Journal of Cleaner Production 336 (2022) 130407

Contents lists available at ScienceDirect

Journal of Cleaner Production

journal homepage: www.elsevier.com/locate/jclepro

A coupled novel framework for assessing vulnerability of water resources using hydrochemical analysis and data-driven models

Abu Reza Md. Towfiqul Islam [a,b,1,*], Subodh Chandra Pal [c,1,**], Rabin Chakrabortty [c], Abubakr M. Idris [d,e,***], Roquia Salam [b], Md Saiful Islam [f], Anwar Zahid [g], Shamsuddin Shahid [h], Zulhilmi Bin Ismail [a]

[a] Centre for Coastal and Ocean Engineering (COEI), Universiti Teknologi Malaysia (UTM), Johor Bahru, 81310, Malaysia
[b] Department of Disaster Management, Begum Rokeya University, Rangpur, 5400, Bangladesh
[c] Department of Geography, The University of Burdwan, Bardhaman, 713104, West Bengal, India
[d] Department of Chemistry, College of Science, King Khalid University, Abha, 62529, Saudi Arabia
[e] Research Center for Advanced Materials Science (RCAMS), King Khalid University, Abha, 62529, Saudi Arabia
[f] Department of Soil Science, Patuakhali Science and Technology University, Dumki, Patuakhali, 8602, Bangladesh
[g] Bangladesh Water Development Board (BWDB), Dhaka, Bangladesh
[h] School of Civil Engineering, Faculty of Engineering, Universiti Teknologi Malaysia (UTM), Johor Bahru, 81310, Malaysia

ARTICLE INFO

Handling Editor: M.T. Moreira

Keywords:
BRT
RF
Classic model
Southern Bangladesh
Coastal plain
Vulnerability map

ABSTRACT

Mapping vulnerability of water resources (VWR) is crucial for the sustainable management of water resources, particularly in freshwater-scarce coastal plains. This research aims to construct a coupled novel framework technique for assessing VWR using hydrochemical properties and data-driven models, e.g., Boosted Regression Tree (BRT), Random Forest (RF) with Support Vector Regression (SVR) as a classic model through k-fold cross-validation (CV). A total of 380 groundwater samples were collected during the dry and wet seasons to construct an inventory map. The models were used to demarcate the vulnerable zones from sixteen vulnerability causal factors using a 4-fold CV approach. Obtained results were validated using the area under the curve (AUC) of receiver operating characteristic (ROC), sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). The results showed the excellent capability of the models to identify the VWR zones in the coastal plain. The RF model showed higher performance (AUC = 0.93, NPV = 0.89, PPV = 0.86, specificity = 0.85, sensitivity = 0.90) than others models. The south-central and southwestern areas had a higher VWR due to salinity, NO_3^-, F^- and As pollution in the coastal plain. Groundwater As, NO_3^- and F^- pollution should be urgently monitored and possibly controlled in areas of high VWR. Decision-makers and water managers can utilize the VWR maps, derived usinga coupled novel framework, to achieve sustainable groundwater management and prevent anthropogenic activities at the regional scale.

1. Introduction

The global trend in freshwater supply is not optimistic. The Organization for Economic Cooperation and Development (OECD) has projected an increase in water demand by around 55% to meet the growing water needs in industry, energy, and other public sectorsby 2050. This would put 240 million global people at risk of freshwater scarcity and retard the economic growth of many countries (OECD-Organization for Economic Co-operation and Development, 2011). Freshwater vulnerability assessment is therefore important to identify the vulnerable zones and causative factors for vulnerability mitigation planning. The vulnerability of water resource (VWR) is a term used to represent the probability of water contamination by geogenic and human activities (Thapa et al., 2018). The nitrogen cycle on the earth's surface has been changed due to anthropogenic activities, resulting in high nitrate concentration in soil and significant water quality degradation. Besides,

* Corresponding author. Centre for Coastal and Ocean Engineering (COEI), Universiti Teknologi Malaysia (UTM), Johor Bahru, 81310, Malaysia.
** Corresponding author.
*** Corresponding author. Department of Chemistry, College of Science, King Khalid University, Abha, 62529, Saudi Arabia.
E-mail addresses: towfiq_dm@brur.ac.bd (A.R.Md.T. Islam), geo.subodh@gmail.com (S.C. Pal), abubakridris@hotmail.com (A.M. Idris).
[1] Abu Reza Md. Towfiqul Islam and Subodh Chandra Pal contributed equally in this work.

https://doi.org/10.1016/j.jclepro.2022.130407
Received 28 August 2021; Received in revised form 30 November 2021; Accepted 4 January 2022
Available online 6 January 2022

unsustainable economic activities have caused groundwater pollution by trace metals and total-dissolved carbon and nitrates, resulting in degradation in aquifer water quality and environmental pollution (Clemens et al., 2020; da Costa et al., 2021). Groundwater vulnerability assessment is generally conducted to identify vulnerable regions and determine the drivers of vulnerability (Wang et al., 2022).

Vulnerability of water resource refers to a state of change the water resource where there is a risk or possibility of any harmful effect to the society (Kaushik et al., 2021). Intrinsic and specific are the two ways of categorizing VWR (NRC, 1993). If the surface contaminants spread into different groundwater areas, it is the intrinsic vulnerability (Vrba and Zaporozec, 1994). On the other hand, if the groundwater is susceptible to a definite contaminant, it is known as the specific vulnerability. The intrinsic vulnerability can be formulated as a weighted sum of hydro-geological factors, such as groundwater depth, net recharge, the aquifer media, soil media, topography, and hydraulic conductivity (Hein et al., 2021). This simple model is very flexible that can incorporate other factors, such as land use and nitrogen inputs, to calculate the specific vulnerability. The vulnerability indicators help monitor and keep track of changing exposure over time and space (Rajput et al., 2020).

Despite technological development, water quality parameters are still fundamental indicators of VWR assessment. Addressing the water quality issues to find solutions to reduce VWR requires a coupled and inclusive management plan with particular consideration for lessening the vulnerability of the global coastal inhabitants (Erostate et al., 2020). The present study recognizes that long-term monitoring is required to address the water quality factors responsible for VWR. This research also suggests the coupled framework approach to find the critical factors that can be managed more efficiently to reduce VWR in the long run. An effort is made to find the VWR of the coastal plains of Bangladesh by identifying the influencing factors, such as salinity, temperature, trace elements, nitrate, fluoride, chloride, etc.,using novel ensemble tree-based models, namely Boosted Regression Tree (BRT) and Random Forest (RF) along with Support Vector Regression (SVR), as a benchmark model.

Various bivariate and multivariate regression methods have been developed to be used with geographical information systems for spatiotemporal modeling of VWR (Anandhi and Kannan, 2018). Several approaches, varying from knowledge-based and data-driven methods to numerical modelling, have been used to assess the VWR (Vu et al., 2021). Numerical models provide a mechanistic understandingof the VWR; however, they require a huge amount of experimental data and computational ability. Therefore, the use of numerical methods for VWR assessment at the regional levelsis very scarce. Data-driven methods, such as ML methods, have emerged as potential tools for assessing water resources and characterizing various physicochemical parameters causing VWR with limited data (Knoll et al., 2019). Researchers have recently paid considerable attention to developing ML methods and adopted them widely to address different constraints associated with the VWR assessment (Liaw et al., 2018). Typical ML toolsthat have been used for the improvement of groundwater resource modellinginclude RF (Roy et al., 2020), Decision Tree (Pham et al., 2017), boosted regression tree (Nolan et al., 2015), the extreme-learning machine technique (Barzegar et al., 2018), support vector machines (Sajedi-Hosseini et al., 2018), artificial neural networks (Nolan et al., 2015), deep learning (Bui et al., 2020) and locally weighted regression (Khalil et al., 2005). However, many of these models have several drawbacks. For example, ANN often shows poor performance (Melesse et al., 2011), and SVM needs suitable kernel function selection for performance enhancement (Sheikh Khozani et al., 2019). In recent years, advanced ensemble tree-based tools, such as data-driven methods, have been used in water resource studies for their inherent advantages. For instance, ensemble tree-based algorithm (e.g. RF) has been utilized to capture complicated nonlinear behaviour to achieve higher predictive precision (Saha et al., 2020). The RF model's application in different fieldsshowed its capability to outperform other methods. For example, Khan et al. (2021) utilized RF, SVM, C4, and Naïve Bayes to assess the VWR at the Al Khatim region of the United Arab Emirates and have found RF as the best-performed model. The RF has also shown it better performance than the traditional artificial intelligence models in groundwater modelling in different regions, including Africa (Ouedraogo et al., 2019), Iran (Band et al., 2020), South and North America (Ransom et al., 2017), and Germany and Spain (Knoll et al., 2019). Boosted regression tree (BRT) is another efficient boosting ensemble tree-based model (Buston and Elith, 2011). The main advantage of BRT is its capacity to spontaneously manage predictor valuesthrough interactions (Elith et al., 2008). Moreover, the SVR is a flexible tool with high efficiency in solving multimodal optimization issues. However, its potential as a spatial modelling approach for VWR mapping is still unexplored. Therefore, the present study employed the BRT and RF models, trained and validatedusing k-fold cross-validation method,for VWR assessment.

Bangladesh is highly vulnerable to extreme weather phenomena, particularly floods, droughts, sea-level rise, cyclones and storm surges. It has nearly 600 km of coastline with three distinct coastal plains: western, central and eastern coastal zones. About 10% of the country is barely 1 m above the mean sea level (MSL),of which one-third is under tidal excursions with anthropogenic activities. Thishas made the country globally most vulnerable to water-related hazards (Islam et al., 2021). Bangladesh's vast low-lying southeastern and southwestern coastal regions are particularly vulnerable to climate change risks that affect the VWR (IPCC, 2014). Besides, increased temperatures, changing rainfall patterns, sea-level rise, land-use change and increased frequency and intensity of extreme climate phenomena have adversely impacted the region's water resources, human health, and vulnerability to coastal ecosystems (GoB, 2018). However, VWR assessment is still lacking in the country due to limited data availability. Therefore, addressing technology-based solutions, such as data-driven machine learning (ML) methods for modelling VWR is extremely important.The primary objective of this research is to assess the VWR in the coastal plains of Bangladesh using hydrochemical analysis and data-driven models, namely BRT and RF and SVR and to compare various data-driven methods for mapping VWR and identifying the factor influencing it. This is the first inclusive VWR assessment study in Bangladesh, where the data scarcity challenge is attempted to overcome using a framework of advanced data-driven models and cross-validation techniques. The proposed study is remarkable by the green concept for the future of water sources. This novel framework can also aid groundwater resource management in other freshwater stressed coastal regions of the world.

The remaining parts of the articleare arranged as follows. Section 2 briefly describes the study area, experimental data and the data-driven methods utilized. Section 3 presents the results of the study. A critical discussion of the results is presented in Section 4 and the conclusion in Section 5.

2. Materials and methods

2.1. Study area description

The low-lying deltaic coastal region, located in the southern plain of Bangladesh along the Bay of Bengal (BoB), is considered the study area. The coastal plain comprises 19 districts out of 64 districts of Bangladesh and covers 24000 km^2 (Fig. 1). The coastal plain contains three distinct areas: (1) Sundarbans is a mangrove area, is distributed in the inactive southwest fluvial delta plain formed by the confluence of the Ganges, Brahmaputra and Meghna Rivers in the BoB; (2) active south-central delta plain; and (3) the southeastern narrow-band coastal plain (Tasnuva et al., 2020). The increase in the sea level, rapid population growth, poor drainage patterns, saline water intrusion, etc., are the key factors for the VWR. It makes the coastal residents prone to severe hazards (Kabir et al., 2021). The primary livelihood pattern in the coastal plain is mostly aquaculture, homestead gardening and agro-farming activities. Aquaculture, mainly shrimp culture, has

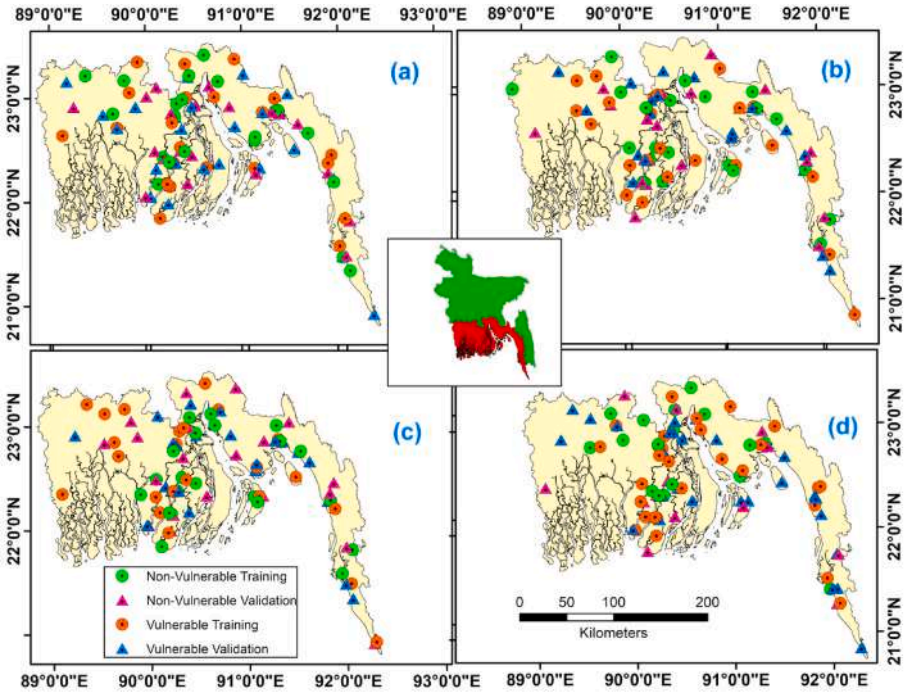

Fig. 1. Location of the study area showing coastal plain, Bangladesh.

increased in recent decades while, crop farming has declined due to a higher level of soil salinity (Islam et al., 2020a).

The coastal region of Bangladesh belongs to a tropical climate, mainly hot, humid and dry, influenced by the southwest monsoonal flow regime from the BoB. The mean annual precipitation and average temperatures are 2000–2500 mm (June–September) and 25 °C, respectively (Islam et al., 2021). During the monsoon period, 10–15% of rainfalloccurs due tothunderstorm. The low-lying topography of the coastal plain varies from west to east. The land use/land cover is the most diverse and dominated mainly by agricultural land, settlements, vegetation, and surface water bodies (Fig. S1). The urban area covers about 25% of the total plain. Generally, the coastal plain faces freshwater scarcity due to severe water quality degradation. The detailed information ongeology and hydrogeology in the coastal plains can be found in supplementary Text S1.

2.2. Experimental data and sample design

Generally, a large dataset is required for modeling using data-driven ML methods. In this research, the datasets were gathered from Islam et al. (2020a, b). A total of 380 water samples from the three coastal multi-aquifers, deep (146–300m), middle (71–145) and shallow (15–70), were obtained based on depth during the dry and wet seasons. Each sample site was well documented with a designatedidentificationnumber. The coordinates were confirmed using a portable GPS (Explorist 200, Magellan). Fig. 1 shows the location of the sampling sites. Three campaign works were conducted during 2015–2017 for collecting samples. Prewashed high-density polypropylene (HDPP) bottles were utilized for gathering the experimental data (APHA, 2005). The samples had obtained, followed by filtering through 0.45 μm membranes (MF-Millipore™, USA). The HDPP bottles were used for preservation in a cooler box at 4 °C. All samples were subsequently analyzed at the lab of the Bangladesh Water Development Board (BWDB). The EC and pH were determined by the portable meters for observing pH/EC (Hanna HI 9811–5). Depth of water, salinity and temperature were also measured for the dry and wet seasonsthrough field survey. Cations (Ca^{2+}, Mg^{2+}, Na^{+}, K^{+}) and anions (Cl^{-}, HCO_3-, NO_3^{-}, PO_4^{2-}, SO_4^{2-}, and F^{-}) were analyzedby ion chromatography using Dionex ICS 90.Arsenic (As) was analyzed using the Thermo Scientific X-Series2 ICP.The detailed information of hydrochemical parameters with the methods used for their determination, types of equipment, and limit of detections used for the present study are presented in Supplementary Table S1. Five standard solutions (1, 5, 10, 15 and 20 mg/L) were used for the calibration procedure. Quality assurance was attained by using a standard laboratory process and quality control checks. Three replicated samples were achieved concurrently to verify the accuracy of the tested results by cross-checking with a certified laboratory. The present study has considered the ion charge balance error (ICBE), utilized for determining accuracy, ranged 3.62–8.56%, with a mean of 8.12%, within the acceptable limit of ±10%.

2.3. Vulnerability of water resource mapping

2.3.1. Proposed coupled novel framework approach

The VWR in the coastal districts is mainly due to elevated nitrate, fluoride and arsenic contaminations in groundwater which are well-established pollutants in the coastal plain (Islam et al., 2021). Thus, we choose to model these pollutants and hydrochemical variables. Groundwater samples collected from the 380 sampling sites were used to give a VRW map in the coastal regions. Among them, 190 sampling points were obtained for the dry season, and 190 sampling points were collected for the wet monsoon season from the coastal plain of southern Bangladesh, which was adopted for modelling and mapping VWR. Subsequently, the maps of sixteen VWR causal factors were generated from the sample data for both wet and dry seasons using the Inverse distance weighted (IDW) interpolation tool in a GIS platform after the hydrochemical analyses of the sixteen variables from each sampling site was done effectively. However, a comprehensive hydrochemical properties assessment is essential to verify a dense monitoring location and large datasets for consequence analysis.

The parameters related to water resources vulnerability were selected considering their prevalence in the study area as found in existing literature (Iqbal et al., 2012; Ukpai and Okogbue, 2017). Then, multicollinearity tests were performed to choose the vulnerability causal factors for prediction of VWR (Salinity, depth, EC, pH, temperature, Ca^{2+}, Mg^{2+}, K^{+}, Na^{+}, Cl^{-}, HCO_3^{-}, NO_3^{-}, SO_4^{2-}, PO_4^{2-}, F^{-} and As).The

dependent variables, i.e., water resource vulnerability, were estimated by considering various qualitative and quantitative parameters, field investigation, and expert knowledge. For analyzing water resource vulnerability, a variety of methods have been utilized or suggested. They vary from complex models of the physical, chemical, and biological processes in the vadose zone and ground water regime to simple models that use statistical methods or expert opinion to weigh important elements impacting susceptibility. Uncertainty is a common feature of all vulnerability assessment methods, whether in the technique itself or the data used.Two ensemble tree-based algorithms, BRT andRF, and a benchmark model, SVR, were utilized to prepare the VWR map. The data in all VWR maps were demarcated into five classes: very low, low, medium, high, and very high VWR using the natural-break classification tool in the GIS platform. The relative contribution of the vulnerability causal factors to the VWR predictions was determined using an RF model. Generally, a weighted integration of a specific model is mainly employed in the ensemble tree-based framework (e.g., RF and boosting) (Sajedi-Hosseini et al., 2018). Subsequently, a resampling technique such as k-fold cross-validation method was used to estimate the generalized performance of three ML algorithms.

Lastly, VWR sampling sites were randomly demarcated into four folds (i.e. Fold-1, Fold-2, Fold-3 and Fold-4) for a 4-fold cross-validation scheme. Three folds (75% = 285 sampling sites) of the dataset were used to train the model, and one fold (25% = 95 sampling sites) was used to validate the proposed model. In such a way, four integrations of training and validation datasets were adopted to construct predictions of VWR. A 75% of the entire dataset was regarded for model construction (training stage) (285 samples), and the rest of 25% was adopted for model assessment (testing stage) (95 samples). A ratio of 75:25 is the most commonly applied approach since there is no agreement on data division for model training and testing (Khosravi et al., 2018; Bui et al., 2020). For validation of the proposed models' performance, receiver operating characteristic (ROC)'s AUC (area under the curve), sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were used (Mallick et al., 2021). The developed tentative methodology is described in the flowchart (Fig. 2). Based on the proposed model's performance, these advanced methods were appraised. Their plausible results are described in the subsequent section.

2.3.2. Vulnerability causal factors in the dry season

Fig. 3a shows the distribution of As during the dry season. The As in groundwater is high in central and southeastern coastal areas during the dry season, especially in Gopalganj, Bagerhat, Lakhmipur and Cox's Bazar districts of the coastal areas. Bhuiyan et al. (2016) described that human activities, includingagro-chemicalpesticdes and animal feeding-induced natural sources, may be responsible for highAs levelsin these areas. The Ca^{2+} was higher in the central and southern districts (Fig. 3b), while Cl^- showed significant spatial variations during the dry season (Fig. 3c). The water depth follows a distinct spatial pattern during the dry season (Fig. 3d), a higher water depth in northern and southeastern and a lower depth in the southern and central coastal plains. The shallow aquifer (depth <100 m) is affected by salinity intrusion, but the deep aquifer (depth >150 m) is free from pollution (Ravenscroft et al., 2013).

The EC values were higher near the BoB during the dry season (Fig. 3e). This is due to the sampling points' depth variation with the coastal rivers' flow regimes. Human activities might also have influenced the enhancement of EC in the coastal plain. Fig. 3f shows a diverse spatial distribution of F^- during the dry season. The high contents of F^- was found in the south and southeast of the coastal plain(Edmunds and Smedley, 2005), especially Bhola and Cox's Bazar districts, exceeding the limits for human consumption set by WHO (2011). A higher distribution of HCO_3^- was detected in south-central and southern regions (Fig. 3g), whilethe K^+ value exhibited higher amounts in the northern and southwestern parts (Fig. 3h). Mg^{2+} showed a similar distribution to K^+ during the dry season (Fig. 3i). More than half of the area showed a lower occurrence of Mg^{2+}. An identical distribution pattern was found for Na^+, i.e., a higher occurrence in central and southwestern regions (Fig. 3j). The NO_3^- values were higher in the south-central part of the study area (Fig. 3k) due to the excessive use of agrochemical fertilizers and infiltration from the sewage systems (Akber et al., 2020). The diverse spatial pattern of pH was identified in the south-central, southwestern and southeastern parts (Fig. 3l). Most portions of the coastal region showed a moderate pH, slight alkalinity (Siddique et al., 2021). The lower concentration of PO_4^{2-} was found in most parts of the study area except over a few small pockets in the southern and northern regions (Fig. 3m). Salinity showed a similar distribution to Na^+ (Fig. 3n).

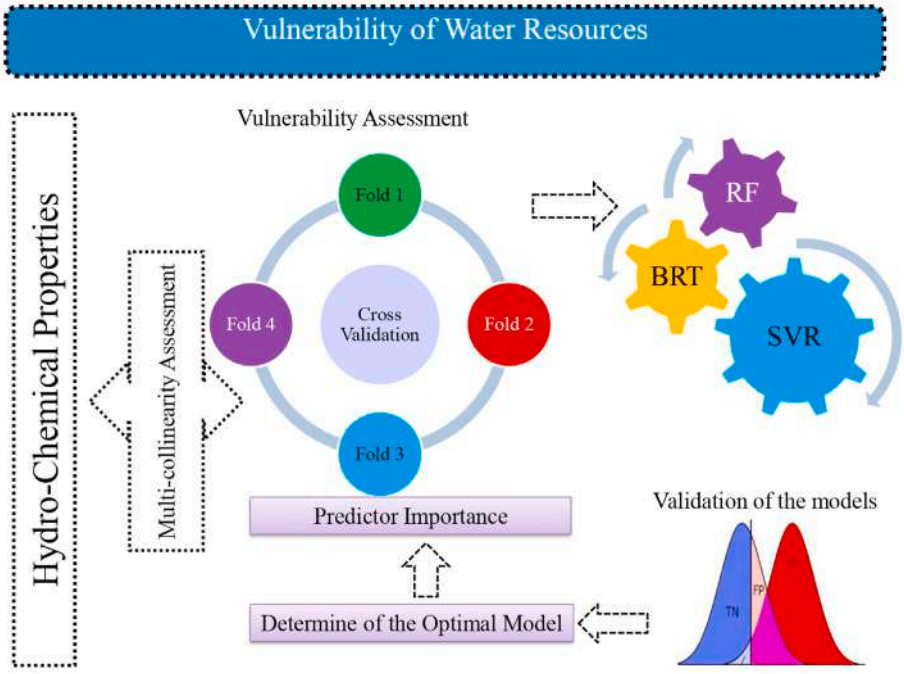

Fig. 2. The proposed coupled novel framework approach for this study.

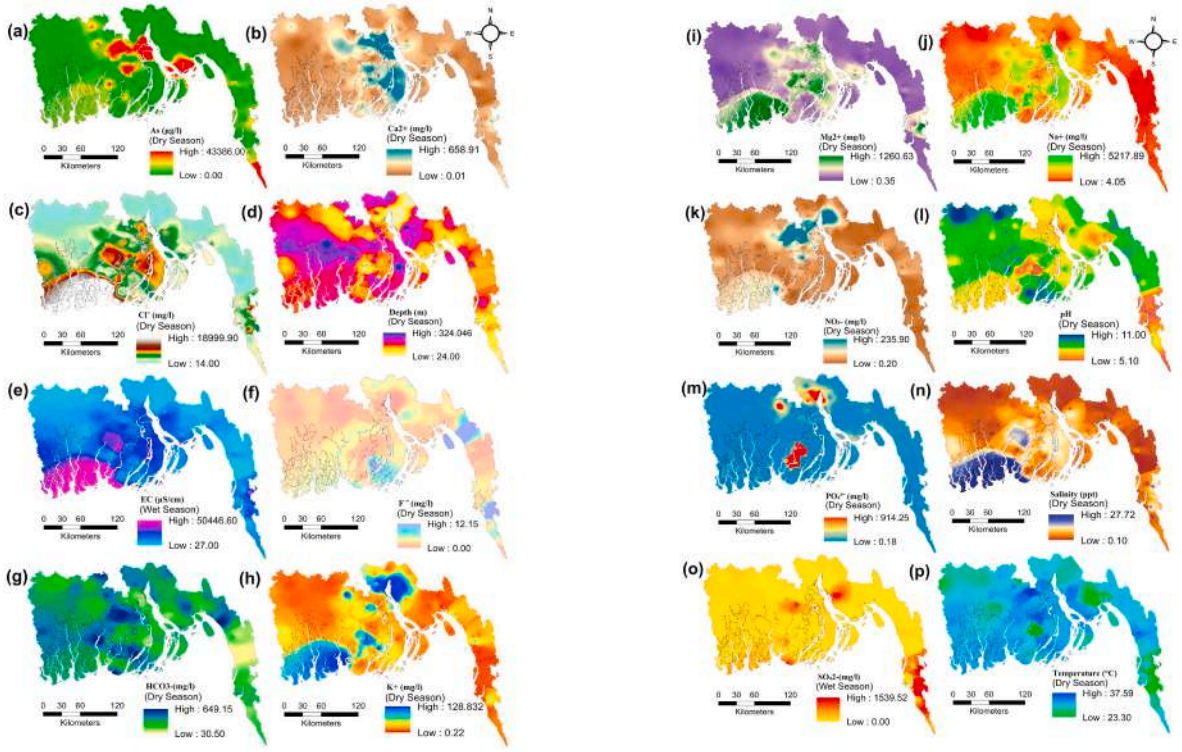

Fig. 3. Vulnerability causal parameters in dry season.

The high salinity contents were detected in the southwestern part, particularly in the coastal districts of Barguna and Patuakhali. This might be associated with the landward saltwater near the BoB, followed by enormous water overexploitation from the aquifers in the dry season. The SO_4^{2-} exhibited higher amounts in the central and southeastern parts of the coastal plain during the dry season (Fig. 3o). The temperature showed significant spatial variations during the dry seasons due to the impact of global climate warming (Islam et al., 2021) (Fig. 3p).

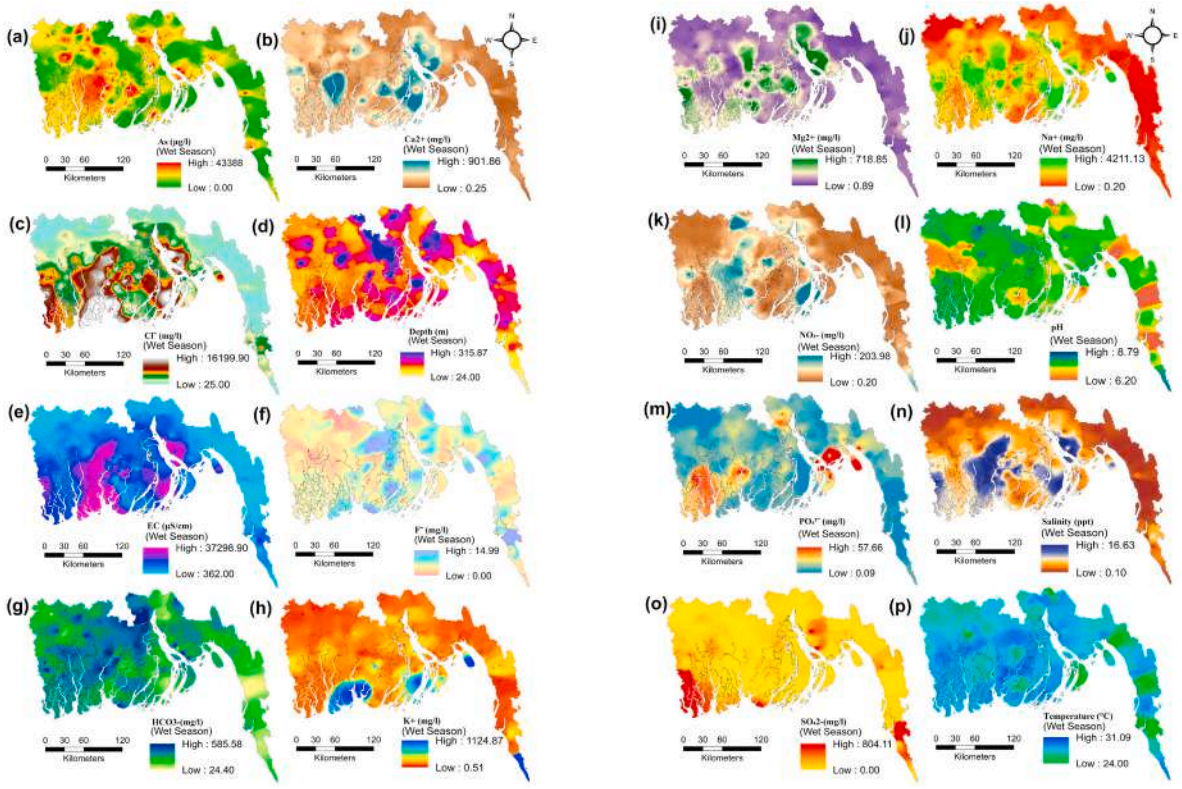

Fig. 4. Vulnerability causal parameters in wet season.

2.3.3. Vulnerability causal factors in the wet season

The heterogeneous distribution of As was noticed in the wet season (Fig. 4a). The high contents of As were observed in thesouthcentral and southwestern parts of the study area (Fig. 4a). The Ca^{2+} were more significant in the southwestern part (Fig. 4b). The Cl^{-} showed a diverse spatial pattern with higher values in the southern and southwestern regions (Fig. 4c). A varied distributionof water depth was also noticed during this season (Fig. 4d). Generally, ahigher depth was found inthesouthcentral and eastern regions, especially Barisal, Lakhmipur, and Noakahli districts, compared to othercoastal plain regions. The EC was higher in the southwestern parts than in the northern and eastern areas (Fig. 4e). This is due to the sampling point's depth variation and the flow regimes of the coastal rivers. Similarly, F^{-} concentration was higher in southcentral and southeast regions, especially Barisal and Cox's Bazar districts, exceeding the drinking quality limits (Fig. 4f). The high F^{-}may be due to hydro-geochemical drivers such as Na–HCO_3^{-} type water usually shows the elevated fluoride content (Edmunds and Smedley, 2005).

On the other hand, a varied irregular pattern of HCO_3^{-} was noticed in the whole coastal plain with a higher amount in southcentral and northern regions (Fig. 4g). However, a revised distribution was found for K^{+} with a higher occurrence in the south and southwest (Fig. 4h). A similar pattern was found for Mg^{2+}, the elevated concentration in the southcentral and western portions of the coastal plain (Fig. 3i). Furthermore, the highest Na^{+} was noticed in the southern and southcentral portions (Fig. 4j). There is also a noticeable spatial variation of NO_3^{-} (Fig. 4k). It was higher in the central and southern parts due to leakage of municipal sewage and the effect of agricultural return flow via the percolation of agro-pesticides and chemical fertilizers (Amiri et al., 2014). As observed from Fig. 4l, the pH was greater over the coastal regions, except for Sathkhira and Chatrogram. The PO_4^{2-} exhibited a higher content in southwestern and southeast portions (Fig. 4m). In contrast, a higher salinity was noticed in the southern and southcentral parts, particularly the Barguna, Pirojpur, and Patuakhali districts (Fig. 4n). Besides, SO_4^{2-} showed higher amounts in the southwestern part of the coastal plain (Fig. 4o). A higher temperature was found in the southcentral and western regionsduring the wet season (Fig. 4p). Overall, the spatial distribution of the vulnerability causal factors implied that the coastal plain is severely overstated by high salinity from the BoB, particularly the southcentral and southern portions of the study area.

2.4. Methodology

2.4.1. Multicollinearity assessment

The causal parameters of VWR must not be relatedto each other (Chowdhuri et al., 2021). In a multivariate regression model, multicollinearity occurs when substantial intercorrelations exist between two or more independent variables. In analyzing an independent variable's role in predicting or explaining the dependent variable, multicollinearity may lead to skewed or misleading conclusions.There are many approaches used for multicollinearity quantification, such as variance decomposition proportions, correlation co-efficient and variance inflation factors (VIF) and tolerances (TOL) (Talukdar et al., 2021). The VIF and TOL were employed in this study to assess multicollinearity. If VIF <10 and TOL >0.1, there are no collinearity, and causal factors are mutually separated (Saha et al., 2021). The TOL and VIF are computed using the following Eqs. (1 and (2).

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

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

where R_J^2 is the determination coefficient.

2.4.2. Background of the proposed method

The BRT model was implemented in R version 4.1.0. There areseveral boosted tree solutions that are not described in this section. The next two sections describe the process of fitting, assessing, and interpreting the BRT modeland the aspects that have made BRT helpful in water resource modelling. Different kernel functions are used to build SVR model, including a) Linear, b) Polynomial, c) Sigmoid, and d) Radial Basis. The selection of an appropriate kernel function is important for developing an SVR model. However, optimum selection of a kernel function needs optimization methods, which is out of the scope of the present study. The Radius Basis Function (RBF) kernel, the default kernel function used in R, was used. The RBF is most widely recommended for developing SVR models considering its capability to guide the SVM to capture nonlinear input-output relationships efficiently. Our data is transformed from nonlinear to linear space using the kernel function. In this stud, the R programming package randomForestwas used to implement the model.

Boosted Regression Tree (BRT). BRT is a hybrid machine learning model that aims to improve performance by fitting many models (Schapire, 2003). This model does not require previous data transformation or outlier removal. One of the advantages of this model is that it resolves problems of correlation between elements automatically (Elith et al., 2008). The regression and boosting are the two models that make the BRT together. Trees are insensitive to outliers and can use surrogates to fill in missing data in predictor variables (Elith et al., 2008). Boosting is a strategy for increasing model accuracy based on the concept that it is simpler to identify numerous rough rules of thumb than to find a single, highly accurate prediction rule (Schapire, 2003). The use of many trees in BRT overcomes the single tree model's main disadvantage of weak predictive power. The 'gbm' package of R was used to implement BRT model (Ridgeway, 2006). The'gbm.step' and 'gbm. simplify' functions were used for fitting and simplifying the model, respectively.

Support Vector Regression (SVR). SVR (Vapnik et al., 1995) is a popular ML algorithm widely used to model and control complicated engineering systems. It maps the relation between input and output variables usingthestructural risk minimization (SRM) technique, which can be expressed using the following equations,

$$y = k(z) = v\varnothing(z) + c \tag{3}$$

where input data represent through $z = (z_1, z_2, \ldots z_n)$ and the resultant value represent by $y_b \in R^l$. In addition to this, $v \in R^l$ indicates the weight factor, $c \in R^l$ indicatestheconstant number of amathematical function, l represents dataset size in the model, and $\varnothing(z)$ is the input datasets mapped using an irregular function. The following equations, developed -based on the SRM principles, can be used to define v and c:

$$Minimize: \left[\frac{1}{2}\left|\left|v\right|\right|^2 + P\sum_{b=1}^{1}\left(\zeta_b + \zeta_b^*\right)\right] \tag{4}$$

$$Subject\ to: \begin{cases} y_b - (v\varnothing(z_b) + c_b) \le \varepsilon + \zeta_b \\ (v\varnothing(z_b) + c_b) - y_b \le \varepsilon + \zeta_b^* \\ \zeta_b, \zeta_b^* \ge 0 \end{cases} \tag{5}$$

where penalty factor is P balances model flatness and its risk, ζ_b *and* ζ_b^* indicate loose variables, and ε represents the optimization performance of the model (Wang et al., 2012).

Random Forest (RF). RF has benefitted from other data mining tool decision trees (DT), to handle a large amountof data (Breiman, 2001). Two variables namely predictor and target variables are needed to train the data by this model (Breiman, 2001). The decision tree is intended to divide the data into homogenized subsets concerning the objective

variable by utilizing the predictor variables (Roy et al., 2020). In the simplest instance, binary classification is the goal variable. The root node contains both the goal and the predictor data. The predictor variable that has done the categorization precisely is picked, and each new node is retrieved. The division process continues until no more division has occurred. A single tree of choice is a poor classification. It often has high variance, i.e. the split at each node is so in line with the training data that it is impossible to use the model to predict fresh data. Besides, it can have high biases of less accuracy for showing correlation of predictor with the target variable. RF mitigates such difficulties by employing a set of decision-making bodies to balance the two causes of mistake. The bagging idea is extended further by RF. In addition to employing the sub-set of data, the random choice of predictor variables is required to cultivate trees instead of using all data. RF offers greater precision, considerably free from overfitting problems (Chowdhuri et al., 2021). The following equations may be represented as an algorithm of RF and the inaccuracy of its generalization.

$$GE = P_{x,y}(mg(x,y) < 0) \tag{6}$$

$$mg(x,y) = av_k I(h_k(x) = y) - max_{j \neq y} av_k I(h_k(x) = j) \tag{7}$$

When x and y are VWRs, *mg* shows the marginal functionandI(*) indicates the probability across x and y spacing.

2.4.3. Resampling method

The resampling method is essential in any ML method because they tune the optimal variables before precision. The higher the number of representatives in the sample size, the less bias in forecasting. This technique significantly impacts spatial dataset analyses, such as groundwater quality parameters with elevated levels of variability. Jackknife resampling tool, bootstrapping process, random sub-sampling technique, and k-fold cross-validation are the most popular resampling methods found in the existing literature. Of these, the k-fold cross-validation is the most robust technique, and therefore, used in this research (Islam et al., 2021).

k-fold cross-validation tool. The k-fold cross-validation tool is a typical method to measure the overall performance of various ML methods. Geisser (1975) proposed this CV method to minimize Leave One's calculation cost (LOOCV). It is achieved by arbitrarily dividing the resulting output into k-parts or folds (Pal et al., 2020), where a single fold is used for testing, and the remaining folds are used to fit the model. The process is repeated to use all the folds sequentiallyfor validation. This allows the whole set of data to be used for model calibration and validation. The average error for each observation helps to pick a suitable model. This study used the K-fold CV method to divide the data set into training and validation. The CV offers the following advantages: firstly, each data point has been examined; secondly, the partiality was reduced; finally, above necessary, the performance result increased.

2.4.4. Methods of validation and precision assessment

The proposed models were validated using five statistical indices, Sensitivity, Specificity, positive predictive value (PPV), negative predictive value (NPV), and area under receiver operating characteristics (ROC) curve (AUC). A model is considered good when the indices are high and vice versa. The ROC is a widely used performance evaluation tool (Pal et al., 2020; Islam et al., 2021). ROC is elucidated on the X- and Y-axis, relying on the sensibility and specificity, respectively (Pal et al., 2020; Saha et al., 2021). Sensitivity evaluates how the predictive ability of the vulnerability model is utilized to categorize water resources. Specificity donates accurate classification of non-vulnerable areas. The success rate and its related permanence of a method can be determined by the ROC curve. PPV is the possible predictive value that is classified as a water resource area. By contrast, NPV is the negative predictive value classified as a non-water resource vulnerable area. Eqs. 8–12 are used to compute the indices (Chen et al., 2021).

$$Sensitivity = \frac{TP}{TP + FN} \tag{8}$$

$$Specificity = \frac{TN}{FP + TN} \tag{9}$$

$$PPV = \frac{TP}{FP + TP} \tag{10}$$

$$NPV = \frac{TN}{TN + FN} \tag{11}$$

$$S_{AUC} = \sum_{k+1}^{n} (X_{k+1} - X_k) \left(S_k + 1 - S_{k+1} - \frac{S_k}{2} \right) \tag{12}$$

whereTP, TN, FP and FN are genuinely positive, genuinely negatively, false-positive and false-non rate, respectively.

Besides, the Taylor diagram (Taylor, 2001) was used to assess model performance. The Tylor diagram uses the correlation of coefficient and the standard deviation to show the contrast between models.

2.4.5. Statistical analyses

All datasets used in the current study were subjected to standard quality control checks. The Wilcoxon rank-sum test was used o measure the difference betweenhydrochemical parameters. Kolmogorov Smirnov (K–S) and ShapiroWilk (S–W) tests were used to determine the normality and homogeneity of hydrochemical data of all groundwater samples during the dry and wet seasons (Table S2). All the hydrochemicaldatawere found to follow a normal distribution at $p < 0.05$ significance level, except for a few variables at some locations. Besides, the parameter selection was reasonable according to the K–S and S–W tests. The spatial map of the elemental distribution was generated using the inverse distance weighting (IDW) interpolation technique within the ArcGIS platform (version 10.5). A significance level of $p < 0.05$ was considered for all the cases.

3. Results

3.1. Multicollinearity assessment

There is a general rule that if TOL is greater than 0.1 and the VIF is less than 10, there is nomulticollinearity among the conditioning factors of water resources. Different folds show different results in this study (Table 1). F- produced a TOL of 0.09 and a VIF of 11.11 in fold-2, indicating multicollinearity. However, it was not observed for other folds, like fold-1, 3 and 4. Therefore, F- was considered in this study for assessing the vulnerability of the water resources. The rest of the factors showed a VIF of less than 5, signifying their validity to carry out further analyses with those factors (Table 1).

3. 2Determination of optimum parameters for model tuning

The optimized parameters of BRT, SVR and RF models are outlined in Tables S3, S4, and S5, respectively. The parameters were estimated based on the lowest error. The parameter finally utilized for developing BRT (n.trees = 50, interaction. depth = 2, shrinkage = 0.1 and n. minobsinnode = 10 for fold-1, 2, 3 except for fold-4 where n.trees = 150, interaction. Depth = 3), SVR (Sigma = 0.0508, cost = 0.5 for fold-1, Sigma = 0.0498, cost = 0.5 for fold-2, Sigma = 0.0677, cost = 0.25 for fold-3 and Sigma = 0.0497, cost = 64 for fold-4) and RF (n. tree = 200 for all folds and m. tree = 8 for fold-1, 2, m. tree = 7 for fold-3 and m. tree = 16 for fold-4).

3.3. Vulnerability assessment of water resources

A total of 12 maps were generated for assessing the VWR of the study

Table 1
Multicollinearity assessment of the present study.

Factors	Collinearity Statistics							
	Fold 1		Fold 2		Fold 3		Fold 4	
	TOL	VIF	TOL	VIF	TOL	VIF	TOL	VIF
Depth	0.24	4.17	0.81	1.23	0.52	1.92	0.60	1.66
Temperature	0.51	1.96	0.71	1.41	0.33	3.03	0.56	1.79
pH	0.89	1.12	0.95	1.05	0.75	1.33	0.25	3.94
EC	0.30	3.33	0.26	3.85	0.94	1.06	0.35	2.86
Salinity	0.42	2.38	0.94	1.06	0.55	1.82	0.76	1.32
Ca^{2+}	0.48	2.08	0.74	1.35	0.55	1.82	0.31	3.22
Mg^{2+}	0.53	1.89	0.53	1.89	0.34	2.94	0.82	1.22
Na^{+}	0.51	1.96	0.43	2.33	0.48	2.08	0.74	1.35
K^{+}	0.49	2.04	0.62	1.61	0.71	1.41	0.22	4.63
Cl^{-}	0.41	2.44	0.35	2.86	0.55	1.82	0.35	2.83
HCO_3^{-}	0.69	1.45	0.76	1.32	0.48	2.08	0.60	1.68
NO_3^{-}	0.35	2.86	0.26	3.85	0.90	1.11	0.55	1.81
SO_4^{2-}	0.21	4.76	0.21	4.76	0.63	1.59	0.67	1.50
PO_4^{2-}	0.33	3.03	0.68	1.47	0.52	1.92	0.58	1.72
F^{-}	0.22	4.55	0.09	11.11	0.80	1.25	2.74	0.37
As	0.60	1.67	0.68	1.47	0.63	1.59	1.18	0.85

area using integrated hydrochemical datasets, BRT, SVR, and RF models with four folds (k = 4) CV, as presented in Fig. 5. The VWR was estimated for both wet and dry seasons. The combined scenario of the wet and dry seasonswas taken into consideration for estimating the overall VWR.The maps were prepared by categorizing the natural breakup classification method into the five vulnerability zones, very low, low, moderate, high, and very high. The results showed that the middle southcentral and southwestern parts of the study area have the highest

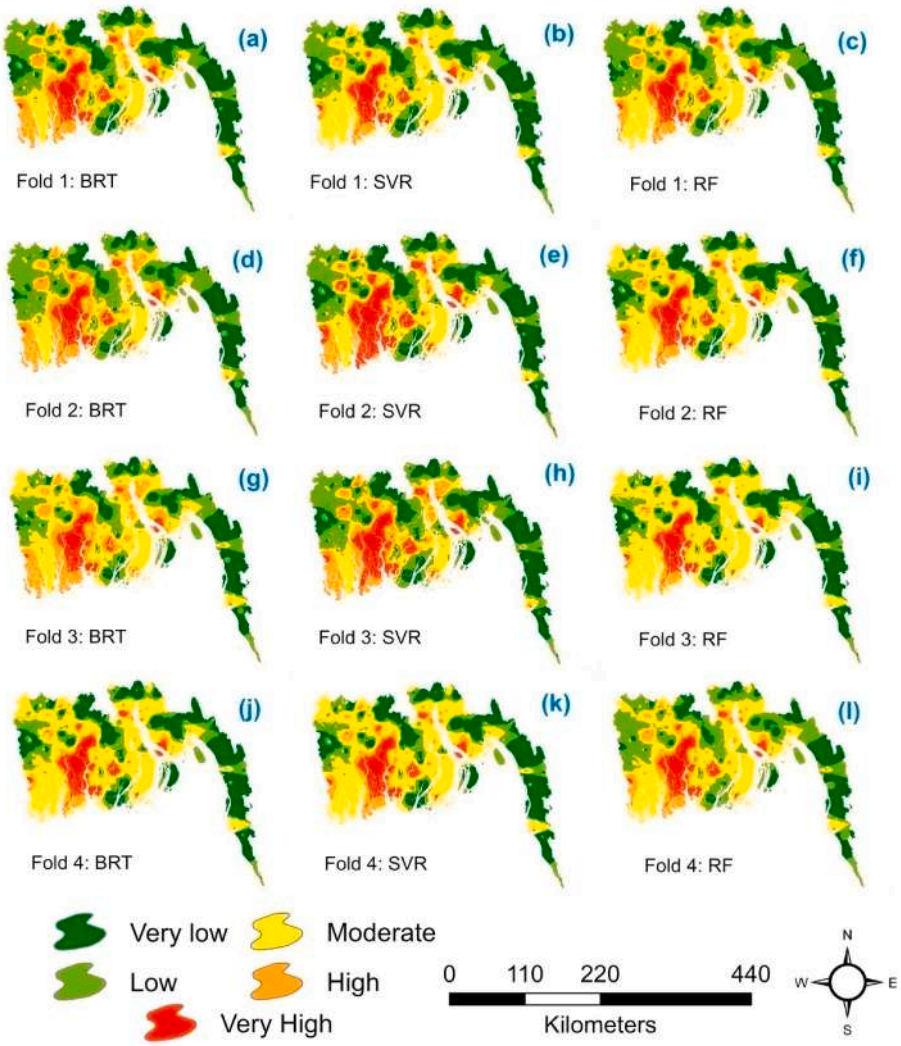

Fig. 5. Vulnerability assessment of water resources.

VWR. On the contrary, the southeastern and northern parts have the lowest vulnerability. The results also showed that among 19 coastal districts, the water resources of Bagerhatis most vulnerable, followed by some parts of Gopalgong, Pirojpur, Borguna, Patuakhali, Shariatpur, Noakhali, and Lakshmipur. On the other hand, some parts of Jashore, Satkhira, Gopalganj, Barishal, Shariatpur, Chandpur, Noakhali, Patuakhali, Lakshmipur, Feni, Chattogram, Rangamati, Khagrachari, and Bandarban are relatively less vulnerable. The rest of the areas are moderate to highly vulnerable.

Table 2 shows the area (%) that belongs to different vulnerability categories identified by the models. All the models identified water resources of most of the area as highly vulnerable, except for fold-3. The BRT, SVM, and RF demarcated 24.3%, 23%, and 22.37% areas, respectively, as very high vulnerable for fold-1. In contrast, the models showed 23.66%, 24.01%, and 21.23% area, respectively, are very high vulnerable for fold-3, and 21.99%, 22.01%, and 24.11% area are very high vulnerable for fold-4. On average, all models demarcated 22.05% of the study area as less vulnerable and 13.45% as very high vulnerable for all the folds. Out of four-folds, fold-4 showed the highest area (29.38% by BRT, 29.11% by SVM, and 29.09% by RF) as moderately vulnerable. Besides, fold-4 demarcated the least area (on average 12.91%) as low vulnerable. On average, 23.58%, 13.45%, 22.05%, 18.58% and 22.33% areas were identified as very high, high, moderate, low and very low VWR. The middle and southwestern parts of the study area showed a very high VWR due to the high salinity and elevated NO_3^- and As (Fig. 5). The ranges of causal parameters of VWR are given in Table 3. Initially, the vulnerability rating was determined based on a review of several literatures in this field. When 70% or more of the parameters (11 or more) were linked to ranges favourable for vulnerability, the mentioned samples were categorized as vulnerable, whereas those with less than 70 percent (less than 11) were classified as non-vulnerable. Then it was divided into two categories: vulnerable and non-vulnerable. It was then subdivided into four categories: vulnerable training, vulnerable validation, non-vulnerable training, and non-vulnerable validation.

3.4. Validation of the proposed models

The demarcated VWR zones can not beconfirmed without a proper statistical assessment (Ghosh et al., 2021). Therefore, findings of the VWR models were tested using the AUC, sensitivity, specificity, PPV, and NPV considering both datasets, i.e. training and testing. The AUC of different models is presented in Fig. 6. The results revealed that all the models have excellent prediction ability for mapping VWR.

Table 4shows the performance of the models based on statistical indices during both model training and testing for all folds. Different models showed the best performance for different folds. For example, the highest sensitivity for BRT, SVM, and RF during trainingwere 0.97, 0.99, 0.98 in fold-1, fold-2 and fold-4, respectively.In contrast, the lowest sensitivity for BRT, SVM, and RF were0.84, 0.84, and 0.77 in fold-1, fold-3 and fold-1, respectively, during testing.The highest specificity of BRT (0.80), SVM (0.97), and RF (0.97) were for fold-4, fold-2 and fold-2, respectively, during training. The lowest specificity of BRT (0.74), SVM (0.72), and RF (0.69) were for fold 1, fold-4 and fold 3, respectively, during testing.

Fig. 6 shows the AUC of ROC for three models fordifferentfolds. The highest AUC of BRT (0.96), SVM (0.99), and RF (0.98) were for fold-3, fold-2, and fold-2, and fold-4, respectively, during training, while the highest AUC of BRT (0.91), SVM (0.96), and RF (0.91) were for fold-4, fold-2, and fold-2, and fold-4, respectively, during testing (Table 4 and Fig. 6). The statistical metrics revealed the superior performance of RF, followed by the SVM and BRT models.

Fig. 7 demonstrates the correlation coefficients of the predicted vulnerability by the models for different folds. The correlation coefficient was found to vary for different models and folds. Fold-4 ($r = 0.78$), fold-2 ($r = 0.75$), and fold-1 ($r = 0.7$) produced the high correlation coefficient compared to other folds for BRT, SVM, and RF, respectively. Overall, the correlation coefficients for different models and folds were in the range of 0.65–7.8, indicating a strong association.

3.5. Relative importance of the causal factors

Table 5depicts the relative importance of the sixteen causal factors considered in this study. The best-performed model (RF) showed salinity as the most influencing factor for making the water resources of the study area vulnerable. Depth and NO_3^-were the second and third most influencing factors. Other most critical variables of VWR were F^- and As, followed by Na^+, Cl^-, and Mg^{2+} (Table 5).

4. Discussion

The geogenic mediated processes, such as saline water intrusion, low-lying topography, hydrogeological setting and its corresponding functions, including marine groundwater discharge, often cause coastal groundwater pollution (Vu et al., 2021). In addition to geogenic processes, other human activities such as solid plastic wastes, intense agricultural practices, land-use changes, population fringe growth, rapid urbanization have made the water resources of coastal Bangladesh vulnerable at different degrees and less suitable for human consumption. Currently, coastal groundwater pollution is increasing gradually due to varying types of waste, frequent coastal hazards, and disasters, making coastal people vulnerable to water resources. However, prediction and spatial modeling of VWR are yet lacking in the coastal plain of southern Bangladesh. Thus, a detailed VWR assessment of the whole coastal plain of Bangladesh is performed in this study. The study's findings can help to

Table 2
Areal coverage of vulnerable zones of water resources.

Models	Cross Validation	Area in Percentage (%)				
		Very Low	Low	Moderate	High	Very High
BRT	Fold 1	22.39	19.73	19.33	14.25	24.3
	Fold 2	21.19	26.33	16.37	11.28	24.83
	Fold 3	22.39	19.3	23.59	11.06	23.66
	Fold 4	25.91	12.21	29.38	8.39	24.11
SVR	Fold 1	22.99	19.39	19.95	14.19	23.48
	Fold 2	21.6	19.01	18.37	15.11	25.91
	Fold 3	20.39	22.27	17.22	16.11	24.01
	Fold 4	23.32	13.23	29.11	12.33	22.01
RF	Fold 1	23.66	18.99	19.99	14.99	22.37
	Fold 2	20.33	18.97	20.34	15.23	25.13
	Fold 3	17.12	20.17	25.25	16.23	21.23
	Fold 4	23.37	13.31	29.09	12.24	21.99

Table 3
Ranges of causal parameters favourable for water resource vulnerability.

Sl. No.	Parameters	Range favourable for vulnerability
1	Depth (m)	150.92–183.69
2	Temperature (°C)	29.01–37.59
3	pH	7.79–11.0
4	EC (μS/cm)	1247.34–50446.60
5	Salinity (ppt)	5.52–27.72
6	Ca^{2+} (mg/l)	0.01–104.91
7	Mg^{2+} (mg/l)	132.19–1260.63
8	Na^+ (mg/l)	1108.72–5217.89
9	K^+ (mg/l)	36.06–128.832
10	Cl^- (mg/l)	2038.65–18999.9
11	HCO_3^-(mg/l)	298.210–649.15
12	NO_3- (mg/l)	12.58–235.90
13	SO_4^{2-} (mg/l)	41.29–1539.52
14	PO_4^{2-} (mg/l)	27.20–914.25
15	F^- (mg/l)	0.92–12.15
16	As (μg/l)	6612–43386.6

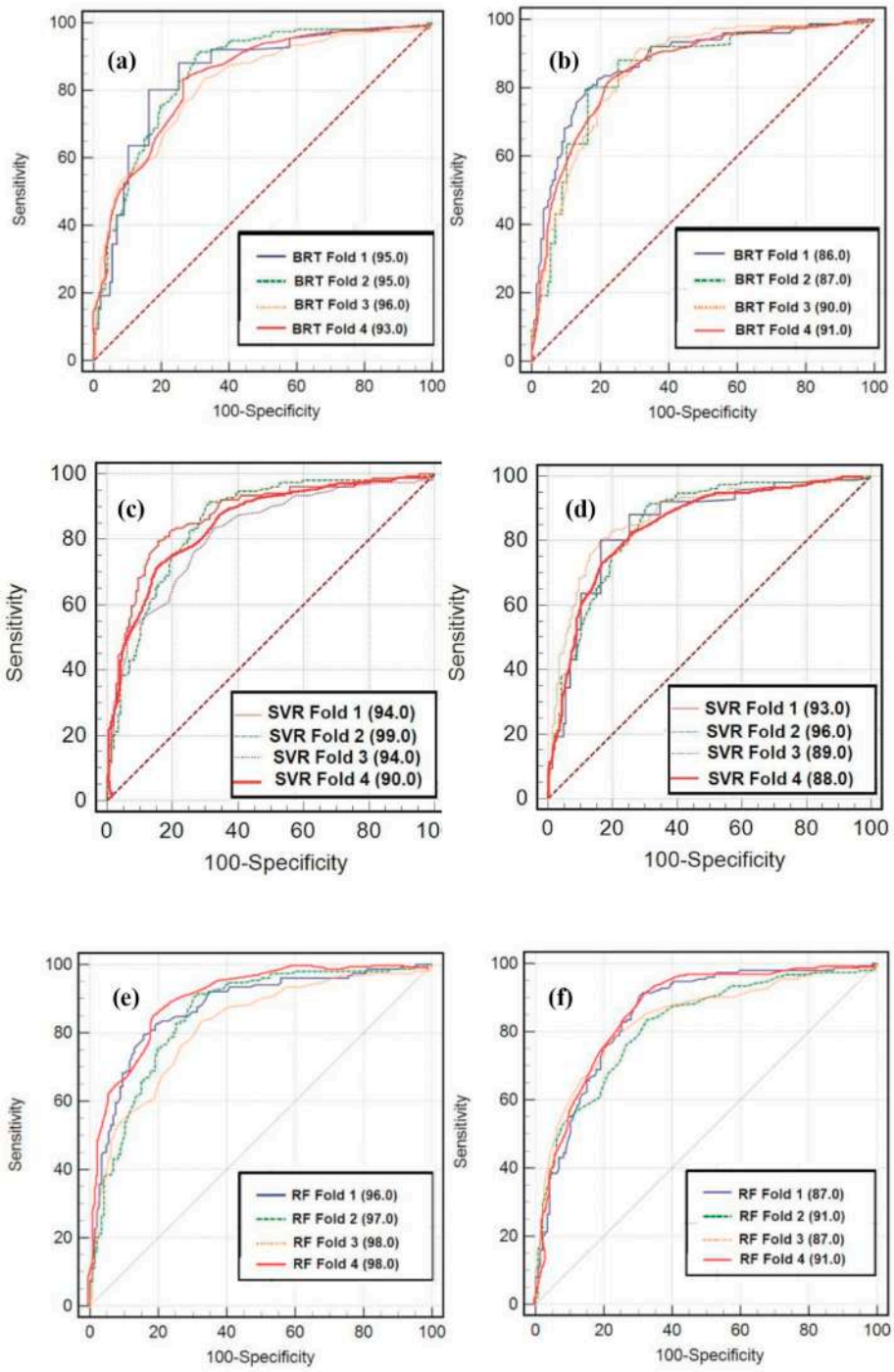

Fig. 6. ROC curve for training (a, c, and e) and testing (b, d, and f) datasets.

take necessary measures to attenuate the vulnerability level in the region.

Previous studies have primarily assessed and mapped the VWR of the study area using some conventional methods in terms of salinity and agricultural practice (Nahin et al., 2020). Heretofore, various types of methods have been constructed, and their performances have been compared in the existing literature to get the optimal tool for a particular area. However, these methods' prediction precision is questionable (Saha et al., 2021). Thus, it is essential to explore new models and tools to evaluate VWR (Barzegara et al., 2021). The present study addressed this issue by employing three data-driven models, namely BRT, SVM, and RF, to assess the VWR in Bangladesh's coastal region. So far, this is the first attempt in Bangladesh to investigate the VWR using machine learning algorithms. Building machine learning models need to identify the relevant causal factors to VWR. The study showed that salinity, depth, and NO_3^- have the most substantial impact on the VWR in the study area. Fluoride and As are also two critical factors for increasing the VWR in the coastal plain. On the other hand, Na^+, Cl^-, and Mg^{2+} contribute little to VWR in the study area.

The validation results revealed the good performance of RF, BRT and SVR in mapping VWR. Considering the performance during both training and testing, RF (AUC = 0.93, NPV = 0.89, PPV = 0.86, specificity = 0.85, sensitivity = 0.90) performed best in assessing the VWR in the study area, followed by SVM (AUC = 0.92, NPV = 0.89, PPV = 0.85,

Table 4
Validation of the predicted models.

Models	Stage	Parameters				
		Sensitivity	Specificity	PPV	NPV	AUC
BRT Fold 1	Train	0.97	0.79	0.82	0.97	0.95
	Test	0.84	0.74	0.77	0.82	0.86
BRT Fold 2	Train	0.95	0.70	0.76	0.94	0.95
	Test	0.87	0.76	0.79	0.85	0.87
BRT Fold 3	Train	0.95	0.72	0.77	0.94	0.96
	Test	0.88	0.76	0.80	0.79	0.90
BRT Fold 4	Train	0.93	0.80	0.82	0.91	0.93
	Test	0.89	0.77	0.77	0.89	0.91
SVR Fold 1	Train	0.87	0.84	0.85	0.87	0.94
	Test	0.86	0.84	0.82	0.87	0.93
SVR Fold 2	Train	0.99	0.97	0.97	0.99	0.99
	Test	0.91	0.87	0.86	0.91	0.96
SVR Fold 3	Train	0.89	0.92	0.91	0.92	0.94
	Test	0.84	0.78	0.82	0.80	0.89
SVR Fold 4	Train	0.88	0.77	0.78	0.88	0.90
	Test	0.93	0.72	0.80	0.89	0.88
RF Fold 1	Train	0.91	0.93	0.92	0.95	0.96
	Test	0.77	0.80	0.82	0.74	0.87
RF Fold 2	Train	0.93	0.97	0.96	0.98	0.97
	Test	0.88	0.86	0.88	0.86	0.91
RF Fold 3	Train	0.97	0.92	0.92	0.97	0.98
	Test	0.88	0.69	0.74	0.84	0.87
RF Fold 4	Train	0.98	0.94	0.93	0.98	0.98
	Test	0.90	0.71	0.75	0.87	0.91

specificity = 0.83, sensitivity = 0.89), and BRT (AUC = 0.91, NPV = 0.88, PPV = 0.78, specificity = 0.75, sensitivity = 0.91). The RF also showed the highest AUC (>0.96) for most of the folds. Besides, the Taylor diagram showed a better performance of the RF model ($r > 0.70$) compared to other models. On average, the RF ensemble model showed 1% more prediction accuracy than the BRT and SVR models. This is because of the RF model's ability to deal with complex datasets, low bias, and minimum generalization errors compared to other tree-ensemble models such as BRT. To the best of the author's knowledge, the BRT and SVR algorithms are first applied in spatial VWR modeling. The existing literature also indicates no groundwater modeling study so far using BRT and SVR models. The present study indicates the reliability and superiority of the RF model over the other two models in predicting VWR. Therefore, it can be recommended for VWR estimation in a setting similar to the present study area.

The resultant maps demonstrated a very high VWR in the southwestern and south-central parts of Bagerhat district and some parts of Gopalgong, Pirojpur, BorgunaPatuakhali, Shariatpur, Noakhali, and Lakshmipur districts. In addition, south-central regions, especially, Shariatpur, Noakhali, Lakshmipur, Barguna, and Jhalkati districts, are the VWR hotspots. Zahid et al. (2018) found similar high VWR across the coastal plain of Bangladesh. Overexploitation of the aquifer has triggered the water table dropand changes in the direction and movement of groundwater flow, playing a crucial role in coastal groundwater salinity. If saline water moved into the shallow and upper shallow aquifer during the wet season, the coastal plain's south-central and southwestern regions should be influenced by aquifer salinization. However, the water resources in these regions have high salinity content, elevated nitrate and arsenic concentration, and thus high VWR. This confirms the findings of Zahid et al. (2018) and Sarker et al. (2021). This outcome is also in line with previous studies (Islam et al., 2020a; Akber et al., 2020). Another possible reason is that higher temperatures during the dry season (>37 °C) cause a rise in evaporation from surface-water bodies existing in those regions, which reduces the interaction region for infiltration and the amount of surface water for irrigation. In addition, the temperature variation may force the biota to alter its biochemical inputs, causing a rise in water demand and creating a high VWR (Bhaskar et al., 2018). Similar findings have also been well documented in Greece, India, and Pakistan (Sahoo et al., 2016; Khan et al., 2019). The VWR in the southeastern and north-central parts, such as Cox'sBazar, Jessore, Khulna districts, were very low to low, while northwestern areas were moderately vulnerable. The water resources in these regions need suitable management plans for maintaining sustainability.

Table 6 shows the range of values of the analyzed physicochemical properties of groundwater collected from the coastal area of Bangladesh. It also compares the values with average world seawater composition and other studies done around the world on coastal groundwater resources to assess aquifer water quality due to seawater intrusion, groundwater salinization, and accumulation of pollutants in the coastal aquifers. The physicochemical parameters of groundwater in the coastal area of Bangladesh indicate that the groundwater is not suitable for drinking purposes. Furthermore, the analyzed groundwater parameters were compared to the drinking water standards of World Health Organization (WHO, 2011), drinking water standards for Bangladesh (DoE, 1997) and Indian standards (Indian Standard IS, 2012). The results showedthat most of the groundwater samples were unsuitable for drinking and domestic purposes due to the high concentrations of the chemical parameters (Table 6). Table 6 shows that the standard pH of drinking water is between 6.5 and 8.6 (WHO, 2011), 6.5 and 8.5 (DoE, 1997), 6.5 and 8.7 (Indian Standard IS, 2012) and 6.5 and 8.8 (USEPA, 1999). The pH of groundwater samples of the present study was in the range from 5.10 to 11.0, higher than the standard values. The range of electrical conductivity (EC) of the groundwater samples was 27.0–50446 μS/cm. The maximum permissible limit of EC is 400 μS/cm for drinking water (EU, 2011). The concentration of Na^+ was ranged from 0.20 to 5218 mg/L (Table 6), while the Na^+ in the drinking water standards of the department of environment (DoE, 1997) and WHO (WHO, 2011) is 200 mg/L. Though the lower value of Na^+ was less than the standard values, the Na + concentration in most of the samples was higher than the recommended values, indicating not suitable for human consumption. The measured Ca^{2+}, Mg^{2+}, and K^+ concentrations ranged from 0.01 to 902 mg/L, 0.35–1261 mg/L, and 0.22–1125 mg/L, respectively. Among the major anions, the average concentration of Cl^- and F^- in drinking water samples were ranged from 14 to 18999 mg/L and 0.00–14.99 mg/L, respectively. The results, as shown in the comparative table, indicate that coastal groundwater samples had a high level of Cl^- and F^-.

Several studies have been conducted to assess groundwater quality in coastal areas of India, Korea, Turkey, Malaysia, Tunisia, Greece, China, Oman, Canada, Morocco, Italy and other countries (Table 6). The studies also demonstrated that the coastal groundwater has high concentrations of chemical properties, especially Cl^-. The spatial distribution map of Cl^-(Figs. 3c and 4c) revealed high concentrations of Cl^-(>16999 and >18999 mg/L), indicating high ions in coastal aquifers, which may be due to mixing with seawater and over-pumping of groundwater resources (Van Camp et al., 2014). The concentration of NO_3^- in the present study was in the range of 0.20–235 mg/L. It is higher than the standard values of 10, 50, 45 and 10 mg/L (DoE, 1997; WHO, 2011, Indian Standard IS, 2012 and USEPA, 1999). It was also found high in other countries (Table 6). It indicates that NO_3^-contamination is a very common and serious problem in groundwater resources in coastal environments around the globe. The spatial distribution map of NO_3^- (Figs. 3k and 4k) showed that the groundwater samples from the coastal area had a high concentration, which may be due to the disposal of domestic sewage and agricultural activities like intensive use of fertilizers (Mahlknecht et al., 2017).

The study presents a promising technique to assist water managers and policymakers in protecting groundwater resources against severe pollution. Constructing precise VWR maps can lead to enhance water resources management and ecological sustainability. However, the present study has some potential limitations. Several factors such as excessive groundwater withdrawal, water flow direction or climate change can influence the groundwater resources. These factors were not considered in this study. Sensitivity assessment of data-driven models is needed to determine the uncertainties linked with causal VWR factors. A

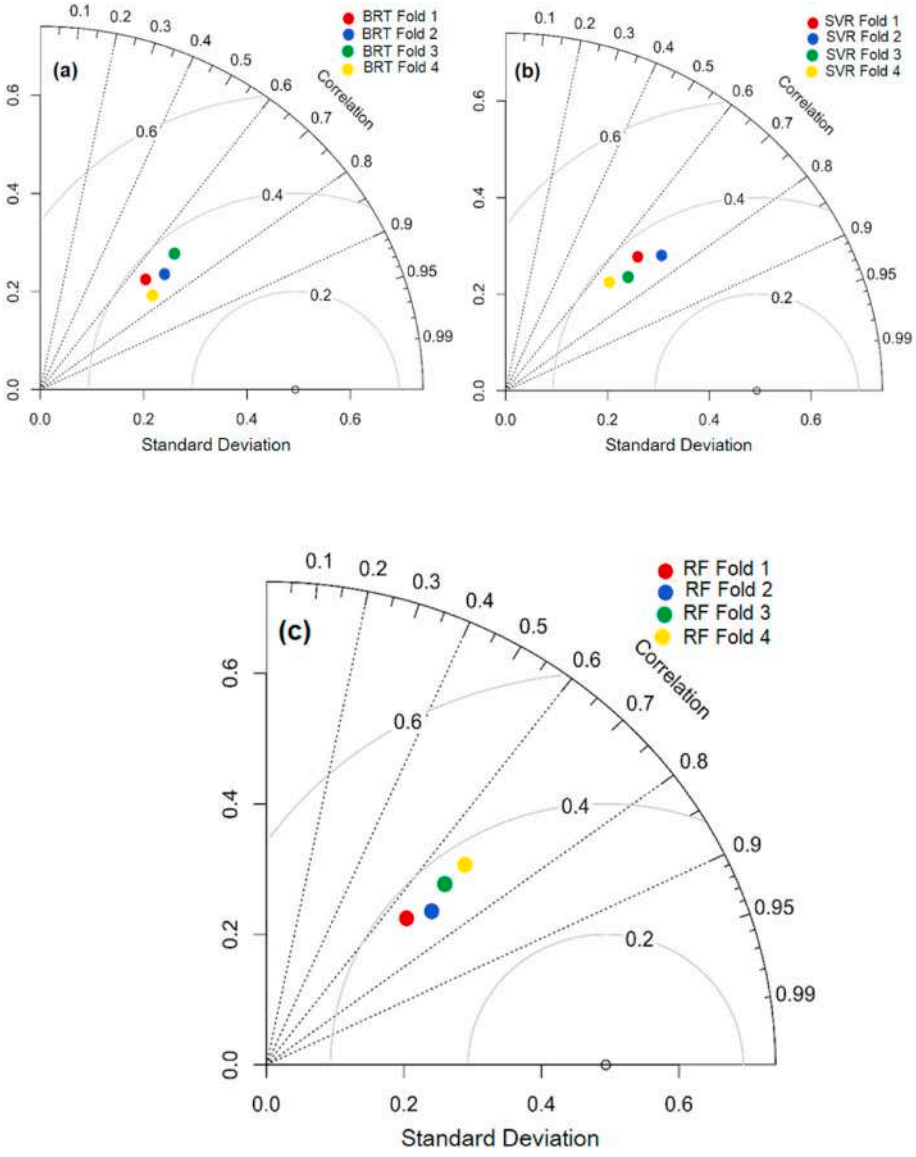

Fig. 7. Taylor diagram for BRT (a), SVR (b) and RF (c) model.

Table 5
Relative importance of the vulnerability causal parameters.

Factors	Fold 1	Fold 2	Fold 3	Fold 4
Depth	15.04	12.35	12.39	15.39
Temperature	7.23	6.33	3.89	8.37
pH	8.00	7.29	7.88	7.01
EC	8.59	12.55	9.99	3.58
Salinity	15.26	15.89	16.77	18.34
Ca^{2+}	9.33	8.29	8.11	8.03
Mg^{2+}	6.66	5.89	8.32	4.21
Na^{+}	5.22	4.99	4.97	4.07
K^{+}	8.36	8.33	8.09	8.16
Cl^{-}	5.23	5.00	5.19	4.92
HCO_3-	7.22	2.93	5.33	6.29
NO_3-	10.36	10.88	10.39	11.36
SO_4^{2-}	5.39	4.33	4.39	4.01
PO_4^{2-}	6.39	7.33	2.39	4.59
F^{-}	10.33	8.98	9.39	8.11
As	8.32	9.36	9.11	9.08

further investigation is required in this regard. Besides, the VWR was estimated without considering long-term temporal changes in the groundwater properties due to the unavailability of data.Future work should focus on using hybrid ensemble models integrated with a physical numerical model to improve the predictive performance of other vulnerability assessment models. Besides, other potential variables, including sea-level rise, aquifer properties, groundwater level, which can influence the VWR in coastal plains, should be considered in future studies to get more consistent VWR maps.

5. Conclusion

The aim of the present research was to construct a coupled novel framework approach using hydrochemical data, ensemble tree-based models, RF and BRT, with a classic model, SVR, through a k-fold CV approach for delineating the VWR zones in the coastal plain of Bangladesh. Our results demonstrated that the ensemble RF model performed superior to the other models. The accuracy of the RF model was 1% higher than the BRT and SVR models. This research on identifying VWR mapping using data-driven modeling have recognized 12.24–25.13% of areas as highly vulnerable for locating groundwater. The south-central and southwestern regions of the study area have high salinity, depth variation, and high nitrate content. These have made the regions highly vulnerable compared to other areas. This study also

Table 6
Comparison of the groundwater parameters concentrations of the coastal aquifer, Bangladesh with other studies around the world on coastal groundwater resources.

Place (Country)	pH	EC (μS/cm)	Ca^{2+}	Mg^{2+}	Na^{+}	K^{+}	Cl^{-}	F^{-}	HCO_3^{-}	NO_3^{-}	SO_4^{2-}	As	References
Coastal plains, Bangladesh	5.10–11.0	27–50446	0.01–902	0.35–1261	0.20–5218	0.22–1125	14–18999	0.0–14.99	24.40–649	0.20–235	0.0–1539.5	0.0–43.4	This study
Andhra Pradesh, India	7.0–8.7	410–15240	24–534	3.0–193	24–2436	1–412	28–4226	NA	144–819	2–184	12.0–3237		Latha and Rao (2012)
Bay of Bengal, India	7.4–8.9	220–23200	12–816	4.0–243	24–7590	2.0–803	57–2912	0.3–10.8*	37–1037	0.5–22	20–1037	0.01–3.2*	Rao et al. (2013)
Southwestern coast of Bangladesh	6.03	7135.66	289.5	340.5	1570	28.54	2940.7	11.85	NA	54.44	181.61	16.55	Rakib et al. (2020)
Coastal areas of Korea	NA	57–47538	0.47–3876	0.39–1249	4.13–8754	0.36–507	4.36–18793	NA	6.10–1525	NA	0.60–3232	NA	Park et al. (2012)
Bafra Plain, Turkey	6.30–8.40	1360–11490	22.04–194	37.67–350	220–2586	2.74–201	291–3474	NA	336–1769	NA	8.65–868	NA	Arslan et al. (2012)
Kapas, Malaysia	4.5–7.7	139–14085	1.1–254.4	1.0–496.2	3.0–1921	0.5–83.3	31–6198	NA	48.8–7320	0.4–10.6	1.0–500	NA	Kura et al. (2013)
Cap Bon Peninsula, Tunisia	6.9–7.9	1100–7800	119.9–551	20–440	73.3–828.1	3.2–150.5	146–2590.9	NA	158.6–610	0.8–590	57.1–1295	NA	Charfi et al. (2013)
Dongguan, South China	4.5–7.3	NA	6.5–170.7	0.4–102.7	2.1–999.6	2.1–86.9	7.1–1666	2.0*	1.8–748.1	0.7–128.1	0.2–493.4	0.05–4.44*	Huang et al. (2013)
Peloponnesus, Greece	6.23–8.03	417–50072	42.6–1151	3.05–1454	4.29–12883	0.05–616.8	7.0–19544	NA	203–1568	3.63–74.8	6.33–2460		Matiatos et al. (2014)
Al Musanaah, Oman	7.1–8.7	1560–31460	40–716	51–1186	157–4780	5.0–50	193–8236	NA	94–600	0.0–76	135–1720		Askri (2015)
Quebec, Canada	5.98–9.66	NA	1.90–290	0.71–57	1.2–860	0.53–26	0.39–1100	NA	17.1–572	NA	0.25–970	NA	Montcoudiol et al. (2015)
Oualidia, Morocco	7.0–7.9	200–7900	62.6–735.0	2.8–406.4	23.0–919.4	1.2–40.1	37.9–2010.7	NA	110.7–370.1	24.0–98.4	10.1–421.7	NA	Fadili et al. (2015)
Gaza, Palestine	6.9–8.3	597–30400	25–657	22.75–665	41.2–5400	1.3–155.5	77.7–10318	NA	101.5–1280	17.8–496	8.0–1604	NA	Abu-alnaeem et al. (2018)
Favignana, Italy	6.8–8.2	364–10920	25–356	3.0–293	33–1835	2.0–138	62–3177	NA	61–502	3.0–287	14–516	NA	Tiwari et al. (2019)
Standard[1]	6.5–8.5	1000	75	30–35	200	12	150–600	1		10	400	50	DoE (1997)
Standard[2]	6.5–8.6	400	100	150	200	NA	250	1.5	200	50	500	10	WHO, 2011, EU, 2011
Standard[3]	6.5–8.7	NA	75	30	NA	NA	250	1		45	200	10	Indian Standard IS, 2012
Standard[4]	6.5–8.8	NA	NA	50	NA	NA	250	2		10	500	10	USEPA 1999

Unit: Concentrations are in mg/L, except pH and EC (μS/cm).

recommends the inclusion of hydrochemical variables for enhancing the precision of the outcomes. Future studies should focus on the integrated hybrid ML models with physical numerical models to estimate VWR for diverse scenarios. Overall, this preliminary research is the first of its type in this coastal plain and will act as a foundation for devising a framework for the further comprehensive study of the groundwater vulnerability assessment over the whole coastal region of Bangladesh. The finding of this study can help the decision-makers and governmental and non-governmental agencies to manage and protect the water resources in the entire coastal plains of Bangladesh. Vulnerability serves as a hinder to sustainable development. The regionalization of the VWR assessment to rationally use and develop water resources strategies and planning for reducing vulnerability have practical implications. The technique developed in the present study can be a promising option for VWR assessment in any coastal region with similar aquifer characteristics and hydrogeologic conditions.

Funding

The authors acknowledge the Universiti Teknologi Malaysia research fellowship under grant number [Q.J130000.21A6.00P11]. The authors also extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through the Research Group Project under grant number [R.G.P.2/33/42].

CRediT authorship contribution statement

Abu Reza Towfiqul Islam: designed, planned, Conceptualization, Writing – original draft, the original manuscript, were involved in, Software, mapping, and proofreading during the manuscript drafting stage. **Subodh Chandra Pal:** designed, planned, Conceptualization, Writing – original draft, the original manuscript, were involved in, Software, mapping, and proofreading during the manuscript drafting stage. **Rabin Chakrabortty:** was involved in statistical, Formal analysis, interpretation, contributed to editing the manuscript, literature review, proofreading. **Abubakr M. Idris:** contributed to editing the manuscript, literature review, proofreading. **Roquia Salam:** was involved in statistical, Formal analysis, interpretation, were involved in, Software, mapping, proofreading during the manuscript drafting stage. **Md Saiful Islam:** contributed instrumental setup, Formal analysis, Validation, contributed to editing the manuscript, literature review, proofreading, were involved in, Software, mapping, proofreading during the manuscript drafting stage. **Anwar Zahid:** contributed instrumental setup, Formal analysis, Validation. **Shamsuddin Shahid:** contributed instrumental setup, Formal analysis, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research has been conducted under the Bangladesh Water Development Board project entitled "Establishment of monitoring network and mathematical model study to assess salinity intrusion in groundwater in the coastal area of Bangladesh due to climate change", led by Dr. Anwar Zahid. The project was financially supported by the Bangladesh Climate Change Trust Fund, Ministry of Forest, Environment and Climate change, Govt. of Bangladesh. We acknowledge the anonymous reviewers for improving the quality of the paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.

org/10.1016/j.jclepro.2022.130407.

References

Abu-alnaeem, M.F., Yusoff, I., Ng, T.F., Alias, Y., Raksmey, M., 2018. Assessment of groundwater salinity and quality in Gaza coastal aquifer, Gaza Strip, Palestine: an integrated statistical, geostatistical and hydrogeochemical approaches study. Sci. Total Environ. 615, 972–989.

Akber, M.A., Islam, M.A., Dutta, M., Billah, S.M., Islam, M.A., 2020. Nitrate contamination of water in dug wells and associated health risks of rural communities in southwest Bangladesh. Environ. Monit. Assess. 192 (3), 1–12.

Amiri, V., Rezaei, M., Sohrabi, N., 2014. Groundwater quality assessment using entropy weighted water quality index (EWQI) in Lenjanat. Iran. Environ. Earth Sci. (72), 3479–3490. https://doi.org/10.1007/s12665-014-3255-0, 2014.

Anandhi, A., Kannan, N., 2018. Vulnerability assessment of water resources – translating a theoretical concept to an operational framework using systems thinking approach in a changing climate: case study in Ogallala Aquifer. J. Hydrol. 557, 460–474.

APHA-AWWA-WEF, 2005. Standard Methods for the Examination of Water and Wastewater, twentieth ed. APHA, AWWA and WEF, Washington DC, USA.

Arslan, H., Cemek, B., Demir, Y., 2012. National scale evaluation of groundwater chemistry in Korea coastal aquifers: evidences of seawater intrusion. Environ. Earth Sci. 66, 707–718.

Askri, B., 2015. Hydrochemical processes regulating groundwater quality in the coastal plain of Al Musanaah, Sultanate of Oman. J. Afr. Earth Sci. 106, 87–98.

Band, S.S., Janizadeh, S., Pal, S.C., et al., 2020. Comparative analysis of artificial intelligence models for accurate estimation of groundwater nitrate concentration. Sensors 20, 5763. https://doi.org/10.3390/s20205763.

Barzegar, R., AsghariMoghaddam, A., Adamowski, J., Ozga-Zielinski, B., 2018. Multi-step water quality forecasting using a boosting ensemble multi-wavelet extreme learning machine model. Stoch. Environ. Res. Risk Assess. 32, 799–813.

Barzegara, R., Razzagh, S., Quilty, J., Adamowski, J., Pour, H.K., Booij, M.J., 2021. Improving GALDIT-based groundwater vulnerability predictive mapping using coupled resampling algorithms and machine learning models. J. Hydrol. 598, 126370.

Bhaskar, Aditi S., Hogan Dianna, M., Nimmo, John, Perkins, R., Kimberlie, S., 2018. Groundwater recharge amidst focused stormwater infiltration. Hydrol. Process. 32, 2058–2068.

Bhuiyan, M.A.H., Bodrud-Doza, M., Islam, A.R.M.T., Rakib, M.A., Rahman, M.S., Ramanathan, A.L., 2016. Assessment of groundwater quality of Lakshimpur district of Bangladesh using water quality indices, geostatistical methods, and multivariate analysis. Environ. Earth Sci. 75 (12), 1020.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Bui, D.T., Khosravi, K., Karimi, M., et al., 2020. Enhancing nitrate and strontium concentration prediction in groundwater by using new data mining algorithm. Sci. Total Environ. 715, 136836.

Buston, P.M., Elith, J., 2011. Determinants of reproductive success in dominant pairs of clownfish: a boosted regression tree analysis. J animal Ecology 80 (3), 528–538.

Charfi, S., Zouari, K., Feki, S., Mami, E., 2013. Study of variation in groundwater quality in a coastal aquifer in north-eastern Tunisia using multivariate factor analysis. Quat. Int. 302, 199–209.

Chen, W., Lei, X., Chakrabortty, R., Pal, S.C., Sahana, M., Janizad, S., 2021. Evaluation of different boosting ensemble machine learning models and novel deep learning and boosting framework for head-cut gully erosion susceptibility. J. Environ. Manag. 284, 112015.

Chowdhuri, I., Pal, S.C., Chakrabortty, R., Malik, S., Das, B., Roy, P., 2021. Torrential rainfall-induced landslide susceptibility assessment using machine learning and statistical methods of eastern Himalaya. Nat. Hazards 1–26.

Clemens, M., Khurelbaatar, G., Merz, R., Siebert, C., van Afferden, M., Roediger, T., 2020. Groundwater protection under water scarcity; from regional risk assessment tolocal wastewater treatment solutions in Jordan. Sci. Total Environ. 706 https://doi.org/10.1016/j.scitotenv.2019.136066.

da Costa, P.C.L., de Azevedo, A.R.G., da Silva, F.C., Cecchin, D., do Carmo, D.F., 2021. Rainwater treatment using an acrylic blanket as a filtering media. J. Clean. Prod. https://doi.org/10.1016/j.jclepro.2021.126964.

DoE (Department of Environment), 1997. The Environment Conservation Rules 1997. Government of the People's Republic of Bangladesh, Dhaka, Bangladesh.

Edmunds, M., Smedley, P., 2005. Fluoride in natural waters. In: Selinus, O., Alloway, B., Centeno, J., Finkelman, R., Fuge, R., Lindh, U., Smedley, P. (Eds.), Esssential of Medical Geology. Impacts of the Natural Environment on Public Health. Elsevier, Amsterdam, pp. 301–329.

Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. J. Anim. Ecol. 77, 802–813.

Erostate, M., Huneau, F., Garel, E., Ghiotti, S., Vystavna, Y., Garrido, M., Pasqualini, V., 2020. Groundwater dependent ecosystems in coastal Mediterranean regions: characterization, challenges and management for their protection. Water Res. 172, 115461. https://doi.org/10.1016/j.watres.2019.115461.

EU, 2011. European and National Drinking Water Quality standards.Northern Ireland Environment Agency. www.doeni.gov.uk/nica.

Fadili, A., Mehdi, K., Riss, J., Najib, S., Makan, A., Boutayab, K., 2015. Evaluation of groundwater mineralization processes and seawater intrusion extension in the coastal aquifer of Oualidia, Morocco: hydrochemical and geophysical approach. Arabian J. Geosci. 8 (10), 8567–8582.

Geisser, S., 1975. The predictive sample reuse method with applications. J. Am. Stat. Assoc. 70, 320–328.

Ghosh, R., Sutradhar, S., Mondal, P., et al., 2021. Application of DRASTIC model for assessing groundwater vulnerability: a study on Birbhum district, West Bengal, India. Model. Earth Syst. Environ. 7, 1225–1239. https://doi.org/10.1007/s40808-020-01047-7.

GoB, 2018. Climate Change and Agriculture in Bangladesh: Information Brief. Ministry of Environment and Forest. Government of the People's Republic of Bangladesh Retreived from. https://cmsdata.iucn.org/downloads/agriculture.pdf. (Accessed 20 January 2021).

Hein, T., Hauer, C., Schmid, M., et al., 2021. The coupled socio-ecohydrological evolution of river systems: towards an integrative perspective of river systems in the 21st century. Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2021.149619.

Huang, G., Sun, J., Zhang, Y., Chen, Z., Liu, F., 2013. Impact of anthropogenic and natural processes on the evolution of groundwater chemistry in a rapidly urbanized coastal area, South China. Sci. Total Environ. 463–464, 209–221.

IPCC, 2014. Climate Change 2014: Fifth Assessment Synthesis Report. Intergovernmental Panel on Climate Change, Geneva, Switzerland.

Indian Standard (IS), 2012. Bureau of Indian Standards Drinking Water Specifications. BIS 10500:2012. New Delhi, India.

Islam, A.R.M.T., Pal, S.C., Chowdhuri, I., Islam, R., Islam, M.S., Rahman, M.M., Zahid, A., Idris, A.M., 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.

Islam, A.R.M.T., Siddiqua, M.T., Zahid, A., Tasnim, S.S., Rahman, M.M., 2020b. Drinking appraisal of coastal groundwater in Bangladesh: an approach of multi-hazards towards water security and health safety. Chemosphere. https://doi.org/10.1016/j.chemosphere.2020.126933.

Islam, A.R.M.T., Mamun, A.A., Rahman, M.M., Zahid, A., 2020a. Simultaneous comparison of modified-integrated water quality and entropy weighted indices: implication for safe drinking water in the coastal region of Bangladesh. Ecol. Indicat. 113, 106229. https://doi.org/10.1016/j.ecolind.2020.106229.

Iqbal, J., Gorai, A.K., Tirkey, P., Pathak, G., 2012. Approaches to groundwater vulnerability to pollution: a literature review. Asian J. Water Environ. Pollut. 9 (1), 105–115.

Kabir, M.M., Akter, S., Ahmed, F.T., Mohinuzzaman, M., Didar-ul-Alam, M., Mostofa, K. M.G., Islam, A.R.M.T., Niloy, N.M., 2021. Salinity-induced fluorescent dissolved organic matter influence co- contamination, quality and risk to human health of tube well water, southeast coastal Bangladesh. Chemosphere 275, 130053.

Kaushik, J., Kumar, V., Garg, A.K., Dubey, P., Tripathi, K.M., Sonkar, S.K., 2021. Biomass derived functionalized graphene aerogel: a sustainable approach for the removal of multiple organic dyes and their mixtures. New J. Chem. 45 (20), 9073–9083.

Khalil, A., Almasri, M.N., McKee, M., Kaluarachchi, J.J., 2005. Applicability of statistical learning algorithms in groundwater quality modeling. Water Resour. Res. 41, W05010. https://doi.org/10.1029/2004WR003608.

Khan, M.S., Qadir, A., Javed, A., Mahmood, K., Amjad, M.R., Shahzad, S., 2019. Assessment of aquifer intrinsic vulnerability using GIS based Drastic model in Sialkot area, Pakistan. Int. J. Econ. Environ. Geol. 7 (1), 73–84.

Khan, Q., Liaqat, M.U., Mohamed, M.M., 2021. A comparative assessment of modeling groundwater vulnerability using DRASTIC method from GIS and a novel classification method using machine learning classifiers. Geocarto Int. 1–19.

Khosravi, K., Panahi, M., Tien Bui, D., 2018. Spatial prediction of groundwater spring potential mapping based on an adaptive neuro-fuzzy inference system and metaheuristic optimization. Hydrol. Earth Syst. Sci. 22, 4771–4792. https://doi.org/10.5194/hess-22-4771-2018.

Knoll, L., Breuer, L., Bach, M., 2019. Large scale prediction of groundwater nitrate concentrationsfromspatial data usingmachine learning. Sci. Total Environ. 668, 1317–1327.

Kura, N.U., Ramli, M.F., Sulaiman, W.N.A., Ibrahim, S., Aris, A.Z., Mustapha, A., 2013. Evaluation of factors influencing the groundwater chemistry in a small tropical island of Malaysia. Int. J. Environ. Res. Publ. Health 10 (5), 1861–1881.

Latha, P.W., Rao, K.N., 2012. An integrated approach to assess the quality of groundwater in a coastal aquifer of Andhra Pradesh, India. Environ. Earth Sci. 66, 2143–2169.

Liaw, R., Liang, E., Nishihara, R., Moritz, P., Gonzalez, J.E., Stoica, I., 2018. Tune: A Research Platform for Distributed Model Selection and Training. CoRR, abs/1807.05118, 2018.URL. http://arxiv.org/abs/1807.05118.

Mahlknecht, J., Merchan, D., Rosner, M., Meixner, A., Ledesma-Ruiz, R., 2017. Assessing seawater intrusion in an arid coastal aquifer under high anthropogenic influence using major constituents, Sr and B isotopes in groundwater. Sci. Total Environ. 587, 282–295.

Mallick, J., Talukder, S., Islam, A.R.M.T., 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.

Matiatos, I., Alexopoulos, A., Godelitsas, A., 2014. Multivariate statistical analysis of the hydrogeochemical and isotopic composition of the groundwater resources in northeastern Peloponnesus (Greece). Sci. Total Environ. 476, 577–590.

Melesse, A.M., Ahmad, S., McClain, M.E., Wang, X., Lim, Y.H., 2011. Suspended sediment load prediction of river systems: an artificial neural network approach. Agric. Water Manag. 98, 855–866. https://doi.org/10.1016/j.agwat.2010.12.012.

Montcoudiol, N., Molson, J., Lemieux, J.M., 2015. Groundwater geochemistry of the outaouais region (quebec, Canada): a regional-scale study. Hydrogeol. J. 23 (2), 377–396.

Nahin, K.T.K., Basak, R., Alam, R., 2020. Groundwater vulnerability assessment with DRASTIC index method in the salinity-affected southwest coastal region of

Bangladesh: a case study in BagerhatSadar, Fakirhat and Rampal. Earth Syst. Environ. 4 (1), 183–195.
Nolan, J.P., Soar, J., Cariou, A., et al., 2015. European resuscitation council and European society of intensive care medicine guidelines for post-resuscitation care 2015: section 5 of the European resuscitation council guidelines for resuscitation 2015. Resuscitation 95, 202–222.
NRC (National Research Council), 1993. Ground Water Vulnerability Assessment: Contamination Potential under Conditions of Uncertainty. National Academy Press, Washington DC.
OECD-Organization for Economic Co-operation and Development, 2011. OECD environmental outlook to 2050: the consequences of inaction. Available in: https://www.oecd.org/g20/topics/energy-environment-green-growth/oecdenvironmentaloutlookto2050theconsequencesofinaction.htm.
Ouedraogo, I., Defourny, P., Vanclooster, M., 2019. Application of randomforest regression and comparison of its performance to multiple linear regression inmodelinggroundwater nitrate concentration at the African continent scale. Hydrogeol. J. 27, 1081–1098. https://doi.org/10.1007/s10040-018-1900-5.
Pal, S.C., Arabameri, A., Blaschke, T., Chowdhuri, I., Saha, A., Chakrabortty, R., Lee, S., Band, S., 2020. Ensemble of machine-learning methods for predicting gully erosion susceptibility. Rem. Sens. 12, 3675.
Park, Y., Lee, J.Y., Kim, J.H., Song, S.H., 2012. National scale evaluation of groundwater chemistry in Korea coastal aquifers: evidences of seawater intrusion. Environ. Earth Sci. 66 (3), 707–718.
Pham, B.T., Khosravi, K., Prakash, I., 2017. Application and comparison of decision tree-based machine learning methods in landside susceptibility assessment at PauriGarhwal Area, Uttarakhand, India. Environ. Process. 4, 711–730.
Rakib, M.A., Sasaki, J., Matsuda, H., Quraishi, S.B., Mahmud, M.J., Bodrud-Doza, M., Ullah, A.K.M.A., Fatema, K.J., Newaz, M.A., Bhuiyan, M.A.H., 2020. Groundwater salinization and associated co-contamination risk increase severe drinking water vulnerabilities in the southwestern coast of Bangladesh. Chemosphere. https://doi.org/10.1016/j.chemosphere.2019.125646.
Rao, V.V.S.G., Rao, G.T., Surinaidu, L., Mahesh, J., Rao, S.T.M., Rao, B.M., 2013. Assessment of geochemical processes occurring in groundwaters in the coastal alluvial aquifer. Environ. Monit. Assess. https://doi.org/10.1007/s10661-013-3171-x.
Rajput, H., Goyal, R., Brighu, U., 2020. Modification and optimization of DRASTIC model for groundwater vulnerability and contamination risk assessment for Bhiwadi region of Rajasthan, India. Environ. Earth Sci. 79 https://doi.org/10.1007/s12665-020-8874-z.
Ransom, K.M., Nolan, B.T., Traum, J.A., Faunt, C.C., Bell, A.M., Gronberg, J.A.M., Wheeler, D.C., Rosecrans, C.Z., Jurgens, B., Schwarz, G.E., et al., 2017. A hybrid machine learning model to predict and visualize nitrate concentration throughout the central valley aquifer, California, USA. Sci. Total Environ. 601, 1160–1172.
Ravenscroft, P., McArthur, J.M., Hoque, M.A., 2013. Stable groundwater quality in deep aquifers of Southern Bangladesh: the case against sustainable abstraction. Sci. Total Environ. 454–455, 627–638.
Ridgeway, G., 2006. gbm: generalized boosted regression models. R Packag. Version 1, 55.
Roy, P., Chandra Pal, S., Chakrabortty, R., Chowdhuri, I., Malik, S., Das, B., 2020. Threats of climate and land use change on future flood susceptibility. J. Clean. Prod. 272, 122757. https://doi.org/10.1016/j.jclepro.2020.122757.
Saha, A., Pal, S.C., Arabameri, A., Blaschke, T., Panahi, S., Chowdhuri, I., Chakrabortty, R., Costache, R., Arora, A., 2021. Flood susceptibility assessment using novel ensemble of hyperpipes and support vector regression algorithms. Water 13, 241.
Saha, S., Saha, A., Hembram, T.K., Pradhan, B., Alamri, A.M., 2020. Evaluating the performance of individual and novel ensemble of machine learning and statistical models for landslide susceptibility assessment at rudraprayag district of Garhwal Himalaya. Appl. Sci. 10 (11), 3772.
Sahoo, S., Dhar, A., Kar, A., Chakraborty, D., 2016. Index- based groundwater vulnerability mapping using quantitative parameters. Environ. Earth Sci. 75, 522.
Sajedi-Hosseini, F., Malekian, A., Choubin, B., Rahmati, O., Cipullo, S., Coulon, F., Pradhan, B., 2018. A novel machine learning-based approach for the risk assessment of nitrate groundwater contamination. Sci. Total Environ. 644, 954–962.
Sarker, M.M.R., Van Camp, M., Hossain, D., Islam, M., Ahmed, N., Karim, M.M., Walraevens, K., 2021. Groundwater salinization and freshening processes in coastal aquifers from southwest Bangladesh. Sci. Total Environ. 779, 146339.
Schapire, R.E., 2003. The boosting approach to machine learning: an overview. Nonlinear Estim. Classif. 149–171.
Sheikh Khozani, Z., Hosseinjanzadeh, H., Wan Mohtar, W.H.M., 2019. Shear force estimation in rough boundaries using SVR method. Appl. Water Sci. 9, 186. https://doi.org/10.1007/s13201-019-1056-z.
Siddique, M.A.B., Khan, R., Islam, A.R.M.T., et al., 2021. Quality assessment of freshwaters from a coastal city of southern Bangladesh: irrigation feasibility and preliminary health risks appraisal. Environ. Nanotechnol. Monit. Manag. 16, 100512. https://doi.org/10.1016/j.enmm.2021.100524.
Tiwari, A.K., Pisciotta, A., Maio, M.D., 2019. Evaluation of groundwater salinization and pollution level on Favignana Island. Italy. Environ. Pollut. 249, 969–981.
Talukdar, S., Eibek, K.U., Akhter, S., Ziaul, S.K., Islam, A.R.M.T., Mallick, J., 2021. 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 126, 107612. https://doi.org/10.1016/j.ecolind.2021.107612.
Tasnuva, A., Hossain, R., Salam, R., Islam, A.R.M.T., et al., 2020. Employing social vulnerability index to assess household social vulnerability of natural hazards: an evidence from southwest coastal Bangladesh. Environ. Dev. Sustain. https://doi.org/10.1007/s10668-020-01054-9.
Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res. Atmos. 106, 7183–7192. https://doi.org/10.1029/2000JD900719.
Thapa, R., Gupta, S., Guin, S., Kaur, H., 2018. Sensitivity analysis and mapping the potential groundwater vulnerability zones in Birbhum district, India: a comparative approach between vulnerability models. Water Sci. 32, 44–66.
Ukpai, S.N., Okogbue, C.O., 2017. Geophysical, geochemical and hydrological analyses of water-resource vulnerability to salinization: case of the Uburu-Okposi salt lakes and environs, southeast Nigeria. Hydrogeol. J. 25 (7), 1997–2014.
Van Camp, M., Mtoni, Y., Mjemah, I.C., Bakundukize, C., Walraevens, K., 2014. Investigating seawater intrusion due to groundwater pumping with schematic model simulations: the example of the Dar es Salaam coastal aquifer in Tanzania. J. Afr. Earth Sci. 96, 71–78.
Vapnik, V., Guyon, I., Hastie, T., 1995. Support vector machines. Mach. Learn. 20, 273–297.
Vrba, J., Zaporozec, A., 1994. Guidebook on Mapping Groundwater Vulnerability, 16. International Contribution to Hydrogeology, Hannover, p. 131.
Vu, T.D., Ni, C.F., Li, W.C., Truong, M.H., Hsu, S.M., 2021. Predictions of groundwater vulnerability and sustainability by an integrated index-overlay method and physical-based numerical model. J. Hydrol. 596, 126082.
Wang, J., Li, L., Niu, D., Tan, Z., 2012. An annual load forecasting model based on support vector regression with differential evolution algorithm. Appl. Energy 94, 65–70.
Wang, W., Mwiathi, N.F., Li, C., Luo, W., Zhang, X., An, Y., Zhang, M., Gong, P., Liu, J., Gao, X., 2022. Assessment of shallow aquifer vulnerability to fluoride contamination usingmodified AHP-DRASTICH model as a tool for effective groundwatermanagement, a case study in Yuncheng Basin, China. Chemosphere. https://doi.org/10.1016/j.chemosphere.2021.131601.
WHO, 2011. Guidelines for Drinking-Water Quality. Recommendations, third ed.1. WHO, Geneva.
Zahid, A., Hossain, A.F.M.A., Ali, M.H., Islam, K., Abbassi, S.U., 2018. Monitoring the coastal groundwater of Bangladesh. In: Mukherjee, A. (Ed.), Groundwater of South Asia. Springer Hydrogeology, pp. 431–451. https://doi.org/10.1007/978-981-10-3889-1_26.
USEPA 1999. Screening Level Ecological Risks Assessment Protocol for Hazardous Waste Combustion Facilities. Appendix E: Toxicity Reference Values. EPA 530-D99-001C, vol.3. Available from: http://www.epa.gov/epaoswer/hazwaste/combust/eco-risk/voume3/appx-e.pdf.