Application of Ensemble Machine Learning Models to Assess the Sub-regional Groundwater Potentiality: A GIS-Based Approach

15

Sunil Saha, Amiya Gayen, and Sk. Mafizul Haque

Abstract

Effective data mining models are powerful tools for the prediction and management of sub-regional groundwater resources. In this work, an integrated attempt is employed to assess the groundwater potentiality in C. D. Block of Birbhum District, India using GIS-based novel ensemble machine learning models of Radial Basis Function neural network (RBFnn) in form of RBFnn-Bagging and RBFnn-Dagging. Fourteen hydro-geomorphological factors were used to find the most potential groundwater area. To support the result, observation data of 86 sites were incorporated empirically. Out of these, 70% were randomly split for the training dataset to develop the model and remaining 30% were used for model validation. Results predict excellent groundwater potentiality by the RBFnn-Bagging and RBFnn-Dagging as they covered 17.38% and 13.97% of the study area, respectively. The prediction capacity of newly built models was established with the root mean square error (RMSE), accuracy, precision, and receiver operating characteristic (ROC) curve which shows a satisfactory result as the RMSE values of 0.05 and 0.07 and AUC values of 82.1% and 81.30% are obtained for RBFnn-Bagging and RBFnn-Dagging models respectively. Well-known mean decrease Gini (MDG) from the random forest (RF) algorithm, implemented to determine the relative importance of the factors, reveals that distance from river, pond frequency, aspect, stream junction frequency, elevation, and geomorphology are most useful determinants of groundwater potentiality in the study area. The adopted approach has a wide scope in effective planning and sustainable management of groundwater resources.

15.1 Introduction

Drinking water crisis and groundwater scarcity are major challenges among the various prevailing contemporary issues of the earth. Groundwater is the most important but fast depleting natural resource whose appropriate delineation and management are momentous at this

S. Saha
Department of Geography, University of Gour Banga, Malda, India
e-mail: sunilgeog@ugb.ac.in

A. Gayen (✉) · Sk.M. Haque
Department of Geography, University of Calcutta, Kolkata, India
e-mail: amiya.gayen@midnaporecollege.ac.in

conjuncture. India is the most groundwater-consuming country in the world, which uses nearly 230 km^3 year^{-1} of groundwater (World Bank 2010). According to the World Bank report of 2010, if India does not reduce the use of groundwater, more than 60% of the aquifers will be dried within 20 years. In India, demand for groundwater has been increasing through the green revolution and the pace of industrialization, urbanization, and agricultural practices (Suhag 2016). There are two different types of aquifers in India, i.e., crystalline aquifers (located in peninsular area) and another are alluvial aquifers (developed in the Indo-Gangetic plain). The former is characterized by low permeability and hard rocks and the latter leads in terms of groundwater resources (Suhag 2016). Thus, groundwater quality and potentiality assessment are important tasks at hand for reasons of sustainability and livelihood.

Therefore, most of the aquifers are in critical situations, particularly in semi-arid and arid regions, which may turn into a severe problem. Several researchers have tried to determine aquifer characteristics with the help of sediments beneath, identifying pore space, and fractures in a rock on the earth's surface, which is not adequate for identifying reliable aquifers (Naghibi et al. 2017). Generally, groundwater potentiality assessment including hydro-geological nature of the region especially porosity, aquifer properties, permeability, storage capacity, groundwater recharge, and hydraulic conductivity of the aquifer materials are very pertinent factors. These are broadly dependent on physical variables like geomorphology, geology, rainfall, soil, drainage, and LULC (Saha 2017; Haque et al. 2020). Presently, the unnecessary use of groundwater and unscientific management strategies are affecting the groundwater recharge level (Chaudhry et al. 2019). Therefore, in such circumstances, it is required that an adequate management strategy for groundwater potentiality assessment is framed (Chen et al. 2019). Thus, a groundwater potential map can help to identify the prospect of groundwater yield, which can guide toward proper management of groundwater.

Different popular and well-accepted models have been developed for preparing groundwater potentiality mapping (Corsini et al. 2009; Ozdemir 2011; Lee et al. 2017; Saha 2017; Chen et al. 2019). For example, the analytical hierarchy process (Razandi et al. 2015; Ghosh et al. 2020), the weight of evidence (Tahmassebipoor et al. 2016), and frequency ratio (Guru et al. 2017; Das 2019), fuzzy logic (Mohamed and Elmahdy 2017). Nowadays those models were not applied by researchers because they are unable to solve multi-criteria decision problems. An examination of the literature reveals that the integration of machine learning models has provided better results (Kenda et al. 2018; Chen et al. 2019). So, machine learning models handle data with high dimensionality and provide more perfect results using geographical information systems and remote sensing data (Gayen et al. 2019; Rudin 2019; Haque et al. 2020). Guzman et al. (2015) applied artificial neural network (ANN) and support vector machine (SVM) to predict groundwater potentiality. Guzman et al. (2015) have explained the superiority of the SVM models over ANN models about prediction. Naghibi et al. (2018) also applied some well-accepted machine learning models, i.e., boosted regression tree, classification and regression tree, and random forest for groundwater potentiality prediction. Their study shows that the boosted regression tree model provides a better result with an AUC value of 0.8103. Sajedi-Hosseini et al. (2018) also implemented a few machine learning models for groundwater risk assessment. Thus, the previous research work confirms the prediction capacity of machine learning models to predict groundwater potentiality. The present study has focused on novel ensemble machine learning models of Radial Basis Function neural network (RBFnn)- Bagging (RBFnn-Bagging) and Dagging (RBFnn-Dagging). The primary objective of this research is to prepare a groundwater potentiality map, along with groundwater quality of Md. Bazar Block in Birbhum District, India. Finally, researchers have tried to predict the groundwater controlling efficiency of the applied factors with mean decrease Gini (MDG).

15.2   Materials and Methods

15.2.1   Study Area

The Md. Bazar is a Jharkhand adjacent western block of Birbhum District located in West Bengal, India. It is extended from 87°25′ E to 87°40′ E and 23°55′ N to 24°50′ N (Fig. 15.1). This block was recognized as drought influenced district of West Bengal. This region is formed of gneisses and associated rocks, older alluvium, and older alluvium with lateritic types of aquifer media. The older alluvium has high to moderate yield potentiality but in the cases of older alluvium with laterite rocks, the yield potentiality is limited between 100 and 700 gpd ft^{-2} hydraulic conductivity in the study area (Thapa et al. 2018). This falls under the warm monsoon climate where annual precipitation is approximately 1200 mm and temperature ranges from 6 to 40 °C (Saha 2017). The maximum precipitation occurs from July to September (monsoon period). The long gap of the rainy season and over-increasing pressure of agriculture leads to continuous updraft of groundwater for irrigation which is one of the major issues of this region. The main routes of groundwater recharge in Md. Bazar block is natural and anthropogenic activities such as artificial canals, hydropower dams, and check dams.

15.2.2   Data Used

In the first instance, dug wells locations were collected from the Central Ground Water Board. A total of 85 dug wells and one piezometer were recognized in Md. Bazar Block of Birbhum District and verified using GPS and field survey and considered for a groundwater inventory map (CGWB, 2017). After that, the well and no-well locations were classified into two sets by maintaining 70:30 ratio. 70% of locations were used as training dataset which was applied to predict the GWPMs. At the same time, the unused 30% locations were considered as a validation dataset

of the modeling result (Naghibi et al. 2017; Chen et al. 2019).

Fourteen groundwater controlling factors viz., aspect, elevation, curvature, topographical positioning index (TPI), topographical wetness index (TWI), slope, stream junction frequency (SJF), geomorphology, distance to a river, rainfall, pond frequency, land use\land cover (LULC), geology, and soil texture were selected for the development of the GWPMs (Fig. 15.2). Thematic data layers of parameters were prepared using the GIS-spatial analysis tool and the PALSAR Digital Elevation Model (DEM) was taken from the Alaska Satellite facility; LULC map was developed by applying the Sentinal-2 data; rainfall data from Indian Meteorological Department (IMD); soil map from NBSS-LUP; and the geological map was collected from Geological Survey of India (GSI).

15.2.3   Preparing Groundwater Influencing Factors

At first, 12.5 × 12.5 m spatial resolution based PALSAR-DEM data was used to prepare the aspect, elevation, curvature, TWI, and TPI maps (Fig. 15.3a–e). Because these parameters are considered by several researchers (Naghibi et al. 2016, 2017; Chen et al. 2019) to be an essential parameters of the GWPM. Aspect and elevation both are associated with soil moisture, sunlight, temperature, wind, soil development, and precipitation therefore both factors can enhance the rate of groundwater recharge (Golkarian et al. 2018; Gayen et al. 2019). The slope is an important terrain factor that increases the velocity of surface runoff wherein a high slope does not allow infiltration of groundwater (Arabameri et al. 2019). The regional slope angle ranges from 0° to 34.21°. The TWI is applied for measuring the influence of topological conditions on hydro-geomorphic processes. It is the integration of slope and the upstream contributing area per unit orthogonal to the direction of flow (Arabameri et al. 2019). The calculation of TWI is represented in Moore et al. (1991):
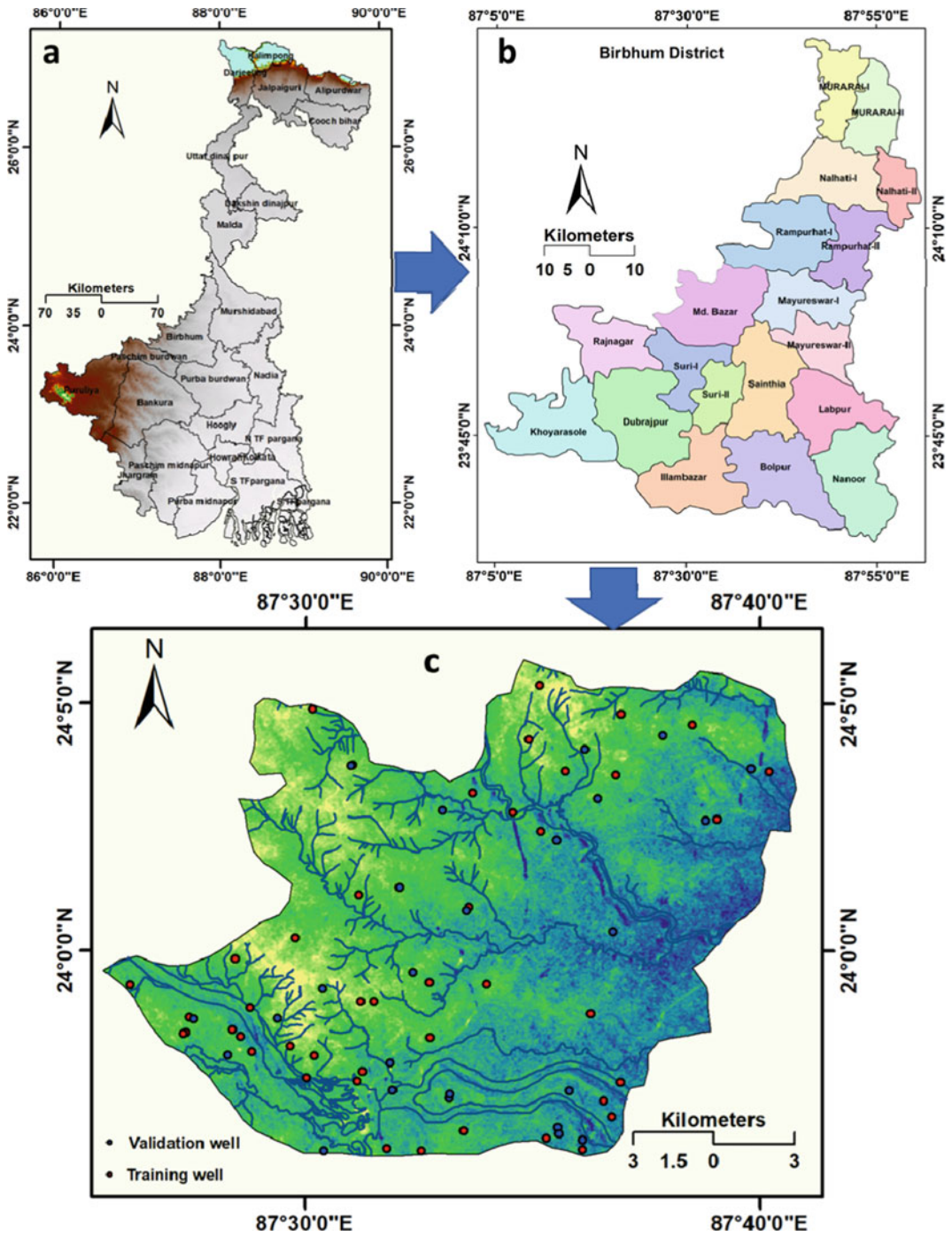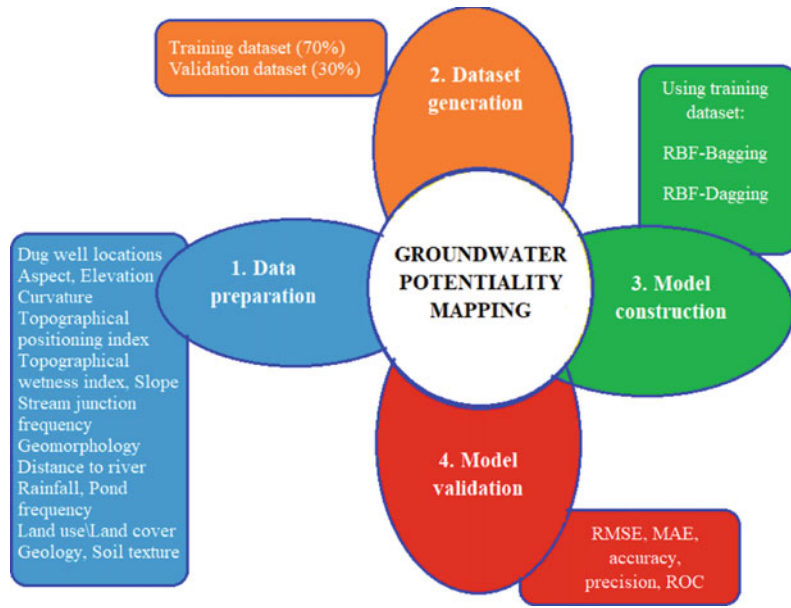
Fig. 15.1 Location map of the study area; **a** West Bengal **b** Birbhum District of West Bengal **c** Md. Bazar block

Fig. 15.2 Flowchart illustrated the applied methodology of groundwater potentiality mapping

$$TWI = \ln\left(\frac{As}{\tan\beta}\right) \qquad (15.1)$$

where, As denotes cumulative catchment area ($m^2\ m^{-1}$) and β defines the slope angle.

The TPI and curvature both are exhibited to affect groundwater potentiality (Grohmann and Riccomini 2009; Arabameri et al. 2019). The TPI and curvature maps were developed with the help of PALSAR-DEM data. The TPI has been calculated by using Eq. (15.2).

$$TPI = Z_0 - \overline{Z} \qquad (15.2)$$

$$\overline{Z} = \frac{1}{n_R} \sum_{i\in R} Z_i \qquad (15.3)$$

where Z_0 denotes the central point altitude, Z represents the mean altitude within a particular radius (R), and small R defines small ridges and valleys (Weiss 2001). The highest and lowest TPI values within the study area are 0.00 and 1.00.

Pond, drainage, and stream dictate structural characteristics and permeability of an area that influences groundwater storage and movement through a hydraulic gradient (Tien Bui et al. 2017). The distance to river and stream junction frequency maps were developed using the extracted drainage output from the 1:50,000

toposheet maps. Junction indicates the confluence areas of two rivers. Generally, chances of groundwater arability are more in the highest pond frequency areas and confluence zone areas because both are enhancing groundwater recharge processes. The LULC can reflect less susceptibility to groundwater potentiality (Saha 2017). A LULC map of study area was developed using Sentinal-2 data and results were affirmed by applying Cohen's Kappa index with 89.6% Kappa value. The Block is covered by eight LULC classes: reservoir, watercourse, sand cover, settlement, agricultural land, mining area, wasteland, and vegetation cover (Fig. 3l). The duration of the Rainfall and its intensity also play a key role in groundwater recharge (Shekhar and Pandey 2014). Jothibasu and Anbazhagan (2016) noted that rainfall influences GPM accuracy and moving water percolation for that reason spatial distribution of rainfall was taken as a predisposing factor for this study (Fig. 3j).

Soil types are most important predisposing factors for the assessment of the infiltration rate of any region. This study area falls under six major soil types like sandy, clay loam, loamy, sandy loam, sandy clay, and sandy clay loam. Maximum areas of Md. Bazar block is covered by sandy loam and clay loam soil types (Fig. 3n). Generally, potentiality of groundwater infiltration rate is
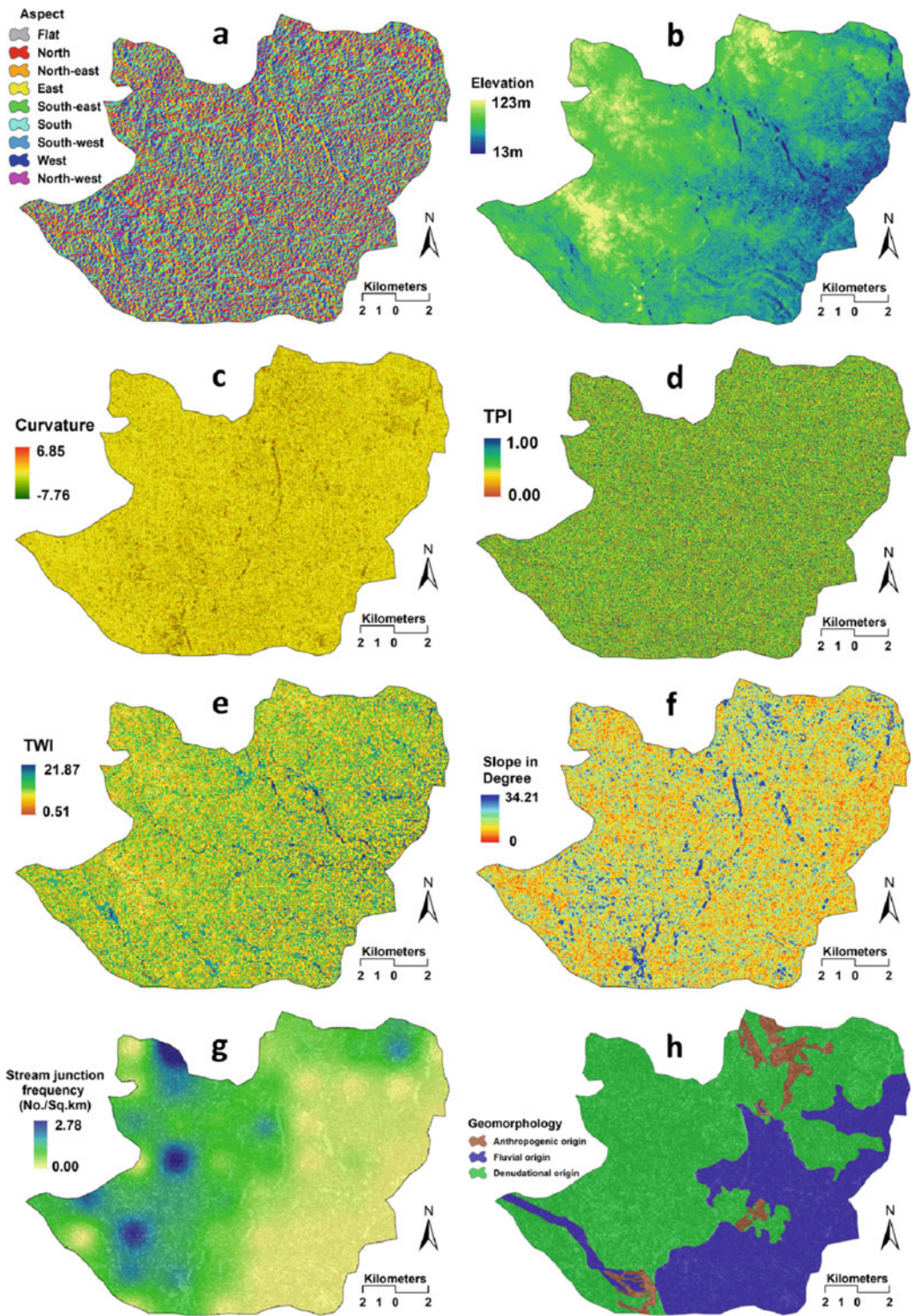
Fig. 15.3 The spatial data layers: **a** aspect, **b** elevation, **c** curvature, **d** TPI, **e** TWI, **f** slope, **g** stream junction frequency, **h** geomorphology, **i** distance to river, **j** rainfall, **k** pond frequency, **l** LULC, **m** geology, **n** soil texture
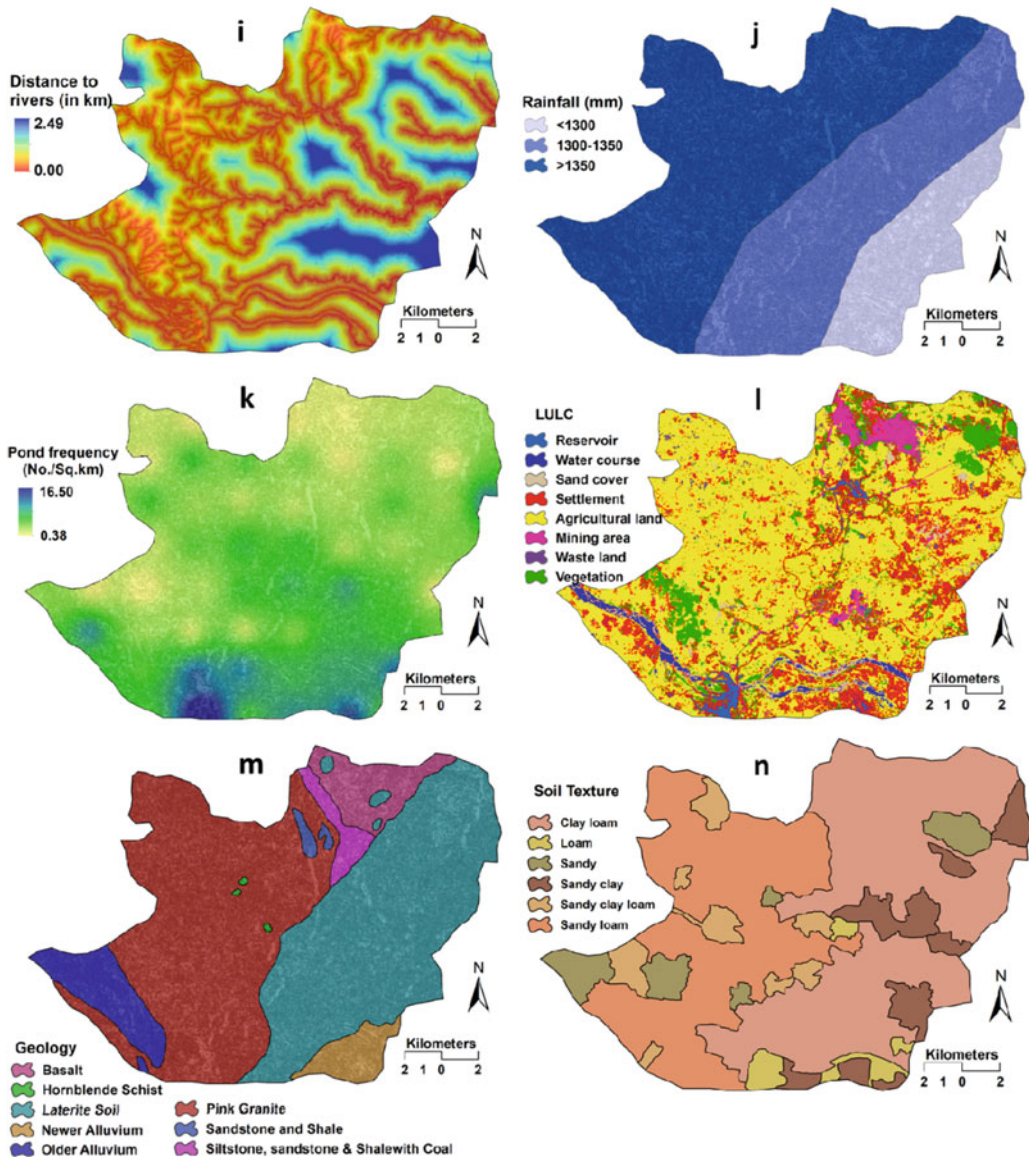
Fig. 15.3 (continued)

higher in sandy regions as compared to loamy or clayey strata. The Md. Bazar block is composed of eight geological formations. The western part is dominated by pink granite whereas rocks belonging to the Vindhyan formation occur to the east (Fig. 3m). The pisolitic and kankar ferruginous concretions are mostly found in the laterite track. Some parts of the block are covered by basaltic rocks and younger alluvium. The block falls under three primary geomorphological regions, i.e., depositional plain, anthropogenic origin, and denudational plain (Fig. 3h).

15.2.4 Machine Learning Ensemble Meta-classifiers Modes for the GWPMs

Novel ensemble models, the RBFnn-Bagging and RBFnn-Dagging, are used for mapping groundwater potentiality in this study. RBFnn originated in the late 1980s is a version of an artificial neural network. In a two-layer neural network, where each hidden unit implements a radial-activated function, RBFs are embedded. A weighted sum of hidden unit outputs is implemented by output units. Although the output is linear, the input into an RBF network is nonlinear. Their exceptional approximation capacities are investigated. RBF networks can model complex mappings due to their nonlinear approximation properties. The RBFnn was used as a base learner in this study. As for the ensemble technique, because of its utility in ensemble estimation, the Bagging and Dagging were applied as the meta-learner.

15.2.4.1 Bagging

The bagging algorithm has introduced by Breiman (1996), is the developer of bootstrapping (Freedman 1981). Several researchers have applied this model to predict susceptibility maps (i.e., flood, landslide, etc.) as this model has excellent performance ability (Hong et al. 2020). The bagging tree is a bagging algorithm comprised of models based on decision trees. This algorithm is selected because it fabricates the decision tree with the help of each produced subset and ultimately, they are assembled within the final model (Hong et al. 2020). It enhances the alignment accuracy by minimizing the inconsistency of the alignment error (Saha et al. 2021; Wu et al. 2020). A bagging classifier is considered a three-step bagging system (Breiman 1996; Yariyan et al. 2020). It is developed as a bootstrap sample through substantive training samples through the displacement approach (Saha et al. 2021). This MLA can promote the success of all arrays of subset by connecting them to the actual feature process for the bagging

classification stage; also, this model is not dependent upon the precision of past models (Breiman 1996; Yariyan et al. 2020).

15.2.4.2 Dagging

The Dagging algorithm was introduced by Ting and Witten (1997), using another sampling method to extract a basic classifier. Dagging is very similar to bagging—name is a portmanteau derived from the phrase "disjoint bagging." In dagging, once data is used for classification the subset is "disjointed" (or set aside). In bagging, each subset is not disjointed and the data is returned to the full set to be used again. Dagging is a well-known group-sampling technique using majority votes to combine several classifiers to improve prediction accuracies of basic classifiers (Kotsianti and Kanellopoulos 2007).

15.2.5 Validation of Groundwater Potentiality Models

Models' evaluation and validation is an important steps in prediction work and without validation, the model does not have any scientific significance (Talukdar and Pal 2020; Pal and Mandal 2021). The applied model's prediction capacity was investigated by ROC curve, RMSE, MAE, accuracy, and precision (Chen et al. 2018). The two categories of ROC curve on prediction and success rate, are developed using validation and training datasets, respectively. It is a graphical illustration of model prediction through a diagnostic test (Chen et al. 2019). The area under the curve (AUC) varies from 0.5 to 1.0 and the value close to 1.0 predicts the power of models (Mishra et al. 2020).

Also, error within the predictive models was calculated through RMSE and MAE tests to identify the prediction capacity (Abedinpour et al. 2012). Each error was calculated with the comparison between model values and field observed values (Rahmati et al. 2017). The precision, RMSE, MAE, and AUC have been calculated by using Eqs. (15.4)–(15.7).

$$\text{Precision} = \frac{TP}{TP + FP} \qquad (15.4)$$

$$\text{RMSE} = \sqrt{\frac{1}{(N)}\sum_{i=1}^{N}(O_i - S_i)^2} \qquad (15.5)$$

$$\text{MAE} = \sqrt{\sum_{i=1}^{n}\frac{(S_i - O_i)}{n}} \qquad (15.6)$$

$$\text{AUC} = \frac{\Sigma TP + \Sigma TN}{P + N} \qquad (15.7)$$

where TN and TP denote true negative and true positive, FP and FN denote false positive and false negative, O_i and S_i are observed and predicted values, n is the number of observations, P and N are the dug wells location points, and N is the total number of non-dug wells location points.

15.3   Results and Analysis

15.3.1   Groundwater Potentiality Models

At first, two accepted meta classifier based MLAs were developed by applying the training dataset. The constructed models were divided into four classes (i.e., high, very high, moderate, and low) to calculate the groundwater potentiality indices (GWPI) (Chen et al. 2018) (Fig. 4a, b). Actually, the user-defined classification of GWPMs is nearly hard for readers to justify and interpret. Therefore, nature break statistics were most convenient for the arrangement of GWPI following the histogram of data distribution (Chen et al. 2019).

The RBFnn-Bagging produced result shows that low potentiality zone has the maximum area (68.64%), followed by the very high (16.92%), moderate (11.51%), and high (2.92%) in the study area. The corresponding area covered by RBFnn-Dagging mode is 68.03%, 13.70%, 13.59%, and 4.68% for the low, very high, moderate, and high zones, respectively. It is manifest through both models GWPMs; the largest GWP area is found in the southern part of the Md Bazar Block because of the more forest cover and presence of water reservoir (Table 15.1).

### 15.3.2   Validation and Comparison of Applied Models

For validation and comparing the applied models; RMSE, MAE, accuracy, precision, and ROC were implemented using validation and training data sets (Fig. 5a, b), as they are important aspects to conclude the prediction capacity of applied models (Pal and Mandal 2021).

The results show that the RBFnn-Bagging algorithm has higher AUC values of 0.837 and 0.847, respectively, for the success and prediction rate curves, followed by the RBFnn-Dagging algorithm with an AUC value of 0.793 and 0.829, respectively. So, it is concluded that both models have excellent GWP prediction capacity. The RMSE and MAE values of RBFnn-Bagging and RBFnn-Dagging were calculated for the training phase as 0.237, 0.057, 0.270, and 0.74 and validation phase as 0.039, 0.198, 0.51, and 0.227, respectively (Table 15.2). Also, results of accuracy and precision tests are presented in Table 15.2 for both the applied models. The accuracy and precision values of both models were 0.88, 0.79, 0.83, and 0.75 for RBF-Bagging and RBFnn-Dagging, respectively which indicates that both the models have uniform prediction capacity for assessment of groundwater potentiality.

15.3.3   Significant Factors Identification by MDGs

The significant factors identification is a challenging task because groundwater recharge is impacted by various groundwater controlling factors (Conforti et al. 2010). The mean decrease Gini was applied to evaluate factor's relative importance by using the random forest (RF) algorithm (Breiman 2001). The MDG varies from 14.09 to 286.01. Distance to a river (286.01), pond frequency (229.96), aspect (103.56), stream

Fig. 15.4 The groundwater
potentiality maps by RBF-
Bagging and RBF-Dagging
models
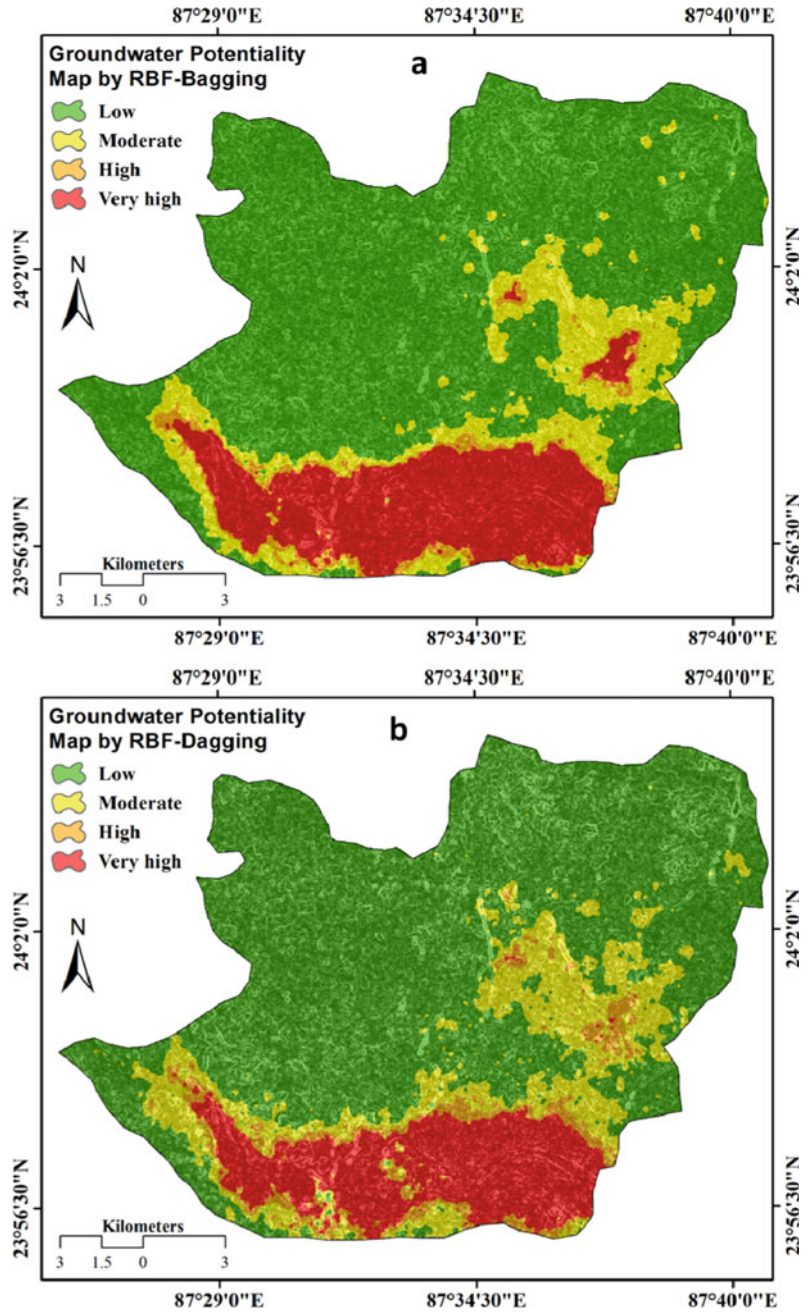
Table 15.1 Areal share
under potentiality classes of
groundwater potentiality
models (area in km^2)

GWP classes	RBF-Bagging	RBF-Dagging
Low	68.64	68.03
Moderate	11.51	13.59
High	2.92	4.68
Very high	16.92	13.70

Fig. 15.5 Validation of groundwater potentiality maps applying ROC curve: **a** success rate curve (applying training dataset) and **b** prediction rate curve (applying validation dataset)
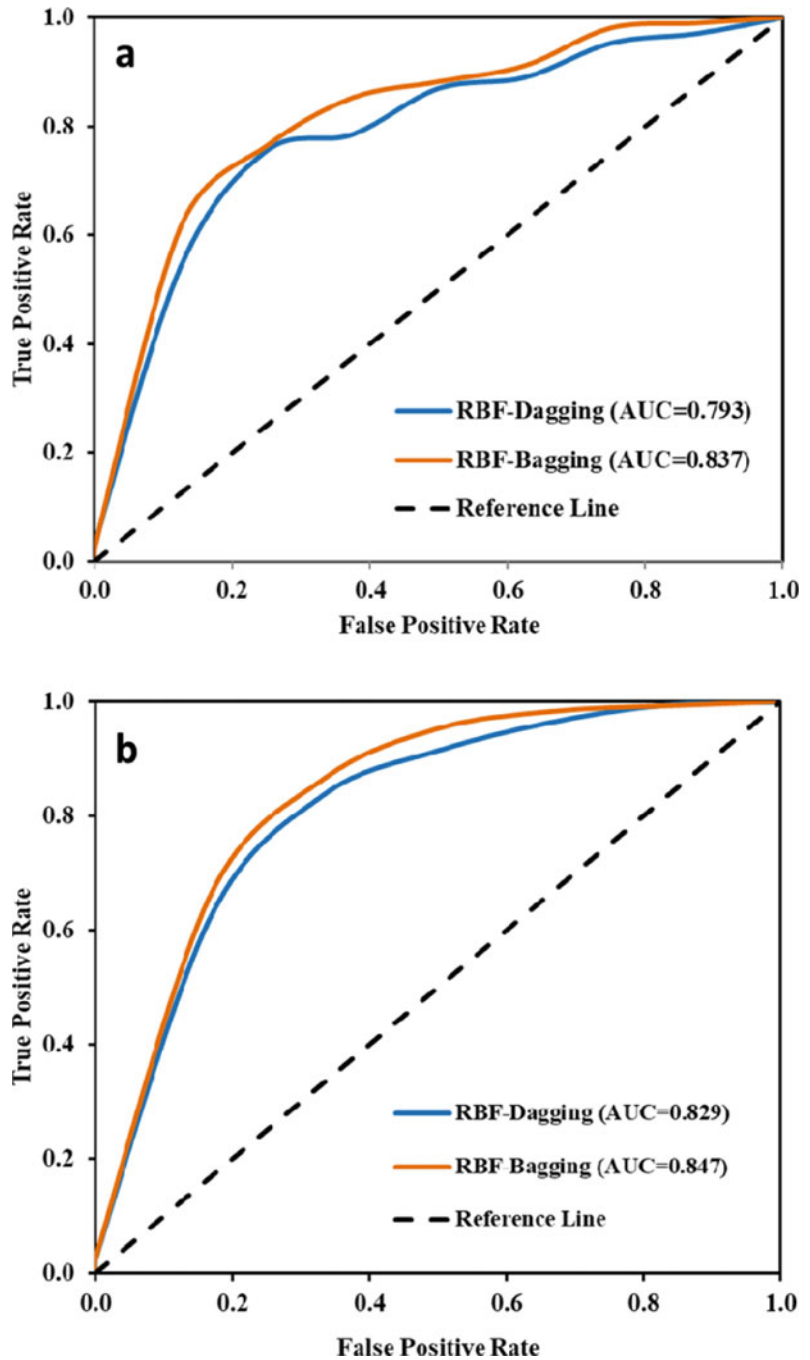
Table 15.2 Estimation of root mean square error (RMSE), MAE, accuracy, and precision for both models

Models	Training dataset				Validation dataset			
	RMSE	MAE	Accuracy	Precision	RMSE	MAE	Accuracy	Precision
RBF-Bagging	0.237	0.057	0.85	0.81	0.039	0.198	0.88	0.83
RBF-Dagging	0.270	0.074	0.71	0.74	0.051	0.227	0.79	0.75
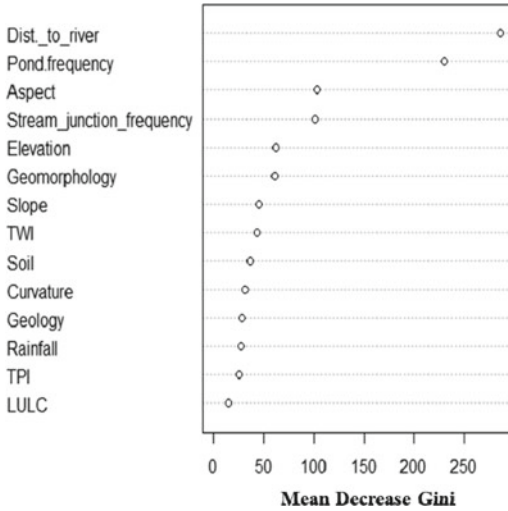
Fig. 15.6 Significant factors identification by mean decrease Gini

Table 15.3 Calculated MDGs values for significant factors identification

Factors	Mean decrease Gini
Curvature	31.45
TPI	24.99
TWI	42.94
Slope	45.32
Distance to river	286.01
Pond frequency	229.96
Aspect	103.56
Elevation	62.06
Stream junction frequency	101.45
Rainfall	26.88
Geology	28.38
Land use/land cover	14.09
Soil types	36.11
Geomorphology	61.10

junction frequency (101.45), elevation (62.06), and geomorphology (61.10) were the most important factors. These were followed in order of influence by the slope (45.32), TWI (42.94), soil types (36.11), curvature (31.45), geology (28.38), rainfall (26.88), TPI (24.99), and LULC (14.09) (Fig. 15.6 and Table 15.3). All the fourteen predisposing factors were subjected to the modelling—purpose because all are contributors to GWP occurrence.

15.4 Discussion

For the groundwater potentiality (GWP) assessment factors like rainfall, land use, slope, elevation, pond frequency, stream junction frequency, distance to a river, TWI, soil texture, geology, geomorphology, curvature, and aspect are used. The elevation and slope are very low in the south-eastern portion of Md. Bazar block. Recharge of groundwater is negatively related to the elevation of study area. Thus, locations that are situated in low elevation areas show high groundwater potentiality at a particular region within the study area rather than being uniformly distributed across it.

In other works (e.g., Corsini et al. 2009; Ozdemir 2011; Rahmati et al. 2017; Naghibi

et al. 2017; Chen et al. 2019), similar factors have been used for assessing GWP and the applied relation between the factor used and the wells are also found to be the same. Usually, there is no algorithm with an extreme prediction capacity that works completely as natural processes, and groundwater modeling is a complex and nonlinear process and cannot be based on normal models with a linear structure (Chen et al. 2019). Several researchers have applied MLAs like Bagging and Dagging, in various fields of research, like gully erosion, landslide, flood hazard, and deforestation susceptibility assessment (Chen et al. 2018, 2019; Arabameri et al. 2020; Hong et al. 2020; Pal et al. 2020; Talukdar et al. 2020; Saha et al. 2021). In every case, prediction capacity of the meta classifier ensemble model's results was extremely appreciable. So, the application of machine learning algorithms (MLAs) is not a new thing, but the implication of these machine learning meta-classifiers models for groundwater potentiality assessment is unique.

Previous research work like that by Corsini et al. (2009), Ozdemir (2011), Lee et al. (2017), Naghibi et al. (2018), concluded that MLAs provided adequate results with respect to

multivariate and bivariate statistical models. In other studies, like floods, landslides, and assessments of spring potential, the RBFnn-Bagging model has also given good results. In the sense that no overfeeding of data is executed, the RBFnn-Bagging model is the most important. It consists of multiple decision trees with an interaction between predisposing factors and non-linearity (Hong et al. 2020; Saha et al. 2021). The results also revealed that the processing speed of RBFnn-Bagging is much higher concerning RBFnn-Dagging mode, which means assignment of input factors is very important. As a matter of fact, concerning percentage of the low and high GWP zones, two models displayed a uniform spatial distribution. So, the RBF-Bagging and RBFnn-Dagging models can be applied for hazard vulnerability and susceptibility mappings such as flood, landslide, forest fire, and gully erosion at a local and regional scale.

15.5   Conclusion

Groundwater potential mapping, applying various predisposing factors, is an important aspect of groundwater research. For the accurate experiment of groundwater conditions, several algorithms have been applied around the globe. In this study, a well-accepted methodology was applied to delineate GWP zones in Md. Bazar Block. After critically evaluating the study, fourteen predisposing factors were overlaid with RBF-Bagging and RBFnn-Dagging models. The RBFnn-Bagging and RBFnn-Dagging models identified 16.92 and 13.70% of areas with very high groundwater potentiality and 68.64 and 68.03% of the block with low groundwater potentiality. The results alert that this block may face vulnerable conditions in the future if the government back-steps from introducing various schemes (i.e., rainwater harvesting, dam construction, etc.) and generating awareness among common people. Based on the experiment results the following conclusions can be summarized. First, the RBFnn-Bagging model has better prediction capacity than the RBFnn-Dagging model because the Bagging algorithm

can be applied to find out reliable features of the real data. Second, researchers can solve the model overfitting problems by applying the RBF-Bagging model. Third, based on the mean decrease Gini, the most effective factors of groundwater potentiality are the distance to a river, pond frequency, aspect, stream junction frequency, elevation, and geomorphology, respectively. Finally, this proposed approach should be useful for the exploration, development, and management of groundwater. At the outset, it is pertinent that groundwater recharge processes along with their management are taken at the earliest in Md. Bazar Block.

References

Abedinpour M, Sarangi A, Rajput TBS, Singh M, Pathak H, Ahmad T (2012) Performance evaluation of AquaCrop model for maize crop in a semi-arid environment. Agric Water Manag 110:55–66

Arabameri A, Chen W, Blaschke T, Tiefenbacher JP, Pradhan B, Tien Bui D (2020) Gully head-cut distribution modeling using machine learning methods—a case study of NW Iran. Water 12(1):16

Arabameri A, Roy J, Saha S, Blaschke T, Ghorbanzadeh O, Tien Bui D (2019) Application of probabilistic and machine learning models for groundwater potentiality mapping in Damghan Sedimentary Plain, Iran. Remote Sens 11(24):3015

Breiman L (1996) Bagging predictors. Mach Learn 24:123–140

Breiman L (2001) Random forests. Mach Learn 45(1):5–32

Chaudhry AK, Kumar K, Alam MA (2019) Mapping of groundwater potential zones using the fuzzy analytic hierarchy process and geospatial technique. Geocarto Int 1–22

Chen W, Panahi M, Khosravi K, Pourghasemi HR, Rezaie F, Parvinnezhad D (2019) Spatial prediction of groundwater potentiality using ANFIS ensembled with teaching-learning-based and biogeography-based optimization. J Hydrol 572:435–448

Chen W, Shahabi H, Zhang S, Khosravi K, Shirzadi A, Chapi K, Pham BT (2018) Landslide susceptibility modeling based on GIS and novel bagging-based kernel logistic regression. Appl Sci 8(12):2540

Conforti M, Aucelli PP, Robustelli G, Scarciglia F (2010) Geomorphology and GIS analysis for mapping gully erosion susceptibility in the Turbolo stream catchment (Northern Calabria, Italy). Nat Hazards 56(3):881–898

Corsini A, Cervi F, Ronchetti F (2009) Weight of evidence and artificial neural networks for potential groundwater spring mapping: an application to the Mt.

Modino area (Northern Apennines, Italy). Geomorphology 111(1–2):79–87

Das S (2019) Comparison among influencing factor, frequency ratio, and analytical hierarchy process techniques for groundwater potential zonation in Vaitarna basin, Maharashtra, India. Groundw Sustain Dev 8:617–629

Freedman DA (1981) Bootstrapping regression models. Ann Stat 9:1218–1228

Gayen A, Pourghasemi HR, Saha S, Keesstra S, Bai S (2019) Gully erosion susceptibility assessment and management of hazard-prone areas in India using different machine learning algorithms. Sci Total Environ 668:124–138

Ghosh D, Mandal M, Banerjee M, Karmakar M (2020) Impact of hydro-geological environment on availability of groundwater using analytical hierarchy process (AHP) and geospatial techniques: a study from the upper Kangsabati river basin. Groundw Sustain Dev 11:100419

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

Grohmann CH, Riccomini C (2009) Comparison of roving-window and search-window techniques for characterising landscape morphometry. Comput Geosci 35:2164–2169

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

Guzman SM, Paz JO, Tagert MLM, Mercer A (2015) Artificial neural networks and support vector machines: contrast study for groundwater level prediction. In: Proceedings of the 2015 ASABE annual international meeting

Haque SM et al (2020) Identification of groundwater resource zone in the active tectonic region of Himalaya through earth observatory techniques. Groundw Sustain Dev 10. https://doi.org/10.1016/j.gsd.2020.100337

Hong H, Liu J, Zhu AX (2020) Modeling landslide susceptibility using LogitBoost alternating decision trees and forest by penalizing attributes with the bagging ensemble. Sci Total Environ 718:137231

Jothibasu A, Anbazhagan S (2016) Modeling groundwater probability index in Ponnaiyar River basin of South India using analytic hierarchy process. Model Earth Syst Environ 2:109

Kenda K, Čerin M, Bogataj M, Senožetnik M, Klemen K, Pergar P, Laspidou C, Mladenić D (2018). Groundwater modeling with machine learning techniques: Ljubljana polje Aquifer. Proceedings 2:697

Kotsianti SB, Kanellopoulos D (2007) Combining bagging, boosting and dagging for classification problems. In: International conference on knowledge-based and intelligent information and engineering systems, pp 493–500. Springer

Lee S, Hong SM, Jung HS (2017) GIS-based groundwater potential mapping using artificial neural network and support vector machine models: the case of Boryeong city in Korea. Geocarto Int 1–33

Mishra SV, Gayen A, Haque SM (2020) COVID-19 and urban vulnerability in India. Habitat Int 103:102230

Mohamed MM, Elmahdy SI (2017) Fuzzy logic and multi-criteria methods for groundwater potentiality mapping at Al Fo'ah area, the United Arab Emirates (UAE): an integrated approach. Geocarto Int 32 (10):1120–1138

Moore ID, Grayson RB, Ladson AR (1991) Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. Hydrol Process 5 (1):3–30

Naghibi SA, Ahmadi K, Daneshi A (2017) Application of support vector machine, random forest, and genetic algorithm optimized random forest model in groundwater potential mapping. Water Resour Manag 31 (9):2761–2775

Naghibi SA, Pourghasemi HR, Dixon B (2016) GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran. Environ Monit Assess 188(1):1–27

Naghibi SA, Pourghasemi HR, Abbaspour K (2018) A comparison between ten advanced and soft computing models for groundwater qanat potential assessment in Iran using R and GIS. Theor Appl Climatol 131 (3):967–984

Ozdemir A (2011) Using a binary logistic regression method and GIS for evaluating and mapping the groundwater spring potential in the Sultan Mountains (Aksehir, Turkey). J Hydrol 405(1):123–136

Pal S, Mandal I (2021) Noise vulnerability of stone mining and crushing in Dwarka river basin of Eastern India. Environ Dev Sustain 1–22

Pal SC, Arabameri A, Blaschke T, Chowdhuri I, Saha A, Chakrabortty R, Lee S, Band S (2020) Ensemble of machine-learning methods for predicting gully erosion susceptibility. Remote Sens 12(22):3675

Rahmati O, Tahmasebipour N, Haghizadeh A, Pourghasemi HR, Feizizadeh B (2017) Evaluating the influence of geo-environmental factors on gully erosion in a semi-arid region of Iran: an integrated framework. Sci Total Environ 579:913–927

Razandi Y, Pourghasemi HR, Neisani NS, Rahmati O (2015) Application of analytical hierarchy process, frequency ratio, and certainty factor models for groundwater potential mapping using GIS. Earth Sci Inf 8(4):867–883

Rudin C (2019) Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 1 (5):206–215

Saha S (2017) Groundwater potential mapping using analytical hierarchical process: a study on Md. Bazar Block of Birbhum District, West Bengal. Spat Inf Res 25(4):615–626

Saha S, Paul GC, Pradhan B, Abdul Maulud KN, Alamri AM (2021) Integrating multilayer perceptron neural nets with hybrid ensemble classifiers for deforestation probability assessment in Eastern India. Geomat Nat Hazards Risk 12(1):29–62

Sajedi-Hosseini F, Malekian A, Choubin B, Rahmati O, Cipullo S, Coulon F, Pradhan B (2018) A novel machine learning-based approach for the risk assessment of nitrate groundwater contamination. Science Total Environ 644:954–962

Shekhar S, Pandey AC (2014) 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(4):402–421

Suhag R (2016) Overview of ground water in India. PRS Legislative Research ("PRS") standing committee report on Water Resources examined 10

Tahmassebipoor N, Rahmati O, Noormohamadi F, Lee S (2016) Spatial analysis of groundwater potential using weights-of-evidence and evidential belief function models and remote sensing. Arab J Geosci 9(1):79

Talukdar S, Pal S (2020) Wetland habitat vulnerability of lower Punarbhaba river basin of the uplifted Barind region of Indo-Bangladesh. Geocarto Int 35(8):857–886

Talukdar S, Ghose B, Salam R, Mahato S, Pham QB, Linh NTT, Costache R, Avand M (2020) Flood susceptibility modeling in Teesta River basin, Bangladesh using novel ensembles of bagging algorithms. Stoch Env Res Risk Assess 34(12):2277–2300

Thapa R, Gupta S, Guin S, Kaur H (2018) Sensitivity analysis and mapping the potential groundwater vulnerability zones in Birbhum district, India: a comparative approach between vulnerability models. Water Sci 32(1):44–66

Tien Bui D, Bui QT, Ngayen QP, Pradhan B, Nanpak H, Trinh PT (2017) A hybrid artificial intelligence approach using GIS-based neural-fuzzy inference system and particle swarm optimization for forest fire susceptibility modeling at a tropical area. Agric for Meteorol 233:32–44

Ting KM, Witten IH (1997) Stacking bagged and dagged models. Working paper 97/09, University of Waikato, Department of Computer Science, Hamilton, New Zealand

Weiss A (2001) Topographic position and landforms analysis. Poster Presentation, ESRI User Conference, San Diego, CA

World Bank (2010) Deep wells and prudence: towards pragmatic action for addressing groundwater overexploitation in India. 51676, Washington, D.C. http://documents.worldbank.org/curated/en/272661468267911138/Deep-wells-and-prudence-towards-pragmaticaction-for-addressing-groundwater-overexploitation-in-India

Wu Y, Ke Y, Chen Z, Liang S, Zhao H, Hong H (2020) Application of alternating decision tree with AdaBoost and bagging ensembles for landslide susceptibility mapping. CATENA 187:104396

Yariyan P, Janizadeh S, Van Phong T, Nguyen HD, Costache R, Van Le H, Pham BT, Pradhan B, Tiefenbacher JP (2020) Improvement of best first decision trees using bagging and dagging ensembles for flood probability mapping. Water Resour Manag 34(9):3037–3053