Application of rotation forest with decision trees as base classifier and a novel ensemble model in spatial modeling of groundwater potential

Seyed Amir Naghibi · Mojtaba Dolatkordestani · Ashkan Rezaei ·
Payam Amouzegari · Mostafa Taheri Heravi · Bahareh Kalantar · Biswajeet Pradhan

Abstract Groundwater resources are facing a high pressure due to drought and overexploitation. The main aim of this research is to apply rotation forest (RTF) with decision trees as base classifiers and an improved ensemble methodology based on evidential belief function and tree-based models (EBFTM) for preparing groundwater potential maps (GPM). The performance of these new models is then compared with three previously implemented models, i.e., boosted regression tree (BRT), classification and regression tree (CART), and random forest (RF). For this purpose, spring locations in the Meshgin Shahr in Iran were detected. The spring locations were randomly categorized into training (70% of the locations) and validation (30% of the locations) datasets. Furthermore, several groundwater conditioning factors (GCFs) such as hydrogeological, topographical, and land use factors were mapped and regarded as input variables. The tree-based algorithms (i.e., BRT, CART, RF, and RTF) were applied by implementing the input variables and training dataset. The groundwater potential values (i.e., spring occurrence probability) obtained by the BRT, CART, RF, and RTF models for all the pixels of the study area were classified into four potential classes and then used as inputs of the EBF model to construct the new ensemble model (i.e., EBFTM). At last, this paper implemented a receiver operating characteristics (ROC) curve for determining the efficiency of the EBFTM, RTF, BRT, CART, and RF methods. The findings illustrated that the EBFTM had the highest efficacy with an area under the ROC curve (AUC) of 90.4%, followed by the RF, BRT, CART, and RTF models with AUC-ROC values of 90.1, 89.8, 86.9, and 86.2%, respectively. Thus, it could be inferred that the ensemble approach is capable of improving the efficacy of the single tree-based models in GPM production.

S. A. Naghibi (✉) · P. Amouzegari
Department of Watershed Management Engineering, Faculty of Natural Resources, Tarbiat Modares University (TMU), Noor, Mazandaran, Iran
e-mail: Amirnaghibi2010@yahoo.com

M. Dolatkordestani
Jiroft University Scholarship, Department of Combat Desertification, College of Natural Resources, Jiroft University, Jiroft, Iran

A. Rezaei
Department of Range and Watershed Management, Faculty of Agriculture and Natural Resources Sciences, University of Hormozgan, Bandar Abbas, Iran

M. T. Heravi
Department of Civil Engineering, Eghbal Lahoori Institute of Higher Education, Khorasan Razavi, Mashhad, Iran

B. Kalantar
RIKEN Center for Advanced Intelligence Project, Goal-Oriented Technology Research Group, Disaster Resilience Science Team, Tokyo 103-0027, Japan

B. Pradhan
The Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), Faculty of Engineering and IT, University of Technology Sydney, Sydney, NSW 2007, Australia

B. Pradhan
Department of Energy and Mineral Resources Engineering, Choongmu-gwan, Sejong University, 209 Neungdong-ro, Gwangjin-gu, Seoul 05006, Korea

Introduction

Iran is located in arid and semi-arid areas and is exposed
to water scarcity (Chezgi et al. 2016). Water scarcity is
considered as the lack of adequate obtainable water
resources, which are required to supply water require-
ments in an area. Water scarcity could be a result of
excessive population growth, climate change, and inap-
propriate water resource management (Asia and
Richman 1991). People's influence on water resources
could be changed by using better water resource man-
agement plan leading to mitigation of the impact of water
scarcity on their lives. One important aspect of water
resource management is groundwater (GW) resource
management. In general, different hydrogeological
mechanisms impact the location and discharge of the
water springs in an area. Modeling GW potential by
different technical methods could be useful to gain a
deeper apprehension of GW resources leading to better
GW management strategies.

Recently, several types of approaches and techniques
have been employed in different countries for generating
GPMs. For example, frequency ratio is calculated in some
works to generate GPMs (Manap et al. 2014; Naghibi
et al. 2015; Mousavi et al. 2017). Weights-of-evidence is
another method, which has been employed for this pur-
pose by Ozdemir (2011a). Similarly, Ozdemir)2011b) and
Pourtaghi and Pourghasemi (2014) implemented logistic
regression for producing GPMs in two basins in Turkey
and Iran, respectively. Evidential belief function (EBF) is
another bivariate approach, which has been implemented
for producing GPMs (Naghibi and Pourghasemi 2015;
Rahmati and Melesse 2016; Tahmassebipoor et al. 2016;
Ghorbani Nejad et al. 2016; Kordestani et al. 2018). More
recently, a number of machine learning models including
tree-based algorithms, i.e., boosted regression trees
(BRT), classification and regression tree (CART), and
random forest (RF); discriminant analysis models; and
neural network and support vector machines have been
employed for obtaining GPMs in several areas regarding
different environmental features (Naghibi and
Pourghasemi 2015; Naghibi and Moradi Dashtpagerdi
2016; Hong et al. 2016a, b; Rahmati et al. 2016; Zabihi
et al. 2016; Naghibi et al. 2017b, 2018a, b; Hong et al.
2018; Golkarian et al. 2018; Rahmati et al. 2018). The

mentioned machine learning models have proved to pro-
duce reliable outputs. Rotation forest is another tree-based
algorithm, which has been implemented in other applica-
tions such as a breast cancer recognition study (Aličković
and Subasi 2017), an image classification study (Xia et al.
2017), a gene expression data classification study (Lu
et al. 2017), and landslide studies (Pham et al. 2017,
2018a; Hong et al. 2018). All those researchers have
approved the acceptable efficacy of the RTF in classifi-
cation issues. In addition, some other researchers have
implemented data mining models in landslide, flood, and
groundwater studies (Chen et al. 2017a, b; Pham et al.
2018a, b, c, d, e; Pham and Prakash 2018; Pham 2018;
Khosravi et al. 2018a, b; Chen et al. 2018a, b, c; Hong et
al. 2019).

A complete investigation of the published articles in
spatial modeling fields such as landslide susceptibility
modeling, flood susceptibility mapping, and ground
subsidence hazard mapping reveals that ensemble
methods are frequently being used, and their outputs
are more reliable than single methods. In a paper, Park
et al. (2014) employed frequency ratio, logistic regres-
sion, and artificial neural network in ground subsidence
hazard mapping. In the next stage, outputs of the men-
tioned models were employed as input of the fuzzy logic
model to create a novel ensemble model. The findings of
their research showed that the ensemble model produced
more accurate results. Lee et al. (2012a, b), Umar et al.
(2014), Althuwaynee et al. (2014), Youssef et al. (2015),
and Aghdam et al. (2016) introduced new ensemble
models in the field of landslide susceptibility mapping
and reported better performance of the ensemble models.
On the other hand, Tehrany et al. (2013, 2014) employed
ensemble models in the field of flood susceptibility
assessment. Tehrany et al. (2013) proposed an ensemble
algorithm by mixing bivariate and multivariate models,
whereas in Tehrany et al. (2014), the outputs of weight of
evidence and support vector machines were mixed
together to build a new ensemble model. Both of those
papers addressed better efficacy of the ensemble
algorithms vs. the single ones. In the field of GW
potential assessment, Naghibi et al. (2017a) used several
data mining methods such as AdaBoost, Bagging, gen-
eralized additive model, and Naive Bayes for determin-
ing GW potential. They combined the outputs of those
models by frequency ratio model and subsequently built
a new ensemble model. The result of this new ensemble
method was reported acceptable. Ensemble models are
reported to give better outputs than single methods since

ensemble models overcome the weakness of single models (Wan et al. 2012; Naghibi et al. 2017a).

By investigating the literature, it can be obviously seen that a shift from statistical models to more sophisticated ensemble models has been occurred in the recent years. However, a limited number of studies have been done on ensemble models in groundwater modeling, which can take the advantages of both statistical and data mining algorithms. With this in mind, the main novelty of this study is the application of a new ensemble model (EBFTM) through a combination of the BRT, CART, RF, and RTF algorithms by the EBF method in GW potential mapping. Another novelty of this research is the application of the RTF in producing GPMs. In general, the objectives of this study are (i) application of two new models i.e., RTF and ensemble model (EBFTM) in producing GPMs, and (ii) defining the importance of the GCFs in modeling GW potential.

Study area

The Meshgin Shahr area is stretched between 38° 38′ 26″ and 38° 13′ 14″N latitudes, and 47° 21′ 11″ and 48° 11′ 25″E longitudes in Iran (Fig. 1). The study region covers an area of about 2032 km^2. The altitude changes between 800 and 4117 m a.s.l. with a mean value of about 1556 m. The average annual precipitation was estimated as 361 mm. Most of the precipitation occurs in winter and spring seasons. The annual average humidity in Meshgin Shahr is measured as 63%. The area comprises six land use classes of agriculture, forest, orchard, range-land, residential area, and water body. Most of the area is covered by rangeland. Geologically, the area comprises 16 different lithological units ranging from andesite to basaltic volcanic to granite to diorite (Table 1).

Material and methods

Figure 2 represents flowchart of the methodology employed in the current research.

Spring inventory

The location of the springs in the Meshgin Shahr area was prepared by importing them into GIS environment from a report by Khorasan Razavi Regional Water Authority (2015). The spring inventory map contains 279 springs. The average discharge of the springs is measured as 2.2 L/

s. Additionally, the average pH of the springs is measured as 6.64. In modeling a phenomenon, it is essential to separate training and validation datasets. The ratio of these datasets is very important in such studies. This study followed Oh et al. (2011), Naghibi et al. (2017a), and Pradhan et al. (2014) in selecting the ratio. According to the mentioned papers, springs were segregated into training (70% of locations) and validation (30% of locations) datasets. It needs to be mentioned that most of the springs are located very close to each other in the west and southwestern parts of the Meshgin Shahr area. Those datasets were implemented in training the models and obtaining the tuned parameters and validating the produced GPMs (Fig. 1).

### Construction of the conditioning factors

In this study, considering the hydrogeological and topographical conditions of the basin and literature (Ozdemir 2011a, b; Naghibi et al. 2016), several GCFs including hydrogeological, topographical, and land use factors were mapped as input variables and employed for the modeling process.

Topographical factors

The topographic parameters, which include slope, aspect, altitude, plan curvature, profile curvature, and slope length (LS), were generated implementing a digital elevation model (DEM) with a spatial resolution of 30×30 m extracted from the topographic maps at a 1:50,000 scale (ASTER-DEM[1]). For those functions, ArcGIS and system for automated geoscientific analysis (SAGA) were used.

The infiltration status is majorly controlled by slope angle of the area, and this factor has a high importance in GW potential mapping (Adiat et al. 2012). In rough slope sections, the overland flow is fast and there is not adequate time for the flow to infiltrate (i.e., an inverse relationship exists among infiltration rate and slope degree) (Prasad et al. 2008). In the studied region, slope angle varies from 0 to 68.79° with an average value of 9.69° (Fig. 3a). Slope aspect impacts the sunlight duration in a given area (Kordestani, et al. 2018), and consequently influences the snowmelt and infiltration rates. This factor was classified into nine categories including

---

[1] Advanced Spaceborne Thermal Emission and Reflection Radiometer Digital Elevation Model
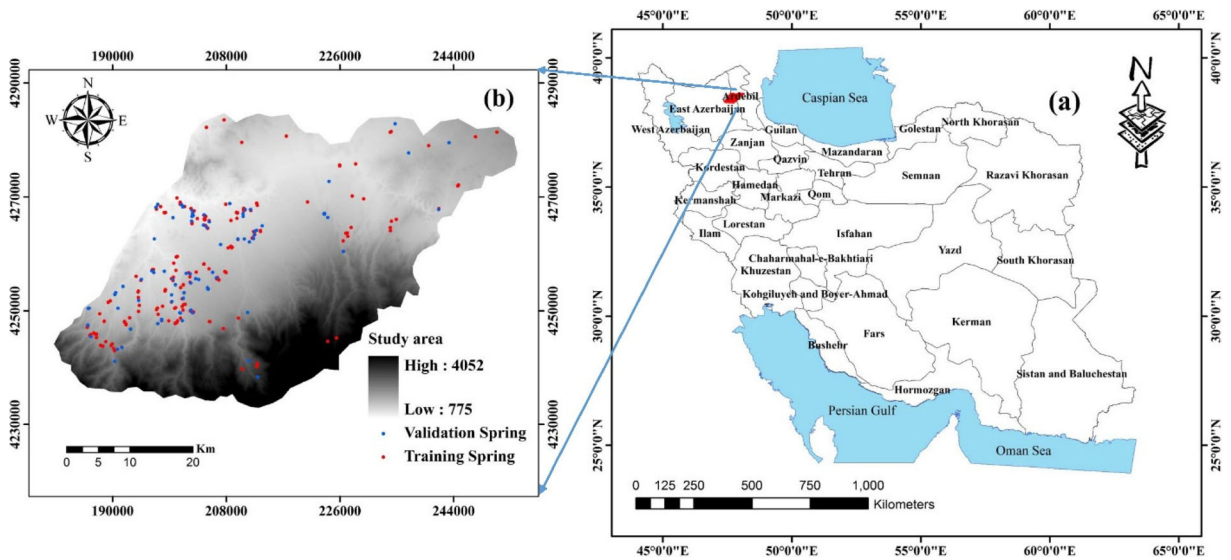
Fig. 1 Location of the study area in Iran (**a**) and location of the training and validation springs in the study area (**b**)

main and sub-main categories (Fig. 3b). Altitude determines the level of drainage system development in an area in a way that normally lower areas have more developed drainage systems. This factor was produced by employing the DEM of the study region and was

Table 1 The lithological characteristics of the study area

Class	Lithological characteristics
Group 1	Andesite to basaltic volcanic
Group 2	Andesitic subvolcanic
Group 3	Andesitic tuff
Group 4	Andesitic volcanic tuff
Group 5	Andesitic volcanic
Group 6	Basaltic volcanic rocks
Group 7	Coarse-grained fanglomerate composed of volcanic lastic materials locally with intercalation of lava flows
Group 8	Dacitic to andesitic tuff
Group 9	Granite to diorite
Group 10	Gypsiferous marl
Group 11	High-level piedmont fan and valley terrace deposits
Group 12	Low-level piedmont fan and valley terrace deposits
Group 13	Pliocene andesitic subvolcanic
Group 14	Polymictic conglomerate and sandstone
Group 15	Rhyolitic to rhyodacitic tuff
Group 16	Rhyolitic to rhyodacitic volcanic tuff

regarded as an important reported factor in GW potential (Naghibi et al. 2015) in this work (Fig. 3c).

Plan and profile curvatures were calculated from a DEM using SAGA software. These two curvatures influence the speed of water flow over the surface and subsequently impact the infiltration rate. Negative plan curvature depicts concave, positive curvature depicts convex, and zero curvature shows flat (Fig. 3d). Profile curvature is parallel to the slope and clarifies the direction of maximum slope (Fig. 3e).

LS as a sign of soil loss potential is a mixture of two factors of slope steepness and slope length (Fig. 3f). A higher value of LS shows higher accumulation of water and erosion, and influences infiltration rate. This factor can be presented mathematically as below (Moore and Burch 1986):

$$LS = \left(\frac{Bs}{22.13}\right)^{0.6} \left(\frac{\sin \alpha}{0.0896}\right)^{1.3} \qquad (1)$$

where Bs shows the specific catchment area (m^2).

Hydrogeological factors

TWI has been implemented as a topographical indicator of saturated sources location in a given area (Beven 1997). TWI shows the water tendency to accumulate at any point in the basin and can be represented mathematically as below (Moore et al. 1991; Fig. 3g):
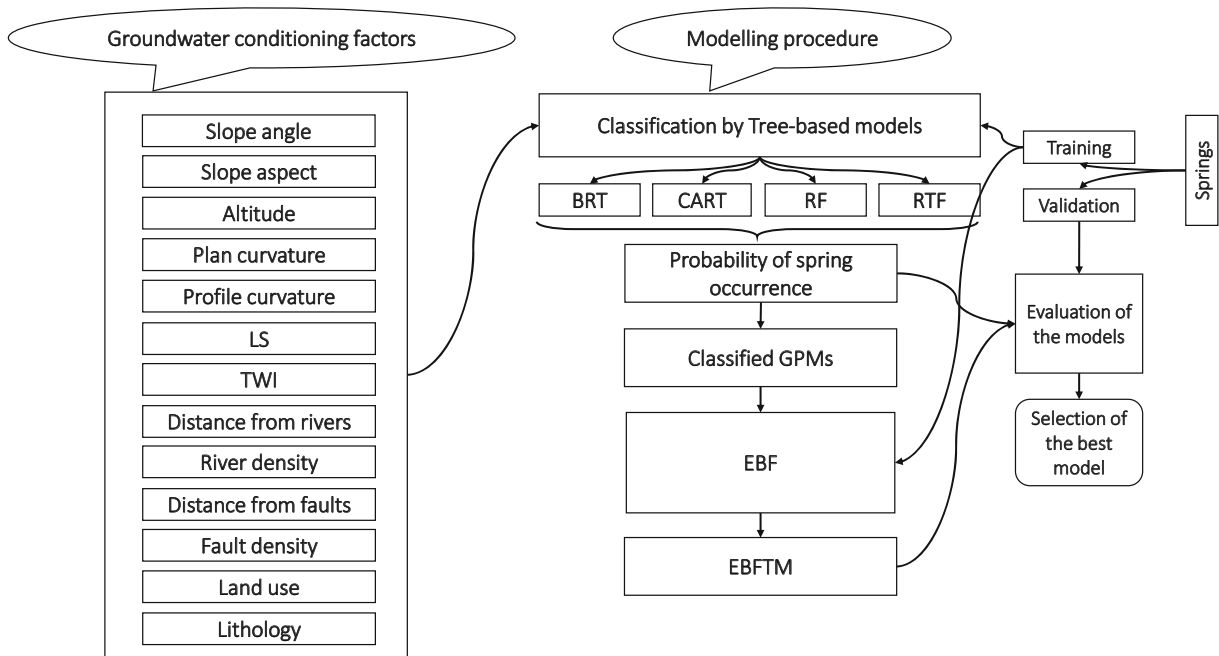
Fig. 2 Flowchart of the methodology implemented in this study

$$TWI = \ln(\alpha/\tan\beta) \qquad (2)$$

where α is the accumulative upslope area draining to a point and tan ß is the slope angle at the point.

Streams, which reflect slope, infiltration rate, and flow direction, were derived from a DEM using ArcGIS. It is known that streams are one of the main sources of groundwater recharge in semi-arid areas such as Iran. Therefore, distance from rivers could show the impact of this recharge source. River density shows the development of the drainage system and impacts groundwater recharge. Distance from rivers was computed by employing the river layer and Euclidean distance module implemented in ArcGIS 10.5 (Fig. 3h). In addition, river density was calculated using ArcGIS 10.5 and line density module (Fig. 3i). In this study, distance and density layers of the faults were built and regarded as GCFs (Fig. 3j, k). Lithology impacts porosity and infiltration of water into the ground (Rahmati et al. 2014). The lithology layer was produced by implementing a 1:100,000 scale geological map (GSI 1997). It was then categorized into 16 groups (Fig. 3l and Table 1).

Land use factor

Landsat 7/ETM images in 2015 were downloaded from USGS (http://earthexplorer.usgs.gov/), and the land use layer of Meshgin Shahr region was generated by implementing a supervised maximum likelihood algorithm in Idris software (Fig. 3m). This map contains six land use classes of agriculture, forest, orchard, rangeland, residential area, and water body.

Methodology

This section presents the methodological steps to obtain the final GPM by the EBFTM. In the first stage, tree-based models, i.e., BRT, CART, RF, and RTF, are explained. It should be noted that these models were applied using the training dataset and produced four GPMs. Then, these maps were classified into four classes of GW potential according to a previous research by Naghibi et al. (2017a). The groundwater potential values (i.e., spring occurrence probability) obtained by the BRT, CART, RF, and RTF models for all the pixels of the study area were classified into four potential classes. These values were used as inputs of the EBF model to construct the new ensemble model (i.e., EBFTM). In other words, probability classes obtained by the BRT, CART, RF, and RTF were extracted for the training dataset and used in the EBFTM to calculate the belief values for each class.
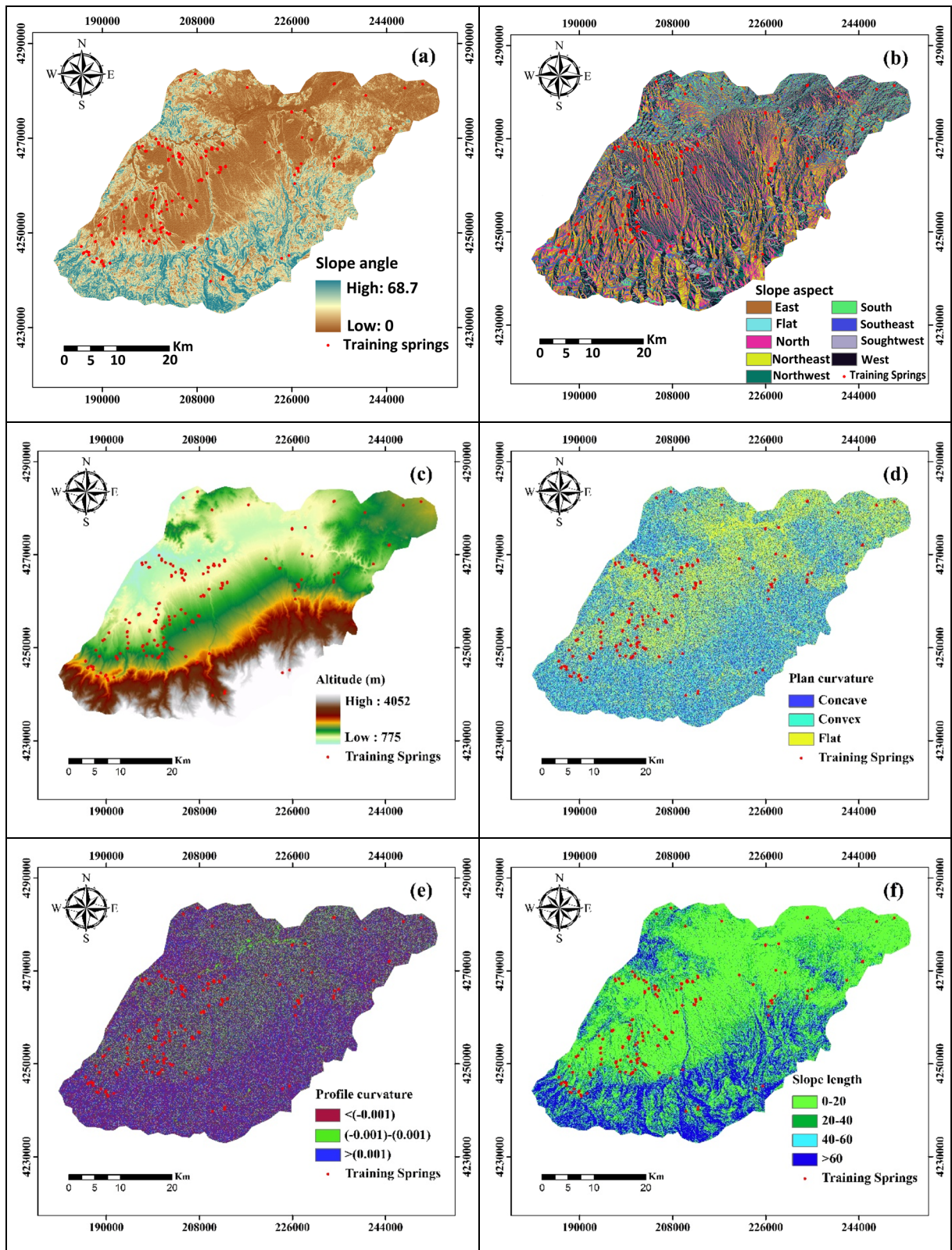
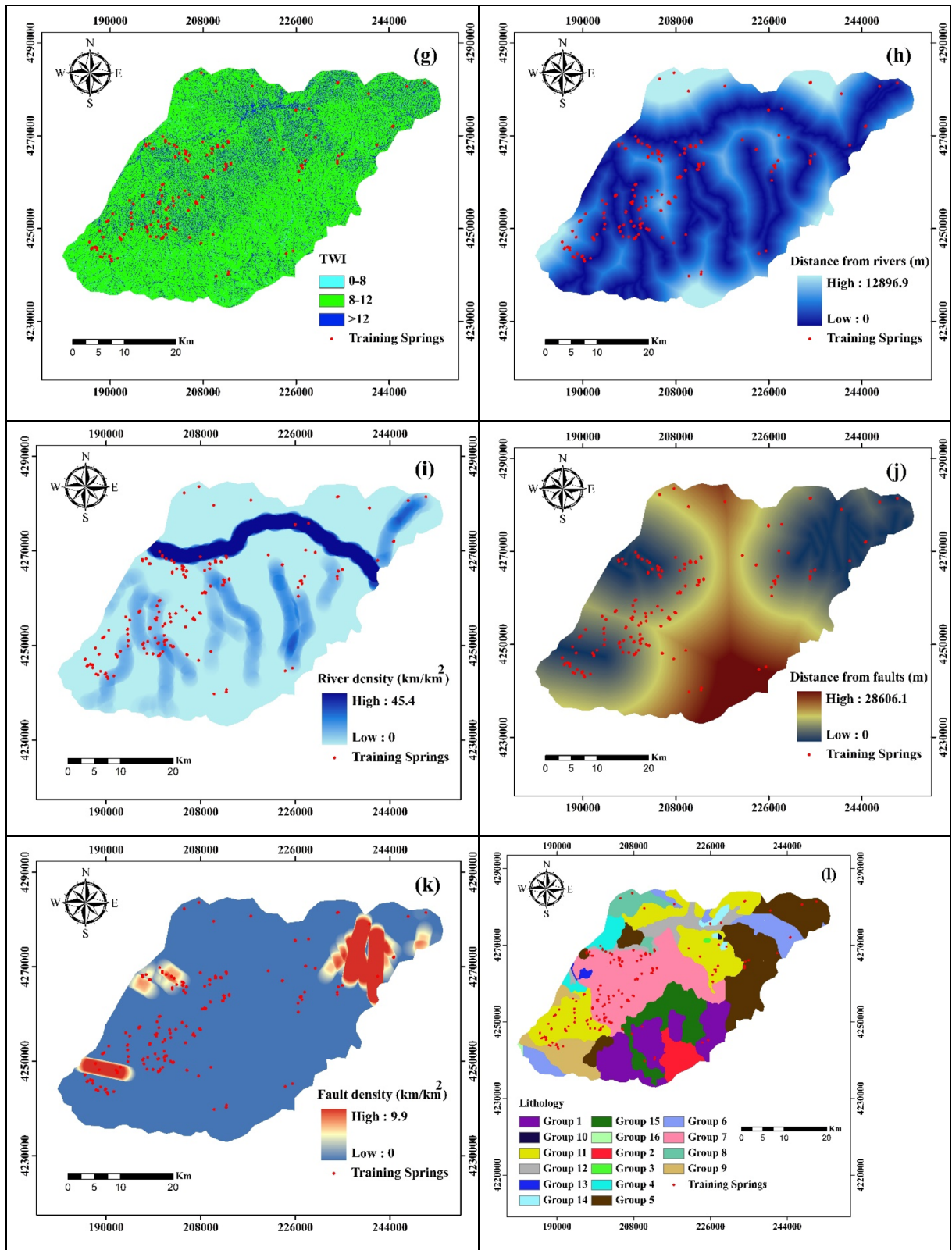**Fig. 3** Groundwater conditioning factors implemented in the modeling process
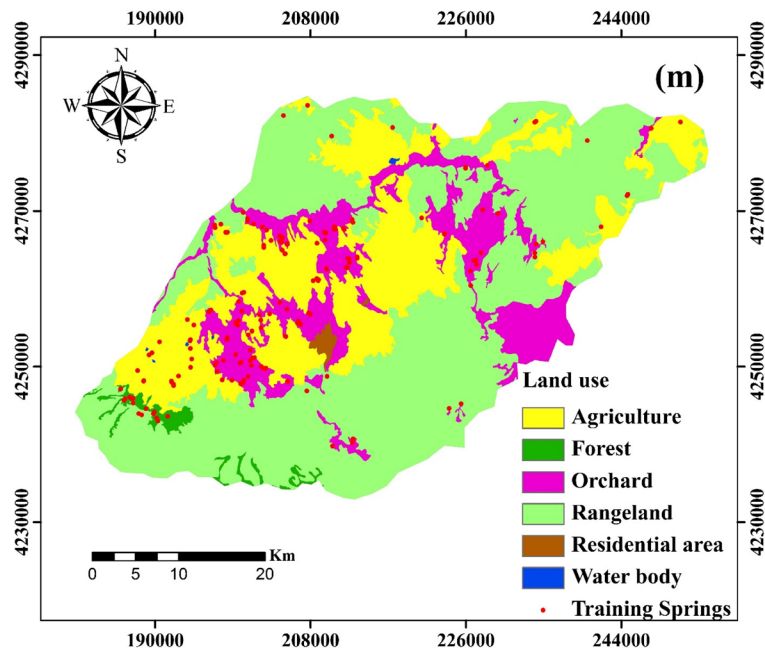
Fig. 3 (continued)

Fig. 3 (continued)

Application of the tree-based models in GW potential modeling

GW potential modeling by boosted regression tree BRT is fundamentally different from classical regression algorithms, which generates a single best choice. It uses boosting feature as a combiner of a large number of comparatively simple trees adaptively for optimizing the predictive efficacy (Elith et al. 2008). Adapting several trees in this algorithm conquers the major issue of single tree methods, which is their comparatively weak predictive efficacy (Aertsen et al. 2010). Elith et al. (2008) clarified that the final model could be considered as an additive regression algorithm with simple trees, adapted in a forward, stepwise form. In the BRT, three factors need to be tuned: number of trees; interaction depth,

Table 2 Importance of groundwater conditioning factors in GPMs using BRT model

GCFs	Relative Influence (%) in BRT	GCFs	Importance in CART	GCFs	Mean decrease Gini in RF
Land use	39.6	Land use	73.73	Distance from faults	22.23
Altitude	15.77	Distance from faults	63.13	Land use	22.12
Distance from faults	11.09	Lithology	60.25	Altitude	19.54
TWI	7.72	Altitude	55.6	Lithology	14.12
LS	5.35	TWI	28.97	TWI	14.06
Distance from rivers	4.73	Aspect	17.44	Slope angle	11.13
Slope angle	4.49	Distance from rivers	13.55	Distance from rivers	10.89
Lithology	4.15	Profile curvature	9.92	Plan curvature	10.13
Plan curvature	2.74	Slope angle	7.21	Profile curvature	9.98
Profile curvature	2.68	Fault density	6.57	LS	9.51
Aspect	1.07	Plan curvature	6.44	River density	5.38
Fault density	0.55	LS	2.79	Aspect	4.61
River density	0	River density	0	Fault density	2.64

which defines the size of trees; and shrinkage as a representative of the tree's contribution in the model (Elith and Leathwick 2013). It needs to be stated that parameter optimization in the BRT was applied by grid search. The implemented grid included interaction depth of 1 to 9 and number of trees of 0 to 1500 with 100 intervals. This algorithm was applied in R statistical software by employing the gradient boosting method (gbm) and caret packages (Naghibi et al. 2016). It needs to be mentioned that a ten-fold cross-validation was employed to define the parameters of this algorithm (Naghibi et al. 2016).

GW potential modeling by classification and regression tree CART is a famous tree-based model as it shows information in such a way that is simple to figure out. CART model is a binary recursive separation process, which is able to process continuous and nominal features as outputs and inputs. Each node (or split) is determined by an easy rule based on a single variable, classifying the data into two classes. In each one of the classes, the response factor has to maximize the homogeneity, intending to minimization of the total deviance (Vorpahl et al. 2012). The estimated amount of a "terminal node" clarifies the mean of the response amount in that node (Breiman et al. 1984). McKenney and Pedlar (2003) stated that classification tree should be pruned for selecting only the most contributing factors, which refers to the nodes explaining the largest amount of deviance. It needs to be clarified that CART was applied in R statistical software using the rpart script (Ripley 2015).

GW potential modeling by random forest Random forest (RF) is a nonparametric technique derived from CART. Each tree is generated on a bootstrapped subset of the dataset implementing CART methodology, with an accidental subspace of factors, which have been chosen for every node (Micheletti et al. 2013). The number of factors is larger than the size of the factors in the input space. The varied trees are decorrelated because of the accidental picking of the characteristics at each node, which enhances the strength of the ensemble model by decreasing the variance with no enhancing effect on the bias (Breiman 2001). The RF tunes two parameters of the number of trees called ntrees, and the number of factors, which is stochastically selected from the accessible

set of characteristics called mtry (Micheletti et al. 2013). One of the capabilities of RF is its effectiveness in defining the contribution of input factors by two indices of mean decrease accuracy and Gini (Pardo and Sberveglieri 2008). It needs to be explained that the RF was carried out in R statistical software employing randomForest and its inner tuning function.

GW potential modeling by rotation forest with decision trees as base classifier Rotation forest draws upon RF idea. RTF ensemble could be implemented to strengthen the performance of weak classifiers such as data mining models (Chen et al. 2017c; Pham et al. 2018a). In RTF, the training set is randomly classified into K sets and a rotation sparse matrix is built by extracting features from each set. RTF implements principal component analysis for extracting features for building sub-training data space to learn base classifiers (Koyuncu and Ceylan 2013). In RTF, every tree is trained using all data with a rotated feature space; however, the base classifiers are separately constructed decision trees (Rodriguez and Kuncheva 2007). Since the algorithm constructs the classification districts implementing hyperplanes parallel to the feature axes, a small rotation of the axes could result in a totally distinct tree. RTF directs to create correct and different classifiers (Rodriguez and Kuncheva 2007). Bootstrap samples are regarded as the training data for separate classifiers, like in bagging. The major heuristic is to conduct feature exploitation and to successively rebuild a complete feature set for each classifier in the ensemble. RTF was applied in the R statistical software employing rotationForest and caret scripts.

EBFTM ensemble approach

In this section, first, EBF model is described and then procedure of the ensemble modeling in this study is explained.

Description of evidential belief function model EBF is constructed on the basis of the Dempster-Shafer theory (Dempster 1968; Shafer 1976). This model comprises degrees of belief, disbelief, uncertainty, and plausibility as Bel, Dis, Unc, and Pls, respectively

(Nampak et al. 2014). EBF regards two different relationships in data including the relationship among a GCF and incidence of the springs as well as the relationship between the classes of each GCF (Pradhan et al. 2014). Degree of belief changes from 0 to 1 (Carranza et al. 2008; Pradhan et al. 2014). In this method, the generalized Bayesian lower and upper probabilities show the belief and plausibility values, respectively. The Unc can be computed from subtraction of the Bel from Pls, which would always be a positive value. The Unc could be determined as the doubt, which the incident supports a preposition (Pradhan et al. 2014). Finally, the Dis could be computed as 1-pls or 1-Unc-Bel (Tien Bui et al. 2012; Pradhan et al. 2014).

GW potential modeling by the new ensemble model EBFTM To construct the ensemble model (i.e., EBFTM) in this work, first, the tree-based models were carried out and their corresponding GPMs were built. In the next stage, the GPMs were grouped into four categories of potentiality (i.e., low, moderate, high, and very high). Afterward, the EBF degrees were computed for each class of the GPMs produced by the tree-based models. Then, Bel values were assigned to the classes of the GPMs, and the EBFTM values were obtained as follows:

$$EBFTM = Bel_{BRT} + Bel_{CART} + Bel_{RF} + Bel_{RTF} \qquad (3)$$

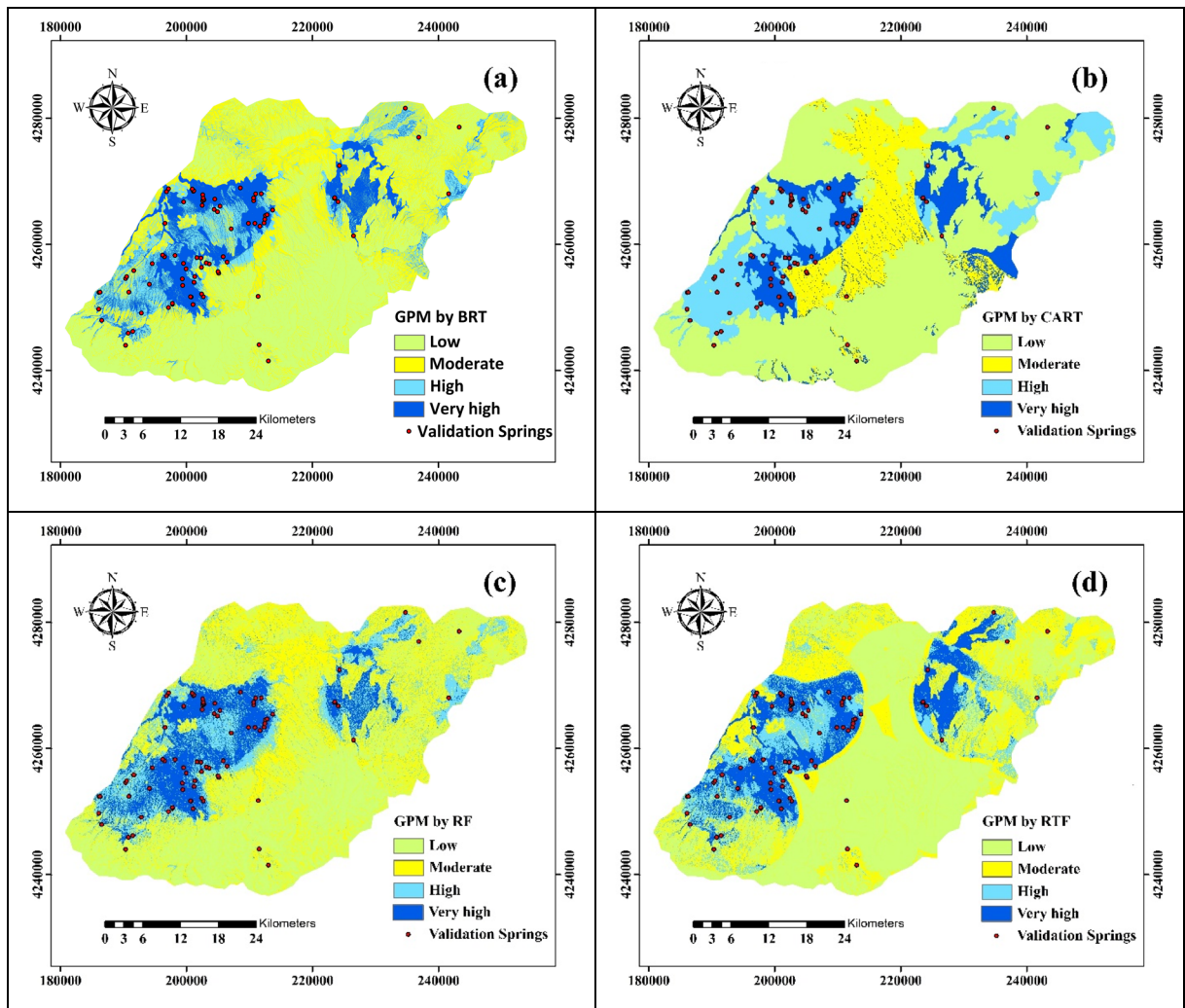where the EBFTM shows the final value for each pixel in the ensemble model; Bel_{BRT} shows Bel values calculated

Fig. 4 Groundwater potential maps produced by **a** BRT, **b** CART, **c** RF, **d** RTF, and **e** EBFTM
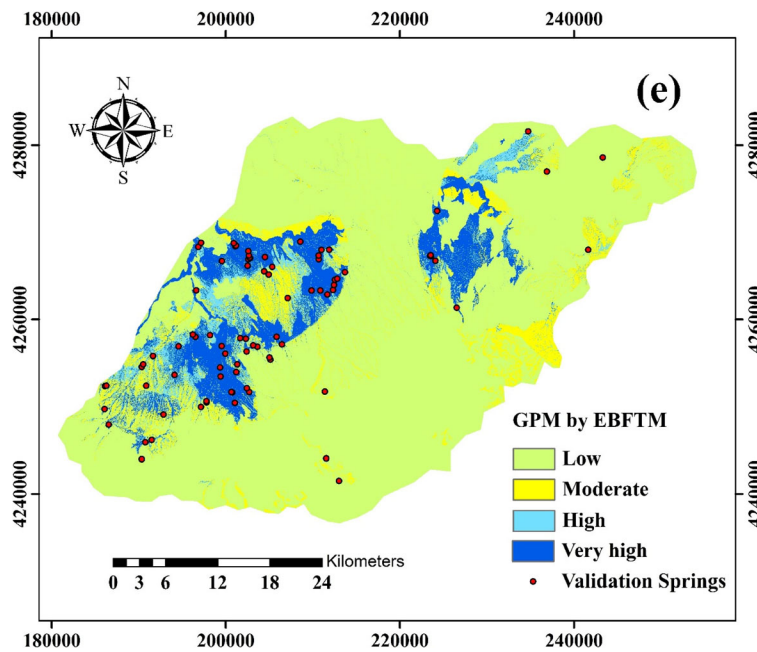
Fig. 4 (continued)

for different classes of the BRT model; and Bel_{CART}, Bel_{RF}, and Bel_{RTF} represent the same value for the CART, RF, and RTF models, respectively.

Validation of the GPMs

The last stage is to evaluate the created GPMs and determine which one provided the best efficiency. In this investigation, receiver operating characteristics (ROC) curve analysis was used for this purpose (Chen et al. 2017a, b; Sangchini et al. 2016). The ROC curve is a plot according to specificity, and sensitivity with diverse cut-off thresholds (Tien Bui et al. 2016). To compare quantitatively, the area under the ROC curves (AUC), which is regarded as a statistical brief of the total efficacy of the implemented algorithms, was employed. An AUC value of one depicts an excellent algorithm that accurately classifies all the spring and non-spring pixels, whereas a value of 0.5 depicts a weaker algorithm (Hong et al. 2016b, 2017a). Additionally, Wilcox signed rank test was conducted in order to define the significance of difference between the new ensemble model, EBFTM, and its elements (i.e., tree-based models such as BRT, CART, RTF, and RF).

Table 3 The distribution of the spring potential values and areas with respect to the groundwater spring potential zones

Model/class		Low	Moderate	High	Very high
BRT	Range	0.04–0.21	0.21–0.40	0.40–0.64	0.64–0.94
	Area %	44.34	27.11	15.65	12.9
CART	Range	0.15–0.17	0.17–0.26	0.26–0.71	0.71–0.89
	Area %	55.44	15.39	17.35	11.82
RF	Range	0.002–0.21	0.21–0.39	0.39–0.61	0.61–0.98
	Area %	36.15	32.14	18.21	13.51
RTF	Range	0.09–0.22	0.22–0.42	0.42–0.65	0.65–0.89
	Area %	43.99	26.51	15.21	14.29
EBFTM	Range	0.02.0.34	0.34–1.11	1.11–2.03	2.03–3.39
	Area %	72.51	9.93	6.58	10.99

Results

Boosted regression tree model

The BRT was tuned by implementing the caret script, and the final BRT model included the number of trees of 100, an interaction depth of 1, a shrinkage of 0.1, and an interaction depth of 1 having accuracy and kappa values of 0.78 and 0.57, respectively. The results of the variable importance are shown in Table 2. Based on the results, land use, altitude, distance from faults, and TWI were the most important GCFs, while river density, fault density, and aspect had the lowest relative influence values. The GPM generated using the BRT algorithm is represented in Fig. 4a and Table 3. Low moderate, high, and very high categories of this GPM included
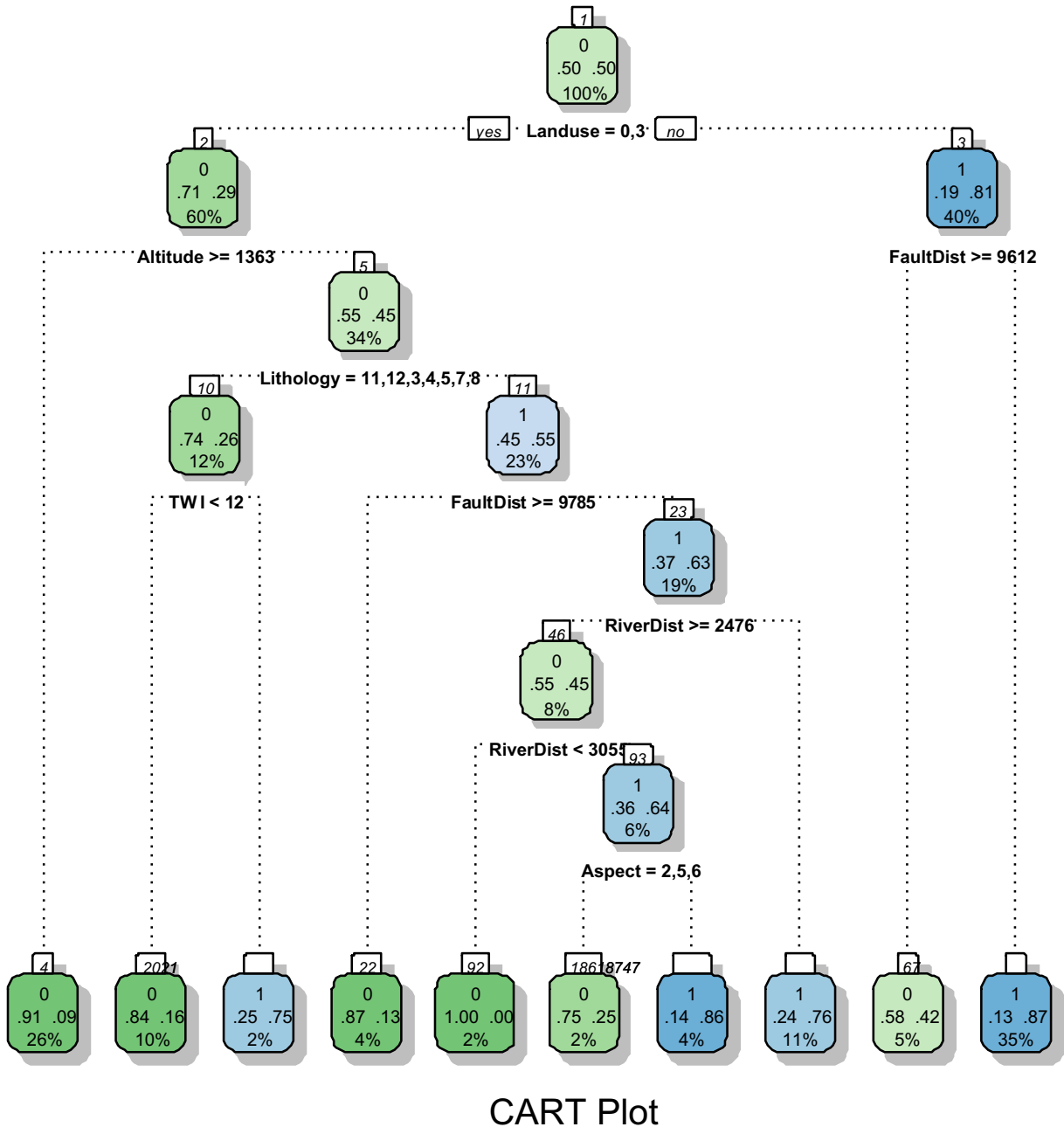
Fig. 5 Pruned classification tree by CART model

44.34, 27.11, 15.65, and 12.90% of the study area, respectively.

Classification and regression tree model

The CART method was trained and pruned by the rpart script in R software having total classified error of 20.4% (Fig. 5). Also, the contribution of the GCFs was determined and displayed in Table 2. The findings showed that land use, distance from faults, and lithology had the highest contributions, while river density and LS had the least contribution in the modeling procedure. The GPM obtained by the CART model is shown in Fig. 4b and Table 3. According to the results, the area allocated to low, moderate, high, and very high categories of the GPM by this algorithm are 55.44, 15.39, 17.35, and 11.82% of the studied area, respectively.

Random forest model

The final model was fitted with a node size of 6, a mtry of 2, and the number of trees of 1000. The "node size" is the minimum size of the nodes. The "mtry" is the number of factors randomly sampled as candidate at each split. Besides, the mean decrease Gini was reported to show the contribution of the GCFs in the concluding model (Table 2). The findings of the mean decrease Gini showed that distance from faults, land use, and altitude had the highest contribution, while fault density and aspect had the least contribution in the modeling procedure. The GPM produced by the RF model is shown in Fig. 4c and Table 3. Low, moderate, high, and very high categories of this GPM cover 36.15, 32.14, 18.21, and 13.51% of the study area, respectively.

Rotation forest model

The RTF model was tuned and fitted using the caret and rotationForest packages in the R software. The final model included a K of 4 and an L of 9 resulting in a ROC value of 0.80 (Fig. 6). K refers to the number of variable subsets and L determines the number of base classifiers. It can be observed that when variable sets range from 1 to 3.5, changes in the number of base classifiers do not have a dramatic effect, while for variable subset of 4, a very distinct result is obtained. The GPM obtained by the RTF algorithm is displayed in Fig. 4d. Considering Table 3, low, moderate, high, and very high categories of this GPM

Fig. 6 The results of cross-validation for rotation forest
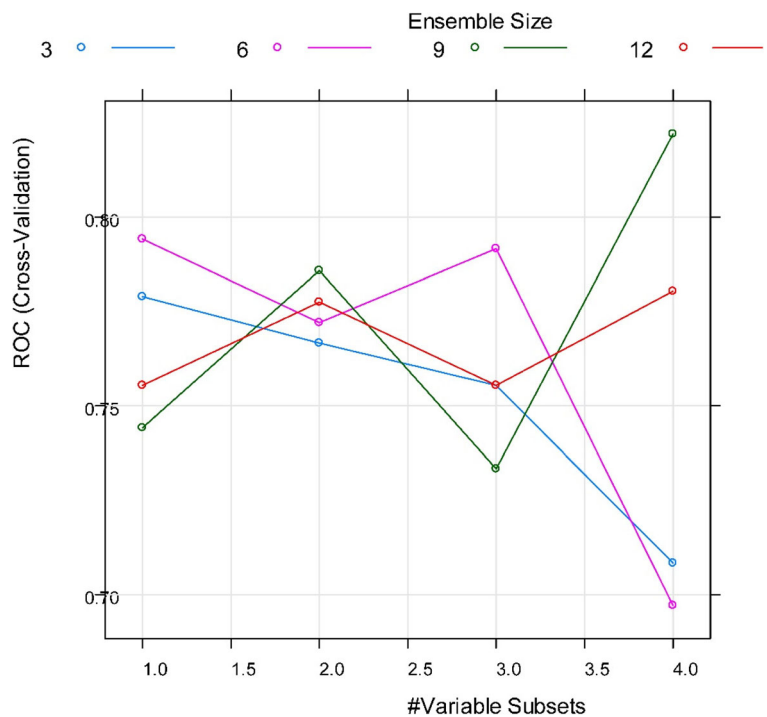
Table 4 Belief values obtained from EBF for different classes of the tree-based models

Class	Evidential belief function (Bel values)			
	BRT	CART	RF	RTF
Low	0.003	0.014	0.002	0.010
Moderate	0.029	0.038	0.015	0.032
High	0.112	0.174	0.086	0.101
Very high	0.856	0.774	0.897	0.857

include 43.99, 26.51, 15.21, and 14.29% of the study region, respectively.
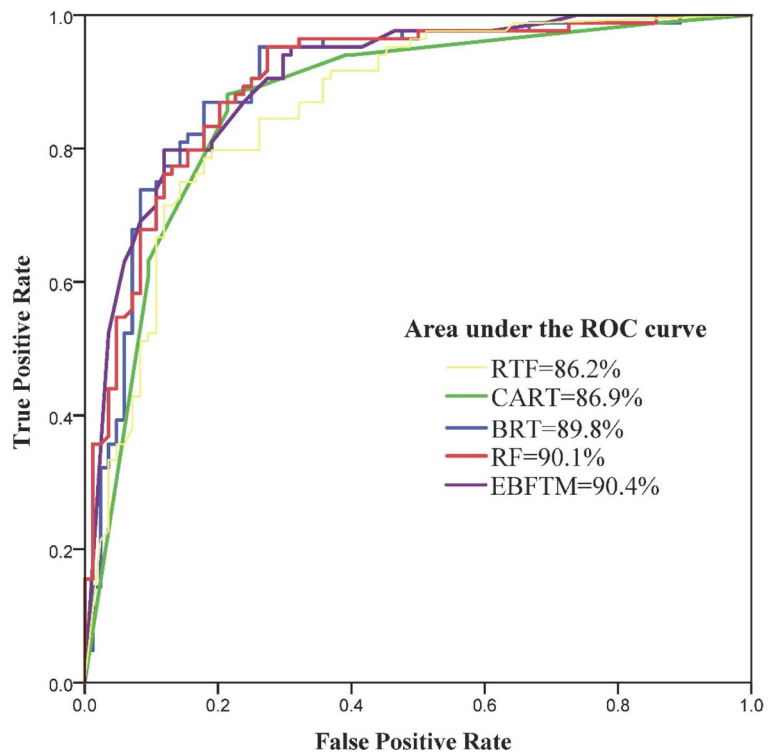
EBFTM results

The findings of the EBFTM are listed in Table 4. According to the table, Bel values for the BRT classes range from 0.003 to 0.856 for low and very high classes of potentiality, respectively. In the case of the CART algorithm, low class has a Bel value of 0.014, while 0.774 is assigned to its very high potential class. In the respect of the RF, it can be observed that low, moderate, high, and very high categories have Bel values of 0.002,

0.015, 0.086, and 0.897, respectively. The findings of the RTF depicted that belief values range from 0.010 to 0.857 for low and very high classes of potentiality, respectively. In addition, the GPM produced by the EBFTM is shown in Fig. 4e and Table 3. It can be observed that a high percent of the area is assigned to the low class (72.51%), whereas a low percent is assigned to the very high potential class (10.99%).

Validation of the models

The AUC-ROC for the models are presented in Fig. 7. As it can be observed, area under the curve of ROC for the implemented methods varies between 86.2 and 90.4%, which shows different prediction capabilities of the methods. The findings showed that the EBFTM, RF, BRT, CART, and RTF algorithms had AUC-ROC values of 90.4, 90.1, 89.8, 86.9, and 86.2%, respectively. The results of the Wilcox signed rank test also showed that there are significant differences between the results of the new ensemble model, EBFTM, with the results of the single tree-based models, i.e., BRT, CART, RF, and RTF (Table 5).

Fig. 7 Prediction rate of the curve for the GPMs produced by tree-based and ensemble models

Table 5 The Wilcox signed rank test results between the accuracy of EBFTM and BRT, CART, RTF, and RF algorithms

Pairwise comparison	BRT vs. EBFTM	CART vs. EBFTM	RTF vs. EBFTM	RF vs. EBFTM
z value	− 5.749	− 6.298	− 5.565	− 5.656
p value	0.000	0.000	0.000	0.000
Significance	Yes	Yes	Yes	Yes

Discussion

GW potential as a hydrogeological phenomenon deals with a wide variety of uncertainties due to the complex nature of GW resources in different areas (Ozdemir 2011a). What makes the situation more complex is the overfitting problem related to the modeling techniques that have been employed in this field of study. Tree-based models face a high amount of uncertainty as they are fitted to datasets as much as possible. The high complexity of GW resources accompanied with different sources of uncertainty in the modeling procedures make the outputs and methodologies regional and reduces the possibility to generalize them. Bartels (1997) recommended model averaging as a recommended solution for this problem. However, there are some more mathematically appropriate methods such as the application of ensemble models. With this in mind, this study applied the EBFTM as a new ensemble model in order to reduce the over fitting issue in each of the tree-based models (i.e., BRT, CART, RF, and RTF algorithms). The results depicted the higher efficacy of the EBFTM comparing with the single models. The findings also clarified that the BRT and RF had similar AUC-ROC values and higher performance capabilities than the CART and RTF models. The findings also depicted that all the tree-based algorithms provided GPMs with higher AUC values than 70%, which shows their acceptable efficacy (Naghibi et al. 2016). The EBFTM produced a more accurate GPM than all the other models since it works as a model averaging technique and could be implemented as a method to reduce the variance. It is known that error could be classified into bias and variance. Naghibi et al. (2017a) stated that ensemble models are believed to combine several weak learners and create a new powerful model. This procedure is regarded as boosting technique in statistical viewpoint. As a matter of fact, since the EBFTM used the same data as the tree-based models, it can be inferred that implementing the ensemble approach could enhance the prediction efficacy with no more data. This finding is in agreement with the results obtained in Lee et al. (2012a, b) in the field of landslide susceptibility mapping.

In the respect of the BRT model, there are some strong characteristics that could result in its high efficacy in classification problems. For instance, BRT can keep important GCFs, detect the interactions, and also it is able to delete the variables with large number of missing values (Elith et al. 2008; Carty 2011). RF algorithm has been reported to have acceptable efficacy in different fields of study (Peters et al. 2007; Vorpahl et al. 2012; Chen et al. 2018d; Hong et al. 2017b, 2018). In this work, RF showed acceptable results. There are some advantages in RF algorithm such as the low aptitude to overfitting, and the capability to support high-dimensional datasets (Caruana and Niculescu-Mizil 2006). In the case of interpreting, it can be seen that among all the tree-based algorithms, which are employed in this research, the CART has the highest interpretability. This kind of information could be very helpful for water resource planners and managers who are not very familiar with the modeling procedures. Breiman et al. (1984) mentioned that CART is able to be conducted on non-linear data and has no requirement to factor transformation. Furthermore, outliers do not have a high influence on the results (Strobl et al. 2008). Besides, another positive feature in CART is that it does not have adverse impact on numerical calculations. In the respect of the RTF algorithm, it produced a GPM with lower accuracy than the other three tree-based models; however, its AUC-ROC is 86.2% and this confirms that the RTF is an appropriate GW predictor especially under the uncertain condition of GW resources and lack of data. The main limitation of this work like other works applied on spatial sciences such as groundwater potential mapping is that the scientists need to investigate the results of their models in more than one study area to make sure that the results are

general. Therefore, the authors suggest the future works to work on two or more study areas to achieve this goal.

This research also investigated the importance of the GCFs in modeling GW potential. It was seen that land use was the most important factor in the BRT and CART, and the second most important factor in the RF. Regarding the land use classes of the study area, it can be seen that most of the springs have occurred in agriculture and orchard categories. This accumulation of the springs could be mentioned as a reason of its importance in the modeling process. Distance from faults, altitude, and lithology were the following most important GCFs in this study. An investigation of the distance from faults layer shows that most of the training springs are occurred in far distances. This shows a direct relationship between GW potential and distance from faults. In another word, springs are not occurred at the location of the faults. Altitude has an inverse influence on spring occurrence as it impacts slope angle, river system development, and agricultural and industrial development. All these factors influence GW potential in an area. An investigation of the location of the training springs regarding the altitude shows that some altitude ranges do not have any spring occurrence. Lithological characteristics of a watershed impact its hydrological as well as hydrogeological features including soil permeability, transmissivity, and river system development. It can be seen that the springs are concentrated in piedmont fan and valley terraces with high amount of permeability. On the other hand, river density, and fault density were the least important GCFs.

In fact, one important characteristic of a classification model could be regarded as how precise it is in determining the very high class of potentiality. The finding of this work depicted that the very high class of potentiality in the GPMs generated by the BRT, CART, RF, RTF, and EBFTM algorithms were calculated as 12.9, 11.82, 13.51, 14.29, and 10.99%, respectively. This finding explains that the new ensemble model (i.e., EBFTM) in addition to its higher performance can determine the very high potential class more precise than the other models.

Conclusions

Water resource-related issues are growing and this fact enhances the necessity to gain a better understanding of these important natural resources. Thus, investigating different aspects of water including surface and GW resources has a high priority. This paper made an attempt to generate a new methodology (i.e., EBFTM method) to model GW potentiality as a useful guideline for water resource managers. Performance of the EBFTM was validated by using ROC curve and accuracy index. The findings of this investigation represented that the EBFTM had better efficacy than all the tree-based models implemented in this work. Another novelty of this study was application of the RTF algorithm in producing a GPM for the first time, and its performance was compared to other data-mining algorithms (i.e., BRT, CART, and RF). The findings depicted that although the RTF had weaker efficacy than the other tree-based models, it gave acceptable efficacy based on area under the ROC curve. Model averaging techniques such as the one implemented in this research (i.e., EBFTM) could be implemented with less uncertainty resulting in more general outputs. Regarding this fact, this study suggests the application of ensemble models like the EBFTM in cross-application studies. These kinds of general models could be more applicable in situations where the managers deal with a larger area with a high variety of hydrogeological, climatic, and topographical features. The outlook of this research in companion with other papers applied on ensemble models could be to lead the research community to work on more advanced ensemble models and obtain better results in order to solve complex geospatial issues. Overall, considering the five constructed GPMs in this study, the authors can suggest west central parts of the study area for water extraction for different water uses. A smaller area at the eastern part of the watershed can also be conservatively considered for water extraction.

References

Adiat, K. A. N., Nawawi, M. N. M., & Abdullah, K. (2012). Assessing the accuracy of GIS-based elementary multi criteria decision analysis as a spatial prediction tool—a case of predicting potential zones of sustainable groundwater resources. *Journal of Hydrology, 440–441*, 75–89. https://doi.org/10.1016/J.JHYDROL.2012.03.028.

Aertsen, W., Kint, V., van Orshoven, J., Özkan, K., & Muys, B. (2010). Comparison and ranking of different modelling techniques for prediction of site index in Mediterranean mountain forests. *Ecological Modelling, 221*(8), 1119–1130. https://doi.org/10.1016/J.ECOLMODEL.2010.01.007.

Aghdam, I. N., Varzandeh, M. H. M., & Pradhan, B. (2016). Landslide susceptibility mapping using an ensemble statistical index (Wi) and adaptive neuro-fuzzy inference system (ANFIS) model at Alborz Mountains (Iran). *Environmental Earth Sciences, 75*(7), 553. https://doi.org/10.1007/s12665-015-5233-6.

Aličković, E., & Subasi, A. (2017). Breast cancer diagnosis using GA feature selection and rotation forest. *Neural Computing and Applications, 28*(4), 753–763. https://doi.org/10.1007/s00521-015-2103-9.

Althuwaynee, O. F., Pradhan, B., Park, H.-J., & Lee, J. H. (2014). A novel ensemble bivariate statistical evidential belief function with knowledge-based analytical hierarchy process and multivariate statistical logistic regression for landslide susceptibility mapping. *Catena, 114*, 21–36. https://doi.org/10.1016/j.catena.2013.10.011.

Asia, S., & Richman, P. (1991). Planning for groundwater protection, 273. https://doi.org/10.1016/S0376-7361(09)70018-4.

Bartels, L. (1997). Specification uncertainty and model averaging. *American Journal of Political Science, 41*(2), 641–674. https://doi.org/10.2307/2111781.

Beven, K. (1997). TOPMODEL: a critique. *Hydrological Process*, *11*, 1069–1085.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32. https://doi.org/10.1023/A:1010933404324.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and Regression Trees. *Wadsworth and Brooks-Cole Advanced Books and Software*. California: Pacific Grove.

Carranza, E. J. M., van Ruitenbeek, F. J. a., Hecker, C., van der Meijde, M., & van der Meer, F. D. (2008). Knowledge-guided data-driven evidential belief modeling of mineral prospectivity in Cabo de Gata, SE Spain. *International Journal of Applied Earth Observation and Geoinformation, 10*(3), 374–387. https://doi.org/10.1016/j.jag.2008.02.008.

Carty, D. (2011). An analysis of boosted regression trees to predict the strength properties of wood composites. *Masters Theses*. http://trace.tennessee.edu/utk_gradthes/954. Accessed 5 April 2018.

Caruana, R., & Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd international conference on Machine learning - ICML '06 (pp. 161–168). New York*. New York: ACM Press. https://doi.org/10.1145/1143844.1143865.

Chen, W., Pourghasemi, H. R., & Naghibi, S. A. (2017a). A comparative study of landslide susceptibility maps produced using support vector machine with different kernel functions and entropy data mining models in China. *Bulletin of Engineering Geology and the Environment, 77*, 1–18. https://doi.org/10.1007/s10064-017-1010-y.

Chen, W., Pourghasemi, H. R., & Naghibi, S. A. (2017b). Prioritization of landslide conditioning factors and its spatial modeling in Shangnan County, China using GIS-based data mining algorithms. *Bulletin of Engineering Geology and the Environment., 77*, 611–629. https://doi.org/10.1007/s10064-017-1004-9.

Chen, W., Shirzadi, A., Shahabi, H., Ahmad, B. B., Zhang, S., Hong, H., & Zhang, N. (2017c). A novel hybrid artificial intelligence approach based on the rotation forest ensemble and naïve Bayes tree classifiers for a landslide susceptibility assessment in Langao County, China. *Geomatics, Natural Hazards and Risk, 8*(2), 1955–1977. https://doi.org/10.1080/19475705.2017.1401560.

Chen, W., Li, H., Hou, E., Wang, S., Wang, G., Panahi, M., Li, T., Peng, T., Guo, C., Niu, C., Xiao, L., Wang, J., Xie, X., & Ahmad, B. B. (2018a). GIS-based groundwater potential analysis using novel ensemble weights-of-evidence with logistic regression and functional tree models. *Science of the Total Environment, 634*, 853–867. https://doi.org/10.1016/J.SCITOTENV.2018.04.055.

Chen, W., Peng, J., Hong, H., Shahabi, H., Pradhan, B., Liu, J., Zhu, A. X., Pei, X., & Duan, Z. (2018b). Landslide susceptibility modelling using GIS-based machine learning techniques for Chongren County, Jiangxi Province, China. *Science of the Total Environment, 626*, 1121–1135. https://doi.org/10.1016/J.SCITOTENV.2018.01.124.

Chen, W., Shahabi, H., Shirzadi, A., Li, T., Guo, C., Hong, H., Li, W., Pan, D., Hui, J., Ma, M., Xi, M., & Bin Ahmad, B. (2018c). A novel ensemble approach of bivariate statistical-based logistic model tree classifier for landslide susceptibility assessment. *Geocarto International, 33*, 1–23. https://doi.org/10.1080/10106049.2018.1425738.

Chen, W., Xie, X., Peng, J., Shahabi, H., Hong, H., Bui, D. T., et al. (2018d). GIS-based landslide susceptibility evaluation using a novel hybrid integration approach of bivariate statistical based random forest method. *CATENA, 164*, 135–149. https://doi.org/10.1016/J.CATENA.2018.01.012.

Chezgi, J., Pourghasemi, H. R., Naghibi, S. A., Moradi, H. R., & Kheirkhah Zarkesh, M. (2016). Assessment of a spatial multi-criteria evaluation to site selection underground dams in the Alborz Province, Iran. *Geocarto International, 31*(6), 628–646. https://doi.org/10.1080/10106049.2015.1073366.

Dempster, A. P. (1968). A generalization of Bayesian inference. *Journal of the Royal Statistical Society. Series B (Methodological). WileyRoyal Statistical Society.* https://doi.org/10.2307/2984504.

Elith, J., & Leathwick, J. (2013). *Boosted regression trees for ecological modeling* (pp. 1–22).

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *The Journal of Animal Ecology, 77*(4), 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x.

Geology Survey of Iran (GSI) (1997) http://www.gsi.ir/Main/Lang_en/index.html

Ghorbani Nejad, S., Falah, F., Daneshfar, M., Haghizadeh, A., & Rahmati, O. (2016). Delineation of groundwater potential zones using remote sensing and GIS-based data-driven models. *Geocarto International*, 1–21. https://doi.org/10.1080/10106049.2015.1132481.

Golkarian, A., Naghibi, S. A., Kalantar, B., & Pradhan, B. (2018). Groundwater potential mapping using C5.0, random forest, and multivariate adaptive regression spline models in GIS, (190:149). https://doi.org/10.1007/s10661-018-6507-8

Hong, H., Naghibi, S. A., Pourghasemi, H. R., & Pradhan, B. (2016a). GIS-based landslide spatial modeling in Ganzhou City, China. *Arabian Journal of Geosciences, 9*(2), 112. https://doi.org/10.1007/s12517-015-2094-y.

Hong, H., Pourghasemi, H. R., & Pourtaghi, Z. S. (2016b). Landslide susceptibility assessment in Lianhua County (China): a comparison between a random forest data mining technique and bivariate and multivariate statistical models.

Geomorphology, 259, 105–118. https://doi.org/10.1016/J.GEOMORPH.2016.02.012.

Hong, H., Naghibi, S. A., Moradi Dashtpagerdi, M., Pourghasemi, H. R. H. R., & Chen, W. (2017a). A comparative assessment between linear and quadratic discriminant analyses (LDA-QDA) with frequency ratio and weights-of-evidence models for forest fire susceptibility mapping in China. Arabian Journal of Geosciences, 10(7). https://doi.org/10.1007/s12517-017-2905-4.

Hong, H., Tsangaratos, P., Ilia, I., Chen, W., & Xu, C. (2017b). Comparing the performance of a logistic regression and a random Forest model in landslide susceptibility assessments. the case of Wuyaun area, China. In Advancing culture of living with landslides (pp. 1043–1050). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-53498-5_118.

Hong, H., Liu, J., Bui, D. T., Pradhan, B., Acharya, T. D., Pham, B. T., et al. (2018). Landslide susceptibility mapping using J48 decision tree with AdaBoost, bagging and rotation forest ensembles in the Guangchang area (China). CATENA, 163, 399–413. doi:https://doi.org/10.1016/J.CATENA.2018.01.005.

Hong, H., Miao, Y., Liu, J., & Zhu, A.-X. (2019). Exploring the effects of the design and quantity of absence data on the performance of random forest-based landslide susceptibility mapping. CATENA, 176, 45–64. https://doi.org/10.1016/J.CATENA.2018.12.035.

Khorasan Razavi Regional Water Authority (2015). http://www.khrw.ir/?l=EN.

Khosravi, K., Pham, B. T., Chapi, K., Shirzadi, A., Shahabi, H., Revhaug, I., Prakash, I., & Tien Bui, D. (2018a). A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran. Science of the Total Environment, 627, 744–755. https://doi.org/10.1016/j.scitotenv.2018.01.266.

Khosravi, K., Sartaj, M., Tsai, F. T.-C., Singh, V. P., Kazakis, N., Melesse, A. M., Prakash, I., Tien Bui, D., & Pham, B. T. (2018b). A comparison study of DRASTIC methods with various objective methods for groundwater vulnerability assessment. Science of the Total Environment, 642, 1032–1049. https://doi.org/10.1016/J.SCITOTENV.2018.06.130s.

Kordestani, M. D., Naghibi, S. A., Hashemi, H., Ahmadi, K., Kalantar, B., & Pradhan, B. (2018). Groundwater potential mapping using a novel data-mining ensemble model. Hydrogeology Journal, 27(1), 211–214. https://doi.org/10.1007/s10040-018-1848-5.

Koyuncu, H., & Ceylan, R. (2013). Artificial neural network based on rotation forest for biomedical pattern classification. In 2013 36th International Conference on Telecommunications and Signal Processing (TSP) (pp. 581–585). IEEE. https://doi.org/10.1109/TSP.2013.6614001.

Lee, M.-J., Choi, J.-W., Oh, H.-J., Won, J.-S., Park, I., & Lee, S. (2012a). Ensemble-based landslide susceptibility maps in Jinbu area, Korea. Environmental Earth Sciences, 67(1), 23–37. https://doi.org/10.1007/s12665-011-1477-y.

Lee, S., Kim, Y. S., & Oh, H. J. (2012b). Application of a weights-of-evidence method and GIS to regional groundwater productivity potential mapping. Journal of Environmental Management, 96(1), 91–105. https://doi.org/10.1016/j.jenvman.2011.09.016.

Lu, H., Yang, L., Yan, K., Xue, Y., & Gao, Z. (2017). A cost-sensitive rotation forest algorithm for gene expression data classification. Neurocomputing, 228, 270–276. https://doi.org/10.1016/J.NEUCOM.2016.09.077.

Manap, M. A., Nampak, H., Pradhan, B., Lee, S., Sulaiman, W. N. A., & Ramli, M. F. (2014). Application of probabilistic-based frequency ratio model in groundwater potential mapping using remote sensing data and GIS. Arabian Journal of Geosciences, 7(2), 711–724. https://doi.org/10.1007/s12517-012-0795-z.

McKenney, D. W., & Pedlar, J. H. (2003). Spatial models of site index based on climate and soil properties for two boreal tree species in Ontario, Canada. Forest Ecology and Management, 175(1–3), 497–507. https://doi.org/10.1016/S0378-1127(02)00186-X.

Micheletti, N., Foresti, L., Robert, S., Leuenberger, M., Pedrazzini, A., Jaboyedoff, M., & Kanevski, M. (2013). Machine learning feature selection methods for landslide susceptibility mapping. Mathematical Geosciences, 46(1), 33–57. https://doi.org/10.1007/s11004-013-9511-0.

Moore, I. D., & Burch, G. J. (1986). Sediment transport capacity ofsheet and rill flow: application of unit stream power theory. Water Research, 22(8), 1350–1360.

Moore, I. D., Grayson, R. B., & Ladson, A. R. (1991). Digital terrain modeling: a review of hydrological, geomorphological and biological applications. Hydrological Processes, 5, 3–30. https://doi.org/10.1002/hyp.3360050103.

Mousavi, S., Golkarian, A., Amir Naghibi, S., Kalantar, B., & Pradhan, B. (2017). GIS-based groundwater spring potential mapping using data mining boosted regression tree and probabilistic frequency ratio models in Iran. AIMS Geosciences, 3(1), 91–115. https://doi.org/10.3934/geosci.2017.1.91.

Naghibi, S. A., Pourghasemi, H. R., Pourtaghi, Z. S., & Rezaei, A. (2015). Groundwater qanat potential mapping using frequency ratio and Shannon's entropy models in the Moghan watershed, Iran. Earth Science Informatics, 8(1), 1–16. https://doi.org/10.1007/s12145-014-0145-7.

Naghibi, S. A., & Moradi Dashtpagerdi, M. (2016). Evaluation of four supervised learning methods for groundwater spring potential mapping in Khalkhal region (Iran) using GIS-based features. Hydrogeology Journal, 25, 169–189. https://doi.org/10.1007/s10040-016-1466-z.

Naghibi, S. A., & Pourghasemi, H. R. (2015). A comparative assessment between three machine learning models and their performance comparison by bivariate and multivariate statistical methods in groundwater potential mapping. Water Resources Management, 29(14), 5217–5236. https://doi.org/10.1007/s11269-015-1114-8.

Naghibi, S. A., Pourghasemi, H. R., & Dixon, B. (2016). GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran. Environmental Monitoring and Assessment, 188(1), 44. https://doi.org/10.1007/s10661-015-5049-6.

Naghibi, S. A., Moghaddam, D. D., Kalantar, B., Pradhan, B., & Kisi, O. (2017a). A comparative assessment of GIS-based data mining models and a novel ensemble model in groundwater well potential mapping. Journal of Hydrology, 548, 471–483. https://doi.org/10.1016/j.jhydrol.2017.03.020.

Naghibi, S. A., Ahmadi, K., & Daneshi, A. (2017b). Application of support vector machine, random Forest, and genetic

algorithm optimized random Forest models in groundwater potential mapping. *Water Resources Management, 31*(9), 1–15. https://doi.org/10.1007/s11269-017-1660-3.

Naghibi, S. A., Pourghasemi, H. R., & Abbaspour, K. (2018a). A comparison between ten advanced and soft computing models for groundwater qanat potential assessment in Iran using R and GIS. *Theoretical and Applied Climatology, 131*(3–4), 967–984. https://doi.org/10.1007/s00704-016-2022-4.

Naghibi, S. A., Vafakhah, M., Hashemi, H., Pradhan, B., & Alavi, S. (2018b). Groundwater augmentation through the site selection of floodwater spreading using a data mining approach (case study: Mashhad plain, Iran). *Water, 10*(10), 1405. https://doi.org/10.3390/w10101405.

Nampak, H., Pradhan, B., & Manap, M. A. (2014). Application of GIS based data driven evidential belief function model to predict groundwater potential zonation. *Journal of Hydrology, 513*, 283–300. https://doi.org/10.1016/j.jhydrol.2014.02.053.

Oh, H.-J., Kim, Y.-S., Choi, J.-K., Park, E., & Lee, S. (2011). GIS mapping of regional probabilistic groundwater potential in the area of Pohang City, Korea. *Journal of Hydrology, 399*(3–4), 158–172. https://doi.org/10.1016/j.jhydrol.2010.12.027.

Ozdemir, A. (2011a). GIS-based groundwater spring potential mapping in the Sultan Mountains (Konya, Turkey) using frequency ratio, weights of evidence and logistic regression methods and their comparison. *Journal of Hydrology, 411*(3–4), 290–308. https://doi.org/10.1016/j.jhydrol.2011.10.010.

Ozdemir, A. (2011b). Using a binary logistic regression method and GIS for evaluating and mapping the groundwater spring potential in the Sultan Mountains (Aksehir, Turkey). *Journal of Hydrology, 405*(1–2), 123–136. https://doi.org/10.1016/j.jhydrol.2011.05.015.

Pardo, M., & Sberveglieri, G. (2008). Random forests and nearest shrunken centroids for the classification of sensor array data. *Sensors and Actuators, B: Chemical, 131*(1), 93–99. https://doi.org/10.1016/j.snb.2007.12.015.

Park, I., Lee, J., & Saro, L. (2014). Ensemble of ground subsidence hazard maps using fuzzy logic. *Central European Journal of Geosciences, 6*(2), 207–218. https://doi.org/10.2478/s13533-012-0175-y.

Peters, J., De Baets, B., Verhoest, N. E. C., Samson, R., Degroeve, S., De Becker, P., & Huybrechts, W. (2007). Random forests as a tool for ecohydrological distribution modelling. *Ecological Modelling, 207*(2–4), 304–318. https://doi.org/10.1016/J.ECOLMODEL.2007.05.011.

Pham, B. T., Tien Bui, D., Prakash, I., & Dholakia, M. B. (2017). Hybrid integration of multilayer perceptron neural networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. *CATENA, 149*, 52–63. https://doi.org/10.1016/J.CATENA.2016.09.007.

Pham, B. T. (2018). A novel classifier based on composite hypercubes on iterated random projections for assessment of landslide susceptibility. *Journal of the Geological Society of India, 91*(3), 355–362. https://doi.org/10.1007/s12594-018-0862-5s.

Pham, B. T., & Prakash, I. (2018). Machine learning methods of kernel logistic regression and classification and regression trees for landslide susceptibility assessment at part of

Himalayan area, India. *Indian Journal of Science and Technology, 11*(12), 1–10. https://doi.org/10.17485/ijst/2018/v11i12/99745.

Pham, B. T., Shirzadi, A., Tien Bui, D., Prakash, I., & Dholakia, M. B. (2018a). A hybrid machine learning ensemble approach based on a radial basis function neural network and rotation forest for landslide susceptibility modeling: a case study in the Himalayan area, India. *International Journal of Sediment Research, 33*(2), 157–170. https://doi.org/10.1016/J.IJSRC.2017.09.008.

Pham, B. T., Tien Bui, D., & Prakash, I. (2018b). Bagging based support vector machines for spatial prediction of landslides. *Environmental Earth Sciences, 77*(4), 146. https://doi.org/10.1007/s12665-018-7268-y.

Pham, B. T., Prakash, I., & Tien Bui, D. (2018c). Spatial prediction of landslides using a hybrid machine learning approach based on random subspace and classification and regression trees. *Geomorphology, 303*, 256–270. https://doi.org/10.1016/J.GEOMORPH.2017.12.008.

Pham, B. T., Jaafari, A., Prakash, I., & Bui, D. T. (2018d). A novel hybrid intelligent model of support vector machines and the MultiBoost ensemble for landslide susceptibility modeling. *Bulletin of Engineering Geology and the Environment*, 1–22. https://doi.org/10.1007/s10064-018-1281-y.

Pham, B. T., Prakash, I., Khosravi, K., Chapi, K., Trinh, P. T., Ngo, T. Q., Hosseini, S. V., & Bui, D. T. (2018e). A comparison of support vector machines and Bayesian algorithms for landslide susceptibility modelling. *Geocarto International*, 1–23. https://doi.org/10.1080/10106049.2018.1489422.

Pourtaghi, Z. S., & Pourghasemi, H. R. (2014). GIS-based groundwater spring potential assessment and mapping in the Birjand Township, southern Khorasan Province, Iran. *Hydrogeology Journal, 22*, 643–662. https://doi.org/10.1007/s10040-013-1089-6.

Pradhan, B., Abokharima, M. H., Jebur, M. N., & Tehrany, M. S. (2014). Land subsidence susceptibility mapping at Kinta Valley (Malaysia) using the evidential belief function model in GIS. *Natural Hazards., 73*, 1019–1042. https://doi.org/10.1007/s11069-014-1128-1.

Prasad, V. K., Badarinath, K. V. S., & Eaturu, A. (2008). Biophysical and anthropogenic controls of forest fires in the Deccan Plateau, India. *Journal of Environmental Management, 86*(1), 1–13. https://doi.org/10.1016/J.JENVMAN.2006.11.017.

Rahmati, O., Nazari Samani, A., Mahdavi, M., Pourghasemi, H. R., & Zeinivand, H. (2014). Groundwater potential mapping at Kurdistan region of Iran using analytic hierarchy process and GIS. *Arabian Journal of Geosciences, 8*(February 2016), 7059–7071. https://doi.org/10.1007/s12517-014-1668-4.

Rahmati, O., & Melesse, A. M. (2016). Application of Dempster–Shafer theory, spatial analysis and remote sensing for groundwater potentiality and nitrate pollution analysis in the semiarid region of Khuzestan, Iran, (June). doi:https://doi.org/10.1016/j.scitotenv.2016.06.176, 568, 1110, 1123.

Rahmati, O., Pourghasemi, H. R., & Melesse, A. M. (2016). Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: a case study at Mehran region, Iran. *Catena, 137*(October), 360–372. https://doi.org/10.1016/j.catena.2015.10.010.

Rahmati, O., Naghibi, S. A., Shahabi, H., Bui, D. T., Pradhan, B., Azareh, A., Rafiei-Sardooi, E., Samani, A. N., & Melesse, A. M. (2018). Groundwater spring potential modelling: comprising the capability and robustness of three different modeling approaches. *Journal of Hydrology, 565*, 248–261. https://doi.org/10.1016/J.JHYDROL.2018.08.027.

Ripley, B (2015). Package "rpart."

Rodriguez, J. J., Kuncheva, L. I., & Alonso, C. J. (2007). Rotation forest: a new classifier ensemble method. *Pattern Analysis Machine Intelligence IEEE Transactions, 28*, 1619–1630. https://doi.org/10.1109/TPAMI.2006.211.

Sangchini, E. K., Emami, S. N., Tahmasebipour, N., Pourghasemi, H. R., Naghibi, S. A., Arami, S. A., & Pradhan, B. (2016). Assessment and comparison of combined bivariate and AHP models with logistic regression for landslide susceptibility mapping in the Chaharmahal-e-Bakhtiari Province, Iran. *Arabian Journal of Geosciences, 9*(3), 201. https://doi.org/10.1007/s12517-015-2258-9.

Shafer, G. (1976). A mathematical theory of evidence. Princeton University Press. https://press.princeton.edu/titles/2439.html. Accessed 5 April 2018.

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics, 9*(1), 307. https://doi.org/10.1186/1471-2105-9-307.

Tahmassebipoor, N., Rahmati, O., Noormohamadi, F., & Lee, S. (2016). Spatial analysis of groundwater potential using weights-of-evidence and evidential belief function models and remote sensing. *Arabian Journal of Geosciences, 9*(1), 79. https://doi.org/10.1007/s12517-015-2166-z.

Tehrany, M. S., Pradhan, B., & Jebur, M. N. (2013). Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS. *Journal of Hydrology, 504*, 69–79. https://doi.org/10.1016/j.jhydrol.2013.09.034.

Tehrany, M. S., Pradhan, B., & Jebur, M. N. (2014). Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. *Journal of Hydrology, 512*, 332–343. https://doi.org/10.1016/j.jhydrol.2014.03.008.

Tien Bui, D., Pradhan, B., Lofman, O., Revhaug, I., & Dick, O. B. (2012). Spatial prediction of landslide hazards in Hoa Binh province (Vietnam): a comparative assessment of the efficacy of evidential belief functions and fuzzy logic models. *Catena, 96*, 28–40. https://doi.org/10.1016/j.catena.2012.04.001.

Tien Bui, D., Tuan, T. A., Klempe, H., Pradhan, B., & Revhaug, I. (2016). Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides, 13*(2), 361–378. https://doi.org/10.1007/s10346-015-0557-6.

Umar, Z., Pradhan, B., Ahmad, A., Jebur, M. N., & Tehrany, M. S. (2014). Earthquake induced landslide susceptibility mapping using an integrated ensemble frequency ratio and logistic regression models in West Sumatera Province, Indonesia. *Catena, 118*(September 2009), 124–135. https://doi.org/10.1016/j.catena.2014.02.005.

Vorpahl, P., Elsenbeer, H., Märker, M., & Schröder, B. (2012). How can statistical models help to determine driving factors of landslides? *Ecological Modelling, 239*, 27–39. https://doi.org/10.1016/j.ecolmodel.2011.12.007.

Wan, S., Lei, T.-C., & Chou, T.-Y. (2012). A landslide expert system: image classification through integration of data mining approaches for multi-category analysis. *International Journal of Geographical Information Science, 26*(4), 747–770. https://doi.org/10.1080/13658816.2011.613397.

Xia, J., Falco, N., Benediktsson, J. A., Du, P., & Chanussot, J. (2017). Hyperspectral image classification with rotation random forest via KPCA. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10*(4), 1601–1609. https://doi.org/10.1109/JSTARS.2016.2636877.

Youssef, A. M., Pradhan, B., Jebur, M. N., & El-Harbi, H. M. (2015). Landslide susceptibility mapping using ensemble bivariate and multivariate statistical models in Fayfa area, Saudi Arabia. *Environmental Earth Sciences, 73*(7), 3745–3761. https://doi.org/10.1007/s12665-014-3661-3.

Zabihi, M., Pourghasemi, H. R., Pourtaghi, Z. S., & Behzadfar, M. (2016). GIS-based multivariate adaptive regression spline and random forest models for groundwater potential mapping in Iran. *Environmental Earth Sciences, 75*(8), 665. https://doi.org/10.1007/s12665-016-5424-9.