Machine Learning and Deep Learning Models for Water Quality Prediction with SHAP-Based Explainable AI

Dr. M. Geetha Jenifel[1], Dr. Mashael Mahmoud Khayyat[2]

[1]Postdoctoral Researcher, Centre of Postgraduate Studies, Lincoln University College, Malaysia; [2]Information Systems and Technology, College of Computer Science and Engineering, University of Jeddah;

pdf.jenifel@lincoln.edu.my, mkhayyat@uj.edu.sa

Abstract: The prediction of water quality has an important role to play in ensuring public health safety and management of the environment. In India, rising pressure on river systems makes water quality prediction necessary. Predictive models built on machine learning algorithms have high efficiency levels, but because of being black boxes, lack of confidence in adopting such models prevails. In this context, three predictive models—SVM, LR, and LSTM networks—have been compared based on the Indian River system dataset to predict water quality. SHapley Additive exPlanations (SHAP) is utilized to interpret predictions made by models and determine those physicochemical parameters that affect the outcome of each prediction. It has been found that LSTM is the most effective of the three models considered in predicting water quality, and it does so due to its ability to capture temporal dependencies in water quality parameters, while SHAP determines the reasons behind each prediction.

Keywords: Decision making; Environment; Pollution; Temporal dependencies; Statistical models

---

Introduction

Clean and safe water is essential for ensuring public health and ecological equilibrium as well as sustainable development. The Indian rivers like the Ganga, Yamuna, and others are under severe pressure because of urbanization, industrial pollution, agricultural waste, and ineffective disposal of waste water. This has resulted in deteriorating water quality conditions and serious threats to human health and the environment. Evaluation of water quality at the appropriate time and with accuracy is very important for environmental monitoring and policy-making. Water quality evaluation through conventional physicochemical analysis, carried out in laboratories, is known for its high level of precision but suffers from time and money expenditures and impracticability for continuous monitoring on a large scale. To overcome this problem, ML techniques like SVM and LR were extensively used due to their easy implementation and explainability, whereas more modern and efficient deep learning models like LSTM have shown better results in understanding the temporal and sequential dependencies characteristic of time series. The main drawback of the majority of ML and deep learning models lies in their black box character, which led to an increasing interest in Explainable Artificial Intelligence (XAI).

In this regard, the current research develops a novel explainable water quality prediction framework that combines three different prediction models, namely SVM, LR, and LSTM, trained on actual datasets collected from Indian River systems. The use of the LSTM model is particularly emphasized to analyze the sequence and time dependency of water quality parameters, which are not inherent features of conventional models such as SVM and LR. Finally, SHAP is used in all three models to make sense out of the predictions made by each model and determine the factors that impact the variation in water quality.

Related work

There is much literature on water quality prediction employing ML; however, more recent studies tend to consider explainability as well to mitigate the issue of black boxes. A number of works use an ensemble tree-based approach and SHAP to predict the water quality index (WQI). Although these works prove that the combination of predictive modelling and XAI is successful, they mainly explore the use of ensemble tree-based models with SHAP or deep sequential models but without any explanation component. Few works provide a direct comparison of classic approaches (SVM, LR) with sequential deep learning (LSTM) by employing SHAP for explainability specifically in the India Rivers.

Table 1. Water quality techniques comparative review

Ref	Techniques used	Dataset	Key findings	Limitations
[1]	Stacked ensemble +SHAP	1987 samples, Indian rivers	Improved WQI forecasting	Computational overhead
[2]	RF, LightGBM, XGBoost+SHAP	River water samples	XGBoost outperformed others	No temporal models considered
[3]	Adaboost, CatBoost, HistGBRT	Illinois	SHAP enabled	Region Specific
[4]	Multiple ML models+SHAP	9 major Indian rivers	Identified the most critical WQ parameters and best-performing models per river	The comparative model set is not clearly benchmarked.
[5]	ANFIS	Santiago River, Mexico	The Best training subset improved prediction	No XAI/explainability integration
[6]	Decision Tree, RF, SVM+Lime,SHAP	River WQI classification dataset	Gradient boosting achieved the highest accuracy	No LSTM/temporal modeling
[7]	Bi-LSTM	Yamuna River, India	Bi-LSTM effectively forecasted WQ factors	No explainability component

Key Contribution

Recent studies have demonstrated that deep learning architectures like LSTM, GRU, CNN, and Transformer models can effectively learn temporal and spatial relationships in data related to the environment [8][9][10]. The main drawback of the majority of ML and deep learning models lies in their black box character, which led to an increasing interest in Explainable Artificial Intelligence (XAI) [11][12][13][14].

Contributions of the study are as follows:

- An ML and DL-based framework for predicting the water quality was built and tested on various water quality parameters.
- Comparison of different ML and DL algorithms was performed, and it was observed that the LSTM algorithm outperforms all others with respect to its predictive capabilities on the chosen dataset.
- Use of the SHAP explainable AI framework is included in the prediction process to estimate the importance of different water quality parameters in the prediction process.
- The framework developed not only gives global importance to features but also gives the importance of features on a per-sample basis.

Thus, this framework offers an interpretable decision support system for water quality prediction and management.

Method, Experiments and Results

The proposed framework follows a five-stage pipeline: data collection, preprocessing, parallel model training (SVM, LR, LSTM), SHAP-based explainability, and performance evaluation, as illustrated in figure 1.

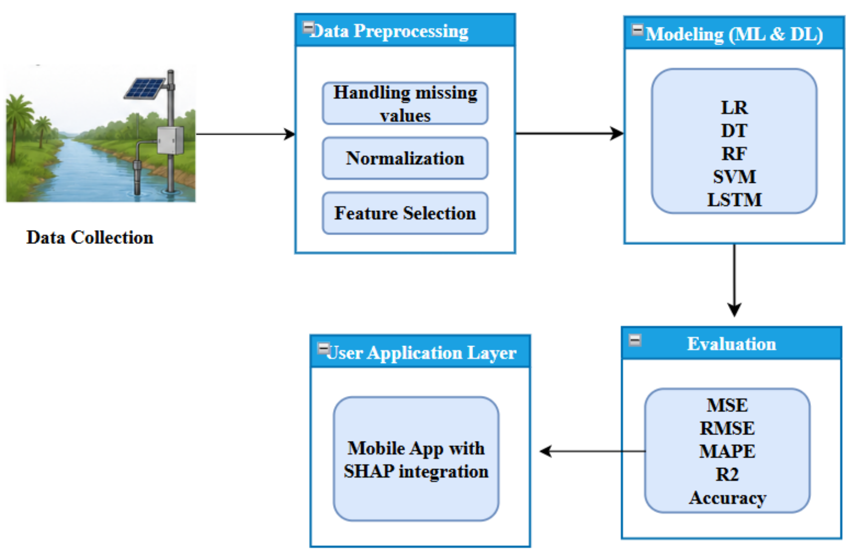

Figure 1. SHAP-based real-time water quality monitoring and forecasting

Data collection and preprocessing

The water quality data is collected using monitoring stations of rivers in India, where various physicochemical factors such as pH, Dissolved Oxygen (DO), Turbidity, BOD, COD, and Temperature are measured. The data is preprocessed using missing value handling and min-max normalization of the features to:

$$x^{\sim} = \frac{(x - x_{\min})}{x_{max} - x_{min}} \quad \text{---- (1)}$$

Data splitting is performed into train and test subsets (80:20) with the LSTM branch transformed into sequences using the sliding window approach as well.

Support Vector Machine (SVM)

SVR fits a function that deviates from the true target by no more than ε for all training data, while remaining as flat as possible:

Minimize:

$$½\|w\|^2 + C \cdot \Sigma_i(\xi_i + \xi_i *) \quad \text{---- (2)}$$

Subject to:

$$y_i - (w \cdot x_i + b) \leq \varepsilon + \xi_i, (w \cdot x_i + b) - y_i \leq \varepsilon + \xi_i *, \xi_i, \xi_i * \geq 0 \quad \text{---- (3)}$$

Where w is the weight vector, b is the bias, C is the regularization parameter, and ξ, ξ* are slack variables permitting deviations beyond ε.

Linear Regression

LR models the target as a linear combination of input features:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \varepsilon \quad \text{---- (4)}$$

Coefficients β are estimated by minimizing the residual sum of squares:

$$\hat{\beta} = \text{argmin\_}\beta\, \Sigma_i(y_i - X_i\beta)^2 \quad \text{---- (5)}$$

Long Short-Term Memory (LSTM)

LSTM captures temporal dependencies in sequential water quality data through gated memory cells:

Forget gate: $f_t = \sigma(Wf \cdot [h_{t-1}, x_t] + bf)$ ---- (6)

Input gate: $i_t = \sigma(Wi \cdot [h_{t-1}, x_t] + bi)$ ---- (7)

Candidate cell state: $\tilde{C}_t = \tanh(Wc \cdot [h_{t-1}, x_t] + bc)$ ---- (8)

Cell state update: $C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$ ---- (9)

Output gate: $o_t = \sigma(Wo \cdot [h_{t-1}, x_t] + bo)$ ---- (10)

Hidden state: $h_t = o_t * \tanh(C_t)$ ---- (11)

Where σ is the sigmoid activation, Wf, Wi, Wc, and Wo are weight matrices, and h_t is the hidden state used for prediction at time step t.

SHAP-based explainability

SHAP assigns each feature an importance value (Shapley value) based on cooperative game theory, quantifying its marginal contribution to a prediction*:*

$$\varphi_i = \Sigma\_\{S \subseteq F\{i\}\} [|S|!\,(|F| - |S| - 1)!]/|F|! \cdot [f(S \cup \{i\}) - f(S)] \quad \text{---- (12)}$$

Where F is the full feature set, S is a subset excluding feature i, and f(S) is the model's prediction using only features in S. The sum of all Shapley values plus the base value equals the model's output:

$$f(x) = \varphi_0 + \Sigma_i\, \varphi_i$$

This is applied independently to the SVM, LR, and LSTM outputs to identify which physicochemical parameters most influence each model's predictions.

Evaluation metrics

Model performance is assessed using:

$$RMSE = \sqrt{[(1/n)\, \Sigma_i(y_i - \hat{y}_i)^2]} \quad \text{---- (13)}$$

$$MAE = (1/n)\, \Sigma_i|y_i - \hat{y}_i| \quad \text{---- (14)}$$

$$R^2 = 1 - [\Sigma_i(y_i - \hat{y}_i)^2 / \Sigma_i(y_i - \bar{y})^2] \quad \text{---- (15)}$$

Where y_i are actual values, $\hat{y}_i$ are predicted values, and $\bar{y}$ is the mean of actual values.

Discussions

Several forecasting models, including linear regression, support vector machine (SVM), and long short-term memory (LSTM), were implemented and compared. Different XAI algorithms were executed: TreeSHAP, LIME, and permutation importance. The experimental results are shown in Figure 2.

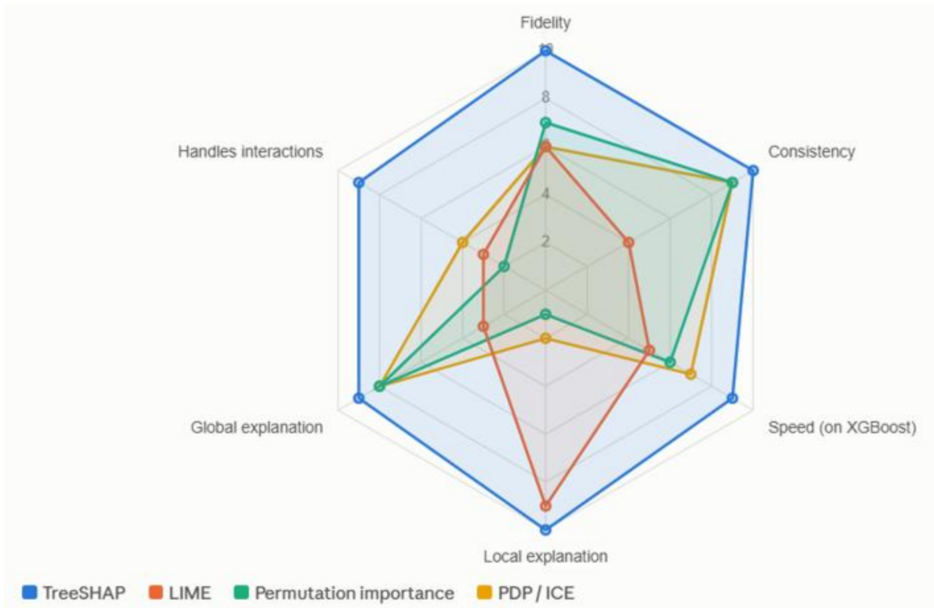

Figure 2. Comparison results of different XAI methods

Comparison results are shown in Figure 2 for TreeSHAP (blue), LIME (orange), and permutation importance (green). TreeSHAP has the largest shape, which is almost equally high on all axes. LIME (orange) has a sharp rise on "local explanation" (9), as it is LIME's main function, but it gets a low score on consistency (4) and interactions (3). Since LIME alters the input randomly and builds a local linear model, running LIME on the same sample may lead to different results, and LIME cannot deal with non-linear interactions the way SHAP can. LIME is not meant to provide global explanations anyway (score 3).

Permutation importance (green) has a high consistency (9) and global explanation (8) because it is a simple global ranking procedure but does a poor job on local explanation (1). Interaction handling by permutation importance (2) is not great either, since it permutes only one feature at a time.

Conclusions

The results confirm that machine learning and deep learning techniques with XAI algorithms,

- Addressed the need for accurate and explainable water quality prediction using Indian River data.
- Compared SVM, linear regression, and LSTM, with SHAP used for model interpretability.
- LSTM effectively captured temporal patterns, while SHAP identified the most influential water quality parameters and improved model transparency.
- Limited by dataset size and coverage. Future work will focus on larger datasets, advanced DL models, and federated learning.

References

1. Merabet, K, Di Nunno, F, Granata, F, "Predicting water quality variables using gradient boosting machine: Global versus local explainability using SHapley Additive Explanations (SHAP)". Earth Science Informatics, 18, Article 298 ,2025
2. Khullar, S., & Singh, N. "Water quality assessment of a river using deep learning Bi-LSTM methodology: Forecasting and validation". Environmental Science and Pollution Research, 2021
3. Choudhary, R., Kumar, A., C., P, "Predicting water quality index using stacked ensemble regression and SHAP based explainable artificial intelligence. Sci Rep 15, 31139, 2025
4. Randika K. Makumbura, Lakindu Mampitiya, Namal Rathnayake, D.P.P. Meddage, Shagufta Henna, Tuan Linh Dang, Yukinobu Hoshino, Upaka Rathnayake, "Advancing water quality assessment and prediction using machine learning models, coupled with explainable artificial intelligence (XAI) techniques like shapley additive explanations (SHAP) for interpreting the black-box nature",Results in Engineering,Volume 23, 102831,ISSN 2590-1230, 2024.
5. Sourav Kundu, Priyangshu Datta, Puja Pal, Kripabandhu Ghosh, Akankshya Das, Basanta Kumar Das, "Unveiling the hidden connections: Using explainable artificial intelligence to assess water quality criteria in nine giant rivers, Journal of Cleaner Production", Volume 492, 144861, ISSN 0959-6526,2025
6. Paneru, B., & Paneru, B. "AI for water sustainability: Global water quality assessment and prediction with explainable AI with LLM chatbot for insights" arXiv. https://arxiv.org/abs/2409.10898, 2025
7. S. Ramya, S. Srinath, Pushpa Tuppad, "Comprehensive analysis of multiple classifiers for enhanced river water quality monitoring with explainable AI,Case Studies" , Chemical and Environmental Engineering,Volume 10, 100822,ISSN 2666-0164,2024
8. M. Zhu, J. Wang, and Y. Deng, "A review of the application of machine learning in water quality evaluation," Environmental Research Communications, vol. 4, no. 11, pp. 1–24, 2022.
9. 9. A. A. Hassan, M. A. M. Hassanien, and H. A. Hefny, "Water Quality Prediction Using Artificial Intelligence Algorithms," Computational Intelligence and Neuroscience, vol. 2021, Art. no. 6659314, pp. 1–14, 2021.
10. 10. M. Y. Shams, E. H. Houssein, M. A. Mahdy, and M. M. Elnahas, "Water quality prediction using machine learning models based on grid search method," Multimedia Tools and Applications, vol. 83, pp. 1–29, 2024.
11. M. Alwateer, "Explainable Machine Learning Models for Water Quality Prediction and Environmental Management," Science of the Total Environment, 2026.
12. M. Zhu, J. Wang, and Y. Deng, "A review of the application of machine learning in water quality evaluation," Environmental Research Communications, vol. 4, no. 11, pp. 1–24, 2022.
13. M. Y. Shams, E. H. Houssein, M. A. Mahdy, and M. M. Elnahas, "Water quality prediction using machine learning models based on grid search method," Multimedia Tools and Applications, vol. 83, pp. 1–29, 2024.
14. J. Han "Predicting Water Quality Index Using Four Advanced Machine Learning Algorithms," Environmental Monitoring and Assessment, 2025.