-Hoz M, Mohsen A, Iaaly A (2014) Assessing groundwater quality in a coastal area using the GIS technique. Desalin Water Treat 52(10–12):1967–1979

Elith J, Leathwick JR, Hastie T (2008) A working guide to boosted regression trees. J Anim Ecol 77(4):802–813

El-Meselhy A, Abdelhalim A, Nabawy BS (2020) Geospatial analysis in groundwater resources management as a tool for reclamation areas of New Valley (El-Oweinat), Egypt. J Afr Earth Sci 162:103720

Evgeniou T, Pontil M (2001) Support vector machines: theory and applications. In: Paliouras G., Karkaletsis V., Spyropoulos C.D. (eds) Machine learning and its applications. ACAI 1999. Lecture Notes in Computer Science, vol 2049. Springer, Berlin, Heidelberg

Gallardo AH (2013) Groundwater levels under climate change in the Gnangara system, Western Australia. Journal of water and climate change 4(1):52–62

Garewal, S.K., Vasudeo, A.D. and Ghare, A.D., 2020. Optimization of the GIS-based DRASTIC model for groundwater vulnerability assessment. *In Nature-inspired methods for metaheuristics optimization* (pp. 489–502). Springer, Cham

Geng X, Boufadel MC (2017) The influence of evaporation and rainfall on supratidal groundwater dynamics and salinity structure in a sandy beach. Water Resour Res 53:6218–6238

Gholami V, Yousefi Z, Zabardast Rostami H (2010) Modeling of ground water salinity on the Caspian southern coasts. Water Resour Manag 24:1415–1424. https://doi.org/10.1007/s11269-009-9506-2

Giannoccaro G, Scardigno A, Prosperi M (2017) Economic analysis of the long-term effects of groundwater salinity: bringing the farmer's perspectives into policy. J Integr Environ Sci 14:59–72

Gil-Márquez JM, Barberá JA, Andreo B, Mudarra M (2017) Hydrological and geochemical processes constraining groundwater salinity in wetland areas related to evaporitic (karst) systems. A case study from southern Spain. J Hydrol 544:538–554

Gowd SS (2004) Electrical resistivity survey to delineate groundwater potential aquifers in Peddavanka watershed, Anantapur District, Andhra Pradesh, India. Environ Geol 46:118–131

Guru B, Seshan K, Bera S (2017) Frequency ratio model for groundwater potential mapping and its sustainable management in cold desert, India. Journal of King Saud University-Science 29(3):333–347

Haselbeck V, Kordilla J, Krause F, Sauter M (2019) Self-organizing maps for the identification of groundwater salinity sources based on hydrochemical data. J Hydrol 576:610–619

Hastie T, Tibshirani R (1996) Discriminant analysis by gaussian mixture. J R Stat Soc 58(1):155–176

Hastie T, Tibshirani R, Buja A (1994) Flexible discriminant analysis by optimal scoring. J Am Stat Assoc 89(428):1255–1270

Hastie T, Tibshirani R, Friedman J (2008). The Elements of Statistical Learning (2nd ed.). Springer. ISBN 0–387–95284-5

He B, Cai Y, Ran W, Jiang H (2014) Spatial and seasonal variations of soil salinity following vegetation restoration in coastal saline land in eastern China. Catena 118:147–153

Hebb DO (1949) The organization of behavior: a neuropsychological theory. J. Wiley; Chapman & Hall

Heidarnejad M, Golmaee SH, Mosaedi A, Ahmadi MZ (2006) Estimation of sediment volume in Karaj dam reservoir (Iran) by hydrometry method and a comparison with hydrography method. Lake Reserv Manag 22:233–239

Ho TK (1995). Random decision forests. Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, 14–16 August 1995. pp. 278–282

Ho TK (1998) The random subspace method for constructing decision forests. IEEE Trans Pattern Anal Mach Intell 20(8):832–844

Hosseini SM, Parizi E, Ataie-Ashtiani B, Simmons CT (2019) Assessment of sustainable groundwater resources management using integrated environmental index: case studies across Iran. Sci Total Environ 676:792–810

Hosseini FS, Choubin B, Mosavi A, Nabipour N, Shamshirband S, Darabi H, Haghighi AT (2020) Flash-flood hazard assessment using ensembles and Bayesian-based machine learning models: application of the simulated annealing feature selection method. Sci Total Environ 711:135161

Institute of Standards and Industrial Research of Iran (ISIRI) (2010) Physical and chemical quality of drinking water, Fifth edn, No. 1053, Tehran. Available from: http://www.isiri.org/std/1053.pdf/

Isazadeh M, Biazar SM, Ashrafzadeh A (2017) Support vector machines and feed-forward neural networks for spatial modeling of groundwater qualitative parameters. Environ Earth Sci 76(17):610

Jović A, Brkić K, Bogunović N (2015) A review of feature selection methods with applications. In 2015 *38th international convention on information and communication technology, electronics and microelectronics (MIPRO)* (pp. 1200–1205). Ieee

JWGFVR (2009) Recommendation on verification of precipitation forecasts. WMO/TD report, no.1485 WWRP 2009–1

Khan S, Mushtaq S, Hanjra MA, Schaeffer J (2008) Estimating potential costs and gains from an aquifer storage and recovery program in Australia. Agric Water Manag 95(4):477–488

Kuhn M (2015) Caret: classification and regression training. Astrophysics Source Code Library. http://adsabs.harvard.edu/abs/2015ascl.soft05003K

Kundzewicz ZW, Doell P (2009) Will groundwater ease freshwater stress under climate change? Hydrol Sci J 54(4):665–675

Leathwick JR, Elith J, Hastie T (2006) Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. Ecol Model 199(2):188–196

Levanon E, Yechieli Y, Gvirtzman H, Shalev E (2017) Tide-induced fluctuations of salinity and groundwater level in unconfined aquifers—field measurements and numerical model. J Hydrol 551: 665–675

Li C, Gao X, Liu Y, Wang Y (2019) Impact of anthropogenic activities on the enrichment of fluoride and salinity in groundwater in the Yuncheng Basin constrained by Cl/Br ratio, δ18O, δ2H, δ13C and δ7Li isotopes, Journal of Hydrology, 579

Lipczynska-Kochany E (2018) Effect of climate change on humic substances and associated impacts on the quality of surface water and groundwater: a review. Sci Total Environ 640:1548–1565

Liu G, Engineer JG, Transportation R, Jin SY (2006) September. Trend analysis of road salt impacts on groundwater salinity at a long-term monitoring site. In 2006 Annual Conference of the Transportation Association of Canada: Transportation Without Boundaries, TAC/ATC, September 17–20

Lu SC (1990) Machine learning approaches to knowledge synthesis and integration tasks for advanced engineering automation. Comput Ind 15(1–2):105–120

Lualdi M, Fasano M (2019) Statistical analysis of proteomics data: a review on feature selection. J Proteome 198:18–26

M'nassri S, Dridi L, Schäfer G, Hachicha M, Majdoub R (2019) Groundwater salinity in a semi-arid region of central-eastern Tunisia: insights from multivariate statistical techniques and geostatistical modelling, Environmental Earth Sciences, 78

Madyaka M (2008) Spatial modeling and prediction of soil salinization using SaltMod in a GIS environment. J. ITC., thesis in Geoinformation science and earth observation

Marston L (2010) Introductory statistics for health and nursing using SPSS. Sage Publications, Ltd., Thousand Oaks, California

Masciopinto C, Liso IS, Caputo MC, De Carlo L (2017) An integrated approach based on numerical modelling and geophysical survey to map groundwater salinity in fractured coastal aquifers, Water (Switzerland), 9

Mas-Pla J, Menció A (2019) Groundwater nitrate pollution and climate change: learnings from a water balance-based analysis of several aquifers in a western Mediterranean region (Catalonia). Environ Sci Pollut Res 26(3):2184–2202

McRobert J, Foley G 1999. The impacts of waterlogging and salinity on road assets: a Western Australian case study (No. 57)

Miraki S, Zanganeh SH, Chapi K, Singh VP, Shirzadi A, Shahabi H, Pham BT (2019) Mapping groundwater potential using a novel hybrid intelligence approach. Water Resour Manag 33(1):281–302

Monserud RA, Leemans R (1992) Comparing global vegetation maps with the kappa statistic. Ecol Model 62(4):275–293

Moore ID, Burch GJ (1986) Sediment transport capacity of sheet and rill flow application of unit stream power theory. Water Resour Res 22: 1350–1360

Mosavi A, Hosseini FS, Choubin B, Goodarzi M, Dineva AA (2020a) Groundwater salinity susceptibility mapping using classifier ensemble and Bayesian machine learning models. IEEE Access 8:145564–145576

Mosavi A, Sajedi-Hosseini F, Choubin B, Taromideh F, Rahi G, Dineva AA (2020b, 1995) Susceptibility mapping of soil water erosion using machine learning models. Water 12(7)

Nahin KTK, Basak R, Alam R (2019) Groundwater vulnerability assessment with DRASTIC index method in the salinity-affected southwest coastal region of Bangladesh: a case study in Bagerhat Sadar, Fakirhat and Rampal, Earth Systems and Environment

Naimi B, Araújo MB (2016) Sdm: a reproducible and extensible R platform for species distribution modelling. Ecography 39:368–375

Naser AM, Wang Q, Shamsudduha M, Chellaraj G, Joseph G (2020) Modeling the relationship of groundwater salinity to neonatal and infant mortality from the Bangladesh demographic health survey 2000 to 2014. GeoHealth 4:e2019GH000229

Newman B, GossK (2000) The Murray-Darling Basin salinity management strategy implications for the irrigation sector, Murray Darling Basin Commission: Proceeding of the 47th annual ANCID Conference, 10–13 September, p: 1–12, Towoomba, Australia

Nozari H, Azadi S (2019) Experimental evaluation of artificial neural network for predicting drainage water and groundwater salinity at various drain depths and spacing. Neural Comput & Applic 31: 1227–1236

Odeh T, Mohammad AH, Hussein H, Ismail M, Almomani T (2019) Over-pumping of groundwater in Irbid governorate, northern Jordan: a conceptual model to analyze the effects of urbanization and agricultural activities on groundwater levels and salinity, Environ Earth Sci, 78

Pourghasemi HR, Pradhan B, Gokceoglu C (2012) Application of fuzzy logic and analytical hierarchy process (AHP) to landslide susceptibility mapping at Haraz watershed. Iran Nat Hazards 63:965–996

Pauw PS, Groen J, Groen MMA, van der Made KJ, Stuyfzand PJ, Post VEA (2017) Groundwater salinity patterns along the coast of the Western Netherlands and the application of cone penetration tests. J Hydrol 551:756–767

Pourghasemi HR, Beheshtirad M (2014) Assessment of a data-driven evidential belieffunction model and GIS for groundwater potential mapping in the Koohrang Water-shed, Iran. Geocarto Int. https://doi.org/10.1080/10106049.2014.966161

Rodríguez-Rodríguez M, Fernández-Ayuso A, Hayashi M, Moral-Martos F (2018) Using water temperature, electrical conductivity, and pH to characterize surface-groundwater relations in a shallow Ponds System (Doñana National Park, SW Spain), Water (Switzerland), 10

Samuel AL (1959) Some studies in machine learning using the game of checkers. IBM J Res Dev 3(3):210–229

Sang S, Zhang X, Dai H, Hu BX, Ou H, Sun L (2018) Diversity and predictive metabolic pathways of the prokaryotic microbial community along a groundwater salinity gradient of the Pearl River Delta, China. Sci Rep 8

Shafapour Tehrany M, Pradhan B, Jebur MN (2013) Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical. Models in GIS. J Hydrol 504:69–79

Sofiyan Abuelaish B, Camacho Olmedo MT (2018) Analysis and modelling of groundwater salinity dynamics in the Gaza strip. In Cuadernos Geograficos, 72–91. University of Granada

Song JH, Shi XY, Cui J, Zhu YQ, Wang HJ (2019) Assessment of the accuracy of a soil salinity model for shallow groundwater areas in Xinjiang based on electromagnetic induction. Appl Ecol Environ Res 17:10037–10057

Statistical center of Iran (2016) Results of the 2016 national population and housing census (industry section). https://www.amar.org.ir/english/Statistics-by-Topic/Industry#2221489-time-series

Stoll S, Hendricks Franssen HJ, Butts M, Kinzelbach WK (2011) Analysis of the impact of climate change on groundwater related hydrological fluxes: a multi-model approach including different downscaling methods. Hydrol Earth Syst Sci 15(1):21–38

Stuart ME, Gooddy DC, Bloomfield JP, Williams AT (2011) A review of the impact of climate change on future nitrate concentrations in groundwater of the UK. Sci Total Environ 409(15):2859–2873

Suzuki K, Kusano Y, Ochi R, Nishiyama N, Tokunaga T, Tanaka K (2017) Electromagnetic exploration in high-salinity groundwater zones: case studies from volcanic and soft sedimentary sites in coastal Japan. Explor Geophys 48:95–109

Tabari H, Aghajanloo MB (2013) Temporal pattern of aridity index in Iran with considering precipitation and evapotranspiration trends. Int J Climatol 33:396–409

Tavakoli-Kivi S, Bailey RT, Gates TK (2019) A salinity reactive transport and equilibrium chemistry model for regional-scale agricultural groundwater systems. J Hydrol 572:274–293

Thiam S, Villamor GB, Kyei-Baffour N, Matty F (2019) Soil salinity assessment and coping strategies in the coastal agricultural landscape in Djilor district, Senegal. Land Use Policy 88:104191

Wang HY, Guo HM, Xiu W, Bauer J, Sun GX, Tang XH, Norra S (2019) Indications that weathering of evaporite minerals affects groundwater salinity and As mobilization in aquifers of the northwestern Hetao Basin, China, Applied Geochemistry, 109

Waqas MM, Shah SHH, Awan UK, Arshad M, Ahmad R (2019) Impact of climate change on groundwater fluctuation, root zone salinity and water productivity of sugarcane: a case study in lower Chenab canal system of Pakistan. Pak J Agric Sci 56:443–450

Wilks DS (1995) Statistical methods in the atmospheric sciences: an introduction. Academic Press, 467pp

Wuest T, Weimer D, Irgens C, Thoben KD (2016) Machine learning in manufacturing: advantages, challenges, and applications. Production & Manufacturing Research 4(1):23–45

Xiao K, Li H, Xia Y, Yang J, Wilson AM, Michael HA, Geng X, Smith E, Boufadel MC, Yuan P, Wang X (2019) Effects of tidally varying salinity on groundwater flow and solute transport: insights from modelling an idealized creek marsh aquifer. Water Resour Res 55:9656–9672

Yihdego Y, Webb JA, Vaheddoost B (2017) Highlighting the role of groundwater in lake–aquifer interaction to reduce vulnerability and enhance resilience to climate change. Hydrology 4(1):10

Yu L, Liu H (2003) Feature selection for high-dimensional data: a fast correlation-based filter solution. In Proceedings of the 20th international conference on machine learning (ICML-03) (pp. 856–863)

Yun P, Huili G, Demin Z, Xioaojuan L, Nobukazu N (2011) Impact of land use change on groundwater recharge in Guishui River basin, China. Chin Geogra Sci 21(6):734–743