EXPLAINABLE ENSEMBLE MACHINE LEARNING FRAMEWORK FOR WATER QUALITY ASSESSMENT USING PHYSICOCHEMICAL INDICATORS

F. Majeed[1], Z. Sohail[1], I. Shabir[1], M. H. Amjad[1], M. M. Aziz[1], M. Azam[1] and T. Ahamd[1]
[1]Institute of Data Science, University of Engineering and Technology (UET), Lahore
[2]Department of Computer Science, University of Engineering and Technology (UET), Lahore
Corresponding Author Email: tauqir_ahmad@uet.edu.pk

ABSTRACT: Water quality monitoring is one of crucial activity for sustainable water resource management and global environmental protection. The conventional assessment techniques mostly rely on the threshold value based analysis of physicochemical data. Such methods may fail to capture complex and nonlinear relationships in the water quality parameters. In our study, a machine learning–based framework has proposed for classification of water quality by using physicochemical parameters. The parameters in dataset includes pH, hardness, solids, conductivity, organic carbon, and turbidity. The proposed methodology for this work includes an enhance data preprocessing strategy for imputation of adaptive missing-values using Mean, KNN and Iterative (MICE). Then evaluation of multiple ML models, logistic regression, support vector machines, random forests, and gradient boosting were performed for optimization of model. The stacking ensemble model combining heterogeneous base learners is developed for the enhancement of model classification and its performance. The efficiency of Model is assessed on evaluation metrics, including accuracies, precision, recall, F1-score, and ROC curves. The results demonstrated an enhanced performance for ensemble-based models compared to individual classifiers. Moreover, explainable artificial intelligence based on Shapley Additive Explanations, known as SHAP, have been adopted to interpret model's prediction. The results show the effectiveness of ensemble machine learning and explainable AI for robust and interpretable water quality assessment. This can be useful offering for a data-driven based decision-support framework for environmental monitoring applications.

Keywords: Water Quality Assessment, Machine Learning, Ensemble Learning, Physicochemical Parameters, Explainable Artificial Intelligence, Environmental Monitoring

(Received 08.10.2025 Accepted 15.12.2025)

INTRODUCTION

Water quality have an important role and significance in managing ecological balance and protecting the public health. The efficient monitoring of the water quality supports a sustainable water resource management. The water quality can be degraded due to natural processes and anthropogenic activities by industrial discharge, agricultural runoff, and urbanization. This imposes a significant risk to aquatic ecosystems and as well as for human societies. For the effective environment management, it is very crucial to continuous monitor and access the water quality for formulation of essential policy implication [1].

Generally, in conventional water quality assessment systems physicochemical parameter, some thresholds and manual laboratory analysis were performed. These systems have only statistical significance. Because often predications generated by such systems fail to capture complex and nonlinear interactions within the dataset features. Also these types of techniques are much time taking and use considerable resources. So their usability becomes limited for real-time predictive analysis. For larger scale environmental data, some modern data-driven approaches provide alternative and efficient solution which can be made scalable for monitoring of water quality [2].

In recent some research machine learning techniques have been applied for the environmental monitoring applications. Many researchers have reported water quality prediction and classification systems based on machine learning models. Some models like support vector machines, and gradient boosting algorithms have shown good performance as compared to statistical methods. Because ML methods can more effectively model the nonlinear relationships within multidimensional data. Some research studies focused on single-model implementation that often neglect issues like missing data handling and class imbalance. To improve robustness and practical applicability of ML based water quality assessment systems, such limitations needs to be reduced [3].

The lack of transparency in model predictions is also one of the major hurdle to use machine learning in environmental sciences. Despite the fact that many of Black-box models provide some predictions but they are limited for providing insights related to the underlying factors that drives the classification. For most of the environmental decision-making systems, understanding the influence of corresponding physicochemical parameters is very important for the development of targeted mitigation strategies. As a result, explainable artificial intelligence (XAI) approaches have drawn interest as a technique to bridge the interpretability and predictive performance gaps [4], [5].

This study proposes an ensemble machine learning based framework for the water quality classification by using physicochemical parameters. The methodology presented in this study employs advanced data preprocessing techniques to enhance robustness of the model. For the preprocessing multiple data imputation techniques were used for handling of missing values which also include adaptive missing-value imputation. Further imbalance-aware model training, has also improve the training efficiency. Various machine learning models were systematically evaluated to develop a stacking ensemble classifier that combines the strengths of heterogeneous learners. Furthermore, explainable AI techniques based on SHAP have been used to interpret model's predictions as well as to identify the most important and influential water quality parameters in the dataset.

LITERATURE REVIEW

Recently the water quality assessment systems have shifted from traditional statistical methods to advanced AI and IoT based real-time monitoring and control systems. Traditional methods like WQI provide useful frameworks but they usually lack the ability to model complex and nonlinear relationships. These systems cannot offer predictive foresights. With the advancement in machine learning and deep learning based research now modern techniques like LSTM and XGBoost are being in use to accurately predict and forecast pollution trends. The integration of IoT and deploying wireless sensor networks have further enhanced their real world applicability [3], [6].

The machine learning (ML) techniques have been increasingly adopted for water quality prediction and classification tasks [1]. For example, Patel et al. [7] reported the use of multiple machine learning based classifiers such as SVM, decision trees, random forests and gradient boosting, for predicting of the water quality classes. For handling of class imbalance, Synthetic Minority Oversampling Technique was employed. For their study Random Forest and Gradient Boosting models have shown the high classification accuracies [7]. In another article, Akhlaq et al. [8] studied the effectiveness of supervised machine learning models for prediction of water quality in glacial lakes and rivers. They implemented the approach using Decision trees, KNN, MLP, SVM, and random forests classifiers. Support Vector Machine and Random Forest models have shown best performance as compared to decision tree and neural network based models [8].

Many researchers recently reported the superior performance of ensemble learning techniques over single machine learning models for similar dataset. Nermeen et al. [9] introduced an ensemble learning framework, EWAIS by combining various classifiers. They used Extra Trees Classifier, KNNs and, AdaBoost based stacked ensemble methodology. High accuracy of 0.89 and an F1-score of 0.85 were achieved. Their methodology utilized SHAP and LIME explainers to show the resilience of ensemble technique and to reveal importance of underlying features.

Elshewey *et al. [10]* conducted an empirical study using a stacking ensemble random forest, extra trees, and XGBoost based learners. The logistic regression meta-learner was developed for water potability classification. The stacking ensemble technique have shown an improved performance accuracy of approximately 69.5 % and F1-score ~70.2 % compared to standalone methods. The results indicated the potential of ensembles have very good even with moderate dataset quality [10].

Even though ensemble approaches can increase prediction accuracy. The addressing of missing data still has been major issue in applied water quality modeling. Makumbura *et al.* [11] reported explainable AI techniques e.g. SHAP, along with ML models. They have shown improved predictive performance and also elucidates the contribution of individual parameters to enhance an interpretable environmental analysis. Li *et al. [12]* also reported SHAP based analysis with an ensemble learning model for water quality prediction. They showed the use of explainable AI helped to identify key predictors like turbidity and rainfall.

These research efforts have indicated a growing recognition and importance of explainability in water quality monitoring models.

METHODOLOGY

This research proposes multi-stage machine learning framework for the water quality classification using a public dataset. Proposed methodology present to have five main stages, e.g. data acquisition, data preprocessing and imputation, feature scaling and selection, model trainings and explainability analysis. The figure 1 shows a proposed workflow.

Unlike existing studies based on a single preprocessing and modeling strategy, this framework presented multiple imputation techniques integrated with explainable AI and stacked ensemble modelling.

Dataset Description: The publically available Water Quality Dataset were downloaded from Kaggle. This dataset contains physicochemical measurements related to 3276 water samples. The dataset contains many characteristic attributes such as pH values of water samples, hardness, amount of solids in sample, chloramines, sulfate concentration, conductivity values, organic carbon, trihalomethanes, and turbidity. This also includes with a binary class prediction variable to indicate the water quality status, '1' for safe /good water quality '0' for Poor water quality.

First of all, the shape and distribution of data were visualized and to see the correct uploading of data in Colab. Figure 2 shows the class distribution of water quality. It indicates that there are more records for '1' against for safe /good water quality as compared to '0' for Poor water quality. This distribution is fair for model training.

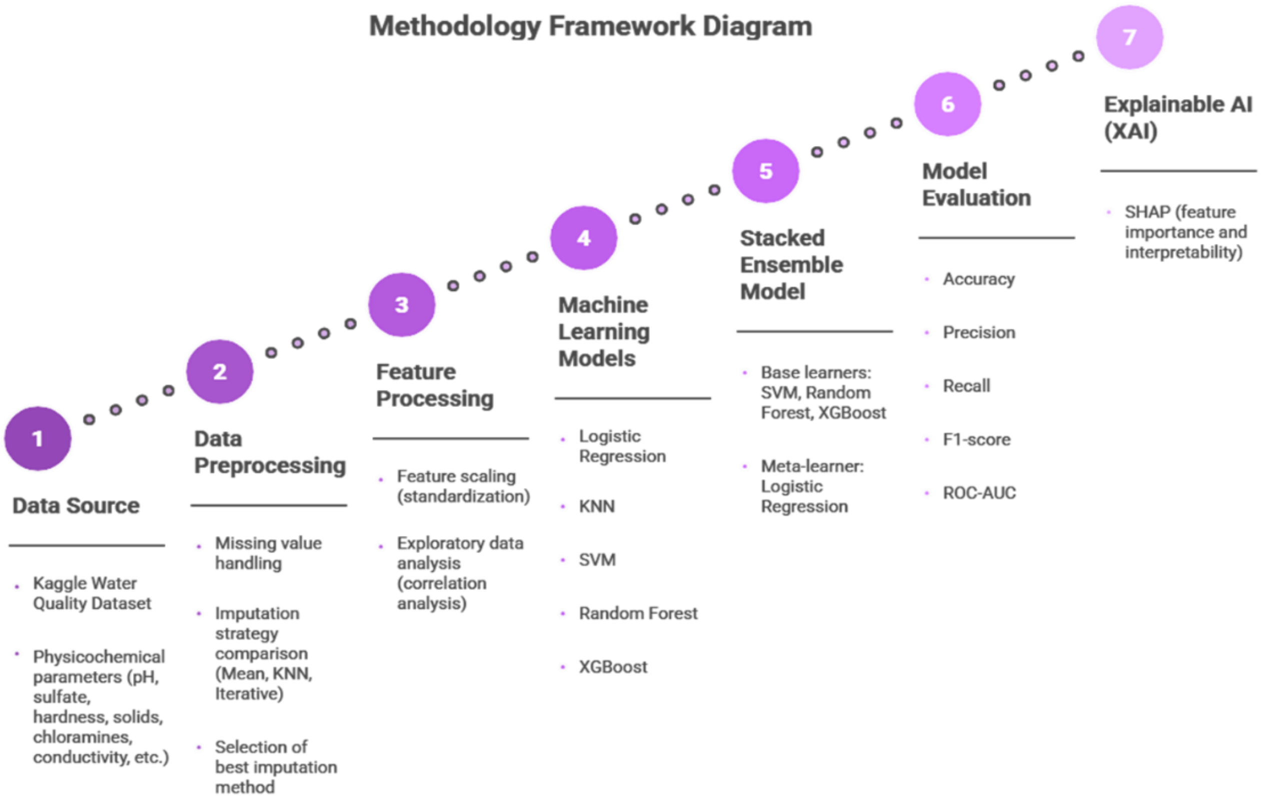
Methodology Framework Diagram
1
Data Source
Kaggle Water Quality Dataset
Physicochemical parameters (pH, sulfate, hardness, solids, chloramines, conductivity, etc.)
2
Data Preprocessing
Missing value handling
Imputation strategy comparison (Mean, KNN, Iterative)
Selection of best imputation method
3
Feature Processing
Feature scaling (standardization)
Exploratory data analysis (correlation analysis)
4
Machine Learning Models
Logistic Regression
KNN
SVM
Random Forest
XGBoost
5
Stacked Ensemble Model
Base learners: SVM, Random Forest, XGBoost
Meta-learner: Logistic Regression
6
Model Evaluation
Accuracy
Precision
Recall
F1-score
ROC-AUC
7
Explainable AI (XAI)
SHAP (feature importance and interpretability)

Figure 1: Methodology Framework Diagram of proposed workflow

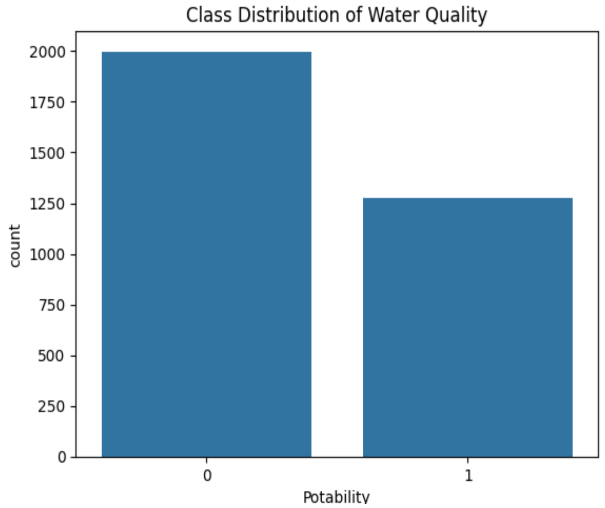

Figure 2: Dataset class distribution

Data Preprocessing: The isnull()function was applied on the dataset to see null values against various features. The table 1 shows the null values in records against each features. The features ph, Sulfate and Trihalomethanes have several missing values/ null records.

Feature	Null Records
ph	491
Sulfate	781
Trihalomethanes	162
Total	1434

This reflects a real-world data collection challenges and motivation to use the robust imputation strategies. To handle missing data efficiently, three imputation strategies are initially evaluated:

- **Mean Imputation:** A baseline approach where missing values are replaced with an approximated value of the feature’s mean.
- **K-Nearest Neighbor (KNN) Imputation:** Missing values are estimated based on feature similarity among nearest neighbors.
- **Iterative Imputation (MICE):** A multivariate imputation approach that models each feature with missing using a sophisticated technique, values is generated as a function of other features.

Each imputation method is then evaluated by using a baseline Random Forest classifier. Then comparative performance analysis was performed to select the best-performing imputation strategy for subsequent experiments.

After the imputation process, feature scaling was performed to standardize all numerical features using z-score normalization as defined by the Eq 1.

$$x' = \frac{x-\mu}{\sigma} \qquad (1)$$

where μ and σ denotes the mean value and value of standard deviation of each feature, respectively.

Model Training: For this study three widely used machine learning classifiers are employed, e.g. Support Vector Machine , Random Forest, and Gradient Boosting (XGBoost). The underlying hyperparameters of each model have been optimized using a grid search with cross-validation.

In this architectural design, a stacked ensemble model is created to take benefit of the complementary strengths of individual classifiers. The stacked ensemble improves generalization by combining heterogeneous decision boundaries. First the base learners consist of SVM, Random Forest, and XGBoost models. Then a logistic regression based model is used as the meta-learner that have been trained on the probabilistic outputs of the base classifiers.

To enhance the interpretability of predictions, Shapley Additive explanations (SHAP) were used on our best-performing model. This quantifies the participation of features to individual predictions and overall model behavior. This also indicates the most influential water quality parameters for model decisions.

All experiments for this study were conducted using Google Colab using Python. Key libraries include NumPy, Pandas, Scikit-learn, XGBoost, and SHAP.

RESULTS AND DISCUSSION

Model performance was evaluated by using the respective results of accuracies, precision, recall, and F1-scores. A train–test splitting was applied to preserve the class distribution. Cross-validation was also applied during hyperparameter tuning to reduce overfitting.

Performance of Individual Machine Learning Models: After performing the suitable imputation, the performance of each model were evaluated. The Table 1 shows the performance of our applied models:

Table 1: Performance metrics of our applied models

Model	Accuracy	Precision	Recall	F1-score
KNN	0.6605	0.5493	0.3603	0.5291
Logistic Regression	0.6684	0	0	0
Random Forest	0.7102	0.6992	0.3681	0.5199
SVM	0.7284	0.6669	0.3759	0.5201
XGBoost	0.7387	0.7916	0.4177	0.5202
Stacking Ensemble	0.7496	0.7312	0.4099	0.5597

Overall, XGBoost and Stacking Ensemble achieved the high performance accuracy as well as F1-score among individual models. This show the benefiting from its gradient-based optimization and Stacking Ensemble for handling of complex nonlinear interactions. The SVM also demonstrated competitive performance with strong robustness. While KNN showed comparatively lower recall due to sensitivity to class overlap in the dataset. These results are consistent with recent studies reporting the effectiveness of tree-based ensemble models for water quality classification tasks.

Feature Importance and Explainable AI (XAI) Analysis: The features correlation in dataset is very important for establishing effective classification as well as to visualize the importance of each feature in decision making. Figure 3 shows a Pearson feature correlation matrix that depicts pairwise linear relationships among the variables in a dataset to highlight the strength and direction of these correlations.

The color scale represents correlation coefficients ranging from −1 to +1. The warmer colors represent stronger positive correlations. The cool colors indicate weak or negligible correlations. The attributes such as solids, conductivity, sulfate, and hardness have shown a mild positive correlation with a warmer blue color around the diagonal at top left corner. This is effect is due to joint association of dissolved minerals content in water. This also reflects that water quality classification depends more on nonlinear interactions among various parameters as compared to any of single dominant variable.

To enhance model's interpretability, explainable artificial intelligence (XAI) technique were employed using Shapley Additive exPlanations (SHAP). SHAP-based explainability analysis was performed on the best-performing model to interpret feature contributions. The SHAP results have shown that sulfate concentration and pH are the most influential features affecting water quality classification. Figure 4 shows a SHAP value indicating impact on model output.

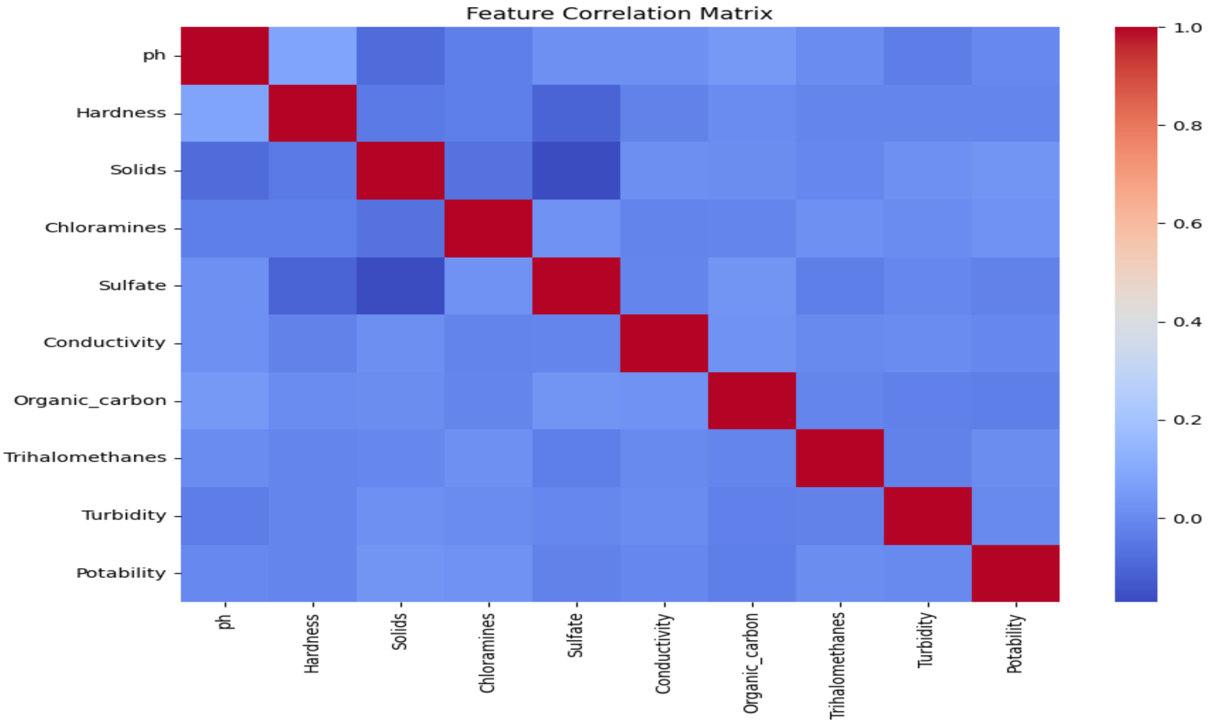

Figure 3: Feature Correlation Matrix

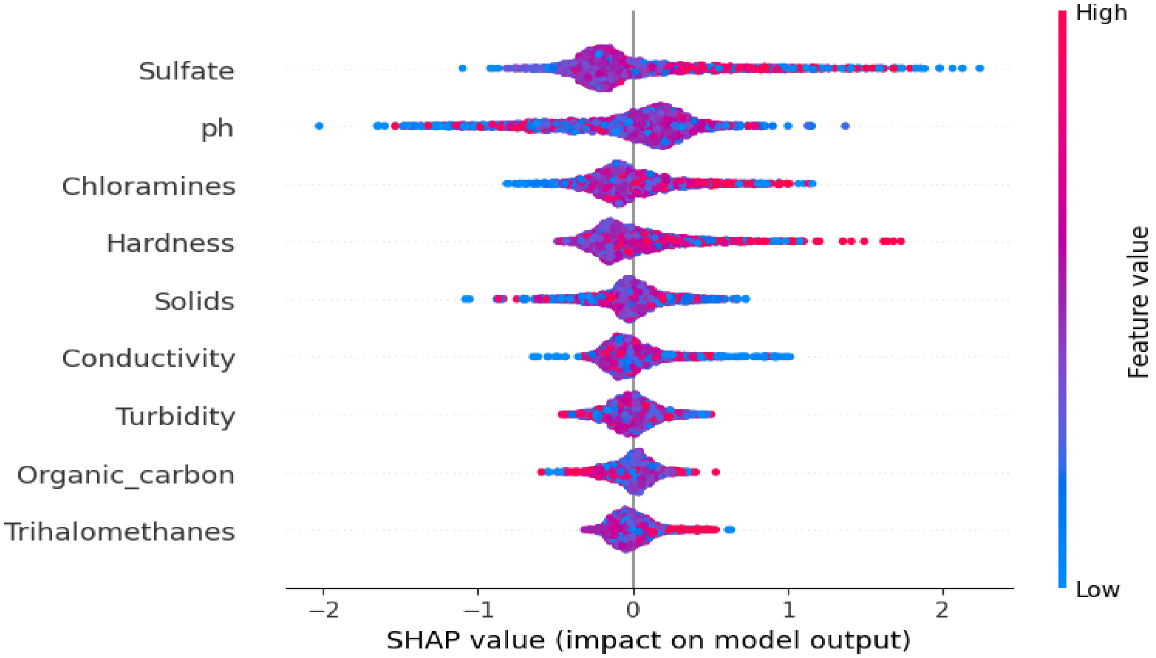

Figure 4:SHAP summary plot illustrating the features importance

Positive SHAP values presented features that increased the probability of a sample being classified as potable. The negative values show that parameters are linked with some degraded water quality. These findings have shown the validity of our framework are well aligned with accepted environmental knowledge. The use of Explainable AI improved the transparency by providing useful information for environmental surveillance. The Figure 5 shows an average impact of features on the output prediction.

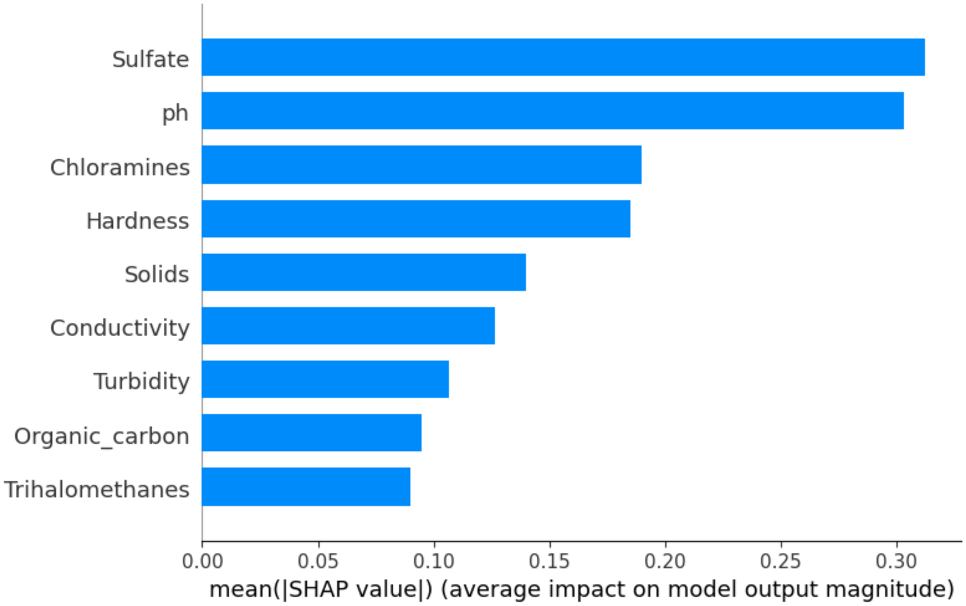

Figure 5:SHAP output magnitude plot illustrating the average impact of features

The Explainable AI bridged the gap between accuracy and interpretability. This enhances its practical adoption for real-world application. This proposed framework has demonstrated potential for scalable and interpretable quality assessment systems.

Conclusion: An explainable machine learning framework for water quality classification were presented. Multiple data imputation techniques were adopted handle the missing data. The iterative imputation method has shown a superior performance over the other two methods by preserving multivariate relations in dataset. Experiments were conducted to train the models using Logistic Regression, KNN, SVM, RF and XGBoost. Their results showed that the ensemble learning models have outperformed compared to other classifiers. The stacked ensemble achieved highest predictive accuracy and robustness, with an accuracy of 0.7496 and F1-score of 0.5597. This indicates a well balance between precision and recall, comparing to individual classifiers.

To visualize the transparency of feature importance, Explainable Artificial Intelligence were used. The SHAP were employed to interpret model's predictions. The XAI analysis identified values of sulfate, pH, chloramines, and hardness are the most influential factors affecting the overall water quality. This approach also offer an effective and interpretable solution for water quality monitering. Such systems and can be useful in making data-driven based decision support systems for marine and freshwater monitoring. Future work can incorporate spatio-temporal data and real-time IoT based sensor neworks to further extend the applicability of this framework.

REFERENCES

[1]. N. A. A. Aziz, “A review of water quality forecasting and classification using machine learning models and statistical analysis,” Water, vol. 17, no. 15, Art. no. 2243, 2025. doi: 10.3390/w17152243.

[2]. Y. Chen, J. Li, and X. Wang, “Machine learning methods for water quality assessment: A comprehensive review,” Water, vol. 14, no. 9, Art. no. 1389, 2022. doi: 10.3390/w14091389.

[3]. S. Shrestha and F. Kazama, “Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan,” Environmental Modelling & Software, vol. 22, no. 4, pp. 464–475, 2007, doi: 10.1016/j.envsoft.2006.02.001.

[4]. M. H. Rahman et al., “EWAIS: An ensemble learning and explainable artificial intelligence approach for water quality classification,” Processes, vol. 12, no. 12, Art. no. 2771, 2024. doi: 10.3390/pr12122771.

[5]. S. Ahmed, A. Khan, and M. Z. Khan, “Water-quality prediction based on H2O AutoML and explainable artificial intelligence techniques,” Water, vol. 15, no. 3, Art. no. 475, 2023. doi: 10.3390/w15030475.

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

[7]. Patel, Jinal, Charmi Amipara, Tariq Ahamed Ahanger, Komal Ladhva, Rajeev Kumar Gupta, Hashem O. Alsaab, Yusuf S. Althobaiti, and Rajnish Ratna. "A Machine Learning-Based Water Potability Prediction Model by Using Synthetic Minority Oversampling Technique and Explainable AI." *Computational Intelligence and Neuroscience* 2022, no. 1 (2022): 9283293.

[8]. Akhlaq, Muhammad, Asad Ellahi, Rizwan Niaz, Mohsin Khan, Saad Sh Sammen, and Miklas Scholz. "Comparative analysis of machine learning algorithms for water quality prediction." *Tellus* 76, no. 1 (2024).

[9]. Rezk, Nermeen Gamal, Samah Alshathri, Amged Sayed, and Ezz El-Din Hemdan. "EWAIS: An Ensemble Learning and Explainable AI Approach for Water Quality Classification Toward IoT-Enabled Systems." *Processes* 12, no. 12 (2024): 2771.

[10]. Elshewey, Ahmed M., Rasha Y. Youssef, Hazem M. El-Bakry, and Ahmed M. Osman. "Water potability classification based on hybrid stacked model and feature selection." *Environmental Science and Pollution Research* 32, no. 13 (2025): 7933-7949.

[11]. Makumbura, Randika K., Lakindu Mampitiya, Namal Rathnayake, D. P. P. Meddage, Shagufta Henna, Tuan Linh Dang, Yukinobu Hoshino, and 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* 23 (2024): 102831.

[12]. Li, Lingbo, Jundong Qiao, Guan Yu, Leizhi Wang, Hong-Yi Li, Chen Liao, and Zhenduo Zhu. "Interpretable tree-based ensemble model for predicting beach water quality." *Water Research* 211 (2022): 118078.