

# Groundwater potentiality mapping using ensemble machine learning algorithms for sustainable groundwater management

Groundwater  
potentiality  
mapping

43

Showmitra Kumar Sarkar

*Department of Urban and Regional Planning,  
Khulna University of Engineering and Technology, Khulna, Bangladesh*

Swapan Talukdar

*Geography, University of Gour Banga, Malda, India and  
Department of Geography, Faculty of Natural Sciences, Jamia Millia Islamia,  
New Delhi, India*

Atiqur Rahman and Shahfahad

*Department of Geography, Faculty of Natural Sciences, Jamia Millia Islamia,  
New Delhi, India, and*

Sujit Kumar Roy

*Institute of Water and Flood Management,  
Bangladesh University of Engineering and Technology, Dhaka, Bangladesh*

Received 10 September 2021  
Revised 20 September 2021  
Accepted 22 September 2021

## Abstract

**Purpose** – The present study aims to construct ensemble machine learning (EML) algorithms for groundwater potentiality mapping (GPM) in the Teesta River basin of Bangladesh, including random forest (RF) and random subspace (RSS).

**Design/methodology/approach** – The RF and RSS models have been implemented for integrating 14 selected groundwater condition parameters with groundwater inventories for generating GPMs. The GPM were then validated using the empirical and bionormal receiver operating characteristics (ROC) curve.

**Findings** – The very high (831–1200 km<sup>2</sup>) and high groundwater potential areas (521–680 km<sup>2</sup>) were predicted using EML algorithms. The RSS (AUC-0.892) model outperformed RF model based on ROC's area under curve (AUC).

**Originality/value** – Two new EML models have been constructed for GPM. These findings will aid in proposing sustainable water resource management plans.

**Keywords** Groundwater potentiality, Data mining, GIS, Remote sensing, Random subspace

**Paper type** Research paper

## 1. Introduction

Groundwater is the world's largest source of freshwater (i.e. one-third of worldwide freshwater consumption) but there is a shortage of data at a micro-spatial level on the

---

© Showmitra Kumar Sarkar, Swapan Talukdar, Atiqur Rahman, Shahfahad and Sujit Kumar Roy. Published in *Frontiers in Engineering and Built Environment*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licences/by/4.0/legalcode>

*Data availability*: The data that support the findings of this study are available from the corresponding author, [swapantalukdar65@gmail.com], upon reasonable request.



potential groundwater source (Ferozur *et al.*, 2019; Adham *et al.*, 2010). Groundwater use is continuously increasing. These increasing demands usually result in overexploitation, putting a pressure on the limited supply of freshwater (Jahan *et al.*, 2019). Furthermore, groundwater problems have worsened, particularly in the tropical and subtropical zones, as a result of unregulated irrigation practices, high population density and climate change.

Groundwater potentiality has been investigated using physical, heuristic and mathematical techniques (Namous *et al.*, 2021). Physical techniques evaluate groundwater potential by examining topography and geology (Mallick *et al.*, 2021a). Heuristic-based techniques are very professional and produce reasonable accuracy (Mallick *et al.*, 2021a). Evidence-based models such as statistical index (SI) (Pande *et al.*, 2020), logistic regression (LR) (Chen *et al.*, 2020; Ozdemir, 2011; Park *et al.*, 2017), evidential belief function (EBF) (Mogaji *et al.*, 2016; Nampak *et al.*, 2014), probability-frequency ratio (FR) (Arshad *et al.*, 2020; Razandi *et al.*, 2015), certainty factors (CF) (Razandi *et al.*, 2015; Ahmadi *et al.*, 2020; Zhao and Chen, 2020), weight of evidence (WoE) (Das *et al.*, 2021; Hembram *et al.*, 2019), index of entropy (IoE) (Al-Abadi and Shahid, 2015; Rahmati *et al.*, 2016) and certainty fact have been used to model groundwater potentiality. Focusing on current groundwater availability regions and related variables makes these methods objective and measurable. However, standard statistical techniques cannot anticipate the dynamic and non-linear interactions between groundwater and the conditioning factors (Mallick *et al.*, 2021b). Since no one technique or methodology works for all situations, machine learning is considered.

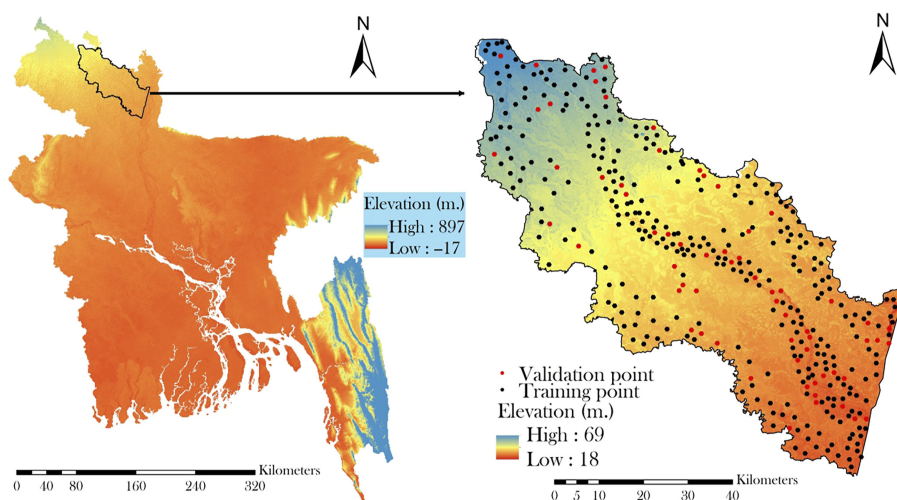
Machine learning has been utilised to predict groundwater potentiality because it can analyse the dynamic relationship between groundwater potentiality and influencing factors (Mallick *et al.*, 2021b). Several methods have been used to assess groundwater potentiality, including artificial neural networks (Lee *et al.*, 2018; Pal *et al.*, 2020), neuro-fuzzy (Termeh *et al.*, 2019; Khosravi *et al.*, 2018), decision trees (Duan *et al.*, 2016; Choubin *et al.*, 2020) and support vector machines (Lee *et al.*, 2018; Naghibi *et al.*, 2018). But until recently, groundwater experts couldn't agree on a model for evaluating groundwater potentiality (Mallick *et al.*, 2021a). Thus, ensemble techniques have lately acquired favour in geohazard susceptibility and potentiality mapping (Mallick *et al.*, 2021b).

Ensemble modelling combines two or more machine learning methods to improve forecast accuracy (Talukdar *et al.*, 2020; Islam *et al.*, 2021a, b). Ensemble modelling may improve an individual model's weaknesses (Talukdar *et al.*, 2021a, b). Researching susceptibility, sensitivity, hazards, potentiality and other issues using multi-model ensembles is a newer trend (Talukdar and Pal, 2020; Mahato *et al.*, 2021). For example, the present research utilised RF and RSS to enhance the model's robustness. The ensemble prediction technique has not been utilised for groundwater potential zone mapping in Bangladesh's northern Teesta sub-catchment.

## 2. Methods and materials

### 2.1 Study area

The Teesta sub-catchment, which covers 2284 km<sup>2</sup> and includes five major districts in Bangladesh's northern region, namely Lalmanirhat, Kurigram, Rangpur, Nilphamary and Gaibandha, is the study's research area (Figure 1). This basin is located in Bangladesh between the latitudes of 25°30'02"N and 26°18'37"N and the longitudes of 88°52'58"E and 89°45'34"E. Bangladesh's largest geomorphic unit is the floodplain, and the drainage basin is made up of several minor rivers that run at elevations ranging from 5 to 110 metres. When floods occur, the river's general slope ranges from 0.47 to 0.55 m/km, suggesting a comparatively flat terrain (Rahman *et al.*, 2011). Since the river basin's morphological depression is narrow and situated in a dormant stream canyon, the river basin's pathways are long and morphologically diverse. With a dense river network and six rivers, including the



**Figure 1.**  
The location of the  
study area with the  
training and validation  
flood points

Naotora, Buri Teesta, Ghagot, Old Brahmaputra, Jamuna and Dharla, the research area has complex hydrological attributes.

The climate in this basin is sub-tropical monsoonal, with two distinct seasons: monsoon (June to September) and dry season (October to December) (October to May). The average annual precipitation in this basin is over 1900 mm (Akter *et al.*, 2019), with over 80% of overall annual precipitation taking place during the monsoon season.

## 2.2 Materials

The groundwater potentiality (GWP) models for this study were prepared using 12 groundwater conditioning parameters. These are land use land cover (LULC), rainfall, distance to road, elevation, slope, topographic roughness index (TRI), stream power index (SPI), sediment transport index (STI), curvature, soil types, topographic wetness index (TWI), aspect. For LULC map, Landsat 8 Operational Land Imager (OLI) image from the United States Geological Survey's (USGS) website (Path/row: 138/42, spatial resolution: 30 m, date:19/03/2019) has been utilised. The advanced spaceborne thermal emission and reflection radiometer (ASTER) global digital elevation model (GDEM) (Version 2, spatial resolution: 30 metre) was utilised to extract topographical and hydrological variables. The rainfall data were given by the Bangladesh Meteorological Department (BMD), Dhaka, Bangladesh. We used a soil taxonomy map from the United States Department of Agriculture's Natural Resources Conservation Service (USDA) (NRCS).

## 2.3 Groundwater potentiality inventory

For GWP mapping, several researchers have utilised the positions of springs, wells and quant for inventory. Well points were taken into account for GWP in this study. The study region's inventory graph includes 220 well points collected from various resources and detailed site inspection. First, non-groundwater data similar to the groundwater data utilised for GWP modelling must be prepared. The selection was made on the basis of the field survey, with equivalent numbers of non-groundwater data (220 points). By arbitrary separation, all groundwater and non-groundwater data have been divided into 80 (352 points):20 (88 points) proportion as calibrating and test datasets (Figure 1). Model calibration is done with

groundwater and non-groundwater training data, while model validation is done with groundwater and non-groundwater testing data (Mallick *et al.*, 2021a). Similarly, inventory maps for other areas have been developed.

#### *2.4 Methods for preparing groundwater potentiality conditioning factors*

Since it requires multiple variables related to topography and hydrology in geospatial layout, the architecture of the spatial groundwater potentiality model is typically very complex and systematic. As a result, identifying variables that affect groundwater potentiality is critical, and scientifically selected criteria can confirm the accuracy of groundwater potentiality modelling charts. All the selected parameters were translated into 30 m spatial resolution using resampling technique.

Topographic influences are critical for GWP modelling because they affect the hydrological characteristics of the research region both directly and indirectly (Panahi *et al.*, 2020). At first, ASTER GDEM data was used to generate digital elevation model to extract slope, curvature, aspect, TWI, SPI, STI and TRI using ArcGIS 10.2 software (Figure 2).

Soil characteristics are one of the most important determining variables in the rainfall-runoff process (Nguyen *et al.*, 2020). While Flügel (1995) reported that other factors such as local weather patterns and erosion processes influence rainfall-runoff generation, soil properties directly govern water penetration, which influences rainfall-runoff generation. Groundwater events are more likely to occur if the degree of penetration is high. The study area has 12 groups of soil as per USDA soil taxonomy (Figure 2j).

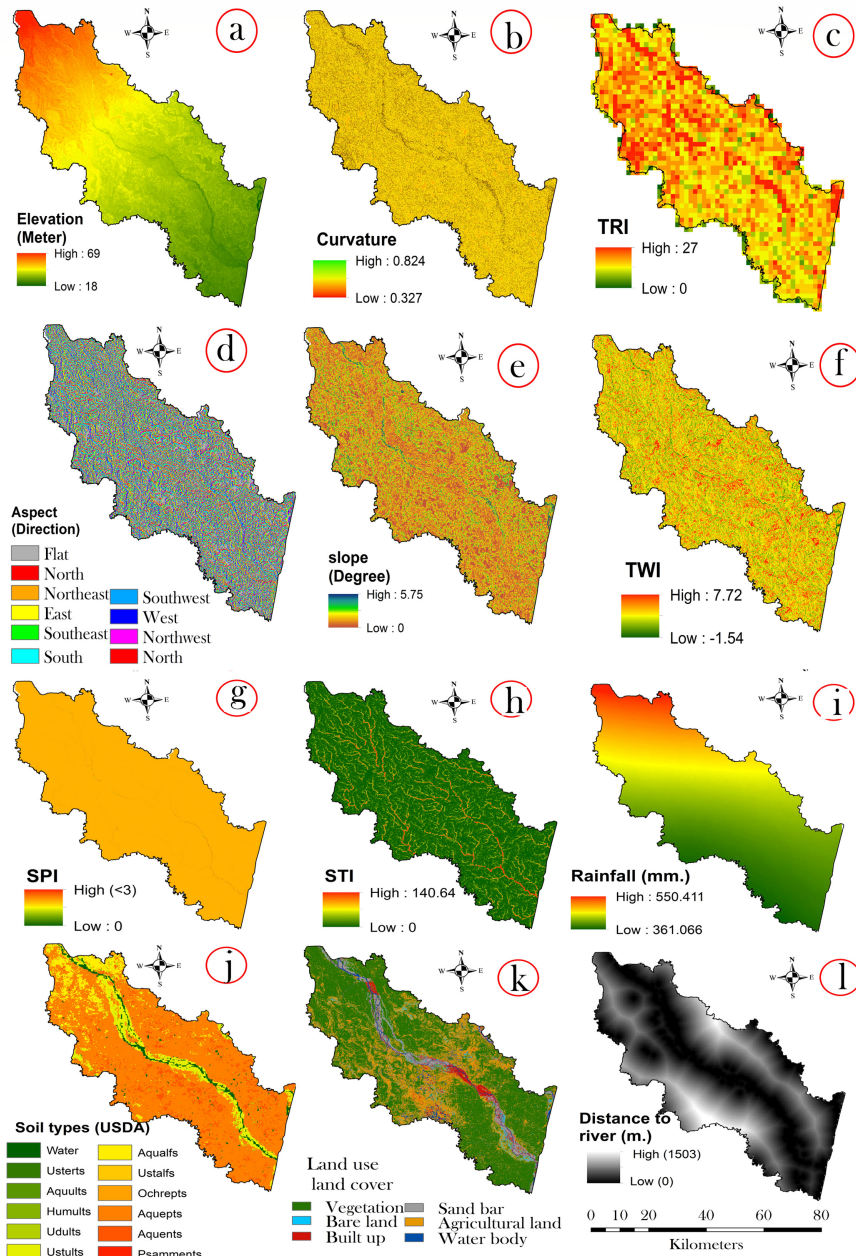
LULC has an effect on surface runoff including a significant impact on the occurrence of groundwater potentiality (Prasad *et al.*, 2020), since the LULC has full control over the generation and penetration of surface runoff. Groundwater potentiality is very less in built-up areas because these zones prevent water from penetrating and producing surface water. In comparison, the woodland region encourages water to infiltrate, resulting in less groundwater potentiality (Mallick *et al.*, 2021a). When comparing hydrological reactions at different temporal scales, the relationship between groundwater potentiality occurrences and plant density is inverse (Tolche, 2021). The artificial neural network (ANN) model was used in Environment for Visualizing Images (ENVI) software (version 5.3) to create a LULC map. Bare ground, forest, sand bar, built-up field, agricultural land and water body were divided into six groups on the LULC map (Figure 2).

#### *2.5 Method for groundwater potentiality modelling*

*2.5.1 Random forest.* RF is a classification and regression approach that uses an ensemble of binary decision trees that have been trained individually (Golkarian *et al.*, 2018). The basic strategy employed by RF for classification issues is to train each decision tree individually with the ultimate conclusion calculated by taking into consideration the findings acquired by each decision tree (Sameen *et al.*, 2019).

Without needing to go through a pruning procedure, RF models can generalise and reduce the danger of overfitting. The training entails producing a number of distinct bootstrap samples from the original dataset, with one-third left out to function as test cases and estimating an unbiased test error, referred to as the out-of-bag-error, which reflects the RF model's prediction performance based on these test cases (Breiman, 2001).

*2.5.2 Random subspace.* RSS was proposed in 1988 as a way to improve the accuracy of weak classifications and the performance of individual classifications. RSS (Ho, 1998; Skurichina and Duin, 2002) is a popular method of random sampling in which the original character varies at random. RSS groups the characteristic series of each sub-classification creation using a majority vote after creating numerous subspaces with tiny dimensions (Skurichina and Duin, 2002; Kuncheva and Plumpton, 2010). RSS has been utilised in a



**Figure 2.** Data layers for groundwater potentiality conditioning factors such as (1) elevation, (2) curvature, (3) TRI and (4) aspect, (5) slope, (6) TWI, (7) SPI, (8) STI, (9) rainfall (10) soil types, (11) LULC and (12) distance to river

variety of disciplines, including economics (Wang and Ma, 2011) and medical (Bertoni *et al.*, 2005) but very seldom in groundwater potential assessment. The optimization of model's parameters has been presented in Table 1.



2.6 Validation of the models

In the ROC, on the horizontal axis (true positive or 1-specificity), the proportion of pixels properly predicted by the presence or absence of groundwater potential is shown, while the proportion of pixels erroneously predicted is represented on the vertical axis (false positive or sensitivity) (Mallick *et al.*, 2021a). The AUC is the area beneath this curve, and the model with the greatest AUC has the best relative performance (Talukdar *et al.*, 2021a, b). Random prediction for a model is shown by AUC values of 0.5 (Talukdar *et al.*, 2021a). The AUC values vary from 0 to 1, with 0 being the lowest and 1 being the highest. AUC values those are greater than 0.7 reflect a model’s prediction effectiveness (Nguyen *et al.*, 2020).

3. Results

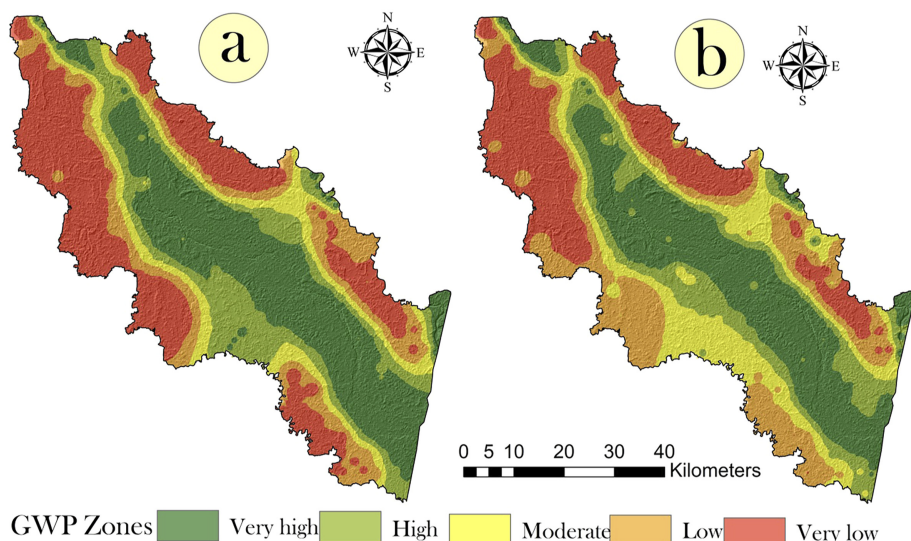
3.1 Description of the parametres

Several conditioning variables can impact a region’s groundwater potentiality (Mukherjee *et al.*, 2021). In this study, the affecting parametres were LULC, distance to river, height, slope, topographic wetness index, stream power index, sediment transport index, curvature, topographic roughness index, curvature and aspect. Low-lying regions, particularly depressed lands in the flood plain region, maintain a high degree of surface moisture and replenish the groundwater aquifer as a result of persistent ponding. The altitudes of the research area varied from 18 to 69 metres (Figure 2). The capacity for recharging water is greatest when the curvature is a concave surface, followed by plain surfaces (Nguyen *et al.*, 2020). Curvature map, which was produced by using the digital elevation model (DEM) ranged from 0.32–0.82 (Figure 2a). The DEM was used to build a curvature map that ranged from 0.32–0.82 (Figure 2b). Also, a flat or moderate slope will help to slow down the flow of water and increase the groundwater recharge (Kumar *et al.*, 2019). In this study, the slopes utilised varied from 0 to 5.75 (Figure 2d). TRI examined the impact of the underlying surface’s conflict on the water flow (Straatsma and Baptist, 2008). The Teesta river was located at the lowest TRI due to the steep hills around the river, generating fast water flow. Lower TRI values imply a larger possibility for groundwater (Chen *et al.*, 2020).

In this analysis, the highest TRI value was 27 (Figure 2). A high TWI also ensures adequate groundwater recharge. The high TWI values are strongly correlated with groundwater potentiality. Figure 3 shows TWI values ranging from –1.54 to –7.72. Furthermore, since higher SPI and STI values mean a higher water level, regions with higher SPI and STI values have a greater chance of experiencing groundwater (Bui *et al.*, 2019). The highest STI value in this study was 140.64 (See Figure 2). LULC is important in modelling groundwater potential zone. Maps showing vegetated land turning to become barren land result in increased runoff, lower infiltration and thereby directly impacting the groundwater (Pal *et al.*, 2020). LULC was divided into six groups in this study, including vegetation, bare land, built up, sand bar, agricultural land and water body (Figure 2). In this place, the greatest distance from the river was 1503 metres, as seen in Figure 2. Soil data played an important role in accounting for excess precipitation and infiltration (Johnson *et al.*, 2000). Water, usterts, aquults, humults, udults, ustults, aqualfs, ustalfs, ochrepts, aquepts, aquents and psamments were amongst the 12 soil types discovered in this study (Figure 2). The sum of

**Table 1.**  
The parametres of the machine learning algorithm used for groundwater potentiality modelling

Model name	Description of parametres
RF	Batch size-100, seed-6, number of iteration-200, max depth-3, calc out of bag-TRUE and compute attribute importance-TRUE
RSS	Classifier-RF, max depth- 3, minimum number-2, minimum proportion of the variance-0.001, executions slots-2, seed-5, iteration- 100 and subspace size-0.5



**Figure 3.**  
Groundwater  
potentiality models  
using (1) RSS and  
(2) RF

rainfall has a major impact in determining the potentiality of groundwater in an area, as the distribution of rainfall strongly controls the recharge volume of a basin (Figure 2).

### 3.2 Groundwater potentiality modelling and validation

Figure 3, represents the groundwater potentiality models as constructed using advance machine learning algorithms, such as, RF and RSS. As shown in Figure 3, the potential zones of groundwater were divided into five categories: very high, high, moderate, low and very low. The potential groundwater zone runs in a northwest–southeast direction, parallel to the drainage direction of the catchment. The south and southeast are dominated by zones with high groundwater potential, whereas the north northwest is dominated by areas with low groundwater potential zones.

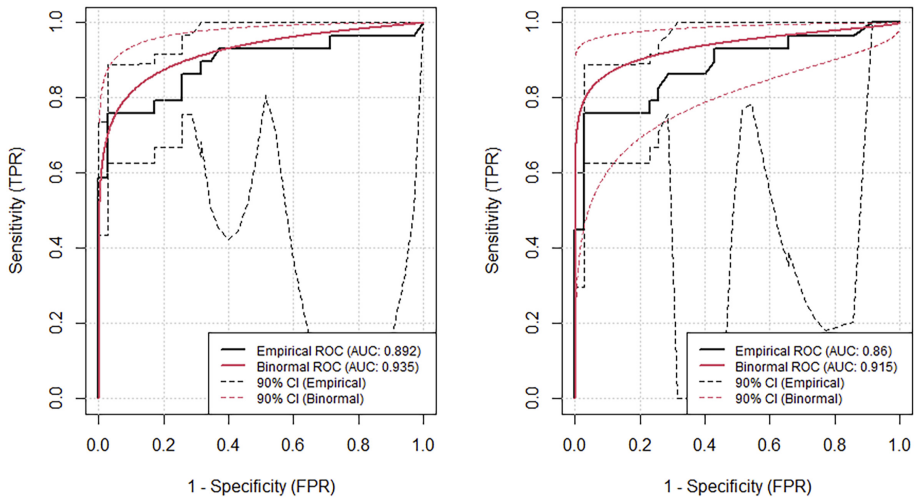
About 2.26 and 36.69% areas to the total basin area are found to have “very high” and “high” potentiality for groundwater, respectively, in case of RSS model (Table 2). While the RF models identified around 30% of the overall basin area as high potential for groundwater. In general, all of the models defined the river catchment area as having a lot of potential for underground water harvesting. However, since there are variations in the size of the region, it is critical to explain the best representative model.

Two different models were used to integrate and define groundwater potential zones in this study. The region ROC curve is used to show how accurate the model is (AUC). The AUC and considerable level of the ROC curve were used to test the evaluation of these models. The AUC calculated using ROC specifically specified the acceptability of all models, as it was greater than 0.8 in all cases (Figure 4). The AUC indicates how accurate the model’s output

GWP zones	Very high	High	Area (km <sup>2</sup> ) Moderate	Low	Very low
RF	1102.21	584.87	361.61	507.79	1103.01
RSS	1023.99	546.11	592.62	722.03	800.01

**Table 2.**  
Computation of area  
coverage under  
different GWP zones

**Figure 4.**  
Validation of  
groundwater  
potentiality models  
using ROC curve  
(1) RSS and (2) RF



can be forecasted. The greater the AUC, the more accurate the model's output can be predicted. The findings of these four human models were statistically important in this study (significant level, 0.5). The RSS (0.89), RepTree (0.898) and M5P (0.89) models had the best results in the test.

#### 4. Discussions

Since machine learning approaches demonstrate potential when working with a variety of geographic data, machine learning modelling of environmental problems has grown in popularity (Panahi *et al.*, 2020; Prasad *et al.*, 2020). As a result, machine learning modelling can successfully address the problem of identifying groundwater potential zones over large-scale regions, which frequently lack reliable and long-term geotechnical and hydrogeological data for the application of physically-based and/or numerical models (Pal *et al.*, 2020; Mallick *et al.*, 2021a; Sameen *et al.*, 2019). Nevertheless, the versatility of various machine learning methods must be thoroughly explored through their implementation in various regions with various geo-environmental settings in order to identify the best model with the highest precision and the least sensitivity to noisy input data (Choubin *et al.*, 2019; Naghibi *et al.*, 2018).

Robust techniques for obtaining very accurate results may be used to propose long-term groundwater management. The goal of this study is to create an EML approach for groundwater potential mapping in Bangladesh's Teesta river basin. RF and RSS models were used to integrate 14 groundwater condition factors with groundwater inventory for GPM production. Based on the ROC AUC, the RSS model (AUC-0.892) outperformed the RF model (AUC). According to the RSS model, about 1024 and 546 km<sup>2</sup> of the overall basin area have "very high" and "high" groundwater potentiality, respectively.

Although the research has mostly focussed on the usefulness of ensemble approaches, these techniques have demonstrated varying levels of success for various issues in various fields. For example, Nguyen *et al.* (2020) found that RSS outperformed bagging and dagging approaches for predicting groundwater potentiality, whereas Mallick *et al.* (2021a) found that the RSS model outperformed rotation forest and bagging for predicting groundwater potential. Using ensemble models to forecast floods, different outcomes have been reported (Mahato *et al.*, 2021; Saha *et al.*, 2021).



## 5. Conclusion

The current research delves into the evolution of EML algorithms for estimating groundwater potentiality. According to the two ensemble models, the very high groundwater potential zone spans an area of 830–21200 km<sup>2</sup>. The ROC curve was used to assess the groundwater potential models. The best representation model for groundwater potentiality modelling was RSS (AUC = 0.892), followed by RF (AUC: 0.86). Distance to river, slope, curvature, elevation, LULC and SPI can be considered as the most dominant and sensitive parameters for groundwater potentiality modelling. Groundwater depletion threatens the survival of natural surface water bodies, agriculture, natural resources and livelihood.

In the case of groundwater potentiality models, RSS model outperformed RF model. This research further proposes that a few other hydrogeological and meteorological variables can be added to the models to increase the accuracy of the outcome. Owing to damming across the river and other anthropogenic problems, the Teesta river basin is notorious for its water shortage. Such findings could aid in the development of long-term water harvesting and cropping strategies. Rapid reclamation of water sources should be stopped at all times, as water bodies have been identified as a good conditioning factor for groundwater potentiality. Land cover and canopy density are also high conditioning influences, according to this report. However, forest loss and destruction are undeniable facts. As a result, forest cover preservation will aid groundwater recharge. Study is needed for scientific assessment of groundwater at various potential zones in order to get a more precise recommendation on how much water can be harvested from each potential zone.

## References

- Adham, M.I., Jahan, C.S., Mazumder, Q.H., Hossain, M.M.A. and Haque, A.M. (2010), "Study on groundwater recharge potentiality of Barind Tract, Rajshahi District, Bangladesh using GIS and remote sensing technique", *Journal of the Geological Society of India*, Vol. 75 No. 2, pp. 432-438.
- Ahmadi, H., Kaya, O.A., Babadagi, E., Savas, T. and Pekkan, E. (2020), "GIS-based groundwater potentiality mapping using AHP and FR models in central antalya, Turkey", *Environmental Sciences Proceedings*, Multidisciplinary Digital Publishing Institute, Vol. 5 No. 1, p. 11.
- Akter, A., Uddin, A.M.H., Wahid, K.B. and Ahmed, S. (2019), "Predicting groundwater recharge potential zones using geospatial technique", *Sustainable Water Resources Management*, Vol. 6 No. 2, pp. 1-13.
- Al-Abadi, A.M. and Shahid, S. (2015), "A comparison between index of entropy and catastrophe theory methods for mapping groundwater potential in an arid region", *Environmental Monitoring and Assessment*, Vol. 187 No. 9, pp. 1-21.
- Arshad, A., Zhang, Z., Zhang, W. and Dilawar, A. (2020), "Mapping favorable groundwater potential recharge zones using a GIS-based analytical hierarchical process and probability frequency ratio model: a case study from an agro-urban region of Pakistan", *Geoscience Frontiers*, Vol. 11 No. 5, pp. 1805-1819.
- Bertoni, A., Folgieri, R. and Valentini, G. (2005), "Bio-molecular cancer prediction with random subspace ensembles of support vector machines", *Neurocomputing*, Vol. 63, pp. 535-539.
- Bui, D.T., Shirzadi, A., Chapi, K., Shahabi, H., Pradhan, B., Pham, B.T., Singh, V.P., Chen, W., Khosravi, K., Bin Ahmad, B. and Lee, S. (2019), "A hybrid computational intelligence approach to groundwater spring potential mapping", *Water*, Vol. 11 No. 10, p. 2013.
- Breiman, L. (2001), "Random forests", *Machine Learning*, Vol. 45 No. 1, pp. 5-32.
- Chen, W., Li, Y., Tsangaratos, P., Shahabi, H., Ilija, I., Xue, W. and Bian, H. (2020), "Groundwater spring potential mapping using artificial intelligence approach based on kernel logistic regression, random forest, and alternating decision tree models", *Applied Sciences*, Vol. 10 No. 2, p. 425.

- Choubin, B., Rahmati, O., Soleimani, F., Alilou, H., Moradi, E. and Alamdari, N. (2019), "Regional groundwater potential analysis using classification and regression trees", *Spatial Modeling in GIS and R for Earth and Environmental Sciences*, Elsevier, pp. 485-498.
- Choubin, B., Hosseini, F.S., Fried, Z. and Mosavi, A. (2020), "Application of Bayesian regularized neural networks for groundwater level modeling", *2020 IEEE 3rd International Conference and Workshop in buda on Electrical and Power Engineering (CANDO-EPE)*, IEEE, pp. 000209-000212.
- Das, N., Sutradhar, S., Ghosh, R. and Mondal, P. (2021), *Applicability of Geospatial Technology, Weight of Evidence, and Multilayer Perceptron Methods for Groundwater Management: A Geoscientific Study on Birbhum District, West Bengal, India*, Groundwater and Society: Applications of Geospatial Technology, pp. 473-499.
- Duan, H., Deng, Z., Deng, F. and Wang, D. (2016), "Assessment of groundwater potential based on multicriteria decision making model and decision tree algorithms", *Mathematical Problems in Engineering*, Vol. 2016.
- Ferozur, R.M., Jahan, C.S., Arefin, R. and Mazumder, Q.H. (2019), "Groundwater potentiality study in drought prone barind tract, NW Bangladesh using remote sensing and GIS", *Groundwater for Sustainable Development*, Vol. 8, pp. 205-215.
- Flügel, W.A. (1995), "Delineating hydrological response units by geographical information system analyses for regional hydrological modelling using PRMS/MMS in the drainage basin of the River Bröl, Germany", *Hydrological Processes*, Vol. 9 Nos 3-4, pp. 423-436.
- Golkarian, A., Naghibi, S.A., Kalantar, B. and Pradhan, B. (2018), "Groundwater potential mapping using C5. 0, random forest, and multivariate adaptive regression spline models in GIS", *Environmental Monitoring and Assessment*, Vol. 190 No. 3, pp. 1-16.
- Hembram, T.K., Paul, G.C. and Saha, S. (2019), "Comparative analysis between morphometry and geo-environmental factor based soil erosion risk assessment using weight of evidence model: a study on Jainti river basin, eastern India", *Environmental Processes*, Vol. 6 No. 4, pp. 883-913.
- Ho, T.K. (1998), "The random subspace method for constructing decision forests", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20 No. 8, pp. 832-844.
- Islam, A.R.M.T., Talukdar, S., Mahato, S., Kundu, S., Eibek, K.U., Pham, Q.B., Kuriqi, A. and Linh, N.T.T. (2021a), "Flood susceptibility modelling using advanced ensemble machine learning models", *Geoscience Frontiers*, Vol. 12 No. 3, p. 101075.
- Islam, A.R.M.T., Talukdar, S., Mahato, S., Ziaul, S., Eibek, K.U., Akhter, S., Pham, Q.B., Mohammadi, B., Karimi, F. and Linh, N.T.T. (2021b), "Machine learning algorithm-based risk assessment of riparian wetlands in Padma River Basin of Northwest Bangladesh", *Environmental Science and Pollution Research*, pp. 1-22.
- Jahan, C.S., Rahaman, M.F., Arefin, R., Ali, M.S. and Mazumder, Q.H. (2019), "Delineation of groundwater potential zones of Atrai-Sib river basin in north-west Bangladesh using remote sensing and GIS techniques", *Sustainable Water Resources Management*, Vol. 5 No. 2, pp. 689-702.
- Johnson, T.M., Roback, R.C., McLing, T.L., Bullen, T.D., DePaolo, D.J., Doughty, C., Hunt, R.J., Smith, R.W., Cecil, L.D. and Murrell, M.T. (2000), "Groundwater 'fast paths' in the Snake River Plain aquifer: radiogenic isotope ratios as natural groundwater tracers", *Geology*, Vol. 28 No. 10, pp. 871-874.
- Khosravi, K., Panahi, M. and Tien Bui, D. (2018), "Spatial prediction of groundwater spring potential mapping based on an adaptive neuro-fuzzy inference system and metaheuristic optimization", *Hydrology and Earth System Sciences*, Vol. 22 No. 9, pp. 4771-4792.
- Kumar, A., Malyan, S.K., Kumar, S.S., Dutt, D. and Kumar, V. (2019), "An assessment of trace element contamination in groundwater aquifers of Saharanpur, Western Uttar Pradesh, India", *Biocatalysis and Agricultural Biotechnology*, Vol. 20, p. 101213.
- Kuncheva, L.I. and Plumptre, C.O. (2010), "Choosing parameters for random subspace ensembles for fMRI classification", *International Workshop on Multiple Classifier Systems*, Springer, Berlin, Heidelberg, pp. 54-63.

- Lee, S., Hong, S.M. and Jung, H.S. (2018), "GIS-based groundwater potential mapping using artificial neural network and support vector machine models: the case of Boryeong city in Korea", *Geocarto International*, Vol. 33 No. 8, pp. 847-861.
- Mahato, S., Pal, S., Talukdar, S., Saha, T.K. and Mandal, P. (2021), "Field based index of flood vulnerability (IFV): a new validation technique for flood susceptible models", *Geoscience Frontiers*, Vol. 12 No. 5, p. 101175.
- Mallick, J., Talukdar, S., Alsubih, M., Almesfer, M.K., Shahfahad, Hoang, T.H. and Rahman, A. (2021a), "Integration of statistical models and ensemble machine learning algorithms (MLAs) for developing the novel hybrid groundwater potentiality models: a case study of semi-arid watershed in Saudi Arabia", *Geocarto International*, pp. 1-35, (just-accepted).
- Mallick, J., Talukdar, S., Alsubih, M., Ahmed, M., Islam, A.R.M.T., Shahfahad and Thanh, N.V. (2021b), "Proposing receiver operating characteristic-based sensitivity analysis with introducing swarm optimized ensemble learning algorithms for groundwater potentiality modelling in Asir region, Saudi Arabia", *Geocarto International*, pp. 1-28.
- Mukherjee, A., Sarkar, S., Chakraborty, M., Duttgupta, S., Bhattacharya, A., Saha, D., Bhattacharya, P., Mitra, A. and Gupta, S. (2021), "Occurrence, predictors and hazards of elevated groundwater arsenic across India through field observations and regional-scale AI-based modeling", *Science of The Total Environment*, Vol. 759, p. 143511.
- Mogaji, K.A., Omosuyi, G.O., Adelusi, A.O. and Lim, H.S. (2016), "Application of GIS-based evidential belief function model to regional groundwater recharge potential zones mapping in hardrock geologic terrain", *Environmental Processes*, Vol. 3 No. 1, pp. 93-123.
- Naghbi, S.A., Pourghasemi, H.R. and Abbaspour, K. (2018), "A comparison between ten advanced and soft computing models for groundwater qanat potential assessment in Iran using R and GIS", *Theoretical and Applied Climatology*, Vol. 131 No. 3, pp. 967-984.
- Namous, M., Hssaisoune, M., Pradhan, B., Lee, C.W., Alamri, A., Elaloui, A., Edahbi, M., Krimissa, S., Eloudi, H., Ouayah, M. and Elhimer, H. (2021), "Spatial prediction of groundwater potentiality in large semi-arid and karstic mountainous region using machine learning models", *Water*, Vol. 13 No. 16, p. 2273.
- Nampak, H., Pradhan, B. and Abd Manap, M. (2014), "Application of GIS based data driven evidential belief function model to predict groundwater potential zonation", *Journal of Hydrology*, Vol. 513, pp. 283-300.
- Nguyen, P.T., Ha, D.H., Avand, M., Jaafari, A., Nguyen, H.D., Al-Ansari, N., Van Phong, T., Sharma, R., Kumar, R., Le, H.V. and Ho, L.S. (2020), "Soft computing ensemble models based on logistic regression for groundwater potential mapping", *Applied Sciences*, Vol. 10 No. 7, p. 2469.
- 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)", *Journal of Hydrology*, Vol. 405 Nos 1-2, pp. 123-136.
- Pal, S., Kundu, S. and Mahato, S. (2020), "Groundwater potential zones for sustainable management plans in a river basin of India and Bangladesh", *Journal of Cleaner Production*, Vol. 257, p. 120311.
- Panahi, M., Sadhasivam, N., Pourghasemi, H.R., Rezaie, F. and Lee, S. (2020), "Spatial prediction of groundwater potential mapping based on convolutional neural network (CNN) and support vector regression (SVR)", *Journal of Hydrology*, Vol. 588, p. 125033.
- Pande, C.B., Moharir, K.N., Singh, S.K. and Dzwairo, B. (2020), "Groundwater evaluation for drinking purposes using statistical index: study of Akola and Buldhana districts of Maharashtra, India", *Environment, Development and Sustainability*, Vol. 22 No. 8, pp. 7453-7471.
- Park, S., Hamm, S.Y., Jeon, H.T. and Kim, J. (2017), "Evaluation of logistic regression and multivariate adaptive regression spline models for groundwater potential mapping using R and GIS", *Sustainability*, Vol. 9 No. 7, p. 1157.
- Prasad, P., Loveson, V.J., Kotha, M. and Yadav, R. (2020), "Application of machine learning techniques in groundwater potential mapping along the west coast of India", *GIScience and Remote Sensing*, Vol. 57 No. 6, pp. 735-752.

- Rahman, A., Tiwari, K.K. and Mondal, N.C. (2020), "Assessment of hydrochemical backgrounds and threshold values of groundwater in a part of desert area, Rajasthan, India", *Environmental Pollution*, Vol. 266, p. 115150.
- Rahmati, O., Pourghasemi, H.R. and Melesse, A.M. (2016), "Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: a case study at Mehran Region, Iran", *Catena*, Vol. 137, pp. 360-372.
- Razandi, Y., Pourghasemi, H.R., Neisani, N.S. and Rahmati, O. (2015), "Application of analytical hierarchy process, frequency ratio, and certainty factor models for groundwater potential mapping using GIS", *Earth Science Informatics*, Vol. 8 No. 4, pp. 867-883.
- Saha, T.K., Pal, S., Talukdar, S., Debanshi, S., Khatun, R., Singha, P. and Mandal, I. (2021), "How far spatial resolution affects the ensemble machine learning based flood susceptibility prediction in data sparse region", *Journal of Environmental Management*, Vol. 297, p. 113344.
- Sameen, M.I., Pradhan, B. and Lee, S. (2019), "Self-learning random forests model for mapping groundwater yield in data-scarce areas", *Natural Resources Research*, Vol. 28 No. 3, pp. 757-775.
- Skurichina, M. and Duin, R.P. (2002), "Bagging, boosting and the random subspace method for linear classifiers", *Pattern Analysis and Applications*, Vol. 5 No. 2, pp. 121-135.
- Straatsma, M.W. and Baptist, M.J. (2008), "Floodplain roughness parameterization using airborne laser scanning and spectral remote sensing", *Remote Sensing of Environment*, Vol. 112 No. 3, pp. 1062-1080.
- Talukdar, S. and Pal, S. (2020), "Wetland habitat vulnerability of lower Punarbhaba river basin of the uplifted Barind region of Indo-Bangladesh", *Geocarto International*, Vol. 35 No. 8, pp. 857-886.
- Talukdar, S., Pal, S., Chakraborty, A. and Mahato, S. (2020), "Damming effects on trophic and habitat state of riparian wetlands and their spatial relationship", *Ecological Indicators*, Vol. 118, p. 106757.
- Talukdar, S., Eibek, K.U., Akhter, S., Ziaul, S., Islam, A.R.M.T. and Mallick, J. (2021a), "Modeling fragmentation probability of land-use and land-cover using the bagging, random forest and random subspace in the Teesta River Basin, Bangladesh", *Ecological Indicators*, Vol. 126, p. 107612.
- Talukdar, S., Mankotia, S., Shamimuzzaman, M. and Mahato, S. (2021b), "Wetland-inundated area modeling and monitoring using supervised and machine learning classifiers", *Advances in Remote Sensing for Natural Resource Monitoring*, pp. 346-365.
- Termeh, S.V.R., Khosravi, K., Sartaj, M., Keesstra, S.D., Tsai, F.T.C., Dijkma, R. and Pham, B.T. (2019), "Optimization of an adaptive neuro-fuzzy inference system for groundwater potential mapping", *Hydrogeology Journal*, Vol. 27 No. 7, pp. 2511-2534.
- Tolche, A.D. (2021), "Groundwater potential mapping using geospatial techniques: a case study of Dhungeta-Ramis sub-basin, Ethiopia", *Geology, Ecology, and Landscapes*, Vol. 5 No. 1, pp. 65-80.
- Wang, G. and Ma, J. (2011), "Study of corporate credit risk prediction based on integrating boosting and random subspace", *Expert Systems with Applications*, Vol. 38 No. 11, pp. 13871-13878.
- Zhao, X. and Chen, W. (2020), "GIS-based evaluation of landslide susceptibility models using certainty factors and functional trees-based ensemble techniques", *Applied Sciences*, Vol. 10 No. 1, p. 16.

**Corresponding author**

Swapan Talukdar can be contacted at: [swapantalukdar65@gmail.com](mailto:swapantalukdar65@gmail.com)

---

For instructions on how to order reprints of this article, please visit our website:

[www.emeraldgroupublishing.com/licensing/reprints.htm](http://www.emeraldgroupublishing.com/licensing/reprints.htm)

Or contact us for further details: [permissions@emeraldinsight.com](mailto:permissions@emeraldinsight.com)