Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Integrated Hydrochemical Assessment and Machine Learning-Based Spatial Prediction of Groundwater Quality in a Coastal Aquifer System of Bangladesh

Md. Hazrat Ali

contact@enghazrat.com

Chittagong University of Engineering and Technology

Saima Zaheen

The University of Texas at El Paso

Md. Ala-Uddin

Chittagong University of Engineering and Technology

Tanvir Haque Tuhin

Technische Universität Berlin

Asiful Hoque

Chittagong University of Engineering and Technology

Md. Reaz Akter Mullick

Chittagong University of Engineering and Technology

Shyamal Acharya

Chittagong University of Engineering and Technology

Research Article

Keywords: Water Quality Index (WQI), Machine Learning (ML), Heatmap, Sensitivity Analysis, Groundwater Quality, Coastal Aquifer

Posted Date: November 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11368045/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Integrated Hydrochemical Assessment and Machine Learning-Based Spatial Prediction of Groundwater Quality in a Coastal Aquifer System of Bangladesh

Md. Hazrat Ali[1,2]*, Saima Zaheen[2,3], Md. Ala- Uddin[2], Tanvir Haque Tuhin[2,4], Asiful Hoque[2], Md. Reaz Akter Mullick[2], Shyamal Acharya[2]

[1]Department of Water Engineering and Management, Asian Institute of Technology, Khlong Luang, Pathum Thani 12120, Thailand

[2]Department of Civil Engineering, Chittagong University of Engineering and Technology, Chattogram-4349, Bangladesh

[3]Department of Civil, Environmental and Construction Engineering, The University of Texas at El Paso, El Paso, Texas, USA

[4]Department of Civil Engineering, Technische Universität Berlin, Berlin, Germany

*Corresponding Author: Md. Hazrat Ali

Email: contact@enghazrat.com; ORCID iD: 0009-0007-7255-1046

Abstract

This study investigates groundwater quality in the coastal aquifers of Patenga, Chattogram, using an integrated approach combining hydrochemical analysis, spatial mapping, and machine learning (ML) techniques. To compare groundwater characteristics, a comparative assessment between shallow tube wells (STWs) and deep tube wells (DTWs) was conducted. The results reveal distinct hydrochemical characteristics: deep aquifers are predominantly affected by salinity intrusion, while shallow aquifers have higher concentrations of total dissolved solids (TDS), hardness, and arsenic due to surface influences and geochemical interactions. Subsequently, the Water Quality Index (WQI) was used to evaluate the suitability of drinking water. The findings indicate that, while a significant portion of the study area falls within excellent to good categories, localized zones exhibit poor to unsuitable conditions, primarily due to elevated Arsenic and Iron concentrations. Further, correlation analysis highlights strong relationships among chloride, salinity, and TDS, confirming seawater intrusion as a dominant process. In contrast, arsenic and iron show strong positive correlations with WQI, indicating they are critical contaminants. Next, machine learning models were developed to predict groundwater quality, with Linear Regression demonstrating superior performance across multiple evaluation metrics ($R^2 \approx 0.99$). Sensitivity analyses using Sobol, Morris, and feature importance methods consistently identify arsenic and iron as the most influential parameters, collectively explaining over 95% of WQI variability. Finally, spatial analysis reveals a deterioration of groundwater quality toward coastal and riverine zones, emphasizing the combined effects of natural hydrogeological processes and anthropogenic activities.

Keywords Water Quality Index (WQI); Machine Learning (ML); Heatmap; Sensitivity Analysis; Groundwater Quality; Coastal Aquifer

1 Introduction

Bangladesh is a low-lying riverine country in the South Asia covering about 147,610 km², of which about 6.4% is water. The country supports nearly 166 million people [1]. Sitting at the confluence of the Ganges, Brahmaputra and Meghna, Rivers, the country is within a large deltaic plain [2]. This unique deltaic setting has resulted in a complex hydrogeological environment where groundwater plays a crucial role in sustaining livelihoods, agriculture, and domestic water supply. Globally, groundwater makes up around 30% of freshwater resources that provides about 50% of drinking water and 20% of irrigation water [3-5]. In Bangladesh, groundwater is the main source of potable water, especially in rural and peri-urban areas, due to its accessibility and reliability.

Bangladesh experiences a tropical monsoon climate, characterized by significant spatial and temporal variation in rainfall and temperature. The country receives about 80% of its annual rainfall during the monsoon season, with an average annual precipitation of approximately 2320 mm, ranging from 1200 mm in the north-western region to nearly 5000 mm in the north-eastern region [6]. Surface water resources originate mainly from rainfall, rivers, standing water bodies, and seasonal wetlands. Rivers alone contribute nearly 1000 billion cubic meters of water annually, while standing water bodies, seasonal

wetlands, and in-stream water storage provide approximately 0.61, 2.69, and 0.5 billion cubic meters, respectively [7]. The Ganges-Brahmaputra River system discharges an estimated 795,000 million m^3 of surface water annually, underscoring the region's hydrological richness [8]. Despite abundant water resources, the availability of safe, potable water remains a major challenge due to spatial variability, seasonal fluctuations, and deteriorating water quality.

Groundwater has become increasingly important in Bangladesh's water supply and agricultural sectors. The total renewable water resource of the country is estimated at around 105 km^3 annually, including 84 km^3 of surface water generated by precipitation and 21 km^3 of groundwater recharge [9]. Of the total annual water withdrawal capacity of approximately 35.87 km^3, around 31.50 km^3 is used for agriculture, 3.60 km^3 for domestic purposes, and 0.77 km^3 for industrial uses. Groundwater alone accounts for nearly 79% of total water withdrawal, reflecting the heavy reliance on subsurface water resources. Irrigation practices in Bangladesh have increasingly shifted toward groundwater; currently, approximately 4.2 million hectares are irrigated with groundwater, while only 1.03 million hectares rely on surface water. The share of surface water irrigation declined from 76% in 1981 to about 23% in 2012, whereas groundwater irrigation increased from 16% to nearly 80% during the same period [10]. This rapid increase in groundwater exploitation underscores its critical role in ensuring food security and economic development.

Bangladesh's hydrogeological framework consists mainly of several aquifer systems: the Upper Holocene, Middle Holocene, Late Pleistocene, and Plio-Pleistocene Aquifers [11]. Groundwater flows from the northern uplands toward the southern coastal regions. There is localized discharge into rivers and streams within shallow aquifers. The hydraulic gradient varies across the country, from 1:1,000 in the northwest to less than 1:20,000 near the coast. Aquifer permeability is high, from 10 to 200 m/day, while transmissivity ranges from 100 to 10,000 m^2/day [12]. These conditions support significant groundwater storage and movement in the Bengal delta. However, shallow coastal aquifers are often contaminated by over-abstraction and saline water intrusion [13]. Salinization leads to higher concentrations of TDS, sodium, magnesium, chloride, and sulfate. It is one of the most widespread forms of groundwater contamination in coastal aquifers [14-15].

Coastal Bangladesh is especially prone to groundwater salinity and contamination. The coastal region covers about 46,271 km^2, about 31% of the country. It stretches 37-195 km inland [16]. Soil and water salinity persist in many coastal districts, especially in the southeast and southwest [17-18]. Seawater intrusion, tidal effects, climate variability, and heavy groundwater use exacerbate these problems. As a result, groundwater from both shallow and deep wells in coastal zones is often brackish or saline. This makes it less suitable for drinking. The Department of Public Health Engineering **(**DPHE) considers salinity above 1000 mg/L (1 ppt) as problematic for drinking water in these areas [19].

Apart from salinity, groundwater in Bangladesh often contains natural contaminants such as arsenic and iron. The country has about 8.6 million tube wells. National surveys show about 20% of these wells have arsenic above safe limits [20]. Studies estimate that 35 to 77 million people have been chronically exposed to arsenic through drinking water. This makes it one of history's largest mass poisoning events [21-22]. In addition, Excessive

iron levels are also common. About 41% of studied areas have iron above 1 mg/L [23]. These contaminants reduce drinking water quality and pose long-term health risks.

Rapid urbanization and population growth have further intensified the pressure on groundwater resources in major cities. Chattogram, Bangladesh's second-largest city and commercial hub, has experienced substantial urban expansion over the past few decades. The city's built-up area is increased by nearly 250% from 1974 to 2017. This growth reflects rapid economic development and population increase [24]. The municipal water supply, managed by the Chattogram Water Supply and Sewerage Authority (CWASA), currently provides about 323 million liters per day (MLD). This meets only about 56% of the estimated demand of 578 MLD. As a result, many residents in peripheral and coastal areas rely heavily on groundwater from tube wells for domestic purposes. However, in several coastal zones, groundwater is often unsuitable for drinking due to high salinity and chemical contamination.

Water quality in Bangladesh is a critical concern due to widespread contamination from heavy metals, pesticides, and fecal coliform, alongside emerging threats like antibiotics and microplastics [25]. Research across regions like Rajshahi and coastal belts utilizes WQI and multivariate statistical tools, such as Principal Component Analysis (PCA) and Cluster Analysis (CA), to determine the suitability of water for drinking and irrigation [26]. Studies indicate that parameters like iron, manganese, and magnesium frequently exceed desirable limits in both rural and urban groundwater, while coastal areas face escalating salinity intrusion (a 26% increase over 35 years) driven by sea-level rise and anthropogenic factors [27-28]. Despite some areas maintaining "good" ratings, approximately 67% of rural households still rely on poor-quality water, emphasizing the urgent need for comprehensive monitoring [29]. A significant research gap remains in the integrated use of ML and Geographic Information Systems (GIS) to predict non-linear relationships among contaminants and visualize spatial distribution in highly vulnerable coastal aquifers. This study addresses this gap by employing ML and GIS-based tools to uncover complex water quality dynamics and support sustainable management. The primary objectives are to compare water quality between shallow tube wells (STWs) and Deep Tube wells (DTWs) in Patenga, estimation of the WQI for drinking suitability, development of predictive ML models, and eventually using GIS to map the spatial distribution of the contaminants.

2 Study area

The study area, in Muslimabad (Ward No. 40 and 41) under Uttar Patenga, a coastal region of Chattogram, includes additional sampling sites in Katghar, from Hindupara to 9 No. Ghat along the Karnaphuli River, to assess groundwater quality variation. The sampling locations' (Fig. 1) coordinates range from 22.2571° to 22.26923° in latitude and from 91.78039° to 91.81136° in longitude. The study area was chosen to capture groundwater salinity variation along a gradient from the Bay of Bengal coastline to the inland Karnaphuli River. This transect-based approach evaluates salinity intrusion and its impact on coastal groundwater quality.

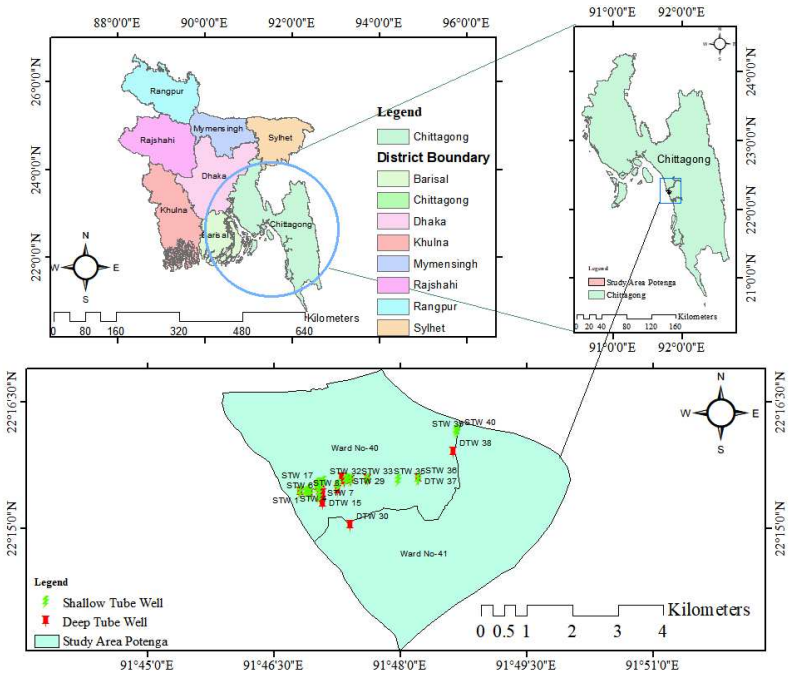


Fig. 1 Study Area and Spatial Distribution of Sampling Sites

To investigate spatial variations, 40 groundwater samples were collected, 32 from Muslimabad and 8 from Katghar, spanning Hindupara to 9 No. Ghat. Muslimabad's higher sampling density reflected its greater population and groundwater use. Among the sampled locations, 22 were identified as STWs and 18 as DTWs.

3 Methodology

This section outlines the methodological framework (Fig. 2) adopted for the assessment of groundwater quality in the coastal area of Patenga. It describes the procedures for sample collection, preservation, and analytical methods employed in the study.

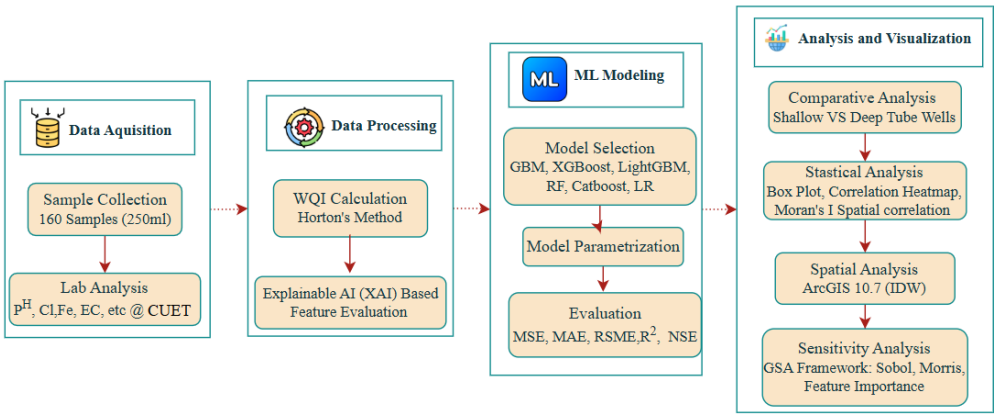


Fig. 2 Methodological Flow Chart of this Study

3.1 Water sample collection

Groundwater samples (Fig. 3) were collected in clean 250 mL screw-capped plastic bottles. All bottles were new or thoroughly washed to ensure freedom from contaminants. Bottles were cleaned with detergent, then rinsed multiple times with distilled water. Bottles were filled completely during sampling to avoid air space, and caps were sealed immediately to prevent contamination. Four samples per location were collected, for a total of 160. Each bottle was clearly labeled with sampling location, date, and time. Samples were transported to the laboratory of Chattogram University of Engineering and Technology under controlled conditions to maintain sample integrity.



Fig. 3 Water sample collection from different locations of the study area

3.1.1 Water quality assessment at lab

To evaluate groundwater characteristics in the study area, several physicochemical parameters were analyzed (Fig. 4) in the laboratory. The analytical procedures followed standard methods; however, certain steps involved visual estimation, which may introduce minor uncertainties in the results. The parameters analyzed in this study include pH, chloride, iron, alkalinity, TDSs, total hardness, electrical conductivity (EC), Salinity, and arsenic.



Fig. 4 Laboratory testing procedures

3.2 Water quality index (wqi) calculation

WQI for the study area was determined using nine parameters: pH, chloride, iron, alkalinity, TDSs, total hardness, EC, salinity, and arsenic. The WQI was calculated using the weighted arithmetic method proposed by Horton through the following three (1,2 and 3) equations [30]:

$$Q_i = \frac{M_i - L_i}{S_i - L_i} * 100 \quad (1)$$

$$W_i = \frac{K}{S_i} \quad (2)$$

$$WQI = \frac{\sum_{i=1}^{n} W_i Q_i}{\sum_{i=1}^{n} W_i} \quad (3)$$

In these equations, Q_i represents the sub-index of the i^{th} parameter, w_i denotes the unit weight of the i^{th} parameter, n is the total number of parameters considered, M_i is the observed value of the parameter, L_i is the ideal value, and S_i is the standard permissible value of the i^{th} parameter. The ideal values were taken as 7 for pH and 14.6 mg/L for dissolved oxygen (DO), while all other parameters were considered zero (0), reflecting the optimal or safe condition for drinking water [31-32].

The unit weight (W_i) for each parameter was calculated as inversely proportional to its corresponding standard value (S_i) recommended by the World Health Organization

(WHO) and environmental conservation rule (ECR). Table 1 presents the classification of water quality into different categories based on the calculated WQI ranges.

Table 1 Classification of Water Quality

WQI Range	*0-25*	*26–50*	*51-75*	*76-100*	*>100*
Water Quality	*Excellent*	*Good*	*Poor*	*Very Poor*	*Unsuitable for drinking*

3.3 Explainable ai (xai) based feature evaluation

This study examines the influence of each feature using XIA with the Linear Regression model. This XIA process aims to make the decision-making process transparent, fostering trust and enhancing the interpretability of complex algorithmic outputs. In this research, one well-known XIA techniques, SHAP, was applied. Lime first introduced by Lundberg and Lee [33], explains the model's prediction by computing the contribution of each feature, considering the overall output. By integrating this technique, the interpretability and reliability of WQI prediction are significantly enhanced.

3.4 Ml models for groundwater quality prediction

This study uses six ensemble regression models-Linear Regression (LR), GBM, XGBoost, Light GBM, Random Forest (RF), and Cat Boost-to predict groundwater quality. GBM is valued for its iterative boost of model accuracy, XGBoost for its regularization and computational efficiency, Light for handling large datasets with low memory consumption, RF regression for its ability to reduce overfitting, and Cat Boost for managing categorical data and minimizing bias. All these models effectively capture complex nonlinear relationships, handle noisy and heterogeneous datasets, support systematic hyperparameter tuning, and enable variable importance assessment, ensuring reliable and robust prediction of groundwater quality in the Patenga coastal aquifers.

3.4.1 Model parameterization

Hyperparameters play a critical role in optimizing model performance and controlling the training process. Table 2 summarizes the hyperparameters employed in this study. Inappropriate selection or tuning of these hyperparameters may lead to overfitting or underfitting, and in some cases, can slow down the training process, ultimately resulting in suboptimal predictive performance.

Table 2 *Parameters of ML models used to predict WQI*

Model Name	Description of Parameters
GBM	n_estimators, learning_rate, max_depth, subsample, max_features, min_samples_split, min_samples_leaf, loss

XGBoost	objective, n_estimators, learning_rate, max_depth, subsample, colsample_bytree, reg_alpha, reg_lambda, random_state, n_jobs
Light GBM	boosting_type, objective, metric, num_leaves, learning_rate, subsample, subsample_freq, colsample_bytree, verbose, random_state, n_estimators
RF Regression	n_estimators, max_depth, random_state, n_jobs
CatBoost	iterations, learning_rate, depth, l2_leaf_reg, border_count, loss_function, verbose, random_state, eval_metric
LR	-

3.4.2 Performance analysis

Model performance was evaluated using multiple statistical metrics, including Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), the coefficient of determination (R^2), and Nash-Sutcliffe Efficiency (NSE). The mathematical formulations of these metrics are presented in Table 3. A model exhibiting lower MSE, MAE, and RMSE values, along with higher R^2, and NSE values, is considered to have superior predictive performance.

Table 3 *Framework for Model Evaluation and Performance Metrics*

Name	Equation	Meaning
MSE	$MSE = \frac{1}{n}\sum_{i=1}^{n}(Yi - Yi)^2$	A lower MSE indicates a better-fitting model.
MAE	$MAE = \frac{1}{n}\sum_{i=1}^{n}(Yi - Yi)$	It's less sensitive to large errors than MSE.
RMSE	$RSME = \sqrt{\frac{\sum_{i=1}^{n}(Yi - Yi)^2}{n}}$	It gives more weight to larger errors.
R^2	$R^2 = 1 - \frac{\sum_{i=1}^{n}(Yi - Yi)^2}{\sum_{i=1}^{n}(Yi - \overline{Y})^2}$	It ranges from 0 to 1, with a value closer to 1 indicating a better fit.
NSE	$NSE = 1 - \frac{\sum_{i=1}^{n}(O_i - S_i)^2}{\sum_{i=1}^{n}(O_i - O^-)^2}$	An NSE value of 1 indicates perfect model performance.

3.5 Spatial Analysis of Groundwater Quality

The analyzed groundwater parameters, including EC, TDSs, chloride, alkalinity, hardness, iron, arsenic, pH, and salinity, were mapped using ArcGIS (version 10.7). The Inverse Distance Weighted (IDW) method was used to create surface maps, in which each grid value reflects the influence of nearby sample points, with points closer to the grid cell having greater impact. This method helps visualize where contaminant levels are higher in the area.

3.6 Sensitivity analysis

To assess how input variables affect model output, a global sensitivity analysis (GSA) framework was used with a trained Gradient Boosting surrogate model. Gradient Boosting was employed as the surrogate model for global sensitivity analysis, as it captures non-linear parameter interactions necessary for meaningful Sobol and Morris index computation [34], [35], [36]. Three complementary methods ensured robust results: Sobol analysis quantified first-order (S1, individual input effects) and total-order (ST, effects including interactions) contributions to output variance; the Morris method provided fast global sensitivity estimates with overall parameter importance (μ*) and interaction/nonlinearity variability (σ); and model-based feature importance reflected how much each input reduced prediction error from the Gradient Boosting algorithm. The surrogate model enabled efficient large-scale Monte Carlo sampling for stable sensitivity estimates. A comparative ranking framework integrated and clarified results across all three methods.

4 Results & discussion

The results of this study provide a comprehensive understanding of groundwater quality characteristics, spatial variability, and predictive modeling performance in the coastal aquifers of Patenga, Chattogram. By integrating hydrochemical analysis, statistical evaluation, spatial mapping, and machine learning, the findings highlight the key factors influencing groundwater quality and their relative importance. The following sections present a detailed discussion of the comparative behavior of shallow and deep aquifers, the interrelationships among physicochemical parameters, spatial distribution patterns, and predictive model performance, offering critical insights into the underlying hydrogeochemical processes and water-quality dynamics of the study area.

The main finding (Fig. 5) reveals that DTWs (strainers >90 ft, up to 160 ft) exhibit higher chloride (5635 mg/L), EC (17260 μS/cm), and salinity (10.18 ppt) levels than STWs, suggesting greater salinity intrusion in Muslimabad's coastal aquifer. Despite minimal differences in most water quality parameters, these specific increases in deep wells highlight the impact of over-extraction, which lowers freshwater hydraulic head and enables denser saline water movement. Additionally, natural salt dissolution from deeper formations may elevate deep well ion concentrations. Elevated salinity and chloride can give water a salty taste and potentially exceed WHO or national drinking water standards. Prolonged consumption may increase the risk of hypertension and other health issues.STWs show higher TDS (13.49 mg/L), hardness (270 mg/L), and arsenic (0.1 mg/L) levels than DTWs because they are closer to surface sources and have different aquifer geochemistry. Shallow aquifers get more recharge from surface water, which

carries dissolved minerals, agricultural runoff, and other contaminants. Less natural filtration means more calcium, magnesium, TDSs, and arsenic are present in these wells. DTWs use confined aquifers, where longer filtration through soil or rock removes many surface contaminants, though mineral content may differ. High hardness causes scaling in pipes and appliances, reducing efficiency. High TDS levels affect taste and suggest the presence of other contaminants. Arsenic is highly toxic and can cause cancer and skin lesions with chronic exposure. These issues degrade water quality for drinking, irrigation, and industrial use, making treatment such as reverse osmosis or softening necessary.

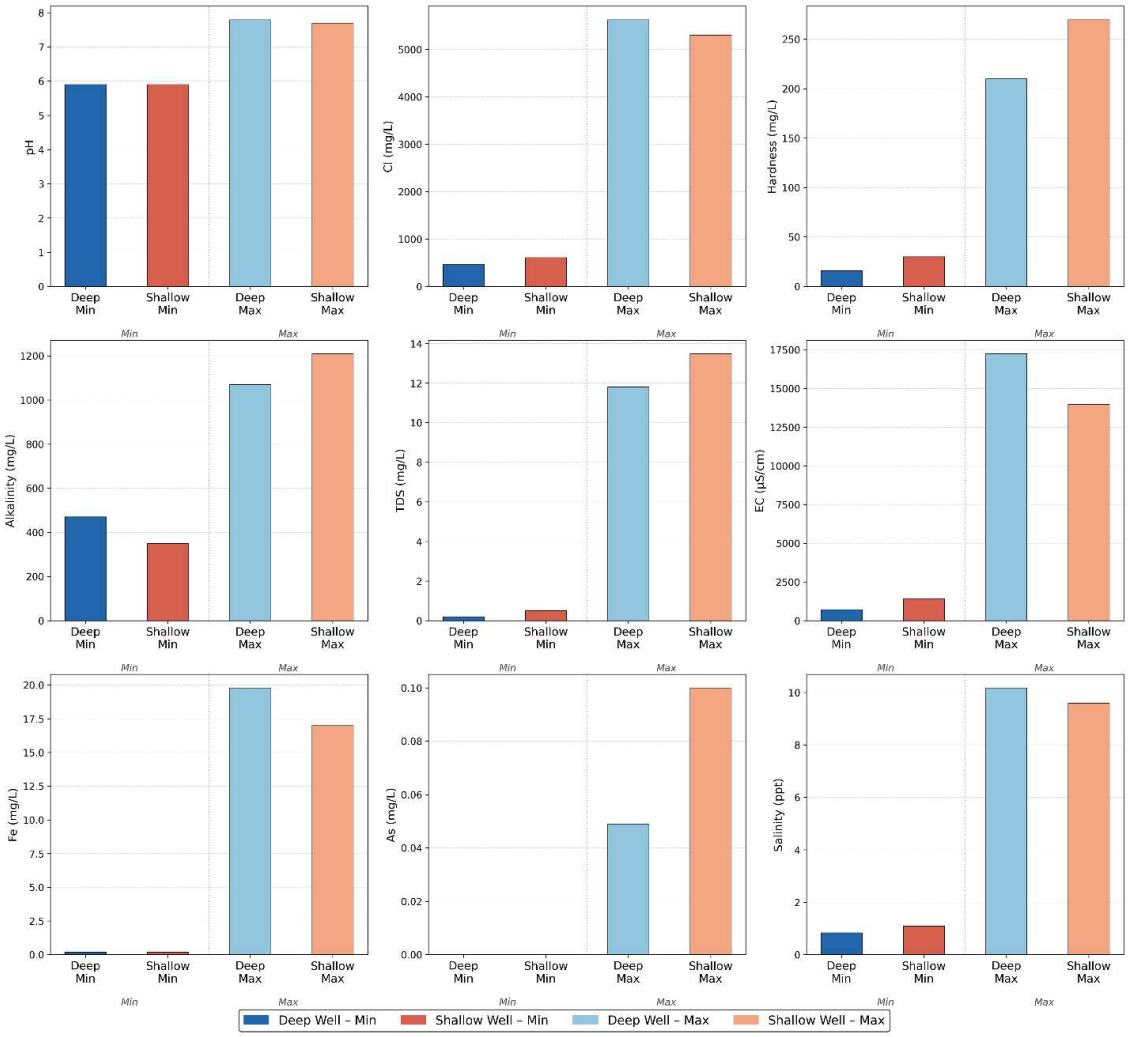


Fig. 5 Comparative Analysis of Groundwater Quality Parameters between Shallow and Deep Tube Wells

The Pearson correlation matrix (Fig. 6) shows the relationships among groundwater quality parameters, including pH, chloride, hardness, alkalinity, TDSs, EC, iron, arsenic, salinity, and the WQI. The correlation coefficients range from - 0.22 to 1.00, reflecting weak to strong associations. Notably, a perfect positive correlation exists between

chloride and salinity (r = 1.00), suggesting that chloride largely controls salinity and indicating seawater intrusion in the coastal aquifer. Furthermore, TDS shows moderate correlations with both chloride (r = 0.54) and salinity (r = 0.54), suggesting that dissolved solids primarily result from mixing with saline water. In addition, chloride shows moderate correlations with hardness (r = 0.44) and EC (r = 0.44), while hardness is moderately correlated with TDS (r = 0.41). Collectively, these patterns suggest the influence of mineral dissolution and ionic enrichment processes in groundwater chemistry.

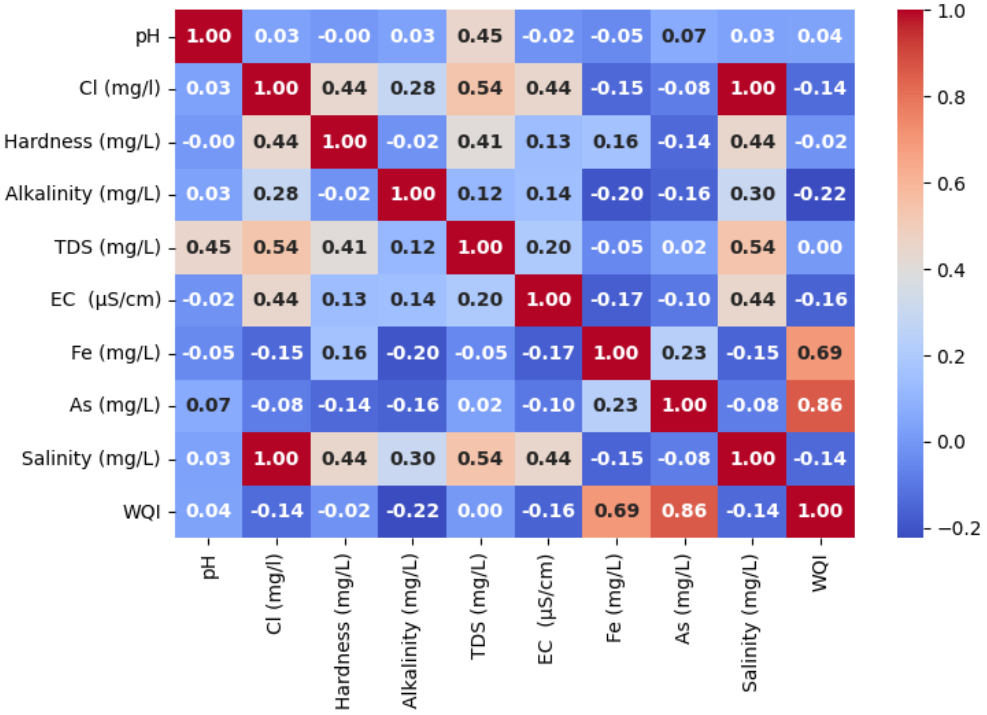


Fig. 6 Correlation Heatmap of Physicochemical Parameters and WQI for Groundwater in Patenga Coastal Aquifer

Among the trace elements, arsenic and iron have strong positive correlations with WQI, with coefficients of 0.86 and 0.69, respectively. This means these parameters mainly contribute to groundwater quality deterioration in the study area. In contrast, TDS shows no correlation with WQI (r = 0.00), suggesting that TDS is not a dominant factor in determining overall water quality. Similarly, pH exhibits very weak correlations with most parameters (r ranging from -0.05 to 0.07), indicating that pH is relatively independent and does not significantly influence other hydrochemical variables. On the other hand, negative correlations observed between WQI and alkalinity (r = -0.22) and EC (r = -0.16) suggest a slight inverse relationship. Likewise, Iron shows weak negative correlations with EC (r = -0.17) and chloride (r = -0.15), whereas Arsenic shows negative correlations with hardness (r = -0.14) and alkalinity (r = -0.16). This indicates that these elements may originate from different geochemical processes and are not directly associated with carbonate chemistry.

Overall, correlation analysis reveals that groundwater quality in Patenga's coastal aquifers is mainly affected by salinity intrusion and trace metal contamination. Specifically, strong associations among chloride, salinity, and TDS confirm seawater mixing. In contrast, high correlations between Arsenic and Iron and WQI indicate their critical role in degrading water quality. Together, these findings offer important insight into the hydrogeochemical processes affecting the aquifers and help select key variables for machine-learning-based groundwater quality prediction.

The box plots (Fig. 7) show the statistical distribution, central tendency, dispersion, and outliers for groundwater quality parameters. These include pH, chloride, total hardness, alkalinity, TDSs, EC, iron, arsenic, salinity, and the WQI. The pH values indicate that the groundwater environment is generally slightly acidic to near-neutral, with a median around 6.4-6.5, while a few localized alkaline conditions are also observed. In addition, chloride concentrations are notably high, with a median near 2500-3000 mg/L, indicating significant salinity intrusion in coastal aquifers and considerable spatial variability. Furthermore, total hardness ranges from approximately 100-120 mg/L, suggesting the water is predominantly moderately hard to hard, with some areas showing higher mineral enrichment. Similarly, alkalinity ranges from 800 to 850 mg/L, indicating a strong buffering capacity of the groundwater system and heterogeneous geochemical conditions. TDS values generally range from 3 to 6 ($\times 10^3$ scale), reflecting moderate to high dissolved solids levels and indicating mineral-rich groundwater in many locations. Correspondingly, EC shows a median of 7000-8000 µS/cm, confirming a high ionic content, a strong relationship with salinity and TDS, and substantial spatial variability. While iron concentrations are typically low, they can be elevated in certain locations, suggesting localized contamination or reducing conditions within the aquifer. Likewise, arsenic concentrations are generally low, although some values exceed standard drinking water limits, indicating potential health risks in specific areas. Overall, salinity ranges from 4500 to 5000 mg/L, clearly confirming significant saline intrusion into the coastal aquifer system. The WQI is highly skewed. Most values are below 50, showing generally acceptable water quality in many locations. However, several extreme outliers exceed 200, indicating severely degraded water quality in certain areas.

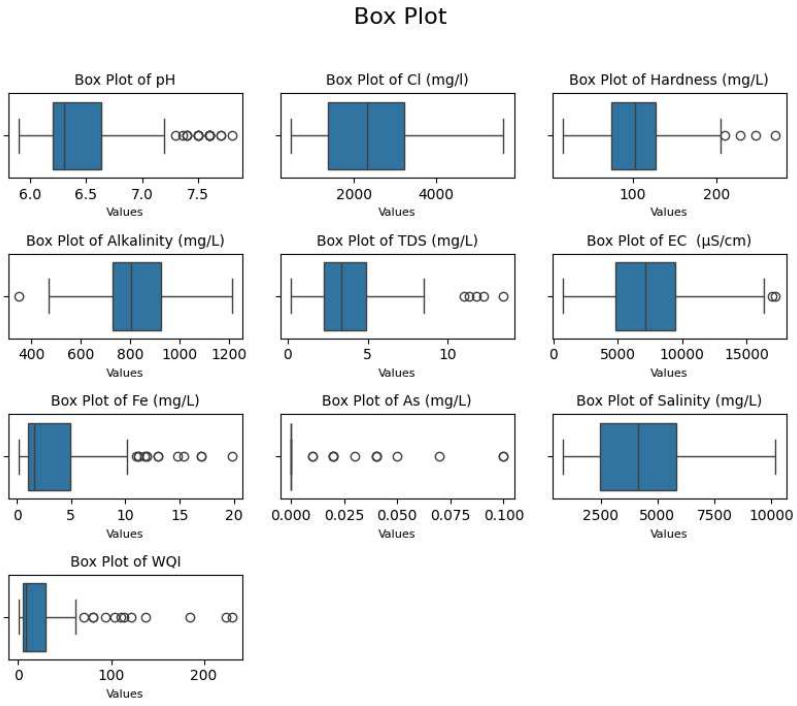


Fig. 7 Statistical Distribution and Variability of Hydrochemical Parameters and WQI Using Box Plot Analysis

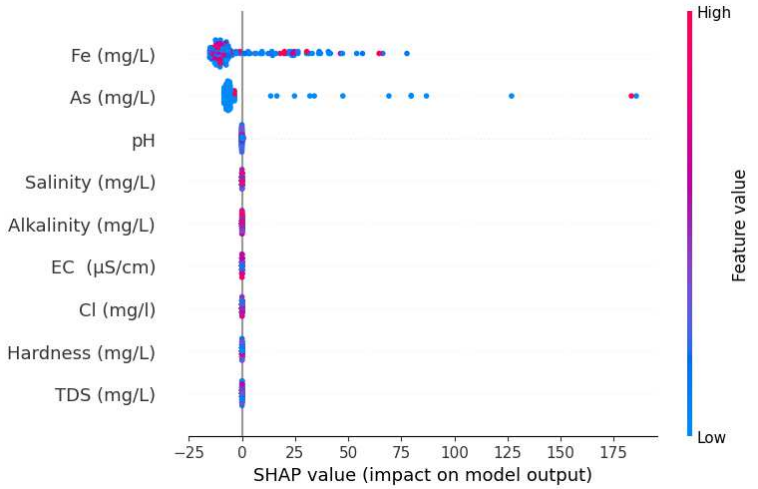


Fig. 8 Feature Importance and Impact Visualization of water quality parameters

Fig. 8 presents the SHAP (Shapley Additive exPlanations) summary plot, which quantifies the contribution of each hydrochemical parameter to the prediction of the Water Quality Index (WQI). The results indicate that arsenic (As) and iron (Fe) are the most influential predictors, with the widest distribution of SHAP values among all input variables. Arsenic has the greatest impact on model output, with SHAP values ranging from approximately −5 to +190, indicating that elevated arsenic concentrations may substantially increase the predicted WQI. Iron similarly exhibits a broad SHAP value range of approximately −15 to

+90, indicating a strong influence on model predictions. The high positive SHAP values for As and Fe suggest that these parameters are the primary drivers of groundwater quality variation within the study area.

In contrast, the remaining parameters, including pH, salinity, alkalinity, electrical conductivity (EC), chloride, hardness, and TDS, display SHAP values concentrated around zero, generally within a narrow range of approximately −2 to +2. This indicates that their marginal contributions to WQI prediction are comparatively small. The color gradient further shows the relationship between feature magnitude and model response, with red points indicating higher feature values and blue points indicating lower values. For arsenic and iron, higher concentrations are predominantly associated with positive SHAP values, indicating a positive contribution to the predicted WQI. Overall, the SHAP analysis shows that the predictive model relies primarily on arsenic and iron concentrations, while the remaining hydrochemical parameters contribute relatively little to the model output. These findings highlight the dominant role of trace metal contamination in controlling groundwater quality conditions in the study area.

Table 4 presents the performance of six machine learning and statistical models in predicting the WQI from groundwater quality parameters, including tube well type, latitude, longitude, elevation (m), distance from coast, iron (mg/L), and arsenic (mg/L). Model performance was assessed using MAE, MSE, RMSE, R^2, and NSE. LR achieved the best predictive performance, with the lowest error values (MAE = 0.14, MSE = 0.03, RMSE = 0.18) and highest accuracy indicators (R^2 = 0.99, NSE = 0.99). This reflects excellent agreement between observed and predicted WQI values. The RF model performed well (R^2 = 0.97, RMSE = 6.34), with CatBoost and GBM following closely (R^2 = 0.97, 0.96 respectively). In contrast, Light GBM exhibited the weakest predictive ability, with the highest errors (MAE = 10.91, RMSE = 20.60) and the lowest R^2 and NSE (0.69), indicating poor generalization on this dataset.

Table 4 *Level one out cross validation Performance Metrics*

Model	MAE	MSE	RMSE	R^2	NSE
Random Forest	1.99	40.17	6.34	0.97	0.97
CatBoost	2.74	47.60	6.90	0.97	0.97
Light GBM	10.91	424.42	20.60	0.69	0.69
GBM	2.39	52.53	7.25	0.96	0.96
XGBoost	3.01	100.50	10.02	0.93	0.93
LR	0.14	0.03	0.18	0.99	0.99

Overall, the results suggest that the relationship between the selected hydrochemical parameters and WQI is highly linear, which explains the superior performance of the Linear Regression model compared to more complex machine learning algorithms.

Moran's I spatial autocorrelation means measuring whether similar values are spatially clustered, dispersed, or randomly distributed across a geographic area. Moran's I spatial autocorrelation analysis revealed (Table 5) varying spatial patterns among the

groundwater quality parameters in the study area. Specifically, chloride, hardness, TDS, and salinity exhibited positive spatial autocorrelation with statistically significant p-values ($p < 0.05$). This indicates clustered spatial distributions and suggests the influence of localized hydrogeochemical or saline intrusion processes. Notably, TDS showed the strongest clustering tendency with a Moran's I value of 0.31. In contrast, EC demonstrated significant negative spatial autocorrelation (Moran's I = -0.19, p = 0.001), indicating a dispersed spatial pattern across the study area. On the other hand, pH, alkalinity, iron, arsenic, and WQI showed no statistically significant spatial autocorrelation ($p > 0.05$), and so their spatial distributions were largely random. Overall, these results indicate that certain groundwater quality parameters are spatially clustered, whereas others exhibit random or dispersed spatial behavior within the coastal aquifer system.

Table 5 *Moran's I Spatial Autocorrelation Indices of Groundwater Quality Parameters*

	pH	Cl	Hardness	Alkalinity	TDS	EC	Fe	As	Salinity	WQI
Moran's I	-0.03	0.27	0.26	0.04	0.31	-0.19	-0.06	-0.06	0.27	-0.08

The spatial distribution of groundwater quality parameters (Fig. 9) reveals significant variability across sampling locations. Since each sampling location was monitored on four occasions, the spatial analysis used averaged measurements to represent stable groundwater conditions.

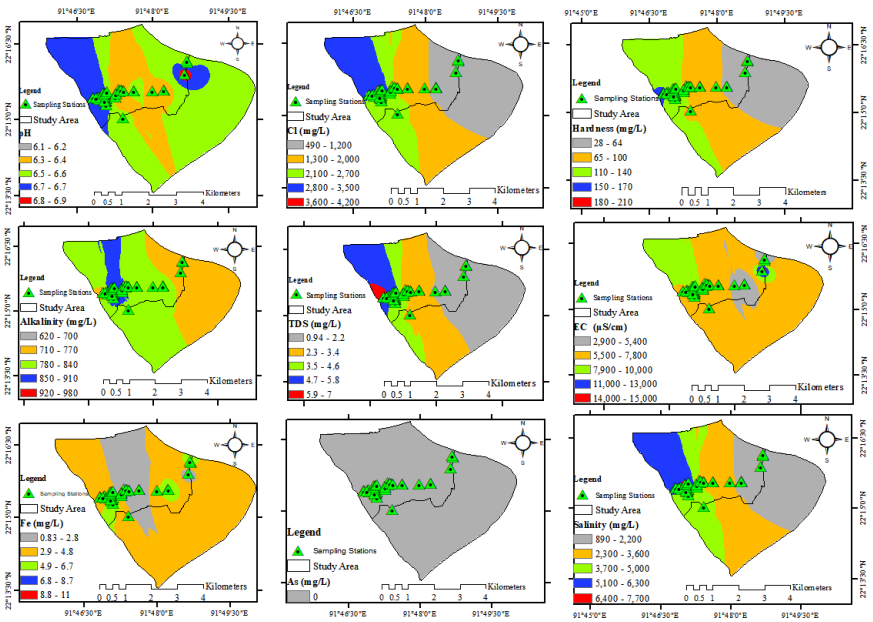


Fig. 9 Spatial distribution of Groundwater Physicochemical Parameters

The pH values ranged from 6.025 to 6.91, indicating a uniformly slightly acidic groundwater environment, with the lowest levels in the north and a gradual increase

toward the southeast. Similarly, chloride concentration varied substantially, ranging from 491 mg/L to 4276.5 mg/L. Elevated concentrations were predominantly observed in the southern coastal zone, a pattern likely attributable to seawater intrusion and proximity to the Karnaphuli River estuary. This distribution is further illustrated by the spatial map, which delineates five distinct concentration zones. Notably, the highest values (3,600-4,200 mg/L) were confined to a small, isolated hotspot in the central-southern portion of the study area. In addition, total hardness ranged from 27.75 to 237.25 mg/L, indicating a transition from soft to moderately hard water and considerable heterogeneity in calcium and magnesium concentrations across the aquifer. Higher values (150-210 mg/L) specifically occurred in the eastern and southeastern portions of the study area. Moving to alkalinity, values ranged from 592.5 mg/L to 990 mg/L, indicating a high groundwater buffering capacity. According to the distribution map, the highest alkalinity values (920-980 mg/L) were concentrated in a small central zone, whereas moderate values (620-700 mg/L) predominated in the northern and western portions of the study area. Furthermore, TDSs concentrations ranged from 0.94 mg/L to 7.33 mg/L, with elevated values in the southern and coastal sectors, consistent with the EC and salinity patterns and saline intrusion. Similarly, EC values ranged from 2,843.75 to 15,325 µS/cm, spanning a fivefold range across the study area. The highest EC values (14,000-15,000 µS/cm) occurred in the eastern sector, coinciding with zones of elevated salinity and TDS. EC zonation closely mirrored the salinity distribution, highlighting the influence of tidal and estuarine processes on groundwater chemistry. Iron concentrations ranged from 0.3 mg/L to 12 mg/L, with localized hotspots exceeding permissible limits. Additionally, arsenic ranged from 0 to 0.08 mg/L, with some locations above recommended standards, posing potential risks. Lastly, salinity values ranged from 887.5 mg/L to 7722.5 mg/L, with a clear gradient increasing toward coastal and river-adjacent regions, confirming the influence of tides and saline intrusion.

Overall, the spatial analysis shows groundwater quality deteriorates toward the coastal and riverine zones, with salinity, chloride, TDS, and EC all elevated in these areas. These patterns indicate that both natural hydrogeological processes and human activities jointly influence groundwater quality in the study area.

The Sobol sensitivity (Fig. 10) analysis showed that arsenic was the most important variable (S1 = 0.7485, ST = 0.7580), with a mainly additive effect. Iron was next (S1 = 0.2139, ST = 0.2184) and showed little interaction. Arsenic and iron together explained over 95% of the total output variance. Other variables-EC, hardness, chloride, alkalinity, TDS, salinity, and pH-had almost no influence (ST < 0.02). Slightly negative or near-zero first-order indices for some variables are due to sampling uncertainty and confirm their minimal contribution.

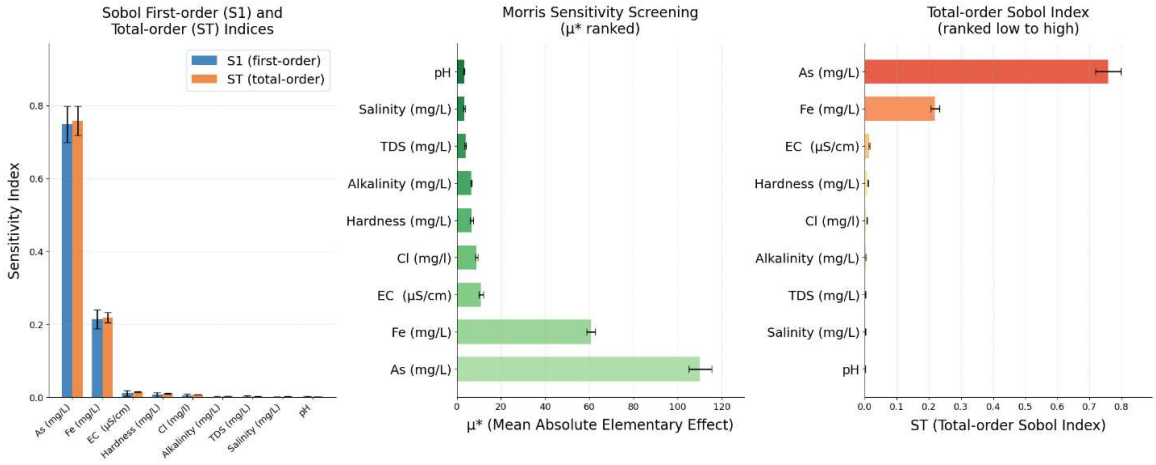


Fig. 10 Sobol sensitivity indices (S1 and ST) and Morris screening (μ*) for groundwater WQI parameters, Patenga, Chattogram

The Morris analysis further underscored the dominance of arsenic and iron, which had the highest mean absolute elementary effects (μ* ≈ 110.28 for arsenic, 60.84 for iron). Their relatively high standard deviation values suggest some nonlinear effects and moderate interactions, though not dominant, consistent with Sobol results. Other variables, such as EC, chloride, and hardness, had moderate μ* values but limited nonlinear effects, while TDS, salinity, and pH had minimal impact.

The Gradient Boosting surrogate model demonstrated strong predictive performance and is suitable for sensitivity analysis. The model achieved a perfect fit on the training dataset (R^2 = 1.00) and generalized well on the test dataset (R^2 = 0.84). Prediction errors were low, with RMSE at 4.10 and MAE at 2.47. The 5-fold cross-validation gave an average R^2 of 0.9252 (±0.035), confirming the model's robustness and stability.

The Gradient Boosting model's feature importance (Fig. 11) aligned with sensitivity results, with arsenic contributing about 50.2% and iron 27.9% of total importance. The other variables each contributed less than 6%, again emphasizing arsenic and iron as dominant. A comparative analysis across Sobol, Morris, and Gradient Boosting methods showed strong agreement, with arsenic and iron consistently ranking as the most influential variables, followed by hardness, EC, and chloride, while alkalinity, TDS, salinity, and pH had lower influence. This cross-method consistency enhances the reliability of the findings.

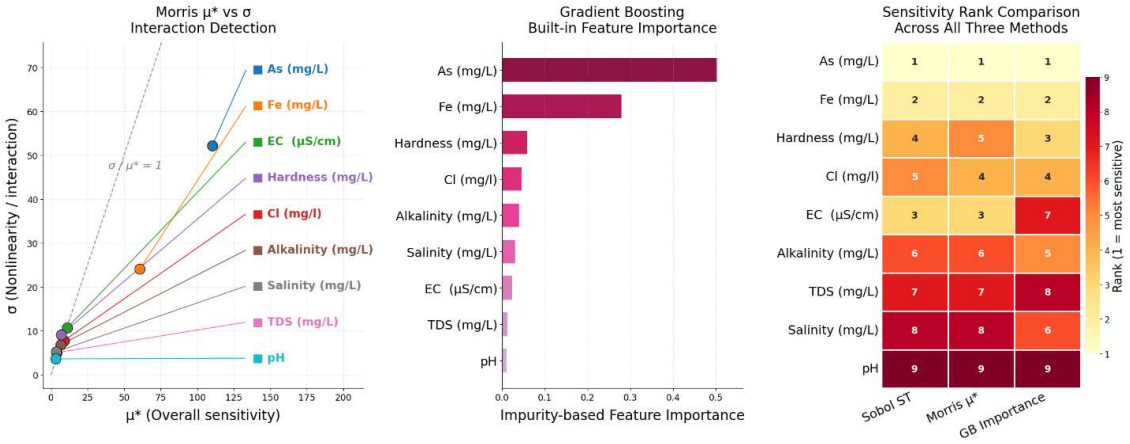


Fig. 11 Feature importance and sensitivity rank comparison of groundwater WQI parameters using Monte Carlo analysis

Overall, the system is highly sensitive to arsenic and iron, with other hydrochemical parameters having minimal impact. The primary additive behavior observed simplifies interpretation. Practically, focusing on these two parameters enables model simplification and efficient data collection. The strong agreement among sensitivity techniques further confirms the robustness of the conclusions and the integrated methodological approach. The spatial distribution of the WQI (Fig. 12), a measure that combines various water quality parameters into a single score to indicate overall suitability for drinking, across the study area demonstrates considerable variability, suggesting differences in groundwater suitability for drinking. The WQI values were classified into five categories: excellent (0-25), good (26-50), poor (51-75), very poor (76-100), and unsuitable for drinking (>100).

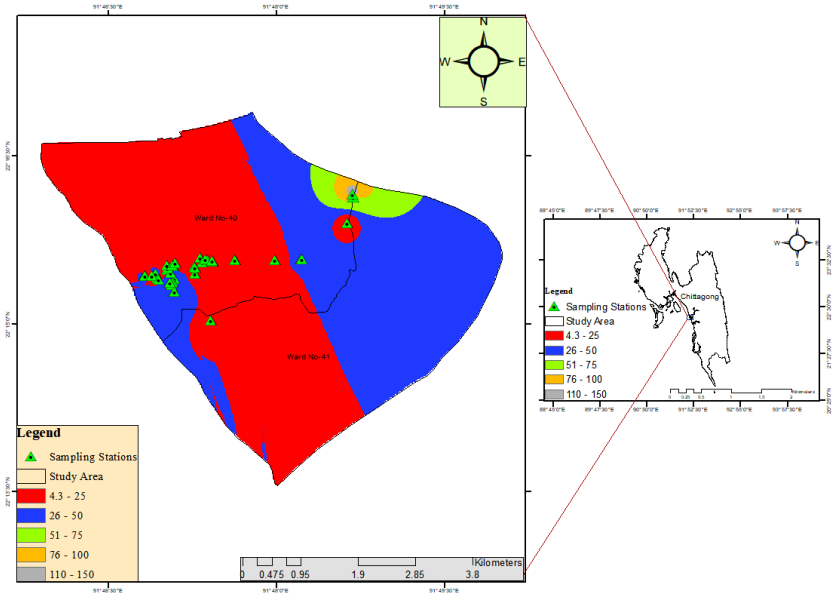


Fig. 12 Spatial Mapping of Drinking Water Suitability Based on WQI Classification

The results reveal that a significant portion of the western and central parts of the study area, particularly in Ward No. 40 and parts of Ward No. 41, falls into the excellent category (WQI: 4.3-25). This indicates that groundwater in these areas is suitable for drinking without requiring major treatment. These regions are likely characterized by relatively low concentrations of dissolved contaminants and stable hydrochemical conditions, where hydrochemical refers to the chemical composition of groundwater, especially concerning dissolved minerals and compounds.

In contrast, the eastern region of the study area is predominantly classified as good quality, suggesting that the groundwater is generally acceptable for consumption, although minor levels of contamination may be present. Additionally, localized zones of poor and very poor water quality are observed, particularly in transitional areas between the western and eastern parts. These variations may be attributed to local hydrogeological (related to groundwater movement and distribution) conditions, anthropogenic (human-caused) influences, or mixing of freshwater with saline (salty) or contaminated sources. A few specific sites have WQI values above 100, making their water unsuitable for drinking. These hotspots likely have high levels of toxic elements such as arsenic and iron, as the sensitivity analysis shows.

Overall, the spatial pattern suggests a gradual deterioration in groundwater quality from west to east, which may be influenced by coastal processes, such as saline water intrusion (the movement of saltwater into freshwater aquifers), as well as by proximity to the Karnaphuli River. The clustering of sampling stations in the central-western region also enhances the reliability of the observed spatial patterns in these areas. In summary, while much of the area has excellent to good groundwater quality, certain eastern and hotspot locations are unsuitable for drinking and need treatment. These findings underscore the importance of targeted management and ongoing monitoring.

5 Conclusion

This study assesses groundwater quality in Patenga's coastal aquifers using spatial analysis, hydrochemical evaluation, and machine learning. Results show groundwater quality varies greatly and is controlled by natural processes and human activities. A comparative analysis of STWs and DTWs reveals distinct hydrochemical behaviour across aquifer depths. DTWs exhibit much higher salinity indicators (chloride, EC, and salinity), confirming more intense seawater intrusion in deeper coastal aquifers, especially in Muslimabad. In contrast, STWs are more vulnerable to surface-derived contamination. This is shown by higher TDS, hardness, and arsenic concentrations. These depth-dependent variations highlight the complexity of groundwater systems in coastal regions. They also challenge the conventional assumption that deeper aquifers always provide safer drinking water. The WQI analysis indicates that much of the study area falls within acceptable drinking water limits. However, some local zones remain critically polluted and unsuitable. Strong positive correlations between arsenic and iron, and between these metals and the WQI, highlight their key roles in degrading groundwater quality and posing health risks. Also, the strong links among chloride, salinity, and TDS support seawater intrusion as a key mechanism in the area. Spatial patterns from GIS analysis show groundwater quality worsens near coastal and river zones. This trend reflects the effects

of saline water encroachment, land-use practices, and possible over-extraction of groundwater. These findings are especially relevant with rapid urbanization and climate-driven sea-level rise, which are expected to worsen salinity intrusion in coastal Bangladesh. From a modeling perspective, the Gradient Boosting algorithm performs best. This shows that machine learning can reliably predict groundwater quality. Consistency across Sobol, Morris, and Gradient Boosting sensitivity analyses also strengthens the findings. It identifies arsenic and iron as the most influential parameters governing groundwater quality.

These findings highlight serious water security challenges for Bangladesh's coastal communities. Arsenic and salinity in drinking water directly threaten health, agricultural productivity, and sustainable livelihoods. Without proper management, these issues may worsen. This could lead to more dependence on alternative water sources and higher economic burdens for vulnerable populations. This study strongly aligns with United Nations Sustainable Development Goal (SDG) 6: Clean Water and Sanitation. It emphasizes the need for safe and sustainable groundwater resources. The work also indirectly supports SDG 3 (Good Health and Well-being) and SDG 13 (Climate Action). This is due to the health risks posed by contaminated water and to climate change's role in accelerating coastal salinity intrusion. A limitation of this study is the relatively small sample size used for training and validating the ML models, which may affect the robustness and generalizability of the predictive performance. Future research should collect groundwater quality data more frequently across seasons to improve model accuracy and reliability, with particular attention to the effects of wet and dry seasons on the WQI in the Muslimabad coastal aquifer.

In conclusion, combining GIS-based spatial mapping with advanced machine learning models provides a robust framework for assessing and predicting groundwater quality. The study recommends targeted mitigation strategies. These include controlled groundwater abstraction, regular monitoring of arsenic and salinity, and the development of alternative, safe water supply systems. Such measures are needed to ensure long-term water security and support sustainable development in vulnerable coastal regions.

Acknowledgments

The authors would like to acknowledge the support provided by the laboratory facilities of Chittagong University of Engineering & Technology (CUET) for conducting the groundwater quality analysis. The authors are grateful for the technical assistance and laboratory support provided during this research.

Author contributions

M.H.A. Conceptualization, data analysis, methodology, writing-original draft preparation, and editing; S.Z. Data collection, writing-original draft preparation, and review; M.A. Data collection and writing-original draft preparation; T.H.T. Review and critical revision of the manuscript; A.H. Review and critical revision of the manuscript; M.R.A.M: Review and critical revision of the manuscript; S.A. Review and critical revision of the manuscript.

All authors read and approved the final version of the manuscript.

Funding

No funding was received for conducting this study.

Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics declaration: not applicable.

Consent for publication

Informed consent was obtained from all the individual participants for publication.

Competing interest

The authors declare that they have no competing interests.

References

1. Leal Filho, W., et al., Climate change adaptation responses among riparian settlements: A case study from Bangladesh, *PLoS One*, 2022, vol. 17, no. 12, p. e0278605.

2. Singh, V., Sharma, N., and Ojha, C.S.P., *The Brahmaputra Basin Water Resources*, Springer Science & Business Media, 2004.

3. Barceló, D., Collection on groundwater to celebrate the World Water Day March 20, 2022, *MethodsX*, 2022, vol. 9, p. 101648.

4. Udmale, P., Shrestha, S., Ichikawa, Y., and Manandhar, S., Assessing groundwater resource and its sustainability in drought prone area of India, *Journal of Japan Society of Civil Engineers, Series B1 (Hydraulic Engineering)*, 2014, vol. 70, no. 4, pp. I_235–I_240.

5. Zektser, I.S. and Everett, L.G., *Groundwater Resources of the World and Their Use*, UNESCO, 2004.

6. Hossain, M.S., Roy, K., and Datta, D.K., Spatial and temporal variability of rainfall over the South-West coast of Bangladesh, *Climate*, 2014, vol. 2, no. 2, pp. 28–46. https://doi.org/10.3390/cli2020028

7. Ahmed, A.M.M., *Utilization and Conservation of Water Resources in Bangladesh*.

8. Rahman, R. and Salehin, M., Flood risks and reduction approaches in Bangladesh, in *Disaster Risk Reduction Approaches in Bangladesh*, Springer, 2013, pp. 65–90.

9. Rajmohan, N. and Prathapar, S.A., *Hydrogeology of the Eastern Ganges Basin: An Overview*, International Water Management Institute (IWMI), 2013. https://doi.org/10.5337/2013.216

10. Boral Doal, D., Rahman, Md.M., and Higuchi, H., Groundwater use efficiency for sustainable agriculture: Prospects in Bangladesh, *Big Data Water Resources Engineering*, 2023, vol. 4, no. 1, pp. 11–15. https://doi.org/10.26480/bdwre.01.2023.11.15

11. Zahid, A. and Ahmed, S.R.U., Groundwater resources development in Bangladesh: Contribution to irrigation for food security and constraints to sustainability, *Groundwater Governance in Asia Series*, 2006, vol. 1, pp. 27–46.

12. Akter, T., et al., Water Quality Index for measuring drinking water quality in rural Bangladesh: A cross-sectional study, *Journal of Health, Population and Nutrition*, 2016, vol. 35, no. 1, p. 4.

13. Chidambaram, S., Senthil Kumar, G., Prasanna, M.V., John Peter, A., Ramanathan, A.L., and Srinivasamoorthy, K., A study on the hydrogeology and hydrogeochemistry of groundwater from different depths in a coastal aquifer: Annamalai Nagar, Tamil Nadu, India, *Environmental Geology*, 2009, vol. 57, no. 1, pp. 59–73. https://doi.org/10.1007/s00254-008-1282-4

14. Dixon, W. and Chiswell, B., The use of hydrochemical sections to identify recharge areas and saline intrusions in alluvial aquifers, southeast Queensland, Australia, *Journal of Hydrology*, 1992, vol. 135, no. 1, pp. 259–274. https://doi.org/10.1016/0022-1694(92)90091-9

15. Giménez, E. and Morell, I., Hydrogeochemical analysis of salinization processes in the coastal aquifer of Oropesa (Castellón, Spain), *Environmental Geology*, 1997, vol. 29, no. 1, pp. 118–131. https://doi.org/10.1007/s002540050110

16. Islam, M.R., Managing diverse land uses in coastal Bangladesh: Institutional approaches, in *Environment and Livelihoods in Tropical Coastal Zones: Managing Agriculture-Fishery-Aquaculture Conflicts*, Comprehensive Assessment of Water Management in Agriculture, 2006, pp. 237–248. https://doi.org/10.1079/9781845931070.0237

17. Shammi, M., et al., Spatio-temporal assessment and trend analysis of surface water salinity in the coastal region of Bangladesh, *Environmental Science and Pollution Research*, 2017, vol. 24, no. 16, pp. 14273–14290. https://doi.org/10.1007/s11356-017-8976-7

18. Bahar, M.M. and Reza, M.S., Hydrochemical characteristics and quality assessment of shallow groundwater in a coastal area of Southwest Bangladesh, *Environmental Earth Sciences*, 2010, vol. 61, no. 5, pp. 1065–1073. https://doi.org/10.1007/s12665-009-0427-4

19. Hepp, P., Nadiruzzaman, M., and Krumeich, A., Contested quantification for planetary health: A sociotechnical analysis of Bangladesh's water salinity monitoring infrastructure, *Social Science & Medicine*, 2024, vol. 360, p. 117312. https://doi.org/10.1016/j.socscimed.2024.117312

20. Johnston, R.B. and Sarker, M.H., Arsenic mitigation in Bangladesh: National screening data and case studies in three upazilas, *Journal of Environmental Science and Health, Part A*, 2007, vol. 42, no. 12, pp. 1889–1896. https://doi.org/10.1080/10934520701567155

21. Kinniburgh, D.G. and Smedley, P.L., *Arsenic Contamination of Groundwater in Bangladesh*, British Geological Survey, 2010.

22. Smith, A.H., Lingas, E.O., and Rahman, M., Contamination of drinking-water by arsenic in Bangladesh: A public health emergency, *Bulletin of the World Health Organization*, 2000, vol. 78, no. 9, pp. 1093–1103.

23. Hossain, M.D. and Huda, M., Study of iron content in groundwater of Bangladesh, *Journal of Civil Engineering*, 1997, vol. 25, pp. 171–179.

24. Hossen, M.A. and Hoque, A., Assessment of rooftop rainwater harvesting system in saline prone area: A case study of Muslimabad Housing Society, Chittagong, 2018.

25. Parvin, F., Haque, M.M., and Tareq, S.M., Recent status of water quality in Bangladesh: A systematic review, meta-analysis and health risk assessment, *Environmental Challenges*, 2022, vol. 6, p. 100416. https://doi.org/10.1016/j.envc.2021.100416

26. Akter, T., et al., Water Quality Index for measuring drinking water quality in rural Bangladesh: A cross-sectional study, *Journal of Health, Population and Nutrition*, 2016, vol. 35, no. 1, p. 4. https://doi.org/10.1186/s41043-016-0041-5

27. Ochuko, U., Thaddeus, O., and Oghenero, A., A comparative assessment of Water Quality Index (WQI) and suitability of River Ase for domestic water supply in urban and rural communities in Southern Nigeria, 2014, vol. 4, no. 1.

28. World Health Organization, *Guidelines for Drinking-Water Quality: Fourth Edition Incorporating the First and Second Addenda*, World Health Organization, Geneva, 2022.

29. Rahaman, Md.F., Ali, Md.S., Arefin, R., Mazumder, Q.H., Majumder, R.K., and Jahan, C.S., Assessment of drinking water quality characteristics and quality index of Rajshahi city, Bangladesh, *Environmental Development and Sustainability*, 2020, vol. 22, no. 5, pp. 3957–3971.
https://doi.org/10.1007/s10668-019-00366-9

30. Horton, R.K., An index number system for rating water quality, *Journal of Water Pollution Control Federation*, 1965, vol. 37, no. 3, pp. 300–306.

31. Abbasi, T. and Abbasi, S.A., Water-quality indices, in *Water Quality Indices*, Elsevier, 2012, pp. 353–356. https://doi.org/10.1016/B978-0-444-54304-2.00016-6

32. Tripathy, J. and Sahu, K., Seasonal hydrochemistry of groundwater in the barrier spit system of the Chilika Lagoon, India, *Journal of Environmental Hydrology*, 2005, vol. 13.

33. Lundberg, S.M. and Lee, S.-I., A unified approach to interpreting model predictions, *Advances in Neural Information Processing Systems*, 2017, vol. 30.

34. Razavi, S., et al., The future of sensitivity analysis: An essential discipline for systems modeling and policy support, *Environmental Modelling & Software*, 2021, vol. 137, p. 104954.
https://doi.org/10.1016/j.envsoft.2020.104954

35. Naghibi, S.A., Pourghasemi, H.R., and Dixon, B., GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran, *Environmental Monitoring and Assessment*, 2016, vol. 188, no. 1, p. 44. https://doi.org/10.1007/s10661-015-5049-6

36. Erdal, D. and Cirpka, O.A., Global sensitivity analysis and adaptive stochastic sampling of a subsurface-flow model using active subspaces, *Hydrology and Earth System Sciences*, 2019, vol. 23, no. 9, pp. 3787–3805. https://doi.org/10.5194/hess-23-3787-2019