Smart Water Quality Prediction System Using Ensemble Machine Learning and Explainable AI

Ruchitha[1], Vaishnavi[2], Sanjana[3], Jameema Anam[4] ,Sangameshwar Kawdimath[5]

[1,2,3,4,5]*Department of Information Science and Engineering, Guru Nanak Dev Engineering College, Bidar, Karnataka, India — 585403*

ruchitakulkarni31@gmail.com

Abstract

Access to safe drinking water is a persistent global challenge, particularly in developing regions where agricultural runoff, industrial discharge, and inadequate sanitation infrastructure continue to degrade river and groundwater quality. Conventional monitoring methods based on laboratory physicochemical analysis are too slow and costly for the real-time, large-scale assessment that effective public health response demands. This paper proposes an Explainable Artificial Intelligence (XAI)-based Water Quality Prediction and Contamination Risk Analysis system using Stacked Ensemble Learning. The framework trains six diverse machine learning regressors — XGBoost, CatBoost, Random Forest, Gradient Boosting, Extra Trees, and AdaBoost — and combines their outputs via a Linear Regression meta-learner to predict the Water Quality Index (WQI) from seven physicochemical parameters. SHAP (SHapley Additive Explanations) is applied to quantify each parameter's contribution to individual predictions, converting black-box model outputs into transparent, parameter-specific contamination diagnoses. Evaluation on 1,987 Indian river water samples from the CPCB national monitoring network demonstrates that the stacked ensemble achieves $R^2 = 0.9952$, MAE = 0.412, and RMSE = 0.589 — outperforming all six individual base models. SHAP analysis identifies BOD and pH as the dominant contamination drivers in the studied systems, with a significant non-linear interaction between BOD and dissolved oxygen carrying direct implications for aeration-based remediation. The system is deployed as a Flask web application providing real-time WQI prediction, five-tier risk categorization, and interactive SHAP visualizations, making advanced water safety intelligence accessible to field-level water authorities without requiring laboratory infrastructure or ML expertise.

Keywords— *Explainable AI; SHAP; Stacked Ensemble Learning; Water Quality Index; XGBoost; Contamination Risk; Environmental Monitoring; Flask*

1. Introduction

Safe drinking water remains one of the most pressing global challenges of the twenty-first century. According to the World Health Organization (WHO), more than two billion individuals worldwide consume water from sources contaminated with bacteria, chemicals, or agricultural pollutants, contributing to nearly half a million deaths annually from waterborne diseases alone [1]. In India, the problem is particularly severe: rapid population growth, unplanned industrialization, and the widespread use of chemical fertilizers have together placed extraordinary pressure on river and groundwater systems, especially in agriculturally intensive states such as Karnataka.

The northern Karnataka region, encompassing districts such as Bidar, Kalaburagi, and Raichur, relies heavily on the Krishna river basin and its tributaries as primary sources of both agricultural and domestic water. Field surveys conducted by the Karnataka State Pollution Control Board have repeatedly documented fluoride concentrations exceeding 3.0 mg/L and nitrate levels surpassing 100 mg/L in shallow groundwater wells across this region — values that significantly exceed the Bureau of Indian Standards (BIS) safety thresholds of 1.0 mg/L and 45 mg/L respectively. Long-term consumption of such water causes physicochemical analysis — a process that typically consumes three to seven working days and costs between Rs. 3,000 and Rs. 8,000 per comprehensive test. This makes continuous large-scale monitoring both economically prohibitive and temporally inadequate for timely public health responses. The Water Quality Index (WQI), a composite metric that aggregates weighted scores from multiple parameter into a single interpretable value, is widely used by environmental agencies to communicate water safety to policymakers. However, manual WQI computation remains slow, error-prone, and inaccessible to field-level water safety officers without specialized laboratory support [3].

Machine learning (ML) techniques have emerged as powerful tools for automating WQI prediction from measured physicochemical inputs, enabling near-instantaneous assessment without laboratory processing. Among available algorithms, ensemble methods — which combine predictions from multiple diverse base models consistently demonstrate superior accuracy on environmental datasets that exhibit non-linear parameter interactions, seasonal variability, and geographic heterogeneity [4]. Gradient boosting approaches such as XGBoost and CatBoost have achieved particular prominence, with studies reporting R^2 values exceeding 0.99 on water quality datasets [5].

skeletal fluorosis, methemoglobinemia in infants and various chronic renal conditions [2].

The conventional approach to water quality assessment involves collecting samples and conducting laboratory

Despite these advances, a fundamental limitation persists across virtually all ML-based water quality systems: their inherent opacity. These models generate accurate WQI predictions while offering no mechanistic insight into which specific parameters are responsible for contamination. For water safety authorities, a prediction alone is insufficient they require clear identification of the primary contaminants to design targeted, cost-effective remediation strategies. Prescribing aeration treatment to address organic waste differs entirely from implementing denitrification processes to reduce agricultural nitrate runoff, yet a black-box model cannot distinguish between these scenarios [6].

Explainable Artificial Intelligence (XAI) addresses this critical transparency gap. The SHAP (SHapley Additive Explanations) framework, derived from cooperative game theory, assigns each input feature a quantified contribution value for every individual model prediction, providing both global feature importance rankings and local prediction-specific explanations [7]. Applied to water quality monitoring, SHAP enables field officers to understand not only the predicted WQI score but also which parameters are driving contamination and to what degree — information directly actionable for targeted intervention.

This paper introduces an integrated XAI-based Water Quality Prediction and Contamination Risk Analysis system that addresses these gaps through three coordinated innovations: (1) a stacked ensemble learning architecture combining six diverse ML regressors with a meta-learner to achieve state-of-the-art WQI prediction accuracy; (2) comprehensive SHAP-based explainability generating four distinct visualization types for transparent parameter-level contamination diagnosis; and (3) a Flask-based web application making real-time, explainable water quality intelligence operationally accessible without requiring specialized ML expertise. The complete system is validated on 1,987 real-world Indian river water samples and demonstrates measurable advantages over conventional monitoring approaches.

2. Related Work

Research on ML-based water quality prediction has grown substantially over the past decade. Early studies applied artificial neural networks (ANN) to predict individual parameters such as dissolved oxygen and biochemical oxygen demand with reasonable accuracy, but these approaches suffered from limited generalizability and poor performance on small regional datasets [8]. Support vector machines (SVM) subsequently gained adoption for WQI classification tasks, demonstrating resilience to high-dimensional parameter spaces, though their computational demands limited scalability for large monitoring networks.

The introduction of ensemble tree methods transformed the field. Random Forest regressors were shown to outperform ANN and SVM on water quality prediction benchmarks across multiple geographic regions, with studies in Indian river systems reporting classification accuracies between 88% and 94% [9]. The development of gradient boosting algorithms — particularly XGBoost — established a new performance ceiling: Chen and Guestrin's seminal work demonstrated that regularized gradient boosting with parallel tree construction Subsequent water quality studies applying XGBoost consistently reported R^2 values above 0.98, making it the preferred base algorithm for ensemble water quality systems.

Stacked ensemble architectures, where predictions from multiple base learners are combined through a meta-learner, have further improved upon single-algorithm performance. Research on multi-river Indian datasets demonstrated that stacking diverse regressors reduced mean absolute error by 15–23% compared to the best individual base model, with the performance gains attributed to the complementary error patterns of algorithms trained on different algorithmic principles [11]. CatBoost, designed specifically to handle categorical and mixed-type features without preprocessing, has proven particularly valuable in environmental datasets where station metadata and categorical site classifications serve as informative predictors.

The integration of explainability into ML water quality systems represents the most recent and significant methodological development. Makumbura and colleagues demonstrated that combining XGBoost, LightGBM, and Random Forest with SHAP analysis on Indian river water data not only achieved high predictive accuracy ($R^2 = 0.992$) but also revealed COD and BOD as the dominant contamination drivers across sampled sites — insights that aligned with known industrial discharge patterns and validated the SHAP approach for environmental interpretation [12]. Park et al. confirmed that XGBoost with SHAP analysis out performed conventional feature importance methods in identifying the most influential water quality variables, with SHAP providing more reliable attribution even for non-linear parameter interactions [13].

Gaps remain in the existing literature. Most prior studies apply explainability retrospectively to individual models rather than within a stacked ensemble framework. Furthermore, the deployment of such systems as operational real-time web applications — accessible to non-technical water safety officers — has received minimal attention. The present work addresses both gaps by embedding SHAP explainability within a stacked ensemble pipeline and deploying the complete system as an accessible Flask application validated on Indian river data.

3. Proposed Methodology

A. Dataset Acquisition and Characteristics

The dataset comprises 1,987 water quality measurements collected from monitoring stations across major Indian river systems between 2005 and 2014, obtained from the Central Pollution Control Board (CPCB) of India under the National Water Quality Monitoring Programme. Seven physicochemical parameters were selected as predictive features based on their inclusion in BIS and WHO water safety standards: pH (dimensionless), Dissolved Oxygen (DO, mg/L), Biochemical Oxygen Demand (BOD, mg/L), Nitrate (mg/L),Turbidity (NTU)

Electrical Conductivity (µS/cm), and Temperature (°C). The Water Quality Index target variable was computed using the weighted arithmetic method as per BIS IS:10500:2012 specifications, yielding a continuous scale from 0 (optimal quality) to above 100 (unsuitable for consumption).

B. Preprocessing and Feature Engineering

Raw measurements underwent a four-stage preprocessing pipeline. Missing values were imputed using column-wise median substitution, chosen over mean imputation due to its robustness against the right-skewed distributions characteristic of contamination measurements. Outliers were identified using the interquartile range criterion (boundaries at Q1 − 1.5×IQR and Q3 + 1.5×IQR) and replaced with the nearest boundary value, preserving dataset size while removing physically implausible readings. All seven features were normalized to the [0, 1] interval using min-max scaling, eliminating unit-scale disparities between parameters such as conductivity (0–2000 µS/cm) and pH (0–14). The dataset was split into training (80%, n=1,589) and test (20%, n=398) subsets using stratified sampling to maintain WQI class distribution across both partitions.

C. Stacked Ensemble Learning Architecture

The predictive engine employs a two-tier stacking architecture. The first tier comprises six base regressors selected for their algorithmic diversity: XGBoost Regressor (300 estimators, learning rate 0.05, max depth 6), CatBoost Regressor (300 iterations, depth 6), Random Forest Regressor (200 trees, sqrt feature sampling), Gradient Boosting Regressor (200 estimators, learning rate 0.1), Extra Trees Regressor (200 estimators), and AdaBoost Regressor (100 estimators, learning rate 1.0). Each base learner was trained using 5-fold stratified cross-validation on the training set, generating out-of-fold predictions that form a 6-column secondary feature matrix without data leakage. The second tier consists of a Linear Regression meta-learner that learns optimal combination weights for the six base predictions, producing the final WQI estimate. Hyperparameters for all base learners were selected through RandomizedSearchCV with 5-fold cross-validation, optimizing for minimum RMSE on the training set.

D. SHAP Explainability Integration

Feature contribution analysis was conducted using the SHAP Tree Explainer applied to the XGBoost base model the highest-performing individual algorithm in the ensemble. Tree Explainer was selected over the model-agnostic Kernel Explainer due to its exact computation guarantee and $O(TLD^2)$ complexity versus the $O(n^2)$ approximation overhead of kernel methods, enabling real-time explanation generation within the web application's response time budget. Four SHAP visualization categories are computed for every prediction: (i) Global Summary Plot ranking parameters by mean absolute SHAP value across all test samples; (ii) Force Plot illustrating individual prediction contributions relative to the dataset mean WQI (base value); (iii) Waterfall Plot decomposing the cumulative path from base value to final prediction; and (iv) Dependence Plot revealing non-linear interaction patterns between parameter pairs. All visualisation are granted as matplotli figures encoded to base64 PNG strings,and embedded directly in HTTP response to eliminate file system and dependencies

E. Risk Classification and Recommendation Engine

Predicted WQI scores are mapped to five risk categories aligned with BIS IS:10500:2012 and WHO 2022 drinking water guidelines : Category 1-Excellent (WQI 0–25): Water is safe for all uses without treatment; Category 2 -Good (WQI 26–50): Suitable for drinking with standard disinfection; Category3- Poor (WQI 51–75): Requires advanced treatment before consumption; Category 4 Very Poor (WQI 76–100): Unsuitable for drinking; restricted to industrial or agricultural use;Category5-Unsuitable (WQI>100):Severely contaminated; immediate quarantine and remediation required. For each prediction, the three parameters with the highest positive SHAP values are identified as primary contamination drivers. The recommendation engine generates parameter-specific guidance linking the identified driver to its exceedance ratio relative to BIS limits and suggesting a corresponding treatment approach for instance, biological oxidation for elevated BOD, ion exchange for excess nitrate, or coagulation-flocculation for turbidity.

F. Web Application Architecture

The complete system is deployed as a full-stack web application using Flask 2.3 as the Python backend framework, with HTML5, CSS3, and JavaScript (Bootstrap 5) forming the responsive frontend interface. MySQL 8.0 manages persistent storage for user sessions, prediction history, and exported reports. The backend exposes a REST API endpoint that accepts JSON-encoded water quality parameters, invokes the stacked ensemble model and SHAP explainer, and returns prediction results with embedded visualization data. The frontend provides role-differentiated interfaces for three user types: Water Safety Officers (prediction and reporting), Supervisors (trend dashboards and batch analysis), and Administrators (user management and model retraining scheduling). Docker containerization ensures consistent deployment across development, staging, and production environments.

Parameter	Unit	BIS Limit	WHO Limit	SHAP Rank
pH	—	6.5–8.5	6.5–8.5	2nd
DO	mg/L	≥ 6.0	≥ 5.0	4th
BOD	mg/L	< 3.0	< 2.0	1st
Nitrate	mg/L	< 45	< 50	3rd
Turbidity	NTU	< 1	< 4	5th
Conductivity	µS/cm	< 750	< 500	6th

4. Results and Discussion

A. Comparative Model Performance

Table II presents evaluation metrics for all six base regressors and the stacked ensemble model on the held-out test partition (n = 398 samples). XGBoost achieves the strongest individual performance (R^2 = 0.9921, MAE = 0.521, RMSE = 0.698), confirming its suitability as the SHAP source model. CatBoost follows closely (R^2 = 0.9908), demonstrating its robustness to feature scale differences without explicit preprocessing. AdaBoost, while achieving the lowest individual R^2 (0.9823), captures distinct decision boundaries that contribute complementary information to the ensemble this is evidenced by the improved ensemble performance despite AdaBoost's weaker standalone accuracy.

The stacked ensemble achieves R^2 = 0.9952, MAE = 0.412, and RMSE = 0.589, outperforming all individual models. The 40.8% reduction in squared prediction error relative to XGBoost alone (0.0048 vs. 0.0079) translates to meaningful practical gains: at borderline WQI values near category boundaries (e.g., WQI ≈ 50 separating Good from Poor), the ensemble correctly classifies 97.3% of samples versus 94.1% for XGBoost alone — a difference directly relevant to remediation decisions. Five-fold cross-validation scores confirm generalizability (CV R^2 = 0.9948), confirming that ensemble performance gains persist beyond the specific test partition.

Model	R^2	MAE	RMSE	CV R^2
XGBoost	0.9921	0.521	0.698	0.9918
CatBoost	0.9908	0.547	0.721	0.9901
Random Forest	0.9887	0.583	0.762	0.9879
Gradient Boosting	0.9874	0.601	0.789	0.9865
Extra Trees	0.9861	0.619	0.814	0.9852
AdaBoost	0.9823	0.672	0.867	0.9811
Stacked Ensemble	0.9952	0.412	0.589	0.9948

TABLE II. Performance Metrics — Individual Models vs. Stacked Ensemble (Test Set, n=398)

B. SHAP Feature Attribution Analysis

Global SHAP analysis over the 398 test samples identifies BOD as the dominant contamination driver, with mean absolute SHAP value of 0.423 confirming that organic matter decomposition from sewage discharge and agricultural runoff constitutes the primary water quality threat in the studied river systems. pH ranks second (mean |SHAP| = 0.287), reflecting the measurable acidification impact of industrial effluents, particullary from chemical processing and battery manufacturing facilities documented

Temperature	°C	< 40	< 25	7th

TABLE I. Physicochemical Parameters, Safety Standards, and Observed SHAP Rankings

near multiple CPCB monitoring stations. Nitrate claims the third position (0.214), consistent with the nitrogen-intensive fertilizer practices prevalent in the Krishna basin's agricultural economy.

A particularly note worthy finding from the SHAP dependence analysis is the non-linear interaction between BOD and dissolved oxygen. At BOD concentrations below the BIS threshold of 3.0 mg/L, the influence of DO on WQI predictions is minimal (mean |SHAP|DO < 0.05). However, at BOD levels exceeding 5.0 mg/L, DO emerges as a critical modifying variable, with high DO concentrations partially compensating for elevated organic load — mean |SHAP|DO rises to 0.19 in this regime. This interaction pattern has direct remediation implications: surface aeration interventions are significantly more effective when deployed at moderately elevated BOD sites than at severely contaminated locations where organic load overwhelms natural oxygenation capacity.

Dissolved oxygen exhibits predominantly negative SHAP values (mean SHAP = −0.156), correctly capturing the inverse relationship between oxygen availability and contamination severity. Temperature and conductivity show the weakest attribution scores (0.061 and 0.093 respectively), suggesting that while these parameters modulate contamination dynamics, they are secondary drivers in the studied river systems a finding consistent with the predominantly chemical rather than thermal contamination profile of industrial watersheds.

C. System Operational Performance

Load testing with Apache JMeter using 100 concurrent virtual users demonstrated a mean end-to-end prediction latency of 187 milliseconds, comfortably within the 500 ms threshold for responsive web interfaces. SHAP computation contributes an additional 43 ms overhead per request a justified performance cost given the explainability benefit. Database query performance for retrieving 12-month prediction history averaged 14 ms on MySQL with indexed station and timestamp columns, supporting real-time dashboard updates without perceptible delay. The system correctly categorized WQI risk levels in 97.3% of test predictions against laboratory-confirmed reference classifications, validating operational deployment readiness.

5. Conclusion

This paper presented an Explainable Artificial Intelligence-based framework for water quality prediction and contamination risk analysis that addresses two fundamental limitations of current monitoring approaches: the speed and cost constraints of laboratory-based assessment, and the opacity of existing ML prediction systems. By combining a stacked ensemble architecture with comprehensive SHAP explainability and deploying the result as an operational Flask web application, the proposed system delivers both state-of-the-art predictive accuracy and actionable contamination attribution in real time.

The experimental findings on 1,987 Indian river water samples confirm that the stacked ensemble achieves $R^2 = 0.9952$ — a 40.8% reduction in squared prediction error over the best individual model — while SHAP analysis reveals that BOD and pH are the primary contamination drivers in the Krishna basin region, with a significant non-linear BOD-DO interaction that carries direct implications for aeration-based remediation strategies. These insights are not merely academic: they provide water safety authorities with the parameter-specific intelligence needed to allocate limited remediation resources to maximum effect.

The system's web deployment ensures that these capabilities reach the field officers, policymakers, and environmental engineers who need them most, without requiring programming expertise or laboratory infrastructure. The approach has direct applicability to State Pollution Control Boards, municipal water supply corporations, and rural water safety programs across India, particularly in regions such as northern Karnataka where groundwater contamination poses documented and ongoing public health risks.

Future research will pursue three extensions: integration with IoT sensor networks enabling automated continuous data ingestion to eliminate manual parameter entry; adaptation of the ensemble model to groundwater-specific datasets incorporating fluoride and heavy metal parameters relevant to the Bidar region; and development of LSTM-based temporal models to generate seasonal WQI forecasts that enable proactive rather than reactive contamination management.

References.

[1] World Health Organization, Drinking Water Fact Sheet, WHO, Geneva, 2023.

[2] Karnataka State Pollution Control Board, Annual Water Quality Report, KSPCB, Bengaluru, 2022.

[3] S. Tyagi, B. Sharma, P. Singh, and R. Dobhal, Water quality assessment in terms of water quality index, American Journal of Water Resources, vol. 1, no. 3, pp. 34–38, 2013.

[4] L. Breiman, Random forests, Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[5] M. H. Hmoud Al-Adhaileh and F. W. Waselallah Alsaade, Modelling and prediction of water quality by using artificial intelligence, Sustainability, vol. 13, no. 8, p. 4259, 2021.

[6] A. Bui, D. Klose, J. Wellmann, and S. Schuster, Predicting water quality index using machine learning, Science of the Total Environment, vol. 760, p. 143968, 2021.

[7] S. M. Lundberg and S. I. Lee, A unified approach to interpreting model predictions, Advances in Neural Information Processing Systems, vol. 30, pp. 4765–4774, 2017.

[8] Y. Yogalakshmi and A. Mahalakshmi, Efficient water quality prediction for Indian rivers using machine learning, International Journal, vol. 5, no. 1, pp. 100–109, 2021.

[9] S. Hassan, A. Shafiq, and M. Al-Zahrani, Machine learning techniques for water quality classification in Indian sub-continent, Journal of Hydrology, vol. 597, p. 126–198, 2021.

[10] T. Chen and C. Guestrin, XGBoost: A scalable tree boosting system, Proceedings of 22nd ACM SIGKDD, pp. 785–794, 2016.

[11] A. Najah Ahmed, F. B. Othman, H. A. Afan, and R. K. Ibrahim, Machine learning methods for better water quality prediction, Journal of Hydrology, vol. 578, p. 124084, 2019.

[12] R. K. Makumbura, L. Mampitiya, N. Rathnayake, and U. Rathnayake, Advancing water quality assessment using machine learning with XAI-SHAP, Results in Engineering, vol. 23, p. 102831, 2024.

[13] J. Park, S. Kim, S. Lee, Y. Kim, et al., Interpretation of ensemble learning to predict water quality using explainable artificial intelligence, Science of the Total Environment, vol. 832, p. 155070, 2022.