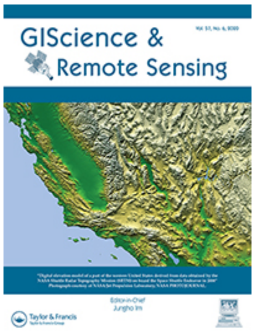

GIScience & Remote Sensing

ISSN: 1548-1603 (Print) 1943-7226 (Online) Journal homepage: www.tandfonline.com/journals/tgrs20

Application of machine learning techniques in groundwater potential mapping along the west coast of India

Pankaj Prasad, Victor Joseph Loveson, Mahender Kotha & Ramanand Yadav

To cite this article: Pankaj Prasad, Victor Joseph Loveson, Mahender Kotha & Ramanand Yadav (2020) Application of machine learning techniques in groundwater potential mapping along the west coast of India, GIScience & Remote Sensing, 57:6, 735-752, DOI: 10.1080/15481603.2020.1794104

To link to this article: https://doi.org/10.1080/15481603.2020.1794104

GISCIENCE & REMOTE SENSING
2020, VOL. 57, NO. 6, 735–752
https://doi.org/10.1080/15481603.2020.1794104

Application of machine learning techniques in groundwater potential mapping along the west coast of India

Pankaj Prasad[a,b], Victor Joseph Loveson[a,c], Mahender Kotha[b] and Ramanand Yadav[a,c]

[a]Geological Oceanography Division, CSIR- National Institute of Oceanography, Dona Paula, Goa, India; [b]School of Earth, Ocean and Atmospheric Sciences, Goa University, Taleigao, Goa, India; [c]Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, India

ABSTRACT

Groundwater potential mapping (GWPM) in the coastal zone is crucial for the planning and development of society and the environment. The current study is aimed to map the groundwater potential zones of Sindhudurg coastal stretch on the west coast of India, using three machine learning models: random forest (RF), boosted regression tree (BRT), and the ensemble of RF and support vector machine (SVM). In order to achieve the objective, 15 groundwater influencing factors including elevation, slope, aspect, slope length (LS), profile curvature, plan curvature, topographical wetness index (TWI), distance from streams, distance from lineaments, lithology, geomorphology, soil, land use, normalized difference vegetation index (NDVI), and rainfall were considered for inter-thematic correlations and overlaid with spring and well occurrences in a spatial database. A total of 165 spring and well locations were identified, which had been divided into two classes: training and validation, at the ratio of 70:30, respectively. The RF, BRT, and RF-SVM ensemble models have been applied to delineate the groundwater potential zones and categorized into five classes, namely very high, high, moderate, low, and very low. RF, BRT, and ensemble model results showed that 33.3%, 35.6%, and 36.8% of the research area had a very high groundwater potential zone. These models were validated with area under the receiver operating characteristics (AUROC) curve. The accuracy of RF (94%) and hybrid model (93.4%) was more efficient than BRT (89.8%) model. In order to further evaluate and validate, four different sites were subsequently chosen, and we obtained similar results, ensuring the validity of the applied models. Additionally, ground-penetrating radar (GPR) technique was applied to predict the groundwater table and validated by measured wells. The mean difference between measured and GPR predicted groundwater table was 14 cm, which reflected the importance of GPR to guide the location of new wells in the study region. The outcomes of the study will help the decision-makers, government agencies, and private sectors for sustainable planning of groundwater in the area. Overall, the present study provides a comprehensive high-precision machine learning and GPR-based groundwater potential mapping.

ARTICLE HISTORY
Received 2 December 2019
Accepted 6 July 2020

KEYWORDS
Groundwater potential; GIS; machine learning; ensemble model; GPR

1. Introduction

Groundwater is the most valuable resource on our planet. It reflects the socio-economic condition and development of an area (Naghibi, Pourghasemi, and Dixon 2016). Since the last century, groundwater has been in high demand for domestic, agricultural, and industrial purposes worldwide (Mogaji, Lim, and Abdullah 2014; Chen et al. 2019). Extensive groundwater extraction has led to a continuous drop of water table (Naghibi, Pourghasemi, and Dixon 2016; Das 2019). Therefore, Groundwater management is necessary for sustainable use of water resources. Groundwater availability and its movement depend on topographical, hydrological, ecological, geological, and atmospheric factors (Oh et al. 2011; Golkarian et al. 2018). As groundwater is a hidden natural resource, the demarcation of groundwater potential zones is essential for planning, management, and sustainable development of an area.

Many studies have been done by researchers on GWPM using different methods. Earlier groundwater mapping was based on field surveys, which was more expensive and time-consuming (Ganapuram et al. 2009; Chen et al. 2018; Das 2019). In the present time, remote sensing (RS), geographic information system (GIS), statistical, and geophysical techniques have been applied to map the groundwater potentiality of a large area in time and cost-effective manner.

In earlier studies, GIS modeling is very successfully applied to identify the groundwater prospect region with a high prediction rate (Das 2019; Chen et al. 2019). The combined use of RS and GIS techniques has been employed in different research for GWPM (Prasad et al. 2008; Oh et al. 2011; Magesh, Chandrasekar, and Soundranayagam 2012; Naghibi, Pourghasemi, and Dixon 2016; Murasingh, Jha, and Adamala 2018; Lee, Hong, and Jung 2018; Golkarian et al. 2018; Chen et al.

CONTACT Pankaj Prasad ppankaj@nio.org
Supplemental data for this article can be accessed here.

2018; Das 2019). Besides, several statistical techniques have been employed along with RS and GIS techniques for GWPM such as frequency ratio (Guru et al., 2017; Oh et al. 2011; Pourtaghi and Pourghasemi 2014; Naghibi et al. 2015; Das 2019), weights of evidence (Lee, Kim, and Oh 2012; Tahmassebipoor et al. 2016; Chen et al. 2018), and logistic regression (Nampak, Pradhan, and Manap 2014; Chen et al. 2018).

In this decade, approaches through machine learning models have continuously been increased for groundwater mapping, such as random forest (Naghibi and Pourghasemi 2015; Rahmati, Pourghasemi, and Melesse 2016; Naghibi, Pourghasemi, and Dixon 2016; Golkarian et al. 2018; Naghibi et al. 2019; Arabameri et al. 2019), boosted regression tree (Naghibi et al. 2015; Naghibi, Pourghasemi, and Dixon 2016; Naghibi et al. 2019; Kordestani et al. 2019), support vector machine (Naghibi, Ahmadi, and Daneshi 2017b; Lee, Hong, and Jung 2018; Naghibi, Pourghasemi, and Abbaspour 2018), C5.0 (Duan et al. 2016; Golkarian et al. 2018), classification and regression tree (Naghibi et al. 2015; Naghibi, Pourghasemi, and Dixon 2016), and artificial neural network (Lee, Hong, and Jung 2018). Such models are also applied in numerous fields, namely landslide susceptibility mapping (Pourghasemi, Moradi, and Aghda 2013; Youssef et al. 2016; Kim et al. 2018), gully erosion susceptibility mapping (Arabameri et al. 2019; Gayen et al. 2019), ecological study (Lek and Guegan 1999; Recknagel 2001; Elith, Leathwick, and Hastie 2008; Crisci, Ghattas, and Perera 2012), flood susceptibility mapping (Tehrany, Pradhan, and Jebur 2014; Tehrany et al. 2015; Khosravi et al. 2018; Shafizadeh-Moghadam et al. 2018), and land use land cover change detection (Friedl, Brodley, and Strahler 1999; Gislason, Benediktsson, and Sveinsson 2006; Rodriguez-Galiano and Chica-Rivas 2014). The prediction rate of machine learning models is very high compared to statistical models as reported in different studies (Naghibi and Pourghasemi 2015; Chen et al. 2018).

The aforementioned studies have used the advantages of a single model. However, the ensemble model is a combination of statistical and machine learning techniques (Naghibi et al. 2017a; Kordestani et al. 2019). Recently, Naghibi et al. (2017a), Kordestani et al. (2019), and Naghibi et al. (2019) have used ensemble models for GWPM with satisfactory accuracy. The present work was applied the ensemble model of RF and SVM for GWPM.

In the previous studies, the researchers used different models, but their applicability was restricted to a specific study region. The objective of the present research is to map the groundwater prospect zones along the Sindhudurg coastal stretch using machine learning techniques and evaluate the results in different regions for the validity of the models. Besides, GPR technology was introduced to identify the groundwater table. Groundwater potential maps of the study region can be helpful for better planning and management of groundwater resources.

2. Study area

Sindhudurg coast stretches from 15°43′11.43″N to 16°33′45.63″N latitude and 73°18′36.53″E to 73°55′50.07″E longitude covering an area around 3177 sq km along the west coast of India. The study region is bounded in the north and south by rivers and in the east and west by Western Ghats and shoreline of the west coast of India (Figure 1a). With these natural boundaries, the area under study represents a typical coastal environment as it is a transition zone between land and sea.

The geology of the study area shows formation from the Archean to the Recent age and the lithology of which mainly consists of granite, basalt, lateritic, and alluvial deposits. The Deccan basalts occupy more than 30% of the study area and the aquifers are mostly associated with fractures and joints. However, the aquifers associated with laterites are substantiated with porous nature. The elevation of the study area varies from 0 to 450 m above mean sea level from seashore to landward dissected hill ranges. From the geomorphological perspective, the study area is reclassified into six distinct regions viz., denudational origin-pediment pediplain complex, coastal origin-younger coastal plain, structural origin medium dissected plateau, denudational origin moderately dissected plateau, structural origin-low dissected plateau, and others. Karli and Gad, the main river networks of the study area, originate from the Western Ghats and debouch into the Arabian Sea. Climatologically, the research area is sub-tropical with a minimum temperature of 15°C to a maximum temperature of 40°C having three seasons (rainy, winter, and summer) throughout a year. The annual average rainfall of the area is around 3000 mm, and the maximum rainfall occurs during the southwest monsoon season (June–September). Monsoon rainfall profoundly influences the groundwater in different parts of the study region. According to the groundwater surveys and development agency and central groundwater board reports, the groundwater level ranges between 0.20 and 21 m/bgl. During the field visit, it was observed that groundwater is the primary source for drinking and irrigation extracted through dug and pumping well. Hence, the groundwater management is a vital concern for the study region.

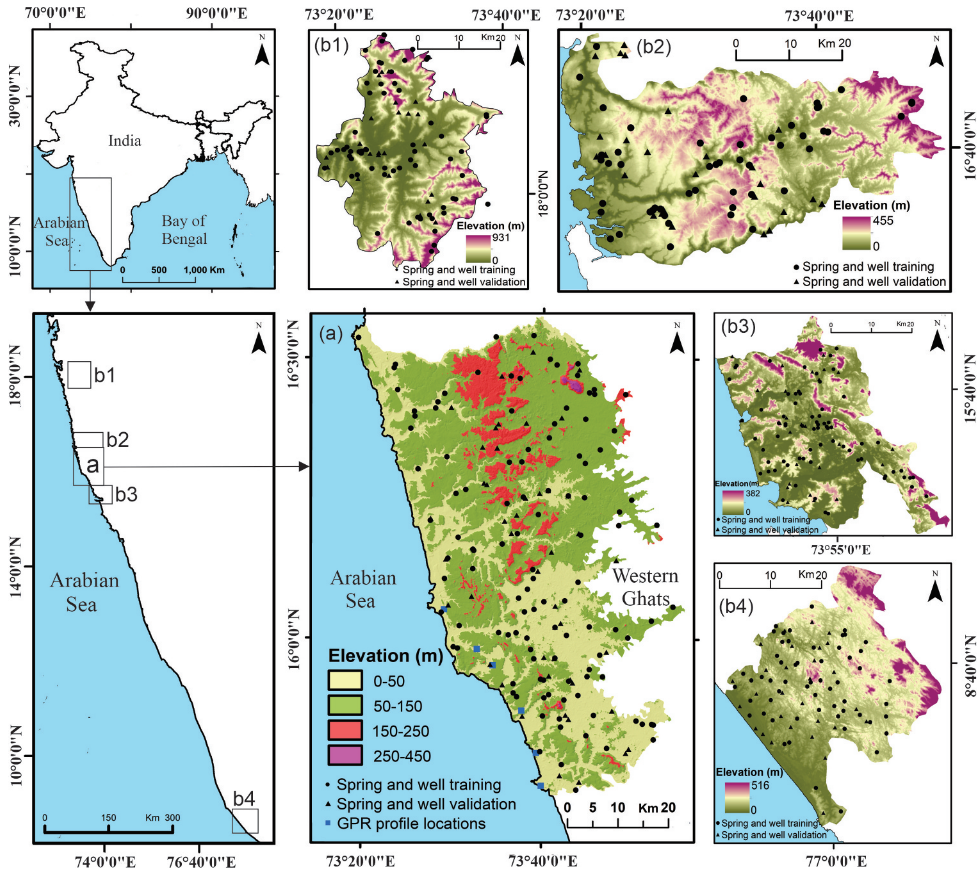

Figure 1. Location maps of the study area (a), and secondary regions (b1, b2, b3, b4).

Apart from this primary study region, four secondary regions were selected, shown in Figure 1 (b1, b2, b3, and b4), for further assessment of the applied models. The geographical details of these areas are given in Table 1.

3. Materials and methods

The methodology applied in the current research is presented in Figure 2, which involves the following stages: Firstly, different thematic layers and inventory map of spring and well were prepared and transformed into the spatial database. Secondly, the groundwater potential maps were produced using machine learning models. Finally, the accuracy of the models was examined by applying the AUROC curve.

3.1. Spring and well inventory map

Many researchers have used the locations of spring, well, and quant as inventory for groundwater potential mapping. In the present research, both spring and well points were considered for GWPM. The inventory map of the study region contains 165 spring and well points, identified from numerous sources (Table 2), and field observation. Random partition algorithm was deployed to separate the spring and well points for training and validation purposes, where 116 (70%) points were

Table 1. Geographical description of selected sites for cross-validation.

Regions	Areal extension: Latitude	Areal extension: Longitude	No. of springs and wells	Area (sq. km)
South-east Raigad (b1)	17°50′57.75″– 18°19′4.39″	73°17′54.76″– 73°40′9.03″	87	1077
South Ratnagiri (b2)	16°29′58.34″– 16°49′44.78″	73°18′0.97″– 73°50′57.34″	76	1221
North Goa (b3)	15°24′22.51″– 15°47′52.15″	73°41′19.26″– 74°7′29.49″	99	933
South Kerala (b4)	8°23′21.62″–8°50′13.49″	76°48′3.46″–77°11′53.69″	87	1036

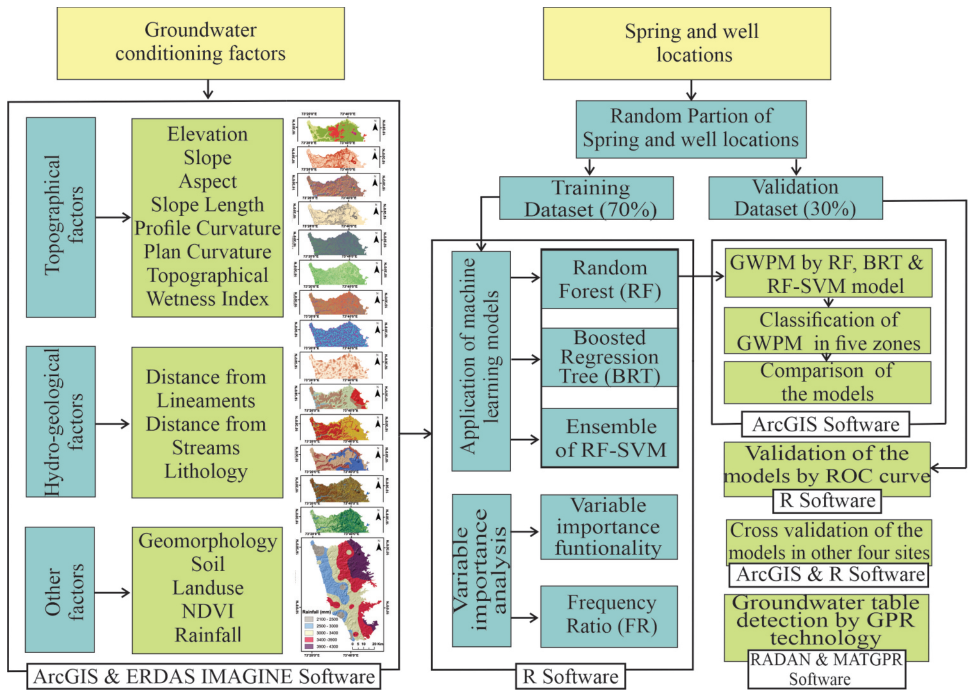

Figure 2. Flowchart of the present study for groundwater potential mapping.

preferred for training and the remaining 49 (30%) for validation of dataset. In the same way, the inventory maps of other regions were prepared.

3.2. Groundwater conditioning factors

It is essential to select the effective parameters for preparing the groundwater prospect map of an area. Based on the previous studies (Naghibi, Pourghasemi, and Dixon 2016; Rahmati, Pourghasemi, and Melesse 2016; Golkarian et al. 2018; Chen et al. 2018) and field examination, 15 groundwater conditioning factors (thematic maps) viz. elevation, slope, aspect, LS, profile curvature, plan curvature, TWI, distance from the streams, distance from the lineaments, lithology, geomorphology, soil, land use, NDVI, and rainfall (Figure 3a–o) were taken into account for GWPM in the study area. These thematic maps were prepared using the ArcGIS 10 software from several data (Table 2). Each thematic layer was resampled into a uniform grid size of 30 × 30 m, and the grid of the research area was prepared by 2201 columns and 3159 rows (3,530,426 pixels; 3177 km^2). Similarly, thematic maps of the selected secondary regions were generated.

SRTM (30 m resolution) digital elevation model (DEM) was used to produce the following topographic factors. Elevation, one of the potential indicators of groundwater, plays a vital role for GWPM (Oh et al. 2011; Naghibi, Pourghasemi, and Dixon 2016; Patra, Mishra, and Mahapatra 2018; Chen et al. 2019). The elevation map was created from the DEM. The slope is considered as the most relevant topographic variable for groundwater potentiality (Naghibi and Pourghasemi 2015; Lee, Hong, and Jung 2018). The slope map was produced in the ArcGIS environment, and the slope values were grouped into six classes using the natural break method. Aspect is one of the important controlling factors for the GWPM. It defines the direction of the slope, which is exposed to sunlight, winds, lineament, and rainfall (Goudie 2013; Chen et al. 2018). The aspect map was used to correlate the groundwater availability at the different directions of the slope. Slope length (LS) defines the length (L) and steepness (S) of the topography that influences the amount of groundwater storage. LS is calculated with the following equation (Moore and Burch 1986).

$$LS = (fa \times cellsize/22.13)^{0.4} \times (sin\theta/0.0896)^{1.3} \quad (1)$$

where *fa* refers to flow accumulation and θ represents the slope in degrees.

Curvature affects the surface and subsurface hydrology (Regmi et al. 2015). Profile curvature is parallel to the maximum slope in a particular direction. The negative value of profile curvature indicates the water flow decelerated in the surface, whereas the positive value indicates the water flow accelerated on the surface, and zero indicates the surface is linear. In another side, plan curvature defines the maximum slope in a perpendicular

Table 2. Details of database used in groundwater potential mapping.

Data layers	Source of data	Scale	Time period
Spring and well locations	Topographical Maps http://www.surveyofindia.gov.in/ Groundwater Surveys and Development Agency https://gsda.maharashtra.gov.in Central Groundwater Board http://cgwb.gov.in/ Field Survey	1:50,000 and 1:25,000	1967 1990–2018 1990–2018
Groundwater level	Groundwater Surveys and Development Agency Central Groundwater Board	–	2007–2016
Rainfall	Department of Agriculture Maharashtra state http://maharain.gov.in/ India Meteorological Department https://mausam.imd.gov.in/	–	2013–2018
Geology	Geological Survey of India https://www.gsi.gov.in	1:250,000	2001
Geomorphology	National Remote Sensing Centre https://bhuvan.nrsc.gov.in SRTM Landsat8 OLI Field Survey	1:50,000	2005–2006
Soil map	National Bureau of Soil Survey and Land Use Planning https://www.nbsslup.in/	1:500,000	1996
Digital elevation model (DEM)	SRTM https://earthexplorer.usgs.gov/	1 arc second,	23 September 2014
Satellite image	Landsat8 OLI https://earthexplorer.usgs.gov/	30 m spatial resolution	19 October 2016, 17 December 2016.

direction. It describes the convergence and divergence of water flow in the earth's surface. Negative values represent the concave slope of the surface, which causes the confluence of water flow. In contrast, positive values indicate the convex slope of the surface that determines the divergence of water flow in the region (ESRI 2016). TWI expresses the effect of topography on the location, which is related to soil conditions of the area. TWI is calculated as follows (Moore, Grayson, and Ladson 1991).

$$TWI = In(fa/tan\beta) \quad (2)$$

Here *fa* is the flow accumulation, and β is the slope angle at the point.

Distance from the stream is inversely related to the identification of groundwater prospects in an area. Lineaments are hydro-geologically very meaningful to control the groundwater movement and storage (Magesh, Chandrasekar, and Soundranayagam 2012). It is in a linear or curvilinear pattern on the earth's surface and identified from the satellite imagery, DEM, and field survey (Pradhan and Lee 2010; Magesh, Chandrasekar, and Soundranayagam 2012; Goudie 2013; Rahmati et al. 2015). For this study, lineaments were extracted from the superimposed shaded relief maps at an interval of 45° azimuth angle. High lineament density indicates more groundwater potentiality in the area (Magesh, Chandrasekar, and Soundranayagam 2012). Euclidean distance method was adopted to examine the relationship between inventory and distance from streams and lineaments. In the study area, the age of the rock traced from the Archean to Recent, which controls the groundwater storage. Based on lithofacies and geological ages, the lithology of the study area was reclassified into six broad classes: Group 1: Archean schist and gneisses (Granite gneiss, quartzite, meta-gabbro, amphibole schist) (Ask), Group 2: Dharwar supergroup (Meta graywacke, metabasalt, granite) (Dsg), Group 3: Kalladgi supergroup (Sedimentary quartzite, shale) (Ksg), Group 4: Sahyadri supergroup (Unclassified flows, Aa flow, Mega cryst flow) (Ssg), Group 5: Laterite (Lat), Group 6: Alluvium (Alv). Geomorphology is another predisposing factor for predicting the potential of groundwater. In this study, the geomorphic unit was categorized into six major groups: 1. Denudational origin-pediment pediplain complex (DoPPc), 2. Coastal origin-younger coastal plain (CoYCp), 3. Structural origin moderately dissected plateau (SoMDp), 4. Denudational origin moderately dissected plateau (DoMDp), 5. Structural origin-low dissected plateau (SoLDp), 6. Others (Oth), using the topographical maps, and Landsat 8 OLI image. Soil is considered one of the most important indicators of the surface and sub-surface runoff, recharge, and infiltration processes (Mogaji, Lim, and Abdullah 2014; Rahmati, Pourghasemi, and Melesse 2016). Soil map from the National Bureau of Soil Survey and Land use planning (NBSSLUP) was used to reclassify the soil into four categories: Ultic Typic Haplustalfs, Lithic Ustorthents, Ultic Haplustalfs, and Typic Ustropepts. Land use and land-use changes influence the groundwater storage and aquifer yield (Ibrahim-Bathis and Ahmed 2016; Guru,

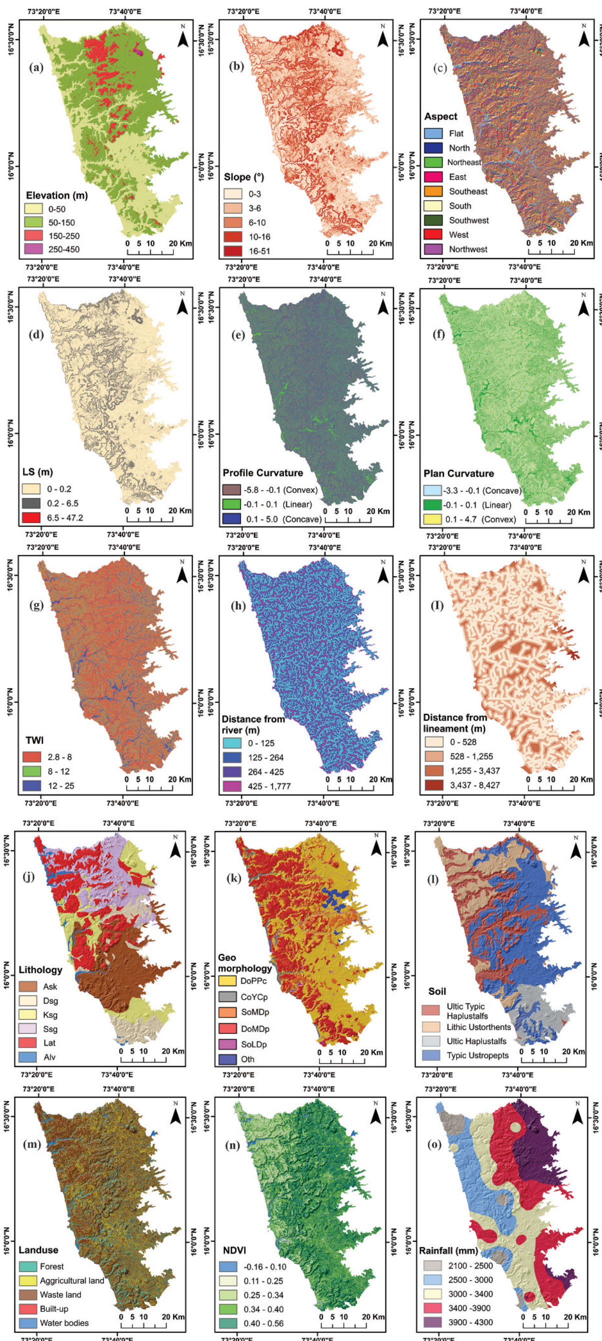

Figure 3. Thematic layers of the groundwater affecting factors (a) elevation, (b) slope, (c) aspect, (d) LS, (e) profile curvature, (f) plan curvature, (g) TWI, (h) distance from rivers, (i) distance from lineaments, (j) lithology, (k) geomorphology, (l) soil, (m) land use, (n) NDVI, (o) rainfall.

Seshan, and Bera 2017; Chen et al. 2018). Unsupervised classification technique was employed to produce the land use map from the Landsat 8 OLI image. The study area was classified into five major land use classes: forest, agricultural land, wasteland, built-up, and waterbodies. The accuracy of the land use classification was calculated as 86% using the Kappa index. NDVI provides the health of vegetation, and it is usually used to relate the vegetation density and groundwater potentiality (Pourghasemi, Moradi, and Aghda 2013; Chen et al. 2018). The range of the NDVI value lying from −1 to 1, with higher NDVI value indicates the healthy vegetation and vice versa. Based on the Landsat 8 OLI image, the NDVI map was created using the following equation:

$$NDVI = (NIR - R)/(NIR + R) \tag{3}$$

Here near-infrared (NIR) and red (R) bands represent the spectral reflectance measurement of these bands.

The distribution, duration, and intensity of rainfall are significant affecting factors for infiltration, runoff, and recharge conditions (Magesh, Chandrasekar, and Soundranayagam 2012). The rainfall map was prepared from the rain gauge data of Maharashtra government using the inverse distance weighted method and classified into five groups.

3.3 Methods

In this research, three machine learning models (RF, BRT, and ensemble of RF-SVM) were employed for GWPM. The relationship between different groundwater conditioning factors and inventory locations was calculated by the frequency ratio method. On the other side, the importance of these groundwater effective factors was measured by "variable importance" function in R software. The raster values of 15 factors of each spring and well location were imported to R software; then, the models were applied using different packages in the R environment. To improve the classification accuracy and avoid the biasness, a 10-fold cross-validation method (with five repetitions) was used in the models. The final output values of the models were transformed into a spatial dataset for GWPM using the ArcGIS software. At last, models were validated by the AUROC curve and also examined in different regions of the west coast of India. In addition, GPR technology was used to measure the groundwater table by RADAN and MATGPR software.

3.3.1 Application of frequency ratio (FR) and variable importance function

FR is a bivariate statistical technique to explain the prospect of occurrence of a certain attribute (Bonham-Carter 1994; Oh et al. 2011; Manap et al. 2014; Naghibi, Pourghasemi, and Dixon 2016; Guru, Seshan, and Bera 2017). It is defined by the relationships between dependent variables (spring and well location) and independent variables (groundwater conditioning factors) (Guru, Seshan, and Bera 2017; Das 2019). In this context, the FR model was used to show the quantitative relationship between the inventory and each sub-class of groundwater variables. FR is calculated as;

$$FR = (P_s/T_s)/(P_a/T_a) \quad (4)$$

where P_s is the number of spring and well under each sub-class of the groundwater effective factors, T_s denotes the total spring and well of the study area, P_a is the number of pixels of each sub-class of the conditioning parameter, and T_a is the total pixels of the study area. FR is the ratio of spring and well occurrences to the total area of each class of the affecting factors. So the FR value 1 indicates the average of the model. If the value is more than 1, it considers the high groundwater prospect and less than 1 value indicates the low groundwater potential in each sub-class of the parameters (Lee and Pradhan 2007; Pradhan and Lee 2010; Oh et al. 2011; Manap et al. 2014; Nampak, Pradhan, and Manap 2014).

"Variable importance" function of the RF model was used for measuring the effectiveness of each factor. It is a generic method for calculating the feature importance by trained methods (Kuhn et al. 2018). Each factor was evaluated individually using a filter approach. The importance values of the features range from 0 to 100. Higher the importance value, greater is the influence of the factor, and vice versa. The details of variable importance function are given in Kuhn et al. (2018).

3.3.2 Application of random forest (RF)

Random forest is an ensemble machine learning technique for both classification and regression tasks (Breiman 2001; Youssef et al. 2016; Naghibi, Ahmadi, and Daneshi 2017b; Kim et al. 2018). For classification, RF uses the resampling technique by randomly changing the predictive variables to increase the diversity in each tree (Youssef et al. 2016; Naghibi, Ahmadi, and Daneshi 2017b). This method consists of multiple decision trees and merges them to explain the spatial relationship between controlling variables of groundwater and inventory of spring and well (Kim et al. 2018). The decision tree is generated by bootstrap samples and leaves few samples for validation to test the accuracy of the decision tree. The mean-squared error of each decision tree with their OOB samples (E_{oob}) is used to calculate the learning error. E_{oob} is expressed as:

$$E_{OOB} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \Delta_i)^2 \quad (5)$$

Here n denotes the total number of OOB samples; y_i is the observed output, and $\hat{y}_i$ is the model output. The main advantages of this approach are; i) it can handle large datasets with high dimensionality as well as avoid the over-fitting of the datasets; ii) this does not need any assumption regarding the explanatory variables and response variables, and iii) it does not require any prior data to transformation and rescaling.

3.3.3 Application of boosted regression tree (BRT)

BRT model is advanced and different from classical regression methods. It uses statistical and machine learning techniques to enhance the performance of a single model by fitting many models and combining them for prediction (Schapire 2003; Elith, Leathwick, and Hastie 2008; Naghibi, Pourghasemi, and Dixon 2016). It is a combination of two algorithms, namely boosting and regression tree (decision tree) (Elith, Leathwick, and Hastie 2008; Youssef et al. 2016; Kim et al. 2018). In the BRT model, the initial decision tree (DT) reduces the loss function. At each iteration, the main target was to decrease the root-mean-square error and the residuals. Then, the next DT is fit for prediction residuals of the first tree. In this stagewise process, the existing trees are unchanged as the model develops increasingly larger. At each step, the fitted value of each observation is reestimated to express the contribution of the recently added tree (Elith, Leathwick, and Hastie 2008; Naghibi, Pourghasemi, and Dixon 2016). DT was used for visualization and explicit decision-making. The advantages of the algorithm are that the predictive variable can be of any type (numeric, binary, and categorical, etc.), and the outcomes of the model are not affected by monotone transformations and different scales of measurement among predictors. It replaces the missing data in predictor variables using surrogates (Breiman 2001; Elith, Leathwick, and Hastie 2008; Youssef et al. 2016). Boosting is a technique to get higher accuracy from the predictive variables of regression trees. It is a sequential procedure to average many rough rules of thumb (Schapire 2003; Elith, Leathwick, and Hastie 2008).

By combining the algorithms, the BRT model establishes a binary tree with general classification and regression tree. The classified data split into two samples. Each sample defines the best point for the data partition, and it formulates the observed deviation and residuals at each partition (Kim et al. 2018). Finally, the model is capable of estimating the observed value.

In the BRT model, three parameters such as number of trees, shrinkage or learning rate, and interaction depth are required for tuning. Interaction depth defines the number of nodes in trees and the learning rate determines the importance of each tree in the built model. Based on these two parameters, the number of trees was decided for optimal prediction (Elith, Leathwick, and Hastie 2008; Naghibi, Pourghasemi, and Dixon 2016).

3.3.4 Application of ensemble of RF and support vector machine (SVM)

SVM model is another popular supervised machine learning technique that is based on the concept of structural risk minimization and statistical learning theory (Tehrany, Pradhan, and Jebur 2014; Mojaddadi et al. 2017; Naghibi, Pourghasemi, and Abbaspour 2018). The principle of the method is to separate the hyperplane formation from the dataset. The hyperplane is defined as the center of the maximum margin of separation (Marjanovic et al. 2011; Tehrany et al. 2015). On the basis of the hyperplane, the point was classified as +1 or −1. For the case of linear separable data, a separating hyperplane can be computed with the help of the following equation (Hong et al. 2017):

$$y_i(w.x_i + b) \geq 1 - \zeta_i \tag{6}$$

where w represents the coefficient vector which expresses the orientation of the hyperplane in the feature space, b is the offset of the hyperplane from the origin, and ζ_i defines the positive slack variables. The following optimization problem can be solved by defining an optimal hyperplane (Samui 2008).

$$Minimize \sum_{i=1}^{n} a_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j y_i a y_j (x_i x_j) \tag{7}$$

$$Subject\ to \sum_{i=1}^{n} a_i y_i = 0, 0 \leq a_i \leq c \tag{8}$$

Here a_i indicates the lag range multiplier and C denotes the penalty. The precision of the successful classification in SVM model depends on the selection of kernel type (Yao, Tham, and Dai 2008). In the present research, radial basis function (RBF) was applied because of its higher capability in interpolation, as reported by many researchers (Tehrany, Pradhan, and Jebur 2014; Tehrany et al. 2015; Gayen et al. 2019). RBF equation is calculated as follows:

$$RBF : K(x_i, x_j) = \exp\left(-yx_i - x_j^2\right) \tag{9}$$

where K (x_i,x_j) is the kernel function, y represents the RBF kernel function. The purpose of applying the algorithm was to reduce the error and model complexity (Naghibi, Pourghasemi, and Abbaspour 2018).

The ensemble model was developed by combining two or more than two different predictive models. In recent studies (Naghibi et al. 2019; Kordestani et al. 2019; Chen et al. 2019), the ensemble model was applied to improve the results. In this work, RF and SVM models were ensembled in R software using the weighted average method. For the implementation of these machine learning models, R statistical software 3.6.1 version was used with the help of different packages.

3.4 Validation of groundwater potential maps

Validation is a fundamental step in modeling for the scientific significance of the research (Naghibi, Pourghasemi, and Dixon 2016; Chen et al. 2019). In this research, the area under the receiver operating characteristic (ROC) curve was opted for evaluation of the models. ROC is a graphical plot, which determines the performance of the models in a diagnostic test (Egan 1975; Golkarian et al. 2018). The curve plots the true-positive rate (sensitivity) on Y-axis and false-positive rate (1 – specificity) on X-axis (Youssef et al. 2016; Golkarian et al. 2018). Model prediction for occurrence and non-occurrence of springs and wells was evaluated using the area under the ROC curve. The area under the curve (AUC) represents the value between 0 and 1, and the higher value represents the better performance of the model (Youssef et al. 2016; Naghibi, Pourghasemi, and Dixon 2016; Golkarian et al. 2018; Chen et al. 2018, 2019). To extract the generalized findings, it is necessary to apply the models in different regions. For this purpose, four different areas along the west coast of India were chosen (b1, b2, b3, and b4). The b1 and b4 regions show slightly different characteristics from the study area in terms of topography, lithology, and climatic conditions whereas b2 and b3 sites lying in close proximity to the study area appear to be similar with respect to aforesaid criteria. Additionally, six GPR profiles were selected to examine the groundwater level with the corresponding water level of wells. GPR is a noninvasive geophysical technique based on propagation and reflection of the transmitted electromagnetic waves (Annan 2003; Neal 2004; Billy et al. 2014). A SIR-4000 model of GSSI (Geophysical Survey System Inc.) with 200 MHz antenna was used to identify the water table depth. GPR data were processed in Radan 7 and MatGPR 3.2 (compatible with Matlab software) software by applying the time zero removal, filtering, background removal, and migration.

4 Results

4.1 Spatial relationship between groundwater conditioning factors and inventory of spring and well

It is necessary to know the relationship between inventory with effective factors for GWPM in an area. The results of the above relation are shown in Table 3 using the FR model. The FR value ranges from 0 to 1.78

Table 3. Spatial relationship between each groundwater effective factor and inventory of spring and well using frequency ratio (FR) model.

Parameters	Classes	% of total area (a)	% of inventory area (b)	Frequency ratio (b/a)
Elevation (m)	0–50	34.29	52.12	1.49
	50–150	56.36	43.64	0.78
	150–250	8.87	4.24	0.48
	250–450	0.48	0.00	0.00
Slope (degree)	0–3	31.78	30.30	0.95
	3–6	30.86	36.36	1.18
	6–10	18.01	18.79	1.04
	10–16	10.96	7.88	0.72
	16–51	8.39	7.27	0.87
Aspect	Flat	4.82	2.42	0.51
	North	11.04	7.88	0.71
	Northeast	10.71	12.73	1.19
	East	10.70	10.30	0.96
	Southeast	11.59	12.12	1.05
	South	12.37	9.09	0.73
	Southwest	13.22	14.55	1.10
	West	13.13	15.76	1.20
	Northwest	12.42	15.15	1.22
LS (m)	0–0.2	88.80	84.24	1.01
	0.2–6.5	16.17	15.76	0.97
	6.5–47.2	0.02	0.00	0.00
Profile curvature	Convex	29.31	20.61	0.70
	Linear	35.09	50.30	1.43
	Concave	35.60	29.09	0.82
Plan curvature	Concave	24.56	16.36	0.67
	Linear	45.71	64.24	1.41
	Convex	29.73	19.39	0.65
TWI	2.8–8	63.26	55.15	0.87
	8–12	22.69	29.09	1.29
	12–25	14.05	15.76	1.10
Distance from river (m)	0–125	38.13	45.45	1.19
	125–264	30.30	28.48	0.94
	264–425	22.80	21.82	0.94
	425–1777	8.77	4.24	0.51
Distance from lineament (m)	0–528	52.52	61.82	1.18
	528–1255	37.79	33.94	0.90
	1255–3437	9.19	4.24	0.46
	3437–8427	0.49	0.00	0.00
Lithology	Ask	27.40	37.58	1.37
	Dsg	11.30	5.45	0.48
	Ksg	14.44	23.64	1.53
	Ssg	20.15	18.79	0.93
	Lat	23.38	13.33	0.57
	Alv	3.33	1.21	0.53
Geomorphology	DoPPc	49.76	76.36	1.55
	CoYCp	2.25	3.03	1.35
	SoMDp	18.65	13.33	0.71
	DoMDp	25.96	5.45	0.21
	SoLDp	2.03	1.82	0.89
	Oth	1.35	0.00	0.00
Soil	Ultic Typic Haplustalfs	25.27	33.94	1.34
	Lithic Ustorthents	18.15	10.91	0.60
	Ultic Haplustalfs	10.92	19.39	1.78
	Typic Ustropepts	45.64	35.76	0.78
Land use	Forest	15.27	18.79	1.23
	Agricultural Land	26.23	39.39	1.50
	Wasteland	55.00	38.79	0.71
	Built-up	1.98	2.42	1.22
	Waterbodies	1.90	0.61	0.32
NDVI	−0.16–0.10	1.84	1.21	0.67
	0.10–0.25	9.06	1.82	0.20
	0.25–0.34	16.10	27.27	1.69
	0.34–0.40	34.23	40.00	1.17
	0.40–0.56	38.77	29.70	0.77
Rainfall(mm)	2100–2500	4.94	45.45	0.99
	2500–3000	19.98	27.88	0.73
	3000–3400	29.14	22.42	1.08
	3400–3900	28.53	4.24	0.98
	3900–4300	17.41	45.45	1.22

in different sub-classes of the groundwater conditioning variables. For elevation, FR value decreases with the higher altitude. In the case of slope, the class from 3°-6° is highly correlated with spring and well occurrences (FR = 1.18). The northwest slope direction with FR value of 1.22 has more spring and well locations compared to other directions of the slope. FR values of LS classes are decreasing with the increasing slope length. In the matter of profile and plan curvature, flat areas have the highest FR value of 1.43 and 1.41, respectively. In TWI class, the second class (8–12) has the maximum FR value (1.29). The FR values increase with decreasing distance from streams, indicating a high groundwater prospect. In relation to lineament, most of the springs and wells occur at 0 to 1.25 km distance from the lineaments. Regarding lithology, spring and well locations have mostly identified in the Kaladgi group of rock with FR value of 1.53. For geomorphology, FR of the denudational origin-pediment pediplain complex has a maximum value (1.55). In the soil groups, Ultic Haplustalfs soil has a strong correlation with spring and well occurrences with FR value 1.78. In relation to land use, agricultural land is the maximum FR value 1.50. In the case of NDVI, the highest FR value is 1.69 in the third class (0.25–-0.34). In the rainfall class, the highest rainfall (3900–4300 mm) shows the maximum FR value (1.22).

4.2 Groundwater potential models

Groundwater potential maps were prepared using the three machine learning algorithms, namely RF, BRT, and the hybrid of RF-SVM models (Figure 4). Based on Min-max normalization, the probability values of the models were normalized between 0 and 1 and subsequently classified into five zones: very low (0–0.20), low (0.20–0.45), moderate (0.45–0.70), high (0.70–0.85), and very high (0.85–1) with same class range for comparison among the models. The higher value represents a very good groundwater prospect of the area and vice versa.

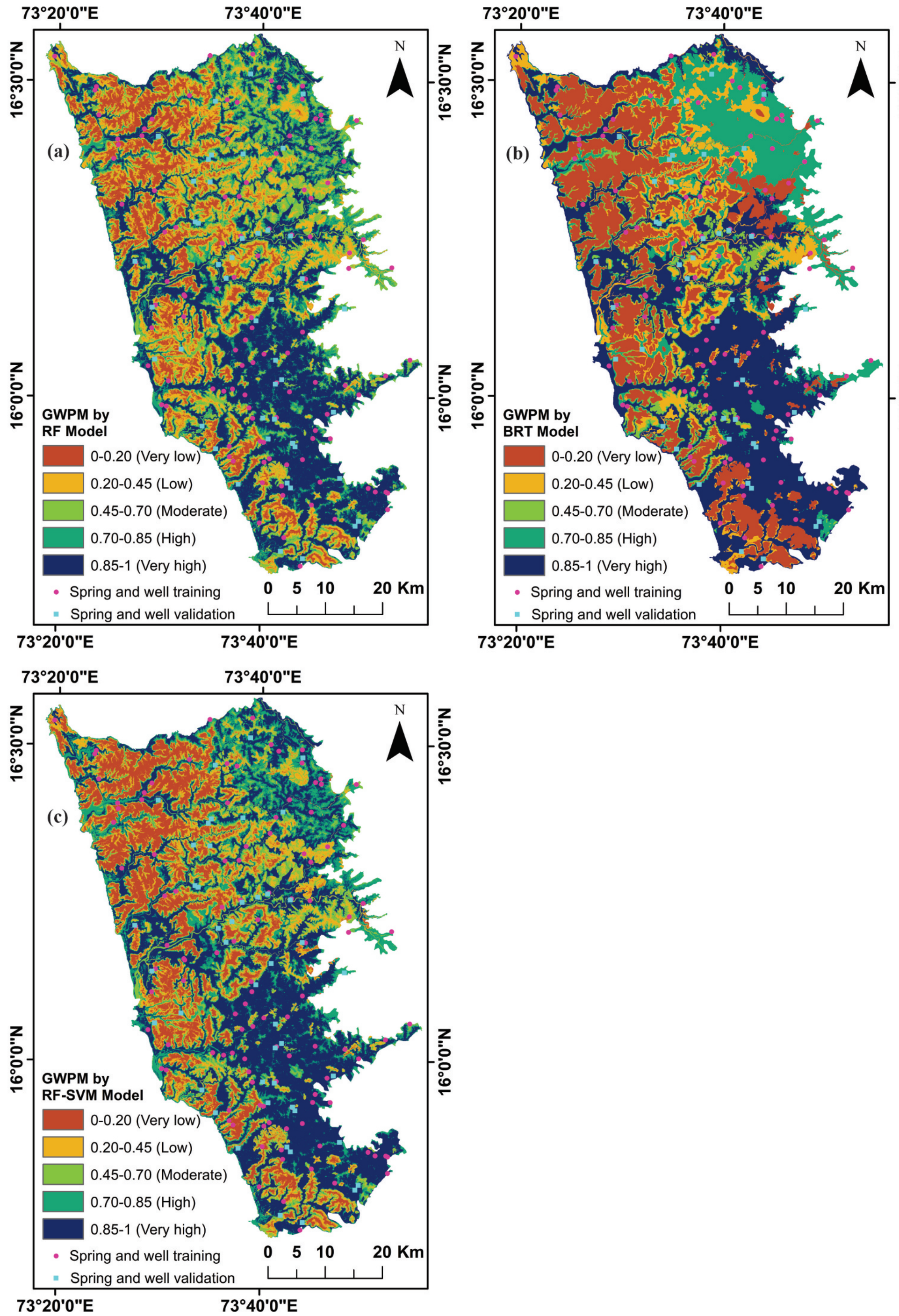

Figure 4. Groundwater potential maps derived from (a) RF, (b) BRT, and (c) RF-SVM models.

The model-wise spatial variation of the groundwater potentiality is shown in Figure 4 and Table 4. According to the RF model, the high and very high prospective groundwater zones covered 16.73% and 33.31% of the study area, respectively. Low and very low groundwater prospect zones comprise 17.98% and 12.48% of the total area, respectively. In the BRT model, it was found that 19.13% and 35.60% of the entire research area were of high and very high groundwater potentiality. Besides, low and very low groundwater potentiality covered 12.93% and 26.37% of the area. Based on the ensemble model result, the study region was classified as very high (36.85%), high (19.18%), moderate (13.14%), low (13.99), and very low (16.84%) groundwater prospect zones (Table 4).

Table 4. Area of different classes (%) of groundwater potential maps using RF, BRT, and RF-SVM ensemble models.

	Area in percentage		
Class	RF	BRT	RF-SVM
0–0.20 (very low)	12.48	26.37	16.84
0.20–0.45 (low)	17.98	12.93	13.99
0.45–0.70 (moderate)	19.50	5.97	13.14
0.70–0.85 (high)	16.73	19.13	19.18
0.85–1 (very high)	33.31	35.60	36.85
Total	100	100	100

4.3 Validation of machine learning models and GPR profiles

The validation of the model is crucial for the assessment of GWPM. In many studies, the ROC curve was used for quantitative validation of the models with a high prediction rate (Golkarian et al. 2018; Chen et al. 2018). In the present study, the validity of RF, BRT, and ensemble models was confirmed using the ROC curve. The AUC values ranging from lower (0) to higher (1) represent the worst to the best model prediction for GWPM. Based on the ROC result, the success rate of RF, hybrid of RF-SVM, and BRT models was measured as 94.0%, 93.4%, and 89.8%, respectively. The accuracy of the models from the selected sites was calculated. In the b1 area, the accuracy of RF, BRT, and RF-SVM was 82.2%, 72%, and 83.3%, respectively. The success rates of RF (94%, 96.5%), BRT (90.5%, 90.7%), and hybrid models (93.8%, 94.7%) were computed for the b2 and b3 regions, respectively. In the case of b4 region, the precision of RF, BRT, and RF-SVM was evaluated as 82.8%, 80.7%, and 82.5%, respectively.

The water table from the GPR profiles (a, b, c, e, and f) was identified at the depth of 3.6, 3.4, 4.9, 4.8, 4.3, and 4.1 m, respectively, and the nearest water level of wells was observed at 3.6, 3.4, 4.9, 5.1, 4.1, and 4.4 m, respectively (Table 5). Only in the case of profile "d," the water table could not be identified. From the five GPR profiles, the average and maximum difference between predicted and measured depths of groundwater were 14 and 30 cm, respectively. Overall, from six GPR profiles, five locations of water table were accurately determined, which means more than 80% accuracy achieved in detecting the groundwater table by GPR technology.

5 Discussion

The presence of spring and well at the particular segments of the study area indicates the potentiality of high groundwater yield (Oh et al. 2011; Naghibi et al. 2017a). However, to assess the groundwater prospect of the entire area, statistical and machine learning methods were used by many researchers with good results (Oh et al. 2011; Chen et al. 2019). The results of the present study are discussed as follows:

5.1 Important conditioning factors for GWPM

The comparative importance of the 15 influencing factors of groundwater was illustrated using "variable importance" function of RF model. In this context, geomorphology had the highest importance followed by elevation, NDVI, and distance from the stream, while soil was of lowest importance followed by slope length rainfall, and land use (Figure 5). Geomorphology is the most effective factor since the geomorphic features of different landforms of the study area control the groundwater potentiality to a maximum extent. Various landforms on the earth's surface are associated with a different kind of groundwater storage (Deepika, Avinash, and Jayappa 2013; Rajaveni, Brindha, and Elango 2017). Structural hill, residual hill, and linear ridge represent the low groundwater potential, whereas pediplain and valley fill have the high groundwater potential due to high infiltration and groundwater recharge (Rajaveni, Brindha, and Elango 2017; Berhanu and Hatiye 2020). Almost 50% of the study area is associated with a pediment-pediplain complex, which indicates the good groundwater potentiality. Elevation is another critical factor for GWPM, which was an agreement with the results of Naghibi and Pourghasemi (2015), Naghibi, Pourghasemi, and Dixon (2016), Rahmati, Pourghasemi, and Melesse (2016), Naghibi, Ahmadi, and Daneshi (2017b), and Naghibi, Pourghasemi, and Abbaspour (2018). The lower elevation of the study area has the highest potentiality of groundwater due to hydraulic gradient and presence of low water table. The NDVI, which was the highest contributing factor for GWPM, found in the study of

Table 5. Water table depths from GPR profiles and measured wells.

GPR profiles	Place name	Latitude (N)	Longitude (E)	Water level depth from the measured wells (m)	Water table depth from GPR profiles (m)	Deviation from well water level to GPR water table (m)
a	Dedoolwada	16.04812	73.47902	3.6	3.8	0.2
b	Betwa	15.98158	73.54611	3.4	3.4	0
c	Mopar	15.95280	73.57527	4.9	4.9	0
d	Daboli	15.87222	73.62916	5.1	Not detectable	-
e	Vengurla	15.86111	73.63194	4.1	4.3	0.2
f	Redi	15.73811	73.66579	4.4	4.1	0.3

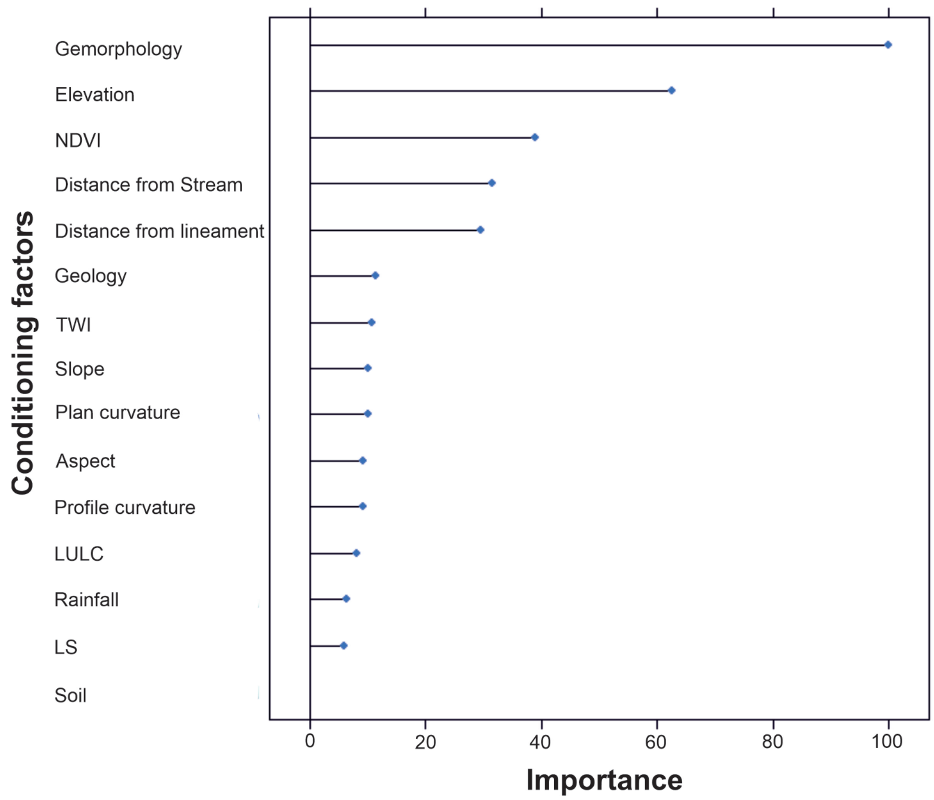

Figure 5. Variables importance in groundwater potential mapping.

Naghibi et al. (2017a) and Chen et al. (2019), is too incorporated in the present study. It is also considered as significant for land use classification, where forest class contains high FR value (1.23). Distance from the stream was a dominant factor for GWPM in the research of Kordestani et al. (2019), which reflects in the current study. The parallel drainage pattern indicates the presence of a fault system (Deffontaines and Chorowicz 1991) in the study area, which controls the groundwater movement and storage. The soil has the lowest controlling factor in GWPM because more than 30% of the area consists of hard rock, and which has negligible for primary porosity. Fractured rock, weathered basement, and depth of soil are favorable for the groundwater occurrence and movement (Prasad et al. 2008; Maiti et al. 2012; Das 2017). In the study area, the groundwater occurrence is mainly controlled by secondary porosity (weathered and fracture rocks) rather than the primary porosity of the soil.

On the other side, the FR model determines the importance of each sub-class of the groundwater affecting variables. The maximum values of the FR ratio were 1.78 (Ultic Haplustalfs), 1.69 (0.25–0.34), 1.55 (Denudational origin-pediment pediplain complex), and 1.53 (Kalladgi), for soil, NDVI, geomorphology, and geology factors, respectively. The importance of geomorphology and NDVI coincides with the result of the variable importance index. In the case of soil, the results of the variable importance index and FR were in contrast due to the consideration of a particular soil class that was conducive for high groundwater potentiality. The geological aspects (lithology, lineament) also influence the distribution and occurrence of groundwater (Berhanu and Hatiye 2020). The Kaladgi supergroup mainly consists of sedimentary rocks having more porosity in comparison to hard rocks. In the case of hard rock, groundwater occurrence is controlled by fault, boundaries between the different lithological units, weathered, and fracture zones. However, the properties of the study area and adopted methods have influenced the effective factors for GWPM.

5.2 Interpretation of the models and GPR profiles

In different studies (Naghibi, Ahmadi, and Daneshi 2017b; Golkarian et al. 2018) on GWPM, the RF model provides an excellent result. The results of the machine learning models from AUC exhibit that the RF is the best-fit model for the current study. The better performance of the RF model (AUC = 94%) may be because of the model consists of the multiple decision trees with no overfitting of the data. Besides, the model provides the interaction ability between effective factors and nonlinearity (Catani et al. 2013). In recent years, ensemble

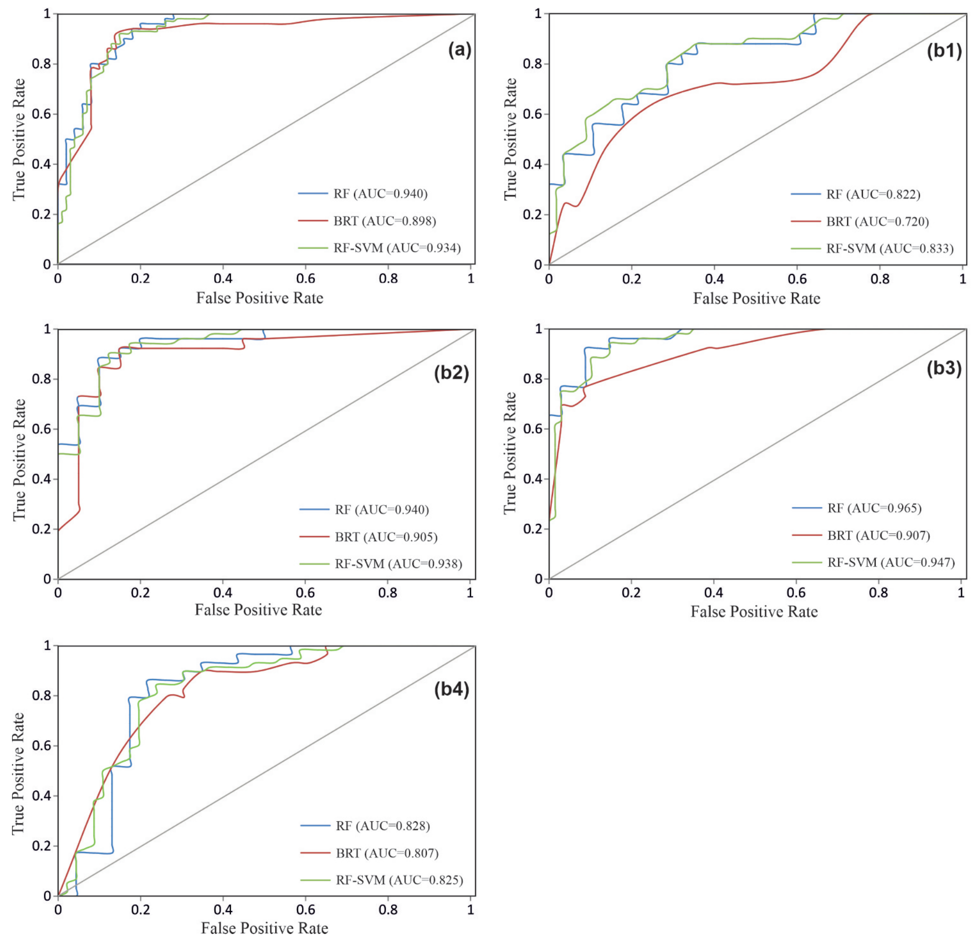

Figure 6. ROC curve of the models in study area (a), secondary areas (b1, b2, b3, b4).

models are increasingly used in groundwater potential mapping with very high accuracy (Chen et al. 2019; Kordestani et al. 2019; Naghibi et al. 2019). It was observed that the RF-SVM model had performed well with 93.40% accuracy from AUC (Figure 6). The advantages of RF and SVM models are the probable reason for high performance in the present work. The result of the BRT model has lesser accuracy compared to RF and ensemble models that may be due to the overfitting of the data. However, since the AUC values of the models are more than 0.7, the GWPM has been reliable for the study region (Naghibi, Pourghasemi, and Dixon 2016; Golkarian et al. 2018). The RF and RF-SVM models were successfully applied in the other selected parts of India. In the b1 and b4 areas, the RF and RF-SVM models performed better than the BRT model (Figure 6). On the other hand, the results from the b2 and b3 regions almost matched the result of the study area due to the homogenous hydro-geologic, topographic, and climatic properties. Moreover, the RF and hybrid models of RF-SVM are promising and sufficient to be advised as the method to prepare groundwater potential maps at the regional scale.

In many research works (Annan, Cosway, and Redman 1991; Nakashima, Zhou, and Sato 2001; Bano 2006; Mahmoudzadeh et al. 2012; Manu and Preko 2014), GPR was successfully used to identify the groundwater table for the better understanding of groundwater condition of an area. In this context, GPR technology was used to detect the groundwater table and advise the location for a new well in the research area. The strong radar reflection and amplitude variation from the groundwater table suggests the different dielectric contrast of the earth materials (Shih et al. 1986; Doolittle et al. 2006; Manu and Preko 2014). The results from GPR profiles revealed that radar reflection, amplitude variation, and high attenuation have prominent signatures of the water table. Groundwater table of the profile a, b, c, e, and f is precisely matched with the nearest water level data of the wells (Table 5). The water table from the GPR profile "d" was not identified due to the presence of different layers interpreted from the multiple radar reflections (Figure 7).

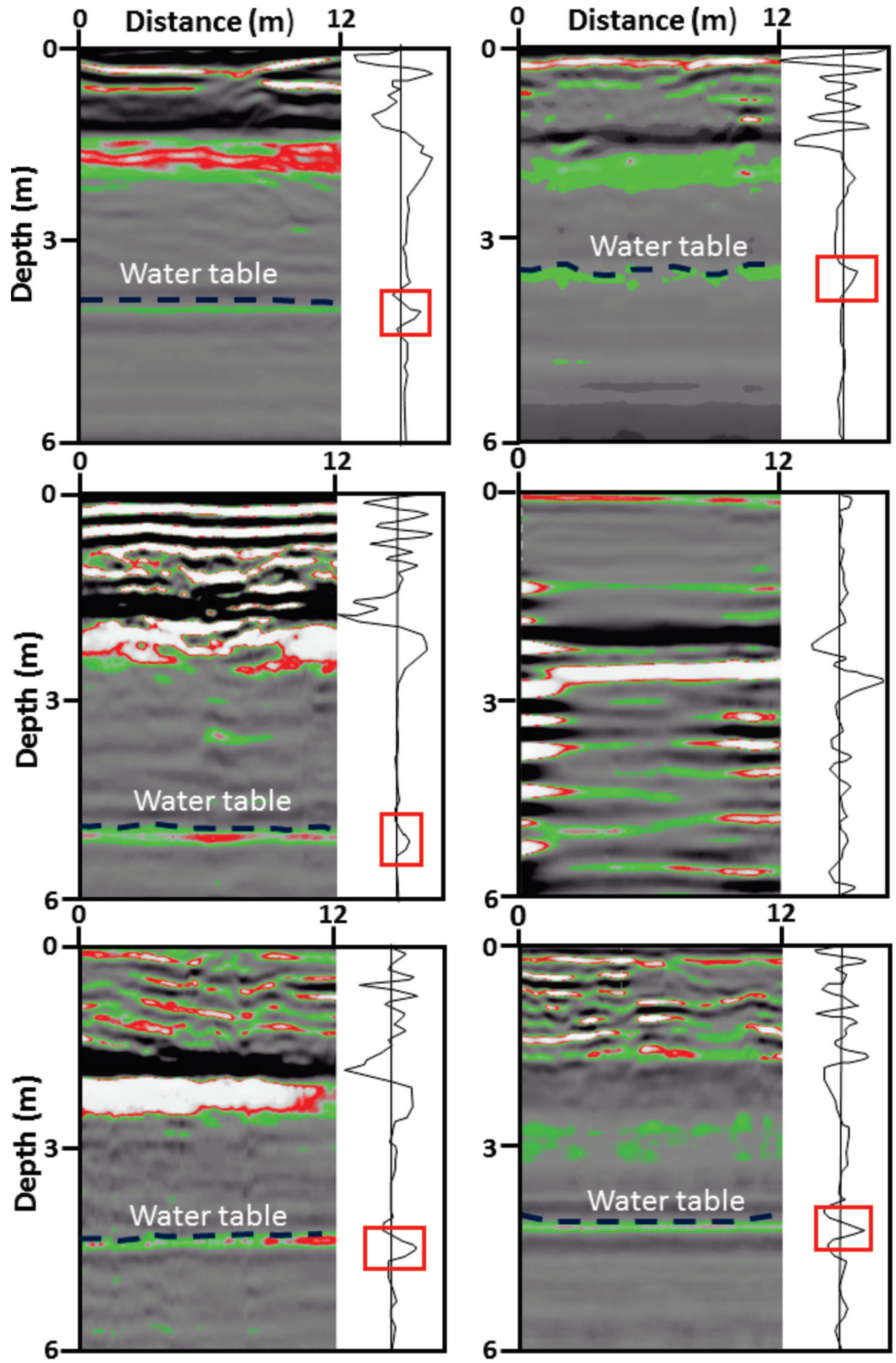

Figure 7. Water table depth from GPR profiles (a) 3.8 m (b) 3.4 m (c) 4.9 m (d) not detectable (e) 4.3 m and (f) 4.1 m.

5.3 Precision of the GWPM

A better model defines that the area of high and very high classes of the models is precisely matched with minimal variation (Naghibi, Ahmadi, and Daneshi 2017b). In this research, the outcomes of the five classes of groundwater potential area are consistent with the lowest percentage of variation for all the models. The validation results from the ROC suggest high preciseness of this model for GWPM. The cross-validation of the RF and RF-SVM models in all the selected sites has ensured the applicability of the models with good precision. For water table detection, the GPR has produced an excellent result which in turn validates the groundwater mapping in the area of interest.

6. Conclusion

The increasing demand for groundwater made concern for groundwater potential mapping, especially on the west coast of India. In this research, machine learning algorithms have been applied to demarcate the groundwater potential zones in Sindhudurg coastal sector of the west coast of India. Based on literature review and field knowledge, 15 groundwater-related thematic layers were superimposed with inventory location in the GIS environment and integrated with RF, BRT, and ensemble of RF-SVM models. According to the results from RF, BRT, and RF-SVM, the very high groundwater potential zone occupies 33.31%, 35.60%, and, 36.85%, respectively, of the research area. The prediction of the models was validated with the AUROC curve. Based on AUROC curves, RF and RF-SVM models exhibited better performance than the BRT model for GWPM in the research area. Likewise, in the other four selected regions, RF and RF-SVM models proved to be superior in comparison to the BRT model. The most influencing factors of the groundwater prospect mapping were geomorphology, elevation, NDVI, distance from streams, and distance from lineament. The predicted depth of groundwater from GPR profiles and measured data during fieldwork on those wells are corroborating to each other, which helps to identify new potential well in the study region using GPR technology. The obtained results of the present study can be useful for government and private agencies in groundwater resource management, land use planning, and environmental protection in the study region. Furthermore, the methodology of this research can be adopted to study other coasts and watersheds with more or less similar hydro-geologic, topographic, and climatic properties.

Highlights

- RF and RF-SVM ensemble models performed very well with AUC>0.9.
- Evaluation of the models in four different regions with high-precision results.
- Geomorphology is the most important variable in groundwater potential mapping.
- GPR technique successfully measured the groundwater table.

Acknowledgements

We acknowledge the financial support from the Indian Rare Earth Limited (Project no. SSP-3232) and University Grant Commission (3160 NET-June 2015). The authors are thankful to the Director CSIR-NIO for support. The authors are also grateful to Prof. Jungho Im and anonymous reviewers for their critical comments and constructive suggestions to improve the manuscript. Field support from the survey team members KM Dubey, Lalit Arya, and Trilochan Kumar is thankfully acknowledged. The NIO contribution number is 6566.

Disclosure statement

The authors declare no conflict of interest.

Funding

This work was supported by the Indian Rare Earth Limited [SSP3232]; University Grants Commission [3160(NET-JUNE 2015)].

ORCID

Pankaj Prasad http://orcid.org/0000-0002-3118-2201

References

Annan, A. P. 2003. *Ground Penetrating Radar Principles, Procedures, and Applications*. Mississauga, ON, Canada: Sensors & Software .

Annan, A. P., S. W. Cosway, and J. D. Redman. 1991. "Water Table Detection with Ground-penetrating Radar." *Society of Exploration Geophysicists*. 494–496. doi:10.1190/1.1888793.

Arabameri, A., K. Rezaei, A. Cerda, L. Lombardo, and J. Rodrigo-Comino. 2019. "GIS-based Groundwater Potential Mapping in Shahroud Plain, Iran. A Comparison among Statistical (Bivariate and Multivariate), Data Mining and MCDM Approaches." *Science of the Total Environment*. 658:160–177. doi:10.1016/j.scitotenv.2018.12.115.

Bano, M. 2006. "Effects of the Transition Zone above a Water Table on the Reflection of GPR Waves." *Geophysical Research Letters*. 33. doi:10.1029/2006GL026158.

Berhanu, K. G., and S. D. Hatiye. 2020. "Identification of Groundwater Potential Zones Using Proxy Data: Case Study of Megech Watershed, Ethiopia." *Journal of Hydrology: Regional Studies*. 28:100676. doi:10.1016/j.ejrh.2020.100676.

Billy, J., N. Robin, C. J. Hein, R. Certain, and D. M. FitzGerald. 2014. "Internal Architecture of Mixed Sand-and-gravel Beach Ridges: Miquelon-Langlade Barrier, NW Atlantic." *Marine Geology*. 357:53–71. doi:10.1016/j.margeo.2014.07.011.

Bonham-Carter, G. F. 1994. *Geographic Information Systems for Geoscientists: Modelling with GIS*. UK: Elsevier.

Breiman, L. 2001. "Random Forests." *Machine Learning*. 45 (1):5–32.

Catani, F., D. Lagomarsino, S. Segoni, and V. Tofani. 2013. "Landslide Susceptibility Estimation by Random Forests Technique: Sensitivity and Scaling Issues." *Natural Hazards and Earth System Sciences*. 13(11):2815–2831.

Chen, W., M. Panahi, K. Khosravi, H. R. Pourghasemi, F. Rezaie, and D. Parvinnezhad. 2019. "Spatial Prediction of Groundwater Potentiality Using ANFIS Ensembled with

Teaching-learning-based and Biogeography-based Optimization." *Journal of Hydrology*. 572:435–448. doi:10.1016/j.jhydrol.2019.03.013.

Chen, W., H. Li, E. Hou, S. Wang, G. Wang, and T. Peng. 2018. "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. doi:10.1016/j.scitotenv.2018.04.055.

Crisci, C., B. Ghattas, and G. Perera. 2012. "A Review of Supervised Machine Learning Algorithms and Their Applications to Ecological Data." *Ecological Modelling*. 240:113–122. doi:10.1016/j.ecolmodel.2012.03.001.

Das, S. 2017. "Delineation of Groundwater Potential Zone in Hard Rock Terrain in Gangajalghati Block, Bankura District, India Using Remote Sensing and GIS Techniques." *Modeling Earth Systems and Environment*. 3:1589–1599. doi:10.1007/s40808-017-0396-7.

Das, S. 2019. "Comparison among Influencing Factor, Frequency Ratio, and Analytical Hierarchy Process Techniques for Groundwater Potential Zonation in Vaitarna Basin, Maharashtra, India." *Groundwater for Sustainable Development*. 8:617–629.

Deepika, B., K. Avinash, and K. S. Jayappa. 2013. "Integration of Hydrological Factors and Demarcation of Groundwater Prospect Zones: Insights from Remote Sensing and GIS Techniques." *Environmental Earth Sciences*. 70(3):1319–1338.

Deffontaines, B., and J. Chorowicz. 1991. "Principles of Drainage Basin Analysis from Multisource Data: Application to the Structural Analysis of the Zaire Basin." *Tectonophysics*. 194 (3):237–263.

Doolittle, J. A., B. Jenkinson, D. Hopkins, M. Ulmer, and W. Tuttle. 2006. "Hydropedological Investigations with Ground-penetrating Radar (GPR): Estimating Water-table Depths and Local Ground-water Flow Pattern in Areas of Coarse-textured Soils." *Geoderma*. 131(3–4):317–329.

Duan, H., Z. Deng, F. Deng, and D. Wang. 2016. "Assessment of Groundwater Potential Based on Multicriteria Decision Making Model and Decision Tree Algorithms." *Mathematical Problems in Engineering*. doi:10.1155/2016/2064575.

Egan, J. P. 1975. *Signal Detection Theory and ROC-analysis*. New York: Academic Press.

Elith, J., J. R. Leathwick, and T. Hastie. 2008. "A Working Guide to Boosted Regression Trees." *Journal of Animal Ecology*. 77 (4):802–813.

ESRI. 2016. "ArcGIS for Desktop." http://desktop.arcgis.com

Friedl, M. A., C. E. Brodley, and A. H. Strahler. 1999. "Maximizing Land Cover Classification Accuracies Produced by Decision Trees at Continental to Global Scales." *IEEE Transactions on Geoscience and Remote Sensing*. 37(2):969–977.

Ganapuram, S., G. V. Kumar, I. M. Krishna, E. Kahya, and M. C. Demirel. 2009. "Mapping of Groundwater Potential Zones in the Musi Basin Using Remote Sensing Data and GIS." *Advances in Engineering Software*. 40(7):506–518.

Gayen, A., H. R. Pourghasemi, S. Saha, S. Keesstra, and S. Bai. 2019. "Gully Erosion Susceptibility Assessment and Management of Hazard-prone Areas in India Using Different Machine Learning Algorithms." *Science of the Total Environment*. 668:124–138. doi:10.1016/j.scitotenv.2019.02.436.

Gislason, P. O., J. A. Benediktsson, and J. R. Sveinsson. 2006. "Random Forests for Land Cover Classification." *Pattern Recognition Letters*. 27(4):294–300.

Golkarian, A., S. A. Naghibi, B. Kalantar, and B. Pradhan. 2018. "Groundwater Potential Mapping Using C5.0, Random Forest, and Multivariate Adaptive Regression Spline Models in GIS." *Environmental Monitoring and Assessment*. 190:149. doi:10.1007/s10661-6507-8.

Goudie, A. S. 2013. *Encyclopedia of Geomorphology*. London: Routledge.

Guru, B., K. Seshan, and S. Bera. 2017. "Frequency Ratio Model for Groundwater Potential Mapping and Its Sustainable Management in Cold Desert, India." *Journal of King Saud University-Science*. 29(3):333–347.

Hong, H., B. Pradhan, D. T. Bui, C. Xu, A. M. Youssef, and W. Chen. 2017. "Comparison of Four Kernel Functions Used in Support Vector Machines for Landslide Susceptibility Mapping: A Case Study at Suichuan Area (China)." *Geomatics, Natural Hazards and Risk*. 8(2):544–569.

Ibrahim-Bathis, K., and S. A. Ahmed. 2016. "Geospatial Technology for Delineating Groundwater Potential Zones in Doddahalla Watershed of Chitradurga District, India." *The Egyptian Journal of Remote Sensing and Space Science*. 19(2):223–234.

Khosravi, K., B. T. Pham, K. Chapi, A. Shirzadi, H. Shahabi, I. Revhaug, I. Prakash, and D. T. Bui. 2018. "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. doi:10.1016/j.scitotenv.2018.01.266.

Kim, J. C., S. Lee, H. S. Jung, and S. Lee. 2018. "Landslide Susceptibility Mapping Using Random Forest and Boosted Tree Models in Pyeong-Chang, Korea." *Geocarto International*. 33(9):1000–1015.

Kordestani, M. D., S. A. Naghibi, H. Hashemi, K. Ahmadi, B. Kalantar, and B. Pradhan. 2019. "Groundwater Potential Mapping Using a Novel Data-mining Ensemble Model." *Hydrogeology Journal*. 27(1):211–224.

Kuhn, M., J. Wing, S. Weston, A. Williams, C. Keefer, A. Engelhardt, T. Cooper, et al., 2018. "Package "Caret" Classification and Regression Training".

Lee, S., and B. Pradhan. 2007. "Landslide Hazard Mapping at Selangor, Malaysia Using Frequency Ratio and Logistic Regression Models." *Landslides*. 4(1):33–41.

Lee, S., S. M. Hong, and H. S. Jung. 2018. "GIS-based Groundwater Potential Mapping Using Artificial Neural Network and Support Vector Machine Models: The Case of Boryeong City in Korea." *Geocarto International*. 33 (8):847–861.

Lee, S., Y. S. Kim, and H. J. Oh. 2012. "Application of a Weights-of-evidence Method and GIS to Regional Groundwater Productivity Potential Mapping." *Journal of Environmental Management*. 96(1):91–105.

Lek, S., and J. F. Guegan. 1999. "Artificial Neural Networks as a Tool in Ecological Modelling, an Introduction." *Ecological Modelling*. 120(2–3):65–73.

Magesh, N. S., N. Chandrasekar, and J. P. Soundranayagam. 2012. "Delineation of Groundwater Potential Zones in Theni District, Tamil Nadu, Using Remote Sensing, GIS and MIF Techniques." *Geoscience Frontiers*. 3(2):189–196.

Mahmoudzadeh, M. R., A. P. Francés, M. Lubczynski, and S. Lambot. 2012. "Using Ground Penetrating Radar to Investigate the Water Table Depth in Weathered granites-Sardon Case Study, Spain." *Journal of Applied Geophysics*. 79:17–26.

Maiti, S., V. C. Erram, G. Gupta, and R. K. Tiwari. 2012. "ANN Based Inversion of DC Resistivity Data for Groundwater Exploration in Hard Rock Terrain of Western Maharashtra (India)." *Journal of Hydrology*. 464:294–308. doi:10.1016/j.jhydrol.2012.07.020.

Manap, M. A., H. Nampak, B. Pradhan, S. Lee, W. N. A. Sulaiman, and M. F. Ramli. 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.

Manu, E., and K. Preko. 2014. "Estimation of Water Table Depths and Local Groundwater Flow Pattern Using the Ground Penetrating Radar." *International Journal of Scientific and Research Publications*. 4(8):1–12.

Marjanovic, M., M. Kovacevic, B. Bajat, and V. Vozenilek. 2011. "Landslide Susceptibility Assessment Using SVM Machine Learning Algorithm." *Engineering Geology*. 123(3):225–234.

Mogaji, K. A., H. S. Lim, and K. Abdullah. 2014. "Regional Prediction of Groundwater Potential Mapping in a Multifaceted Geology Terrain Using GIS-based Dempster-Shafer Model." *Arabian Journal of Geosciences*. 8(5):3235–3258.

Mojaddadi, H., B. Pradhan, H. Nampak, N. Ahmad, and A. H. B. Ghazali. 2017. "Ensemble Machine-learning-based Geospatial Approach for Flood Risk Assessment Using Multi-sensor Remote-sensing Data and GIS." *Geomatics, Natural Hazards and Risk*. 8(2):1080–1102.

Moore, I. D., and G. J. Burch. 1986. "Physical Basis of the Length-slope Factor in the Universal Soil Loss Equation 1." *Soil Science Society of America Journal*. 50(5):1294–1298.

Moore, I. D., R. B. Grayson, and A. R. Ladson. 1991. "Digital Terrain Modelling: A Review of Hydrological, Geomorphological, and Biological Applications." *Hydrological Processes*. 5(1):3–30.

Murasingh, S., R. Jha, and S. Adamala. 2018. "Geospatial Technique for Delineation of Groundwater Potential Zones in Mine and Dense Forest Area Using Weighted Index Overlay Technique." *Groundwater for Sustainable Development*. 7:387–399. doi:10.1016/j.gsd.2017.12.001.

Naghibi, S. A., D. D. Moghaddam, B. Kalantar, B. Pradhan, and O. Kisi. 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. doi:10.1016/j.jhydrol.2017.03.020.

Naghibi, S. A., and H. R. Pourghasemi. 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.

Naghibi, S. A., H. R. Pourghasemi, and B. Dixon. 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*. 188:44. doi:10.1007/s10661-015-5049-6.

Naghibi, S. A., H. R. Pourghasemi, and K. Abbaspour. 2018. "A Comparison between Ten Advanced and Soft Computing Models for Groundwater Qanat Potential Assessment in Iran Using R and GIS." *Theoretical and Applied Climatology*. 131:967–984. doi:10.1007/s00704-016-2022-4.

Naghibi, S. A., H. R. Pourghasemi, Z. S. Pourtaghi, and A. Rezaei. 2015. "Groundwater Qanat Potential Mapping Using Frequency Ratio and Shannon's Entropy Models in the Moghan Watershed, Iran." *Earth Science Informatics*. 8 (1):171–186.

Naghibi, S. A., K. Ahmadi, and A. Daneshi. 2017b. "Application of Support Vector Machine, Random Forest, and Genetic Algorithm Optimized Random Forest Models in Groundwater Potential Mapping." *Water Resources Management*. 31(9):2761–2775.

Naghibi, S. A., M. Dolatkordestani, A. Rezaei, P. Amouzegari, M. T. Heravi, B. Kalantar, and B. Pradhan. 2019. "Application of Rotation Forest with Decision Trees as Base Classifier and a Novel Ensemble Model in Spatial Modeling of Groundwater Potential." *Environmental Monitoring and Assessment*. 191:248. doi:10.1007/s10661-019-7362-y.

Nakashima, Y., H. Zhou, and M. Sato. 2001. "Estimation of Groundwater Level by GPR in an Area with Multiple Ambiguous Reflections." *Journal of Applied Geophysics*. 47 (3–4):241–249.

Nampak, H., B. Pradhan, and M. A. Manap. 2014. "Application of GIS Based Data Driven Evidential Belief Function Model to Predict Groundwater Potential Zonation." *Journal of Hydrology*. 513:283–300. doi:10.1016/j.jhydrol.2014.02.053.

Neal, A. 2004. "Ground-penetrating Radar and Its Use in Sedimentology: Principles, Problems and Progress." *Earth-Science Reviews*. 66(3–4):261–330.

Oh, H. J., Y. S. Kim, J. K. Choi, E. Park, and S. Lee. 2011. "GIS Mapping of Regional Probabilistic Groundwater Potential in the Area of Pohang City, Korea." *Journal of Hydrology*. 399 (3–4):158–172.

Patra, S., P. Mishra, and S. C. Mahapatra. 2018. "Delineation of Groundwater Potential Zone for Sustainable Development: A Case Study from Ganga Alluvial Plain Covering Hooghly District of India Using Remote Sensing, Geographic Information System and Analytic Hierarchy Process." *Journal of Cleaner Production*. 172:2485–2502.

Pourghasemi, H. R., H. R. Moradi, and S. F. Aghda. 2013. "Landslide Susceptibility Mapping by Binary Logistic Regression, Analytical Hierarchy Process, and Statistical Index Models and Assessment of Their Performances." *Natural Hazards*. 69(1):749–779.

Pourtaghi, Z. S., and H. R. Pourghasemi. 2014. "GIS-based Groundwater Spring Potential Assessment and Mapping in the Birjand Township, Southern Khorasan Province, Iran." *Hydrogeology Journal*. 22(3):643–662.

Pradhan, B., and S. Lee. 2010. "Regional Landslide Susceptibility Analysis Using Back-propagation Neural Network Model at Cameron Highland, Malaysia." *Landslides*. 7(1):13–30.

Prasad, R. K., N. C. Mondal, P. Banerjee, M. V. Nandakumar, and V. S. Singh. 2008. "Deciphering Potential Groundwater Zone in Hard Rock through the Application of GIS." *Environmental Geology*. 55(3):467–475.

Rahmati, O., A. N. Samani, M. Mahdavi, H. R. Pourghasemi, and H. Zeinivand. 2015. "Groundwater Potential Mapping at Kurdistan Region of Iran Using Analytic Hierarchy Process and GIS." *Arabian Journal of Geosciences*. 8(9):7059–7071.

Rahmati, O., H. R. Pourghasemi, and A. M. Melesse. 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.

Rajaveni, S. P., K. Brindha, and L. Elango. 2017. "Geological and Geomorphological Controls on Groundwater Occurrence in a Hard Rock Region." *Applied Water Science*. 7(3):1377–1389.

Recknagel, F. 2001. "Applications of Machine Learning to Ecological Modelling." *Ecological Modelling*. 146(1–3):303–310.

Regmi, N. R., J. R. Giardino, E. V. McDonald, and J. D. Vitek. 2015. "A Review of Mass Movement Processes and Risk in the Critical Zone of Earth." *Developments in Earth Surface Processes*. 19:319–362.

Rodriguez-Galiano, V. F., and M. Chica-Rivas. 2014. "Evaluation of Different Machine Learning Methods for Land Cover Mapping of a Mediterranean Area Using Multi-seasonal Landsat Images and Digital Terrain Models." *International Journal of Digital Earth*. 7(6):492–509.

Samui, P. 2008. "Slope Stability Analysis: A Support Vector Machine Approach." *Environmental Geology*. 56(2):255.

Schapire, R. E. 2003. "The Boosting Approach to Machine Learning: An Overview." In *Nonlinear Estimation and Classification*. Denison, D.D., Hansen, M.H., Holmes, C.C., Mallick, B., Yu.B. (eds.) New York: Springer; p. 149–171. doi:10.1007/978-0-387-21579-2_9.

Shafizadeh-Moghadam, H., R. Valavi, H. Shahabi, K. Chapi, and A. Shirzadi. 2018. "Novel Forecasting Approaches Using Combination of Machine Learning and Statistical Models for Flood Susceptibility Mapping." *Journal of Environmental Management*. 217:1–11. doi:10.1016/j.jenvman.2018.03.089.

Shih, S. F., J. A. Doolittle, D. L. Myhre, and G. W. Schellentrager. 1986. "Using Radar for Groundwater Investigation." *Journal of Irrigation and Drainage Engineering*. 112(2):110–118.

Tahmassebipoor, N., O. Rahmati, F. Noormohamadi, and S. Lee. 2016. "Spatial Analysis of Groundwater Potential Using Weights-of-evidence and Evidential Belief Function Models and Remote Sensing." *Arabian Journal of Geosciences*. 9:79. doi:10.1007/s12517-015-2166-z.

Tehrany, M. S., B. Pradhan, and M. N. Jebur. 2014. "Flood Susceptibility Mapping Using a Novel Ensemble Weights-of-evidence and Support Vector Machine Models in GIS." *Journal of Hydrology*. 512:332–343. doi:10.1016/j.jhydrol.2014.03.008.

Tehrany, M. S., B. Pradhan, S. Mansor, and N. Ahmad. 2015. "Flood Susceptibility Assessment Using GIS-based Support Vector Machine Model with Different Kernel Types". *Catena*. 125:91–101. doi:10.1016/j.catena.2014.10.017.

Yao, X., L. G. Tham, and F. C. Dai. 2008. "Landslide Susceptibility Mapping Based on Support Vector Machine: A Case Study on Natural Slopes of Hong Kong, China." *Geomorphology*. 101(4):572–582.

Youssef, A. M., H. R. Pourghasemi, Z. S. Pourtaghi, and M. M. Al-Katheeri. 2016. "Landslide Susceptibility Mapping Using Random Forest, Boosted Regression Tree, Classification and Regression Tree, and General Linear Models and Comparison of Their Performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia." *Landslides*. 13(5):839–856.