



EXPLAINABLE MACHINE LEARNING FOR WATER QUALITY ASSESSMENT AND POLLUTION RISK PREDICTION: AN INTEGRATED FRAMEWORK FOR SUSTAINABLE WATER RESOURCE MANAGEMENT

Priyanka Jeetendra Patil

Assistant Professor, Department- CSE-AIML, College/University Name-A.P.Shah Institute of Technology, City-Thane, Email- pjpatil@apsit.edu.in

Received:- 11/05/2026, Revised:- 22/07/2026, Accepted:- 02/09/2026, Published:- 19/09/2026

ABSTRACT

Water-quality degradation threatens aquatic ecosystems, public health, and sustainable resource management, creating a need for accurate and interpretable predictive approaches. This study developed an explainable machine-learning framework for water-quality assessment and pollution-risk prediction using physicochemical, hydroclimatic, spatial, and seasonal variables. A total of 1,495 observations from 58 sampling locations across 13 rivers and streams were analyzed. Water Quality Index (WQI) was modeled as a continuous outcome using Linear Regression, Random Forest, Extra Trees, Gradient Boosting, and Support Vector Regression, while pollution status was classified into Good, Poor, and Very Poor categories using Logistic Regression, Random Forest, Extra Trees, and Support Vector Machine. Five-fold location-grouped cross-validation was applied to reduce spatial leakage. The overall mean WQI was 91.40, and 94.25% of observations were classified as Very Poor. Gradient Boosting achieved the strongest WQI prediction, with a mean R^2 of 0.9993, MAE of 0.1215, and RMSE of 0.3485. Random Forest provided the best pollution classification, attaining 0.9940 accuracy, 0.9061 balanced accuracy, 0.9394 macro-F1, and 0.9414 MCC. Permutation importance identified phosphate as the dominant predictor of WQI, followed by pH and dissolved oxygen. The framework demonstrates that explainable machine learning can provide accurate, transparent, and management-relevant support for sustainable water-resource monitoring.

Key words: Water Quality Index; Explainable Machine Learning; Pollution Risk Prediction; Gradient Boosting; Sustainable Water Resource Management

1. INTRODUCTION

The degradation of water quality is now a significant environmental issue, as the surface-water systems are more affected by urbanisation, industrial effluents, agricultural runoff, hydrological variability and climate related pressures. Assessment that is reliable is therefore required to describe current conditions as well as to anticipate deterioration and support timely management decisions. While most monitoring methods are based on single measurements or composites of measurements (physicochemical), water-quality variables often interact in a nonlinear fashion and are spatially variable. Recent advances in machine learning (ML) present the opportunity to combine multiple indicators and develop prediction systems, which can be more flexible in their analytical approach for estimating water quality conditions. The River Quality Modelling reviews have been carried out, and it has been proved that the relevance of ML in Water Quality Index (WQI) based Prediction and Environmental Monitoring is increasing (Shaheed et al., 2025).

The WQI is a good tool for reducing multidimensional information concerning water quality to a single metric of overall environmental condition. But, to predict the WQI, accurate analytical approaches must be able to detect the nonlinear relationships and complex interactions in the physicochemical and hydroclimatic characteristics. In recent years, gradient boosting and other ML methods have shown promising performance in WQI estimation and water monitoring sustainability (Islam, Salekin, Anzum, et al., 2024). Collaboration of predictive algorithms has been demonstrated to enhance the WQI estimation and aid in the understanding of the water-quality variables that are influential in the process (Choudhary et al., 2025).

Although complex ML systems have the predictive power, they are often criticized for its black box nature. This is a significant issue in environmental applications because environmental managers must know why a given water body is classified as a particular quality or pollution classification. The integration of ML with explainable AI (XAI) can enhance groundwater-quality prediction and create a clearer understanding of information for environmental decision making (Alshehri & Rahman, 2023). Through systematic evidence, transparency, interpretability, and trust have been identified as key prerequisites for water-quality monitoring using XAI (Aderemi et al., 2025b). Predictive models can also be interpreted in a parametric way, which can further show how individual physicochemical properties contribute to variations in the predicted water quality (Nallakaruppan et al., 2024).

The relevance of interpretability to pollution management is particularly important because the ability to predict is not sufficient to provide information on which environmental parameters are related to deterioration. Explainable data-driven approaches have thus been used to water quality assessment to deliver more actionable data for pollution management decisions (Mallick et al., 2024). Additionally, ensemble learning has been suggested to obtain accurate and interpretable water-quality prediction alongside XAI (Hellen et al., 2023). Model calibration and model explainability are equally crucial for the confidence of the environmental decision support application with predictive water-quality systems (Hellen, 2022).

The importance of ML is not just for predicting WQI, it can also be used for ecological assessment and for sustainable resource planning. ML-based ecological-quality prediction can help assess the conditions of rivers and provide guidance for management interventions based on scenarios (Politikos et al., 2024). Explainable modelling has also been used to explore the potential effects of climate change and urban-policy scenarios on water-quality conditions, illustrating the need to consider the environmental context in predictive assessment (Tian et al., 2024). The integration of XAI and ML, on the other hand, can provide a more comprehensive and holistic understanding of water pollution and can help to make more manageable and actionable decisions based on the predictive results. The integration of XAI and ML, however, can create a wider scope of understanding water pollution and can help in transforming complex predictive outcomes into evidence relevant for management decisions (Talukdar et al, 2024).

The purpose of this study was therefore to create an explanatory machine learning-based water-quality assessment and pollution-risk prediction framework. The range included physicochemical, hydroclimatic, seasonal, spatial and river-level characteristics where WQI was considered as a continuous prediction result and pollution status was considered as a three-class classification result. Overall and seasonal water-quality patterns were characterized; several water-quality models based on multiple regression were compared in terms of their water-quality prediction capability; water-quality classification models for pollution-risk discrimination were assessed while dealing with severe class imbalance; and variables contributing most strongly to the prediction capability were identified using permutation-based explainability. The overall goal was to create a clear analytical approach to facilitate sustainable water-resource monitoring and management.

2. MATERIALS AND METHODS

2.1 Study Data and Analytical Variables

Secondary water quality and hydroclimatic observations from rivers and streams in Dhaka, Bangladesh were used for the study, consisting of 1495 observations from 58 sampling locations spread across 13 rivers and streams. The records spanned the time interval from January 2017 to July 2021 and comprised physicochemical, hydroclimatic, spatial, temporal and water-quality classification variables. The data were from the publicly available repository reported by Sinha et al. (2026).

The physicochemical variables measured were: water temperature (°C), pH, dissolved oxygen (mg/L), chemical oxygen demand (mg/L), nitrate (mg/L), turbidity (NTU), and phosphate (mg/L). The hydroclimatic variable included was rainfall (mm). Contextual variables were sampling date, river or stream identity, sampling location, latitude, longitude and season. Water-quality predictions were conducted using the Water Quality Index (WQI), while recorded pollution label was used as the categorical outcome for pollution-risk classification. The outputs of WQI and pollution status were both classified as "separate outcomes" and WQI was not directly used as an input predictor in the pollution-classification models to reduce target leakage.

2.2 Data Preprocessing and Descriptive Analysis

There were no missing values coded in the original data frame and 1,495 data rows. Prior to performing statistical analysis, data-quality screening was performed. Two pairs of labels identified the same season, so the seasonal labels were standardized. The phrase 'Pre monsoon' was merged with 'Pre-monsoon', and 'Post monsoon' with 'Post-monsoon'. The seasonal variability was thus obtained as four categories namely Winter, Pre-monsoon, Monsoon and Post-monsoon.

Six of the COD measurements were negative, with the lowest of these at −59 mg/L, which was physically implausible as the concentration of COD. The six measurements were not interpreted as valid observations, but were recorded as missing for analyses involving COD. Thus, descriptive statistics calculated for COD were 1,489 valid observations and the denominator for the overall study was 1,495 observations. For predictive modelling, the records with invalid COD measurements were retained as they were preprocessed with data from the training folds to avoid transferring information from the validation folds into training the model.

The mean, standard deviation, minimum, median, and maximum were used to summarize continuous variables. Percentages for pollution categories were determined by dividing the

number of observations in each category by the total sample ($N = 1,495$) and multiplying by 100. WQI, DO, COD, nitrate, turbidity, phosphate and rain were calculated and presented in descriptive summaries for each season. The means (standard deviation) were used to summarize the data for WQI, and the means were used for the remaining seasonal variables. The comparisons were descriptive and no inferential seasonal hypothesis test was used because the Results section provided information about the distribution of the data by season rather than the results of tests for significance.

2.3 Water Quality Index Prediction

The continuous dependent variable was WQI and supervised regression analysis was conducted. The predictor variables included all the available physicochemical, hydroclimatic, spatial, temporal and contextual variables, apart from those that explicitly included the target variable. The following regression algorithms were tested: Linear Regression, Random Forest Regression, Extra Trees Regression, Gradient Boosting Regression and Support Vector Regression with a radial basis function kernel.

Five-fold grouped cross validation with sampling location as grouping variable was used for evaluating the predictive ability. Each fold had only one of the two subsets (training or validation) that contained records of one sampling location. This method was used to achieve this by adjusting performance estimates to be less optimistic due to spatial dependence in repeated observations from the same location.

Regression performance was assessed using the coefficient of determination (R^2), mean absolute error (MAE), and root mean squared error (RMSE):

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

y_i Mean $\backslash(R^2\backslash)$, MAE, and RMSE values across the five validation folds were used for model comparison, while the standard deviation of fold-specific $\backslash(R^2\backslash)$ values was used to evaluate

performance stability. Predictions from all held-out folds were subsequently combined to generate out-of-fold predictions for all 1,495 observations, which were used for the observed-versus-predicted WQI assessment.

2.4 Pollution-Risk Classification

Pollution risk prediction has been designed as a three-class supervised classification problem where the dependent variable is the recorded categories Good, Poor, and Very Poor. There were 52 Good observations, 34 Poor observations and 1,409 Very Poor observations, which represent 3.48% and 2.27% and 94.25% of the total observations, respectively.

Due to the significant class imbalance, accuracy alone was not used to assess model performance. The four classification algorithms evaluated were: Logistic Regression, Random Forest, Extra Trees, and Support Vector Machine (SVM) – radial basis function (RBF) kernel. Where supported, imbalance-aware class weighting or other algorithm-specific weighting applied.

The physicochemical, hydroclimatic, seasonal, river and geographic variables were included in the predictor set. To avoid leakage of a variable that was highly correlated with the pollution-category outcome, the pollution-risk predictor was not included in WQI.

Pollution classification was performed using the same five-fold sampling-location-grouped cross validation as performed for the WQI regression. As a result, there were no records available for training and validation at the same sample site within a fold.

Classification performance was assessed using accuracy, balanced accuracy, macroaveraged F1-score, and Matthews correlation coefficient (MCC). Accuracy was calculated as:

$$\text{Accuracy} = \frac{\text{Correctly classified observations}}{\text{Total evaluated observations}}$$

Balanced accuracy was calculated as:

$$\text{Balanced Accuracy} = \frac{1}{K} \sum_{k=1}^K \text{Recall}_k$$

where $K = 3$ represents the number of pollution classes.

For each class, the F1-score was calculated as:

$$\text{F1}_k = 2 \times \frac{\text{Precision}_k \times \text{Recall}_k}{\text{Precision}_k + \text{Recall}_k}$$

Macro-F1 was obtained as the unweighted mean of the three class-specific F1-scores. MCC was additionally reported because it incorporates information from the complete classification confusion structure and provides a more informative evaluation under severe class imbalance.

Predictions from all held-out folds were combined to generate out-of-fold classifications for all 1,495 observations. For the best-performing classifier, the pooled confusion matrix and class-specific precision, recall, and F1-score were examined to evaluate predictive performance separately for the minority Good and Poor classes and the dominant Very Poor class.

2.5 Model Explainability and Statistical Reporting

Permutation feature importance was used to explore model explainability for the top performing WQI regression model. The validation data were then repeatedly permuted for each predictor, while the model fitted was not changed. For this, the decrease in the predictive R^2 compared to the unpermuted model was used as the importance measure. Larger decrease in R^2 were deemed to be more influential in the determination of model performance. The errors bars show the estimated variability of the importance estimates, using repeated permutations. The results of the importance test were used as a sign of the lack of independence of the model instead of as evidence of a causal relationship. All comparisons of machine-learning models were computed using out-of-sample predictions obtained using location-grouped cross-validation.

3. Results

3.1 Overall Water-Quality Characteristics and Pollution Status

Results from an analysis of 1495 observations were obtained. As summarized in Table 1, mean water temperature was 27.76 ± 3.42 °C and mean pH was 7.38 ± 0.36 . The average value of the dissolved oxygen was 2.27 ± 2.45 mg/L (excluding the six negative COD measurements), while the average value of COD was 71.98 ± 27.85 mg/L. Nitrogen (nitrate), turbidity and phosphate were highly variable. The overall WQI was high, with a mean of 91.40 ± 13.63 and a median of 94.80.

Table 1. Descriptive characteristics of water-quality and hydroclimatic parameters

Parameter	Mean	SD	Minimum	Median	Maximum
Temperature (°C)	27.76	3.42	20.20	28.90	39.60
pH	7.38	0.36	4.47	7.38	9.12
DO (mg/L)	2.27	2.45	0.00	1.07	12.28
COD (mg/L)	71.98	27.85	0.00	70.00	380.00
Nitrate (mg/L)	9.31	17.13	0.10	5.10	367.80
Turbidity (NTU)	45.47	49.82	0.00	30.10	501.00
Phosphate (mg/L)	3.98	6.06	0.00	2.00	84.20
Rainfall (mm)	8.20	16.94	0.00	1.14	118.92
WQI	91.40	13.63	7.26	94.80	99.65

The distribution of pollution category is very skewed as depicted in Figure 1. Of the 1,495 observations, 1,409 (94.25%) were classified as Very Poor, 52 (3.48%) as Good, and 34 (2.27%) as Poor. The WQI values also exhibited good separation between the pollution classes, with the mean of WQI values in the Good observations ranged from 7.26 to 49.20; Poor observations 52.75 to 68.16, and Very Poor observations 76.02 to 99.65.

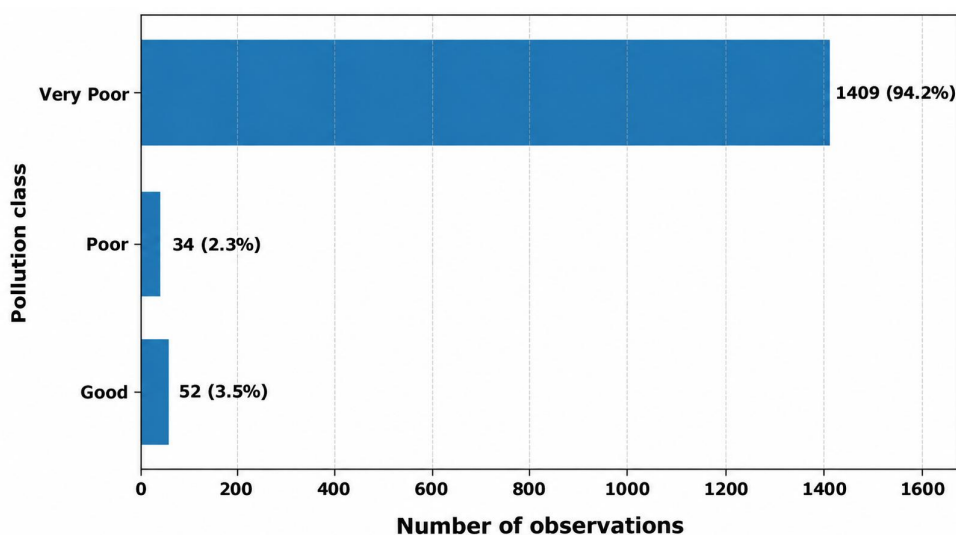


Figure 1. Pollution-class distribution across all water samples

3.2 Seasonal Variation in Water Quality

Table 2 shows seasonal water-quality patterns. Mean WQI was highest during the pre-monsoon season (93.10 ± 10.98), followed by winter (92.92 ± 13.39), post-monsoon (91.20 ± 11.19), and monsoon (89.26 ± 16.20). The highest mean dissolved oxygen concentration was observed during monsoon (3.78 mg/L), mean phosphate concentration was the highest during winter (6.77 mg/L), and mean turbidity was the highest during monsoon with a value of 72.33 NTU. The

dissolved oxygen concentration was highest in the monsoon (3.78 mg/L), the phosphate concentration was highest in winter (6.77 mg/L), and the turbidity concentration showed the highest monsoonal increase (72.33 NTU). Monsoon rainfall was also significantly higher with mean of 19.10 mm.

Table 2. Seasonal profile of WQI and major water-quality parameters

Season	n	WQI Mean ± SD	DO	COD	Nitrate	Turbidity	Phosphate	Rainfall
Winter	426	92.92 ± 13.39	1.18	69.68	11.19	31.28	6.77	0.03
Pre- monsoon	258	93.10 ± 10.98	0.74	66.26	7.63	39.92	5.78	7.96
Monsoon	476	89.26 ± 16.20	3.78	77.30	8.76	72.33	1.66	19.10
Post- monsoon	335	91.20 ± 11.19	2.71	71.84	8.98	29.64	2.32	3.27

The seasonal WQI distributions are visualized in Figure 2. Although all four seasons were dominated by relatively high WQI values, monsoon observations exhibited greater dispersion and a somewhat lower central WQI profile than the other seasonal groups.

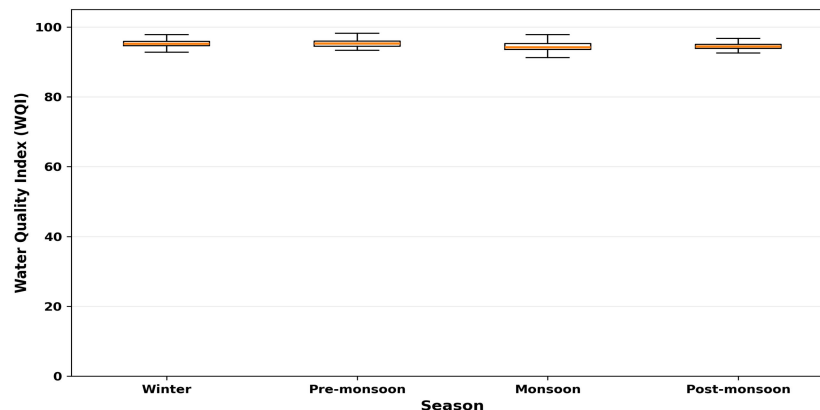


Figure 2. Seasonal distribution of the Water Quality Index

3.3 Machine-Learning Prediction of Water Quality Index

The five regression algorithms were tested for prediction of WQI via location-grouped five-fold cross-validation. As evident in Table 3, Gradient Boosting regressor was the best performing model with 0.9993 as Mean R^2 , 0.1215 as MAE, and 0.3485 as RMSE. Random Forest also gave

excellent R² value of 0.9989 while Extra Trees achieved an R² value of 0.8331, which was still significant. The performance of linear Regression and RBF-SVR was significantly poorer.

Table 3. Location-grouped five-fold cross-validation performance for WQI prediction

Model	Mean R ²	SD of R ²	MAE	RMSE
Linear Regression	0.1361	0.0864	6.6398	12.2826
Random Forest	0.9989	0.0008	0.1597	0.4441
Extra Trees	0.8331	0.0432	1.9413	5.3875
Gradient Boosting	0.9993	0.0006	0.1215	0.3485
SVR (RBF)	0.0853	0.0503	3.6324	12.7238

The close agreement between observed and cross-validated predicted WQI values is illustrated in Figure 3. Based on pooled out-of-fold predictions, Gradient Boosting produced an overall R² of 0.9991, MAE of 0.1214, and RMSE of 0.4067.

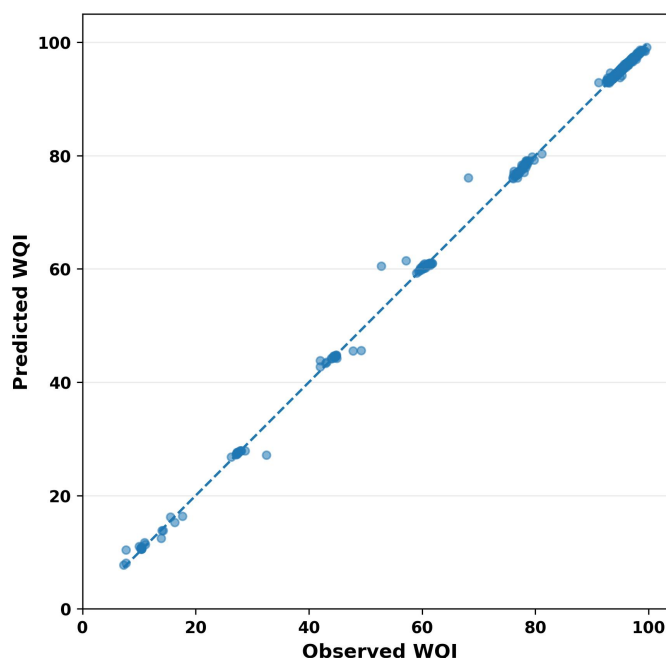


Figure 3. Observed versus cross-validated WQI predicted by the Gradient Boosting model

The results of the model interpretability analysis showed that the models for predicting WQI were dominated by certain physicochemical variables. When looking at the changes in predictive R² after the permutation of each individual variable (Figure 4), phosphate was the variable that caused the greatest decrease, indicating that it is the strongest controller of the fitted WQI model. The other variables that gave some explanatory information were the pH, the dissolved oxygen, and the turbidity, followed by nitrate. The other physicochemical predictors accounted for

relatively little extra predictive information when added to the major physicochemical predictors, which were season, COD, river identity, and spatial coordinates.

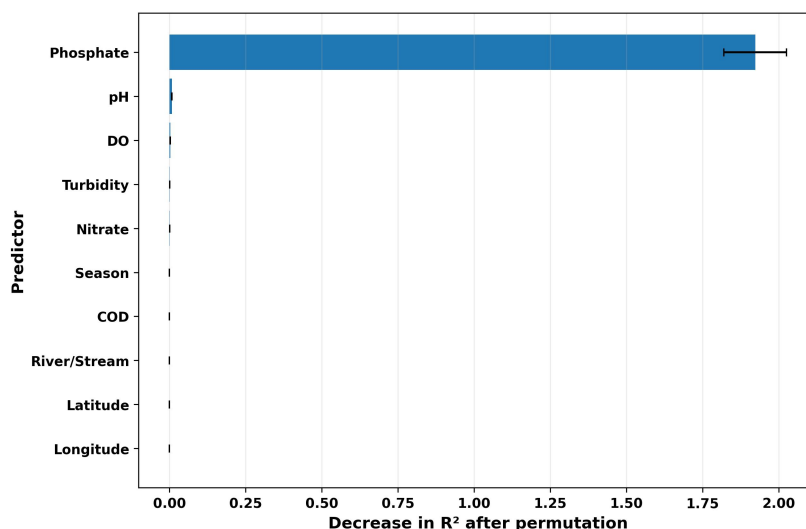


Figure 4. Permutation-based predictor importance for Gradient Boosting WQI prediction

3.4 Explainable Pollution-Risk Classification

Pollution-risk classification was done independently from WQI with physicochemical, hydroclimatic, seasonal, river and geographic predictors. Class-weighted models and imbalance-sensitive performance measures were highlighted, as 94.25% of observations fell in the Very Poor category. Table 4 shows that the Random Forest classifier performed best overall, with a mean accuracy of 0.9940, balanced accuracy of 0.9061, macro-F1 of 0.9394 and MCC of 0.9414 over the location-grouped folds.

Table 4. Location-grouped five-fold cross-validation performance for pollution-risk classification

Model	Accuracy	Balanced Accuracy	Macro-F1	MCC
Logistic Regression	0.8120	0.7685	0.5049	0.3779
Random Forest	0.9940	0.9061	0.9394	0.9414
Extra Trees	0.9391	0.6595	0.6189	0.5018
SVM (RBF)	0.8735	0.5901	0.4886	0.3288

The pooled out-of-fold predictions for the Random Forest model accurately identified 51 of 52 Good samples, 26 of 34 Poor samples and all 1,409 Very Poor samples. Class-specific F1-scores were 0.9808 for Good, 0.8525 for Poor, and 0.9975 for Very Poor. The comparatively lower

recall value (0.7647) suggests that this is still an intermediate pollution category, and that this is still the most difficult to separate, while the most predominant pollution category (Very Poor) recovered with complete sensitivity.

4. DISCUSSION

The present study showed significant degradation of the water systems observed with a mean WQI of 91.40 and 94.25% of the observations were classified as Very Poor. The dissolved oxygen was relatively low, while large variations were found for COD, nitrate, phosphate and turbidity. The results show that water-quality condition was related to interactions between several physicochemical components, which is not related to one pollution characteristic. Interpretable ML also has shown that quantifying the effect of single water qualities is also significant under multiple pollution sources (Zhang et al., 2025).

Further, the seasonal trends were observed, with minimum WQI during the monsoon season and maximum turbidity and dissolved oxygen during this period. Phosphate levels were higher, in contrast, in winter and pre-monsoon periods. These differences emphasize the need to include temporal and hydroclimatic data in monitoring water quality rather than assuming that all aspects of the environment are at the same state throughout the year. Water Quality Index (WQI) based ML evaluation has also been suggested as a tool to capture the complexity of water quality variations using integrated predictive modelling (Islam, Salekin, Abdullah, et al., 2024).

The Gradient Boosting regression algorithm produced the highest mean WQI prediction results (0.9993), the lowest MAE (0.1215), and best RMSE (0.3485) of the regression algorithms evaluated. Similarly, there was a strong performance from Random Forest while Linear Regression and RBF-SVR gave poor performance. This difference indicates that the relationship between the environmental variables measured and the underlying relationships with WQI was more well-represented by nonlinear tree-based ensembles. Similarly, robust ML techniques have been used to make good performance in the prediction of WQI under complex environmental conditions (Han et al., 2025). Explainable ensemble models also showed consistent performance in predicting water-quality when using various forms of water-quality indicators (Doan & Nguyen, 2025). Similar to the interpretable models described, models with high predictive accuracy have been coupled with the identification of contamination-related variables (Aldrees et al., 2025).

Phosphate was found to provide the most important information for WQI, followed by pH and

dissolved oxygen and then turbidity and nitrate. Permutation importance indicated that these variables were dominant and provided the most important information regarding the WQI. The pattern indicates that nutrient enrichment, oxygen availability, acidity/alkalinity balance and suspended material all played a role in determining the modelled water-quality condition. Interpretable ML has also been employed to reduce monitoring protocols while retaining highly informative water-quality parameters, without discarding them (Xue et al., 2025). The XAI-based assessment has also shown the benefit of revealing influential predictors to water-quality predictions, which are otherwise model outputs without explanatory power (Maurya et al., 2025). However, permutation importance provides a measure of model dependence, but should not be used to interpret causal environmental relationships.

The classification for pollution risk was especially difficult as only 52 observations were classified as Good and 34 as Poor, while 1,409 were classified as Very Poor. The accuracy of Random Forest was 0.9940, the balanced accuracy was 0.9061, the macro-F1 was 0.9394 and the MCC was 0.9414, despite this, the imbalance. Modelling of riverine pollution indices using data has also been done to support the use of multiple ML algorithms for discriminating pollution types (Chowdhury & Debnath, 2025). Explainable pollution-index prediction has also been shown to be beneficial for bridging between environmental measurements and resource-suitability decisions (Abba et al., 2023).

The RF model correctly classified all of the Very Poor observations and 51 of 52 Good observations, yet only 26 of 34 Poor observations were correctly classified. The middle category was therefore the most challenging to differentiate, and highlighted the need for accuracy to be misleading under the class imbalance observed. With the assessment of water-quality prediction systems beyond headline parameters, robustness evaluation is becoming an increasingly important aspect (Zalte et al., 2025). The automation of interpretation of predictive water-quality models can further improve the confidence in environmental-management applications (Mallick et al., 2025). When predictive evidence is used for management planning, the interpretation of black-box pollution models has been encouraged, too (Talukdar et al., 2025). At a larger scale, recent XAI literature reviews highlight the importance of transparent interpretation of the models for reliable water quality monitoring systems (Aderemi et al., 2025a). Overall, the results go to validate the use of nonlinear prediction, imbalance-sensitive validation, and interpretable feature analysis in water-resource management frameworks that are sustainable in the long run.

5. CONCLUSION

This study showed the feasibility of an explainable machine-learning system for water-quality analysis and forecasting of water pollutants based on physicochemical, hydroclimatic, spatial and seasonal data. Generally poor water quality with an average WQI of 91.40 and 94.25% of observations rated as Very Poor was found in the monitored waters. Seasonal differences also suggested that monsoon conditions were related to low mean WQI and high turbidity, and nutrients displayed significant seasonal trends. The Gradient Boosting regression model was the most reliable model for predicting the WQI with a mean cross-validated R^2 of 0.9993 and minor MAE and RMSE values. Random Forest had the highest BAC, macro-F1, and MCC scores and high balanced accuracy despite significant class imbalance in the data, making it the best for classification of pollution-risk among all the algorithms. The explainability analysis showed that phosphate was the most important predictor of the WQI, followed by pH, dissolved oxygen, turbidity and nitrate. Combining the two approaches, predictive modelling and model interpretability, can give an accurate and interpretable environmental assessment, the findings reveal. This approach can assist in early warning of poor water quality, focus monitoring to important pollution factors, and more specific monitoring approaches. Overall the framework has practical applicability to sustainable water-resource management, but further studies would be needed to establish its generalizability to other regions, monitoring durations, and contaminant inventories and suites.

REFERENCES

1. Abba, S. I., Yassin, M. A., Mubarak, A. S., Shah, S. M. H., Usman, J., Oudah, A. Y., Naganna, S. R., & Aljundi, I. H. (2023). Drinking water resources suitability assessment based on pollution index of groundwater using improved explainable artificial intelligence. *Sustainability*, 15(21), 15655. <https://www.mdpi.com/2071-1050/15/21/15655>
2. Aderemi, I. A., Kehinde, T. O., Okwor, U. D., Ahmad, K. H., Adjei, K. Y., & Ekechi, C. C. (2025a). Explainable AI for Water Quality Monitoring: A Systematic Review of Transparency Interpretability and Trust. *IEEE Sensors Reviews*, 2, 419–443. https://www.researchgate.net/profile/Khalid-Ahmad-15/publication/397269056_Explainable_AI_for_Water_Quality_Monitoring_A_Systematic_Review_of_Transparency_Interpretability_and_Trust/links/690a5c4d9708d52f2da4e0cc/Explainable-AI-for-Water-Quality-Monitoring-A-Systematic-Review-of-Transparency-Interpretability-

and-Trust.pdf

3. Aderemi, I. A., Kehinde, T. O., Ugochukwu, D. O., Ahmad, K. H., Adjei, K. Y., & Chijioke, C. E. (2025b). Beyond the black box: a systematic review of explainable AI for transparent and trustworthy water quality monitoring. *IEEE Sensors Reviews*. <https://ieeexplore.ieee.org/abstract/document/11112533/>
4. Aldrees, A., Jibrin, A. M., Dan'azumi, S., Mahmoud, I. A., Aliyu, U. U., & Abba, S. I. (2025). Explainable machine learning framework for assessing groundwater quality and trace element contamination in Eastern Saudi Arabia. *Scientific Reports*, 15(1), 45333. <https://www.nature.com/articles/s41598-025-29598-8>
5. Alshehri, F., & Rahman, A. (2023). Coupling machine and deep learning with explainable artificial intelligence for improving prediction of groundwater quality and decision-making in arid region, Saudi Arabia. *Water*, 15(12), 2298. <https://www.mdpi.com/2073-4441/15/12/2298>
6. Choudhary, R., Kumar, A., C, P., Naik, M. M., Choudhury, M., & Khan, N. A. (2025). Predicting water quality index using stacked ensemble regression and SHAP based explainable artificial intelligence. *Scientific Reports*, 15(1), 31139. <https://www.nature.com/articles/s41598-025-09463-4>
7. Chowdhury, M. A. H., & Debnath, P. (2025). Predictive Multiple Machine Learning Models of Riverine Water Pollution Index: A Comprehensive Data-Driven Study.. https://www.researchgate.net/profile/Md-Anamul-Chowdhury-3/publication/396063024_Predictive_Multiple_Machine_Learning_Models_of_Riverine_Water_Pollution_Index_A_Comprehensive_Data-Driven_Study/links/68dd69a902d6215259b7fe22/Predictive-Multiple-Machine-Learning-Models-of-Riverine-Water-Pollution-Index-A-Comprehensive-Data-Driven-Study.pdf
8. Doan, C. T., & Du Nguyen, H. (2025). Robust water quality prediction across multiple indicator formulations using an explainable ensemble learning model. *Water Resources and Industry*, 100329. <https://www.sciencedirect.com/science/article/pii/S2212371725000538>
9. Han, Z., Zhang, S., & He, L. (2025). Predicting and investigating water quality index by robust machine learning methods. *Journal of Environmental Management*, 381, 125156. <https://www.sciencedirect.com/science/article/pii/S0301479725011326>
10. Hellen, N. (2022). Explainable artificial intelligence and model calibration for water quality

prediction.. <https://dspace.bracu.ac.bd/items/22b3adac-3142-4bc0-94f7-f098df1e60a4>

11. Hellen, N., Sabuj, H. H., & Ashraful Alam, M. (2023). Explainable AI and Ensemble Learning for Water Quality Prediction. *Proceedings of International Conference on Information and Communication Technology for Development*, 235-250. https://doi.org/10.1007/978-981-19-7528-8_19
12. Islam, M. J., Salekin, S. U., Abdullah, M. S., Zaman, N., & Khan, A. A. A. (2024). Evaluation of water quality assessment through machine learning: A water quality index-based approach.. <https://assets-eu.researchsquare.com/files/rs-4370890/v1/6047f324-ea97-4edb-902f-64fb858786f1.pdf>
13. Islam, M. J., Salekin, S. U., Anzum, A., Zaman, N., Khan, A. A. A., Sarkar, D., Rabbani, M. L., & Hossain, M. T. (2024). Machine learning-driven water quality index prediction: Enhancing accuracy with gradient boosting and explainable AI for sustainable water monitoring. *Applied Agriculture Sciences*, 2(1), 1–14. <https://publishing.emanresearch.org/Journal/abstract/agriculture-2110031>
14. Mallick, J., Alqadhi, S., Alsubih, M., Ahmed, M. F. M., & Abdo, H. G. (2025). Automated interpretation of deep learning-based water quality assessment system for enhanced environmental management decisions. *Applied Water Science*, 15(5), 110. <https://doi.org/10.1007/s13201-025-02452-y>
15. Mallick, J., Alqadhi, S., Hang, H. T., & Alsubih, M. (2024). Interpreting optimised data-driven solution with explainable artificial intelligence (XAI) for water quality assessment for better decision-making in pollution management. *Environmental Science and Pollution Research*, 31(30), 42948-42969. <https://doi.org/10.1007/s11356-024-33921-7>
16. Maurya, R., Saha, S., Misra, B., Chatterjee, D., Chakraborty, S., & Sarkar, P. (2025). Employing Explainable Artificial Intelligence (XAI) for Enhanced Water Quality Assessment. *2025 International Conference on Computing, Intelligence, and Application (CIACON)*, 1–5. <https://ieeexplore.ieee.org/abstract/document/11189765/>
17. Nallakaruppan, M. K., Gangadevi, E., Shri, M. L., Balusamy, B., Bhattacharya, S., & Selvarajan, S. (2024). Reliable water quality prediction and parametric analysis using explainable AI models. *Scientific Reports*, 14(1), 7520. <https://www.nature.com/articles/s41598-024-56775-y>

18. Politikos, D., Stefanidis, K., Varlas, G., Papadopoulos, A., & Dimitriou, E. (2024). Predicting the Ecological Quality of Rivers: A Machine Learning Approach and a What-if Scenarios Tool. *Environmental Modeling & Assessment*, 29(6), 1059-1077. <https://doi.org/10.1007/s10666-024-09980-y>
19. Shaheed, H., Zawawi, M. H., & Hayder, G. (2025). The development of a river quality prediction model that is based on the water quality index via machine learning: A review. *Processes*, 13(3), 810. <https://www.mdpi.com/2227-9717/13/3/810>
20. Sinha, P., Debnath, N., & Bhattacharjee, P. (2026). Water quality and hydroclimatic dataset for rivers and streams of Dhaka, Bangladesh (Version 1) [Data set]. Mendeley Data. <https://doi.org/10.17632/xftpv7jvx3.1>
21. Talukdar, S., Ahmed, I. A., Shahfahad, Baig, M. R. I., Rihan, M., & Rahman, A. (2025). Water Pollution Assessment and Management Through Interpreting Black Box Deep Learning Algorithms in the Loktak Lake. *Advanced GIScience in Hydro-Geological Hazards*, 359-375. https://doi.org/10.1007/978-3-031-76189-8_17
22. Talukdar, S., Shahfahad, Ahmed, I. A., Rihan, M., & Rahman, A. (2024). Integrating Explainable Artificial Intelligence and Machine Learning for Understanding Water Pollution and Its Management. *Environmental Risk and Resilience in the Changing World*, 115-128. https://doi.org/10.1007/978-3-031-62442-1_7
23. Tian, P., Xu, Z., Fan, W., Lai, H., Liu, Y., Yang, P., & Yang, Z. (2024). Exploring the effects of climate change and urban policies on lake water quality using remote sensing and explainable artificial intelligence. *Journal of Cleaner Production*, 475, 143649. <https://www.sciencedirect.com/science/article/pii/S0959652624030981>
24. Xue, Y., Niu, Z., Zhang, P., Zhang, R., Jia, L., & Duan, K. (2025). Optimizing water quality monitoring in an arid wetland using interpretable machine learning: From identifying dominant parameters to simplifying the water quality index. *Ecological Indicators*, 181, 114397. <https://www.sciencedirect.com/science/article/pii/S1470160X25013299>
25. Zalte, J., Rai, A., Shah, H., & Fulekar, M. H. (2025). Adversarial susceptibility analysis for water quality prediction models. *Scientific Reports*, 15(1), 37828. <https://www.nature.com/articles/s41598-025-21797-7>
- Zhang, T., Wu, J., Chu, H., Liu, J., & Wang, G. (2025). Interpretable machine learning based

quantification of the impact of water quality indicators on groundwater under multiple pollution sources.