



OPEN

Hydro-chemical based assessment of groundwater vulnerability in the Holocene multi-aquifers of Ganges delta

Asish Saha¹, Subodh Chandra Pal¹✉, Abu Reza Md. Towfiqul Islam^{2,3}, Aznarul Islam⁴, Edris Alam^{5,6} & Md. Kamrul Islam⁷

Determining the degree of high groundwater arsenic (As) and fluoride (F⁻) risk is crucial for successful groundwater management and protection of public health, as elevated contamination in groundwater poses a risk to the environment and human health. It is a fact that several non-point sources of pollutants contaminate the groundwater of the multi-aquifers of the Ganges delta. This study used logistic regression (LR), random forest (RF) and artificial neural network (ANN) machine learning algorithm to evaluate groundwater vulnerability in the Holocene multi-layered aquifers of Ganges delta, which is part of the Indo-Bangladesh region. Fifteen hydro-chemical data were used for modelling purposes and sophisticated statistical tests were carried out to check the dataset regarding their dependent relationships. ANN performed best with an AUC of 0.902 in the validation dataset and prepared a groundwater vulnerability map accordingly. The spatial distribution of the vulnerability map indicates that eastern and some isolated south-eastern and central middle portions are very vulnerable in terms of As and F⁻ concentration. The overall prediction demonstrates that 29% of the areal coverage of the Ganges delta is very vulnerable to As and F⁻ contents. Finally, this study discusses major contamination categories, rising security issues, and problems related to groundwater quality globally. Henceforth, groundwater quality monitoring must be significantly improved to successfully detect and reduce hazards to groundwater from past, present, and future contamination.

Groundwater is a crucial source of drinking water, and its availability is essential for economic growth in urban and rural areas worldwide^{1,2}. Groundwater is less vulnerable to contamination and pollution than surface water and is widely used for domestic purposes^{3,4}. Due to its high percentage, reduced sensitivity to pollution, and large storage capacity, groundwater is more important than surface water at a socioeconomic level worldwide. Groundwater undergoes a natural filtration process that removes bacteria and odors, making it suitable for drinking⁵. Groundwater has many advantages, including meeting water supply needs for industrial, agricultural, and other sectors. In many parts of the world, groundwater is the primary source of fresh water with 50% of portable water demands being met by groundwater, 40% of which is used for industry, and the remaining portion used for irrigation⁶. As the world's population grows, its dependence on groundwater also increases, with 33% of people depending on it to meet their daily needs^{7,8}. Unfortunately, more groundwater is being consumed than replenished or recharged, which stresses on the availability of this precious natural resource. As a result, groundwater overuse has led to declining water tables, declining water quality, and ongoing frequent land subsidence activities⁹.

Despite all the advantages of groundwater, many nations, especially developing countries like India and Bangladesh, are quickly experiencing a crisis of diminishing groundwater quality due to misuse and contamination¹⁰. Groundwater contamination can result from both natural and human causes. However, the quantity and

¹Department of Geography, The University of Burdwan, Purba Bardhaman, West Bengal 713104, India. ²Department of Disaster Management, Begum Rokeya University, Rangpur 5400, Bangladesh. ³Department of Development Studies, Daffodil International University, Dhaka 1216, Bangladesh. ⁴Department of Geography, Aliah University, 17 Gorachand Road, Kolkata 700014, India. ⁵Faculty of Resilience, Rabdan Academy, 22401 Abu Dhabi, United Arab Emirates. ⁶Department of Geography and Environmental Studies, University of Chittagong, Chittagong 4331, Bangladesh. ⁷Department of Civil and Environmental Engineering College of Engineering, King Faisal University, 31982 AlAhsa, Saudi Arabia. ✉email: geo.subodh@gmail.com

groundwater quality are highly vulnerable today. Anthropogenic activities have accelerated the rate at which the quality of groundwater is declining. Unplanned land-use activities owing to industrialization and subsequent urbanization have led to rising groundwater contamination in recent decades¹¹. Urbanization increases impervious surfaces, worsens ephemeral runoff, increases flood risk, and reduces subsequent groundwater recharge. In addition, saltwater intrusion exacerbates the situation in coastal areas, severely threatening the city's water supply and lowering the living standards of residential homes¹². Human activities such as irrigation and climate change are currently impacting groundwater quality and increasing its sensitivity to contamination on a broader scale¹³. Chemical fertilizers are exacerbating the critical issue of nitrate poisoning of aquifers¹⁴.

One of the largest natural groundwater catastrophes for humanity has been reported to be arsenic (As) pollution in the groundwater. A study revealed that only five Asian nations, namely “Taiwan, China, India, Bangladesh, and Thailand”, were acknowledged as having groundwater contamination due to As in the late twentieth century¹⁵. At least 100,000 individuals in these impacted nations are exposed to As poisoning through their drinking water. In India, several places in the Brahmaputra and Ganges River's floodplain have been affected by groundwater contamination with arsenic at levels higher than the permitted limit of 10 µg/L¹⁶. Currently, Bangladesh and India have the highest number of As contaminated areas and associated health issues¹⁷. Specifically, the Ganges delta of Indo-Bangladesh region is highly affected by groundwater contamination due to As^{14,18,19}. These floodplains are made of recent alluvial aquifers from the Holocene period that originated in the Himalayan region²⁰. Consequently, people in these afflicted areas have been regularly exposed to drinking water from hand tube wells contaminated with arsenic. In Bangladesh, using contaminated surface water resources, such as, ponds, rivers, and shallow dug-wells has led to water-borne illnesses like cholera, diarrhea, and dysentery²¹. Recent estimates suggest that 5 million people in West Bengal's North 24 Paraganas district consume water with an arsenic concentration of more than 50 µg/l and As rich causing approximately 50,000 people to develop skin sores in West Bengal. The population impacted in the geographic area of the issue are alarmingly growing each year²². Moreover, As has also entered the food chain through rice (paddy) production in the Indo-Gangetic plains via irrigation water carrying As²³.

From a methodological viewpoint, statistical, machine learning (ML), and artificial intelligence have been utilized to evaluate groundwater vulnerability globally. ML algorithms offer several advantages over statistical methods, as they can efficiently analyze large datasets^{24,25}. Additionally, geospatial approaches provide quick, efficient spatial, temporal, and spectral analysis of data over a wide area²⁶. As a result, numerous researchers have combined geospatial technology with ML algorithms, such as the deep learning network used by Elzain et al.²⁷ in south Korea, the BRT model used by²⁸ in Iran, RF used by Pal et al.¹⁴ in coastal areas of West Bengal, and the Bayesian model averaging (BMA) used by Gharekhani et al.²⁹ in West Azerbaijan, Iran, to assess groundwater vulnerability.

Considering the ongoing phenomena related to groundwater resources worldwide, particularly in the Ganges delta, very few studies have been undertaken that couple hydrochemical factors with ML algorithms^{30–32}. Literature review on groundwater vulnerability highlights that the Ganga–Brahmaputra delta in the Indo-Bangladesh region stands out as a significant area globally affected by arsenic contamination^{17,33}. The widespread use of tube wells for water supply in the Ganges Delta is a critical concern, leading to severe arsenic poisoning. In addition to arsenic, some regions in India also face challenges with elevated fluoride levels in groundwater. The Ganges Delta, marked by high population density and robust agricultural activity, necessitates sustainable water resource management to ensure optimal utilization. Hence, our study focuses on assessing groundwater vulnerability in the Ganges delta, emphasizing the urgency of effective water resource management in this crucial region. Therefore, researching groundwater vulnerability in this area is crucial for managing groundwater effectively and making it safe for consumption. In this regard, the presents study uses geospatial techniques and ML algorithms, including the LR, RF, and ANN models, to evaluate groundwater vulnerability in the study region. The distinctive aspect of this research resides in integrating statistical, ML, and neural network algorithms with hydrochemical factors. This fusion aims to comprehend the fluctuations in modeling outcomes and their corresponding spatial distribution in such a vast region. Furthermore, quality assessment for irrigation water in this study region has been assessed using USSS and Wilcox's diagram. This study has distinctive contribution in its novel outcomes and insights, which contribute optimal perspectives to the current body of existing literatures. Furthermore, the research insight into regional disparities, illuminating differences or distinctions within a specific geographical area. The outcomes of this study will be helpful to environmentalists and policy-makers in planning for the local people regarding the safe consumption of water resources.

Study area

The Ganges and Brahmaputra delta, known as the Ganges delta, is one of the mega-deltas in the world, covering an area of approximately 105,000 km². It consists of Bangladesh and parts of India's state of West Bengal, formed by sedimentation of the Ganga, Meghna and Brahmaputra rivers at the Bay of Bengal during the late Holocene to recent times²⁰. The delta stretches from 21° 10' 42" to 24° 50' 39" N latitude and 87° 30' 21" to 91° 26' 46" E longitude (Fig. 1) and has a shoreline of nearly 350 km along the Bay of Bengal. The Ganges delta has been divided into three parts from a geological perspective i.e., “Moribund delta, Active delta and Mature delta”³⁴. The delta's stratigraphic section shows alternating sand-dominated and fine-grained phases with intricate interfingerings between them³³. This delta enclosed by “Precambrian crystalline rocks” to the north and west and the “Assam-Arakan Neogene fold belt” to the east, signifies a comprehensive sedimentation history during the late Quaternary period³⁵. Literatures indicates that numerous elevated terraces from the Pleistocene era are present both within and along the periphery of its alluvial plain³⁶. Present evidence considering remote sensing supporting neotectonics activities in the Gangetic plain³⁷. Salinity has impacted aquifers in the coastal regions of Bangladesh, reaching depths of up to 350 m, furthermore the salinity levels in the upper aquifers of the coastal

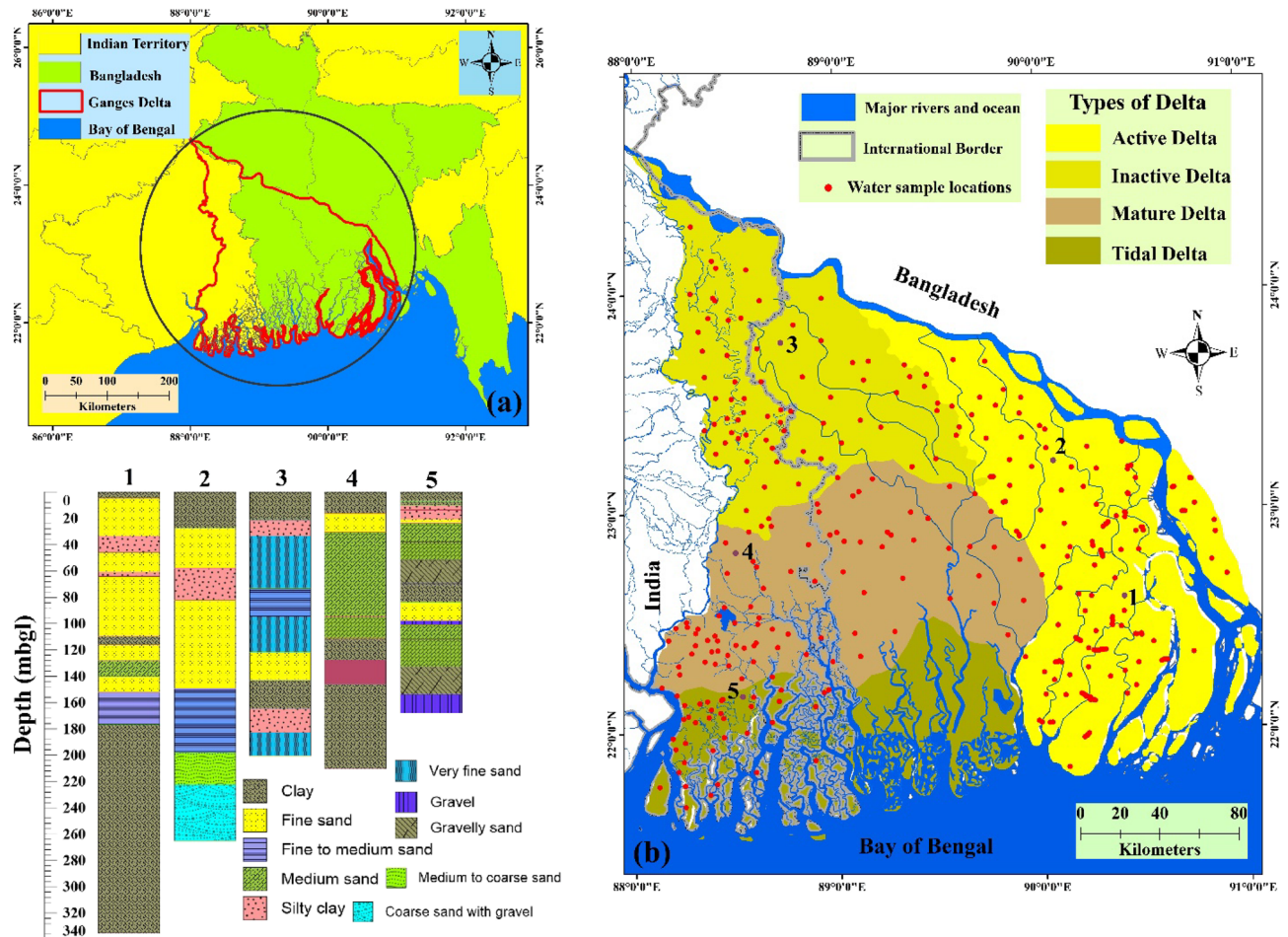


Figure 1. Details about the study area: (a) Ganges delta in a transnational boundary of Indo-Bangladesh region, (b) Ganges delta and its morphological types and (c) litholog profile of some selected points (this map was generated using ArcGIS, version: 10.3.1, www.esri.com/arcgis).

region, reaching depths of 200–250 m, demonstrate notable fluctuations and experience abrupt changes over short distances³⁸. The monsoon season (June–October) accounts for more than 80% of the annual rainfall, which ranges from 1500 to 2000 mm³⁹. During the monsoon months, high rainfall and frequent tropical cyclones cause catastrophic flooding and saltwater intrusion in the land areas. The minimum seasonal temperature of the region varies from 12 to 24 °C, and the maximum ranges from 25 to 35 °C. The area has the largest population density compared to other deltaic regions due to the high soil fertility³⁰. The Sundarbans, the world's largest mangrove forest, covers the southernmost part of this deltaic region also known as the Sunderban delta. Borehole data indicates that sediment primarily consists of sand and clay types.

Materials and methods

Methodology

In this study, the following methodological steps have been followed to fulfill the current research objectives:

- In the initial stage, 352 water samples were collected from the existing tube-wells in the field to assess different hydro-geochemical properties. Additionally, 352 non-sample points were created for modelling purposes. Furthermore, it is necessary to divide the dataset into train and test to assess how well machine learning models are performed. Where, train dataset is used to fit the model and the test dataset is used for validation of the respective model. The entire dataset was split into two categories in a “70:30 ratio for training and validation” of the respective models.
- A total of fifteen hydro-geochemical parameters were identified for modelling groundwater vulnerability. These parameters are “Depth (m), pH, EC ($\mu\text{S}/\text{cm}$), Salinity (ppt), Ca^{2+} (mg/l), Mg^{2+} (mg/l), Na^+ (mg/l), K^+ (mg/l), Cl^- (mg/l), HCO_3^- (mg/l), NO_3^- (mg/l), SO_4^{2-} (mg/l), PO_4^{2-} (mg/l), F^- (mg/l), As ($\mu\text{g}/\text{l}$)”.
- Statistical analysis, including “Pearson’s correlation coefficient, principal component analysis (PCA) and multicollinearity (MC)” test, was conducted to understand the nature of data.
- Statistical, ML and neural network algorithms i.e., “logistic regression (LR), random forest (RF) and artificial neural network (ANN)” were used for groundwater vulnerability assessment.

- Statistical evaluation metrics, such as “sensitivity, specificity, AUC-ROC, F score, Kappa coefficient, and graphical measures such as the Taylor diagram” were used to optimize the assessment of modelling output.
- “USSL and Wilcox’s diagram” used to assess groundwater quality.

The following sub-section described in details regarding the methods used in this study.

Sampling and inventory dataset

Field-based water sample collection was the primary task to prepare several hydro-chemical parameters for assessing groundwater vulnerability. In this regard, a “random stratified” sampling method was used to collect water sample across the study region. A total of 352 water samples were collected to prepare the inventory dataset (Fig. 1). Standard procedures were followed during the collection of water samples. Sampling was done by running wells for 5 min as it removes the stagnant water from bore wells as well as hand pumps. The sample tube well was kept pumping until the pH and EC achieved stable conditions. Two independent (dry and clean) sample kits were used each with its own collection methods and safety measures, to keep the water samples that were taken. In order to transport each water sample from the field to the lab and keep it at 4 °C, we stored it in a water sample kit during sample collection. Measurements were made to analyze the groundwater samples obtained both on-site and off-site. The analyzed samples were split into two categories on the ArcGIS 10.4.1 platform based on a ratio of 70:30. The sample was used for training (70%) while the other was used for validation (30%).

In our current research, we opted for dry season (March–early June) data to model and map groundwater vulnerability in this susceptible region, excluding wet season data. Existing literature indicates a prevalent use of dry season data in studies related to arsenic-induced vulnerability studies⁴⁰, as it is deemed more suitable for assessing vulnerability to arsenic-related risks. In the wet season, groundwater contamination occurs through the percolation and infiltration of surface water, facilitated by ample rainfall. This leads to the transfer of various particles, metals, and ions from surface water bodies to groundwater, resulting in temporary water contamination, which is not ideal for assessing water-related health hazards. In contrast, during the dry season, water levels remain normal, and there is no risk of water contamination through surface metals or other substances. Therefore, based on these considerations, we have exclusively utilized dry season data in our study.

MC test

To ensure the accuracy of the model’s output, it is crucial to select appropriate parameters for any vulnerability assessment. To achieve this, MC analysis is one of the most important techniques. Correlation analysis has shown that a link between two or more input variables can create deviations. “Tolerance (TOL) and Variance Inflation Factor (VIF)” are two statistical measures often used to test multi-collinearity among distinct components. The predictor variables have a high degree of multicollinearity when the “TOL value is < 0.10 and the VIF value is > 5”. If the MC result exceeds this limit, the highly correlated factors are not suitable for modelling purposes and should be removed from the dataset; otherwise, the output result will not be optimal. The equations for TOL and VIF are presented below:

$$\text{TOL} = 1 - R_j^2 \quad (1)$$

$$\text{VIF} = \frac{1}{\text{TOL}} \quad (2)$$

where R_j^2 is the R-squared value of regression using the j on all other variables regression model.

Adopted methods for groundwater vulnerability modelling

LR

One can create a multivariate regression relationship between a dependent variable and several independent factors using LR. LR is a multivariate analysis model that can be used to forecast the existence or absence of a characteristic or result based on the values of many response variables. Many studies used LR as a standard or conventional way to verify the effectiveness of a new algorithm in vulnerability studies. The benefit of LR is that, unlike traditional linear regression, where the variables must all have normal distributions, it can use any combination of continuous and discrete variables as well as appropriate link functions⁴¹. The challenge in conducting vulnerability analysis using a LR model is choosing the appropriate sample size for the dependent and independent variables⁴². The components in multi-regression analysis must be numerical, and the variables in discriminant analysis, a related statistical model, should have a normal distribution. After converting the dependent variable into a logit variable, the LR procedure uses maximum likelihood estimation⁴³. This is how LR calculates the likelihood of a specific event occurring⁴⁴. The fundamental idea behind LR is investigating a problem in which a result assessed using dichotomous variables i.e., true or false (0 and 1) is determined based on a single or a series of independent factors⁴⁵. The LR can expressed by the following equation:

$$f(z) = \frac{1}{1 + e^{-z}} \quad (3)$$

where z indicates a linear combination of a constant and the independent variables’ product, and their corresponding coefficients. The value of z varies from $-\infty$ to ∞ , subsequently $f(z)$ ranges from 0 to 1”:

$$z = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \quad (4)$$

where α indicates constant; $\beta_1, \beta_2, \dots, \beta_n$ represent the coefficients and X_1, X_2, \dots, X_n are the independent variables”.

RF

The RF model is a reliable AI method for classifying various natural hazards, including groundwater vulnerability. Breiman⁴⁶ proposed a potent ensemble-learning method called random forest, which is one of the most widely used classifier ensemble techniques for feature selection, regression, and classification applications. RF is a tree-based ensemble learning technique that builds several decision trees while constructing models. Each tree structure in the ensemble model uses the original input data to train a bootstrapped sample⁴⁷. Decision trees use a collection of binary rules to select a target variable. The data used to train the model comprises the target variable being predicted and a set of predictor variables. Using the predictor variables, the decision tree divides the data into homogenous datasets based on the target variable. The programme then assesses each predictor variable's ability to categorize the predicted value into the two groups. The splitting process continues until there are no more splits to be made⁴⁸. RF prediction is viewed as the unweighted majority of class votes when solving classification issues. The bagging approach is used to select random samples of variables as part of the training dataset for model calibration⁴⁹. The algorithm for RF is expressed as follows:

$$h(x, i_k), k = 1, 2, \dots, n \quad (5)$$

where i_k represents flood occurrence conditioning factors; 1, 2, ..., n are input vector x .

In a RF the general errors can be defined as follows:

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

where x and y indicate the different flood occurrence conditioning factors, and mg represents the margin function. Again, margin function” can be described as follows

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

ANN

The ANN is a computational method that can obtain, display, and compute mapping from one multivariate data space to another. The objective of the ANN model is to provide a technique for forecasting results from inputs that have not been used in the modelling process⁵⁰. An artificial neural network is trained using a series of examples of related input and output values. The goal of an artificial neural network is to create a model of the data-generation process in order to generalize and predict outcomes from inputs that it has never seen before. Back-propagation learning is the neural network approach that is most often utilized in the ANN model⁵¹. This neural network learning technique has three levels: an input layer, hidden layers, and an output layer. The network is trained using the back-propagation technique until a predetermined minimal error between the network's desired and actual output values is reached. When training is complete, the network is utilized as a feed-forward structure to provide a classification for the entire database (Paola and Schowengerdt⁵²). The ANN assigns each input element a specific weight, multiplies the results, adds them up, and then uses a nonlinear transfer function to construct the outcomes. The back propagation of the ANN model is expressed by the following equations:

$$net_j^l(t) = \sum_{i=0}^p (y_i^{l-1}(t) w_{ji}^l(t)) \quad (8)$$

The net input of j th neuron of layer l and I iteration

$$y_j^l(t) = f(net_j^l(t)) \quad (9)$$

$$f(net) = \frac{1}{1 + e^{-net}} \quad (10)$$

$$e_j(t) = c_j(t) - a_j(t) \quad (11)$$

$$\delta_j^l(t) = e_j^l(t) a_j(t) [1 - a_j(t)] \quad (12)$$

δ Factor for neuron j th in the output layer l th

$$\delta_j^l(t) = y_j^l(t) [1 - y_j^l(t)] \sum \delta_j^l(t) w_{kj}^{(l+1)}(t) \quad (13)$$

δ factor for neuron j th in the hidden layer l th

$$w_{ji}^l(t+1) = w_{ji}^l(t) + \alpha [w_{ji}^l(t) - w_{ji}^l(t-1)] + n \delta_j^{(l)}(t) y_j^{(l-1)}(t) \quad (14)$$

where α is the momentum rate and n is the learning rate within this model.

Selected evaluation measures

Evaluating a model's performance, which establishes whether it is relevant or not, is one of the key goals of model comparison. In the geoscientific discipline, assessment metrics for applied models are crucial to estimating their best-case performance in making predictions, especially for modelling approaches based on machine learning. Henceforth, several evaluation measures have been used by many researchers in different fields of study to optimally assess the modelling output^{14,24,53,54}. After a rigorous literature survey, five prevalent evaluations metrics i.e., "sensitivity, specificity, PPV, NPV, ROC-AUC, Kappa-coefficient and F-score", were selected for this study. Alongside, the Taylor diagram is also applied in this study, which is a graphical representation of evaluation measures expressing the relationship. A useful tool for displaying and assessing classifiers is the "receiver operating characteristics (ROC) curve" the common name for a performance indicator for classification problems at different threshold levels is the AUC-ROC curve. The ROC curve, which is a graph based on the true positive rate (sensitivity) and the false positive rate (1-specificity), may be thought of as a statistic that measures how well the model performed overall⁵⁵. The AUC-ROC value ranged from 0 to 1 and indicates a poor and good performance accordingly⁵⁶. The following formulas were used to create the performance evaluation criteria for this study:

$$\text{Sensitivity} = TP / (TP + FN) \quad (15)$$

$$\text{Specificity} = TN / (FP + TN) \quad (16)$$

$$PPV = \frac{TP}{FP + TP} \quad (17)$$

$$NPV = \frac{TN}{TN + FN} \quad (18)$$

$$AUC = \frac{(\sum TP + \sum TN)}{(P + N)} \quad (19)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (20)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (21)$$

$$Fscore = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (22)$$

$$k = \frac{P_o - P_e}{1 - P_e} \quad (23)$$

Here, "TP is true positive, TN is true negative, FN is false negative, FP is false positive, and kappa coefficient is represented by k, observed samples by P_o and predicted result by P_e ".

Result

Statistical measures of selected hydrochemical parameters

In this study, three statistical tests were conducted on the selected hydrochemical dataset: MC, correlation coefficient, and PCA. The MC test (Table 1) showed that all factors were within the threshold value of MC, and therefore suitable for modelling purposes. The depth factor had the highest TOL and lowest VIF (0.66 and 1.515 respectively), while the Ca^{2+} factor had the lowest TOL and highest VIF (0.38 and 0.632 respectively). Pearson's correlation coefficient was used to understand the nature of the substantial association between physical and chemical properties. The correlation coefficient (r) ranges from -1 to +1, with values of 0.5, 0.5–0.8 and 0.8 indicating weak, moderate, and strongly correlation, respectively. The highest correlation values were found between pH and K^+ (0.952) and EC and CI (0.973), while moderate relationships were found between pH and salinity (0.546), pH and Mg^{2+} (0.506), EC and Na^+ (0.644), Mg^{2+} and K^+ (0.593), Na^+ and CI (0.613), and the lowest values were found between Ca^{2+} and Mg^{2+} (0.422), EC and Ca^{2+} (0.365), depth and HCO_3^- (0.359), etc. Details about the correlation coefficient map and table are presented in Fig. 2 and Table 2. PCA analysis showed that PC 1 consisted of 43.21% eigenvalue, followed by PC 2 and PC 3, which had 31.02% and 17.08% eigenvalue, respectively. In PC 1, the dominant factors were EC (0.933), salinity (0.927), Mg^{2+} (0.874) and CI (0.924), while important factors in PC 2 important factors were F^- (0.765), As (0.599) and HCO_3^- (0.582) and in PC 3 dominant factors were PO_4^{2-} (0.620), NO_3^- (0.582) and K^+ (0.339). The biplot map of PC 1, PC 2 and PC 3 is presented in Fig. 3.

Assessment of groundwater vulnerability

Groundwater vulnerability in the aquifers of Ganges delta was assessed using LR, RF and ANN models, and the results are presented in Fig. 4. Statistical, ML, and neural network algorithms were used to understand the spatial distribution of groundwater vulnerability in the vulnerable mega-delta region. We used ArcGIS 10.5 software

Factors	VIF	TOL
Depth (m)	1.515	0.66
pH	2.564	0.39
EC ($\mu\text{S}/\text{cm}$)	2.041	0.49
Salinity (ppt)	1.852	0.54
Ca^{2+} (mg/l)	2.632	0.38
Mg^{2+} (mg/l)	1.818	0.55
Na^+ (mg/l)	2.128	0.47
K^+ (mg/l)	2.564	0.39
Cl^- (mg/l)	1.852	0.54
HCO_3^- (mg/l)	2.381	0.42
NO_3^- (mg/l)	2.778	0.36
SO_4^{2-} (mg/l)	1.852	0.54
PO_4^{2-} (mg/l)	2.083	0.48
F^- (mg/l)	1.724	0.58
As ($\mu\text{g}/\text{l}$)	2.326	0.43

Table 1. Multi-collinearity analysis of selected factors.

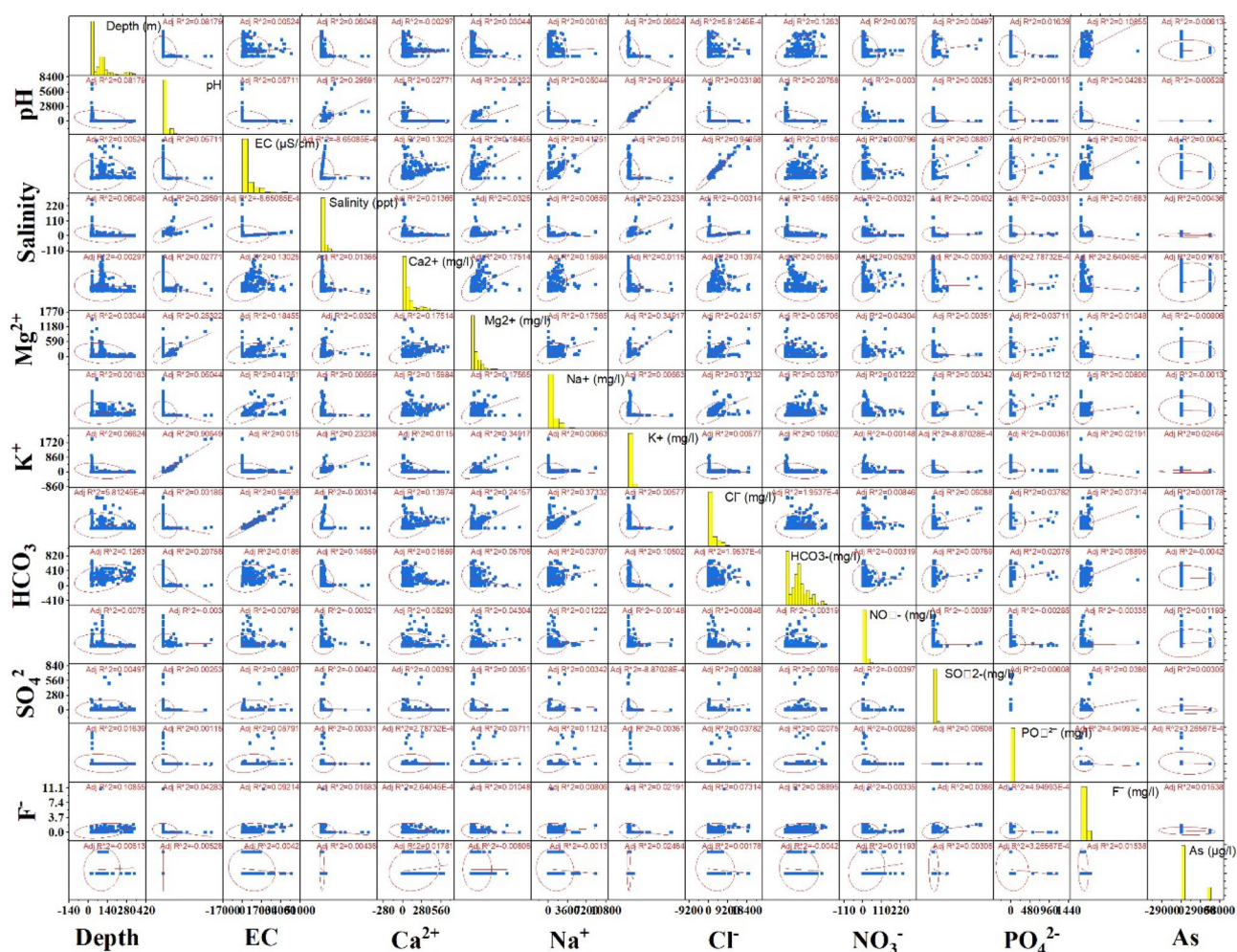


Figure 2. Pearson's correlation among the factors.

Correlations																
Factors	Depth (m)	Temperature (°C)	pH	EC (µS/cm)	Salinity (ppt)	Ca ²⁺ (mg/l)	Mg ²⁺ (mg/l)	Na ⁺ (mg/l)	K ⁺ (mg/l)	Cl ⁻ (mg/l)	HCO ₃ ⁻ (mg/l)	NO ₃ ⁻ (mg/l)	SO ₄ ²⁻ (mg/l)	PO ₄ ²⁻ (mg/l)	F ⁻ (mg/l)	As (µg/l)
Depth	1	0.192**	-0.048	0.044	-0.014	-0.009	-0.011	0.040	0.025	0.044	0.033	-0.063	-0.009	0.015	0.414**	0.027
pH	-0.048	-0.102*	1	-0.264**	-0.299**	-0.213**	-0.271**	-0.329**	-0.108*	-0.300**	0.172**	-0.068	-0.054	-0.012	0.046	-0.164**
EC	0.044	-0.157**	-0.264**	1	0.835**	0.251**	0.551**	0.541**	0.384**	0.842**	0.016	0.250**	0.143**	0.087*	0.239**	0.176**
Salinity	-0.014	0.079	-0.299**	0.835**	1	0.303**	0.584**	0.615**	0.237**	0.861**	-0.014	0.294**	0.101*	0.092*	0.187**	0.188**
Ca ²⁺	-0.009	0.071	-0.213**	0.251**	0.303**	1	0.366**	0.302**	0.142**	0.313**	-0.273**	0.055	0.169**	-0.088*	-0.040	0.075
Mg ²⁺	-0.011	0.164**	-0.271**	0.551**	0.584**	0.366**	1	0.587**	0.264**	0.675**	-0.106*	0.112**	0.146**	0.095*	-0.061	0.213**
Na ⁺	0.040	0.129**	-0.329**	0.541**	0.615**	0.302**	0.587**	1	0.267**	0.633**	0.011	0.239**	0.127**	0.130**	-0.001	0.289**
K ⁺	0.025	0.132**	-0.108*	0.384**	0.615**	0.142**	0.264**	0.267**	1	0.469**	-0.125**	0.180**	-0.040	-0.041	0.059	0.027
Cl ⁻	0.044	0.168**	-0.300**	0.842**	0.861**	0.313**	0.675**	0.633**	0.469**	1	-0.088*	0.280**	0.178**	0.083	0.192**	0.204**
HCO ₃ ⁻	0.033	-0.022	0.172**	0.016	-0.014	-0.273**	-0.106*	0.011	-0.125**	-0.088*	1	0.001	0.014	0.179**	0.197**	-0.061
NO ₃ ⁻	-0.063	-0.029	-0.068	0.250**	0.294**	0.055	0.112**	0.239**	0.180**	0.280**	0.001	1	-0.029	0.049	-0.035	0.010
SO ₄ ²⁻	-0.009	0.010	-0.054	0.143**	0.101*	0.169**	0.146**	0.127**	-0.040	0.178**	0.014	-0.029	1	-0.031	0.130**	-0.006
PO ₄ ²⁻	0.015	-0.005	-0.012	0.087*	0.092*	-0.088*	0.095*	0.130**	-0.041	0.005	0.179**	0.049	-0.031	1	0.005	0.192**
F ⁻	0.414**	0.125**	0.046	0.239**	0.187**	-0.040	-0.061	-0.001	0.059	0.192**	0.197**	-0.035	0.130**	0.005	1	-0.104*
As	0.027	0.019	-0.164**	0.176**	0.188**	0.075	0.213**	0.289**	0.027	0.204**	-0.061	0.010	-0.006	0.192**	-0.104*	1

Table 2. Correlation test among the selected factors. **Correlation is significant at the 0.01 level (2-tailed). *Correlation is significant at the 0.05 level (2-tailed).

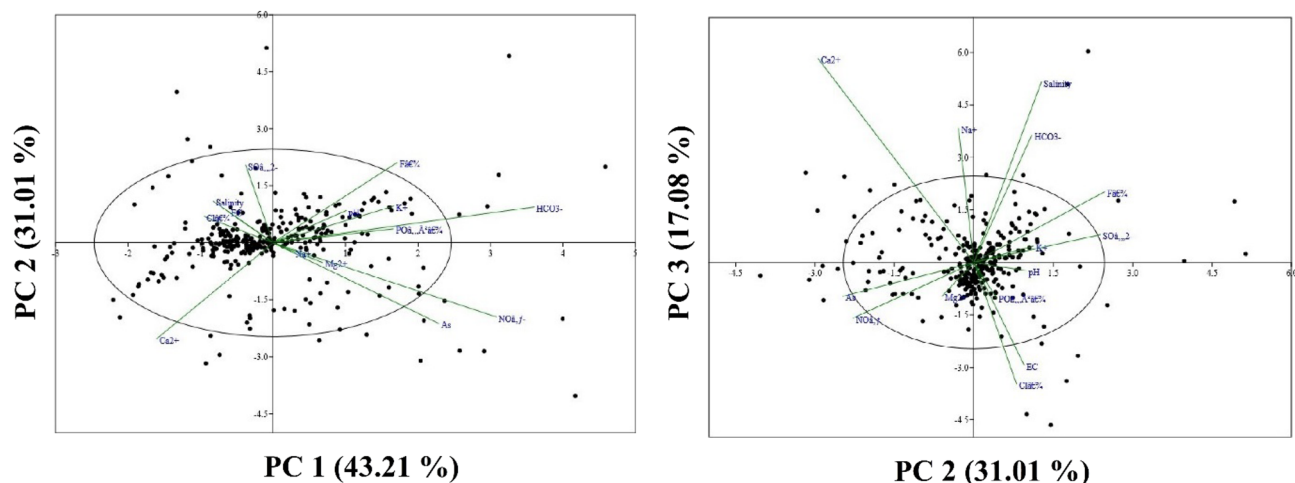


Figure 3. Bi-plot among the applied hydro-chemical factors.

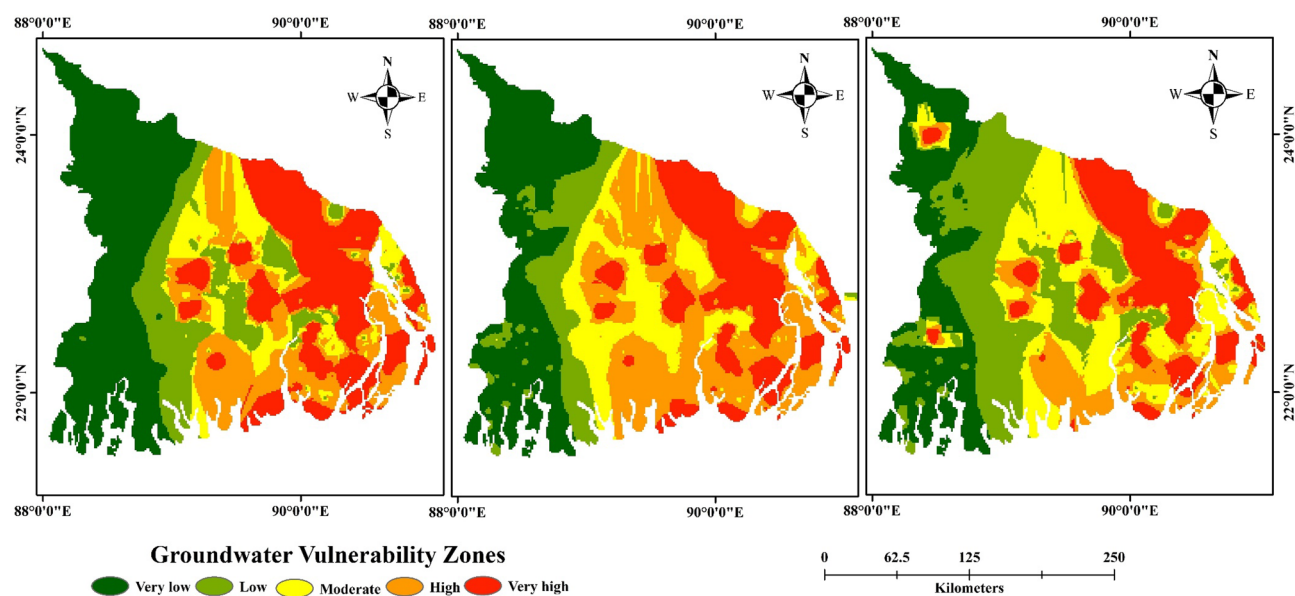


Figure 4. Groundwater vulnerability maps of Ganges delta: (a) LR, (b) RF and (c) ANN.

to map the final spatial distribution of vulnerability using the respective modelling outcomes. Each map was classified into five vulnerability zones namely “very low, low, moderate, high and very high” using “Jenks natural break method”. The final vulnerability maps show that very high groundwater vulnerability zones are found in the eastern and some isolated south-eastern and central middle portions. Conversely, very low groundwater vulnerability zones are found in the north-western, eastern, and south-western parts. The moderate vulnerability zone is found in the central part and isolated patches of the south-eastern and southern parts of the study area. Due to the high concentration of As and other contaminated factors in the groundwater, the eastern part of the Ganges delta, i.e., the region of Bangladesh, is very vulnerable to groundwater compared to the western part of the delta region, i.e., the state of West Bengal in India. Although two isolated patches are found to be in the very high vulnerable zone in the western region of the Ganges delta i.e., part of India in RF and ANN models (Fig. 4).

Importance hydrochemical parameters for groundwater vulnerability

It is a fact that all selected hydrochemical parameters have not equal responsibility for groundwater vulnerability assessment in this study. Therefore, it is fundamental to determine the dominant factors in each applied learning model for groundwater vulnerability. The most dominant factors were identified for the three applied models, i.e., LR, RF, and ANN. The results of the dominant factors for groundwater vulnerability are presented in Table 3 for the three applied models. Factors such as F^- (0.74), Na^+ (0.77), As (0.69), Mg^{2+} (0.58) and HCO_3^- (0.54) are more dominant, while SO_4^{2-} (0.2), pH (0.21), EC (0.31) and PO_4^{3-} (0.32) are less dominant in the LR model. The “mean decrease accuracy (MDA) method of RF algorithm” revealed that Na^+ (0.84), F^- (0.77) and As (0.72) are the most influential factors on groundwater resources followed by HCO_3^- (0.55), and Mg^{2+} (0.54). In ANN, the

Factors	LR	RF	ANN
NO ₃	0.44	0.35	0.41
AS	0.69	0.72	0.78
PO ₄ ²⁻	0.32	0.21	0.38
SO ₄ ²⁻	0.2	0.22	0.19
Salinity	0.32	0.24	0.31
HCO ₃	0.54	0.55	0.67
Cl	0.41	0.38	0.47
K+	0.24	0.25	0.28
Na+	0.77	0.84	0.88
Mg ²⁺	0.58	0.54	0.51
F ⁻	0.74	0.77	0.81
pH	0.21	0.28	0.24
EC	0.31	0.39	0.33
Depth	0.4	0.41	0.41
Ca ²⁺	0.33	0.31	0.35

Table 3. Variables importance of selected factors through applied three models.

dominant factors are Na⁺ (0.88), F⁻ (0.81), As (0.78) and HCO₃ (0.67), and the less dominant factors are SO₄²⁻ (0.19), pH (0.24), salinity (0.31), and EC (0.33).

Evaluation assessment

All three models were evaluated using various metrics such as “sensitivity, specificity, NPV, PPV, ROC-AUC, Kappa-coefficient, and F-score”. Among the three models, the ANN model is the most suitable for modelling groundwater vulnerability, with a ROC value of 0.912 and 0.902, for training and validation, respectively. This is followed by the RF model with 0.817 and 0.792 for training and validation, and then the LR model with 0.749 and 0.712 for training and validation. The PPV and NPV are also high in the ANN model, with values of 0.883 and 0.885 in the validation stage. The sensitivity analysis showed that the ANN model had the highest result at 0.889, followed by RF and LR, with 0.782 and 0.721, respectively, in the validation stage. The Kappa and F-score also indicate that the ANN model is the best fit with values of 0.643 and 0.882 in the validation stage, followed by RF and LR (Table 4). The Taylor diagram in Fig. 5 also shows that the ANN is optimal based on standard deviation and correlation.

Quality assessment of groundwater

Piper, USSL, and Wilcox diagrams were used to assess the hydrochemical properties and quality of groundwater in the Ganges delta region. The Piper diagram (Fig. 6a) showed that alkaline earth (Ca²⁺+Mg²⁺) dominates over alkalis (Na+K) and that Cl and NO₃ dominate over HCO₃. The Wilcox diagram (Fig. 6b) showed that two samples were unsuitable, while the others fell into the doubtful to unsuitable, permissible to doubtful, good to permissible and excellent categories. The USSL diagram (Fig. 6c) revealed a high salinity, low sodium and alkali hazard dominance. The collected datasets were grouped and analyzed using a hierarchical clustering method, which showed that the second cluster, and to a lesser extent, the first cluster, significantly influenced the state and the groundwater quality. The dendrogram (Fig. 6d) showed that the first cluster covered approximately 32% of the datasets, the second cluster covered the maximum dataset (47%) and the third cluster covered the lowest dataset (21%).

Models	Stage	Parameters					Kappa	F-Score
		Sensitivity	Specificity	PPV	NPV	AUC		
LR	Train	0.741	0.749	0.752	0.721	0.749	0.532	0.811
	Validation	0.721	0.722	0.711	0.701	0.712	0.513	0.792
RF	Train	0.821	0.812	0.813	0.825	0.817	0.63	0.863
	Validation	0.782	0.751	0.794	0.744	0.792	0.601	0.841
ANN	Train	0.907	0.903	0.901	0.901	0.912	0.651	0.912
	Validation	0.889	0.891	0.883	0.885	0.902	0.643	0.882

Table 4. Evaluation metrics of applied models.

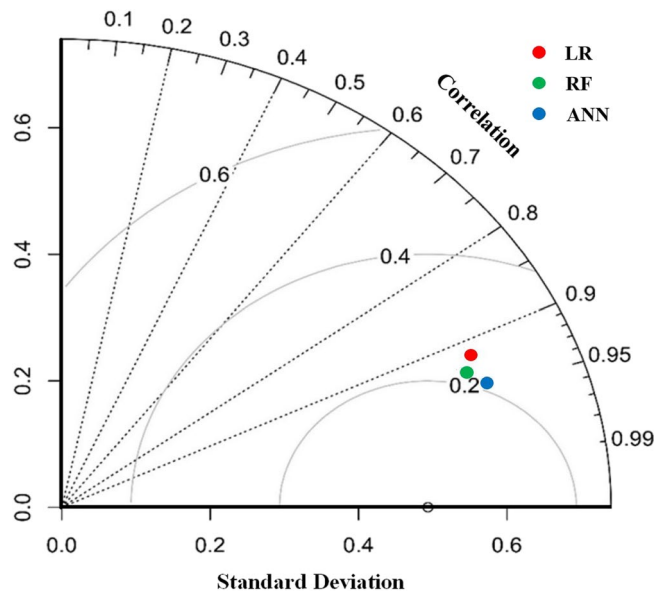


Figure 5. Graphical evaluation measure of applied models using Taylor diagram.

Discussion

The fundamental reasons for spatio-temporal fluctuations in the groundwater supply are increasing water demand across all sectors and changing climatic conditions⁵⁷. These factors present a significant challenge to water resource planners. This study has demonstrated that considerable concentrations of elevated arsenic and nitrate in groundwater, as well as salinization, are among the groundwater quality issues in the coastal areas of the multi-aquifers of the Ganges delta. The quality of groundwater along the coast primarily depends on geological conditions, hydrogeological processes, and chemical activities⁵⁸. Therefore, a trustworthy assessment of groundwater vulnerability is a crucial first step in choosing the best design or framework for future water resource development.

It can be challenging to select a preferred model for assessing inherent vulnerability that can effectively match the research topic's features and the study area's geo-environmental characteristics. The literature reveals that many academics have compared two or more vulnerability indices to create a meticulously tailored intrinsic vulnerability model for their research, aiming to achieve optimal output^{59,60}. In the current scenario, statistical and machine learning (ML) algorithms are widely employed in groundwater-related studies worldwide. For example, Yu et al. (2022) Vu⁶¹ applied an integrated Variable Weight Model (VWM) and DRASTIC model to assess groundwater vulnerability in China and found that the VWM-DRASTIC combination provided optimal predictive analysis. Vu et al.⁶² used a numerical model and the index-overlay method in conjunction with climate scenarios (RCPs) to evaluate groundwater vulnerability and associated sustainability in Taiwan, and they recommended optimal predictive analysis. Furthermore, several machine learning models have been utilized in various groundwater-related studies, including groundwater vulnerability^{4,14,18,63}, nitrate concentration in groundwater^{64–66}, and more. The random forest (RF) model is well-known for its numerous advantages and has been employed in various geoscientific fields, including groundwater vulnerability studies. Lahjouj et al.⁶⁷ utilized the RF algorithm in a survey of groundwater vulnerability to nitrate concentration in Morocco and achieved an accuracy assessment of 0.822 in terms of AUC-ROC. Similarly, Saha et al.¹⁸ used RF to assess hydrochemical-based groundwater vulnerability in parts of the Ganges delta and achieved optimal accuracy rates of 0.849 and 0.812 in the training and validation data of the ROC. Various statistical techniques are available, ranging from straightforward descriptive statistics of concentrations of specific contaminants to more complex regression analyses that consider the impacts of multiple predictor variables⁶⁸. Binary logistic regression, sometimes known as logistic regression (LR), is a frequently used statistical technique for estimating groundwater vulnerability. LR models relate the potential influencing factors to the likelihood that a pollutant concentration will exceed a threshold value. Mohammaddost et al.⁶⁹ employed DRASTIC, EBF, and LR models in the Kabul basin of Afghanistan to assess groundwater vulnerability, and they found that LR provided 66% accuracy in AUC-ROC prediction analysis. Adiat et al.⁷⁰ applied LR for the same assessment in the Ilesa gold mining area of Nigeria, achieving an 85.7% accuracy in model prediction. Recently, with the significant advantages of neural network algorithms, several neural network models have also been used in groundwater studies. For instance, Elzain et al.²⁷ used the DLNN model in aquifer vulnerability studies in South Korea, while Elzain et al.⁷¹ employed the RBNN model to assess groundwater vulnerability to nitrate contamination in the southern part of Korea.

Based on the discussion above and considering the significant advantages of statistical, machine learning (ML), and neural network algorithms, three popular learning algorithms, namely logistic regression (LR), random forest (RF), and artificial neural network (ANN), were selected for the optimal assessment of groundwater vulnerability in the mega delta of the Ganges delta, taking into account field-based hydrochemical parameters. The findings of this study demonstrate that among the applied models, ANN yields the most optimal results,

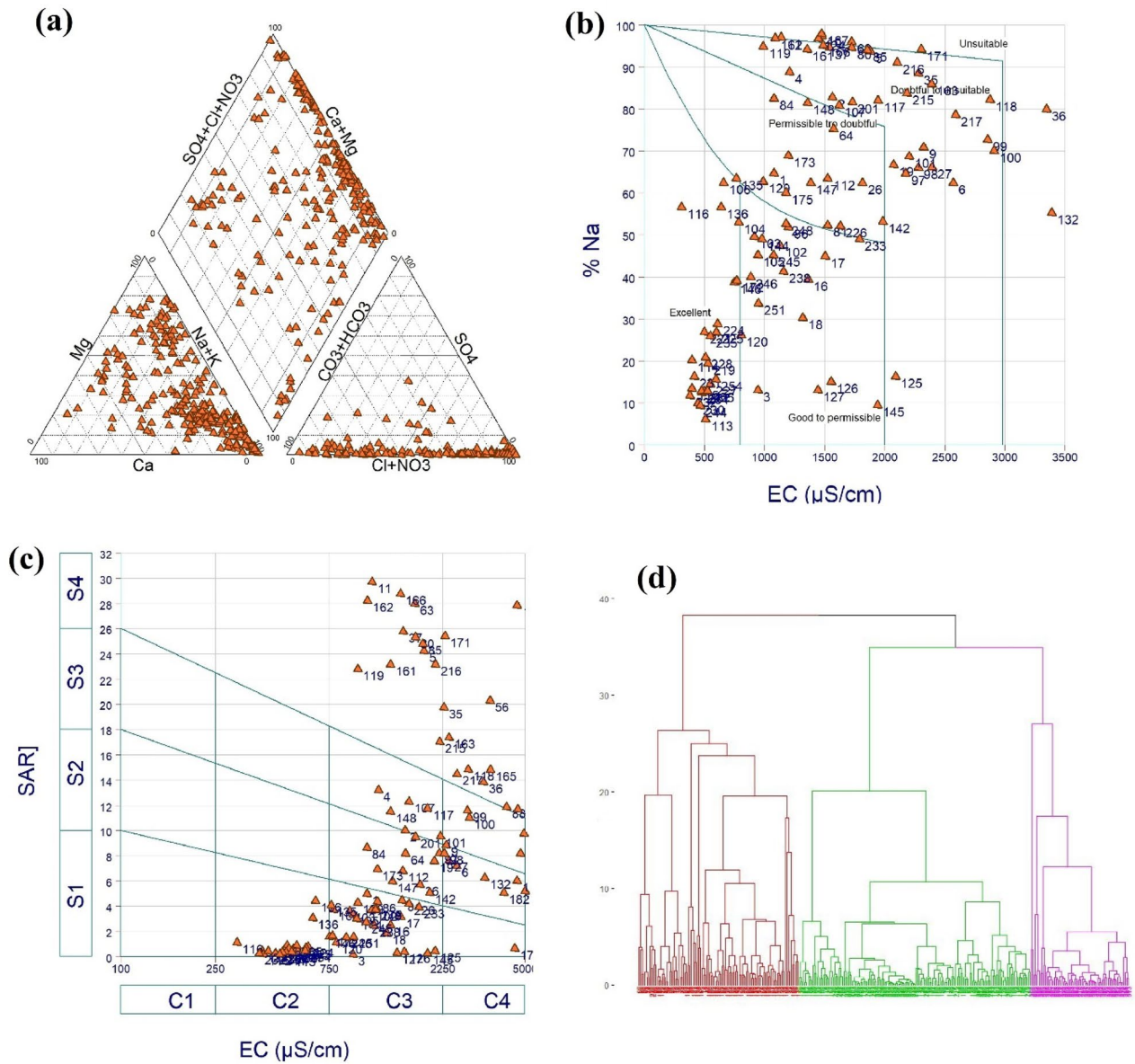


Figure 6. Quality assessment of groundwater: (a) Piper diagram, (b) Wilcox's diagram, (c) USSL diagram, and (d) cluster analysis using dendrogram.

with AUC-ROC scores of 0.912 and 0.902 in training and validation, respectively, for groundwater vulnerability studies. RF follows with scores of 0.817 and 0.792 in training and validation, and LR with scores of 0.749 and 0.712 in training and validation. The high performance of the ANN model can be attributed to its capacity for parallel processing, enabling it to handle multiple tasks simultaneously. The statistical analysis of all selected hydrochemical parameters reveals that pH and K^+ (0.952) and EC and Cl^- (0.973) are highly correlated, while pH and salinity (0.546), pH and Mg^{2+} (0.506), EC and Na^+ (0.644), Mg^{2+} and K^+ (0.593), and Na^+ and Cl^- (0.613) show moderate correlations. It is also found that pH, NO_3^- , As, and K^+ are the most influential factors for groundwater vulnerability in this study region.

Henceforth, studies on groundwater vulnerability serve as crucial measurements for the sustainable management of water resources, environmental preservation, and the guarantee of a secure and uncontaminated drinking water supply for both present and future generations.

Nonetheless, it is a fact that employing combined techniques and methodologies can aid in resolving ambiguities related to GIS-based vulnerability assessment frameworks in geoscientific fields. The approaches presented in this research can be tested in various hydrogeological and geo-environmental contexts to understand the spatial distribution of vulnerability. Evaluating groundwater vulnerability studies requires careful consideration of the data and tools used for validation. Furthermore, the limitations of this study are not considered various important factors, such as the hydrogeological process of groundwater, land use land cover, and aquifer and soil characteristics, as all of these factors affect groundwater quality. In the future, other neural networks and deep learning algorithms can be beneficial for the optimal assessment of groundwater vulnerability in the mega-delta,

considering changing climate and land use land cover. Therefore, the results of this study will be valuable to land use planners and provide fundamental information for the optimal assessment and management of groundwater risk zones accordingly.

Conclusion

Globally, assessing susceptibility to groundwater contamination is crucial for proactive management aimed at safeguarding groundwater resources for various uses. Creating more effective sustainable development policies regarding potential groundwater pollution by utilizing more precise vulnerability maps. In the Ganges deltaic region, the high concentrations of contaminants, such as arsenic (As), are primarily responsible for groundwater vulnerability, and the associated human health hazards are a significant concern for global researchers. In the present research, there is a focus on creating an effective vulnerability map for a mega-delta, specifically the Ganges delta. This involves the application of LR, RF, and ANN models in the modelling and mapping process. Sensitivity analysis indicates that the ANN output is the most optimal, followed by RF and LR. The study reveals that the neural network algorithm is the best suited for assessing groundwater vulnerability related to contamination in the study region, surpassing traditional statistical analysis. Hydrochemical parameters such as pH, NO_3^- , As, and K^+ dominate this deltaic aquifer, contributing to vulnerability. Overall, all vulnerability maps indicate that the study area's western, central, south, and eastern parts are highly vulnerable. Due to elevated levels of As and various ion contaminations, most groundwater samples from the Ganges delta are unsuitable for drinking and irrigation. Consequently, the improper implementation of government policies, a lack of awareness, and inadequate management are the primary concerns leading to groundwater deterioration in this region. Therefore, immediate action is necessary to sustain and conserve groundwater resources in the world's largest and most densely populated deltaic region. Henceforth, in future application of deep learning and both the dataset i.e., dry and wet season for sampling procedure will be helpful for better understanding of groundwater vulnerability in this vulnerable region.

Data availability

The datasets used and/or analyzed during the current study are available from the reasonable request.

Received: 22 August 2023; Accepted: 11 January 2024

Published online: 13 January 2024

References

- Korrai, S., Gangu, K. K., Rao, P. V. V. & Jonnalagadda, S. B. Assessment of groundwater vulnerability to seawater intrusion using multiple approaches. *Arab. J. Geosci.* **15**, 1–13 (2022).
- Biswas, T., Chandra Pal, S., Saha, A. & Ruidas, D. Arsenic and fluoride exposure in drinking water caused human health risk in coastal groundwater aquifers. *Environ. Res.* **238**, 117257 (2023).
- Jaydhar, A. K., ChandraPal, S., Saha, A., Islam, A. R. M. D. T. & Ruidas, D. Hydrogeochemical evaluation and corresponding health risk from elevated arsenic and fluoride contamination in recurrent coastal multi-aquifers of eastern India. *J. Clean. Prod.* **369**, 133150 (2022).
- Ruidas, D., Pal, S. C., Md. Towfiqul Islam, A. R. & Saha, A. Hydrogeochemical evaluation of groundwater aquifers and associated health hazard risk mapping using ensemble data driven model in a water Scars Plateau region of Eastern India. *Expo Health* <https://doi.org/10.1007/s12403-022-00480-6> (2022).
- Jayaprakash, M., Giridharan, L., Venugopal, T., KrishnaKumar, S. P. & Periakali, P. Characterization and evaluation of the factors affecting the geochemistry of groundwater in Neyveli, Tamil Nadu, India. *Environ. Geol.* **54**, 855–867 (2008).
- Khan, Q., Liaqat, M. U. & Mohamed, M. M. A comparative assessment of modeling groundwater vulnerability using DRASTIC method from GIS and a novel classification method using machine learning classifiers. *Geocarto Int.* **20**, 1–19 (2021).
- Kirlas, M. C., Karpouzou, D. K., Georgiou, P. E. & Katsifarakis, K. L. A comparative study of groundwater vulnerability methods in a porous aquifer in Greece. *Appl. Water Sci.* **12**, 1–21 (2022).
- Oh, H.-J., Kim, Y.-S., Choi, J.-K., Park, E. & Lee, S. GIS mapping of regional probabilistic groundwater potential in the area of Pohang City, Korea. *J. Hydrol.* **399**, 158–172 (2011).
- Ahmad, W., Choi, M., Kim, S. & Kim, D. Detection of land subsidence and its relationship with land cover types using ESA Sentinel satellite data: A case study of Quetta Valley, Pakistan. *Int. J. Remote Sens.* **40**, 9572–9603 (2019).
- Kulkarni, H., Shah, M. & Shankar, P. V. Shaping the contours of groundwater governance in India. *J. Hydrol. Regional Stud.* **4**, 172–192 (2015).
- Nair, A. M., Prasad, K. R. & Srinivas, R. Groundwater vulnerability assessment of an urban coastal phreatic aquifer in India using GIS-based DRASTIC model. *Groundw. Sustain. Dev.* **19**, 100810 (2022).
- Zhang, Q., Li, P., Lyu, Q., Ren, X. & He, S. Groundwater contamination risk assessment using a modified DRATICL model and pollution loading: A case study in the Guanzhong Basin of China. *Chemosphere* **291**, 132695 (2022).
- Mohamed, M. M., Parimalarenganayaki, S., Khan, Q. & Murad, A. Review on the use of environmental isotopes for groundwater recharge and evaporation studies in the GCC countries. *Groundw. Sustain. Dev.* **12**, 100546 (2021).
- Pal, S., Ruidas, D., Saha, A., Islam, A. R. M. D. T. & Chowdhuri, I. Application of novel data-mining technique-based nitrate concentration susceptibility prediction approach for coastal aquifers in India. *J. Clean. Prod.* **346**, 131205 (2022).
- Raju, N. J. Arsenic in the geo-environment: A review of sources, geochemical processes, toxicity and removal technologies. *Environ. Res.* **203**, 111782 (2022).
- Ghosh, N. C. & Singh, R. D. Groundwater arsenic contamination in India: Vulnerability and scope for remedy (2009).
- Chakraborty, M., Mukherjee, A. & Ahmed, K. M. Regional-scale hydrogeochemical evolution across the arsenic-enriched transboundary aquifers of the Ganges River Delta system, India and Bangladesh. *Sci. Total Environ.* **823**, 153490 (2022).
- Saha, A., Pal, S. C., Chowdhuri, I., Roy, P. & Chakraborty, R. Effect of hydrogeochemical behavior on groundwater resources in Holocene aquifers of moribund Ganges Delta, India: Infusing data-driven algorithms. *Environ. Pollut.* **314**, 120203 (2022).
- Ruidas, D., ChandraPal, S., Biswas, T., Saha, A. & Md. Towfiqul Islam, R. A. Extreme exposure of fluoride and arsenic contamination in shallow coastal aquifers of the Ganges delta, transboundary of the Indo-Bangladesh region. *Geosci. Front.* <https://doi.org/10.1016/j.gsf.2023.101725> (2023).
- Acharyya, S. K. & Shah, B. A. Groundwater arsenic pollution affecting deltaic West Bengal, India. *Curr. Sci. (Bangalore)* **99**, 1787–1794 (2010).

21. Hoque, B. A. *et al.* Demand-based water options for arsenic mitigation: An experience from rural Bangladesh. *Public Health* **118**, 70–77 (2004).
22. Saha, R., Dey, N. C., Rahman, M., Bhattacharya, P. & Rabbani, G. H. Geogenic arsenic and microbial contamination in drinking water sources: Exposure risks to the coastal population in Bangladesh. *Front. Environ. Sci.* **7**, 57 (2019).
23. Chakraborti, D., Singh, S. K., Rashid, M. H. & Rahman, M. M. Arsenic: Occurrence in groundwater. *Encyclop. Environ. Health* **2**, 1e17 (2011).
24. Arabameri, A. *et al.* K-fold and state-of-the-art metaheuristic machine learning approaches for groundwater potential modelling. *Water Resour. Manage.* **35**, 1837–1869 (2021).
25. Band, S. S. *et al.* Using computational-intelligence algorithms and remote sensing data to optimize the locations of check dams to control sediment and runoff in Kandolus watershed, Mazandaran, Iran. *Geocarto Int.* **20**, 1–23 (2022).
26. Adimalla, N. & Taloor, A. K. Hydrogeochemical investigation of groundwater quality in the hard rock terrain of South India using Geographic Information System (GIS) and groundwater quality index (GWQI) techniques. *Groundw. Sustain. Dev.* **10**, 100288 (2020).
27. Elzain, H. E. *et al.* Modeling of aquifer vulnerability index using deep learning neural networks coupling with optimization algorithms. *Environ. Sci. Pollut. Res.* **28**, 57030–57045 (2021).
28. Mosavi, A. *et al.* Susceptibility prediction of groundwater hardness using ensemble machine learning models. *Water* **12**, 2770 (2020).
29. Gharekhani, M., Nadiri, A. A., Khatibi, R., Sadeghfam, S. & Moghaddam, A. A. A study of uncertainties in groundwater vulnerability modelling using Bayesian model averaging (BMA). *J. Environ. Manage.* **303**, 114168 (2022).
30. Biswas, T. *et al.* Hydro-chemical assessment of groundwater pollutant and corresponding health risk in the Ganges delta, Indo-Bangladesh region. *J. Clean. Prod.* **382**, 135229 (2023).
31. Chakraborty, M. *et al.* Modeling regional-scale groundwater arsenic hazard in the transboundary Ganges River Delta, India and Bangladesh: Infusing physically-based model with machine learning. *Sci. Total Environ.* **748**, 141107 (2020).
32. Verma, S. & Sinha, A. Appraisal of groundwater arsenic on opposite banks of River Ganges, West Bengal, India, and quantification of cancer risk using Monte Carlo simulations. *Environ. Sci. Pollut. Res.* **20**, 1–21 (2022).
33. Kar, S. *et al.* Arsenic-enriched aquifers: Occurrences and mobilization of arsenic in groundwater of Ganges Delta Plain, Barasat, West Bengal, India. *Appl. Geochem.* **25**, 1805–1814 (2010).
34. Bagchi, K. & Mukerjee, K. Diagnostic survey of West Bengal (s). *Dept. Geography Calcutta Univ. Pantg. Rareh Bengal* **42**, 17–19 (1983).
35. Sarkar, A. *et al.* Evolution of Ganges-Brahmaputra western delta plain: Clues from sedimentology and carbon isotopes. *Quatern. Sci. Rev.* **28**, 2564–2581 (2009).
36. Srivastava, P., Singh, I. B., Sharma, M. & Singhvi, A. K. Luminescence chronometry and Late Quaternary geomorphic history of the Ganga Plain, India. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **197**, 15–41 (2003).
37. Kanhaiya, S., Singh, B. P., Tripathi, M., Sahu, S. & Tiwari, V. Lithofacies and particle-size characteristics of late Quaternary flood-plain deposits along the middle reaches of the Ganga river, central Ganga plain, India. *Geomorphology* **284**, 220–228 (2017).
38. Saha, D. Groundwater resources (2018).
39. Sarangi, S. K., Mainuddin, M. & Maji, B. Problems, management, and prospects of acid sulphate soils in the Ganges Delta. *Soil Syst.* **6**, 95 (2022).
40. Rani, N. N. V. S., Satyanarayana, A. N. V., Bhaskaran, P. K., Rice, L. & Kantamaneni, K. Assessment of groundwater vulnerability using integrated remote sensing and GIS techniques for the West Bengal coast, India. *J. Contam. Hydrol.* **238**, 103760 (2021).
41. Lee, S. & Pradhan, B. Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. *Landslides* **4**, 33–41 (2007).
42. Wubalem, A. & Meten, M. Landslide susceptibility mapping using information value and logistic regression models in Goncha Siso Eneses area, northwestern Ethiopia. *SN Appl. Sci.* **2**, 807 (2020).
43. Bai, S.-B. *et al.* GIS-based logistic regression for landslide susceptibility mapping of the Zhongxian segment in the Three Gorges area, China. *Geomorphology* **115**, 23–31 (2010).
44. Dai, F. C. & Lee, C. F. Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong. *Geomorphology* **42**, 213–228 (2002).
45. Saha, G. Applying logistic regression model to the examination results data. *J. Reliabil. Stat. Stud.* **20**, 105–117 (2011).
46. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
47. Chen, W., Li, X., Wang, Y., Chen, G. & Liu, S. Forested landslide detection using LiDAR data and the random forest algorithm: A case study of the Three Gorges, China. *Remote Sens. Environ.* **152**, 291–301 (2014).
48. Saha, A. *et al.* Land degradation risk dynamics assessment in red and lateritic zones of eastern plateau, India: A combine approach of K-fold CV, data mining and field validation. *Ecol. Inform.* **69**, 101653 (2022).
49. Kohestani, V. R., Hassanlourad, M. & Ardakani, A. Evaluation of liquefaction potential based on CPT data using random forest. *Nat. Hazards* **79**, 1079–1089 (2015).
50. Lee, S., Ryu, J.-H., Lee, M.-J. & Won, J.-S. Use of an artificial neural network for analysis of the susceptibility to landslides at Boun, Korea. *Environ. Geol.* **44**, 820–833 (2003).
51. Kawabata, D. & Bandibas, J. Landslide susceptibility mapping using geological data, a DEM from ASTER images and an Artificial Neural Network (ANN). *Geomorphology* **113**, 97–109 (2009).
52. Paola, J. D. & Schowengerdt, R. A. A review and analysis of backpropagation neural networks for classification of remotely-sensed multi-spectral imagery. *Int. J. Remote Sens.* **16**, 3033–3058 (1995).
53. Ruidas, D., Pal, S., Saha, A., Chowdhuri, I. & Shit, M. Hydrogeochemical characterization based water resources vulnerability assessment in India's first Ramsar site of Chilka lake. *Mar. Pollut. Bull.* **184**, 25 (2022).
54. Saha, A. *et al.* Land degradation risk dynamics assessment in red and lateritic zones of eastern plateau, India: A combine approach of K-fold CV, data mining and field validation. *Ecol. Inform.* **20**, 101653 (2022).
55. Beguería, S. Validation and evaluation of predictive models in hazard assessment and risk management. *Nat. Hazards* **37**, 315–329 (2006).
56. Hanley, J. A. & McNeil, B. J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143**, 29–36 (1982).
57. Swain, S., Mishra, S. K. & Pandey, A. A detailed assessment of meteorological drought characteristics using simplified rainfall index over Narmada River Basin, India. *Environ. Earth Sci.* **80**, 221 (2021).
58. Han, D. & Currell, M. J. Review of drivers and threats to coastal groundwater quality in China. *Sci. Total Environ.* **806**, 150913 (2022).
59. Hester, E. T. & Fox, G. A. Preferential flow in Riparian groundwater: Gateways for watershed solute transport and implications for water quality management. *Water Resour. Res.* **56**, e2020WR028186 (2020).
60. Torkashvand, M., Neshat, A., Javadi, S. & Pradhan, B. New hybrid evolutionary algorithm for optimizing index-based groundwater vulnerability assessment method. *J. Hydrol.* **598**, 126446 (2021).
61. Yu, H. *et al.* Integrated variable weight model and improved DRASTIC model for groundwater vulnerability assessment in a shallow porous aquifer. *J. Hydrol.* **608**, 127538 (2022).

62. Vu, T.-D., Ni, C.-F., Li, W.-C., Truong, M.-H. & Hsu, S. M. Predictions of groundwater vulnerability and sustainability by an integrated index-overlay method and physical-based numerical model. *J. Hydrol.* **596**, 126082 (2021).
63. Barzegar, R. *et al.* Improving GALDIT-based groundwater vulnerability predictive mapping using coupled resampling algorithms and machine learning models. *J. Hydrol.* **598**, 126370 (2021).
64. Adimalla, N. & Qian, H. Groundwater chemistry, distribution and potential health risk appraisal of nitrate enriched groundwater: A case study from the semi-urban region of South India. *Ecotoxicol. Environ. Saf.* **207**, 111277 (2021).
65. Castaldo, G., Visser, A., Fogg, G. E. & Harter, T. Effect of groundwater age and recharge source on nitrate concentrations in domestic wells in the san Joaquin valley. *Environ. Sci. Technol.* **55**, 2265–2275 (2021).
66. Nakagawa, K., Amano, H., Persson, M. & Berndtsson, R. Spatiotemporal variation of nitrate concentrations in soil and groundwater of an intensely polluted agricultural area. *Sci. Rep.* **11**, 1–13 (2021).
67. Lahjouj, A., El Hmaid, A., Bouhafa, K. & Boufala, M. Mapping specific groundwater vulnerability to nitrate using random forest: Case of Sais basin, Morocco. *Model. Earth Syst. Environ.* **6**, 1451–1466 (2020).
68. Worrall, F. & Kolpin, D. W. Direct assessment of groundwater vulnerability from single observations of multiple contaminants. *Water Resour. Res.* **39**, 25 (2003).
69. Mohammaddost, A., Mohammadi, Z., Rezaei, M., Pourghasemi, H. R. & Farahmand, A. Assessment of groundwater vulnerability in an urban area: A comparative study based on DRASTIC, EBF, and LR models. *Environ. Sci. Pollut. Res.* **29**, 72908–72928 (2022).
70. Adiat, K. A. N., Akeredolu, B. E., Akinlalu, A. A. & Olayanju, G. M. Application of logistic regression analysis in prediction of groundwater vulnerability in gold mining environment: A case of Ilesa gold mining area, southwestern, Nigeria. *Environ. Monit. Assess.* **192**, 577 (2020).
71. Elzain, H. E. *et al.* Comparative study of machine learning models for evaluating groundwater vulnerability to nitrate contamination. *Ecotoxicol. Environ. Saf.* **229**, 113061 (2022).

Acknowledgements

This publication was supported by the Deanship of Scientific Research at the King Faisal University, Saudi Arabia (Grant: 3422).

Author contributions

A.S.: conceptualization, methodology, investigation, formal analysis, visualization, writing—original draft, writing—review and editing; S.C.P.: supervision, conceptualization, methodology, investigation, formal analysis, visualization, writing—original draft, writing—review and editing; A.R.M.T.I.: investigation, formal analysis, visualization, writing—original draft, writing—review and editing; A.I.: investigation, formal analysis, visualization, writing—original draft, writing—review and editing; E.A.: investigation, formal analysis, visualization, writing—original draft, writing—review and editing; M.K.I.: investigation, formal analysis, visualization, writing—original draft, writing—review and editing.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.C.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



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

© The Author(s) 2024