Characterization of groundwater potential zones in water-scarce hardrock regions using data driven model

Dipankar Ruidas[1] · Subodh Chandra Pal[1] · Abu Reza Md. Towfiqul Islam[2] · Asish Saha[1]

Abstract

The deficiency of freshwater has become a global issue in the recent era, especially in water-scarce hard rock region including India. Groundwater (GW) as a natural resource is decreasing at an alarming rate in West Bengal, India. Sustainable use and planning for better management of groundwater resources are essential; thus, spatial modelling of GW distribution requires proper assessment to conserve and manage the groundwater resource. Machine learning algorithms in RS-GIS environment plays a crucial role in exploration, assessing, monitoring and conserving groundwater resource in this regard. Logistic regression (LR), support vector machine (SVM) and random forest (RF) were used to develop groundwater potential zone (GWPZ) of water-stressed district Purulia with the help of 10 GW controlling factors including geology, geomorphology, lineament density, slope, soil texture, drainage density, GW level, rainfall, NDVI and NDWI. Multi-collinearity analysis was also used to eliminate collinearity issues among all controlling factors. In this study, the total area has been divided into five groups under the very low, low, moderate, high and very high groundwater potential zone categories. It has been calculated that most of the area has moderate groundwater potentiality, which is 29% of the total regions. About 8%, 19%, 25% and 18% of the study area fall under the very low, low, high and very high zones. Finally, all the adopted models were validated through ROC–AUC using GW depth data from CGGB and 484 validating point datasets in this area. The AUC values of adopted LR, SVM and RF models are 0.801, 0.849 and 0.878, respectively; implies that RF is a more reliable algorithm with better predictive ability than other models in the study area. This study's findings will help decision-makers take proper strategies and planning of groundwater resource management for this particular water-scare hard rock region.

Keywords Sustainable development · Purulia · RS-GIS · SVM · AUC–ROC · CGWB · Groundwater management

Introduction

Water is one of the essential constituents for sustaining all forms of life among all renewable resources on earth, which exists in three forms, i.e., solid, liquid and gaseous states. It is a unique gift to mortality from nature, and it is tough to de-pollute, expensive to transport, and impossible to substitute its most uses (Kumar et al. 2005). Three fourth area of the earth's surface is covered with water, out of which the ocean holds about 96.54% that, importantly saline in nature, only 2.8% is available as freshwater (USGS). Over

time, these available water sources are also decreasing at an alarming rate due to the rapid increase of water demand resulting in an increasing number of water-scarce regions. It derives the most significant challenge of supplying adequate water to meet social needs (Ayob and Rahmat 2017). According to (McDonald and Mitchell 2019), worldwide remarkable dropping of freshwater availability put several countries to suffer from water scarcity. Water scarcity is now a global issue; significant numbers of people live in severe water-scarce regions. (Seckler et al. 1999) was estimated that several regions of developing countries, having one-third of the total population, will experience severe water scarcity in the near future. Out of 2.8% freshwater, only 0.76% is available as fresh groundwater (GW), stored beneath the earth's surface through GW recharge (USGS). Water infiltration from unsaturated zone to saturated zone is identified as GW recharge (Yeh et al. 2016); usually formed by entry of rain and snow melts into underlying rock structure through

✉ Subodh Chandra Pal
geo.subodh@gmail.com

1 Department of Geography, The University of Burdwan, Bardhaman, West Bengal 713104, India

2 Department of Disaster Management, Begum Rokeya University, Rangpur 5400, Bangladesh

soil (Nampak et al. 2014).It is a prime source of fresh water in many parts of the world and is essential for human well beings.

As a dynamic and essential renewable natural resource, GW has a crucial role in global climate change and satisfying human needs (Chatterjee et al. 2010; Duan et al. 2016). Seckler et al. (1999) exposed that about 30% of freshwater resources are regarded as GW resources, the second most abundant freshwater resource of the globe. It is the only available water free from pollution due to no direct contact with the earth's surface, contributing around 34% of the total annual water supply (Magesh et al. 2012). The consumption of GW has been increasing rapidly day by day due to the failure of surface water in meeting necessary demands in the daily life of humankind (de Vries and Simmers 2002; Das and Pal 2019). In recent years, the world has been facing threatening conditions regarding the availability of groundwater due to its excessive use. The large dependency of several regions in India and fast consumption of this valuable resource in comparison to its natural recharge leads to a significant decline in the water table beneath the surface (Rodell et al. 2009). In India, 90% of rural domestic water usage is based upon GW. In comparison, 70% of the freshwater used for agricultural purposes is extracted from aquifers, and about 50% of urban water usage is based upon GW (Das et al. 2019). GW is replenishable in nature but finite in distribution; therefore, we should be more concerned about sustainable use (Chatterjee and Purohit 2009). At present, the resiliency of the aquifer system is under a threatening condition due to unsustainable use of GW and evokes a dubitation about the ability to provide a long-term water source (Richey et al. 2015). In the recent era, the problem of rural water supply, drought problem, irrigation project and low cost of development draws attention towards GW management (Kamila et al. 2018); therefore, identification of GW potential zone as well as GW recharge area is critical to maintaining water quality and better management of GW system (Waikar et al. 2014). The GW occurrence and movements of a particular region are affected by several topographical (geology, surface slope, geomorphology, lineament density, NDVI, soil texture) and hydrological factors (drainage density, rainfall, NDWI, GW level); considering all above controlling factors groundwater potential zone (GWPZ) can be identified of an area (Jha and Peiffer 2006; Chowdhury et al. 2009; Singh et al. 2019). Until now, several methods have been adopted to identify the GWPZ of a particular area; traditionally, test drilling and stratigraphy analysis (Sander et al. 1996), soil moisture model and hydrogeological field investigation model (Thorpe and Scott 1999; White et al. 2003). However, all those in-situ investigation methods are cost-effective and notably time-consuming. Remote sensing (RS) and GIS integrated platforms have become a quicker and cost-efficient alternative to identify

GWPZ (Saraf et al. 2004; Solomon and Quiel 2006; Gupta and Srivastava 2010; Hatti 2011; Ghosh et al. 2015; Thapa et al. 2018). Several statistical techniques (MCDA, AHP and Fuzzy logic) were also applied by different researchers (Machiwal et al. 2011; Jhariya et al. 2016; Arabameri et al. 2020; Pal et al. 2020; Mallick et al. 2021) by using expert knowledge or literature values. In groundwater data analysis, strong and flexible analytical methods are required (Rahmati et al. 2015). In such a case, advance data-driven model can aid to delineate groundwater potential zone. Recently various machine learning (ML) algorithms and artificial intelligence such as logistic regression (LR), decision tree, ANN (artificial neural network) random forest, SVM (support vector machine) get huge importance due to large data handling capacities and remarkable accuracy level. Proper approach and modeling are essential to finding out groundwater potentiality and planning for the water-scarce region.

This study has applied a comprehensive GIS technique with the three most significant ML algorithms like LR, random forest, and SVM to delineate GWPZ of Purulia district, India. Stage of groundwater development is an important index to explain groundwater availability in a particular area; based on it, several Safe Category blocks are 215, Semi-Critical blocks are 53, and Critical block is 1 in West Bengal and same way Purulia has a good position in the stage of groundwater development which is 13.78% (GEC 1997). Despite that, Purulia is notoriously famous as a water-scarce region in every summer season. The village women of this area walk a very long distance along village streets in search of water. Various factors are responsible for water scarcity in Purulia. The rapid growth of the population in this area is highly responsible for widening the gap between demand and supply of water in recent times. Very few studies have been done regarding the GW issues of Purulia (Das et al. 2019). Our study intends to close the research gap in the existing literature. The primary objective of our study is to find out the groundwater potential zone (GWPZ) of the Purulia district through LR, Random forest and SVM with the help of the RS-GIS platform. The final zonation map will help predict potential groundwater areas and will also be helpful in proper planning to manage this valuable resource for the future.

Study area

The Purulia district is situated in the south-western part of West Bengal, between 22°43′ and 23°42′ North latitude and between 85°49′ and 86°54′ East longitudes (Fig. 1). It occupies 5th position in the state according to its area, which is 6259. The district is surrounded by several districts of Jharkhand and West Bengal. Hazaribagh–Dhanbad, Singbhum, and Ranchi districts of Jharkhand are situated
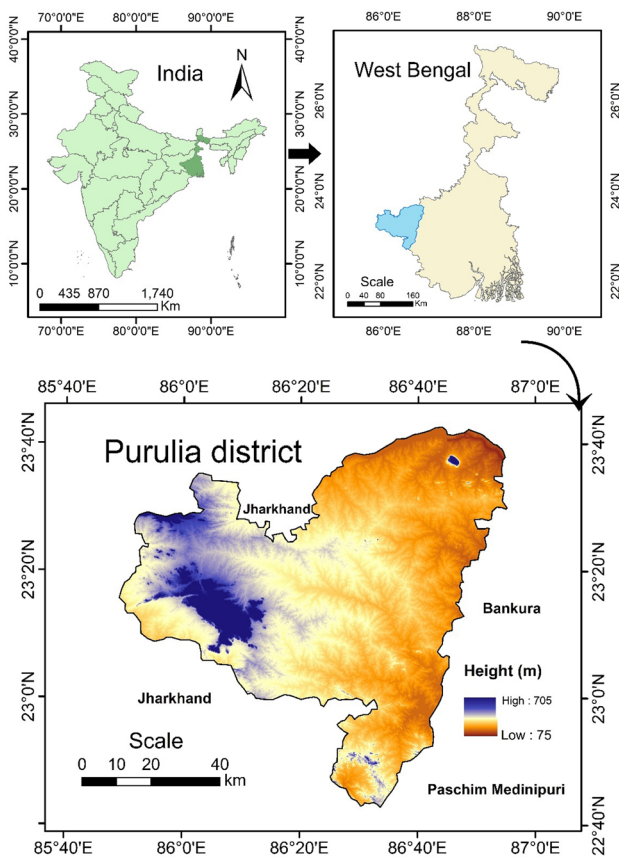
Fig. 1 Location map of study area

in the North, South, and West of Purulia, respectively. On the eastern side, it is bounded by the Bankura, Burdwan, and Midnapore districts of West Bengal. Physio graphically Purulia is located at a transition zone between the plateau of Chotanagpur and the young alluvial plains of West Bengal (Mahato and Gupta 2016). As a part of the Chotanagpur plateau, undulating rugged topography, hills, high ridges, and low valleys are the main characteristics of this study area. The eastern part consists of a rolling upland which changes ultimately into a rugged and dissected topography in the west. The study area has numerous streams namely Kangsabati, Damodar, Darakeswar, Kumari, and Subarnarekha. All the rivers are east and south east-flowing, except for Subarnarekha which flows in the south direction. There are some alluvial areas found in very narrow strips along the rivers. Kangsabati is the master stream of the area. Though many rivers flow through the region, more than 50% of it runs off due to undulating topography, resulting in low infiltration. Porosity and permeability are very low due to hard rocky terrain surfaces; that is mainly dominated by metamorphic rocks. Shallow low fractures and weathered mantle are suitable for GW storage which remains in unconfined or semi-confined condition. Purulia district mainly falls under

sub-tropical climate and is characterized by low precipitation and high evaporation, making it a drought-prone area of West Bengal. The district represents extreme climatic conditions with a very low temperature of 5 °C and maximum temperature of 46 °C in winter and summer months, respectively. Average annual rainfall of 1000–1600 mm occurs in the month of mid-June to September by southwest monsoons. There are three types of soil namely residual types derived from weathering of igneous rocks and metamorphic rocks, lateritic soils in the upland areas, clayey loam to clay in the valleys. The soil pH ranges between 5.5 and 7.2 (Lok Kalyan Parishad, Purulia); 75.05 thousand hectares of area under forest (District Industrial Profile Purulia, 2018–19). The study area is endowed with natural forest in the North-Western part and some areas of Ajodhya Hills and Panchet hills in the Northeast. The forest is mainly composed of Sal trees with different other species like Kusum, Palash, Mahua, Sirish, Neem, Simul, etc.

Materials and methodology

Figure 2 represents the methodological framework used in this study.

Materials used

A truthful GWPZ map and its accuracy level depend on the size and availability of the data sets as well as used modeling approaches to produce the maps. Some factors which affect the groundwater potentiality have been considered in this study through literature review. They are Geology, Slope, Lineament Density, Normalized Difference Vegetation Index (NDVI), Geomorphology, Drainage Density, Normalized Difference Water Index (NDWI), Rainfall, GW level, and Soil texture. The maps of all these factors are prepared by the collection of data from different sources. Geology and geomorphology map of the study area has been collected from the Geological Survey of India (GSI). Slope, lineament density, and drainage density maps are prepared from Shuttle Radar Topography Mission 1 arc-second for global coverage (USGS, 30 m resolution). NDVI and NDWI maps were prepared from Landsat 8 (USGS) data. Rainfall and groundwater level maps are prepared by Inverse Distance Weighting (IDW) method with the help of Indian Meteorological Department rainfall data and Central Groundwater Control Board (CGWB) respectively. Soil texture data have been collected from the National Bureau of Soil Survey and Land Use Planning (NBSS and LUP), Government of India. Arc-GIS 10.4 and ERDAS IMAGINE 2014 software have been used for the representation of the data. R software and SPSS statistical software were used for adopted ML algorithms to delineate GWPZ with the help of available data sets.

Fig. 2 Methodological framework adopted for this study
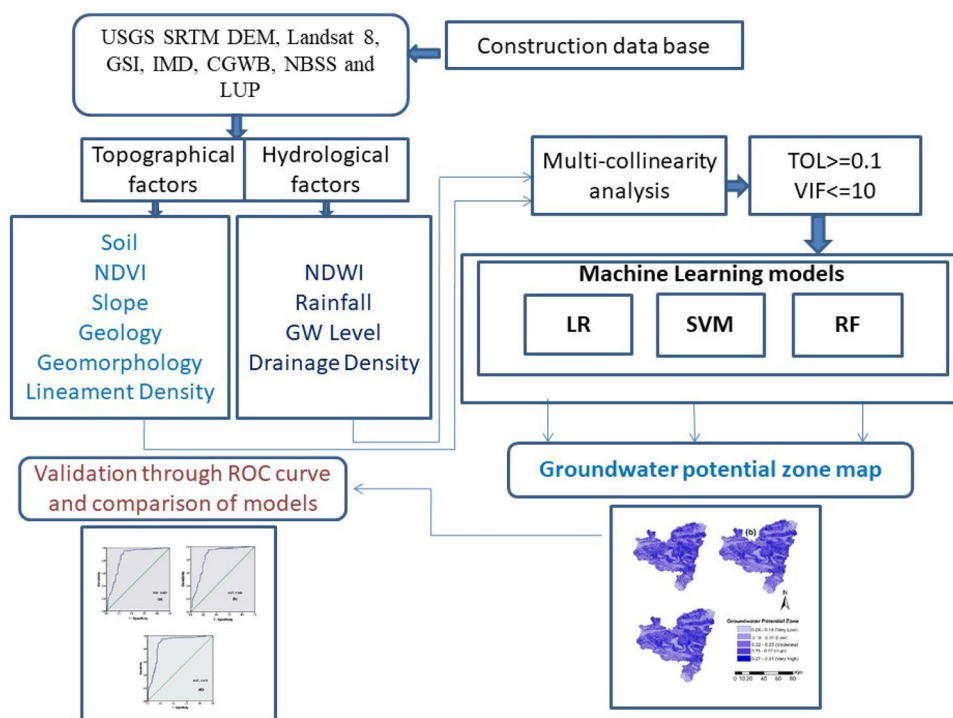
Methodology

Multi-collinearity analysis

In a regression model to ensure the accuracy, testing of independence of each factor is very important (Chen et al. 2011). Multi-collinearity occurs when more than one variable is significantly correlated to each other in a regression model; a big problem can be resulted through a small percentage of multicollinearity (Daoud 2017). For this study, several causative factors were used, therefore it is essential to identify the linear relationship among the variables through this test which helps to exclude associated variables to eliminate the error chances and increase the accuracy in adopted models. In this study, variance inflation factors (VIF) and Tolerance (TOL) techniques play a significant role in defining errors. The TOL and VIF were calculated as follows:

$$TOL = 1 - r^2 \tag{1}$$

$$VIF = \frac{1}{TOL} \tag{2}$$

where the VIF value > 10 and TOL value < 0.1 shows multi-collinearity issue (Khosravi et al. 2019). For present study, the threshold value for GWPZ were considered to be less than 5 VIF value.

Logistic regression

Logistic regression is the most widely used multivariate statistical techniques in diverse field of science, significantly in environmental science (Nguyen et al. 2020; Chowdhuri et al. 2020). In a generalized linear method this empirical model is based on logit link function; by considering several conditioning factors it enhances the prediction accuracy (Yariyan et al. 2020). In LR, categorical outcomes can be predicted i.e., occurrence and non-occurrence; possibility of occurrence of an event is assessed within the range of 0–1. The presence and absence of attribute can be predicted by considering several independent variables in binomial LR when the dependent variables are in binomial nominal level whereas multivariate LR is used to measure the relationship between various independent variables with one dependent variable. Generally, LR model expressed in following equations (Ozdemir 2011; Park et al. 2017):

$$P = \frac{e^z}{(1 + e^z)} \tag{3}$$

where the probability of occurrence is indicated by P and the linear combination of the independent variable by Z. Z is expressed as follows:

$$Z = \alpha + \beta_1 x_1 + \beta_2 x_2 \ldots \ldots \ldots + \beta_n x_n \tag{4}$$

where α defines the intercept, n is the number of independent variables whereas β_n and x_n represent the regression coefficients and independents variables, respectively. Z value ranges from $-\infty$ to $+\infty$; the positive regression coefficient defines the positive correlation of dependent variable with independent variables whereas the negative regression indicates the negative relation in between dependent and independent variables.

Support vector machines

Support vector machine learning is one of the most significant supervised algorithms used as a classifier besides regression problems; it is performed based on statistical learning theory and structural risk minimization (SRM) principle (Tien Bui et al. 2012; Tehrany et al. 2015). The SRM principle is an important factor in SVR in identification of the relationship between input and output variables (Saha et al. 2021). However, these learning algorithms are mostly used for classification purposes by developing hyperplane which is decision boundary to segregate n-dimensional space into classes. (Vapnik 1998) proposed that by maximizing the margin between the classes, hyperplane classify the linear data. The epsilon tolerance margin is considered, that makes the difference between SVM and SVR which is the regression version of SVM (Smola and Schölkopf 2004).Generally, the SVR modelling approach applied in case of very complex dataset; several curved margins are developed to solve this dataset (Kalantar et al. 2018).

Following mathematical formula can describe the SVM model:

$$\frac{1}{2}||W||^2 \tag{5}$$

The following constraints of subject is

$$y_i = (w \cdot x_i) + b \geq 1 \tag{6}$$

where, $||W||$ is the hyperplane, b is the scalar base, ($\bullet$) indicates the scalar product. The cost function of SVM can be defined by using Lagrangian multiplier:

$$L = \frac{1}{2}||W||^2 - \sum_{i=1}^{n} \lambda_i \big(y_i \big((w \cdot x_i) + b \big) - 1 \big) \tag{7}$$

where,λ_i indicates Lagrangian multiplier. In the case of nonseparable function the constraints can be modified by introducing slack variables

$$y_i(w \cdot x_i) + b \geq 1 - \zeta_i \tag{8}$$

And finally the equation becomes as follow

$$L = \frac{1}{2}||W||^2 - \frac{1}{vn} \sum_{i=1}^{n} \zeta_i \tag{9}$$

where, v (0, 1) represent the generated to account for misclassification. In addition to this, kernel function $K(x_i, x_j)$ was introduced by Vapnik in the year of 1995 to explanation for non-linear decision boundary.

Random forest

Random forest (RF) decision tree algorithm was first developed by Breiman (2001); as an advanced version of bagging, randomness added to it consisted of tree predictors and each tree depends on the values of a random vector sampled autonomously with the similar distribution for all trees in the forest (Breiman 2001). Two types of trees such as classification trees and regression trees are separated from decision trees (Rodriguez-Galiano et al. 2014). RF as a flexible ensemble learning algorithm constructs a set of classifiers rather than one classifier based on multiple decision tree helps to produce repeated predictions of similar phenomenon (Ok et al. 2012; Micheletti et al. 2014; Zabihi et al. 2016; Islam et al. 2021). RF classification algorithm helps to develop new training sites by choosing sample training sites. In this study, RF machine learning approach was used to identify key variables for mapping GWPZ. Random forest requires two components (Goel and Abhilasha 2017):

$$GE = P_{x,y}(mg(x, y) < 0) \tag{10}$$

$$mg(x, y) = a_{vk}I\big(h_k(x) = y\big) - \max_{j \neq k} av_k I\big(h_k(x) = j\big) \tag{11}$$

where x and y are groundwater potential parameters reflecting the probability of x and y, the edge part of the indicator is mg and $I(*)$.

Model validation

Model validation is a prominent part to evaluate a model that can prove the scientific reliability of the result. Therefore, the widely applied Receiver Operating Characteristics (ROC) and Area Under Curve (AUC) curve are used to determine the accuracy of the models(Chen et al. 2018; Gayen et al. 2019). The key focus of this spatial modelling approach is to quantify the successful and unsuccessful events on the basis of model testing data; it works independently to assess the model's predictive ability of a specific probability threshold. According to Mandrekar (2010) this curve is regarded as the average value of the specificity of a test over all possible values of a sensitivity or vice versa. AUC values range between 0.5 and 1, where values closer 1 denote excellent performance and nearer to 0.5 indicate the

very poor predicting accuracy (Sameen et al. 2020). In this study, three models were adopted to delineate GWPZ, therefore comparative review of the adopted models executed by this validation technique. Dou et al. (2019) asserts that the trapezoidal rule integral calculus helps to calculate AUC value:

$$AUC = \sum_{k=1}^{n} (X_{k+1} - X_k)(S_{k+1} - S_k - S_k/2) \qquad (12)$$

where area under curve represented by AUC and in case of specificity and sensitivity is represented by X_k and S_k, respectively.

Result and discussion

Multi-collinearity analysis

Extensive literature review helps to identify several groundwater conditionings factors. In the present study, 10 groundwater affecting factors were identified after positive multi collinearity analysis. The results of multi collinearity analysis were shown in Table 1. The results display that the lowest TOL is 0.27 and the highest value of VIF is 3.71; that indicate there are no multi-collinearity issues among adopted 10 groundwater conditioning factors.

Geology

Geology is a major controlling factor in the groundwater potentiality of a particular area that generally affects groundwater recharge by supervising water flow percolation(Elewa and Qaddah 2011) and subsequently it is associated with water permeability (Oikonomidis et al. 2015). The occurrence, movement, and qualities of groundwater are significantly affected by the geological units of an area (Rajaveni et al. 2017). According to the Geological Survey of India, the

Purulia district consists of seven separable major lithostratigraphic units which are Chotanagpur gneissic complex, unclassified metamorphic, Singhbhum GP, Dalma volcano, Manbhum granite, Kuilapal granite series, and some places covered with sediments (Fig. 3). The study area is mainly characterized by numerous rocks and minerals such as granite, gneiss, schist, phyllite, quartzite, sandstone, shale, mica, feldspar, china clay which are abundant in nature. Chotanagpur gneissic complex is present throughout the whole district with an area of 3465 (55.38%). The southern portion of the district is mostly dominated by the Singhbhum GP series which covers 1159 (18.53%) area. Some part of this district is covered with Dalma volcanic series which have an area of 338 (5.41%). Some part of the northern and central region is dominated by manbhum granite and unclassified metamorphic which are distributed in a scattered way with 552 (8.82%) and 531 (8.5%) area respectively. Sedimentary rocks are also present in a very small part of the central region of the district with an area of 135 (2.16%). According to Dolui et al. (2016), Proterozoic hard granite gneiss including soft phyllite and mica-schist of the Singhbhum group are the dominant rocks in the Purulia district as a part of the Chotanagpur plateau. Hardy rocks like granite, gneiss, quartzite, feldspar, etc. are mostly found in this area; that makes the barrier to infiltrate the surface water and increase the surface runoff which is not suitable for groundwater potentiality.

Table 1 Multi-collinearity test of groundwater potential factors

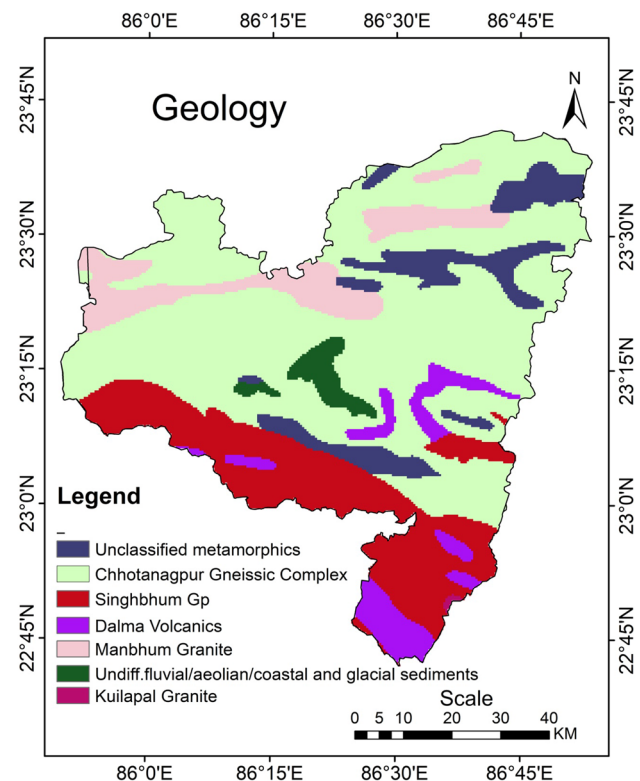| Variables	VIF	TOL
Geology	1.43	0.70
Geomorphology	2.1	0.47
Lineament density	2.32	0.43
Slope	2.51	0.39
Soil	1.59	0.63
NDVI	2.41	0.41
Drainage density	2.36	0.42
Rainfall	3.71	0.27
NDWI	1.87	0.53
GW level	2.26	0.44

Fig. 3 Geology map

But there is some strip area covered with sandstone which is very preferable for groundwater potentiality in this district.

Geomorphology

Geomorphology of a particular area plays a dominant role in estimating the storage and movement of groundwater (Sinha et al. 1990; Dinesh Kumar et al. 2007). Many researchers (Waters et al. 1990; Krishnamurthy et al. 1996; Sahu and Sahoo 2006) have successfully adopted geomorphology as a principal parameter in delineating groundwater potentiality. The identification of the geomorphologic features is very important to delineate groundwater potential zone. Pediment pediplain, dissected hills, valleys are the main geomorphic features of this region but water bodies, active flood plain, rivers are situated in a very small part of this region (Fig. 4), (Geological Survey of India). Most of the areas of the district are dominated by high hills and sloppy land. Pediment and Pediplain complex are present in almost the entire district covering an area of 5458 (85.97%). Apart from it, 9% of the district is covered with high hills and valleys which are not suitable for groundwater potentiality. Ramaiah et al. (2012) stated that structural hill, residual hill, and linear ridge landforms are not suitable for GWP due to its un-fractured rock characteristics having a low infiltration rate. Only a few areas of this district are covered with water

bodies, dams and reservoirs, rivers, and ponds consisting of only around 7% of the study area. The water bodies play a very significant role in the groundwater potentiality of this rugged and hardy terrain. As a result, only a few areas have high potentiality.

Lineament density

Generally geologic structures such as fractures, faults and discontinuous surfaces are defined as lineaments that can be identified by RS-GIS techniques (O'Leary et al. 1976); that is architecture of the rock basement (Hobbs 1904) and significantly important in hydrological studies which works as pathways of groundwater movement (Abdalla 2012) and as a curvilinear feature of earth having a significant role in GW exploration (Pothiraj and Baskaran 2013). Lineament is basically a permeable zone and has positive relation with high GWP (Haridas et al. 1998). According to Krishnamurthy et al. (2000), the buffer area of 300 m around lineament and faults is considered as suitable groundwater recharge and potential zone. Lineament density (L_d) is expressed by following equation:

$$L_d = \frac{\sum_{i=1}^{i=n} L_i}{A} \tag{13}$$

where $\sum_{i=1}^{i=n} L_i$ defines the total length of lineaments and A is the total area (L^2).

The distribution of lineament concentration is not uniform throughout the district. This has been classified into five groups namely very low, low, medium, high and very high-density area. Among them maximum portion of the study area has very low lineament density zone with around 5273 km^2 (83%). Some part of the northern region, southern region and western hilly region are fall under low 475 km^2 (7.48%), moderate 300 km^2 (4.73%), high 205 km^2 (3.23%) and very high density 93 km^2 (1.43%) area (Fig. 5). Very small region of western hilly area of the district at Bagmundi, have greater lineament density which is suitable for groundwater recharge and increases groundwater potentiality. Similarly, the entire study area having very low lineament density is mostly responsible for least groundwater recharge.

Slope

Slope gradient is one of the most important parameters in zone identification of groundwater potentiality of an area; it has a direct influence on the infiltration of surface water (Selvam et al. 2015). Generally, the slope is the ratio of change in elevation between two points and has an inverse relation with GWP. On a steep slope, water flows downward quickly having inadequate time to infiltrate that resulting
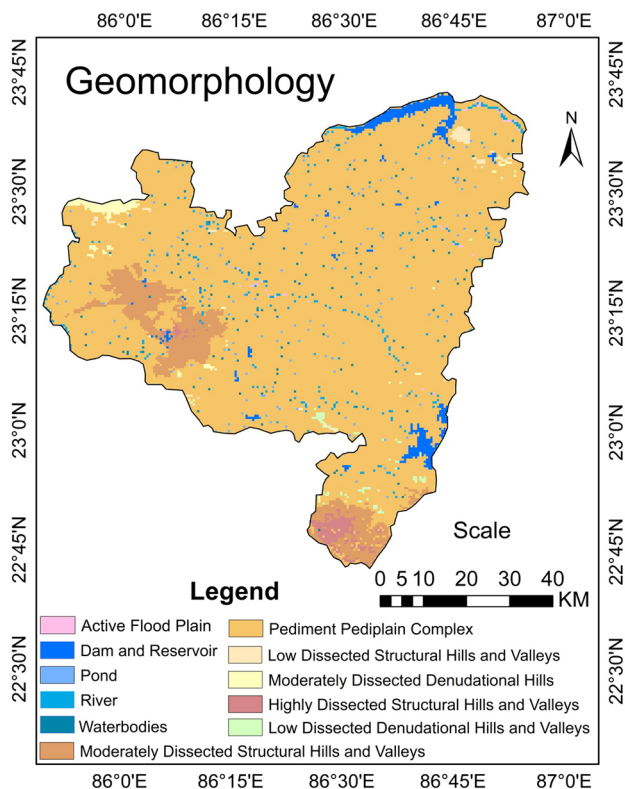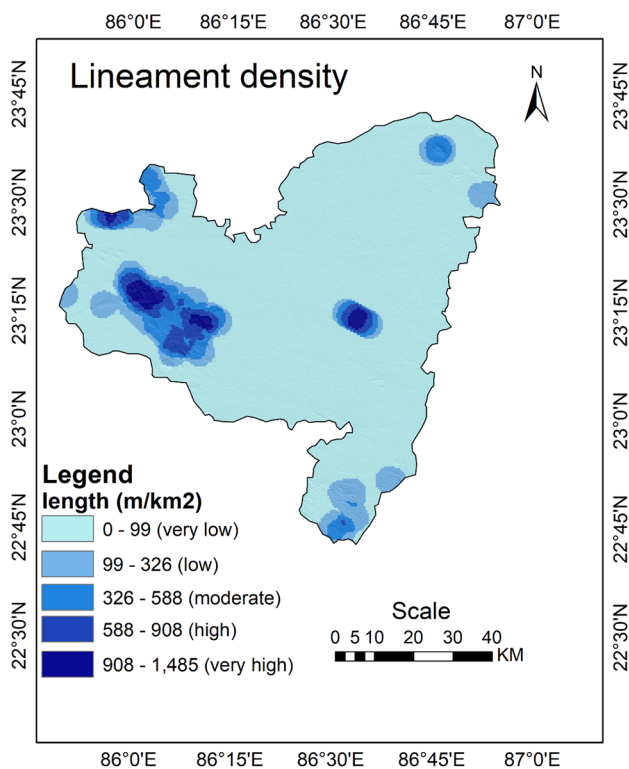

Fig. 4 Geomorphology

Fig. 5 Lineament density map

Fig. 6 Slope map

in very low GW recharge (Machiwal et al. 2011; Magesh et al. 2012). Whereas, gentle slopes provide an environment for GW recharge due to widespread high retention of surface water. According to Saraf and Choudhury (1998), a break in the slope accelerates infiltration leading to more groundwater recharge. Here, the study area is categorized into five zones viz, very low, low, moderate, high, and very high. Among them, the greater part of the study area is characterized by undulating topography which falls under very low (0–2 degrees) and low (2–5 degrees) categories covering 3408 (53%) and 2275 (35%) areas, respectively. This region with a very low degree of slope is very suitable for groundwater recharge. Western, Southern and NE parts of the district are under moderate (5–12 degree), high (12–22 degree) and very high (22–65 degree) category with an area of 461 (7.27%), 142 (2.25%) and 59 (0.93%) respectively (Fig. 6). These three regions are not suitable for groundwater recharge and can be designated as low groundwater potentiality areas.

Soil texture

Punmia and Jain (2005) asserts that to delineate GWPZ, the soil cover is an essential factor; Porosity and permeability of soil plays a significant role in the infiltration of surface water which is based on soil characteristics
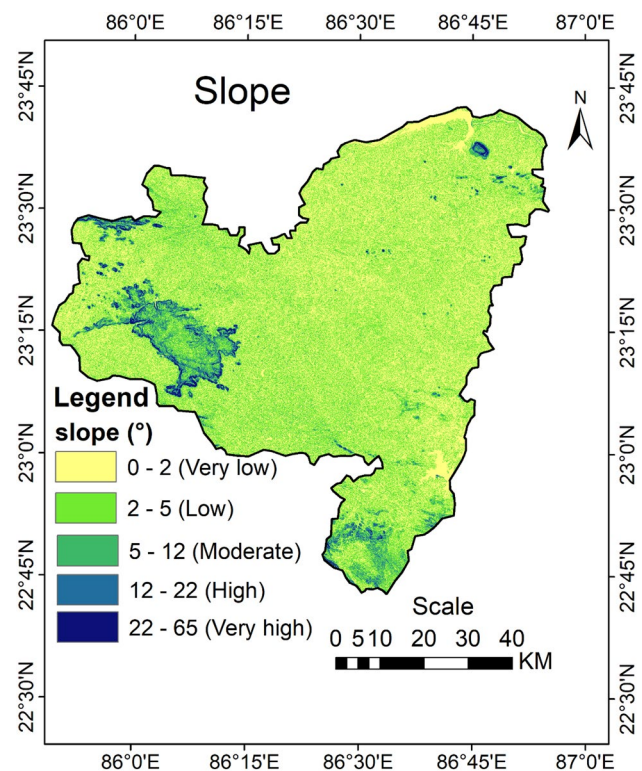
such as texture, depth, and water transmission capacity of an area that control the magnitude of groundwater recharge (Terence and Viessman 1977). Generally, soil texture means the composition of soil that refers to small (clays), medium (silts), and large (sands) size particles. Groundwater recharge is significantly controlled by soil texture because different soil texture has different infiltration rate; fine-grained soil has very low GWP compared to Coarse-grain soil because of its low level of porosity and permeability. This study area is characterized by very thin soil cover, containing sandy and reddish laterite material resulting from granite and gneiss through weathering process (Das et al. 2019). Different parts of this district are characterized with different soil textures such as gravelly loam–loam soil constitutes 1704 (27%), fine loam soil covered 1989 (31%), fine soil fall under the area of 1418 (22.45%), gravelly loam and fine loamy- coarse loamy soil present in the 378 (5.98%) and 772 (12.23%) area, respectively (Fig. 7). This gravelly loam to loam, gravelly loam, and coarse loam soil has moderate porosity and permeability which is preferable to groundwater potentiality and fine and fine loam soil have the least porosity and permeability which leads to the least potential zone.
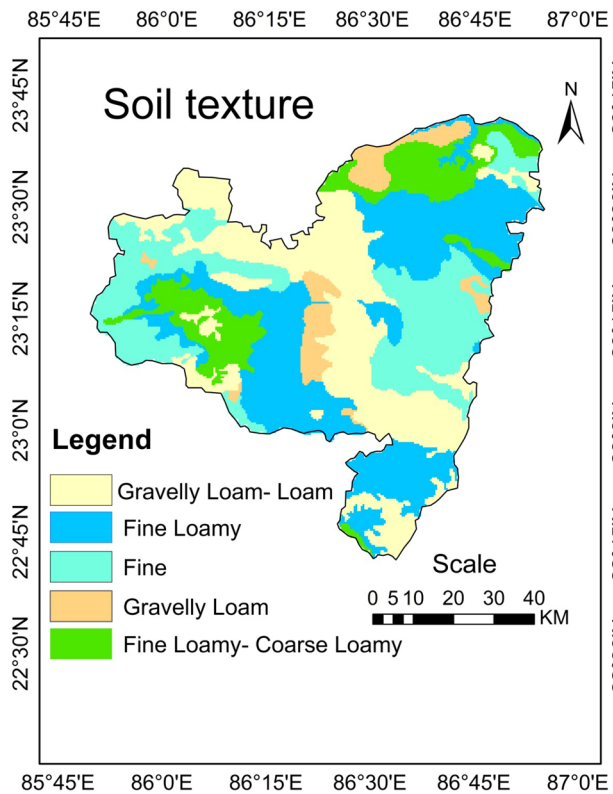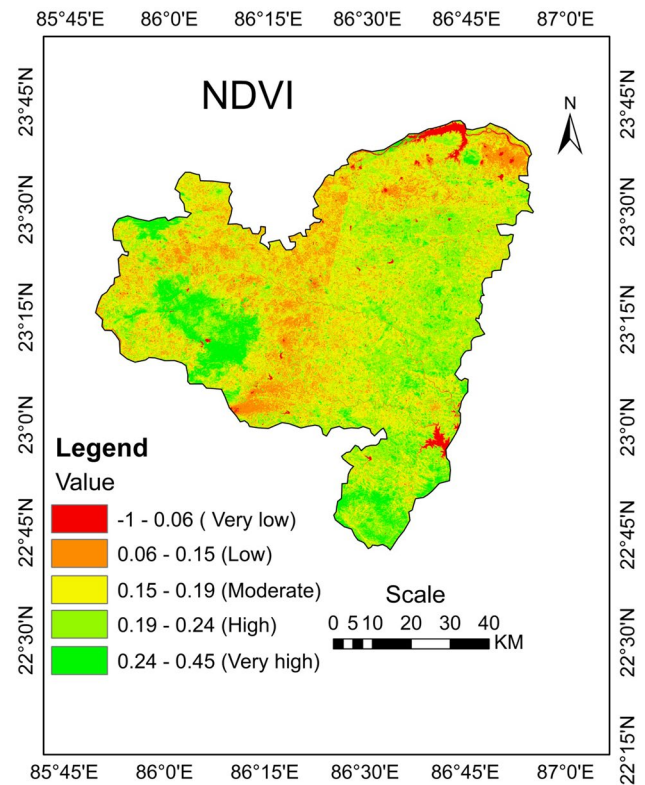
Fig. 7 Soil texture map

Fig. 8 NDVI map

Normalized difference vegetation index (NDVI)

NDVI is a very common and broadly used remote sensing index (Bhandari et al. 2012); generally, it helps to measure live green vegetation and also analyzes the characteristics of vegetation cover of a certain area. The groundwater enrichment can be directly reflected by surface vegetation (Huajie et al. 2016). The quantity of green vegetation in a particular area is an essential factor of groundwater recharge; if a large area is covered with vegetation, it will be suitable for groundwater recharge because the area will gain water holding capacity and get sufficient time to infiltrate the surface water in the groundwater layer. As a result, the groundwater availability of that area increases. This spatial index derived through following equation:

$$NDVI = \frac{NIR - RED}{NIR + RED} \tag{14}$$

where NIR represent the near infrared band and RED is red band in satellite image.

The largest part of the study area is covered with light vegetation and few parts have very high vegetation cover in a scattered way. The entire area is classified into five NDVI zones: very low, low, moderate, high and very high region. Among them maximum part of Purulia district (Fig. 8) fall under moderate, high and very high vegetation

type covering 2594 km^2 (41%), 2097 km^2 (33%) and 755 km^2 (11%) area respectively. These areas are responsible for moderate to high groundwater potentiality. The study area also has low to very low vegetation which covers 778 km^2 (12%) and 122 km^2 (2%) area respectively but they are not suitable and results low ground water potentiality.

Drainage density

Avtar et al. (2011) defined drainage density as the total length of the entire stream in a particular basin divided by its total area; that is a quantitative measurement of the length of stream channels (Singh et al. 2014) and also as spacing of stream channels in a specific area. It helps to measure the drainage concentration in a specific area which denotes how well and poorly drained the area is by stream channels. A lot of work (Bagyaraj et al. 2013; Jenifer and Jha 2017; Thomas and Duraisamy 2018; Andualem and Demeke 2019) established the inverse relation between drainage density and permeability of aquifers; which plays an important role in the runoff distribution and level of infiltration. There is a close relationship between drainage density and infiltration rate of the surface water in a specific area. It has a direct relationship with slope but is inversely proportional to permeability. Based on the result, this study area is categorized into five drainage density zones. They are very low, low, moderate,
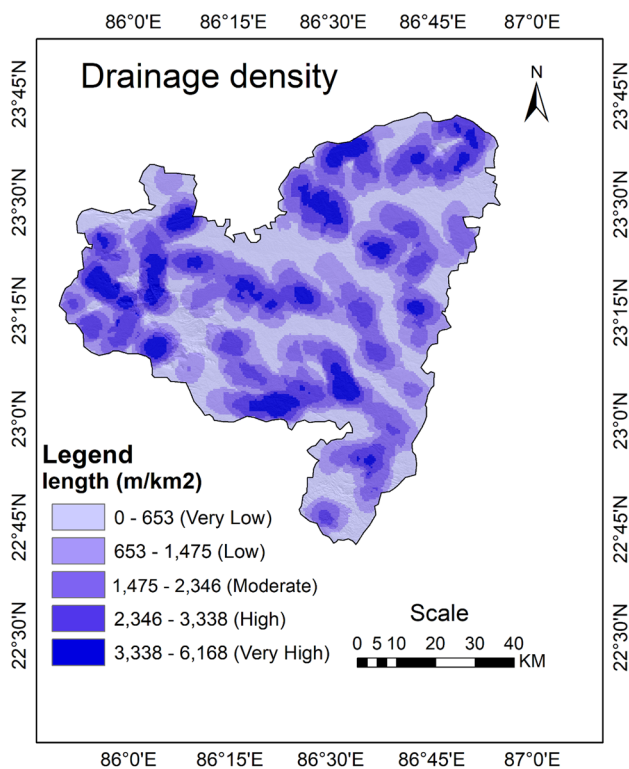
Fig. 9 Drainage density

Fig. 10 GW level map in the study area

high and very high zones covering 1736 (27.3%), 1529 (24.9%), 1515 (23.86%), 1071 (16.87%), and 495 (7.81%) of the study area (Fig. 9) respectively. The result depicts that average drainage densities are found throughout the district. High drainage density is the result of more surface runoff which is not suitable for the groundwater potential zone but indirectly it helps in groundwater recharge in those areas where the slope is very low. The western part of this district, as well as some areas of the northern and southern parts have steep slopes causing a high concentration of drainage which is not suitable for GWP. But most of the area is characterized by a gentle slope which is very suitable for the groundwater potentiality of the Purulia district.

Groundwater level

The groundwater depth of an area fully depends on the recharge and discharge of the subsurface water and it shows the spatial differentiation in the depth of water level of a specified area. This is an important factor for GWPZ (Jhariya et al. 2021); the connection between the geological formation and average water depth carries the ability of its water transmission, storage, and discharge capacity (Indhulekha et al. 2019). Therefore, where the shallow water depth indicates high water availability and deeper water levels define water scarcity instead of its abundance. Based on groundwater depth data, this area is categorized into five groups such as very low, low, medium, high,
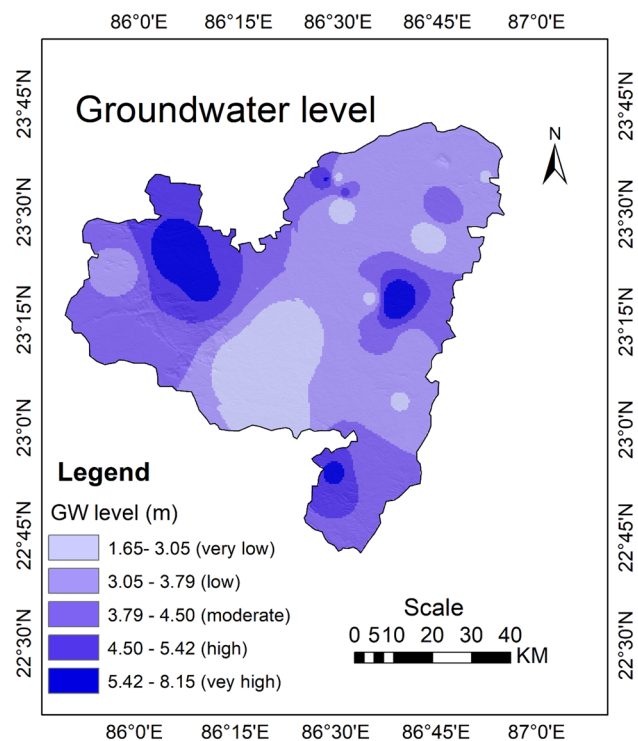
and very high groundwater level zone; which ranges from 1.65 to 8.15 m (Fig. 10). Higher water depth area denotes lower groundwater potentials and lower depth denotes higher groundwater potentials in this area.

Rainfall

Rainfall plays a vital role in groundwater recharge of a particular region; without rainfall, recharge of groundwater is impossible. Purulia district is a part of the Western Plateau region as a result it falls under the dry tropical climatic regions. The study area experiences maximum rainfall only in the monsoon period (Jun–Sep) (Fig. 11). According to the Govt. of West Bengal (1985), the average annual rainfall of the study area is around 1200–1400 mm. The key source of rainfall in this region is the southwest monsoon. The study area experienced a drought situation throughout the year except for monsoonal months. Eastern and South-Eastern parts of this district receive a high amount of rainfall compared to other regions. It indicates high groundwater potentiality in these particular areas. The western part of this district receives the least amount of rainfall and is not suitable for groundwater potentiality. This little variation of rainfall does not make any greater variability of groundwater in this district because here the major dominating factor is the geology and geomorphology which control the surface runoff of rainwater due to hardy rock terrain.
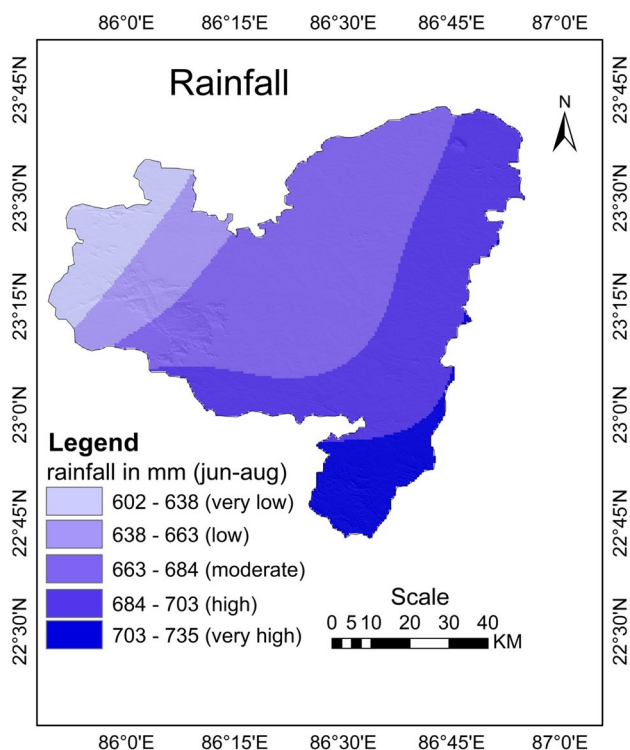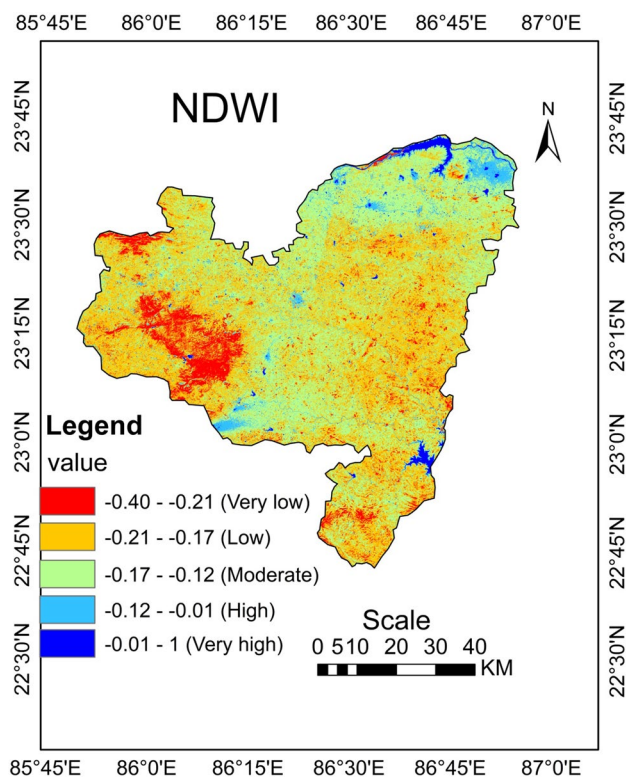
Fig. 11 Rainfall map

Fig. 12 NDWI map

Normalized difference water index (NDWI)

NDWI is mainly used to delineate the different water bodies as surface water viz, ponds, reservoirs, rivers, lakes etc. which are more responsible for sub surface water in a region with the location of permeable rocks. There is a strong relationship between surface and sub-surface water; the reasonable number of surface water source makes good groundwater resource by normal hydrologic cycle. The communication between surface water and groundwater is mostly controlled by an intricate relationship between hydrological, geological, geomorphological, climate and landscape factor. This index is calculated as follows:

$$\text{NDWI} = \frac{\text{GREEN} - \text{NIR}}{\text{GREEN} + \text{NIR}} \quad (15)$$

NDWI calculated through green band and NIR band of satellite image.

In this study, Purulia district is divided into five areas on the basis of the result which can help to understand the amount of surface water situation in this region. Those five categories are very low (9%), low (42%), moderate (41%), high (6%) and very high (2%) which cover the entire area (Fig. 12). This is clear that very few areas are under the high and very high surface water condition but it is least influence to groundwater due to its small areal extent throughout

the district. Majority of the study area have less amount of NDWI index which implies least amount of surface water present in this area which have negative impact on GWPZ.
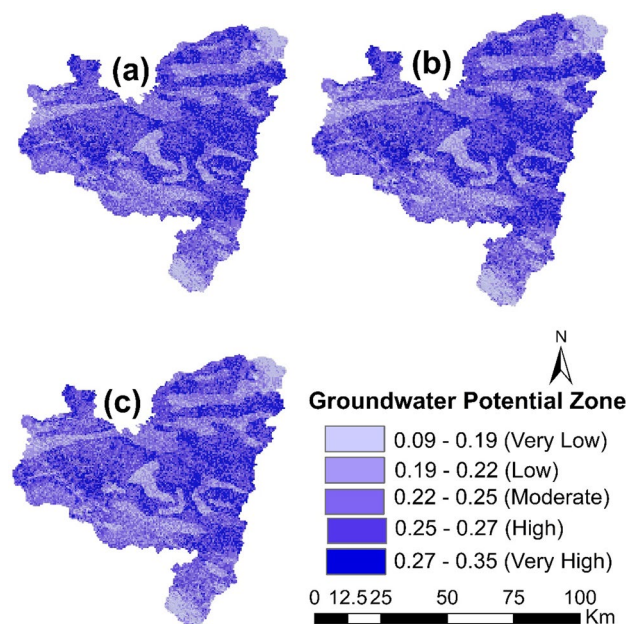

Fig. 13 Delineation of groundwater potentiality by using—**a** LR, **b** SVM, **c** RF

Affecting factors	Classes	Coefficients
Table 2 Groundwater potential factors, their classes and respective logistic regression coefficients		
Geology	Unclassified metamorphics	0.632
	Chotanagpur gnesis complex	2.158
	Singhbhum GP	9.632
	Panchet/pachmarhifm	1.253
	Dalma volcanics	−1.547
	Undiff. fluvial/aeolian/coastal/glacial sediments	0.512
	Manbhum Granite	3.102
Geomorphology	Pediment pediplain complex	1.235
	River	−0.258
	Moderately dissected structural hills/valley	3.014
	Highly dissected structural hills and valleys	5.023
	Dam & reservoir	0.953
	Pond/water bodies	0.254
Lineamnet density	0–99 (very low)	0.279
	99–326 (low)	
	326–588 (moderate)	
	588–908 (high)	
	908–1485 (very high)	
Slope	0–2 (very low)	0.649
	2–5 (low)	
	5–12 (moderate)	
	12–22 (high)	
	22–65 (very high)	
Soil texture	Gravelly Loam–Loam	2.314
	Fine loamy	0.023
	Fine	−0.623
	Gravelly loam–loam	3.025
	Fine loamy–coarse loamy	1.542
NDVI	− 1–0.06 (very low)	2.314
	0.06–0.15 (low)	0.023
	0.15–0.19 (moderate)	−0.623
	0.19–0.24 (high)	3.025
	0.24–0.45 (very high)	1.542
Drainage density	0–653 (very low)	0.349
	653–1475 (low)	
	1475–2346 (moderate)	
	2346–3338 (high)	
	3338–6168 (very high)	
Rainfall	602–638 (very low)	−0.01
	638–663 (low)	
	663–684 (moderate)	
	684–703 (high)	
	703–735 (very high)	
GW level	1.65–3.05 (very low)	1.245
	3.05–3.79 (low)	
	3.79–4.50 (moderate)	
	4.50–5.42 (high)	
	5.42–8.15 (very high)	

Table 2 (continued)

Affecting factors	Classes	Coefficients
NDWI	− 0.40--0.21 (very low)	0.943
	− 0.21--0.17 (low)	
	− 0.17--0.12 (moderate)	
	− 0.12--0.01 (high)	
	− 0.01–1 (very high)	
Intercept = 0.241		

Fig. 14 Determining the relative importance of causative factors using RF
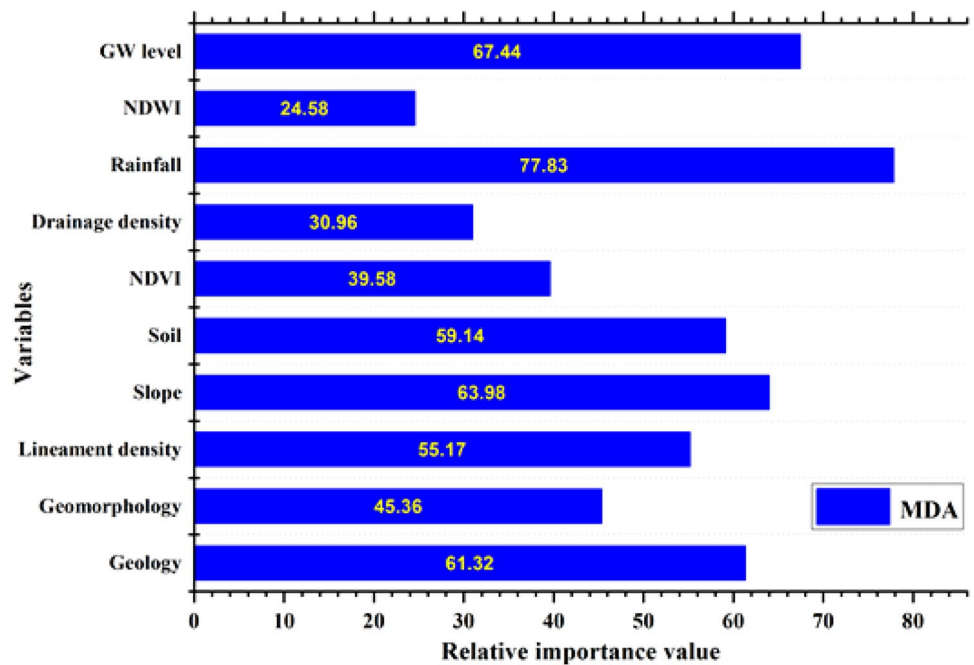
Table 3 Area coverage (%) of different zones by different methods

Groundwater potential zone	LR	SVM	RF
Very low	7.95	8.89	8.79
Low	19.72	19.94	19.91
Moderate	28.71	28.82	28.86
High	25.54	24.8	25.16
Very high	18.06	17.53	17.25

Groundwater potential zone and validation

Groundwater potential zone

Groundwater potential zonation mapping is an essential task for sustainable use and planning of groundwater management. As Purulia district is a drought-prone area, it is necessary to know about the exact location and zone of groundwater storage which in turn helps to make proper water management. In this study, three machine learning methods such as LR, RF, and SVM methods have been used to delineate the GWPZ of the Purulia district. In the current study, the stepwise LR model was used to analyze the data and the model coefficient was calculated by using SPSS software; therefore, that was used to produce GWPZ (Fig. 13a). According to the LR model, the Positive coefficient plays a significant role to make groundwater recharge whereas negative values indicate insignificant roles in groundwater potentiality (Table 2). Dominating groundwater affective factors are as follows Singbhum GP (9.632), Highly dissected structural hills and valleys (5.023) and Gravelly Loam–Loam (3.025) whereas, Dalma volcanics (−1.547), fine soil (−0.623), and rainfall (−0.01) have an insignificant role in GW recharge in this area. SVM method helps to reclassify all the data layers and make a reliable result in GWPZ mapping (Fig. 13b). RF modeling approach also gave the most suitable and acceptable result in this study area due to its small spatial coverage (Fig. 13c); by calculating relative importance values (Fig. 14). The results illustrate
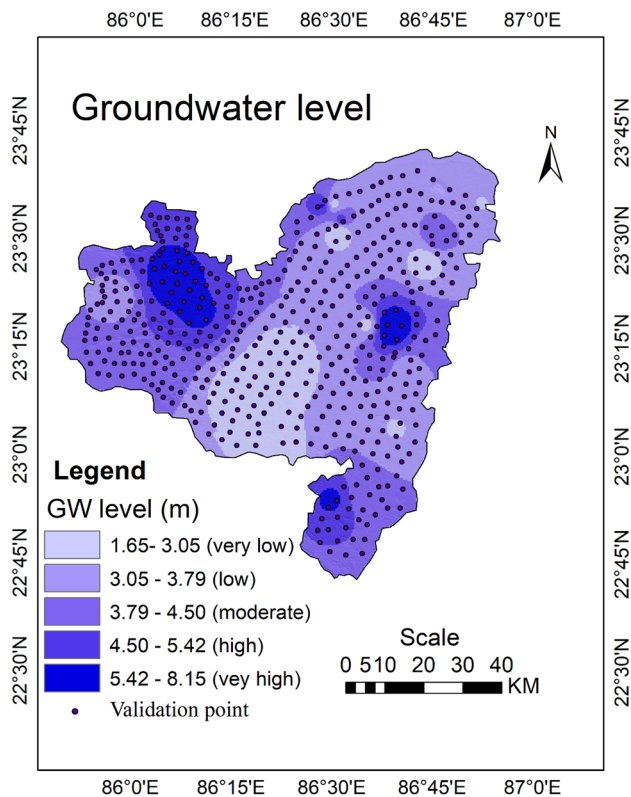
Fig. 15 Validation point of groundwater potentiality

that rainfall (77.83), GW level (67.44), slope (63.98), geology (61.32), soil (59.14) and lineament density (55.17) have great impact whereas, NDVI (39.58), NDWI (24.58), drainage density (30.96) and geomorphology (45.36) have the least impact on groundwater potentiality. All the adopted models in this study construct more or less similar results for groundwater potentialities in this region. Based on this result, the Purulia district has been categorized into five potential zones, those are very low, low, moderate, high, and very high area; spatial coverage all the models are shown in (Table 3). Very low, low potentiality zones are scattered in the Northern, North Eastern, Western, Southern, and Central parts of the district. The potentiality areas under high and very high zones are also found throughout the district in a scattered way which is more or less half of the study area. This assessment indicates that groundwater potentiality is restricted throughout the majority portion of this district while it is relatively high in a small segment of the northeastern and north-western part of the study area; water scarcity is a common thing here because of rugged terrain and hardy rock surface. Though sufficient water potentiality is present in some places of the district, we have to manage it appropriately.

Validation of groundwater potentiality

The Receiver Operating Characteristics (ROC) curve has been used in this study for reviewing the result of groundwater potentiality zone with the help of the current water depth data of groundwater which has been taken from CGWB as a substitute for groundwater storage data and 484 data points also used to validate the produced result (Fig. 15). Assessment of the models reveals that AUC values are 0.801, 0.841, and 0.878 for the LR, SVM, and RF, respectively (Fig. 16). Narkhede (2021) claimed that where the AUC value is 0.7, there is a 70% chance to differentiate among positive and negative classes conversely where the value is greater than 0.7 that means the reliability of this study is more than 70%. This evaluation established that RF modeling approach is most effective to delineate GWPZ in the present area followed by SVM and LR.

Conclusion

A precise assessment of the GWPZ of a particular area especially in water-scare hard rock region plays a significant role in proper water resource management, land use planning, and also in environmental protection. In recent times, RS-GIS technique with data-driven algorithm has been emerged and increasingly implemented as a very powerful and cost-effective technique. It serves as a useful tool for delineating the groundwater potential zone of Purulia district which is one of the water-scarce regions of West Bengal. The assessment was performed by applying different kinds of ML algorithms such as LR, SVM, and RF; which gives more or less similar results. These coupled algorithms help to identify GWPZ based on different causative factors; multi-collinearity analysis was also used to eliminate any collinearity issues in this study. GW level, rainfall, slope and geology are the most influencing factors for prospective GWPZ in the study area. The study indicates that some parts of this district are not suitable for groundwater recharge and results in low groundwater potentialities because of rugged terrain; very low potentiality area is minimal in comparison to other categories and moderate potentiality is more common in the entire Purulia district. It is evident that porous soil texture, low slope, permeable geological strata, high rainfall, the highest frequency of water bodies, sufficient vegetation cover, and higher lineament concentration are the favorable conditions for the development of GWPZ. The success rate of the ROC curve also depicts the efficiency of all adopted models and the resulting output is significantly reliable. So, this assessment will be helpful in the successful detection of suitable locations for extraction of further GW and
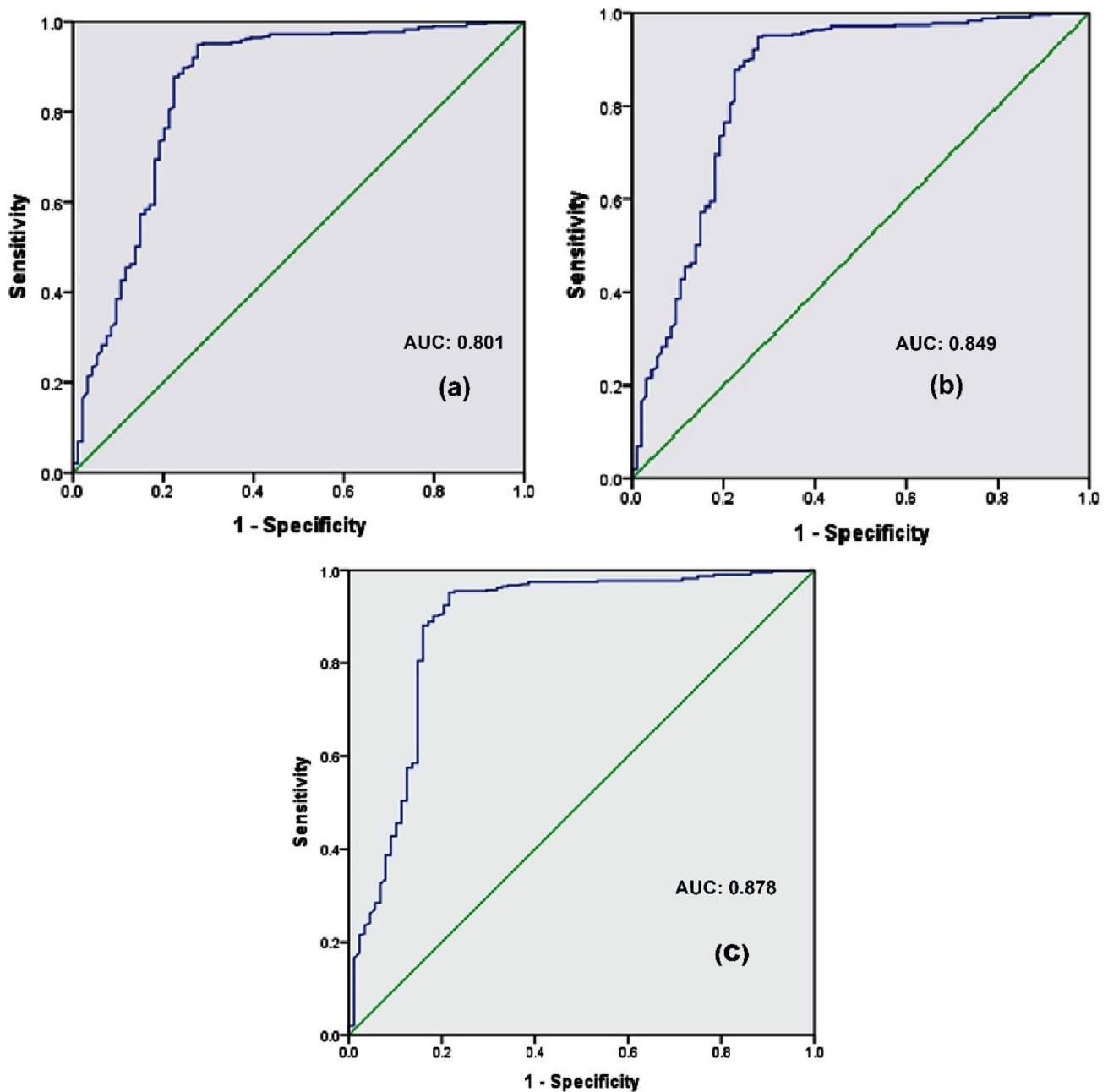
Fig. 16 Validation of the models, using ROC–AUC curve, **a** LR, **b** SVM, **c** RF

to make planning for better management of this valuable groundwater resource in the Purulia district.

Declarations

Conflict of interest The corresponding author states on behalf of all authors that there is no conflict of interest in this work.

References

Abdalla F (2012) Mapping of groundwater prospective zones using remote sensing and GIS techniques: a case study from the Central Eastern Desert, Egypt. J Afr Earth Sci 70:8–17. https://doi.org/10.1016/j.jafrearsci.2012.05.003

Andualem TG, Demeke GG (2019) Groundwater potential assessment using GIS and remote sensing: a case study of Guna tana landscape, upper blue Nile Basin, Ethiopia. J Hydrol Region Stud 24:100610. https://doi.org/10.1016/j.ejrh.2019.100610

Arabameri A, Lee S, Tiefenbacher JP, Ngo PTT (2020) Novel ensemble of MCDM-artificial intelligence techniques for groundwater-potential mapping in arid and semi-arid regions (Iran). Remote Sens 12:490. https://doi.org/10.3390/rs12030490

Avtar R, Singh CK, Shashtri S, Mukherjee S (2011) Identification of erosional and inundation hazard zones in Ken-Betwa river linking area, India, using remote sensing and GIS. Environ Monit Assess 182:341–360. https://doi.org/10.1007/s10661-011-1880-6

Ayob S, Rahmat SN (2017) Rainwater harvesting (RWH) and groundwater potential as alternatives water resources in Malaysia: a review. MATEC Web Conf 103:04020. https://doi.org/10.1051/matecconf/201710304020

Bagyaraj M, Ramkumar T, Venkatramanan S, Gurugnanam B (2013) Application of remote sensing and GIS analysis for identifying groundwater potential zone in parts of Kodaikanal Taluk, South India. Front Earth Sci 7:65–75. https://doi.org/10.1007/s11707-012-0347-6

Bhandari AK, Kumar A, Singh GK (2012) Feature extraction using normalized difference vegetation index (NDVI): a case study of Jabalpur city. Proc Technol 6:612–621. https://doi.org/10.1016/j.protcy.2012.10.074

Breiman L (2001) Random forests. Mach Learn 45:5–32. https://doi.org/10.1023/A:1010933404324

Chatterjee R, Purohit RR (2009) Estimation of replenishable groundwater resources of India and their status of utilization. Curr Sci 96:1581–1591

Chatterjee R, Tarafder G, Paul S (2010) Groundwater quality assessment of Dhanbad district, Jharkhand, India. Bull Eng Geol Environ 69:137–141. https://doi.org/10.1007/s10064-009-0234-x

Chen Y-R, Yeh C-H, Yu B (2011) Integrated application of the analytic hierarchy process and the geographic information system for flood risk assessment and flood plain management in Taiwan. Nat Hazards J Int Soc Prevent Mitig Nat Hazards 59:1261–1276

Chen W, Li H, Hou E et al (2018) GIS-based groundwater potential analysis using novel ensemble weights-of-evidence with logistic regression and functional tree models. Sci Total Environ 634:853–867. https://doi.org/10.1016/j.scitotenv.2018.04.055

Chowdhury A, Jha MK, Chowdary VM, Mal BC (2009) Integrated remote sensing and GIS-based approach for assessing groundwater potential in West Medinipur district, West Bengal, India. Int J Remote Sens 30:231–250. https://doi.org/10.1080/01431160802270131

Chowdhuri I, Pal SC, Chakrabortty R (2020) Flood susceptibility mapping by ensemble evidential belief function and binomial logistic regression model on river basin of eastern India. Adv Space Res 65:1466–1489. https://doi.org/10.1016/j.asr.2019.12.003

Daoud JI (2017) Multicollinearity and regression analysis. J Phys Conf Ser 949:012009. https://doi.org/10.1088/1742-6596/949/1/012009

Das B, Pal SC (2019) Combination of GIS and fuzzy-AHP for delineating groundwater recharge potential zones in the critical Goghat-II block of West Bengal, India. Hydroresearch 2:21–30. https://doi.org/10.1016/j.hydres.2019.10.001

Das B, Pal SC, Malik S, Chakrabortty R (2019) Modeling groundwater potential zones of Puruliya district, West Bengal, India using remote sensing and GIS techniques. Geol Ecol Landsc 3:223–237. https://doi.org/10.1080/24749508.2018.1555740

de Vries JJ, Simmers I (2002) Groundwater recharge: an overview of processes and challenges. Hydrogeol J 10:5–17. https://doi.org/10.1007/s10040-001-0171-7

Dinesh Kumar PK, Gopinath G, Seralathan P (2007) Application of remote sensing and GIS for the demarcation of groundwater potential zones of a river basin in Kerala, southwest coast of India. Int J Remote Sens 28:5583–5601. https://doi.org/10.1080/01431160601086050

Dolui G, Chatterjee S, Das Chatterjee N (2016) Geophysical and geochemical alteration of rocks in granitic profiles during intense weathering in southern Purulia district, West Bengal, India. Model Earth Syst Environ 2:132. https://doi.org/10.1007/s40808-016-0188-5

Dou J, Yunus AP, Tien Bui D et al (2019) Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan. Sci Total Environ 662:332–346. https://doi.org/10.1016/j.scitotenv.2019.01.221

Duan H, Deng Z, Deng F, Wang D (2016) Assessment of groundwater potential based on multicriteria decision making model and decision tree algorithms. Math Probl Eng 2016:e2064575. https://doi.org/10.1155/2016/2064575

Elewa HH, Qaddah AA (2011) Groundwater potentiality mapping in the Sinai Peninsula, Egypt, using remote sensing and GIS-watershed-based modeling. Hydrogeol J 19:613–628. https://doi.org/10.1007/s10040-011-0703-8

Gayen A, Pourghasemi HR, Saha S et al (2019) Gully erosion susceptibility assessment and management of hazard-prone areas in India using different machine learning algorithms. Sci Total Environ 668:124–138. https://doi.org/10.1016/j.scitotenv.2019.02.436

Ghosh PK, Bandyopadhyay S, Jana NC (2015) Mapping of groundwater potential zones in hard rock terrain using geoinformatics: a case of Kumari watershed in western part of West Bengal. Model Earth Syst Environ 2:1. https://doi.org/10.1007/s40808-015-0044-z

Goel E, Abhilasha E (2017) Computer Science & Engineering & GZSCCET Bhatinda, Punjab, India (2017) Random forest: a review. IJARCSSE 7:251–257. https://doi.org/10.23956/ijarcsse/V7I1/01113

Gupta M, Srivastava PK (2010) Integrating GIS and remote sensing for identification of groundwater potential zones in the hilly terrain of Pavagarh, Gujarat, India. Water Int 35:233–245. https://doi.org/10.1080/02508061003664419

Haridas VR, Aravindan S, Gopinath G (1998) Remote sensing and its application for ground water favorable area identification. J Geol Assoc Res Center 6:18–22

Hatti B (2011) Identification of groundwater potential zone using geoinformatics in Ghataprabha basin. North Karnataka, India

Hobbs WH (1904) Lineaments of the Atlantic Border region. GSA Bull 15:483–506. https://doi.org/10.1130/GSAB-15-483

Huajie D, Zhengdong D, Feifan D (2016) Classification of groundwater potential in Chaoyang area based on QUEST algorithm. In: 2016 IEEE international geoscience and remote sensing symposium (IGARSS), pp 890–893

Indhulekha K, Chandra Mondal K, Jhariya DC (2019) Groundwater prospect mapping using remote sensing, GIS and resistivity survey techniques in Chhokra Nala Raipur district, Chhattisgarh, India. J Water Supply Res Technol AQUA 68:595–606. https://doi.org/10.2166/aqua.2019.159

Islam ARMT, Pal SC, Chowdhuri I, Islam R, Islam MS, Rahman MM, Zahid A, Idris AM (2021) Application of novel framework approach for prediction of nitrate concentration susceptibility in coastal multi-aquifers, Bangladesh. Sci Total Environ 801:149811. https://doi.org/10.1016/j.scitotenv.2020.149811

Jenifer MA, Jha MK (2017) Comparison of analytic hierarchy process, catastrophe and entropy techniques for evaluating groundwater prospect of hard-rock aquifer systems. J Hydrol 548:605–624. https://doi.org/10.1016/j.jhydrol.2017.03.023

Jha MK, Peiffer S (2006) Applications of remote sensing and GIS technologies in groundwater hydrology: past, present and future. Bayreuth Center of Ecology and Environmental Research (BayCEER), Bayreuth

Jhariya DC, Kumar T, Gobinath M et al (2016) Assessment of groundwater potential zone using remote sensing, GIS and multi criteria

decision analysis techniques. J Geol Soc India 88:481–492. https://doi.org/10.1007/s12594-016-0511-9

Jhariya DC, Khan R, Mondal KC et al (2021) Assessment of groundwater potential zone using GIS-based multi-influencing factor (MIF), multi-criteria decision analysis (MCDA) and electrical resistivity survey techniques in Raipur city, Chhattisgarh, India. J Water Supply Res Technol AQUA 70:375–400. https://doi.org/10.2166/aqua.2021.129

Kalantar B, Pradhan B, Naghibi SA et al (2018) Assessment of the effects of training data selection on the landslide susceptibility mapping: a comparison between support vector machine (SVM), logistic regression (LR) and artificial neural networks (ANN). Geomat Nat Haz Risk 9:49–69. https://doi.org/10.1080/19475705.2017.1407368

Kamila A, Bal A, Sen TS et al (2018) An Integrated GIS approach for assessment of ground water potential zone: a case study on the upper catchment area of Subarnarekha river. Indian J Geogr Environ Manage 15–16:1–11

Khosravi K, Shahabi H, Pham BT et al (2019) A comparative assessment of flood susceptibility modeling using multi-criteria decision-making analysis and machine learning methods. J Hydrol 573:311–323. https://doi.org/10.1016/j.jhydrol.2019.03.073

Krishnamurthy J, Srinivas G, Jayaraman V, Chandrasekhar MG (1996) Influence of rock types and structures in the development of drainage networks in typical hardrock terrain. Int J Appl Earth Obs Geoinf 3/4(3):252–259

Krishnamurthy J, Mani A, Jayaraman V, Manivel M (2000) Groundwater resources development in hard rock terrain—an approach using remote sensing and GIS techniques. Int J Appl Earth Obs Geoinf 2:204–215. https://doi.org/10.1016/S0303-2434(00)85015-1

Kumar R, Singh RD, Sharma KD (2005) Water resources of India. Curr Sci 89:794–811

Machiwal D, Jha MK, Mal BC (2011) Assessment of groundwater potential in a semi-arid region of India using remote sensing, GIS and MCDM techniques. Water Resour Manage 25:1359–1386. https://doi.org/10.1007/s11269-010-9749-y

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 3:189–196. https://doi.org/10.1016/j.gsf.2011.10.007

Mahato S, Gupta F (2016) Fluvial characteristics of the Rivers of the Purulia District, West Bengal. UMR J 1(4):22–33

Mallick J, Talukder S, Islam ARMT et al (2021) Proposing receiver operating characteristic-based sensitivity analysis with introducing swarm optimized ensemble learning algorithms for groundwater potentiality modelling in Asir region, Saudi Arabia. Geocarto Int. https://doi.org/10.1080/10106049.2021.1878291

Mandrekar JN (2010) Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol 5:1315–1316. https://doi.org/10.1097/JTO.0b013e3181ec173d

McDonald AT, Mitchell G (2019) Water demand planning and management. In: Water resources, 2nd edn. Routledge, Milton Park

Micheletti N, Foresti L, Robert S et al (2014) Machine learning feature selection methods for landslide susceptibility mapping. Math Geosci 46:33–57. https://doi.org/10.1007/s11004-013-9511-0

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

Narkhede S (2021) Understanding AUC–ROC Curve. In: Medium. https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5. Accessed 9 Sep 2021

Nguyen PT, Ha DH, Avand M et al (2020) Soft computing ensemble models based on logistic regression for groundwater potential mapping. Appl Sci 10:2469. https://doi.org/10.3390/app10072469

O'Leary DW, Friedman JD, Pohn HA (1976) Lineament, linear, lineation: some proposed new standards for old terms. GSA Bull 87:1463–1469. https://doi.org/10.1130/0016-7606(1976)87%3c1463:LLLSPN%3e2.0.CO;2

Oikonomidis D, Dimogianni S, Kazakis N, Voudouris K (2015) A GIS/remote sensing-based methodology for groundwater potentiality assessment in Tirnavos area, Greece. J Hydrol 525:197–208. https://doi.org/10.1016/j.jhydrol.2015.03.056

Ok AO, Akar O, Gungor O (2012) Evaluation of random forest method for agricultural crop classification. Eur J Remote Sens 45:421–432. https://doi.org/10.5721/EuJRS20124535

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 405:123–136. https://doi.org/10.1016/j.jhydrol.2011.05.015

Pal SC, Ghosh C, Chowdhuri I (2020) Assessment of groundwater potentiality using geospatial techniques in Purba Bardhaman district, West Bengal. Appl Water Sci 10:221. https://doi.org/10.1007/s13201-020-01302-3

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 9:1157. https://doi.org/10.3390/su9071157

Pothiraj P, Baskaran R (2013) Mapping of lineaments for groundwater targeting and sustainable water resource management in hard rock hydrogeological environment using RS-GIS. pp 235–247

Punmia BC, Jain AK (2005) Soil mechanics and foundations. Firewall Media

Rahmati O, Nazari Samani A, Mahdavi M et al (2015) Groundwater potential mapping at Kurdistan region of Iran using analytic hierarchy process and GIS. Arab J Geosci 8:7059–7071. https://doi.org/10.1007/s12517-014-1668-4

Rajaveni SP, Brindha K, Elango L (2017) Geological and geomorphological controls on groundwater occurrence in a hard rock region. Appl Water Sci 7:1377–1389. https://doi.org/10.1007/s13201-015-0327-6

Ramaiah S, Gopalakrishna G, Vittala S, Najeeb K (2012) Geomorphological mapping for identification of ground water potential zones in hard rock areas using geo-spatial information- a case study in Malur Taluk, Kolar District, Karnataka, India

Richey AS, Thomas BF, Lo M-H et al (2015) Uncertainty in global groundwater storage estimates in a total groundwater stress framework. Water Resour Res 51:5198–5216. https://doi.org/10.1002/2015WR017351

Rodell M, Velicogna I, Famiglietti JS (2009) Satellite-based estimates of groundwater depletion in India. Nature 460:999–1002. https://doi.org/10.1038/nature08238

Rodriguez-Galiano V, Mendes MP, Garcia-Soldado MJ et al (2014) Predictive modeling of groundwater nitrate pollution using random forest and multisource variables related to intrinsic and specific vulnerability: a case study in an agricultural setting (Southern Spain). Sci Total Environ 476–477:189–206. https://doi.org/10.1016/j.scitotenv.2014.01.001

Saha A, Pal SC, Arabameri A et al (2021) Flood susceptibility assessment using novel ensemble of hyperpipes and support vector regression algorithms. Water 13:241. https://doi.org/10.3390/w13020241

Sahu PC, Sahoo H (2006) Targeting ground water in tribal dominated Bonai area of drought-prone Sundargarh District, Orissa, India—a combined geophysical and remote sensing approach. J Hum Ecol 20:109–115. https://doi.org/10.1080/09709274.2006.11905912

Sameen MI, Pradhan B, Lee S (2020) Application of convolutional neural networks featuring Bayesian optimization for landslide susceptibility assessment. CATENA 186:104249. https://doi.org/10.1016/j.catena.2019.104249

Sander P, Chesley MM, Minor TB (1996) Groundwater assessment using remote sensing and GIS in a rural groundwater project in Ghana: lessons learned. HYJO 4:40–49. https://doi.org/10.1007/s100400050086

Saraf AK, Choudhury PR (1998) Integrated remote sensing and GIS for groundwater exploration and identification of artificial recharge sites. Int J Remote Sens 19:1825–1841. https://doi.org/10.1080/014311698215018

Saraf AK, Choudhury PR, Roy B et al (2004) GIS based surface hydrological modelling in identification of groundwater recharge zones. Int J Remote Sens 25:5759–5770. https://doi.org/10.1080/0143116042000274096

Seckler D, Barker R, Amarasinghe U (1999) Water scarcity in the twenty-first century. Int J Water Resour Dev 15:29–42. https://doi.org/10.1080/07900629948916

Selvam S, Magesh NS, Chidambaram S et al (2015) A GIS based identification of groundwater recharge potential zones using RS and IF technique: a case study in Ottapidaram taluk, Tuticorin district, Tamil Nadu. Environ Earth Sci 73:3785–3799. https://doi.org/10.1007/s12665-014-3664-0

Singh P, Gupta A, Singh M (2014) Hydrological inferences from watershed analysis for water resource management using remote sensing and GIS techniques. Egypt J Remote Sens Space Sci 17:111–121. https://doi.org/10.1016/j.ejrs.2014.09.003

Singh SK, Zeddies M, Shankar U, Griffiths GA (2019) Potential groundwater recharge zones within New Zealand. Geosci Front 10:1065–1072. https://doi.org/10.1016/j.gsf.2018.05.018

Sinha BK, Kumar A, Shrivastava D, Srivastava S (1990) Integrated approach for demarcating the fracture zone for well site location: a case study near Gumla and Lohardaga, Bihar. J Ind Soc Remote Sens 18:1–8. https://doi.org/10.1007/BF03030727

Smola AJ, Schölkopf B (2004) A tutorial on support vector regression. Stat Comput 14:199–222. https://doi.org/10.1023/B:STCO.0000035301.49549.88

Solomon S, Quiel F (2006) Groundwater study using remote sensing and geographic information systems (GIS) in the central highlands of Eritrea. Hydrogeol J 14:1029–1041. https://doi.org/10.1007/s10040-006-0096-2

Tehrany MS, Pradhan B, Mansor S, Ahmad N (2015) Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. CATENA 125:91–101. https://doi.org/10.1016/j.catena.2014.10.017

Terence E, Viessman W (1977) Introduction to hydrology, 2nd edn. Harper & Row, New York

Thapa R, Gupta S, Gupta A et al (2018) Use of geospatial technology for delineating groundwater potential zones with an emphasis on water-table analysis in Dwarka River basin, Birbhum, India. Hydrogeol J 26:899–922. https://doi.org/10.1007/s10040-017-1683-0

Thomas R, Duraisamy V (2018) Hydrogeological delineation of groundwater vulnerability to droughts in semi-arid areas of western Ahmednagar district. Egypt J Remote Sens Space Sci 21:121–137. https://doi.org/10.1016/j.ejrs.2016.11.008

Thorpe HR, Scott DM (1999) An evaluation of four soil moisture models for estimating natural ground water recharge. J Hydrol (n Z) 38:179–209

Tien Bui D, Pradhan B, Lofman O, Revhaug I (2012) Landslide susceptibility assessment in Vietnam using support vector machines, decision tree, and naïve Bayes models. Math Probl Eng 2012:e974638. https://doi.org/10.1155/2012/974638

Vapnik V (1998) The support vector method of function estimation. In: Suykens JAK, Vandewalle J (eds) Nonlinear modeling: advanced black-box techniques. Springer, US, Boston, MA, pp 55–85

Waikar ML, Nilawar AP, Head P (2014) Identification of groundwater potential zone using remote sensing and GIS technique. Int J Innov Res Sci Eng Technol 3:12163–12174

Waters P, Greenbaum D, Smart PL, Osmaston H (1990) Applications of remote sensing to groundwater hydrology. Remote Sens Rev 4:223–264. https://doi.org/10.1080/02757259009532107

White PA, Hong Y-S, Murray DL et al (2003) Evaluation of regional models of rainfall recharge to groundwater by comparison with lysimeter measurements, Canterbury, New Zealand. J Hydrol (n Z) 42:39–64

Yariyan P, Zabihi H, Wolf ID et al (2020) Earthquake risk assessment using an integrated fuzzy analytic hierarchy process with artificial neural networks based on GIS: a case study of Sanandaj in Iran. Int J Disast Risk Reduc 50:101705. https://doi.org/10.1016/j.ijdrr.2020.101705

Yeh H-F, Cheng Y-S, Lin H-I, Lee C-H (2016) Mapping groundwater recharge potential zone using a GIS approach in Hualian River, Taiwan. Sustain Environ Res 26:33–43. https://doi.org/10.1016/j.serj.2015.09.005

Zabihi M, Pourghasemi HR, Pourtaghi ZS, Behzadfar M (2016) GIS-based multivariate adaptive regression spline and random forest models for groundwater potential mapping in Iran. Environ Earth Sci 75:665. https://doi.org/10.1007/s12665-016-5424-9