



Groundwater potential mapping using multi-criteria decision, bivariate statistic and machine learning algorithms: evidence from Chota Nagpur Plateau, India

Md Hasanuzzaman¹ · Mehedi Hasan Mandal² · Md Hasnine³ · Pravat Kumar Shit¹

Received: 8 September 2021 / Accepted: 3 February 2022 / Published online: 9 March 2022
© The Author(s) 2022

Abstract

Increased consumption of water resource due to rapid growth of population has certainly reduced the groundwater storage beneath the earth which leads certain challenges to human being in recent time. For optimal management of this vital resource, exploration of groundwater potential zone (GWPZ) has become essential. We have applied Analytical Hierarchy Process (AHP), Frequency Ratio (FR) and two machine learning techniques specifically Random Forest (RF) and Naïve Bayes (NB) here to delineate GWPZ in Gandheswari River Basin in Chota Nagpur Plateau, India. To achieve the goal of the study, twelve factors that determine occurrence of groundwater have been selected for inter-thematic correlations and overlaid with location of wells. These factors include elevation, drainage density, slope, lithology, geomorphology, topographical wetness index (TWI), distance from the river, rainfall, lineament density, Normalized Difference Vegetation Index (NDVI), soil, and Land use and Land cover (LULC). A total 170 points including 85 in well site and 85 in non-well site have been selected randomly and allocated into two parts: training and testing at the share of 70:30. The implemented methods have significantly provided five GWPZs specifically Very Good (VG), Good (G), Moderate (M), Poor (P) and Very Poor (VP) with high and acceptable accuracy. The study also finds that geomorphology, slope, rainfall and elevation have greater importance in shaping GWPZs than LULC, NDVI, etc. Model performance has been tested with receiver operator characteristics (ROC), Accuracy (ACC), Kappa Coefficient, MAE, RMSE, etc., methods. Area under curve (AUC) in ROC curve has revealed that accuracy level of AHP, FR, RF and NB is 78.8%, 81%, 85.3% and 85.5, respectively. The machine learning techniques coupled with AHP and FR unveil effective delineation of groundwater potential area in said river basin which by genetically offers low primary porosity due to lithological constrains. Therefore, the study can be helpful in watershed management and identifying appropriate location wells in future.

Keywords Analytical Hierarchy Process · Frequency Ratio · Machine learning · Random Forest · Naïve Bayes · ROC curve

✉ Pravat Kumar Shit
pravatgeo2007@gmail.com

Md Hasanuzzaman
hasan20geo@gmail.com

Mehedi Hasan Mandal
hasanmhm86@gmail.com

Md Hasnine
hasnine.geo@gmail.com

¹ PG Department of Geography, Raja N. L. Khan Women's College (Autonomous), Gope Palace, Midnapore, West Bengal 721102, India

² Department of Geography, Krishnagar Government College, Krishnagar, West Bengal, India

³ Department of Geography, Aliah University, Kolkata, India

Introduction

Groundwater is among the most indispensable resources of the earth that takes place below the surface of the earth (Naghibi et al. 2015) on which near about 2.5 billion human beings depend on these fresh water resources in daily basis (Alcaide and Santos 2019). Groundwater varies spatially in both quality and quantity; however, it is very important for socio-economic development because groundwater meets certain demands of mankind, namely water for drinking, for irrigation, for forestry, for industrial purpose and to support livestock (Naghibi et al. 2016). Utilization of groundwater is hygienic and more reliable than surface water because groundwater is less exposed to environmental degradation (Kim et al. 2019; Lee et al. 2020). In most part of the globe, uncontrolled use of groundwater has

depleted this resource. Since the last few decades, the availability of freshwater resource has become challenging issue because of its high demand for domestic, agricultural, industrial purposes (Chakraborty et al. 2021; Shit et al. 2019, Chen et al. 2019), insufficient rainfall, surface water scarcity and population growth (Panahi et al. 2020) which can lead the shortage of groundwater globally by 2025 (Nguyen et al. 2020). Being world's leading groundwater consumer, the consumption rate of India has been stated 230 cubic km per year (Fienen and Arshad 2016). Thus, mapping the GWPZ has become an essential and central part in the management system of watershed (Verma et al. 2018; Bhunia et al. 2018; Kulkarni et al. 2018).

Groundwater mapping has been carried out with direct filed surveys in recent past in expensive and time-consuming manner (Prasad et al. 2020). But now the integration of remote sensing and GIS is capable of accumulating, maneuvering and demonstrating various forms of data which result into the construction of thematic maps (Band et al. 2020; Rukhsana 2020; Karimi-Rizvandi et al. 2021). Besides, this platform is time as well as cost-effective and also applicable in large area (Prasad et al. 2020). The occurrence of groundwater varies over place to place in accordance with hydrology, climate, topography, geology, ecology, soil, slope, etc., of the region (Karimi-Rizvandi et al. 2021). Therefore, such factors are used in GIS to prepare the GWPZs.

Review of the literature suggests that researchers across the globe have used various methods to delineate GWPZs. Among them Analytical Hierarchy Process (Maity and Mandal 2019), Logistic regression (Park et al. 2017), Frequency Ratio (Ozdemir 2011), Weights of evidence (Madani and Niyazi, 2015) are very commonly used for this purpose. Besides, various techniques under machine learning are now broadly accepted in order to delimit GWPZs. These include Random Forest (Naghbi et al. 2016), SVM (Support vector machine) (Lee et al. 2018), BRT (boosted regression trees) (Naghbi and Pourghasemi 2015), linear discriminant analysis (Naghbi et al. 2017), Naïve Bayes (Miraki et al. 2018), classification and regression tree (Naghbi et al. 2016) and artificial neural network (Lee et al. 2018). Despite being used in different parts of the planet all these techniques have some drawbacks. Identification of groundwater potential zones based on one single method is now not justifying the study.

AHP reduces the mathematical complexity in decision making (Abhijit 2020), thereby widely used. Frequency Ratio has been also successfully used with very high and precise accuracy by Ozdemir 2011. Moreover, hypothesis or postulation is not obligatory in the allocation of revealing factors in RF model and enables mixed use of categorical data and numeric data (Aertsen et al. 2010). Even NB model is very simple and does not necessitate for estimation of parameter (Wu et al. 2008). Both RF (Naghbi et al. 2016) and NB (Miraki et al. 2018) models have been successfully

implemented by several researchers across the globe with high accuracy. Among the machine learning model, Random Forest (RF) and Naïve Bayes (NB) are the most acceptable and high accuracy models depicted in previous studies' results (Naghbi et al., 2017; Pham et al., 2021; Miraki et al., 2018). It helps the model selection for GWPZs. Therefore, present study tries to map the probable groundwater sites by using with AHP, Frequency Ratio (FR), Random Forest (RF) and Naïve Bayes (NB) in Gandheswari River Basin of Bankura District, West Bengal. Gandheswari Watershed is composed with hard crystalline rock mainly granite gneiss which is not preamable; therefore, occurrence groundwater is not widely spread over the region. Thus, the main objective of the current work is to compare among multi-criteria decision approach, bivariate statistic method and machine learning algorithms for the delineation of groundwater potential zone (GWPZ) of the study area.

Description of study area

Gandheswari Watershed has been selected to delineate the GWPZs. Gandheswari River is the 32-km-long tributary of Dwarakeshwar River and flows through the four CD Blocks of Bankura district of West Bengal after originating from Santuri CD Block of Purulia district of West Bengal. The study area extends between 86° 53' 20.526" E and 87° 08' 20.681" E longitudes and 23° 13' 43.376" N and 23° 31' 15.417" N latitudes. The watershed occupies nearly 394.96 km² (Fig. 1). This watershed is mainly situated in the peripheral region of Chota Nagpur Plateau; thereby, the studied region consists with undulating plane (below 120 m), an eroding plateau (120–220 m) and the Susunia Hill Zone (220–437 m) (Sinha 2016). Thick layer of 'mottled clay' is very abundant in Gandheswari basin, and most part of the study area consist granitic gneissic of Pre-Cambrian which results into moderate to low storage of groundwater (Ghosh et al. 2020).

Material and methods

Data from different sources are used for spatial modeling and GWPZ analysis (Table 1). After converted the data into spatial database in accordance with our requirements AHP, FR, RF and NB methods have been applied to conduct the study. Figure 2 represents overall framework of the study.

Preparation of inventory map

Researchers, across the globe, have prepared inventory dataset for groundwater mapping by using location of springs, wells and quant. However, present study selects 85 well points and 85 non-well points (where occurrence

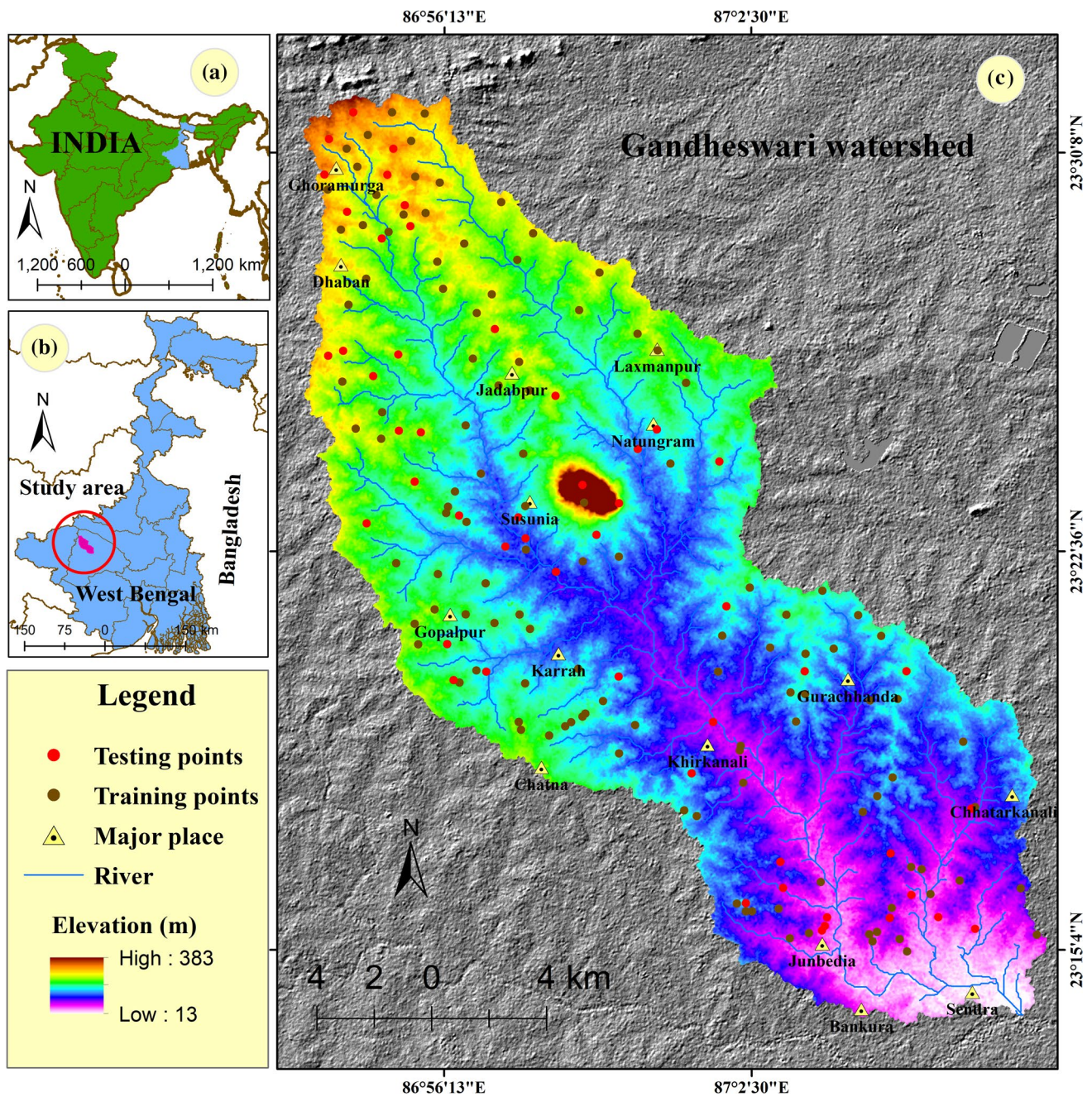


Fig. 1 Location of the study area: a India, b West Bengal, and c Gandheswari Watershed

of groundwater is minimum) to construct the inventory map. SOI toposheets (73I/15, 73 M/3 and 73 M/4) and Central Ground Water Board (CGWB) data have been used here. Of the 170 sites, 70% (119) have been randomly used for modeling and 30% (51) have been randomly used for validation purpose.

Factors affecting groundwater potential zone

Selection of effective parameters of GWPZ is crucial task for researchers (Naghibi et al. 2016). Literature review (Table 2) has helped to identify twelve such parameters. The thematic maps (Fig. 3a–i) based on the selected parameters have been

Table 1 Data sources and type of data required in the research

Factor	Resolution	Data out type	Data source
SRTM DEM (elevation)	30 m	Raster	https://earthexplorer.usgs.gov
Slope	30 m	Raster	Extracted from DEM
Drainage density (DD)	30 m	Raster	Extracted from DEM
Topographic weightiness index (TWI)	30 m	Raster	Extracted from DEM
Distance from the river (DFR)	30 m	Raster	Extracted from DEM
Lineament	30 m	Raster	Extracted from DEM
NDVI	30 m	Raster	https://earthexplorer.usgs.gov (Landsat 8)
Soil	30 m	Raster	Soil map of the Soil Survey and Land Use Planning (NBSS&LUP)
Rainfall	885.37×885.37	Raster	WorldClim website
Lithology	30 m	Raster	Geological survey of India, (R.F.1:250,000)
Geomorphology	30 m	Raster	Geological survey of India, (R.F.1:250,000)
Land Use and Land Cover (LULC)	30 m	Raster	Extracted from satellite image (Landsat 8)
Groundwater Point	Randomly	Vector	The Survey of India (Toposheet 73I/15, 73 M/3 and 73 M/4 of 1:50,000), Central Ground Water Board (CGWB)

prepared by using ArcGIS software. Details of these factors are as follows:

Elevation has tremendous impact on groundwater potential mapping (Naghibi et al. 2016) as it is contrariwise related to the reserve of groundwater (Karimi-Rizvandi et al. 2021). Figure 3a reveals that elevation of Gandheswari Watershed varies from 13 to 383 m.

Slope is another important factor that controls rate of infiltration and run-off in any part of the globe. Higher slope adversely effects on groundwater storage; thereby, groundwater potential zones are generally associated with lower slope region (Maskooni et al. 2020). Highest slope in the study area is recorded as 43.24 degree, and lowest is recorded as zero degree (Fig. 3b). Drainage density is directly related to run-off and inversely related to groundwater storage (Magesh et al. 2012). In this area, drainage density extends from 0 to 0.75 km² (Fig. 3c). TWI value ranges from 26.21 to 2.66 in the study area (Fig. 3d). TWI uncovers the saturated portions in said watershed. The index indicates effects of topography on accumulation of water in a region (Biswas et al. 2020), and hence, steep slope and higher elevation have greater run-off and thus reduce the capacity of water accumulation; on contrary, low-lying area has greater potential of topographical wetness or accumulation of water in the study area. The formula, given by Moore et al. (1991), is used to compute the TWI in present research. Distance from the river can be a vital controlling factor of groundwater storage. In this specific research, the distance from river ranges from 0 to 1511.67 m (Fig. 3e). Lineament density is among the most influential variables as it is positively related to groundwater storage. Lineaments act as the place of secondary porosity (Ghosh et al. 2020) and thereby

very important in this study because most parts of the Gandheswari River Basin are composed with granite gneiss whose primary porosity is assumed to be low. The lineament density varies 0–0.59 km² (Fig. 3f) in the studied watershed. Rainfall acts as natural sources of groundwater which helps the amount of infiltration (Karimi-Rizvandi et al. 2021). The mean annual rainfall in mentioned area fluctuates between 97.5 and 114.83 cm (Fig. 3i). Nature of soil also determines the storage of groundwater because soil properties determine the permeability of the region (Karimi-Rizvandi et al. 2021). Figure 3h unveils that current study area consists of four types of soil group, namely coarse loamy, clayey loamy, fine loamy and fine silt. Among those groups, coarse loamy soil can recharge the groundwater more efficiently than the others. Storage of groundwater is also shaped by geomorphology of any region (Biswas et al. 2020). Figure 3k uncovers five distinct features explicitly residual hill, pediment, pediplain, valley fill and water bodies. These features may be advantageous (valley fill, pediplain) for groundwater storage and residual hill and pediment may retard groundwater storage. Water bodies in the selected region act as the direct source of GWPZs. Lithological configuration of the studied watershed can be considered as primary controlling factor that determines the permeability and porosity of the region. Figure 3j reveals that most part of the watershed is composed with granite gneiss. This lithological constrain reduces the primary infiltration here, and thereby, groundwater storage is heavily depended on either secondary infiltration (through the cracks and joints) or the area having recent deposits (Ghosh et al. 2020). NDVI also significantly affects the groundwater storage capacity. Higher value of NDVI suggests thick coverage of vegetation coverage,

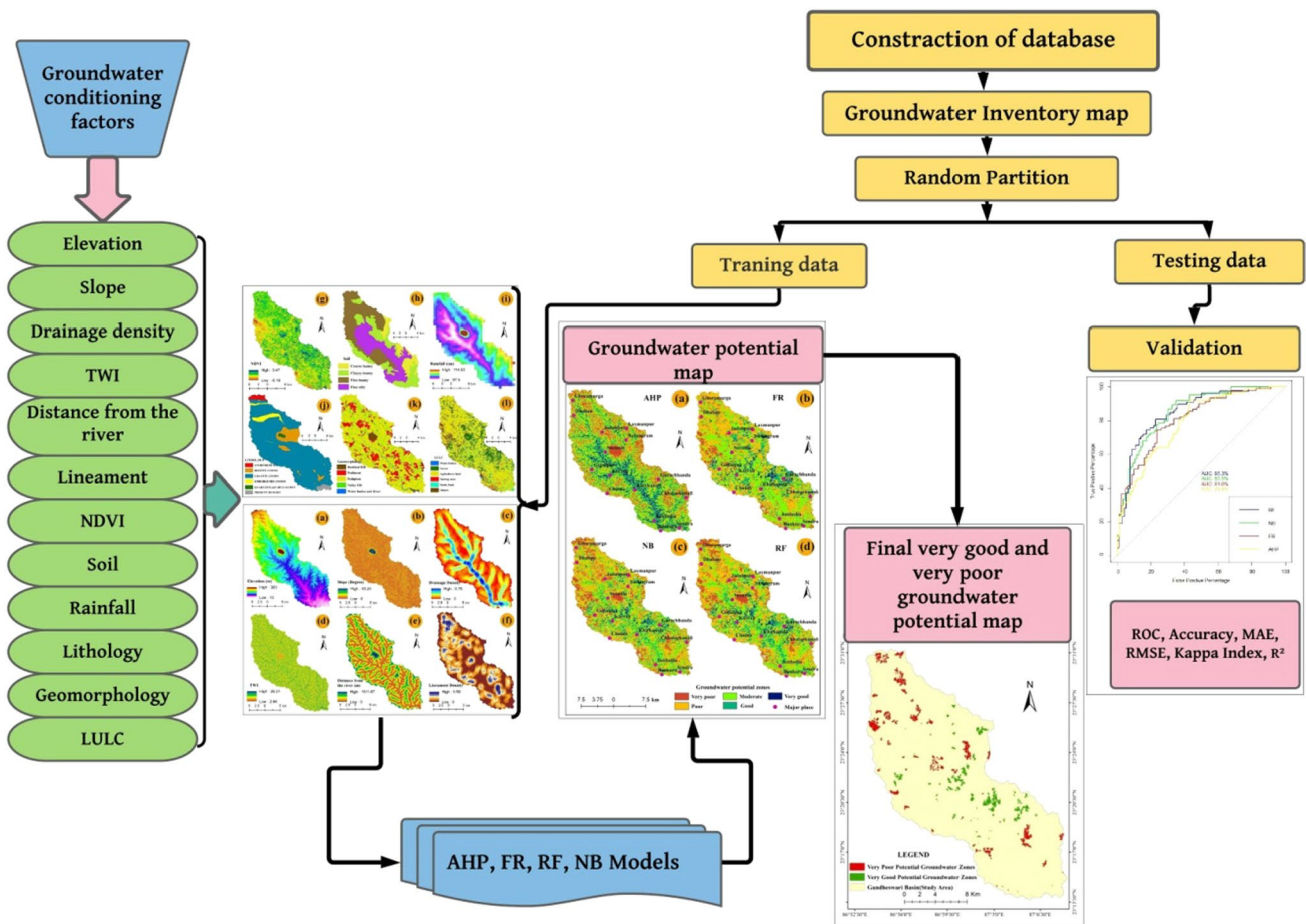


Fig. 2 Methodological design of the study: starting from criteria selection to model validation

and vegetation reduces run-off and helps in recharging the groundwater. In our area of interest, the NDVI value ranges from 0.47 to -0.19 (Fig. 3g). LULC of any region controls the groundwater movements. Evapotranspiration, surface runoff and groundwater recharge are largely controlled by LULC (Karimi-Rizvandi et al. 2021). Our study (Fig. 3i) has divided entire basin into six prominent LULC classes, namely water bodies, forests, agricultural lands, built-up area, sandy lands and other lands.

Accuracy assessment of groundwater-influencing factors

The important part of the research work is selection of the groundwater-influencing factors. The current work has used two methods for the selection of factors that influences groundwater storage. Firstly, variance inflation factors (VIF) (Dormann et al. 2013) method uncovers the multicollinearity among the selected parameters. In the current research, multicollinearity validates the possibility of association among the twelve parameters. Multicollinearity between parameters

specifies that variables which are linked can be estimated by other factors. Therefore, the multicollinearity affected variable is needed to be removed from the model. The VIF values of > 10 and < 0.1 denote such problems (Khosravi et al. 2019).

Secondly, Information Gain Ratio (IGR) method unveils the relative importance of every influencing parameter (Chen et al. 2017). The Average Merit is computed through this method which quantifies the pattern of influence. Greater Average Merit signifies greater effect on the groundwater availability and vice versa.

Methods for GWPZ

AHP method

Analytical Hierarchical Process (AHP), invented by Saaty (1971), is the hierarchical additive weighting approaches for multi-criteria decision problems, and it is broadly used by researchers across the globe. This method analyzes parameters based on their relative relevance when compared to

Table 2 Literature review of factors used to delineate groundwater potential zones (GWPZ)

Literature review	Elv	Slope	DD	TWI	DFR	RF	GM	Geology	LD	Soil	NDVI	LULC
Abijith et al (2020)		✓	✓			✓	✓	✓	✓	✓		✓
Allafta et al. (2020)	✓	✓	✓		✓	✓	✓		✓	✓		✓
Arabameri et al. (2019)	✓	✓		✓	✓	✓		✓		✓	✓	✓
Arulbalaj et al. (2019)		✓	✓	✓		✓	✓	✓	✓	✓		✓
Chakraborty et al. (2018)		✓	✓			✓	✓	✓	✓	✓		✓
Das et al. (2018)		✓				✓		✓	✓	✓		✓
Haghizadeh et al. (2017)	✓	✓	✓	✓	✓			✓	✓			✓
Hazra, Mondal, Sanjib (2018)	✓	✓		✓		✓		✓		✓	✓	✓
(Karimi-Rizvandi, et al. 2021)		✓	✓	✓	✓	✓		✓		✓		✓
Kolli et al. (2020)		✓	✓			✓	✓		✓	✓		✓
Lee, Hyun, Lee, Lee (2020)	✓	✓						✓		✓		✓
Mir, Bhat, Rather, and Mattoo (2021)		✓	✓	✓		✓		✓	✓			✓
Naghibi, et al. (2016)	✓	✓	✓		✓			✓				✓
Owolabi et al. (2020)			✓	✓		✓		✓	✓			✓
Pal, Ghosh, and Chowdhuri (2020)					✓	✓				✓		✓
Pothiraj and Rajagopalan (2013)		✓	✓				✓	✓	✓	✓		✓
Pourghasemi et al. (2020)	✓	✓	✓	✓	✓	✓		✓				✓
Prasad et al. (2020)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Rao et al. (2021)			✓			✓	✓	✓	✓	✓		✓
Sinha et al. (2018)	✓	✓					✓	✓	✓	✓		✓
Thapa et al. (2017)	✓	✓	✓			✓	✓	✓	✓	✓		✓
Tolche (2021)		✓	✓			✓		✓	✓	✓		✓

1. Elv (elevation) 2. Slope 3. Drainage density (DD) 4. Topological Wetness Index (TWI) 5. Distance from river (DFR) 6. Rainfall 7. Geomorphology (GM) 8. Geology 9. Lineament density (LD) 10. Soil 11. NDVI 12. Land use and Land cover (LULC)

one another. Moreover, it is able to determine the subject, along with their rank and precedence, which is computed by pairwise comparison matrix to arrange the criteria in hierarchical order. Each parameter is given a set of weights (Table 3). Next step is to normalize the data. The consistency index (CI) coupled with consistency ratio (CR) is then computed to test the constancy of these weights. This AHP method has been gone through several steps. First of all, formation of a hierarchy is necessary from the problems. AHP begins with identifying the criteria to be used in evaluating several options, which are arranged in a treelike hierarchy. After that, data have been collected by comparing criteria at each level of the hierarchy and alternatives in pairs. Then estimation of the relative importance of selected criteria and alternatives is taken places, which is followed by validating the constancy in the pairwise comparisons (Table 4). The weights of each criterion were then normalized, and their average weights were determined (Table 5). The consistency vector has been calculated by multiplying the average weight of each criterion. The following equations have widely been used to check the CI and CR from the pairwise comparison matrix of all the parameters.

$$CI = \frac{(\lambda_{max} - n)}{(n - 1)} \tag{1}$$

Here, *n* is the total number of criteria and λ_{max} (**lambda**) is simply the average value of consistency vector.

$$CR = \frac{CI}{RI} \tag{2}$$

Here, *RI* is the random index from Table 3

The present research finds the followings: maximum eigen value (λ_{max}) = 13.673, consistency index ($CI = (\lambda_{max} - n) / (n - 1) = 0.15209$, random index ($RI = 1.54$ (for $n = 12$), consistency ratio ($CR = (CI/RI) = 0.0987$ or 9.9 (acceptable).

The weighted overlay analysis is very much useful tools for any suitable area analysis. This method has the ability to assigning and combining the multilayers to create an integrated analysis. The weighted values calculated by AHP method are used in weighted overlay tools to identify prominent factor through this process (Parimala and Lopez 2012).

$$S = \sum_{i=1}^n W_i * X_i \tag{3}$$

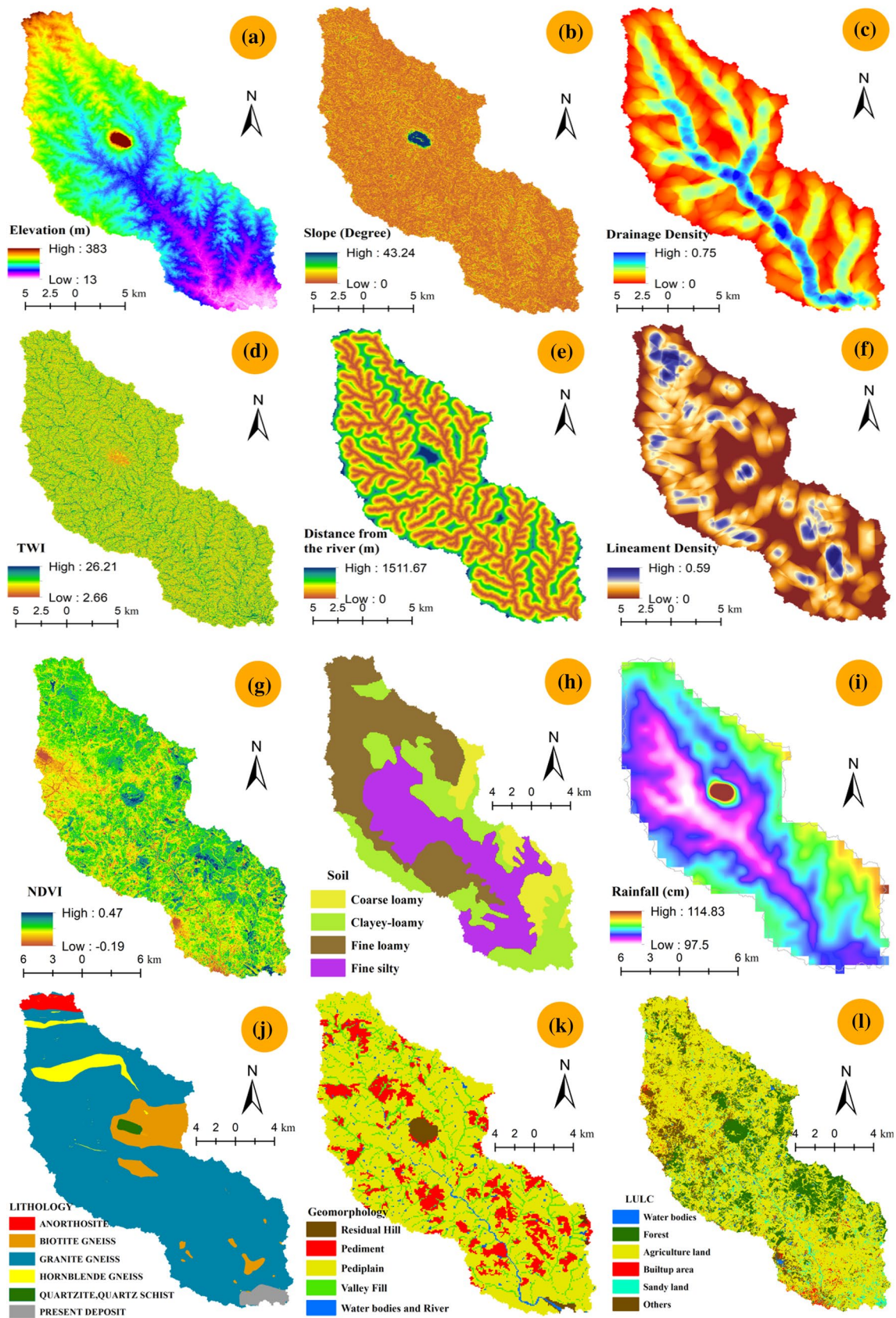


Fig. 3 Distribution of six causative factors used in this study: **a** elevation, **b** slope, **c** drainage density, **d** TWI, **e** distance from the river, **f** lineament, **g** NDVI, **h** soil, **i** rainfall, **j** lithology, **k** geomorphology and **l** LULC

where S is the suitability index for each pixel map. W_i is the weight of the i th layer and X_i score of the i th criteria layer. n is the number of suitability layer.

Frequency Ratio (FR)

The Frequency Ratio (FR) is a statistic-based bivariate approach and has been developed to discover the groundwater potential area by evaluating the relationships among the controlling factors (Oh et al. 2011; Naghibi et al. 2016). The model has been applied here to uncover the quantitative link between distribution of well occurrence and predictor factors. Frequency Ratio has been calculated based on the following equation:

$$FR = \frac{W/TW}{CP/TP} \tag{4}$$

where W represents the number of pixels having linked with well from each thematic map, whereas TW represents the total number of pixels across the area under concern. CP and TP represent number of pixels in each thematic map and in area under concern, respectively.

Random Forest (RF)

Random Forest (RF) is a very popular and accurate machine learning algorithm (Wang et al. 2021). RF is basically a tree-based method, which has an authentic and great expectation execution by joining an enormous number of decision trees to determine the relationship between the factors affecting groundwater and dug well occurrence (Kim et al. 2018). Random forest creates many trees for making a ‘forest,’ where trees are created by bootstrapped data (Rahmati et al. 2017). The data are produced by the aid of classification and regression tree methods followed by Rahmati et al. (2017). RF method is further carried out by following the works of Naghibi et al. (2016), Lee et al. (2017) and Wang et al. (2021). The advantage of this method in comparison with other methods is as follows: (i) the overfitting problems of the datasets, (ii) manage big datasets with various dimensionality in nature, (iii) it does not need any hypotheses within the response variable and explanatory variables, (iv) it does not require any previous data to rescale and transform the datasets (Arabameri et al. 2019). The RF classification adopted resampling methods by randomly transferring the predictive factors to enhance the diversity in every tree (Naghibi et al. 2017). The notation of the predictive variable

Table 3 Random inconsistency indices for $n = 15$

Order	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
R.I	0	0	0.52	0.89	1.11	1.25	1.35	1.4	1.45	1.49	1.52	1.54	1.56	1.58	1.59
First order difference		0	0.52	0.37	0.22	0.14	0.1	0.05	0.05	0.04	0.03	0.02	0.02	0.02	0.01

Saaty (1980)

Table 4 Pairwise comparison matrix

	Rainfall	GM	Elevation	DD	Soil	Lithology	LD	TWI	NDVI	DFR	LULC	Slope
Rainfall	1	3	7	1	5	4	5	4	5	7	6	5
GM	0.33	1	5	1	1	2	3	4	3	4	4	6
Elevation	0.14	0.2	1	3	0.5	1	3	5	5	3	3	7
DD	1	1	0.33	1	2	3	3	5	4	6	5	3
Soil	0.2	1	2	0.5	1	2	2	2	3	3	3	5
Lithology	0.25	0.5	1	0.33	0.5	1	2	2	1	3	4	3
LD	0.2	0.33	0.33	0.33	0.5	0.5	1	5	2	2	3	3
TWI	0.25	0.25	0.2	0.2	0.5	0.5	0.2	1	0.5	2	1	1
NDVI	0.2	0.33	0.2	0.25	0.33	1	0.5	2	1	2	1	1
DFR	0.14	0.25	0.33	0.17	0.33	0.33	0.5	0.5	0.5	1	1	3
LULC	0.17	0.25	0.33	0.2	0.33	0.25	0.33	1	1	1	1	1
Slope	0.2	0.17	0.14	0.33	0.2	0.33	0.33	1	1	0.33	1	1
Sum	4.08	8.28	17.86	8.31	12.19	15.91	20.86	32.5	27	34.33	33	39

1. Elevation 2. Slope 3. Drainage density (DD) 4. Topological Wetness Index (TWI) 5.Distance from river (DFR) 6. Rainfall 7. Geomorphology (GM) 8. Geology 9. Lineament density (LD) 10. Soil 11. NDVI 12. Land use and Land cover (LULC)

Table 5 Normalized pairwise comparison matrix

	Rainfall	GM	Elevation	DD	Soil	Lithology	LD	TWI	NDVI	DFR	LULC	Slope	Sum	Weight
Rainfall	0.245	0.362	0.392	0.120	0.410	0.251	0.240	0.123	0.185	0.204	0.182	0.128	2.843	0.237
GM	0.081	0.121	0.280	0.120	0.082	0.126	0.144	0.123	0.111	0.117	0.121	0.154	1.579	0.132
Elevation	0.034	0.024	0.056	0.361	0.041	0.063	0.144	0.154	0.185	0.087	0.091	0.179	1.420	0.118
DD	0.245	0.121	0.018	0.120	0.164	0.189	0.144	0.154	0.148	0.175	0.152	0.077	1.706	0.142
Soil	0.049	0.121	0.112	0.060	0.082	0.126	0.096	0.062	0.111	0.087	0.091	0.128	1.125	0.094
Lithology	0.061	0.060	0.056	0.040	0.041	0.063	0.096	0.062	0.037	0.087	0.121	0.077	0.801	0.067
LD	0.049	0.040	0.018	0.040	0.041	0.031	0.048	0.154	0.074	0.058	0.091	0.077	0.721	0.060
TWI	0.061	0.030	0.011	0.024	0.041	0.031	0.010	0.031	0.019	0.058	0.030	0.026	0.372	0.031
NDVI	0.049	0.040	0.011	0.030	0.027	0.063	0.024	0.062	0.037	0.058	0.030	0.026	0.457	0.038
DFR	0.034	0.030	0.018	0.020	0.027	0.021	0.024	0.015	0.019	0.029	0.030	0.077	0.345	0.029
LULC	0.042	0.030	0.018	0.024	0.027	0.016	0.016	0.031	0.037	0.029	0.030	0.026	0.326	0.027
Slope	0.049	0.021	0.008	0.040	0.016	0.021	0.016	0.031	0.037	0.010	0.030	0.026	0.303	0.025
Σ														1

is defined as $\log_2(M+1)$, where M is the total input number within the algorithm. The RF model determines the split at each node with the help of predictive variables and the number of trees (Kim et al. 2018). The average prediction of the tree is computed as:

$$Gp = \frac{1}{k} \sum k^{th} v^{response} \tag{5}$$

where Gp is any groundwater prediction and k represents the separate trees in the method.

Naïve Bayes (NB)

Naïve Bayes (NB) model is based on postulation that there are no dependent attributes to capitalize on the subsequent possibility in determination of the class for categorization (Soni et al. 2011). NB classification scheme is a term in Bayesian statistics which supervises an easy probabilistic classifier determined by Bayes' hypothesis (Bhargavi and Jyothi, 2009). The major benefit of the NB classifier is that it is simple to build and iterative parameter estimation schemes are not needed in it (Wu et al. 2008).

x_1 is the vector of the 12 controlling factors of groundwater potential zone, and y_i is the vector of classifier variable (potential zone or non-potential zone). The NB is based on following equations.

$$Y_{NB} = \prod_{[y_i=potential_zone_or_non_potential_zone]}^{arg \max p(y_i)} \prod_{i=1}^{12} P\left[\frac{x_i}{y_i}\right] \tag{6}$$

where $P(y_i)$ is the prior probability of y_i that can be estimated based on the proportion of the observed cases with

output class y_i in the training dataset. $P(x_i/y_i)$ is the conditional probability that can be calculated by the following equation:

$$p\left(\frac{x_i}{y_i}\right) = \frac{1}{\sqrt{2\pi}\alpha} e^{-\frac{(x_i-\eta)^2}{2\alpha^2}} \tag{7}$$

where η is the mean and α is the standard deviation of x_i .

Model validation

Validation of any model is fundamental steps for scientific research (Naghbi et al 2016). The performance of GWPM by four methods has been evaluated by ROC curve and the statistical measures of accuracy (ACC), mean absolute error (MAE), root-mean-square error (RMSE), Kappa index (K) and coefficient of determination (R^2). The formulas that are used here are as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{i=n} (X_{ei} - X_{oi})^2} \tag{9}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{i=n} |X_{ei} - X_{oi}| \tag{10}$$

$$Kappa(k) = \frac{P_c - P_{cyp}}{1 - P_{cyp}} \tag{11}$$

where P_c indicates numeral of pixels to be matched accurately as well or non-well pixels; P_{cyp} denotes estimated

results. X_{oi} and X_{ei} are the i^{th} observed and model predicted values, respectively, and n is the amount of data point (Khosravi et al. 2019).

The present study also uses ROC curve to unveil overall validity of the models applied here. The ROC curve significantly predicts the occurrence or non-occurrence of wells by sensitivity on Y-axis and specificity on X-axis (Prasad et al. 2020). The region below the curve is called area under curve. AUC is very much essential for model efficiency (Karimi-Rizvandi et al. 2021). The value of AUC ranges from 0 to 1 and near to 1 represents higher accuracy of the models (Naghieb et al. 2016; Chen et al. 2018; Prasad et al. 2020).

Results

Importance of factors

IGR and VIF technique have been employed to identify the influence of selected parameters in groundwater potential map (GPM) and to unveil the multicollinearity issues in the selected parameters, respectively. The results of IGR and VIF are portrayed below (Table 6). The table discloses that VIF values of all factors are smaller than 10; therefore, no multicollinearity problem is existed among the selected parameters. Apart from VIF, IGR values also uncover the factorwise influence upon GWPZ.

Table 6 also demonstrates that for the river basin, geomorphology has the highest (0.94) importance in GWPZ, followed by slope (0.88) and rainfall (0.87). Besides, distance from the river (0.69), elevation (0.67) has moderate influence in the storage of groundwater. Moreover, LULC (0.08) has the least effect on groundwater storage and followed by soil (0.22) and topographical wetness index (0.23). So, the results unveil that all the selected factors have some impact on GWPZ; therefore, all these factors have been included in model development.

Groundwater potential zone mapping

Based on four different models GWPZ has been prepared for the Gandheswari Watershed (Fig. 5 a-d). ArcGIS has helped to classify GWPZ into five different classes such as Very good (VG), Good (G), Moderate (M), Poor (P) and Very Poor (VP). Based on expertise thoughts, pairwise comparison matrix and normalized pairwise comparison matrix are computed in Tables 4 and 5, respectively, to make decisions via AHP model. Weight overlay analysis techniques have been performed based on the result in ArcGIS, and GWPM has been created by AHP model (Fig. 4a).

Table 7 demonstrates percentagewise area of each class in each GWPM. According to the AHP model (Table 7), the percentages for the class VP, P, M, G and VG potential zones are 12.76, 27.88, 26.33, 26.81 and 6.21%, respectively. In case of the FR technique 9.66, 29.07, 28.41, 27.55 and 5.31% area falls into the class of VP, P, M, G and VG, respectively. RF model depicts (Table 7) that 12.66, 29.09, 28.87, 25.59 and 3.68 percentages area falls under the class of VP, P, M, G and VG potential categories, respectively. Finally, NB technique uncovers that percentages for the class VP, P, M, G and VG potential categories are 14.16, 29.52, 27.21, 25.98 and 3.02%, respectively.

Based on the very good potential and very poor potential zone a final overlay map has been created in ArcGIS platform to show the common area across the four model under the category of very good and very poor category. This overlay map (Fig. 5) presents the location where water can be easily accessible in near future. This map depicted that 10.41 km² areas are under the very good and 20.77 km² areas is under the very poor category of groundwater probability. This result may help in watershed management as the result provides the sites where wells are to be drilled and sites where well should not be drilled.

Model validation

The analytical performance of four GWPZ models has been measured by several measures, namely accuracy, Kappa coefficient, RMSE, MAE and R² (Table 8). The results clearly unveil that proposed machine learning-based Naïve Bayes model has the highest value of accuracy (87.36%), Kappa coefficient (0.85), coefficient of determination (0.86)

Table 6 The evolution of the influencing factors using VIF and IGR test (Average Merit)

Sl. No	Influencing factors	VIF	Average merit (AM)
1	Elevation	1.84	0.67
2	Slope	1.47	0.88
3	Drainage Density	1.95	0.46
4	TWI	1.08	0.23
5	Distance from the river	2.95	0.69
6	Lineament	1.98	0.36
7	NDVI	1.29	0.41
8	Soil	1.03	0.22
9	Rainfall	3.11	0.87
10	Lithology	1.21	0.28
11	Geomorphology	3.15	0.94
12	LULC	2.10	0.08

and lowest value of MAE and RMSE as 0.16 and 0.19, respectively, in the validation phase. This result significantly represents a very high level of satisfaction in mapping of GWPZ through this model. The performance analysis of the four models in the validation stage follows the descending order: NB > RF > FR > AHP.

The ROC curve (Fig. 4) unveils that NB model has (AUC = 85.5%) outperformed the RF (AUC = 85.3%), FR (AUC = 0.81.0%) and AHP (AUC = 78.8%) models in the validation phase (Fig. 6). The prediction percentage depicts that all the models have performed well, but machine learning-based RF and NB models show highest prediction effectiveness over statistical-based FR and MCDM-based AHP models.

Discussion

The groundwater potentiality mapping is expected to very useful for water resource management in the studied Gandheswari river basin because most parts of the basin consist of hard rock and thereby exhibit very low primary porosity. Methodological approach for the study having high accuracy is based on logical consideration among twelve commonly used groundwater contributing factors. The elevation and slope were very low in the southeastern portion of this Gandheswari watershed. Groundwater recharge is negatively related to the elevation (Pham et al. 2021). Thus, locations that are located in low-elevation

Fig. 4 Result of AHP (a), FR (b), NB (c) and RF (d) model for GWPZ

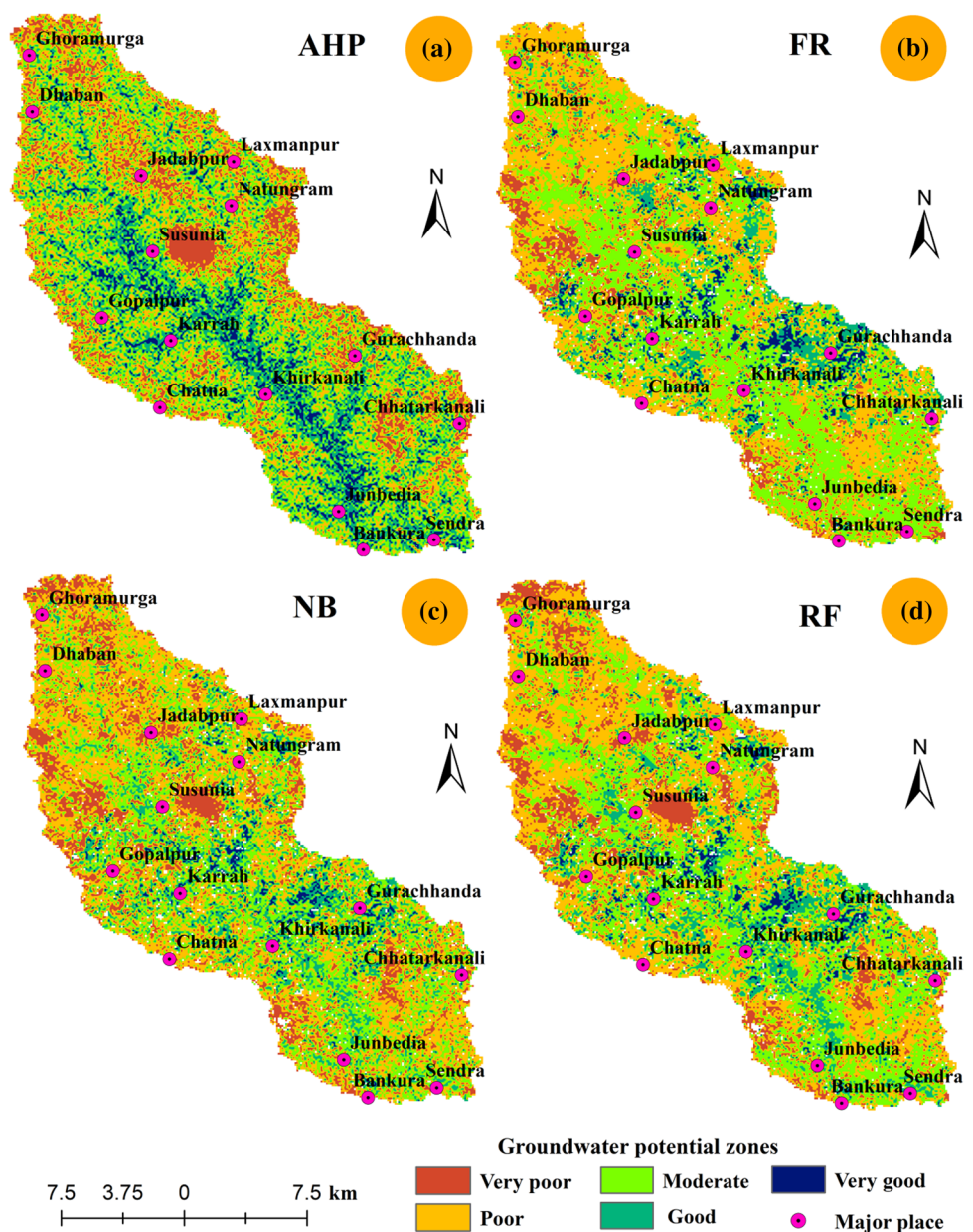


Table 7 Area under groundwater potential zones of different models

Categories	Weight base		Machine learning base	
	AHP (km ²)	FR (km ²)	RF (km ²)	NB (km ²)
Very poor	49.1	37.19	48.74	54.51
Poor	107.24	111.45	111.96	113.61
Moderate	101.26	109.35	111.09	104.71
Good	103.15	106.01	98.47	99.98
Very good	23.70	20.43	14.19	11.64
Total	384.45	384.45	384.45	384.45

areas represent high groundwater potential in particular regions of the study area rather than the overall study area. Since the Gandheswari watershed is situated on the Pre-Cambrian granitic and gneissic rocks, the movement and occurrence of groundwater are found to be moderate to low (Etikala et al. 2019). In the current study area, shallow

aquifers are of great importance as source of water (Central Ground Water Board 2017). Groundwater supports various sectors, namely agriculture, industry and many more to the human society. But recently irrational exploitation of this resource has led water shortage (Miraki et al. 2018). Reduction of surface water along with the misuse of existing groundwater has brought some key challenges to planet earth. Thus, managing the groundwater has become necessary. The current study has aimed at the exploration of GPZ in Gandheswari Watershed with the help of widely used AHP, statistical-based method FR and two machine learning algorithms, namely RF and NB. During model building for the study, the VIF has showed there is no multicollinearity problem and thus all the selected twelve parameters have been used during model building. Furthermore, InGR method has revealed that geomorphology followed by slope have the highest impact in the mapping of GWPZ.

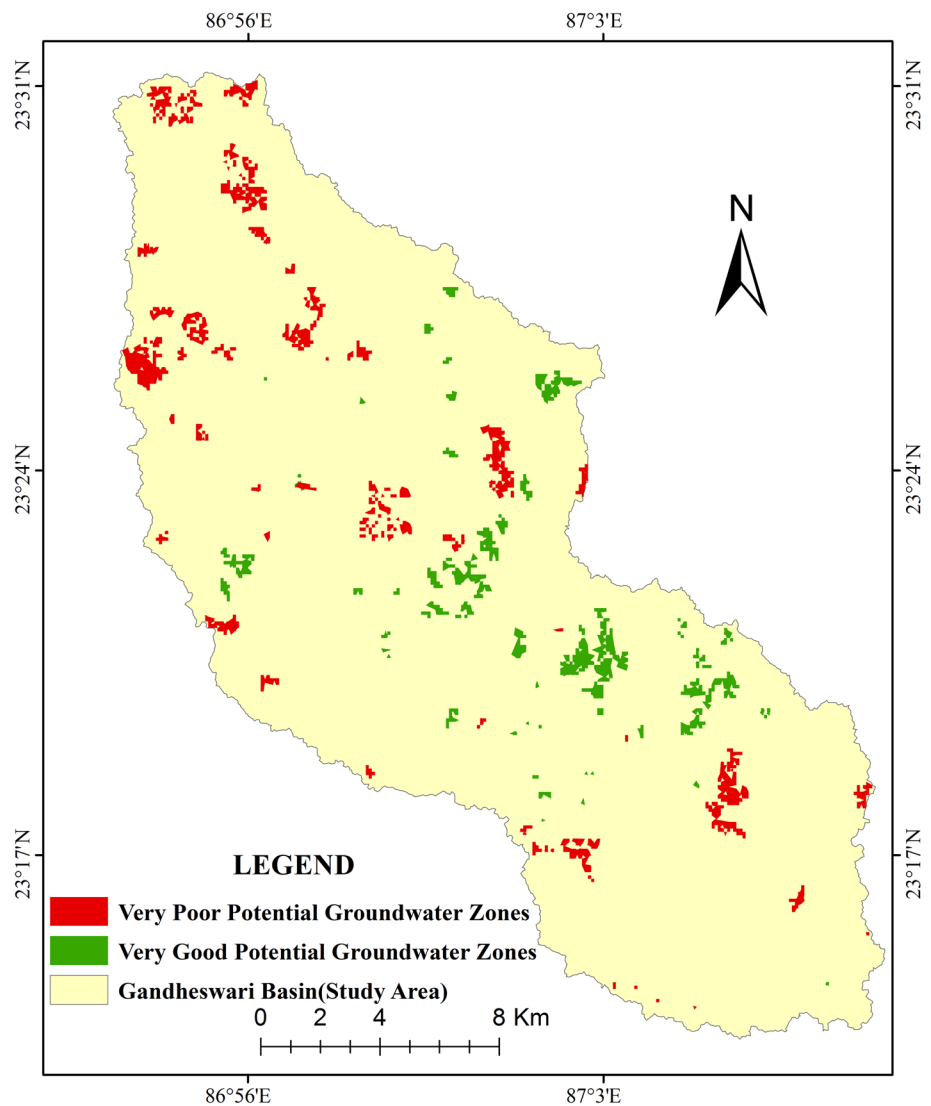
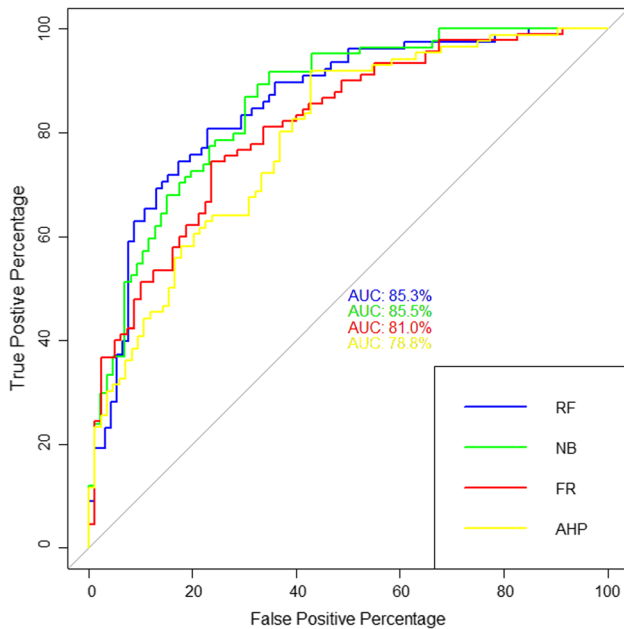
Fig. 5 Final very good and very poor groundwater potential map

Table 8 The accuracy assessment of AHP, FR, NB and RF model for training and testing data using error measures

Methods	AHP	FR	NB	RF
Accuracy (%)	76.21	81.32	87.36	86.24
Kappa index (K)	0.78	0.80	0.85	0.83
MAE	0.41	0.29	0.16	0.18
RMSE	0.36	0.27	0.19	0.21
R ²	0.79	0.81	0.86	0.84

**Fig. 6** ROC for models' validation

The study unveils that the selected techniques have made a substantial contribution to map the potential groundwater sites into following categories: VP, P, M, G and VG with high accuracy. The result reveals that less than 2.71% area of Gandheswari Watershed is very good potential zone for easy access to groundwater across all models and nearly 50 to 55% area indicate moderate to good potential zone. The watershed is mostly composed of granite gneiss of Archean era; therefore, porosity and permeability are assumed to be low, besides geomorphology of the area also suggests existence of residual hill (for example Susunia Hill) which may negatively affect groundwater storage. The ROC curve uncovers that the accuracy level for AHP, FR, RF and NB is 78.8, 81.0, 85.3 and 85.5%, respectively. That definitely depicts that NB method has more accurately identified the potential groundwater sites followed by RF method. Furthermore, the research can be used by engineers and decision-makers to the refill of world's most vital and precious resources.

Conclusions

Groundwater potential mapping using various factors is one of the significant aspects in groundwater studies. In the current research, the performance of four relatively new data mining models such as AHP, Frequency Ratio (FR), Random Forest (RF) and Naïve Bayes (NB) models has been assessed. Therefore, multi-criteria decision approach, bivariate statistic method and machine learning algorithms were employed and investigated in groundwater potential mapping. Accordingly, area under curve for prediction dataset was computed as 78.8, 81.0, 85.3 and 85.5% for AHP, FR, RF and NB models, respectively. Therefore, it can be concluded that NB had the best performance. Also, it can be suggested that data mining models performed generally well and could be considered in this field of study. This research showed that among the various approaches of the delineation of groundwater potential zone, machine learning algorithms are the most accurate and acceptable method. Moreover, it was seen that geomorphology, slope and rainfall had high importance in groundwater potential mapping, while LULC had the lowest importance. The output of the study showed that less than 2.71% area of Gandheswari Watershed is very good potential zone for easy access to groundwater across all models and nearly 50–55% area indicate moderate to good potential zone. Moreover, this work may lead appropriate selection of drilling wells and augmentation of available water resource by sustainable aquifer management. Apart from this, the present research may be further modified with the integration of some factors, i.e., the rate of abstraction of groundwater, amount of groundwater used by domestic purpose, quality of groundwater, etc., in order to find out the future potential sites for collecting water resource. Therefore, this approach can be applied in other parts of fringe area of Chota Nagpur Plateau having similar type of lithological features with or without necessary modifications.

Acknowledgements The authors show their kind acknowledgment to the Dept. of Geography and Microbiology, Raja N. L. Khan Women's College (Autonomous), and Department of Geology & Geophysics, Indian Institute of Technology (IIT), Kharagpur, West Bengal, India, for their laboratory facilities and kind encouragement.

Author contributions MH conceptualized and planned the study and reviewed and edited the manuscript. MHM conducted the survey, analyzed the data and interpreted the results. MH analyzed the data and interpreted the results. PKS supervised the study and reviewed and edited the manuscript. All authors have read and approved the final manuscript.

Funding This research was supported by the Department of Geography, Raja N. L. Khan Women's College (Autonomous), affiliated to Vidyasagar University, Midnapore, West Bengal, India. The author (P. K. Shit) grateful acknowledges West Bengal DSTBT for financial support through R&D Research Project Memo no. 104(Sanc.)/ST/P/S&T/10G-5/2018.

Data availability The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abijith D, Saravanan S, Singh L, Jennifer JJ, Saranya T, Parthasarathy K (2020) GIS-based multi-criteria analysis for identification of potential groundwater recharge zones - a case study from Ponniyaru watershed Tamil Nadu, India. *Hydro Research* 3:1–14. <https://doi.org/10.1016/j.hydres.2020.02.002>
- Allafta H, Opp C, Patra S (2020) Identification of groundwater potential zones using remote sensing and GIS techniques: a case study of the Shatt Al-Arab Basin. *Remot Sens*. <https://doi.org/10.3390/rs13010112>
- Arabameri A, Roy J, Saha S, Blaschke T, Ghorbanzadeh O, Bui DT (2019) Application of probabilistic and machine learning models for groundwater potentiality mapping in damghan sedimentary plain Iran. *Remot Sens* 11(24):3015. <https://doi.org/10.3390/rs11243015>
- Arulbalaj P, Padmalal D, Sreelash K (2019) GIS and AHP Techniques Based Delineation of Groundwater Potential Zones: a case study from Southern Western Ghats. *India Scientific Report* 9:(2082). <https://doi.org/10.1038/s41598-019-38567-x>
- Band SS, Janizadeh S, Chandra Pal S, Saha A, Chakraborty R, Mm S, Mosavi A (2020) Novel ensemble approach of deep learning neural network (dlmn) model and particle swarm optimization (ps) algorithm for prediction of gully erosion susceptibility. *Sensors* 20(19):5609. <https://doi.org/10.3390/s20195609>
- Bhargavi P, Jyothi S (2009) Applying naive bayes data mining technique for classification of agricultural land soils. *Int J Comp Sci Net Secur* 9(8):117–122
- Bhunia G, Keshavarzi A, Shit P, Omran E, Bagherzadeh A (2018) Evaluation of groundwater quality and its suitability for drinking and irrigation using GIS and geostatistics techniques in semiarid region of Neyshabur, Iran. *Appl Water Sci*. <https://doi.org/10.1007/s13201-018-0795-6>
- Biswas S, Mukhopadhyay BP, Bera A (2020) Delineating groundwater potential zones of agriculture dominated landscapes using GIS based AHP techniques: a case study from Uttar Dinajpur district West Bengal. *Environ Earth Sci*. <https://doi.org/10.1007/s12665-020-09053-9>
- Central Ground Water Board (2017) Groundwater year book—India 2016–2017. Ministry of Water Resources, River Development and Ganga Rejuvenation. Government of India, New Delhi
- Chakraborty R, Pal CS, Malik S, Das B (2018) Modeling and mapping of groundwater potentiality zones using AHP and GIS technique: a case study of Raniganj Block, Paschim Bardhaman, West Bengal. *Model Earth Sys Environ* 4:1085–1110. <https://doi.org/10.1007/s40808-018-0471-8>
- Chakraborty B, Roy S, Bera A, Adhikary P, Bera B, Sengupta D et al (2021) Groundwater vulnerability assessment using GIS-based DRASTIC model in the upper catchment of Dwarakeshwar river basin, West Bengal, India. *Environ Earth Sci*. <https://doi.org/10.1007/s12665-021-10002-3>
- Chen W et al (2017) A novel hybrid artificial intelligence approach based on the rotation forest ensemble and naïve Bayes tree classifiers for a landslide susceptibility assessment in Langao County, China. *Geomat Nat Haz Risk* 8(2):1955–1977
- Chen W, Panahi M, Khosravi K, Pourghasemi HR, Rezaie F, Parvinnezhad D (2019) “Spatial prediction of groundwater potentiality using ANFIS ensemble with teaching-learning-based and biogeography-based optimization. *J Hydrol* 572:435–448. <https://doi.org/10.1016/j.jhydrol.2019.03.013>
- Das B, Pal SC, Malik S, Chakraborty R (2018) Modeling groundwater potential zones of Puruliya district, West Bengal, India using remote sensing and GIS techniques. *Geol, Ecol, Landsc* 3(3):223–237. <https://doi.org/10.1080/24749508.2018.1555740>
- Díaz-Alcaide S, Martínez-Santos P (2019) Review: advances in groundwater potential mapping. *Hydrogeol J* 27:2307–2324. <https://doi.org/10.1007/s10040-019-02001-3>
- Dormann CF et al (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36(1):27–46
- Etikala B, Golla V, Li P, Renati S (2019) Deciphering groundwater potential zones using MIF technique and GIS: a study from Tirupati area, Chittoor District, Andhra Pradesh, India. *Hydro Res* 1:1–7. <https://doi.org/10.1016/j.hydres.2019.04.001>
- Fienen MN, Arshad M (2016) The international scale of the groundwater issue. In: Jakeman AJ, Barreteau O, Hunt RJ, Rinaudo JD, Ross A (eds) *Integrated groundwater management*. Springer, Cham
- Ghosh D, Mandal M, Karmakar M, Banerjee M, Mandal D (2020) Application of geospatial technology for delineating groundwater potential zones in the Gandheswari watershed, West Bengal. *Sustain Water Res Manage* 6:14. <https://doi.org/10.1007/s40899-020-00372-0>
- Haghizadeh A, Moghaddam DD, Pourghasemi HR (2017) GIS-based bivariate statistical techniques for groundwater potential analysis (an example of Iran). *J Earth Syst Sci*. <https://doi.org/10.1007/s12040-017-0888-x>
- Hazra N, Mondal M, Sanjib S (2018) Demarcation of groundwater potentiality zones using analytical hierarchy process (Ahp) model with RS & GIS techniques of Paschim Medinipur District in West Bengal. *India Int J Cur Adv Res* 7(4(M)):12193–12201
- Karimi-Rizvandi S, Goodarzi HV, Afkoeieh JH, Chung I-M, Kisi O, Kim S et al (2021) Groundwater-potential mapping using a self-learning bayesian network model: a comparison among metaheuristic algorithms. *Water* 13:658. <https://doi.org/10.3390/w13050658>
- Khosravi K, Shahabi H, Pham BT, Adamowski J, Shirzadi A, Pradhan B, Dou J, Ly H-B, Gróf G, Ho HL (2019) A comparative assessment of flood susceptibility modeling using Multi-Criteria Decision-Making Analysis and Machine Learning Methods. *J Hydrol* 573:311–323
- Kim JC, Lee S, Jung HS, Lee S (2018) Landslide susceptibility mapping using random forest and boosted tree models in Pyeong-Chang Korea. *Geocarto Int* 33(9):1000–1015

- Kim JC, Jung H-S, Lee S (2019) Spatial mapping of the groundwater potential of the geum river basin using ensemble models based on remote sensing images. *Remote Sens* 11(19):2285. <https://doi.org/10.3390/rs11192285>
- Kolli MK, Opp C, Groll M (2020) Mapping of potential groundwater recharge zones in the kolleru lake catchment, india, by using remote sensing and gis techniques. *Natural Res* 11:127–145. <https://doi.org/10.4236/nr.2020.113008>
- Kulkarni H, Aslekar U, Patil S (2018) Groundwater management in india: status, challenges and a framework for responses. In: Mukherjee A (ed) *Groundwater of South Asia*. Springer Hydrogeology. Springer, Singapore
- Lee S, Hyun Y, Lee S, Lee M-j (2017) Spatial prediction of flood susceptibility using random-forest and boosted-tree models in Seoul metropolitan city, Korea. *Geomat Nat Haz Risk* 8(2):1185–1203. <https://doi.org/10.1080/19475705.2017.1308971>
- Lee S, Hong S-M, Jung HS (2018) GIS-based groundwater potential mapping using artificial neural network and support vector machine models: the case of boryeong city in Korea. *Geocarto Int* 33(8):847–861. <https://doi.org/10.1080/10106049.2017.1303091>
- Lee S, Hyun Y, Lee S, Lee M-j (2020) Groundwater potential mapping using remote sensing-based and gis-based machine learning techniques. *Remote Sens*. <https://doi.org/10.3390/rs12071200>
- Madani A, Niyazi B (2015) Groundwater potential mapping using remote sensing techniques and weights of evidence GIS model: a case study from Wadi Yalamlam basin, Makkah Province Western Saudi Arabia. *Environ Earth Sci* 74:5129–5142
- Magesh NS, Chandrasekar N, Soundranayagam JP (2012) Delineation of groundwater potential zones in Theni district, Tamil Nadu, using remote sensing GIS and MIF techniques. *Geosci Front*. <https://doi.org/10.1016/j.gsf.2011.10.007>
- Maity DK, Mandal S (2019) Identification of groundwater potential zones of the Kumari river basin, India: an RS & GIS based semi quantitative approach. *Environ Dev Sustain* 21:1013–1034. <https://doi.org/10.1007/s10668-017-0072-0>
- Maskooni EK, Naghibi SA, Hashemi H, Berndtsson R (2020) Application of advanced machine learning algorithms to assess groundwater potential using remote sensing-derived data. *Remote Sens* 12:2742
- Mir SA, Bhat MS, Rather GM, Mattoo D (2021) Groundwater potential zonation using integration of remote sensing and AHP/ANP approach in north kashmir Western Himalaya India. *Remote Sens Land* 5(1):41–58
- Miraki S, Zanganeh SH, Chapi K, Singh VP, Shirzadi A, Shahabi H, Pham BT (2018) Mapping groundwater potential using a novel hybrid intelligence approach. *Water Resour Manag* 33(1):1–22. <https://doi.org/10.1007/s11269-018-2102-6>
- 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, Pourghasemi HR (2015) A comparative assessment between three machine learning models and their performance comparison by bivariate and multivariate statistical methods in groundwater potential mapping. *Water Res Manag* 29(14):5217–5236. <https://doi.org/10.1007/s11269-015-1114-8>
- Naghibi SA, Pourghasemi HR, Pourtaghi ZS et al (2015) Groundwater qanat potential mapping using frequency ratio and Shannon's entropy models in the Moghan watershed Iran. *Earth Sci Inform* 8:171–186
- 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* 188:44
- Naghibi SA, Pourghasemi HR, Abbaspour K (2017) A comparison between ten advanced and soft computing models for groundwater qanat potential assessment in Iran using R and GIS. *Theor Appl Clim* 131(3–4):967–984. <https://doi.org/10.1007/s00704-016-2022-4>
- Nguyen PT, Ha DH, Jaafari A, Nguyen HD, Van Phong T, Al-Ansari N, Prakash I, Van Le H, Pham BT (2020) Groundwater potential mapping combining artificial neural network and real adaboost ensemble technique: the daknong province case-study. *Vietnam* 17(7):2473. <https://doi.org/10.3390/ijerph17072473>
- Oh H-J, Kim Y-S, Choi J-K, Park E, Lee S (2011) GIS mapping of regional probabilistic groundwater potential in the area of Pohang City, Korea. *J Hydrol* 399(3–4):158–172
- Owolabi ST, Madi K, Kalumba AM, Orimoloye IR (2020) A groundwater potential zone mapping approach for semi-arid environments using remote sensing (RS), geographic information system (GIS), and analytical hierarchical process (AHP) techniques: a case study of Buffalo catchment, Eastern Cape South Africa. *Arab J Geosci* 13:1184
- Ozdemir A (2011) GIS-based groundwater spring potential mapping in the Sultan Mountains (Konya, Turkey) using frequency ratio, weights of evidence and logistic regression methods and their comparison. *J Hydrol* 411:290–308. <https://doi.org/10.1016/J.JHYDROL.2011.10.010>
- Pal SC, Ghosh C, Chowdhuri I (2020) Assessment of groundwater potentiality using geospatial techniques in Purba Bardhaman district West Bengal. *Appl Water Sci*. <https://doi.org/10.1007/s13201-020-01302-3>
- Panahi M, Sadhasivam N, Pourghasemi HR, Rezaie F, Lee S (2020) Spatial prediction of groundwater potential mapping based on convolutional neural network (CNN) and support vector regression (SVR). *J Hydrol* 588:125033. <https://doi.org/10.1016/j.jhydrol.2020.125033>
- Parimala M, Lopez D (2012) Decision making in agriculture based on land suitability- spatial data analysis. *Appl J Theoret Appl Info Technol* 46:1
- Park S, Hamm S-Y, Jeon H-T, Kim J (2017) Evaluation of logistic regression and multivariate adaptive regression spline models for groundwater potential mapping using R and GIS. *Sustainability* 9(7):1157
- Pham B, Jaafari A, Phong T, Mafi-Gholami D, Amiri M, Van Tao N et al (2021) Naïve Bayes ensemble models for groundwater potential mapping. *Eco Inform* 64:101389. <https://doi.org/10.1016/j.ecoinf.2021.101389>
- Pothiraj P, Rajagopalan B (2013) A GIS and remote sensing based evaluation of groundwater potential zones in a hard rock terrain of Vaigai sub-basin India. *Arab J Geosci*. <https://doi.org/10.1007/s12517-011-0512-3>
- Pourghasemi HR, Sadhasivam N, Yousefi S, Tavangar S, Nazarlou HG, Santosh M (2020) Using machine learning algorithms to map the groundwater recharge potential zones. *J Environ Manage* 265:110525
- Prasad P, Loveson VJ, Kotha M, Yadav R (2020) Application of machine learning techniques in groundwater potential mapping along the west coast of India. *Giscience Remote Sens* 57(6):735–752
- Rahmati O, Tahmasebipour N, Haghizadeh A, Pourghasemi HR, Feizizadeh B (2017) Evaluation of different machine learning models for predicting and mapping the susceptibility of gully erosion. *Geomorphology* 298:118–137
- Rao PV, Subrahmanyam M, Raju AB (2021) Groundwater exploration in hard rock terrains of East Godavari district Andhra Pradesh, India Using AHP and WIO Analyses Together with Geoelectrical Surveys. *AIMS Geosci* 7(2):243–267. <https://doi.org/10.3934/geosci2021015>
- Rukhsana HM (2020) Modelling of Potential Sites for Residential Development at South East Peri-Urban of Kolkata. In: Monprapussorn S, Lin Z, Sitthi A, Wetchayont P (eds) *Geoinformatics for*

- sustainable development in Asian Cities ICGGS 2018 springer geography. Springer, Cham
- Shit P, Bhunia G, Bhattacharya M, Patra B (2019) Assessment of domestic water use pattern and drinking water quality of Sikkim, North Eastern Himalaya, India: a cross-sectional Study. *J Geol Soc India* 94(5):507–514. <https://doi.org/10.1007/s12594-019-1348-9>
- Sinha M (2016) Gandeshwari rivulet: a geomorphic study, West Bengal. *India Social Science Review* 2:2
- Sinha AK, Kumar V, Singh P (2018) Delineation of groundwater potential zones using remote sensing and geographic information system techniques: a case study of Udaipur district, Rajasthan, India. *Int Conf Food Security Sustain Agri* 4:265–273
- Soni J, Ansari U, Sharma D, Soni S (2011) Predictive data mining for medical diagnosis: an overview of heart disease prediction. *Int J Comp Appl* 17:43–48
- Thapa R, Gupta S, Guin S, Kaur H (2017) Assessment of groundwater potential zones using multi-influencing factor (MIF) and GIS: a case study from Birbhum district, West Bengal. *Appl Water Sci* 7:4117–4131
- Tolche AD (2021) Groundwater potential mapping using geospatial techniques: a case study of Dhungeta-Ramis sub-basin Ethiopia *Geology. Ecol, Landsc* 5(1):65–80. <https://doi.org/10.1080/2474950820201728882>
- Verma D, Bhunia G, Shit P, Tiwari A (2018) Assessment of groundwater quality of the central gangetic plain area of India using geospatial and WQI techniques. *J Geol Soc India* 92(6):743–752. <https://doi.org/10.1007/s12594-018-1097-1>
- Wang H, Zhang L, Yin K, Luo H, Li J (2021) Landslide identification using machine learning. *Geosci Front* 12(1):351–364
- Wu X, Kumar V et al (2008) Top 10 algorithms in data mining. *Knowl Infor Sys* 14:1–37
- Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.