




# Mapping Groundwater Potential Using a Novel Hybrid Intelligence Approach

Shaghayegh Miraki<sup>1</sup> · Sasan Hedayati Zanganeh<sup>2</sup> · Kamran Chapi<sup>3</sup>  · Vijay P. Singh<sup>4</sup> · Ataollah Shirzadi<sup>3</sup> · Himan Shahabi<sup>5</sup> · Binh Thai Pham<sup>6</sup>

Received: 26 September 2017 / Accepted: 6 September 2018 /

Published online: 10 September 2018

© Springer Nature B.V. 2018

## Abstract

Identifying areas with high groundwater potential is important for groundwater resources management. The main objective of this study is to propose a novel classifier ensemble method, namely Random Forest Classifier based on Random Subspace Ensemble (RS-RF), for groundwater potential mapping (GWPM) in Qorveh-Dehgolan plain, Kurdistan province, Iran. A total of 12 conditioning factors (slope, aspect, elevation, curvature, stream power index (SPI), topographic wetness index (TWI), rainfall, lithology, land use, normalized difference vegetation index (NDVI), fault density, and river density) were selected for groundwater modeling. The least square support vector machine (LSSVM) feature selection method with a 10-fold cross-validation technique was used to validate the predictive capability of these conditioning factors for training the models. The performance of the RS-RF model was validated using the area under receiver operating characteristic curve (AUROC), success and prediction rate curves, kappa index, and several statistical index-based measures. In addition, Friedman and Wilcoxon signed-rank tests were used to assess statistically significant level among the new model with the state-of-the-art soft computing benchmark models, such as random forest (RF), logistic regression (LR) and naïve Bayes (NB). Results showed that the new hybrid model of RS-RF had a very high predictive capability for groundwater potential mapping and exhibited the best performance among other benchmark models (LR, RF, and NB). Results of the present study might be useful to water managers to make proper decisions on the optimal use of groundwater resources for future planning in the critical study area.

**Keywords** Mapping groundwater potential · Least square support vector machine · Random forest random subspace ensemble · Logistic regression · Kurdistan

---

## Highlights

- A new hybrid machine learning model named RS-RF was introduced for mapping groundwater potential.
- The new model was successfully evaluated and validated compared to the RF, LR, and NB models.
- The new model can be used as a tool for groundwater resources management in semi-arid areas.

---

✉ Kamran Chapi  
k.chapi@uok.ac.ir

Back Affiliation

## 1 Introduction

As one of the most critical resources throughout the world, groundwater supplies required water for agriculture, industry, animal husbandry, and human communities (Neshat et al. 2014). About one third of world population is fully dependent on groundwater (Rahmati et al. 2018) such that it is the main source of drinking water in many countries, such as the U.S., Canada and Germany and the only source for community uses in some other countries, such as Austria, Denmark, and Lithuania (Manap et al. 2013). In Iran, over 70% of the people, living in rural and urban areas, are reliant on groundwater resources for their drinking and domestic requirements (Rahmati 2013). Recently, many regions of Iran have become dried up due mainly to climate change and intensive withdrawal of available groundwater resources, resulting in a serious lack of water in many provinces of the country (Osati et al. 2014).

Driven by rainwater, surface water, and snow melting, groundwater occurrence and movement are affected by many factors which are related to topography, lithology, geologic structures, fracture density, aperture and connectivity, secondary porosity, groundwater table distribution, groundwater recharge, slope, drainage pattern, landforms, land cover, climatic condition, and their interrelationships and interactions (Oh et al. 2011). Due to the plurality of affecting factors, groundwater assessment and identifying its potential areas are complicated and still a challenge to water resources managers.

To identify groundwater potential areas, it is necessary to apply methods that will assist managers to use these sources more effectively (Rahmati et al. 2015). These methods are needed for future development, management, and prevention of declining groundwater resources. Many studies have used GIS and RS techniques for assessing groundwater potential mapping (Jha et al. 2007), including frequency ratio (FR) (Naghbi et al. 2015), multi-criteria decision analysis (Chenini et al. 2010), weight of evidence (WOE) (Tahmassebpour et al. 2016), and analytical hierarchy processing (AHP) (Shekhar and Pandey 2015).

During the past decade, machine learning algorithms, such as random forest (Rahmati et al. 2015), extreme learning machine (ELM) (Lian et al. 2014), support vector machine (SVM) (Micheletti et al. 2014), logistic regression (LR) (Mair and El-Kadi 2013), Naive Bayes (NB) (Farid et al. 2014) and decision tree (Tehrany et al. 2013) have been applied to groundwater potential mapping. More recently, machine learning ensemble techniques, known as efficient techniques, are becoming popular for enhancing the prediction accuracy of weaker single classifier (Pham et al. 2015). A full range of these various techniques have been comprehensively reviewed in Golkarian et al. (2018), Rahmati et al. (2018), Golkarian et al. (2018), and Falah et al. (2017). Even though these new ensemble techniques have been successfully applied for studying landslides (Shirzadi et al. 2017a; Shirzadi et al. 2017b), they have rarely been assessed in groundwater studies. Thus, the main objective of this study was to propose a novel classifier ensemble method, which is a hybrid intelligence approach of two state-of-the-art machine learning methods, namely Random Subspace Ensemble based on Random Forest classifier (RF-RS) for groundwater potential mapping in Qorveh-Dehgolan plain, Kurdistan province, Iran. Applying this approach in the study area, where the area is a part of semi-arid areas of Iran, can be considered as a new attempt for identifying groundwater potential areas.

## 2 Study Area and Geological Setting

The Qorveh–Dehgolan plain was selected for this study, since the area has experienced a dramatic average decline of 85.5 m in its groundwater level during the last two decades (Rahmati 2013). The aquifer of Qorveh–Dehgolan plain is located in the southeastern part of Kurdistan province (Iran) (Fig. 1). It lies between 47°10' E and 48°8' E longitudes and 34°55' N and 35°25' N latitudes, covering an area of about 890.3 Km<sup>2</sup>. The elevation of the study area ranges from 1700 m to 2800 m. The average temperature ranges from 10 to 13 °C and the average annual precipitation is 345 mm. The study area is part of Sanandaj–Sirjan geological structural zone in Iran, which its 87.2% is covered by the quaternary deposits (Alavi 1994). The most common land uses within the study area are dry-farming agriculture (37.4%) and irrigated agriculture (22.7%). Other types of land uses, such as pastures (19.2%), barren lands (16.9%), residential areas, and gardens are also present in the study area. The plain uses two sources of water supplies; surface water and groundwater. Most of the surface water is used for irrigation purposes; while, groundwater in the study area is intensively used for domestic purposes as well as agricultural production.

## 3 Data and Methodology

### 3.1 Groundwater Inventory

Groundwater data, including piezometric well locations, number of wells, discharge, well diameter, and groundwater level and depth, were collected from the Kurdistan Regional Water Authority (KRWA). A total of 47 groundwater piezometric wells data were identified in the study area. The data were randomly divided into two parts. One part included 70% of the total

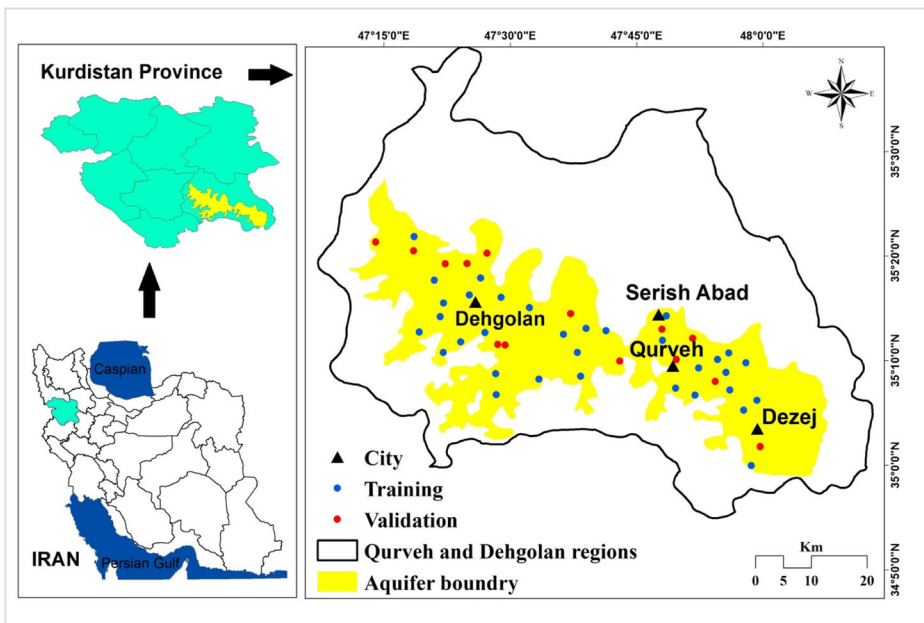
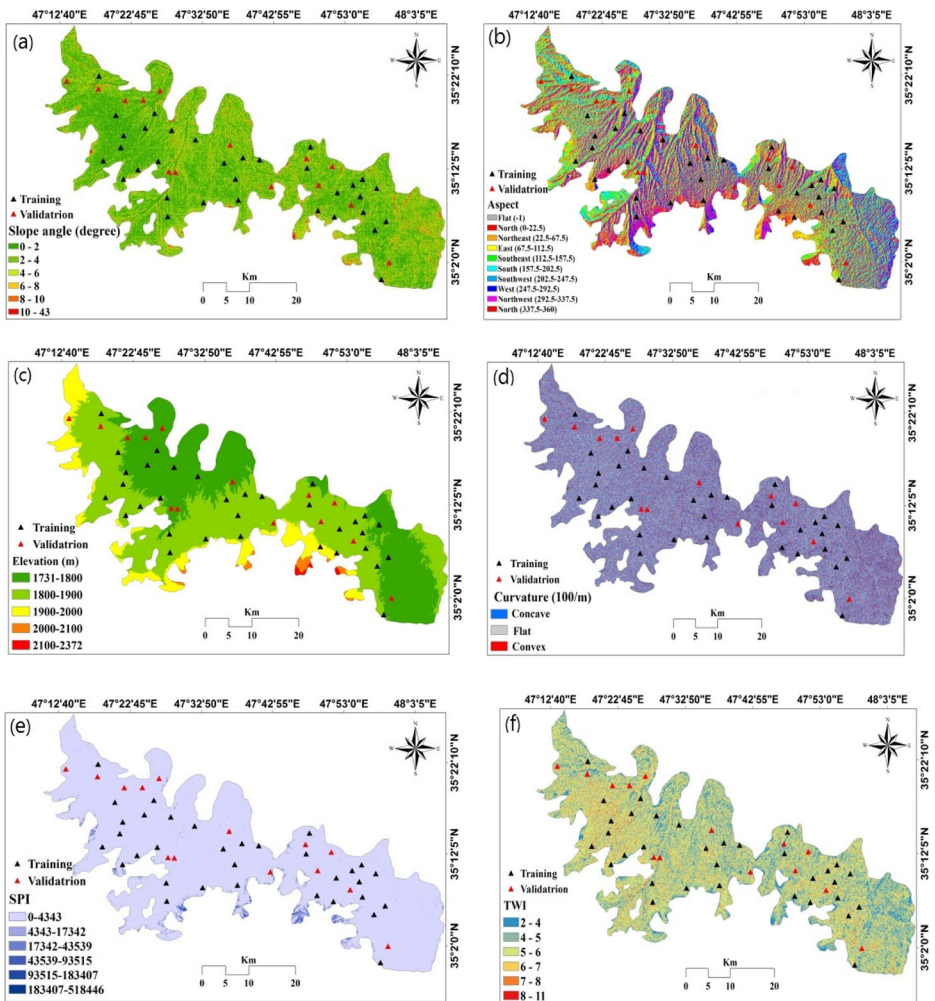


Fig. 1 Study area on Iran and Kurdistan province maps

piezometric wells data (33 groundwater wells) which were then used for generating the training dataset. The other part consisted of the remaining 30% piezometric wells data (14 groundwater wells) utilized later for generating the validation dataset. Additionally, as groundwater modeling was concerned, groundwater potential locations were considered as a binary classification issue. Therefore, a total of 47 non-groundwater piezometric wells data (the locations where there was no groundwater well) were also extracted from the study area for generating training dataset (33 locations), and testing dataset (14 locations).

### 3.2 Geo-environmental Factors Affecting Groundwater Potential

**Slope angle** is a land characteristic which can assist the identification of groundwater conditions through controlling infiltration (Al Saud 2010). In the present study, slope angle



**Fig. 2** Groundwater conditioning factors: **a** Slope angle, **b** Slope aspect **c** Elevation, **d** Curvature, **e** SPI, **f** TWI, **g** Rainfall, **h** River density, **i** Land use, **j** Lithology, **k** Fault density, and **l** NDVI

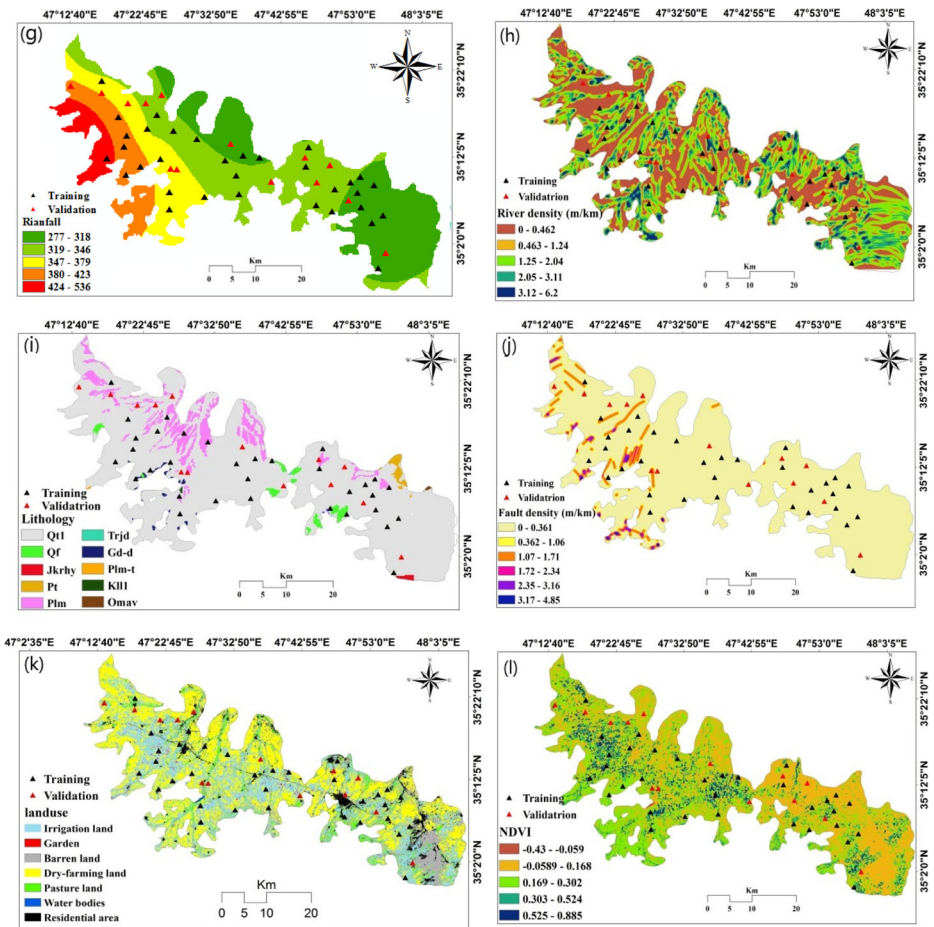


Fig. 2 (continued)

varied from 0 to 43 degrees which was then divided into 4 categories to prepare a slope map (Fig. 2a).

**Slope aspect** indirectly impacts groundwater investigation and is divided into nine categories to prepare its map (Fig. 2b).

**Elevation** is known as terrain ruggedness, playing the same role as slope angle such that the higher the elevation is, the lower the infiltration and recharge will be (Manap et al. 2014). Elevation in the study area ranged from 1731 m to 2372 m. The elevation map was constructed with seven categories as shown in (Fig. 2c).

**Curvature** indirectly influences groundwater recharge (Dar et al. 2010; Oh et al. 2011). It varied from -11.25 to 12.5 (m/100 m) which was then divided into three categories to construct the curvature map (Fig. 2d).

Extracted from Digital Elevation Model (DEM), **SPI** can affect groundwater recharge through the concept of variable source area (Chapi et al. 2015; Naghibi et al. 2015; Nampak et al. 2014). The SPI ranged from 0 to 518,446 which was divided into 6 categories for



preparing the SPI map of the study area (Fig. 2e). SPI was computed as follow (Moore and Wilson 1992):

$$SPI = A_s \tan \beta \quad (1)$$

where  $A_s$  is the specific basin area, and  $\beta$  is the local slope gradient (in degree).

**TWI** positively affects groundwater potential mapping (Oh et al. 2011) through the effect of topography on the location and magnitude of saturated source areas of runoff generation. Equation (2) proposed by (Moore et al. 1991) was used for TWI computation:

$$TWI = Ln(A_s / \tan \beta) \quad (2)$$

where  $A_s$  is the specific basin area, and  $\beta$  is the local slope gradient (in degree). TWI, ranging from 2 to 11, was used to generate the TWI map with 8 categories (Fig. 2f).

**Rainfall** is a hydrologic process for recharging aquifers as it increases, groundwater potentiality correspondingly increases (Oikonomidis et al. 2015). The rainfall map of the study area was generated by virtue of five classes of rainfall ranging between 277 mm and 536 mm, averaged from 30 years data of 10 weather stations (Fig. 2g).

**River density** represents a watershed drainage condition (Oikonomidis et al. 2015) though which affecting groundwater recharge (Oh et al. 2011). River density values ranged from 0 to 1.82 which was then divided into four categories to construct the river density map (Fig. 2h).

**Lithology** controls soil porosity and water permeability (Chowdhury et al. 2010) which they, in turn, affects the specific storage of groundwater. The lithology map (Fig. 2i) was extracted from the geological map at a 1:100,000 scale obtained from the Geological Survey & Mineral Exploitation of Iran (GSI).

**Fault density** is a form of lineament which affects the storage and movement of groundwater (Devi et al. 2001; Nampak et al. 2014). The fault density map was produced from the geological map of Sanandaj with a 1:100000 scale. The fault density value in the study area change from 0 to 0.63, divided into six categories for the map (Fig. 2j).

**Land use/cover** directly influences infiltration and surface runoff (Dinesh Kumar et al. 2007). Land use was generated from OLI sensor images from Landsat 8 satellite using the supervised Maximum Likelihood (MLC) model in ENVI 5.1. Seven land use layers were recognized, including irrigated-farming lands, gardens, barren lands, dry-farming lands, pastures, water bodies, and residential areas (Fig. 2k).

**Normalized Difference Vegetation Index (NDVI)** investigates long-term changes in vegetated areas (Fu and Burgher 2015). The changes in groundwater levels (Aguilar et al. 2012) and groundwater flow discharge (Petus et al. 2012) can be correlated to the NDVI. The NDVI map was prepared through OLI sensor images from Landsat 8 in ENVI5.1. The NDVI values varied between -0.45 and 0.88 which were then divided into five categories (Fig. 2l).

### 3.3 Feature Selection Method of Least Square Support Vector Machine (LSSVM)

The aim of feature selection is to remove irrelevant factors in order to increase the generalization performance of a given learning algorithm (Guyon and Elisseeff 2003). To achieve more accurate prediction of a groundwater potential map, not only selecting a model but also the quality of conditioning factors is important (Pradhan 2013). It is possible that in the modeling process, some of the conditioning factors may play ineffective roles, due to the effect of noise on the predictive ability. Therefore, recognizing and removing conditioning factors

with low or null predictive ability is one of the most significant stages before conducting the learning process (Bui et al. 2015). For this purpose, there are some techniques, including Fuzzy-Rough sets (Dubois and Prade 1990), Relief (Kononenko 1994), Information Gain Ratio (Quinlan 1996), and Least Square Support Vector Machine (LSSVM) (Suykens et al. 2002). In this study, feature selection was carried out using the LSSVM method, a standard SVM technique which has been modified (Tang et al. 2005). Consider a training dataset of  $n$  training sample pairs  $(x_i, y_i)$ ,  $i = 1, 2, \dots, n$ , where  $x_i \in \mathbb{R}$  the  $i^{\text{th}}$  training sample is and  $y_i = (\pm 1)$  is the class label of well (+1) and non-well (-1). LSSVM was computed using a mapping function  $\Phi$  to map the input data into regeneration kernel Hilbert space (defined by kernel function;  $k(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$ ). The general framework of the LSSVM can be expressed as follows:

$$\min J(w, e) = \frac{1}{2} w^T w + \frac{1}{2} \sum_{i=1}^n \gamma e_i^2 \quad (3)$$

$$y_i = [w \cdot \Phi(x_i) + b] = 1 - e_i \quad (4)$$

$$w = \sum_{i=1}^n \alpha_i y_i \Phi(x_i) \quad (5)$$

$$f(x) = \text{sign} \left[ \sum_{i=1}^n \alpha_i y_i k(x, x_i) + b \right] \quad (6)$$

$$f(x) = \text{sign} (w^T \alpha + b) \quad (7)$$

where  $e_i$  is the regression error,  $\gamma$  is a positive constant,  $w^T$  is the inverse matrix of weight matrix assigned to each groundwater well conditioning factor,  $a = (a_1, a_2, \dots, a_{11})$  is the vector of inputs that contains eleven groundwater well conditioning factors, and  $b$  is offset from the origin of the hyper-plane.

### 3.4 Accuracy Assessment and Comparison

#### 3.4.1 Statistical Index

To assess the reliability of model prediction, validation is a significant phase (Chung and Fabbri 1993; Nampak et al. 2014). For achieving this target, some metric predictions are usually used, including sensitivity, specificity, accuracy, kappa, and area under the receiver operating characteristic (AUROC) curve. The concepts and definitions of these metric predictions have been fully reviewed in Rahmati et al. (2018). The outcome of a modeling in the machine learning techniques is a cross-table or confusion matrix involving four types of possible outcomes which are true positive (TP), true negative (TN), false positive (FP), and false negative (FN) (Althuwaynee et al. 2014). The abovementioned criteria are obtained as:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (8)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (9)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

$$\text{Kappa index (K)} = \frac{P_C - P_{\text{exp}}}{1 - P_{\text{exp}}} \quad (11)$$

$$P_C = (TP + TN) / (TP + TN + FN + FP) \quad (12)$$

$$P_{\text{exp}} = \left( (TP + FN)(TP + FP) + (FP + TN)(FN + TN) / \sqrt{(TP + TN + FN + FP)} \right) \quad (13)$$

$$\text{RMSE} = \sqrt{\left[ \left( \frac{1}{n} \right) \sum_{i=1}^n (X_{\text{obs}} - X_{\text{est}})^2 \right]} \quad (14)$$

where  $X_{\text{obs}}$  is the value observed;  $X_{\text{est}}$  is the value estimated using the four groundwater potential models.  $P_C$  is the proportion of number of pixels that are correctly classified as groundwater well or non-groundwater well and is computed as  $(TP / TN)$ /total number of pixels.  $P_{\text{exp}}$  is the expected agreement and is calculated as  $((TP + FN) + (TP + FP) + (FP + TN) + (FN + TN)) / \text{Sqrt}$  (total number of training pixels).

### 3.4.2 Receiver Operating Characteristic Curve (ROC)

The ROC curve is an off use technique to identify the reliability and quality of deterministic and probabilistic models (Swets 1988) and also for the validation of accuracy of groundwater potential map (Pradhan 2013). In the ROC curve, the sensitivity of a model is plotted against 1-specificity. The ROC curve can be plotted according to the existing events (groundwater wells) defined as success rate curve (SRC). Additionally, it would be defined as a prediction rate curve (PRC) for events that will possibly occur in the future.

What is remarkable in the ROC excavation is the area under the ROC curve (AUROC). AUROC is the best mensuration component in the ROC analysis (Simpson and Fitter 1973). It has been considered as a standard and general indicator to perform a test (Walter 2002). It ranges between 0 and 1, so that if AUROC is closer to 1, the ability of prediction accuracy of the model increases; therefore, a perfect forecast gives an area of 1 (Centor and Keightley 1989). Overall, if the AUROC value is greater than 0.8, the performance of the groundwater model is appropriate. AUROC was computed as:

$$\text{AUC} = \sum TP + \sum TN / P + N \quad (15)$$

(Yesilnacar 2005) declared that if the value of AUROC ranges 0.5–0.6, 0.6–0.7, 0.7–0.8, 0.8–0.9, and 0.9–1, the prediction accuracy of a quantitative–qualitative model is poor, average, good, very good, and excellent, respectively.



### 3.4.3 Statistical Analysis

Since a new model is introduced, it is first better to evaluate it based on parametric and non-parametric statistical procedures (Bui et al. 2012). In this study, Friedman test (Friedman 1937) as a non-parametric statistical test was considered to validate a new model and to compare it with two or more models (Beasley and Zumbo 2003). It is first assumed that there is no difference between the performances of the models at the significance level of 0.05. This hypothesis is either rejected or accepted when  $p$  (value) of the test is less or more than 0.05, respectively (Bui et al. 2015). Of course, if the significance level of this test for all models is less than 0.05, the results cannot be interpreted. In this case, Friedman test is not a proper approach to compare models (Bui et al. 2015). Therefore, the models must be compared as pairwise. Eventually, Wilcoxon signed-rank test is applied to statistically assess differences between the models. The null hypothesis is similar to Friedman test. Moreover, the performance of groundwater potential models is different when  $\text{Sig} < 0.05$  and  $z$ -value is more than the critical values of  $z$  ( $-1.96$  and  $+1.96$ ) (Bui et al. 2015).

## 4 Background of Methods Used

### 4.1 Logistic Regression (LR)

Logistic regression which is a statistical model for creating a relationship between dependent and independent variables (Shahabi et al. 2015), was used in this study for the prediction of presence or absence of groundwater in relation to a set of conditioning factors (Nampak et al. 2014). Logistic regression model can give a relationship between the logistic function  $f(z)$  and the potential groundwater which can be expressed as:

$$P = \frac{1}{1 + e^{-Z}} \quad (16)$$

$Z$  factor can be computed using the following equation:

$$Z = \log \text{it} (p) = \ln \left( \frac{p}{1-p} \right) = b_0 + b_1x_1 + \dots + b_nx_n \quad (17)$$

where  $P$  is the probability of an event (potential groundwater) occurrence,  $Z$  is the linear logistic factor that is erratic from  $-\infty$  to  $+\infty$ . It ranges between 0 and 1 such that the closer to 1, the higher the probability of occurrence of the event will be and vice versa.  $b_0$  is the constant coefficient of the model,  $n$  is the number of independent variables,  $b_i$  ( $i = 1, 2, 3, \dots, n$ ) is coefficient of the logistic regression model, and  $x_i$  ( $i = 1, 2, 3, \dots, n$ ) is the independent variable or conditioning factor.

### 4.2 Random Forest (RF) Classifier

Random Forest (RF) is a powerful machine learning algorithm for classification (Rodriguez-Galiano et al. 2015). RF was first created by Ho (1998), and then expanded by Breiman in 2001 (Breiman 2001; Ho 1998; Miao and Wang 2015; Micheletti et al. 2014). The RF model is a compound of many decision trees in which each tree is a set of multiple bootstrap samples

constructed by original samples called bagging (Miao and Wang 2015) creating numerous values of training data by randomly resampling the original dataset with replacement (Rodriguez-Galiano et al. 2015). Bagging develops a variety of trees by performing different training data subsets to avoid the correlation between different trees (Rodriguez-Galiano et al. 2015). Additionally, about one-third of all samples are determined as out-of-bag (OOB) error and they are used for validation of data set (Rahmati et al. 2016). The OOB error is an unbiased estimate of the generalization error that gives an estimate of important variables (Micheletti et al. 2014; Rahmati et al. 2016). Moreover, it can obtain the variance and covariance between grid cells (Kuhnert et al. 2010). Each individual tree among the forest will be constructed on a bootstrap sample in which at each node a subset of features is selected. After splitting nodes according to the Gini Impurity (Criminisi and Shotton 2013), the last node is converted to leaves with a number of samples. For tree of each class, the posterior probabilities ( $P_t(C|f)$ ) are then calculated as:

$$P_t(C|f_j) = \frac{1}{T} \sum_{t=1}^T P_t(C|f) \tag{18}$$

The posterior probability ( $P_t(C|f)$ ) is the probability that a selected case belongs to class  $C_i$  of the training dataset  $f$ . There are some advantages of random forest classifier, such as (i) low computational cost, (ii) powerful performance in large data, (iii) performing many input features without feature elimination, (iv) determining the most important variables in classification, (v) identifying the relationship between variables and classification, and (vi) avoiding over-fitting (Rahmati et al. 2015).

### 4.3 Naïve Bayes (NB) Classifier

One of the main goals of decision tree is to create a suitable model tree for describing the relationship between predictive and class variables (Wang et al. 2006). Naïve Bayesian (NB) classifier is a probabilistic graphical model for classification (Farid et al. 2014). It works based on the Bayes theorem which, in turn, is based on independent distribution and discretization of continuous attribute values by constructing a probability curve for each class in the dataset (Farid et al. 2014). NB has some advantages such as it is fast to train and classify, very simple and easy to understand, very strong to irrelevant features, and requires a small amount of training dataset for classification. Therefore, this model helps predict the future results based on the probabilities of observations from past observations and finding the state of query feature among other variables in the dataset (Ho 1998). There are some steps for performing the NB model. The first step is the collection of data, followed by estimating the probability and mean for each class, creating variance and covariance matrix, and constructing the discriminant function for each class (Ho 1998; Pham et al. 2015).

If  $x = (x_1, x_2, \dots, x_n)$  is the twelve vector of the conditioning factors and  $y = (y_1, y_2)$  is the vector of the classifier variables (well, non-well), the NB classifier will be computed as follow:

$$y_{NB} = \operatorname{argmax} P(y_i) \prod_{i=1}^{12} P(x_i, y_i) \tag{19}$$

$$y_i = [well, non-well]$$

where  $P(y_i)$  is the prior probability of  $y_i$  which can be estimated based on the proportion of the observed cases with output class  $y_i$  in the training dataset, and  $P(x_i, y_i)$  is the conditional

probability which can be calculated as:

$$P(x_i, y_i) = \frac{1}{\sqrt{2\pi}\alpha} e^{-\frac{(x_i - \eta)^2}{2\alpha^2}} \quad (20)$$

where  $\eta$  and  $\alpha$  are the mean and standard deviation, respectively.

#### 4.4 Random Subspace (RS)

Random Subspace is an ensemble learning technique first developed by Ho (1998). It is known as an efficient ensemble technique in which multiple classifiers are combined, trained, and performed on the modified feature space (Pham et al. 2017). It generates sub-training for training base classifiers. As an advantage, different samples on feature space are applied instead of the instance space (Skurichina and Duin 2002).

Considering each training object  $X_i (i = 1, 2, \dots, n)$  in the training dataset,  $X = (X_1, X_2, \dots, X_n)$  is a  $p$ -dimensional vector,  $X_i = (X_{i1}, X_{i2}, \dots, X_{ip})$ . In this method, from the  $p$ -dimensional dataset  $X$ , one randomly chooses  $r < p$ . In this way,  $r$ -dimensional RS from the  $p$ -dimensional feature space can be obtained. The modified training dataset  $\tilde{X} = (\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n)$  comprises of  $r$ -dimensional training objects  $\tilde{X}_i = (\tilde{X}_{i1}, \tilde{X}_{i2}, \dots, \tilde{X}_{ir})$ , ( $i = 1, 2, \dots, n$ ) is achieved where  $r$ -components  $\tilde{x}_{ij} (j = 1, 2, \dots, r)$  are randomly selected from  $p$ -components  $x_{ij} (j = 1, 2, \dots, p)$  of the training vector,  $X_i$ . Classifiers are ultimately constructed in the random subspaces  $\tilde{X}$  and combined by the majority voting using:

$$\beta(x) = \operatorname{argmax}_y \sum_b \delta_{b,y} \operatorname{sgn}(C^b(x)), y \in \{-1, 1\} \quad (21)$$

where  $\delta_{i,j}$  is the Kronecker symbol, and  $y \in \{-1, 1\}$  is a decision (well and non-well) of the classifier.

#### 4.5 Novel Hybrid Approach of RS-RF for Groundwater Potential Assessment

This paper attempts to introduce a novel classifier ensemble method, namely random subspace based on Random Forest (RS-RF), in order to enhance the prediction accuracy of a base classifier and groundwater potential mapping. The main aim of ensemble modeling is to build an effective method by integrating multiple outputs from a set of models (Rokach 2005). Therefore, an ensemble model can make the decision easier with further increase in accuracy and reliability. The framework of this technique is shown briefly in Fig. 3. It can be constructed by five main steps, (i) data collection and interpretation, (ii) dataset preparation, (iii) random subspace (RS) ensemble construction, (iv) random forest (RF) algorithm construction, and (v) RS-RF model construction.

- i. Data collection and interpretation: data has been collected from various sources including Google Earth images; available thematic maps, meteorological data, and groundwater location reports.
- ii. Dataset preparation: 70% (33 locations) and 30% (15 locations) of groundwater well locations were used to generate training and validation datasets, respectively. In addition to groundwater well locations, non-groundwater well locations were randomly considered to construct datasets. The groundwater and non-groundwater well locations were converted to pixels (20 × 20 m) to overlay with conditioning factors to construct the final dataset.

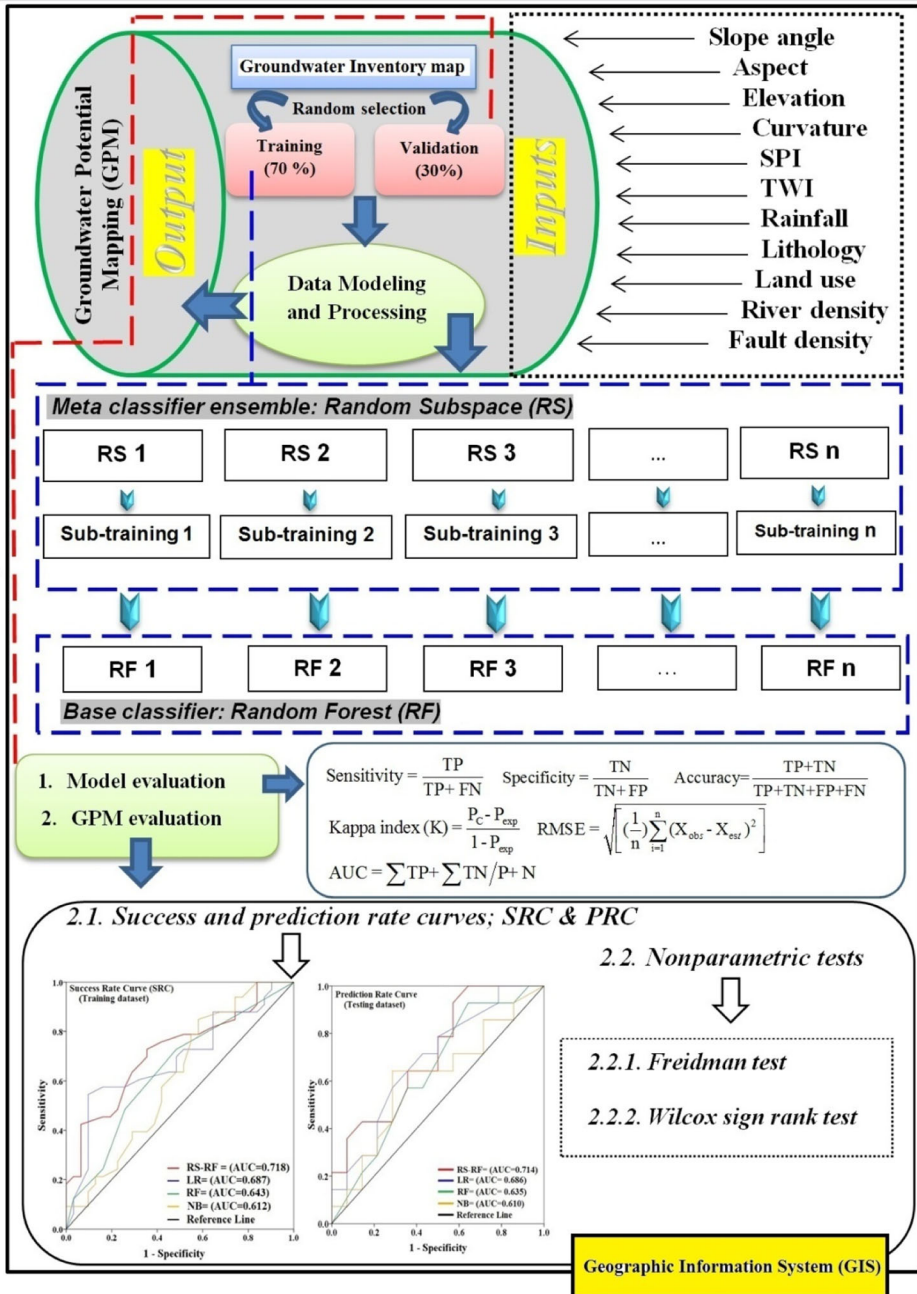


Fig. 3 Novel classifier ensemble of random subspace based on random forest (RS-RF) model framework used in this study

iii. Meta classifier ensemble: Random subspace ensemble was constructed for each sub-training dataset. Optimal sub-training was generated after training the random subspace ensemble. These sub-training datasets were then used to train a base classifier, namely

- random forest (RF). Finally, the RS-RF model was constructed using a combination of all sub-training based on the RS.
- iv. Base classifier: Random forest (RF) algorithm was used as a base classifier for each sub-training dataset to generate groundwater well potential mapping.
  - v. The proposed model is a novel classifier ensemble method constructed by a combination of random subspace ensemble and random forest algorithm classifier (RS-RF).

## 5 Results and Discussion

### 5.1 Selection of the Most Significant Groundwater Conditioning Factors

The selection of the most important affecting factors for groundwater potential mapping (GWPM) using the LSSVM method are shown in Fig. 4. These results show that factors with higher weights were more important to groundwater models than others. It could be observed that TWI had the highest predictive capability for groundwater models (Average Merit (AM)=10.89). River density factor was ranked as the second among the affecting factors with AM=10.2, and also slope angle had a remarkable contribution to groundwater models (AM=7.32). Other factors, such as curvature (AM=6.32), elevation (AM=6.12), lithology (AM=6.02), rainfall (AM=5.86), and NDVI (AM=5.63) held almost the same predictive capability for groundwater models. Land use (AM=5.02), aspect (AM=4.32), and SPI (AM=3.2) showed low predictive capabilities, and fault density (AM=2.1) had the lowest contribution to groundwater models. These results imply that since all twelve groundwater affecting factors had AM >0, they were considered for building groundwater potential models in the present study.

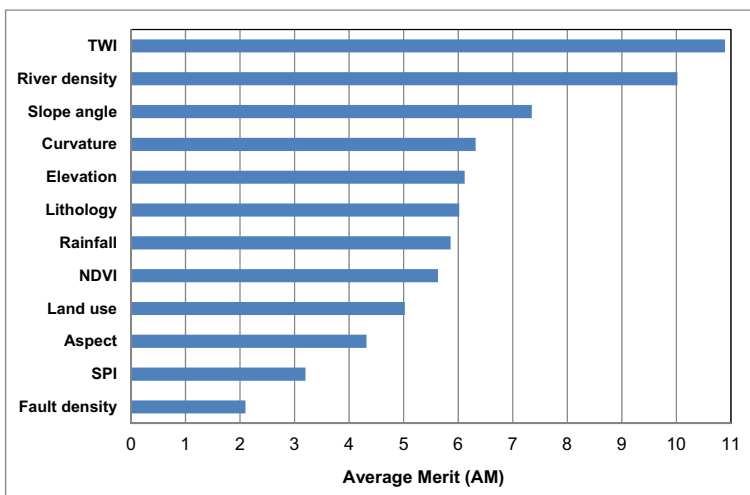


Fig. 4 Prediction capability of the twelve groundwater conditioning factors

**Table 1** Performance of the RS-RF model in training and validation datasets

Parameter	Training	Validation
True positive	31	13.00
True negative	32	12.00
False positive	3	1.00
False negative	2	2.00
Sensitivity (%)	93.9	86.7
Specificity (%)	91.4	92.3
Accuracy (%)	92.6	89.3
Kappa	0.792	0.689
RMSE	0.380	0.397
AUROC	0.995	0.878

## 5.2 Training the RS-RF Model and Validation

The four groundwater potential models were evaluated by criteria that were extracted from the confusion matrix, including sensitivity, specificity, accuracy, kappa, RMSE and AUROC. Table 1 shows the performance of RS-RF model for training and validation datasets. The results, according to the value of sensitivity for training (93.9%) and validation (86.7%) datasets, indicated that the new hybrid model was acceptable; while, the values of this criterion were, 90.9%, 87.9 and 69.7%, respectively, for the RF, LR and NB models in the training dataset and were 80%, 81.3 and 75%, respectively, in the validation dataset (Table 2).

Table 2 shows that in training and validation datasets, the specificity of the new model had the values of 91.4 and 92.3%, respectively. This index was 87.9% for RF, 87.9% for LR, and 90.9% for NB in the training dataset; while, in the validation dataset, it was 84.6%, 84.6 and 83.3%, respectively. The RS-RF had the highest accuracy in training (92.6%) and validation (89.3%) datasets, indicating that 92.6 and 89.3% of groundwater well and non-groundwater well pixels had been correctly classified, followed by RF (89.4%), LR (87.9%) and NB (80.3%) models for training and 82.1%, 82.8%, and 78.6% for validation datasets, respectively (Table 2).

The results indicated that the RS-RF model in training (0.792) and validation (0.689) datasets had the highest value of kappa index, followed by RF (0.754), LR (0.754) and NB

**Table 2** Performance of RF, LR, and NB groundwater models

Parameter	RF		LR		NB	
	Training	Validation	Training	Validation	Training	Validation
True positive	30	12	29	13	23	12
True negative	29	11	29	11	30	10
False positive	4	2	4	1	3	2
False negative	3	3	4	3	10	4
Sensitivity (%)	90.9	80.0	87.9	81.3	69.7	75.0
Specificity (%)	87.9	84.6	87.9	84.6	90.9	83.3
Accuracy (%)	89.4	82.1	87.9	82.8	80.3	78.6
Kappa	0.754	0.658	0.754	0.678	0.702	0.632
RMSE	0.410	0.439	0.418	0.449	0.486	0.466
AUROC	0.929	0.809	0.947	0.825	0.891	0.800



(0.702) in the training dataset and 0.658, 0.678 and 0.632 in the validation dataset, respectively (Table 2). The results of this section revealed that all models showed substantial agreement between observed and predicted locations of groundwater wells.

The RMSE values for training and validation datasets were 0.380 and 0.397, respectively. The RF, LR, and NB models were ranked thereafter, since they obtained 0.410, 0.418, and 0.486 for the training dataset, and 0.439, 0.449 and 0.466 for the validation dataset, respectively, indicating that the introduced model showed a higher performance (Table 2).

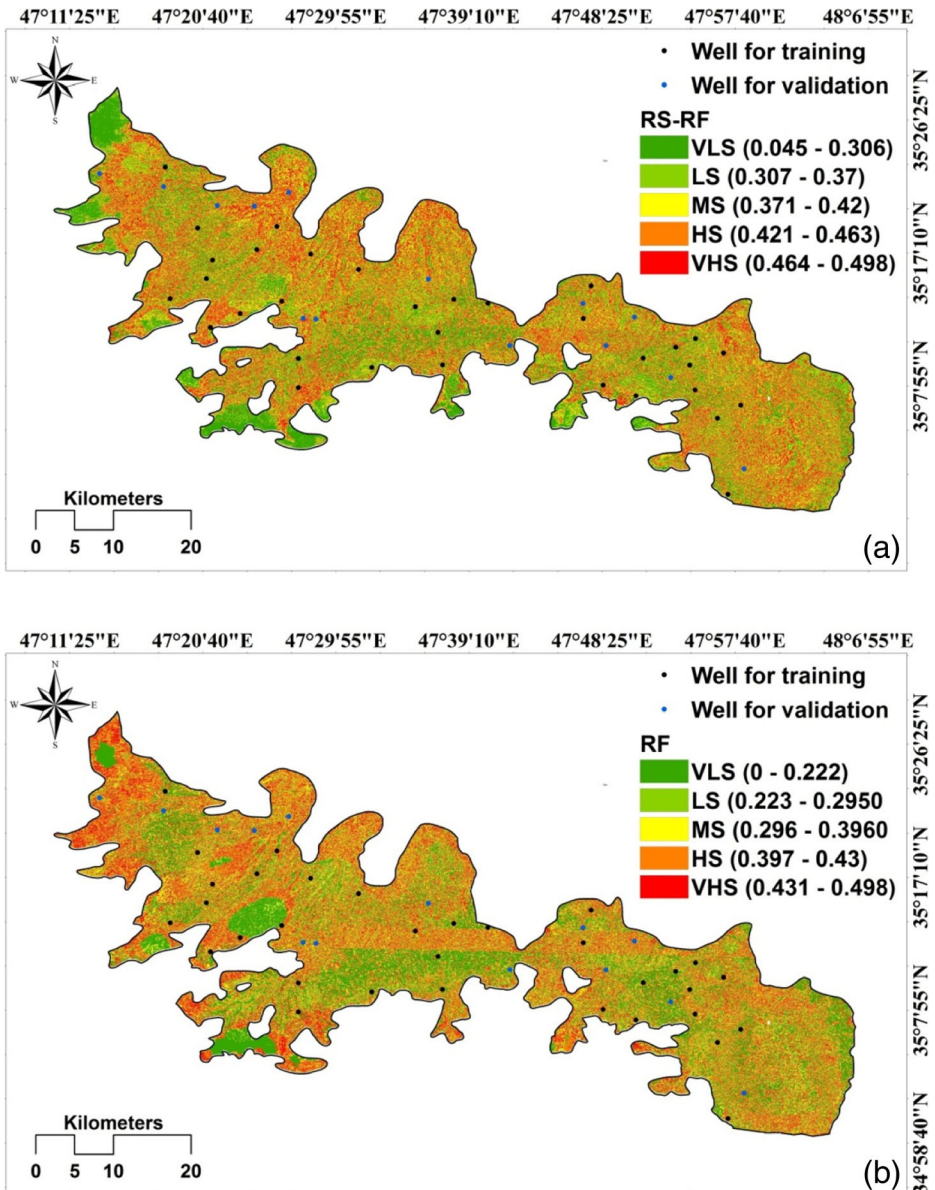


Fig. 5 Groundwater potential maps derived from a RS-RF, b RF, c LR, and d NB

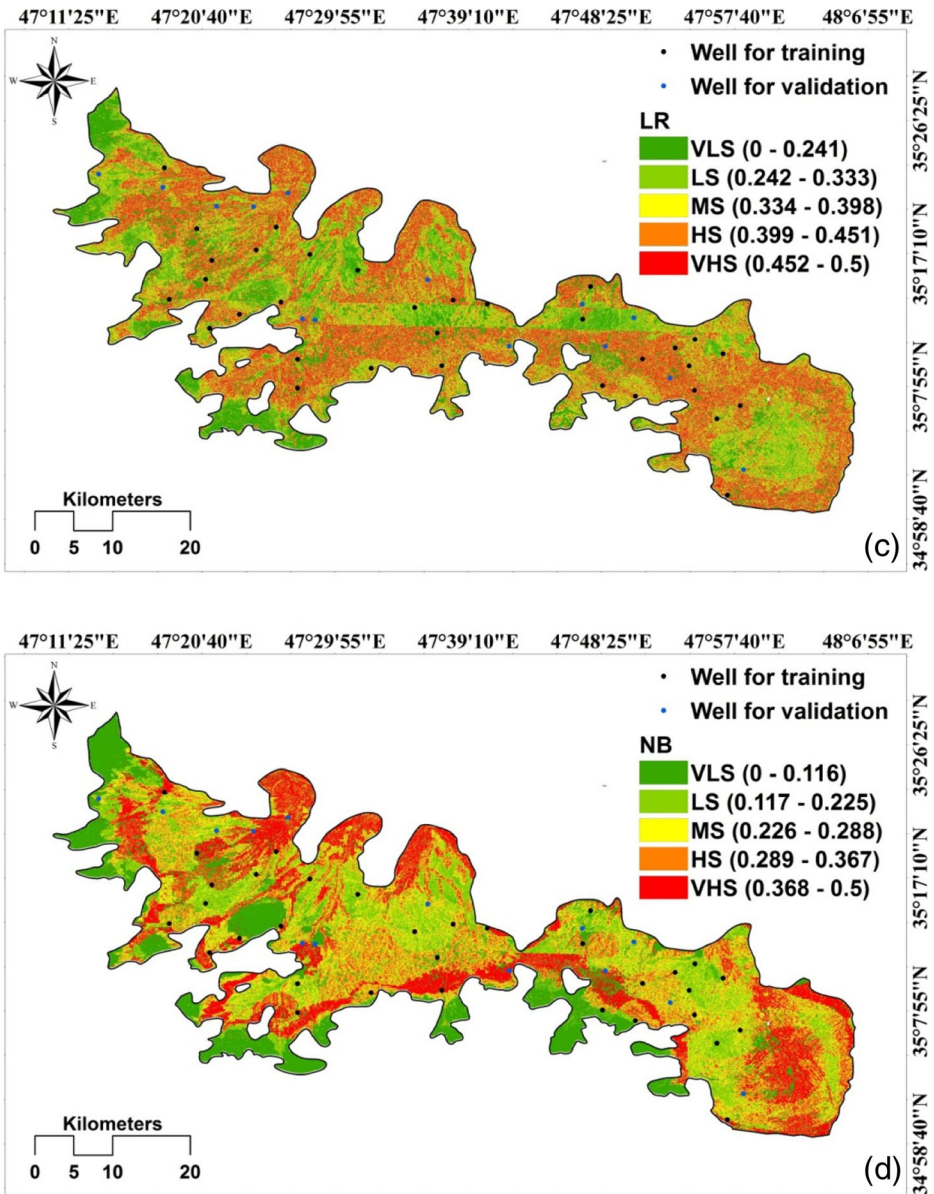


Fig. 5 (continued)

The results demonstrated that AUROC of the new model for the training dataset was 0.995, followed by the RF (0.929), LR (0.947) and NB (0.891) models. However, in the validation dataset AUROC was 0.878, ranked before the RF (0.809), LR (0.825) and NB (0.800) models.

Overall, comparison of models showed that in the training and validation datasets, the new proposed hybrid model (RS-RF) had the highest performance in terms of sensitivity, specificity, accuracy, kappa and AUROC, followed by the LR, the RF and the NB models. Therefore,

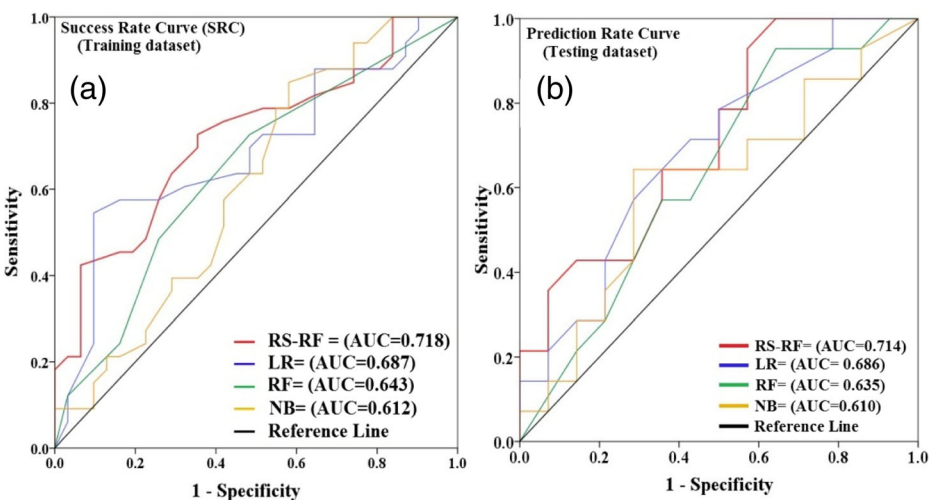
it can be concluded that this new model had a higher capability for preparing the groundwater well potential mapping than the other state-of-the-art benchmark machine learning models used in this study.

### 5.3 Preparing Groundwater Potential Mapping

Preparation of groundwater potential maps is one of the most important key elements of groundwater modeling studies; hence, maps were constructed after training and validation in two main steps. The first was the generation of groundwater potential indices for all pixels of the study area, and the second was the reclassification of these indices according to the natural break method. Groundwater potential classes were reclassified into five categories with respect to susceptible index intervals as very low (0 – 0.106), low (0.106 – 0.298), moderate (0.298 – 0.537), high (0.537 – 0.787), and very high (0.787 – 0.999). All groundwater potential maps are shown in Fig. 5.

It can be observed that the very high class had the largest area (44.6%), followed by high (32.78%), moderate (16.85%), low (5.196%), and very low (0.564%), respectively. Obviously, the groundwater potential map using the RS-RF model indicated very good performance, since highest and lowest numbers of wells were located in areas with very high (29.787) and very low potential (2.127), followed by high (23.404) and low (14.893) classes, respectively.

ROC curves of all groundwater potential models were prepared for each map in both the training dataset (for success rate curve) and the testing dataset (for prediction rate curve). The results of ROC and AUC of the four groundwater potential maps are shown in Fig. 6. The success rate curve varied between 0.718 to 0.612, indicating that all four studied models had suitable prediction capabilities. The highest prediction capability was provided by the RS-RF model with AUROC equal to 0.718 in the training dataset, followed by LR (0.687), RF (0.643), and NB (0.612). It can also be observed that the RS-RF model had the highest AUROC value (0.714) in the testing dataset; while, LR (0.686), RF (0.635), and NB (0.610) were ranked thereafter.



**Fig. 6** Comparison of groundwater well susceptibility models using ROC curve technique, **a** Success rate curve, and **b** Prediction rate curve

**Table 3** Groundwater well density for the new hybrid model (RS-RF) obtained in this study

Class	Intervals	Number of class pixels	Number of well pixels	% Class pixels	% well pixels	well density (W.D)
Very low	0.045 - 0.306	21,152	1	0.564	2.127	0.265
Low	0.306 - 0.370	194,786	7	5.196	14.893	0.349
Moderate	0.370 - 0.420	631,807	14	16.853	29.787	0.566
High	0.420 - 0.463	1,229,093	11	32.785	23.404	1.406
Very high	0.463 - 0.498	1,672,028	14	44.600	29.787	1.497

These outcomes also pointed out that RS-RF had the highest predictive capability among other groundwater potential models in both success and prediction rate curves. In addition to the AUROC, the well density (WD) was applied to the new hybrid model to more appropriately evaluate groundwater potential mapping (Table 3). The WD was obtained by overlaying groundwater well potential maps with the location of wells. The amount of WD significantly increased from “very low” class (WD = 0.265) to “very high” class (WD = 1.497).

The four groundwater models were finally assessed using Friedman test at the 5% significant level to answer the question whether or not there was a statistically significant difference among them. The results showed that the null hypothesis was rejected because the Sig. factor was less than 0.05. Even though Friedman test was able to determine significant differences among models, it was unable to recognize which model made this difference. Thus, to better depict the systematic pairwise differences, Wilcoxon signed-rank test was used (Table 4). The Sig < 0.05 and Z-value higher than the critical value (−1.96 and + 1.96) revealed that the new hybrid model (RF-RS) had a statistical significant difference with other groundwater well models. Overall, the RS-RF model outperformed the RF, LR, and NB models for purposes of this study.

## 6 Conclusion

The main objective of this study was to present a new artificial intelligent approach which is a hybrid of random subspace and random forest (RS-RF) for mapping groundwater potential. Twelve conditioning factors were selected for groundwater analysis of the aquifer of Qorveh–Dehgolan plain in the west of Iran based on LSSVM. Proper selection of these twelve factors increased the accuracy of groundwater well mapping by reducing noise and over-fitting for the training dataset. This implies that the selected factors are very suitable for groundwater potential mapping in the study area which might be appropriate in similar areas as well. To state the efficiency of the new hybrid model, three state-of-the-art soft computing benchmark

**Table 4** Performance of the RS-RF model compared to other groundwater susceptibility models using Wilcoxon signed-rank test (two-tailed)

NO	Pair wise comparison	Number of positive differences	Number of negative differences	z-value	p value	Significance level (Sig.)
1	RS-RF vs. RF	20	43	- 2.664	0.008	Yes
2	RS-RF vs. LR	43	21	- 3.248	0.001	Yes
3	RS-RF vs. NB	19	39	- 2.560	0.010	Yes

The standard p value is 0.05

models, including random forest (RF), logistic regression (LR) and naïve bayes (NB), were utilized for comparison. These comparisons were made using commonly criteria, including sensitivity, specificity, accuracy, kappa, RMSE and area under the ROC curve. The prediction capability of groundwater potential maps was validated and compared using the success rate and predictive rate curves. This study revealed that random subspace significantly improved the performance of single random forest such that among all 4 models studied for groundwater potential mapping, the new hybrid model performed by far better than LR, RF, and NB.

The RS-RF hybrid model would be a promising technique that can appropriately be used for groundwater potential mapping which means it can accurately recognize the area with higher potential of groundwater at the study area and in similar regions, maybe with caution. This methodology would suggest farmers to find suitable places for digging wells to avoid spending a great deal of money. In addition, the groundwater potential map can be useful to water managers for making proper decisions on the optimal use of groundwater resources for future planning specifically in arid and semi-arid climates such as the study area.

## Compliance with Ethical Standards

**Conflict of Interest** None.

## References

- Aguilar C, Zinnert JC, Polo MJ, Young DR (2012) NDVI as an indicator for changes in water availability to woody vegetation. *Ecol Indic* 23:290–300
- Al Saud M (2010) Mapping potential areas for groundwater storage in Wadi Aurnah Basin, western Arabian peninsula, using remote sensing and geographic information system techniques. *Hydrogeol J* 18: 1481–1495
- Alavi M (1994) Tectonics of the Zagros orogenic belt of Iran: new data and interpretations. *Tectonophysics* 229: 211–238
- Althuwaynee OF, Pradhan B, Park H-J, Lee JH (2014) A novel ensemble decision tree-based CHI-squared automatic interaction detection (CHAID) and multivariate logistic regression models in landslide susceptibility mapping. *Landslides* 11:1063–1078
- Beasley TM, Zumbo BD (2003) Comparison of aligned Friedman rank and parametric methods for testing interactions in split-plot designs. *Comput Stat Data Anal* 42:569–593
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32
- Bui DT, Pradhan B, Lofman O, Revhaug I, Dick OB (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
- Bui DT, Pradhan B, Revhaug I, Nguyen DB, Pham HV, Bui QN (2015) A novel hybrid evidential belief function-based fuzzy logic model in spatial prediction of rainfall-induced shallow landslides in the Lang Son city area (Vietnam). *Geomat Nat Haz Risk* 6:243–271
- Centor R, Keightley G (1989) Receiver Operating Characteristics (ROC) curve area analysis using the ROC ANALYZER. In: *Proceedings/the... Annual Symposium on Computer Application [sic] in Medical Care. Symposium on Computer Applications in Medical Care. American Medical Informatics Association*, p 222–226
- Chapi K, Rudra RP, Ahmed SI, Khan AA, Gharabaghi B, Dickinson WT, Goel PK (2015) Spatial-temporal dynamics of runoff generation areas in a small agricultural watershed in southern Ontario. *J Water Resour Protect* 7:14–40
- Chenini I, Mammou AB, El May M (2010) Groundwater recharge zone mapping using GIS-based multi-criteria analysis: a case study in Central Tunisia (Maknassy Basin). *Water Resour Manag* 24:921–939



- Chowdhury A, Jha MK, Chowdary V (2010) Delineation of groundwater recharge zones and identification of artificial recharge sites in West Medinipur district, West Bengal, using RS, GIS and MCDM techniques. *Environ Earth Sci* 59:1209–1222
- Chung C-JF, Fabbri AG (1993) The representation of geoscience information for data integration. *Nonrenewable Resources* 2:122–139
- Criminisi A, Shotton J (2013) Decision forests for computer vision and medical image analysis. Springer Science & Business Media, Berlin
- Dar IA, Sankar K, Dar MA (2010) Remote sensing technology and geographic information system modeling: an integrated approach towards the mapping of groundwater potential zones in Hardrock terrain, Mamundiyyar basin. *J Hydrol* 394:285–295
- Devi PS, Srinivasulu S, Raju KK (2001) Hydrogeomorphological and groundwater prospects of the Pageru river basin by using remote sensing data. *Environ Geol* 40:1088–1094
- Dinesh Kumar P, Gopinath G, Seralathan P (2007) Application of remote sensing and GIS for the demarcation of groundwater potential zones of a river basin in Kerala, southwest coast of India. *Int J Remote Sens* 28:5583–5601
- Dubois D, Prade H (1990) Rough fuzzy sets and fuzzy rough sets. *Int J Gen Syst* 17:191–209
- Falah F, Ghorbani Nejad S, Rahmati O, Daneshfar M, Zeinivand H (2017) Applicability of generalized additive model in groundwater potential modelling and comparison its performance by bivariate statistical methods. *Geocarto Int* 32:1069–1089
- Farid DM, Zhang L, Rahman CM, Hossain MA, Strachan R (2014) Hybrid decision tree and naïve Bayes classifiers for multi-class classification tasks. *Expert Syst Appl* 41:1937–1946
- Friedman M (1937) The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J Am Stat Assoc* 32:675–701
- Fu B, Burgher I (2015) Riparian vegetation NDVI dynamics and its relationship with climate, surface water and groundwater. *J Arid Environ* 113:59–68
- Golkarian A, Naghibi SA, Kalantar B, Pradhan B (2018) Groundwater potential mapping using C5.0, random forest, and multivariate adaptive regression spline models in GIS. *Environ Monit Assess* 190:149
- Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. *J Mach Learn Res* 3:1157–1182
- Ho TK (1998) The random subspace method for constructing decision forests. *IEEE Trans Pattern Anal Mach Intell* 20:832–844
- Jha MK, Chowdhury A, Chowdary V, Peiffer S (2007) Groundwater management and development by integrated remote sensing and geographic information systems: prospects and constraints. *Water Resour Manag* 21:427–467
- Kononenko I (1994) Estimating attributes: analysis and extensions of RELIEF. In: European conference on machine learning. Springer, p 171–182
- Kuhnert PM, Martin TG, Griffiths SP (2010) A guide to eliciting and using expert knowledge in Bayesian ecological models. *Ecol Lett* 13:900–914
- Lian C, Zeng Z, Yao W, Tang H (2014) Ensemble of extreme learning machine for landslide displacement prediction based on time series analysis. *Neural Comput & Applic* 24:99–107
- Mair A, El-Kadi AI (2013) Logistic regression modeling to assess groundwater vulnerability to contamination in Hawaii, USA. *J Contam Hydrol* 153:1–23
- Manap MA, Sulaiman WNA, Ramli MF, Pradhan B, Surip N (2013) A knowledge-driven GIS modeling technique for groundwater potential mapping at the upper Langat Basin, Malaysia. *Arab J Geosci* 6:1621–1637
- Manap MA, Nampak H, Pradhan B, Lee S, Sulaiman WNA, Ramli MF (2014) Application of probabilistic-based frequency ratio model in groundwater potential mapping using remote sensing data and GIS. *Arab J Geosci* 7:711–724
- Miao T, Wang M (2015) Susceptibility analysis of earthquake-induced landslide using random forest method
- Micheletti N, Foresti L, Robert S, Leuenberger M, Pedrazzini A, Jaboyedoff M, Kanevski M (2014) Machine learning feature selection methods for landslide susceptibility mapping. *Math Geosci* 46:33–57
- Moore ID, Wilson JP (1992) Length-slope factors for the revised universal soil loss equation: simplified method of estimation. *J Soil Water Conserv* 47:423–428
- Moore ID, Grayson R, Ladson A (1991) Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. *Hydrol Process* 5:3–30
- Naghibi SA, Pourghasemi HR, Pourtaghi ZS, Rezaei A (2015) Groundwater qanat potential mapping using frequency ratio and Shannon's entropy models in the Moghan watershed, Iran. *Earth Sci Inf* 8:171–186
- Nampak H, Pradhan B, Manap MA (2014) Application of GIS based data driven evidential belief function model to predict groundwater potential zonation. *J Hydrol* 513:283–300
- Neshat A, Pradhan B, Pirasteh S, Shafri HZM (2014) Estimating groundwater vulnerability to pollution using a modified DRASTIC model in the Kerman agricultural area, Iran. *Environ Earth Sci* 71:3119–3131



- 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. *J Hydrol* 399:158–172
- Oikonomidis D, Dimogianni S, Kazakis N, Voudouris K (2015) A GIS/remote sensing-based methodology for groundwater potentiality assessment in Timavos area, Greece. *J Hydrol* 525:197–208
- Osati K, Koeniger P, Salajegheh A, Mahdavi M, Chapi K, Malekian A (2014) Spatiotemporal patterns of stable isotopes and hydrochemistry in springs and river flow of the upper Karkheh River basin, Iran. *Isot Environ Health Stud* 50:169–183
- Petus C, Lewis M, White D (2012) Using MODIS Normalized Difference Vegetation Index to monitor seasonal and inter-annual dynamics of wetland vegetation in the Great Artesian Basin: a baseline for assessment of future changes in a unique ecosystem. In: International Society for Photogrammetry and Remote Sensing
- Pham BT, Tien Bui D, Indra P, Dholakia M (2015) A comparison study of predictive ability of support vector machines and naive bayes tree methods in landslide susceptibility assessment at an area between Tehri Garhwal and Pauri Garhwal, Uttarakhand state (India) using GIS. In: national symposium on geomatics for digital India and annual conventions of ISG & ISRS, Jaipur (India)
- Pham BT, Bui DT, Prakash I, Dholakia M (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
- Pradhan B (2013) A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Comput Geosci* 51:350–365
- Quinlan JR (1996) Improved use of continuous attributes in C4. 5. *J Artif Intell Res* 4:77–90
- Rahmati O (2013) An investigation of quantitative zonation and groundwater potential (case study: Ghorveh-Dehghan plain). M. Sc. thesis, Tehran University
- Rahmati O, Samani AN, Mahdavi M, Pourghasemi HR, Zeinivand H (2015) Groundwater potential mapping at Kurdistan region of Iran using analytic hierarchy process and GIS. *Arab J Geosci* 8:7059–7071
- Rahmati O, Pourghasemi HR, Melesse AM (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* 137:360–372
- Rahmati O, Naghibi SA, Shahabi H, Bui DT, Pradhan B, Azareh A, Rafiei-Sardooi E, Samani AN, Melesse AM (2018) Groundwater spring potential modelling: comprising the capability and robustness of three different modeling approaches. *J Hydrol* 565:248–261
- Rodriguez-Galiano V, Sanchez-Castillo M, Chica-Olmo M, Chica-Rivas M (2015) Machine learning predictive models for mineral prospectivity: an evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geol Rev* 71:804–818
- Rokach L (2005) Ensemble methods for classifiers. In: Data mining and knowledge discovery handbook. Springer, p 957–980
- Shahabi H, Hashim M, Ahmad BB (2015) Remote sensing and GIS-based landslide susceptibility mapping using frequency ratio, logistic regression, and fuzzy logic methods at the central Zab basin, Iran. *Environ Earth Sci* 73:8647–8668
- Shekhar S, Pandey AC (2015) Delineation of groundwater potential zone in hard rock terrain of India using remote sensing, geographical information system (GIS) and analytic hierarchy process (AHP) techniques. *Geocarto Int* 30:402–421
- Shirzadi A, Chapi K, Shahabi H, Solaimani K, Kavian A, Ahmad BB (2017a) Rock fall susceptibility assessment along a mountainous road: an evaluation of bivariate statistic, analytical hierarchy process and frequency ratio. *Environ Earth Sci* 76:152
- Shirzadi A, Bui DT, Pham BT, Solaimani K, Chapi K, Kavian A, Shahabi H, Revhaug I (2017b) Shallow landslide susceptibility assessment using a novel hybrid intelligence approach. *Environ Earth Sci* 76:60
- Simpson AJ, Fitter MJ (1973) What is the best index of detectability? *Psychol Bull* 80:481–488
- Skurichina M, Duin RP (2002) Bagging, boosting and the random subspace method for linear classifiers. *Pattern Anal Appl* 5:121–135
- Suykens JA, Van Gestel T, De Brabanter J (2002) Least squares support vector machines. World Scientific
- Swets JA (1988) Measuring the accuracy of diagnostic systems. *Science* 240:1285–1293
- Tahmasebipoor 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. *Arab J Geosci* 9:79
- Tang X, Ou Z, Su T, Sun H, Zhao P (2005) Robust precise eye location by adaboost and svm techniques. In: International Symposium on Neural Networks. Springer, p 93–98
- Tehrany MS, Pradhan B, Jebur MN (2013) Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS. *J Hydrol* 504:69–79
- Walter S (2002) Properties of the summary receiver operating characteristic (SROC) curve for diagnostic test data. *Stat Med* 21:1237–1256

- Wang L-M, Li X-L, Cao C-H, Yuan S-M (2006) Combining decision tree and naive Bayes for classification. *Knowl-Based Syst* 19:511–515
- Yesilnacar EK (2005) The application of computational intelligence to landslide susceptibility mapping in Turkey. University of Melbourne, Department, 200

## Affiliations

**Shaghayegh Miraki<sup>1</sup> · Sasan Hedayati Zanganeh<sup>2</sup> · Kamran Chapi<sup>3</sup> · Vijay P. Singh<sup>4</sup> · Ataollah Shirzadi<sup>3</sup> · Himan Shahabi<sup>5</sup> · Binh Thai Pham<sup>6</sup>**

<sup>1</sup> Department of Watershed Sciences Engineering, Faculty of Natural Resources, University of Agricultural Science and Natural Resources of Sari, Sari, Iran

<sup>2</sup> Department of GIS and RS, Faculty of Geography, University of Tabriz, Tabriz, Iran

<sup>3</sup> Department of Rangeland and Watershed Management, Faculty of Natural Resources, University of Kurdistan, Sanandaj, Iran

<sup>4</sup> Department of Biological and Agricultural Engineering, and Zachry Department of Civil Engineering, Texas A & M University, College Station, TX 77843-2117, USA

<sup>5</sup> Department of Geomorphology, Faculty of Natural Resources, University of Kurdistan, Sanandaj, Iran

<sup>6</sup> Department of Geotechnical Engineering, University of Transport Technology, 54 Trieu Khuc, Thanh Xuan, Hanoi, Vietnam