Groundwater Salinity Susceptibility Mapping Using Classifier Ensemble and Bayesian Machine Learning Models

AMIRHOSEIN MOSAVI1,2, FARZANEH SAJEDI HOSSEINI3, BAHRAM CHOUBIN4, MASSOUD GOODARZI5, AND ADRIENN A. DINEVA6,7

1Environmental Quality, Atmospheric Science and Climate Change Research Group, Ton Duc Thang University, Ho Chi Minh City, Vietnam
2Faculty of Environment and Labour Safety, Ton Duc Thang University, Ho Chi Minh City, Vietnam
3Reclamation of Arid and Mountainous Regions Department, Faculty of Natural Resources, University of Tehran, Karaj 31585-77871, Iran
4West Azerbaijan Agricultural and Natural Resources Research and Education Center, Soil Conservation and Watershed Management Research Department, AREEO, Urmia 57169-63963, Iran
5Soil Conservation and Watershed Management Research Institute (SCWMRI), AREEO, Tehran 13898-17611, Iran
6Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam
7Kalman Kando Faculty of Electrical Engineering, Obuda University, 1034 Budapest, Hungary

Corresponding authors: Bahram Choubin (b.choubin@areeo.ac.ir) and Adrienn A. Dineva (adrienndineva@duytan.edu.vn)

ABSTRACT Risk and susceptibility mapping of groundwater salinity (GWS) are challenging tasks for groundwater quality monitoring and management. Advancement of accurate prediction systems is essential for the identification of vulnerable areas in order to raise awareness about the potential salinity susceptibility and protect the groundwater and top-soil in due time. In this study, three machine learning models of Stochastic Gradient Boosting (StoGB), Rotation Forest (RotFor), and Bayesian Generalized Linear Model (Bayesglm) are developed for building prediction models and their performance evaluated in the delineation of salinity susceptibility maps. Both natural and human effective factors (16 features) were used as predictors for groundwater salinity modeling and were randomly divided into the training (80%) and testing (20%) datasets. The models were evaluated using testing datasets after calibration using the selected features by recursive feature elimination (RFE) method. The RFE indicated that modeling with 8 features had better performance among 1 to 16 features (Accuracy = 0.87). Results of the groundwater salinity prediction highlighted that StoGB had a good performance, whereas the RotFor and Bayesglm had an excellent performance based on the Kappa values (> 0.85). Although spatial prediction of the models was different, all of the models indicated that central parts of the region have a very high susceptibility which matches with agricultural areas, lithology map, the locations with low depth to groundwater, low slope, and elevation. Additionally, areas near to the Maharlu lake and locations with a high decline in groundwater are also located in the very high susceptibility zone, which can confirm the effects of saltwater intrusion. The susceptibility maps produced in this study are of utmost importance for water security and sustainable agriculture.

INDEX TERMS Groundwater salinity, hazard, recursive feature elimination, stochastic gradient boosting, rotation forest, Bayesian generalized linear model.

I. INTRODUCTION

In arid areas, the essentials of life, economic growth, agriculture, and industries highly rely on groundwater as essential natural storage to make up against a shortage of surface water [1]–[4]. While the climate change, rapid urban development, unsewered disposal, constructions, mining, and other industrial pollutants are constantly threatening the quality and quantity of groundwater; generally the oil and gas production, as well as agricultural practices of overusing pesticides and fertilizers more than any anthropogenic activities, are contributing to the groundwater contamination and scarcity [5]–[16]. Various salts (including nontoxic and toxic ions), minerals, heavy metals, microorganisms, pathogens, petroleum pollutants, and nutrients are among the major contaminants [17]–[22]. The salinity of groundwater may have various origins of natural and man-made causes [23]. As one of the main natural causes, the rock weathering
phenomenon realizes minerals, and over time, it is washed to the groundwater causing salinity [24], [25]. However, during the past decades, the excessive withdrawal of groundwater using numerous deep wells for irrigation purposes has been part of the major contributor to the rise of unnatural salinity level [26] and also saltwater intrusion (SWI) [27] resulting in a depletion of groundwater resources for sustainable and quality irrigation which has become a major environmental concern [24], [28]. Furthermore, irrigation using water with a high level of groundwater salinity (GWS) can permanently damage the top-soil through continuous salinization of the land surface [29]–[32]. Thus, risk mapping and salinity susceptibility assessment of the groundwater in the early stages is very important to protect soil, groundwater, and crop production [33]–[35].

Producing risk maps and susceptibility prediction models, particularly in arid areas, is essential for effective groundwater quality monitoring and management so as to regulate and enforce policies in due time. As the recovery from salinity and other contaminations may take a long time (due to the slow movement of groundwater), the environmental protection policies have been focused on prevention [36]. Delineation of high-risk zones, mapping vulnerable areas to salinity, and hazard prediction have become essential for sustainable land use planning, agricultural strategies and crop selection, water supply management, water security, and policy-making [37]–[42]. Literature includes various scenarios of risk maps and hazard prediction models used for informed policy-making in arid areas around the world [43]–[50].

Physically-based salinity modeling includes the geochemical, hydrogeochemical, hydro-chemical as well as solute transport modeling tools which are often used for identifying the source and dynamics of salinity and risk assessment of groundwater through considering isotopic or non-isotopic geochemical and physicochemical parameters of aquifer [51]–[55]. Nassery and Kayhomayoon [56] and Stephens et al. [57] present models based on geophysical log analysis and aqueous geochemical measurements for salinity mapping in the arid area of Isfahan, Iran, and California, USA, respectively. Geochemical and geostatistical models integrated with GIS-based mapping have recently been used for mapping, monitoring the salinity, and assessment of the groundwater governance [51], [55], [58], [59]. Computational hydrologic modeling platforms, e.g. DRAINMOD [60], have been widely used to predict salinity of drainage water used in irrigation [61]. A number of numerical models and finite element analyses have also been used for salinity prediction, e.g., SUTRA, RMA-10 [62]. Statistical based models such as multivariate statistical analysis [63], [64], cluster analysis and multiple linear regressions (MLR) [59], [65]–[67], linear discriminant analysis [68], transient modeling [69], Bayesian kriging [70] have been used to study groundwater salinity and its origin. Sofiyan Abuelaish and Camacho Olmedo [71] developed a statistical model based on linear and multiple regression analysis to predict the 5-year ahead salinity value in the arid areas of Gaza, Palestine. Amiri-Bourkhani et al. [72] and Ashrafzadeh et al. [73] used geostatistical data and kriging methods to develop predictive models for probabilistic and zoning map of salinity in the arid areas of Iran. Time series analysis for modeling of groundwater salinity can be found in the literature [74], [75]. For instance, Cook et al. [76] predicted the groundwater salinity using a distributed parameter recharge model in the SW Murray Basin. They used variables of depth to groundwater, recharge, water content of the unsaturated zone, and soil salinity as input data. Gundogdu and Akkaya Aslan [77] in a similar case predicted the salinity maps. However, time-series prediction for salinity is associated with a number of drawbacks including accuracy and generalization [78], [79]. Artificial intelligence (AI) methods and applications have already been reported as efficient and popular methods in groundwater modeling. However, among the AI methods, a limited number of soft computing models are available in literature for modeling the salinity hazard or irrigation risk mapping, e.g. Abdi et al. [80] proposed a fuzzy logic model and Attwa et al. [81] genetic algorithm (GA) in the arid areas of Iran and Egypt respectively. The use of machine learning methods also has been limited in a few modeling experiments even though they appeared more convenient and performed with higher accuracy, better generalization ability, more reliability, and efficiency compared to the statistical, numerical, and physical models. For instance, Alagha et al. [82] used artificial neural networks (ANNs) and support vector machine (SVM) to model salinization processes with higher performance and more simplicity compared to statistical methods. Yu et al. [53] also reported machine learning models of back propagation ANN (BP-ANN) and neuro-fuzzy (NF) superior to the conventional linear models in the prediction of groundwater salinity hazards. Although applications of machine learning methods in various realms of hydrological modeling have already been established, they have not received adequate attention for groundwater modeling, particularly for salinity mapping and hazard prediction. Thus, the literature misses the application of various machine learning methods for modeling spatial groundwater salinity and the evaluation studies on their performance and accuracy.

This study in order to fill an existing gap in literature aimed to predict groundwater salinity probability for understanding susceptible areas and supporting manager communities to adopt the best decisions. Therefore, the main purposes of the study were: (i) to identify key variables related to the groundwater salinity, (ii) to predict the spatial probability of the groundwater salinity, (iii) to compare the performance of machine learning models in prediction of the groundwater salinity probability, and (iv) to determine susceptible regions of the study area in view of the groundwater salinity.

II. MATERIAL AND METHODS
A. DESCRIPTION OF STUDY AREA
Sarvestan plain, which is a coastal plain in the south of Maharlu lake in the Fars province, Iran, is extended between
The literature, values of EC more than 2250 µS/cm highlight an unsuitable water quality and it is even not suitable for irrigation [89]. Therefore, a threshold equal to 2250 µS/cm for EC was considered to separate saline and non-saline wells and assignment of the values 0 and 1 to model groundwater salinity. The mean yearly EC data (from 2003 to 2017) for 83 monitoring wells varies from 736.4 to 70257.5 µS/cm. Due to the high level of EC in the study area, understanding susceptible and safe zones in view of groundwater salinity could help managers to have better planning for groundwater utilization and assignments, as well as to control susceptible areas.

C. GROUNDWATER SALINITY CONDITIONING FACTORS (GSCF)

The origin of the groundwater salinity is because of the co-existence of both natural factors and human activities [54], [81]. According to a survey conducted by researchers within literature, e.g. [88], [91], both natural and human effective factors on groundwater salinity including topographic factors, topographic wetness index (TWI), distance from stream (DFS), distance from lake (DFL), distance from fault (DFF), depth of groundwater (DTGW), groundwater withdrawal (GWW), decline of groundwater level (DGWL), evaporation, precipitation, land use, lithology, and soil type of the study area were considered as GSCF in this study.

Topographic factors such as elevation, slope, aspect, and curvature (Fig. 2a,b,c,d) are effective in flushing and exporting the salty materials from the soil into fluvial plains, and transporting and accumulating it into lowlands. An ALOS PALSAR Digital Elevation Model (DEM) 12.5 × 12.5 m (https://vertex.daac.asf.alaska.edu/) was used to extract topographic factors. TWI (Fig. 2e) and DFS (Fig. 2f) can affect the groundwater through waterlogging and infiltration of surface water, leaching and dissolving of solid materials within the unsaturated zone into groundwater [92]. Also, DFL (Fig. 2g) as an index indicating proximity to a lake for considering effects of the intrusion probability of saline water from the lake was considered. Existence of fault in the study area can affect the quality of the groundwater through conjunction with surface water, so, DFF (Fig. 2h) was considered. Layers of DFS, DFL, and DFF were calculated using the Euclidean tool in ArcGIS. Also, TWI was calculated using the SAGA GIS software.

From the groundwater, three indices DTGW (Fig. 2i), GWW (Fig. 2j), and DGWL (Fig. 2k) were calculated. DTGW (also called water-table depth) could be affected by the evaporation, irrigation water, and conjunction with rivers [88]. Also, groundwater salinity can be increased using GWW and DGWL [93], this increases the potential for saline water intrusion into the groundwater. The groundwater data for each well were received from the IWRMC, and a yearly average map from 2003 to 2017 was produced in the ArcGIS software.

Climate data such as evaporation and precipitation are the natural factors that have effects on the groundwater. Precipitation can cause groundwater salinity through
FIGURE 2. Groundwater salinity conditioning factors: (a) elevation, (b) slope, (c) aspect, (d) curvature, (e) topographic wetness index (TWI), (f) distance from stream (DFS), (g) distance from lake (DFL), (h) distance from fault (DFF), (i) depth to groundwater (DTGW), (j) groundwater withdrawal (GWW), (k) decline of groundwater level (DGWL), (l) evaporation, (m) precipitation, (n) landuse, (o) lithology, and (p) soil type.
recharge, infiltration, and dissolving materials within the soil [91]. Also, evaporation can affect the groundwater because some parts of the study have low DTGW and were near to the earth’s surface (about 3 m, Fig. 2i). The evaporation and precipitation data were received from IWRMC and a yearly average map from 2003 to 2017 for them was produced using the IDW method (Fig. 2i and m). Landuse such as domestic and industrial wastewaters, agricultural fertilizers, and irrigation water can be affected the groundwater salinity [33], [40], [52], [56]. Soil type and lithology, directly and indirectly, are effective in groundwater salinity. Direct effects include the presence of salts and evaporative rocks in the saturated and unsaturated zones, as well as water-rock interactions [91], and indirect effects are an influence on recharge and infiltration rates. Landuse, lithology, and soil type maps of the study area were received from IWRMC.

D. FEATURE SELECTION

Although features considered in this study (GSCF, Fig. 2) were based on a survey and influences on the groundwater, however, the existence of the redundant features could affect results. On the other hand, various combinations of features would have different results. So, a successful feature selection method was considered to identify key features. In this study, recursive feature elimination (RFE) [94] method was used to find and select the key feature combinations among the GSCF that was used in this study.

RFE is a machine learning technique that builds a model according to the predictors’ subsets and computes them individually and then scores them through feature selection (FS). The model is reconstructed only based on the most important predictors and the less important predictors are eliminated. The number and the size of predictors’ subsets can be adjusted as a mean for the model performance optimization. RFE has been widely used in model construction in various scientific fields. It has also been recently used for groundwater contamination prediction where it is used for identifying the best variable combination for each machine learning model [95]. However, the application of RFE has not been explored for salinity modeling and risk mapping. As the RFE is a wrapper, the random forest (RF) algorithm is used to train it [96]. Furthermore, selection of the best features was established upon the cross-validation with 5-fold and metric of accuracy using Caret R package [97]–[99].

E. MODELING APPROACHES

In this study, three machine learning models of probability prediction including stochastic gradient boosting, rotation forest, and Bayesian generalized linear model were employed to model groundwater salinity susceptibility:

**Stochastic Gradient Boosting (StoGB):** StoGB has been developed by Friedman [100] to improve the performance of his gradient boosting introduced earlier, through including randomness to the function estimation procedures. The randomness has been included using the adaptive bagging [101] for the least-squares fitting of additive expansions which replaces the base learner in gradient boosting procedures with the bagging learning. Further, at each boosting step, it substitutes random residuals for the ordinary ones. At each iteration, a subsample of the training data is drawn randomly from the training data. This randomly selected subsample is then used, instead of the full sample, to fit the base learner and compute the model update for the next iteration [100]. The model has four parameters including a number of trees (n.trees), maximum nodes per tree or number of leaves (interaction.d), learning rate (Shrinkage), and a minimum number of samples in tree terminal nodes (n.minobsinn), which are tuned within a 5-fold cross-validation procedure. The StoGB has been previously used for groundwater modeling for the purpose of prediction of the hydro-chemical composition of groundwater with promising results [102]. Further application of StoGB has been limited in hydrological modeling [103]. In this study, the R package of StoGB, which is known as gbm was used to implement this model. A more detailed description of the model is presented in Friedman [100].

**Rotation Forest (RotFor):** Rodríguez et al. [104] advanced the feature extraction machine learning method of RotFor for generating classifier ensembles. To handle the training data used for classifier ensembles, the feature set is randomly split into (K) number of subsets which are prepared by principal component analysis (PCA) [105] to retain the data variability. Then, to encourage simultaneous individual accuracy of the ensemble, equal to the number of subsets, axis rotations take place to form the new features for a base classifier. RotFor uses decision trees that perform well in showing a high level of sensitivity to the rotation of the feature axes. The RotFor has two parameters including the number of variable subsets (K) and Ensemble size (L). RotFor has been very recently used as the base classifier in spatial modeling of groundwater with promising results [106], and its popularity in other hydrological susceptibility mappings [107]–[109]. In this study, the model parameters are tuned within the 5-fold cross-validation procedure. The rotationForest R package [104] was applied to execute the RotFor. More description of the model can be found in Rodríguez et al. [104].

**Bayesian Generalized Linear Model (Bayesglm):** Generalized Linear Models (GLMs) [110] represents a group of modeling techniques e.g. Bayesian, logistic regression, through a general modeling approach to several useful distributions, including the gamma, Poisson, and binomial, suitable for a broad range of response modeling problems [111]. While modeling with GLM, the response can be specified with a higher degree of flexibility in the modeling, resulting in a more convenient structure. However, it needs parametrizations that are difficult to interpret and solutions that are not comprehensive. The observation model, linear predictor, and link function are the three building block components of GLM [110]. In a Bayesian GLM, or so-called Bayesglm, a normal prior is often chosen for the linear predictor’s parameters. This conjugate prior to the normal linear model is a GLM with a Gaussian observation model and identity link
function. A normal prior resulting from expert opinion is therefore sought for the unknown coefficients in Bayesian GLMs. In this study, R’s arm package [112] was applied to perform the Bayesglm. See Hosack et al. [111] and Hosseini et al. [113] for more details of the Bayesglm model.

**F. MODEL CALIBRATION AND PERFORMANCE ESTIMATION**

The database including predictors and predictand is randomly divided into the training (80 %) and testing (20 %) datasets. The models were evaluated using testing datasets after calibration using the selected features by RFE. The researchers used 5-fold cross-validation to optimize the models. The process was repeated 50 times and two key metrics of classification including Accuracy (Acc) and Kappa (K) were used to assess the models’ performance by calculating a contingency table and hit and miss analysis. Accuracy indicates the percentage of correct classification. Kappa highlights the likelihood of agreement by chance using probabilities of the model classification [114]. This statistic is classified into five classes: poor \((K < 0.4)\), moderate \((0.4 < K < 0.55)\), good \((0.55 < K < 0.85)\), excellent \((0.85 < K < 0.99)\), and perfect \((0.99 < K < 1.00)\); indicating the performance of the classification models [115]. These metrics are calculated as follows:

\[
Acc = \frac{H + CN}{H + FA + M + CN} 
\]

where \(H\), \(FA\), \(M\), and \(CN\) are calculated using the contingency table and respectively correspond with the number of hits, false alarms, misses, and correct negatives [116].

\[
K = \frac{Acc - Pe}{1 - Pe} 
\]

where \(Pe\) is the expected probability of chance agreement [117], [118], which is calculated as follow:

\[
Pe = \frac{(H + FA)(H + M) + (M + CN)(FA + CN)}{(H + FA + M + CN)^2} 
\]

**III. RESULTS AND DISCUSSION**

**A. RESULTS OF FEATURE SELECTION**

Results of the recursive feature elimination (RFE) method for determining the best number of features are reprinted in Fig. 3. The accuracy metric was considered to evaluate the RFE performance. Each box indicates aggregated model performance for 5-fold cross validation during many times of model runs (about 700 runs). As can be seen, the RFE indicates that modeling with 8 features had better performance among 1 to 16 features (Accuracy is about 0.87) (Fig. 3). The average values of Accuracy and Kappa are represented in Table 1. The Accuracy and Kappa values for the number of 8 features are equal to 0.87 and 0.64, respectively, which are the highest in comparison with other numbers of features (Table 1).

Importance of the features was determined based on their occurrence in the model runs (Fig. 4). As can be seen from Fig. 4, the depth of groundwater and elevation contributes as the input features for more than 96 % of the model runs. According to the RFE results, the number of 8 features must be selected for groundwater salinity modeling (Fig. 3). So, variables of depth to groundwater (DTGW), elevation, distance from fault (DFF), lithology, decline of groundwater level (DGWL), distance from lake (DFL), groundwater withdrawal (GWW), and distance from stream (DFS) were selected as key features respectively with occurrence frequency equal to 97.5%, 96.3%, 81.3%, 78.8%, 71.3%, 61.3%, 60%, and 50 % in all model runs (Fig. 4).

**B. RESULTS OF MODEL EVALUATION**

The StoGB model was optimized with 50 runs, and results indicated that the best value for learning rate (Shrinkage)
was equal to 0.1, also, 200 for the number of trees (n.trees), 3 for maximum nodes per tree (interaction.d), and 10 for the minimum number of samples in the tree terminal nodes (n.minobsinn). Fig. 5 shows the variation of Accuracy vs the number of trees (n.trees) and maximum nodes per tree (interaction.d). Predictive performance of the StoGB was calculated using the training data set, and results indicated good performance according to the Accuracy (80.01) and Kappa (0.79) (Table 2).

According to the models’ evaluation, the RotFor and Bayesglm indicate the superior performance of the StoGB. Although StoGB had a good performance, the RotFor and Bayesglm had an excellent performance based on the Kappa values (K > 0.85) [115].

C. SPATIAL PREDICTION OF GROUNDWATER SALINITY

Three individual machine learning algorithms (i.e., StoGB, RotFor, and Bayesglm) were applied to predict groundwater salinity in the area of the region. The models produced the

### Figure 4. The frequency of features occurrence (%) in the RFE runs.

| Feature                          | Frequency |
|----------------------------------|-----------|
| Depth to groundwater             | 97.5      |
| Elevation                       | 96.3      |
| Distance from fault              | 81.3      |
| Lithology                        | 78.8      |
| Decline of groundwater level     | 71.5      |
| Distance from lake               | 61.3      |
| Groundwater withdrawal           | 60.0      |
| Distance from stream             | 50.0      |
| Evaporation                      | 41.3      |
| Slope                            | 41.2      |
| TWI                              | 38.8      |
| Precipitation                    | 37.8      |
| Landuse                          | 32.5      |
| Aspect                           | 28.8      |
| Curvature                        | 16.3      |
| Soil type                        | 16.3      |

### Table 3. The coefficients estimated by Bayesglm model for each variable.

| Variable                          | Coefficient |
|-----------------------------------|-------------|
| Intercept                         | 1.70        |
| Depth to groundwater              | -1.30       |
| Elevation                         | -1.03       |
| Distance from fault               | -0.67       |
| Lithology                         | -0.54       |
| Decline of groundwater level      | 0.72        |
| Distance from lake                | 0.16        |
| Groundwater withdrawal            | -0.08       |
| Distance from stream              | -0.19       |
probability of groundwater salinity maps with a cell size of 12.5 m. The probability maps vary between zero to 1, which respectively indicates the lowest and highest probability of groundwater salinity (Fig. 7). The probability maps were classified using equal interval classification method with an interval 0.2 into five classes: very low (0 – 0.2), low (0.2 – 0.4), medium (0.4 – 0.6), high (0.6 – 0.8), and very high (0.8 – 1).

As can be seen, the very low class is highest for the StoGB model (1210.8 km$^2$) and includes the most area of the region, whereas the medium susceptibility class predicted by this model is lowest (70.6 km$^2$) (Table 4). Most of the study area was placed in medium (mostly marginal parts with an area equal to 704.6 km$^2$) and very high (mostly central parts with an area equal to 628.7 km$^2$) susceptibility classes by the RotFor model, whereas the class of high susceptibility was lowest (56.9 km$^2$) (Table 4). The Bayesglm predicted that most of the study area has very low (702.2 km$^2$) and very high (596.4 km$^2$) susceptibility (Table 4), respectively located in the boundary and central of the study area (Fig. 7).

Spatial modeling results highlighted that two models of the RotFor and Bayesglm produced strong similarities in the development of the probability and susceptibility maps. It should be noted that these two models have performed.
TABLE 4. Area (km²) of susceptibility class for groundwater salinity map of predictive models.

| Hazard class | StoGB  | RotFor | Bayesglm |
|--------------|--------|--------|----------|
| Very low     | 1210.8 | 121.6  | 702.2    |
| Low          | 144.8  | 176.4  | 102.3    |
| Medium       | 70.8   | 704.6  | 114.6    |
| High         | 113.9  | 56.9   | 172.7    |
| Very high    | 147.9  | 628.7  | 596.4    |

better than the StoGB model (Table 2). Although prediction of the models was different, all of the models indicated that central parts of the region have a very high susceptibility in which this class matches through both probability and susceptibility maps predicted by models of the RotFor and Bayesglm (Fig. 7). The very high class of the susceptibility maps predicted by models, especially the RotFor and Bayesglm (Fig. 7), are matched with agricultural areas, lithology map, the locations with low depth to groundwater, low slope, and elevation (Fig. 2). Additionally, areas near to the Maharlu Lake and locations with a high decline in groundwater are located in the very high susceptibility zone.

However, it is obvious that a set of the abovementioned factors which may affect the salinity of the groundwater in the region are: (i) topographic effects, which can cause flushing and exporting the salty materials from the soil and accumulating it in lowlands; (ii) decline in groundwater and proximity to lake, which can cause intrusion of saltwater [86], [93]. In this regard, Jahanshahi and Zare [119] confirmed the contribution of the saltwater intrusion from Maharlu lake to deteriorate the Sarvestan plain groundwater; (iii) low depth to groundwater along with the agricultural area in the central parts, which can cause infiltration of irrigation water, leaching and dissolving solid material into groundwater [35], [92]. Also because of low depth to groundwater, Jahanshahi and Zare [119] demonstrated the effects of evaporation from shallow groundwater on the quality of groundwater in some parts of this plain; and (v) lithology effects, which includes salty formations of the Sachoun (lithology unit PeEsa includes the evaporate minerals of halite and gypsum), Razak (lithology unit Omr includes the marl interbedded with limestone), Aghajari (lithology unit MPlfgp which includes calcareous sandstone and red marl with interbedded gypsum), and Hormoz (lithology unit pC-Ch includes mainly the salts and other evaporate rocks) (Fig. 2o) [119]. Although mostly, these formations are located in the border of the very high susceptibility zones, however, drainage of the plain causes the salty materials to be flushed and deposited in lowlands (which have mostly Qfr2 lithology: low-level piedmont fan, and valley terrace deposits) and penetrate into groundwater. This is confirmed by Raeisi et al. [84] who, using optical and chemical methods, indicated that the dissolution of gypsum and other evaporate materials, from surrounding formations of the plain, with water expelled from those are the main reasons for sodium-chloride water type and the poor quality of groundwater in this area.

IV. CONCLUSION

Producing the susceptibility map of salinity for groundwater is of utmost importance for the water security of arid areas and sustainable agriculture. In this study, the best features are selected by the RFE method, and the three machine learning models of StoGB, RotFor, and Bayesglm showed promising results in the prediction of susceptibility maps. All of the models indicated that central parts of the region have a very high susceptibility which matches with agricultural areas, lithology map, the locations with low depth to groundwater, low slope, and elevation.

In this study, due to irregular monitoring and lack of a continuous sampling for each month and season, modeling was conducted using mean yearly EC of wells during a period of 15 years (2003-2017). However, variations of EC during seasons and overtime may be effective on a susceptibility map, so it is recommended considering impacts of climate variability and land-use change on susceptibility map by dividing a longer time period into sub-periods for future studies in other regions. Although the sodium adsorption ratio (SAR), as well as the EC, affects groundwater salinity, investigation of the observed groundwater data in this study indicated low values of the SAR. However, integration of salinity susceptibility (from EC) and sodicity (alkalinity) susceptibility (from SAR) maps during the next studies can be worthwhile to indicate the status of the groundwater quality.

The susceptibility maps produced in this study can increase the awareness of farmers from susceptible areas to control their practices and to adapt cropping systems in order to deal with the groundwater salinity, and it also helps decision-makers to change both land use and land management.

REFERENCES

[1] W. Xu and X. Su, “Challenges and impacts of climate change and human activities on groundwater-dependent ecosystems in arid areas—A case study of the Nalenggele Alluvial fan in NW China,” J. Hydrol., vol. 573, pp. 376–385, Jun. 2019, doi: 10.1016/j.jhydrol.2019.03.082.
[2] J. Xue, Z. Huo, Q. Huang, F. Wang, J. Boll, G. Huang, and Z. Qu, “Assessing sustainability of agricultural water saving in an arid area with shallow groundwater,” Irrigation Drainage, vol. 68, no. 2, pp. 205–217, Apr. 2019, doi: 10.1002/id.2295.
[3] N. Barreiras and L. Ribeiro, “Estimating groundwater recharge uncertainty for a carbonate aquifer in a semi-arid region using the Kessler’s method,” J. Arid Environ., vol. 165, pp. 64–72, Jun. 2019, doi: 10.1016/j.jaridenv.2019.04.005.
[4] R. P. Sishodia, S. Shukla, W. D. Graham, S. P. Wani, J. W. Jones, and J. Heaney, “Current and future groundwater withdrawals: Effects, management and energy policy options for a semi-arid Indian Watershed,” Adv. Water Resour., vol. 110, pp. 459–475, Dec. 2017, doi: 10.1016/j.advwatres.2017.05.014.
[5] A. E. Akay Demir, F. B. Dilek, and U. Yetis, “A new screening index for pesticides leachability to groundwater,” J. Environ. Manage., vol. 231, pp. 1193–1202, Feb. 2019, doi: 10.1016/j.jenvman.2018.11.007.
[6] C. Q. Dai, “Analysis and prevention of the impacts of metro construction on groundwater environment,” Appl. Ecology Environ. Res., vol. 17, no. 2, pp. 4839–4854, 2019, doi: 10.15666/aer/1702_48394854.
[113] F. S. Hosseini, B. Choubin, A. Mosavi, N. Nabipour, S. Shamsirband, H. Darabi, and A. T. Haghighi, “Flash-flood hazard assessment using ensembles and Bayesian-based machine learning models: Application of the simulated annealing feature selection method,” Sci. Total Environ., vol. 711, Apr. 2020, Art. no. 135161, doi: 10.1016/j.scitotenv.2019.135161.

[114] H. Darabi, B. Choubin, O. Rahmati, A. Torabi Haghighi, B. Pradhan, and B. Kløve, “Urban flood risk mapping using the GARP and QUEST models: A comparative study of machine learning techniques,” J. Hydrol., vol. 569, pp. 142–154, Feb. 2019, doi: 10.1016/j.jhydrol.2018.12.002.

[115] R. A. Monserud and R. Leemans, “Comparing global vegetation maps with the Kappa statistic,” Ecol. Modell., vol. 62, no. 4, pp. 275–293, 1992, doi: 10.1016/0304-3800(92)90003-W.

[116] L. E. Johnson and B. G. Olsen, “Assessment of quantitative precipitation forecasts,” Weather Forecast., vol. 13, no. 1, pp. 75–83, 1998, doi: 10.1175/1520-0434(1998)013<0075:AOQPF>2.0.CO;2.

[117] A. Beucher, A. B. Møller, and M. H. Greve, “Artificial neural networks and decision tree classification for predicting soil drainage classes in Denmark,” Geoderma, vol. 352, pp. 351–359, Oct. 2019, doi: 10.1016/j.geoderma.2017.11.004.

[118] A. Mosavi, F. Sajedi-Hosseini, B. Choubin, F. Taromideh, G. Rahi, and A. Dineva, “Susceptibility mapping of soil water erosion using machine learning models,” Water, vol. 12, no. 7, p. 1995, Jul. 2020, doi: 10.3390/w12071995.

[119] R. Jahanshahi and M. Zare, “Hydrochemical investigations for delineating salt-water intrusion into the coastal aquifer of Maharlou Lake, Iran,” J. Afr. Earth Sci., vol. 121, pp. 16–29, Sep. 2016, doi: 10.1016/j.jafrearsci.2016.05.014.

FARZANEH SAJEDI HOSSEINI received the B.S. degree in rangeland and watershed management from Shiraz University, Shiraz, Iran, in 2011, and the M.S. degree in watershed management and engineering from Sari Agricultural Sciences and Natural Resources University, Sari, Iran, in 2015. She is currently pursuing the Ph.D. degree in watershed management with the University of Tehran, Tehran, Iran. Her research interests include hydrology and spatial and natural hazard modeling.

BAHRAM CHOUBIN received the B.S. degree in rangeland and watershed management from Shiraz University, Shiraz, Iran, in 2011, the M.S. degree in watershed management and engineering from the University of Tehran, Tehran, Iran, in 2013, and the Ph.D. degree in watershed management from Sari Agricultural Sciences and Natural Resources University, Sari, Iran, in 2017. From 2018 to 2019, he was a Postdoctoral Researcher at the University of Tehran. He has been an Assistant Professor at the Soil Conservation and Watershed Management Research Department, AREEO, Urmia, Iran, since 2019. He has published more than 30 articles in JCR-indexed journals. His research interests include hydrometeorology, natural hazard, prediction in ungauged basin (PUB), remote sensing, and hazard modeling. He was a recipient of the Academic Award from the National Elites Foundation, Iran, in 2016. He is a scientific reviewer for more than 50 ISI journals.

MASSOUD GOODARZI received the Ph.D. degree in physical geography (climatology and planning) from the University of Tabriz, in 2011. He is currently an Assistant Professor at the Soil Conservation and Watershed Management Research Institute, AREEO, Iran. His research interests include climate change, drought, dust, flood, and other environmental issues.

ADRIENN A. DINEVA received the B.Sc. and M.Sc. degrees in mechatronics engineering from Obuda University, in 2013, and the Ph.D. degrees in applied informatics from the Universita degli Studi di Milano and Obuda University, in 2017. She is currently an Assistant Professor with Obuda University. Her research interests include the area of adaptive control of nonlinear systems and applied engineering, including machine learning and soft computing techniques, and also the area of adaptive signal processing. She has authored more than 20 ISI articles in these fields.

***