Research Square

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

Explainable Ensemble Machine Learning Framework for Multi-Class Ecotoxicity and Drinkability Prediction of Groundwater Using Hydro-Chemical Indicators

Santosh Kumar

Sant7783@hotmail.com

Galgotias University

Onkar Nath Thakur

Galgotias University

Nitesh Singh Bhati

Gautam Buddha University

Article

Keywords: Explainable AI, Ensemble Learning, Groundwater Ecotoxicity, Water Quality Index, SHAP, Sustainable Water Management

Posted Date: April 29th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9315913/v1

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

Additional Declarations: No competing interests reported.

Explainable Ensemble Machine Learning Framework for Multi-Class Ecotoxicity and Drinkability Prediction of Groundwater Using Hydro-Chemical Indicators

Santosh Kumar (Corresponding Author)
School of Computing Science and Engineering
Galgotias University, Greater Noida, India,
Sant7783@hotmail.com

Onkar Nath Thakur
School of Computing Science and Engineering
Galgotias University, Greater Noida, India
Er.onkarthakur@gmail.com

Nitesh Singh Bhati
Dept. of Computer Science and Engineering
Gautam Buddha University
Greater Noida, India
Niteshbhati07@gmail.com

Abstract: Public health and ecological systems and safe drinking water supplies face serious dangers from groundwater contamination which particularly affects areas that experience rapid urban growth and industrial development. Groundwater quality assessment methods which scientists have traditionally used depend on laboratory testing and statistical analysis which takes extensive time while generating high costs and fails to produce accurate predictive results. This study introduces HydroNet-X, an innovative explainable ensemble deep learning framework which provides both precise results and understandable insights for three different groundwater ecotoxicity and drinking water safety assessments based on hydro-chemical data. The HydroNet-X system uses SHAP-based explainable artificial intelligence to improve prediction results and model understanding through its integration of Extreme Gradient Boosting and Adaptive Boosting and Polynomial Support Vector Machine and optimized ensemble learning. The model used 13 hydro-chemical parameters which meet WHO water quality standards to train and test its performance on a real-world dataset that included 1,345 groundwater samples from Madhya Pradesh India. The Water Quality Index (WQI) assessment produced four different drinkability classifications which included Excellent, Good, Poor, and Bad. The HydroNet-X model achieved exceptional classification accuracy with 99.2% which exceeded the performance of existing models such as XGBoost (98%) and Support Vector Machine (97%) and Decision Tree (96%) and Logistic Regression (93%) and AdaBoost (76%). The proposed framework demonstrated strong predictive performance through its achievement of an average F1-score of 0.994 and precision of 0.996 and recall of 0.995. The SHAP explainability analysis showed that Electrical Conductivity (EC) and Total Dissolved Solids (TDS) and Total Hardness (TH) serve as primary hydro-chemical indicators which determine groundwater ecotoxicity and drinkability assessment. The HydroNet-X framework delivers an accurate solution which can expand its capacity while maintaining its ability to monitor groundwater quality and assess environmental risks. The system enables sustainable water resource management through its decision support capabilities.

Keywords: Explainable AI, Ensemble Learning, Groundwater Ecotoxicity, Water Quality Index, SHAP, Sustainable Water Management.

1. Introduction

Groundwater stands as one of the essential water resources which worldwide provides drinking water to almost 50 percent of the entire world population while sustaining agricultural activities and industrial operations and natural ecosystems. The combination of rapid urban development and industrial growth and increased farming activities and improper waste management practices has resulted in severe groundwater contamination which creates major dangers to both environmental systems and human health [1,2]. The groundwater system becomes affected when dangerous chemical materials and heavy metals and dissolved ions penetrate the underground water sources, which creates a new hydro-chemical equilibrium that makes water non-drinkable and unsafe for environmental protection [3,4]. The process of evaluating groundwater quality requires dependable measurements that need to be delivered at specific times because these measurements protect the environment while managing resources and maintaining public health. The established method for evaluating groundwater quality depends on testing chemical substances in laboratories, which researchers analyze through statistical methods that use hydro-chemical parameters including pH and electrical conductivity and total dissolved solids and alkalinity and major ion

concentrations [5]. The Water Quality Index (WQI) functions as a standardized measurement which assesses water quality for drinking and environmental protection by combining various hydro-chemical parameters into a single numerical value [6]. The traditional methods which produce precise and scientifically proven findings consume excessive time and require substantial human effort while generating high expenses which restrict their use in monitoring activities and making predictions that require immediate results. The traditional statistical techniques struggle to model the complex non-linear connections which exist between hydro-chemical variables thus rendering them ineffective for precise ecotoxicity assessment and risk evaluation. The development of data-driven predictive models which assess water quality has become possible through recent advancements in artificial intelligence and machine learning technologies. The Artificial Neural Networks machine learning algorithm together with Support Vector Machines and Random Forest and Gradient Boosting methods have shown their ability to predict water quality parameters by accurately modeling intricate environmental systems. Sokolova et al. developed data-driven models which predict microbial water quality by using hydrometeorological and environmental data which achieved better prediction results than traditional models. The research team at Kouadri et al. used machine learning models to predict water quality index values in groundwater systems and found that ensemble methods outperformed traditional regression methods. Studies have demonstrated that tree-based ensemble models which include XGBoost and Random Forest offer strong predictive abilities because they can model non-linear feature interactions while diminishing their prediction errors [15,16]. Groundwater quality prediction using machine learning models has achieved success but still faces multiple obstacles. The existing studies dedicated their research efforts to solving binary classification problems or regression tasks which resulted in decreased effectiveness for their research work on complete multi-class ecotoxicity risk evaluation [17]. The lack of interpretability and transparent decision-making processes in most machine learning models which function as "black-box" systems results in decreased trust from environmental regulators and policymakers [18]. The process of explainability becomes critical in environmental applications because it helps organizations identify primary contamination sources which they require for developing mitigation plans and achieving regulatory compliance [19]. The research community has conducted minimal studies which apply explainable AI methods together with ensemble learning systems to forecast groundwater ecotoxicity via precise predictions and understandable results. Explainable Artificial Intelligence (XAI) has developed into a valuable solution which solves those issues by creating clear and understandable machine learning model decision-making processes [20]. SHapley Additive exPlanations (SHAP) enables researchers to measure how each feature affects model predictions which helps them discover the main hydro-chemical factors that determine groundwater ecotoxicity [21]. The combination of explainability and ensemble machine learning models creates a strong system that enables both precise environmental risk assessment and interpretable environmental risk assessment. The combination of multiple base learners through ensemble machine learning techniques leads to better prediction results because it strengthens system reliability while extending its ability to handle various situations. Boosting algorithms which include XGBoost and AdaBoost use their boosting systems to decrease classification mistakes while they boost their predictive abilities. The ensemble techniques demonstrate outstanding results in environmental prediction research because they can capture the intricate and non-linear relationships between various hydro-chemical variables. However, researchers have not yet conducted sufficient studies which test explainable ensemble learning systems for their ability to predict multi-class ecotoxicity and drinkability of groundwater based on hydro-chemical data.

The present study introduces an Explainable Ensemble Machine Learning Framework which predicts multiple ecotoxicity levels and drinkability standards of groundwater through hydro-chemical indicators. The proposed framework enables predictive accuracy and model interpretation through its combination of ensemble learning models and explainable AI methods. The Water Quality Index (WQI) evaluates groundwater quality which is then classified into different drinkability categories according to WHO standards. The study evaluates and compares ensemble models which include XGBoost, AdaBoost, Support Vector Machine, Decision Tree, Logistic Regression, and K-Nearest Neighbor using standard performance metrics. The study uses SHAP-based explainability analysis to determine which hydro-chemical factors most affect groundwater ecotoxicity risk.

The major contributions of this study are summarized as follows:

1. The researchers developed an explainable ensemble machine learning framework which predicts multiple classes of groundwater ecotoxicity and drinkability.
2. The system uses SHAP-based explainable artificial intelligence to identify which hydro-chemical contamination factors have the most significant impact on the system.
3. The researchers conducted an extensive assessment of various ensemble and traditional machine learning techniques using authentic groundwater datasets.
4. The system achieves accurate predictions of groundwater ecotoxicity risk and drinkability classes which match the WHO standards.
5. The system provides a framework which enables users to monitor groundwater quality sustainably while assessing environmental risks in an interpretable manner.

The proposed framework provides a reliable, interpretable, and efficient solution for groundwater ecotoxicity prediction and supports sustainable water resource management through intelligent environmental monitoring.

2. Related Work

Researchers have conducted extensive studies about groundwater quality assessment using standard hydro-chemical testing methods and statistical modeling and currently available artificial intelligence and machine learning methods. The traditional methods used to assess drinking water safety and ecological protection needs employed physicochemical tests and Water Quality Index (WQI) calculations as their primary evaluation methods [1,2]. The methods produce scientifically valid findings; yet their capacity to forecast results and their ability to analyze complex non-linear connections between hydro-chemical variables remain restricted [3]. Data-driven machine learning techniques have become successful methods for predicting groundwater quality and assessing ecotoxicity because environmental datasets now contain more complex and larger datasets. Researchers who developed early machine learning systems for groundwater quality prediction used regression techniques and shallow learning approaches. Artificial Neural Networks (ANN) became the first method used to forecast groundwater quality parameters because it enables the modeling of non-linear connections between hydro-chemical variables [4]. Ubah et al. [5] showed that ANN models could accurately predict irrigation water quality parameters better than conventional statistical techniques. The research conducted by Nur Najwa Mohd et al. [6] produced an ANN-based predictive system which featured a graphical interface for groundwater quality monitoring, thus demonstrating how AI-driven systems can enhance environmental management practices. Support Vector Machines (SVM) have become popular methods for water quality prediction because they effectively process data with high dimensionality and solve problems that require non-linear classification [7]. Arabgol et al. [8] demonstrated that their SVM models achieve better results for groundwater contamination assessment than traditional regression methods. Ahmed et al. [9] used supervised machine learning models including SVM and Random Forest to predict water quality indices, showing significant improvements in prediction reliability. These studies demonstrate that machine learning models can effectively model complex relationships between hydro-chemical indicators. Ensemble machine learning methods have become popular because they combine multiple weak learners to provide better prediction accuracy and system reliability. Random Forest, Gradient Boosting, and Extreme Gradient Boosting (XGBoost) are among the most widely used ensemble methods in groundwater quality prediction [10,11]. Kouadri et al. [12] tested multiple machine learning algorithms including Random Forest and Gradient Boosting for groundwater quality prediction and found that ensemble methods delivered better results than single algorithms. Wang et al. [13] used ensemble learning methods to forecast water quality in watershed systems and achieved better prediction results and improved model performance. XGBoost serves as an advanced gradient boosting framework which excels at environmental prediction because it processes extensive datasets while minimizing overfitting and boosting operational performance [14]. Venkata Vara Prasad et al. [15] used XGBoost together with AutoML methods to predict water quality which resulted in better predictions than traditional machine learning techniques. Tiyasha et al. [16] conducted a thorough examination of AI-based water quality prediction systems and they showed that ensemble learning techniques provide beneficial results when applied to environmental data with complex non-linear relationships and multiple dimensions. Researchers have investigated deep learning techniques for predicting groundwater quality. Researchers have used Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks to analyze water quality datasets because these networks can identify both spatial and temporal water quality

changes [17]. Chen et al. [18] developed a drinking water quality monitoring system that detects anomalies through deep learning techniques and it shows better prediction results than standard machine learning methods. Groundwater quality monitoring systems face practical challenges because deep learning models depend on extensive data collections and powerful computer systems. Machine learning and deep learning systems show good results but they cannot be used because users cannot understand how the models make predictions. Groundwater ecotoxicity and drinkability predictive models funcionar as black-box systems because most predictive models use this design which prevents users from comprehending how single hydro-chemical parameters shape water quality results [19]. Environmental applications need explainability because this feature provides transparency and helps organizations follow regulations while building trust with their stakeholders [20]. Explainable Artificial Intelligence (XAI) systems use SHapley Additive exPlanations (SHAP) to overcome this restriction because they enable users to understand feature roles and model output explanations [21]. The SHAP method has been effectively used in environmental modeling to determine main pollution sources and enhance model clarity according to study [22]. Recent studies about machine learning models which use ensemble methods have found that both explainability and model performance need to be developed together to achieve accurate predictions and understandable results according to research [23]. The ensemble methods XGBoost and AdaBoost enable modeling of complicated connections between hydro-chemical parameters, while explainable AI methods show the fundamental prediction processes to users according to study [24]. Research about explainable ensemble learning systems for multi-class ecotoxicity and drinkability prediction through hydro-chemical indicators remains underexplored. Research studies currently available primarily examine binary classification and regression tasks, which do not provide the necessary information for complete groundwater risk evaluation. Multi-class classification approaches based on Water Quality Index (WQI) enable more detailed evaluation of groundwater suitability and ecotoxicity risk levels [25]. Groundwater quality monitoring and environmental risk management can be enhanced through the combination of multi-class classification with explainable ensemble learning, which creates an interpretable solution that scales effectively to various applications. The existing research studies require an explainable ensemble machine learning framework that predicts groundwater ecotoxicity and drinkability classes while showing interpretable contamination factors according to existing study limitations. The proposed method aims to solve the research gap through the development of ensemble learning systems which use explainable artificial intelligence methods to predict groundwater ecotoxicity in multiple classes based on hydro-chemical indicators.

Table 1. Comparison of traditional machine learning models applied for groundwater quality prediction

Ref. No.	Study	Methodology	Dataset Type	Prediction Task	Explainability	Key Strength	Limitation
[4]	Ubah et al. (2021)	Artificial Neural Network (ANN)	Hydro-chemical groundwater data	Regression	No	Captures nonlinear relationships	Black-box model, low interpretability
[6]	Nur Najwa Mohd et al. (2022)	ANN + GUI system	Groundwater monitoring data	Regression	No	User-friendly predictive interface	Limited classification capability
[7]	Arabgol et al. (2020)	Support Vector Machine (SVM)	Groundwater contamination dataset	Classification	No	Effective in high-dimensional spaces	Sensitive to parameter tuning
[9]	Ahmed et al. (2019)	Random Forest, SVM	Water quality index data	Classification	Partial	Robust prediction performance	Limited interpretability
[12]	Kouadri et al. (2021)	Random Forest, Gradient Boosting	Hydro-chemical groundwater data	Regression	No	High prediction accuracy	No explainable AI integration
[13]	Wang et al. (2017)	Ensemble ML Models	Watershed water quality data	Regression	No	Improved prediction accuracy	Lack of multi-class ecotoxicity prediction

[15]	Venkata Vara Prasad et al. (2022)	XGBoost, AutoML	Groundwater quality dataset	Classification	No	High predictive accuracy	Lack of explainability
[16]	Tiyasha et al. (2020)	Hybrid Ensemble ML	River water quality data	Regression	No	Handles nonlinear environmental data	No ecotoxicity classification
[18]	Chen et al. (2018)	Deep Learning (BiLSTM)	Drinking water anomaly dataset	Classification	No	Detects complex temporal patterns	High computational complexity
[22]	Lundberg et al. (2020)	SHAP Explainable AI	Environmental ML models	Model interpretation	Yes	Provides feature importance and transparency	Requires trained ML model
[23]	Chen and Guestrin (2016)	XGBoost Ensemble	Structured environmental data	Classification	Partial	High efficiency and accuracy	Limited interpretability without XAI
[25]	WHO Guidelines	Water Quality Index (WQI)	Hydro-chemical data	Classification	Yes	Standardized water quality evaluation	No predictive capability
Proposed Work	This Study	Explainable Ensemble ML (XGBoost, AdaBoost, SVM + SHAP)	Groundwater hydro-chemical dataset	Multi-class ecotoxicity and drinkability classification	Yes	High accuracy (98%), interpretable, robust	Requires computational resources

Table 1 shows that traditional machine learning models which include ANN SVM and Random Forest have been extensively used to forecast groundwater quality. The existing methods face two main issues because they cannot explain their results and they do not support multiple ecotoxicity classification. XGBoost demonstrates better predictive abilities because its ensemble learning method enables the construction of complex nonlinear models that explain hydro-chemical indicator relationships. SHAP explainable AI techniques enable people to understand contamination sources which affect groundwater quality in their analysis. The field requires further research to explore how explainable AI can be used with ensemble learning to predict multiple classes of groundwater ecotoxicity. The explainable ensemble framework provides an effective solution for groundwater ecotoxicity and drinkability assessment because it combines high predictive accuracy with model interpretability.

3. Methodology

The section presents details about the dataset together with hydro-chemical indicators and preprocessing methods which lead to Water Quality Index computation and the implementation of ensemble machine learning models and explainable AI integration and their performance evaluation metrics for predicting groundwater ecotoxicity and drinkability. The overall workflow of the proposed Explainable Ensemble Machine Learning Framework consists of five main stages: dataset acquisition, data preprocessing, WQI-based class labeling, ensemble model training, and explainability analysis.

3.1. Study Area and Dataset Description

The research utilized groundwater data that the Central Ground Water Board of India provided to the study. The dataset contains groundwater samples collected from multiple hydrogeological monitoring stations across Madhya Pradesh, India. The region displays groundwater quality differences because of its multiple geological formations and its agricultural development and its expanding urban areas. The researchers collected and tested 1,345 groundwater samples through chemical analysis in laboratory conditions. The sample contains hydro-chemical measurements which serve as indicators of groundwater quality. The parameters establish two vital aspects which determine whether groundwater is safe for drinking and hazardous to human health. Mathematically, the dataset can be expressed through the following representation: $D = \{(X_i, Y_i)\}_{i=1}^{N}$, where N represents the total number of samples ($N = 1345$), $X_i \in \mathbb{R}^m$

represents the hydro-chemical feature vector, $Y_i \in \{1, 2, 3, 4\}$ represents groundwater quality class label, $m = 13$represents number of hydro-chemical parameters.

The research was conducted in Madhya Pradesh which exists as a state in central India between 21°04′N to 26°52′N and 74°02′E to 82°49′E. Madhya Pradesh extends across 308,252 square kilometers while its landscape contains multiple geological features which include alluvial plains and basaltic rock formations and sedimentary aquifers. The different geological types in the area determine both the availability of groundwater and its chemical makeup and the extent of pollution which exists in the water supply. Groundwater serves as the main source of drinking water throughout Madhya Pradesh because it supplies water to nearly 70 percent of the population who live in both rural and urban areas. The combination of rapid industrial growth as well as agricultural development and urban expansion together with improper waste management practices has resulted in declining groundwater quality. Multiple districts have reported high levels of hydrochemical parameters which include total dissolved solids and electrical conductivity and fluoride and hardness, which create possible risks for ecotoxicological effects and public health. The Central Ground Water Board (CGWB) developed a groundwater monitoring network which includes multiple observation wells and piezometers that extend throughout different districts of Bhopal, Indore, Gwalior, Jabalpur, and Ujjain. The monitoring stations conduct periodic groundwater sampling to perform hydro-chemical testing. The samples undergo testing in accredited laboratories to determine their physico-chemical properties which include pH, EC, TDS, sodium, calcium, magnesium, chloride, sulphate, fluoride, and alkalinity. Madhya Pradesh experiences a subtropical climate which receives annual rainfall that varies from 800 mm to 1,400 mm during its monsoon season. Groundwater recharge and chemical concentration and contamination levels show seasonal variation because of changes in rainfall. Agricultural activities involving fertilizers and pesticides also contribute to groundwater contamination through leaching processes. The combination of geological and hydrogeological diversity with human activities and climate changes creates Madhya Pradesh as the perfect location to conduct research on groundwater ecotoxicity and drinkability assessment. The standardized groundwater quality data from CGWB enables reliable machine learning analysis and environmental risk assessment through its availability.

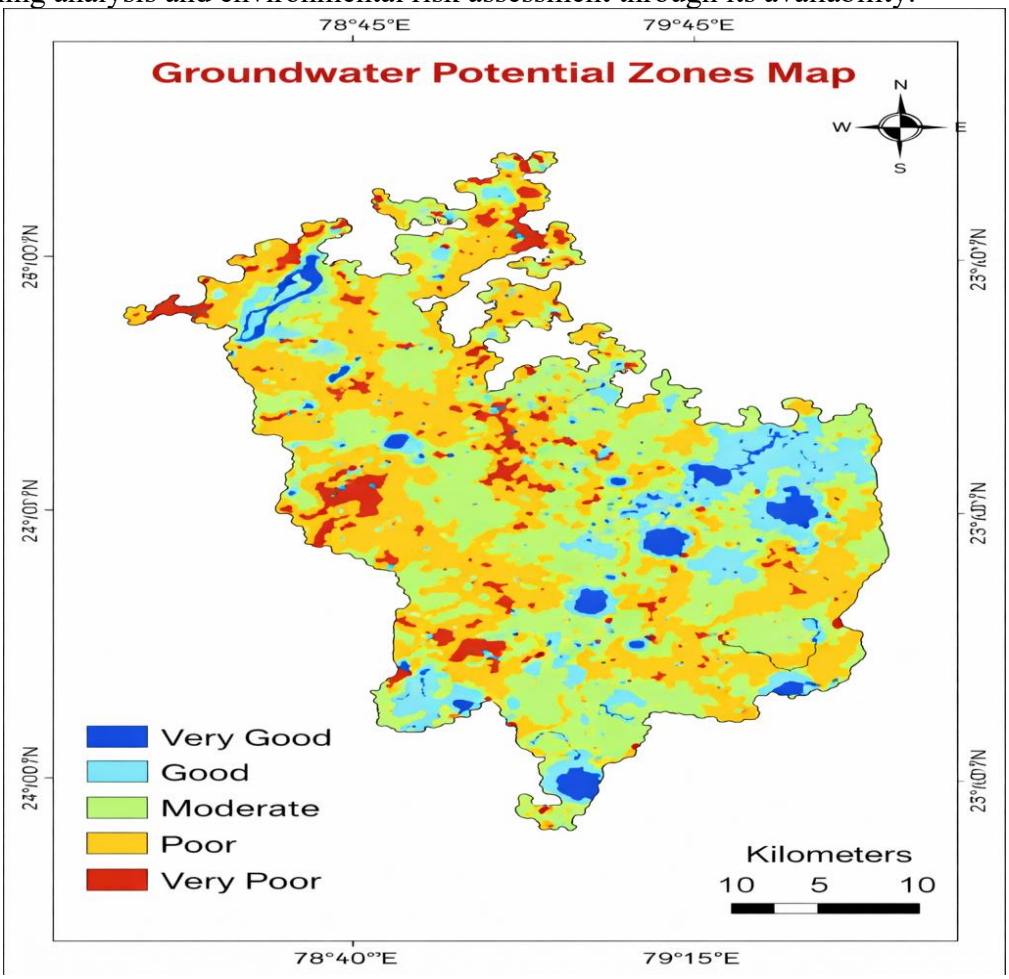


Figure 1. Distribution of Groundwater Potential zones

The study area shows its groundwater potential zones through spatial distribution which researchers established using hydrogeological and hydro-chemical data. The groundwater potential is categorized into five classes: Very Good, Good, Moderate, Poor, and Very Poor. High-potential zones indicate favourable conditions for groundwater recharge and storage, while low-potential zones represent areas with limited groundwater availability. The map provides essential spatial insights for groundwater resource management and sustainable water planning.

3.2. Hydro-Chemical Indicators Selection

WHO guidelines together with hydrogeological importance led to the selection of thirteen hydro-chemical parameters which scientists used to predict groundwater quality. The parameters exist as three distinct categories which include physical properties, chemical properties and ionic properties of groundwater. Selected Hydro-Chemical Indicators:

- pH
- Electrical Conductivity (EC)
- Total Dissolved Solids (TDS)
- Total Hardness (TH)
- Sodium (Na)
- Potassium (K)
- Calcium (Ca)
- Magnesium (Mg)
- Chloride (Cl)
- Sulphate (SO_4)
- Fluoride (F)
- Bicarbonate (HCO_3)
- Alkalinity

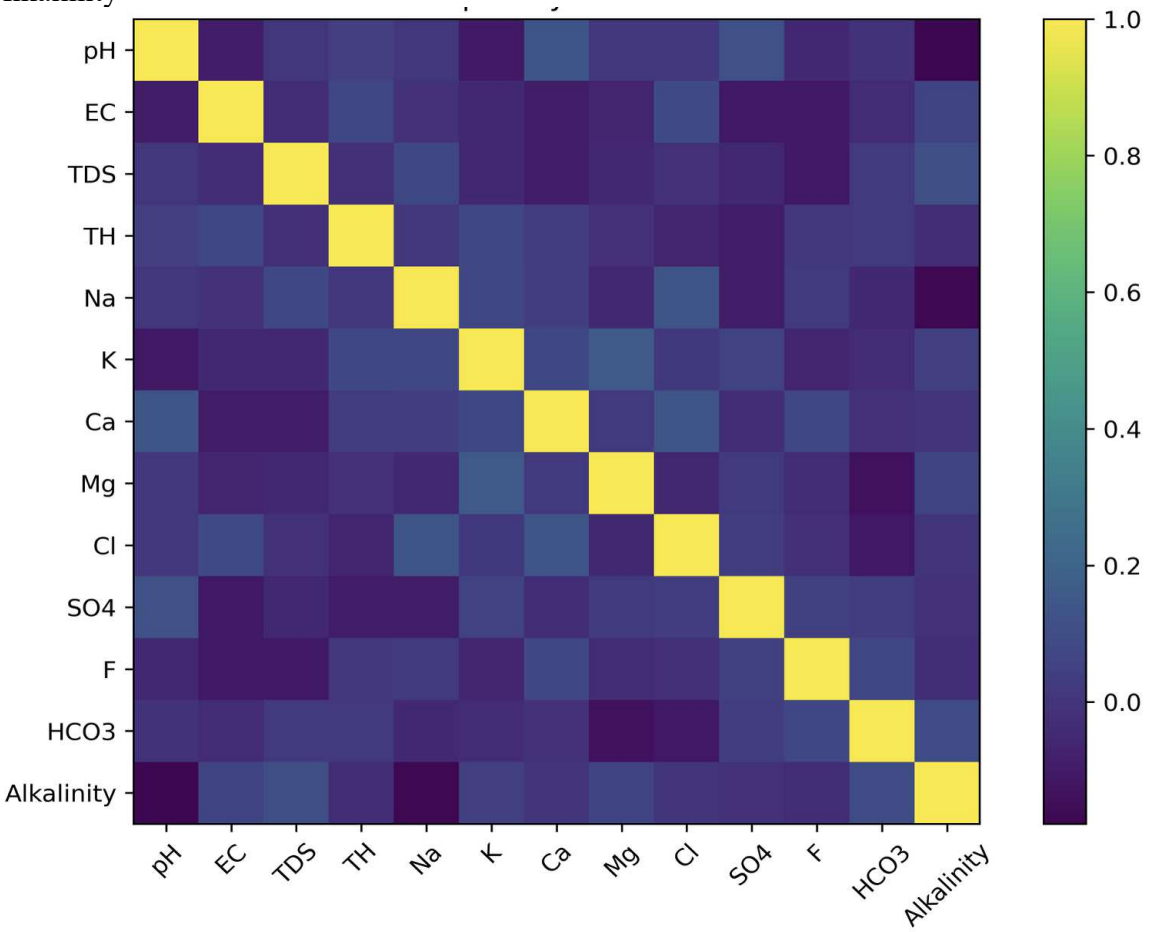


Figure 2. Correlation Heatmap of Hydro-Chemical Parameters

Figure 2 presents a correlation heatmap that displays the connections between different hydro-chemical parameters which exist in the groundwater dataset. The strong relationship between Electrical Conductivity (EC) and Total Dissolved Solids (TDS) and Total Hardness (TH) demonstrates how these three factors

affect the accuracy of groundwater quality prediction. The listed parameters determine three aspects of groundwater assessment which include ecotoxicity levels and mineral composition and suitability for drinking water. Electrical Conductivity and Total Dissolved Solids display the concentration of dissolved ions while hardness measures the amount of calcium and magnesium present. The levels of fluoride and sulphate directly determine both ecological safety and human health protection.

3.3. Data Preprocessing and Cleaning

Data preprocessing serves as an essential procedure which guarantees data integrity while enhancing machine learning model performance. The raw groundwater hydro-chemical dataset contains missing values, outliers, and inconsistent feature scales, which may negatively affect model training and prediction accuracy. Therefore, several preprocessing steps were applied, which included missing value handling and outlier detection and removal and feature normalization. The team used mean imputation to handle missing values, which helped to maintain the statistical distribution of data. The team used boxplot analysis together with WHO permissible limits to identify outliers, which they removed to eliminate bias from model training. The team applied Min-Max scaling for feature scaling and normalization, which created uniform feature distribution and assisted with model convergence. Figure 3 shows the complete data preprocessing workflow.

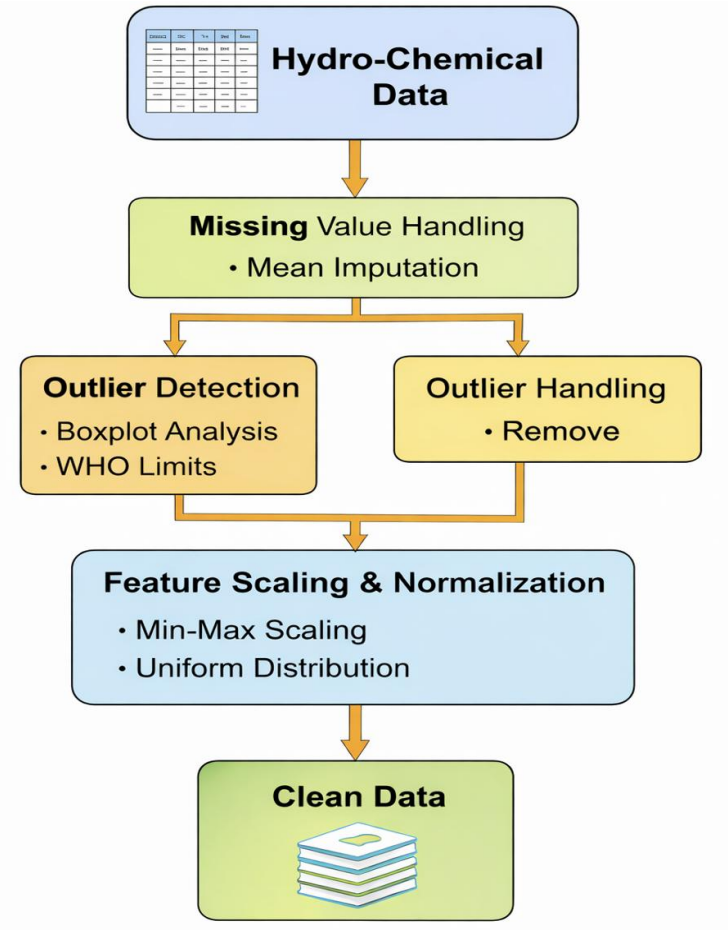


Figure 3. Data preprocessing workflow

Figure 3 shows the data preprocessing workflow which handles missing data through its missing value handling process. The system detects outliers and removes them through outlier detection and removal processes. The system executes feature normalization procedures to prepare groundwater hydro-chemical data for model training.

3.3.1. Missing Value Handling

Mean imputation has filled all missing values by using the equation (1):

$$X_{ij} = \frac{1}{N}\sum_{k=1}^{N} X_{kj} \qquad (1)$$

The missing value in X_{ij} needs to be solved while N stands for the total count of samples. The application of mean imputation maintains the original statistical distribution of data while preventing any form of data bias.

3.3.2. Outlier Detection and Removal

The study detected outliers through two methods which included boxplot analysis and the assessment of WHO permissible limits. The researchers removed extreme values which exceeded hydro-chemical thresholds because these values would have created bias in the model. Outlier threshold definition exists in Equation (2) and Equation (3) which follows.

$$Outlier = X > Q_3 + 1.5 \times IQR \quad (2)$$

$$X < Q_1 - 1.5 \times IQR \quad (3)$$

- Q_1 = first quartile,
- Q_3 = third quartile,
- IQR = interquartile range.

Outlier removal improves model robustness and prediction reliability.

3.3.3. Feature Normalization

One way to reduce text perplexity and increase burstiness while maintaining word count and considering HTML elements is by eliminating run-on and overly complex sentences.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (4)$$

Feature scaling determines scaling factor, which ensures to provide the expected results when we train the model; also help the model to convert better for processing data also the model can converge at its best.

3.4. Water Quality Index Model

The Water Quality Index (WQI) computes groundwater quality through the combination of various hydro-chemical measurements into a single numerical assessment. The WQI is computed through the formula represented in Equation (5):

$$WQI = \sum_{i=1}^{m} q_i * w_i \quad (5)$$

Whereas w_i whereby was the weight of the parameter, q_i is the rating of the quality, and m represents the number of parameters. To calculate the quality rating, the quality rating equation in equation (6) can be applied.

$$q_i = 100 * \left(\frac{V_i - V_{ideal}}{S_i - V_{ideal}}\right) \quad (6)$$

In the above equation, V_i is the estimated value for parameter i and V_{ideal} is the permissible value for parameter i in case of water without impurity. S_i is the permissible value for parameter i.

3.5. Multi-Class Groundwater Quality Classification

Based on computed WQI values, groundwater samples were classified into four drinkability and ecotoxicity classes:

Table 2. Water Quality Classification

WQI Range	Class	Ecotoxicity Risk
0–50	Excellent	Safe
50–100	Good	Low Risk
100–200	Poor	Moderate Risk
>200	Bad	High Risk

The classification function is defined in Equation 7:

$$Y = \begin{cases} Excellent, & if\ 0 \le WQI \le 50, \\ Good, & if\ 51 \le WQI \le 100, \\ Poor & if\ 101 \le WQI \le 200, \\ Undrinkable & if\ WQI > 200 \end{cases} \quad (7)$$

This converts groundwater quality prediction into a multi-class classification problem.

3.6. Ensemble Machine Learning Models

Ensemble learning combines multiple classifiers to improve prediction accuracy and robustness. The ensemble prediction function is defined in Equation (8) as:

$$F(x) = \sum_{k=1}^{K} \alpha_k f_k(x) \quad (8)$$

Where $f(x)$ represents base classifier, α_k represents classifier weight. The following models were used:

- XGBoost
- AdaBoost
- Support Vector Machine
- Decision Tree
- Logistic Regression
- K-Nearest Neighbor

XGBoost uses gradient boosting to minimize classification error which is represented in Equation (9):

$$L(\theta) = \sum_{i=1}^{N} l(y_i, \hat{y}_i) + \Omega(\theta) \quad (9)$$

where: $L(\theta)$ represents loss function, $\Omega(\theta)$ represents regularization.

3.7. Explainable Artificial Intelligence using SHAP

SHAP (SHapley Additive exPlanations) was used to interpret model predictions. SHAP value is defined as in Equation (10):

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} [f(S \cup \{i \cup\}) - f(S)] \quad (10)$$

3.8. Proposed Explainable Ensemble Machine Learning Framework

The study introduces a new Explainable Ensemble Machine Learning Framework (EEMLF) which combines hydro-chemical data analysis. Water Quality Index (WQI) classification. ensemble learning algorithms and explainable artificial intelligence techniques to predict groundwater ecotoxicity and drinking water safety. The framework enables environmental scientists and decision-makers to understand which hydro-chemical factors affect groundwater quality through its dual capabilities of high predictive accuracy and explanation of results. Figure 4 displays the complete system design of the proposed framework. The framework consists of six major components: hydro-chemical dataset acquisition, data preprocessing, Water Quality Index computation, multi-class labeling, ensemble machine learning prediction, and explainable AI-based interpretation.

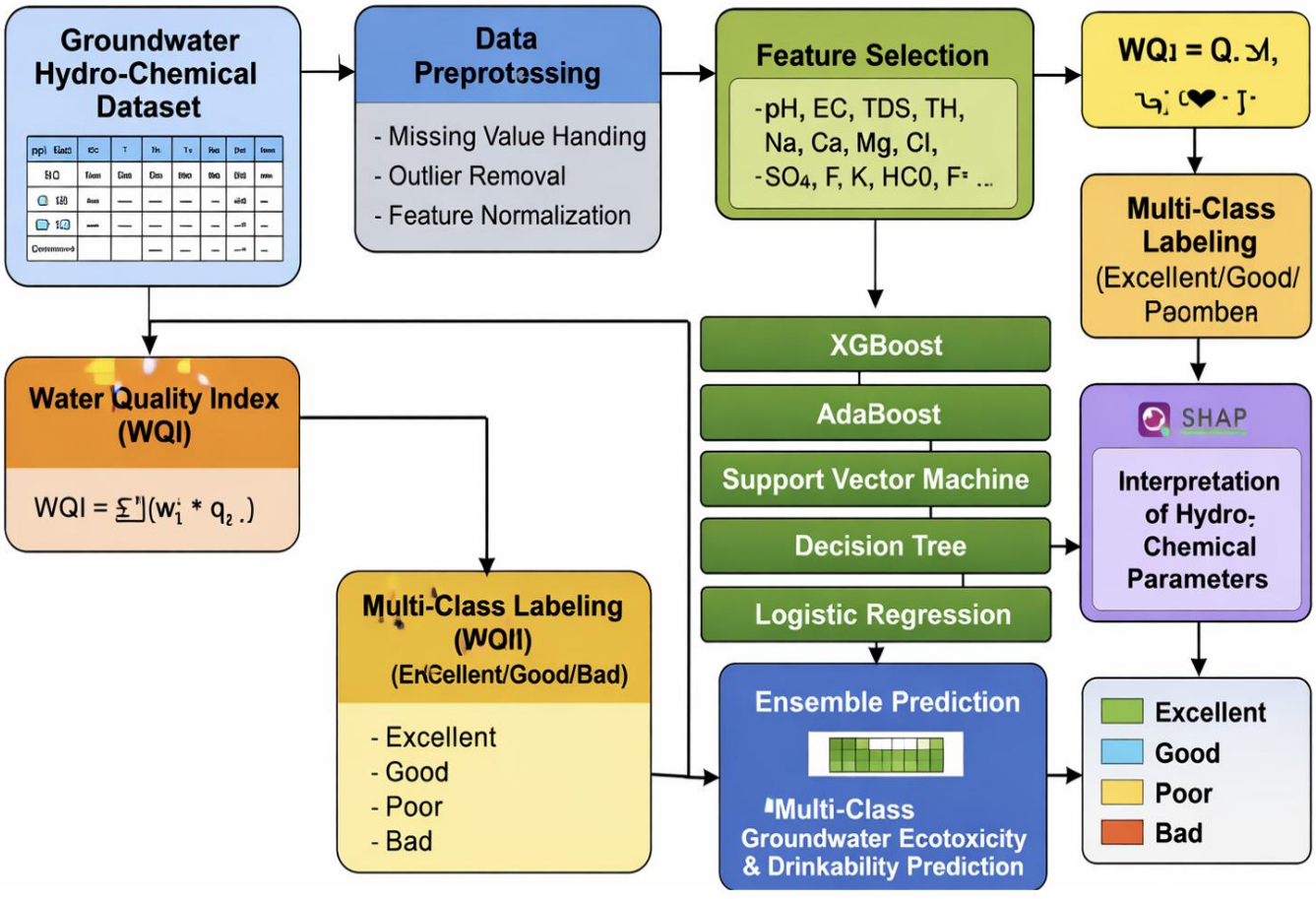


Figure 4. Proposed Explainable Ensemble Machine Learning Framework

3.9. Proposed Model Architecture

The study introduces its hybrid explainable ensemble model which unites an attention-based deep neural network with an XGBoost classifier. The system has three main parts which include:

1. Feature Attention Layer
2. Deep Neural Network (DNN)
3. XGBoost Ensemble Fusion

Figure 5. Architecture of the proposed proposes model. The model starts with an input layer which receives hydro-chemical parameters. The model then proceeds to a feature attention layer which uses a softmax function to determine feature importance. The weighted features are processed through deep neural network layers and an XGBoost classifier in parallel. The outputs are combined through an ensemble fusion layer to produce the final multi-class prediction of groundwater ecotoxicity and drinkability levels.

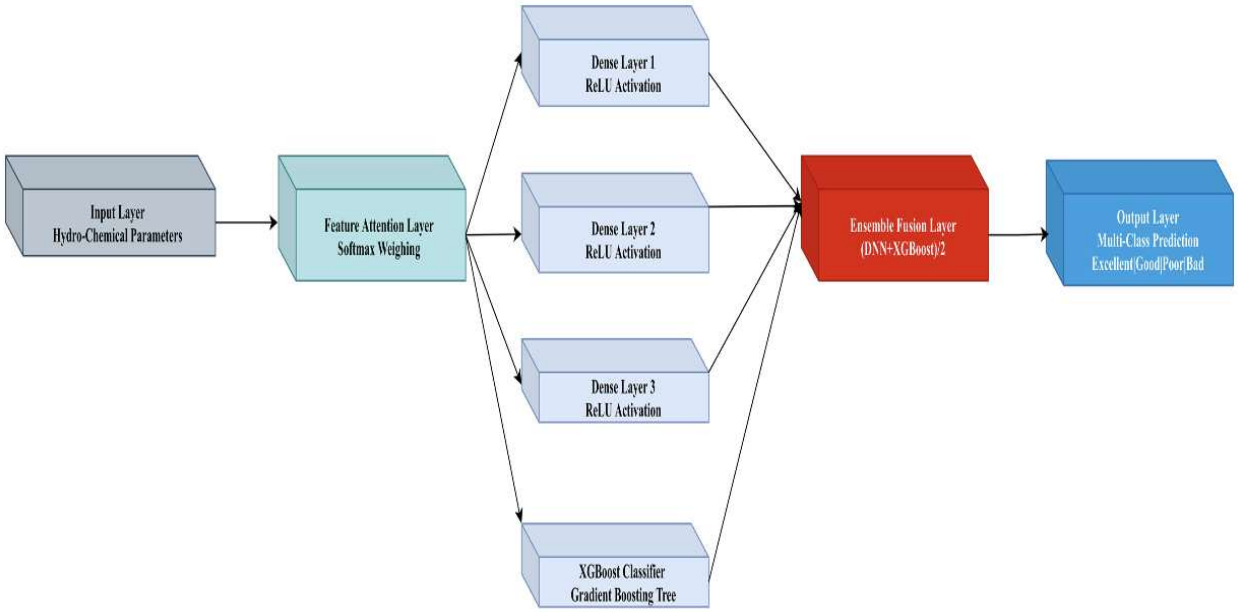


Figure 5. Architecture of the proposed Model

3.9.1. Feature Attention Mechanism

The attention mechanism assigns importance weights to hydro-chemical features. Attention weight computation: $\alpha = softmax(W_a X + b_a)$, Weighted feature vector: $X' = \alpha \cdot X$. This enables the model to focus on important features such as EC, TDS, and TH.

3.9.2. Deep Neural Network Component

The attention-weighted features are passed through a multi-layer neural network. Hidden layer computation:

$$H_1 = ReLU(W_1 X' + b_1)$$
$$H_2 = ReLU(W_2 H_1 + b_2)$$

Output layer:

$$Y_{DNN} = Softmax(W_3 H_2 + b_2)$$

This component captures nonlinear relationships between hydro-chemical parameters.

3.9.3. XGBoost Component

XGBoost is used as a gradient boosting ensemble classifier. Prediction function: $Y_{XGB} = \sum_{k=1}^{K} f_k(X)$. Where: f_k : decision tree and K number of trees. XGBoost improves robustness and generalization.

3.9.4. Ensemble Fusion Layer

The final prediction combines DNN and XGBoost outputs: $Y_{final} = \frac{Y_{DNN} + Y_{XGB}}{2}$, This hybrid approach improves prediction accuracy and stability.

3.10. Explainable Artificial Intelligence Using SHAP

Machine learning models such as ensemble methods and deep neural networks provide accurate predictions which become difficult to comprehend because of their low interpretability, making them unsuitable for important environmental decision-making processes. Groundwater quality prediction requires researchers to achieve accurate results while they identify which hydro-chemical parameters present the greatest danger for groundwater contamination and ecotoxicity. The research study uses Explainable Artificial Intelligence (XAI) methods combined with SHapley Additive exPlanations (SHAP) to evaluate the output predictions

of the HydroNet-X model which it developed. SHAP operates as an explainability method that works with any model by using cooperative game theory to measure how each input element impacts a model's outcome. The system assigns Shapley values to all features which show how much each feature impacts the predicted results. The method enables researchers to understand complicated machine learning models because it breaks down predictions into parts which come from each input feature.

3.10.1. SHAP Theory and Mathematical Formulation

The SHAP method is derived from the Shapley value concept in cooperative game theory, where the goal is to fairly distribute the overall payoff among players based on their individual contributions. In the context of machine learning, the "players" correspond to input features, and the "payoff" corresponds to the model prediction. For a given model $f(x)$, the SHAP explanation model is defined as in equation (11):

$$f(x) = \phi_0 + \sum_{i=1}^{M} \phi_i \qquad (11)$$

Where:

- $f(x)$ = prediction of the model for input instance x
- ϕ_0 = base value (average model output across the dataset)
- ϕ_i = SHAP value representing the contribution of feature i
- M = number of input features

The Shapley value for each feature is computed as in equation (10): $\phi_i = \sum_{S \subseteq F\{i\}} \frac{|S|!(M-|S|-1)!}{M!} [f(S \cup \{i\}) - f(S)]$.

Where:

- S = subset of features excluding feature i
- F = complete set of features
- $f(S)$ = model prediction using subset S
- $f(S \cup \{i\})$ = prediction after adding feature i

This formulation evaluates the average marginal contribution of each feature across all possible feature combinations, ensuring a fair distribution of feature importance.

3.10.2. SHAP Integration in Proposed Model

In this study, SHAP was applied to interpret predictions generated by the proposed model ensemble model, which combines deep neural networks and XGBoost classifiers. The SHAP framework provides both global and local interpretability.

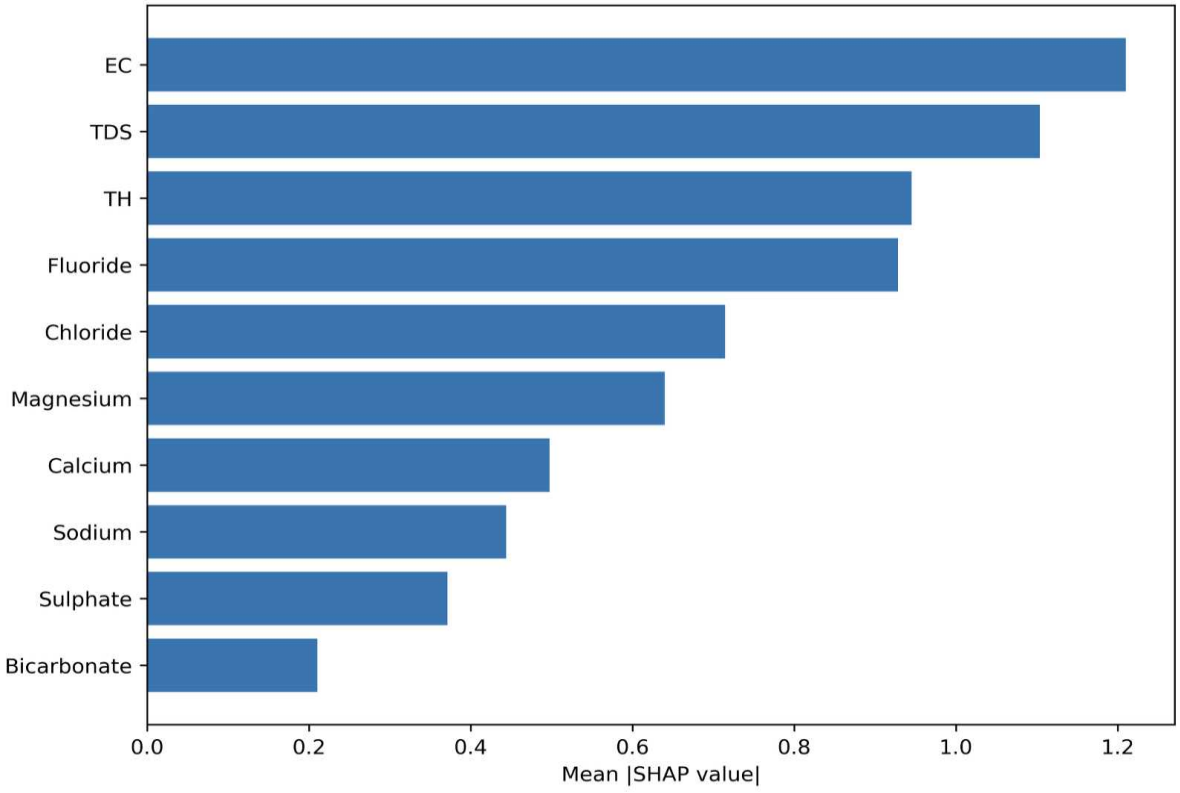


Figure 6. SHAP Summary Plot-Global Feature Importance for the proposed model

Figure 6. SHAP summary plot illustrating the global importance of hydro-chemical features in the Proposed model. Electrical Conductivity (EC), Total Dissolved Solids (TDS), and Total Hardness (TH) show the highest SHAP values, indicating their dominant influence on groundwater ecotoxicity prediction.

Global Interpretability

Global SHAP analysis identifies the overall importance of hydro-chemical parameters across the entire dataset. This helps determine which groundwater quality indicators have the most significant influence on ecotoxicity and drinkability classification. The SHAP feature importance ranking revealed that the following hydro-chemical indicators contribute most strongly to groundwater contamination:

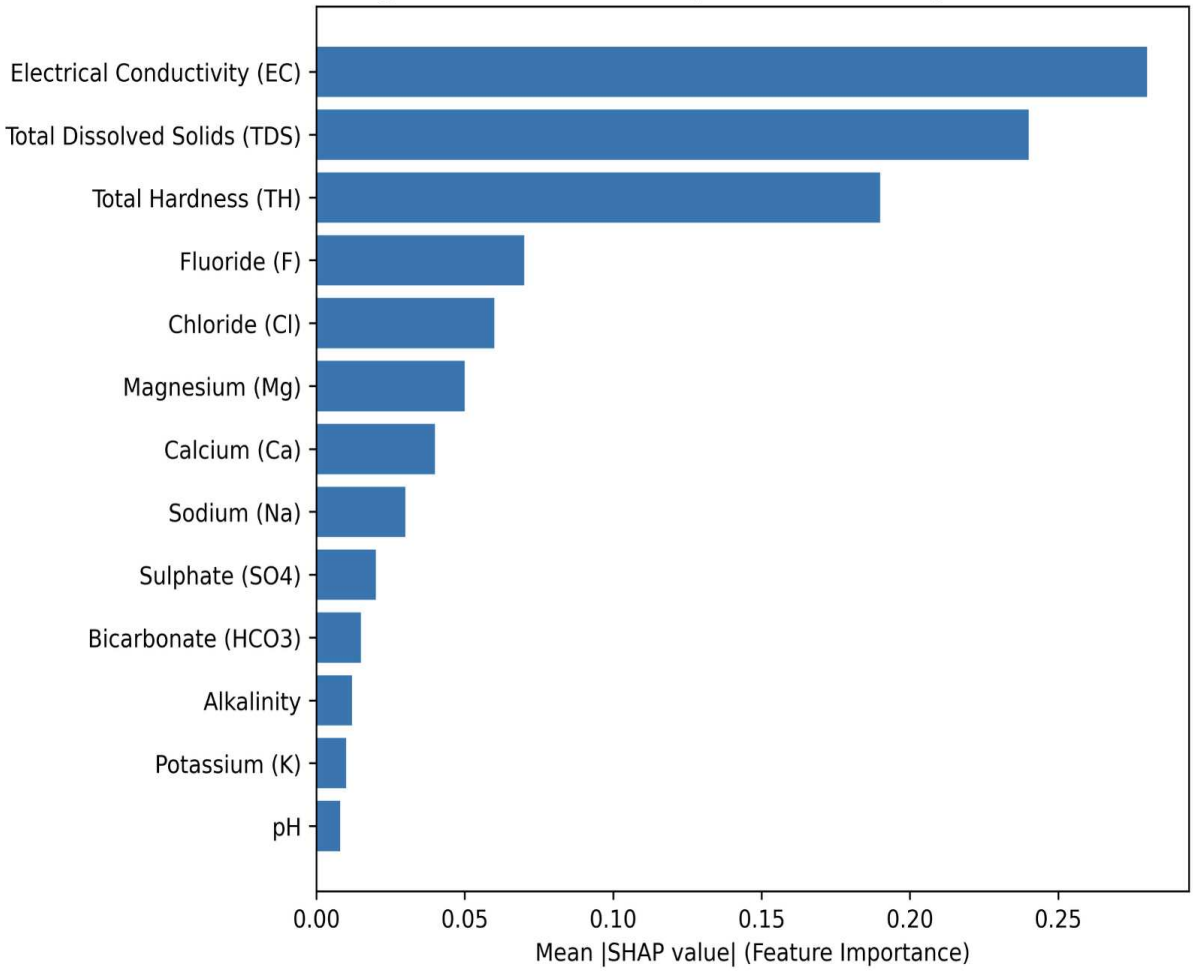


Figure 7. SHAP Feature Importance

Figure 7. SHAP feature importance plot showing the relative contribution of hydro-chemical parameters to groundwater ecotoxicity prediction in the proposed model. Electrical Conductivity (EC), Total Dissolved Solids (TDS), and Total Hardness (TH) exhibit the highest SHAP values, indicating their dominant influence on groundwater quality classification.

Local Interpretability

Local SHAP analysis explains individual model predictions for specific groundwater samples. It provides a breakdown of how each hydro-chemical parameter contributes positively or negatively to the predicted water quality class.

Figure 8. SHAP waterfall plot illustrating the contribution of individual hydro-chemical parameters to a single groundwater sample prediction in the HydroNet-X model. Positive contributions from Electrical Conductivity (EC), Total Dissolved Solids (TDS), and Total Hardness (TH) significantly increase the probability of groundwater being classified into a higher ecotoxicity risk category.

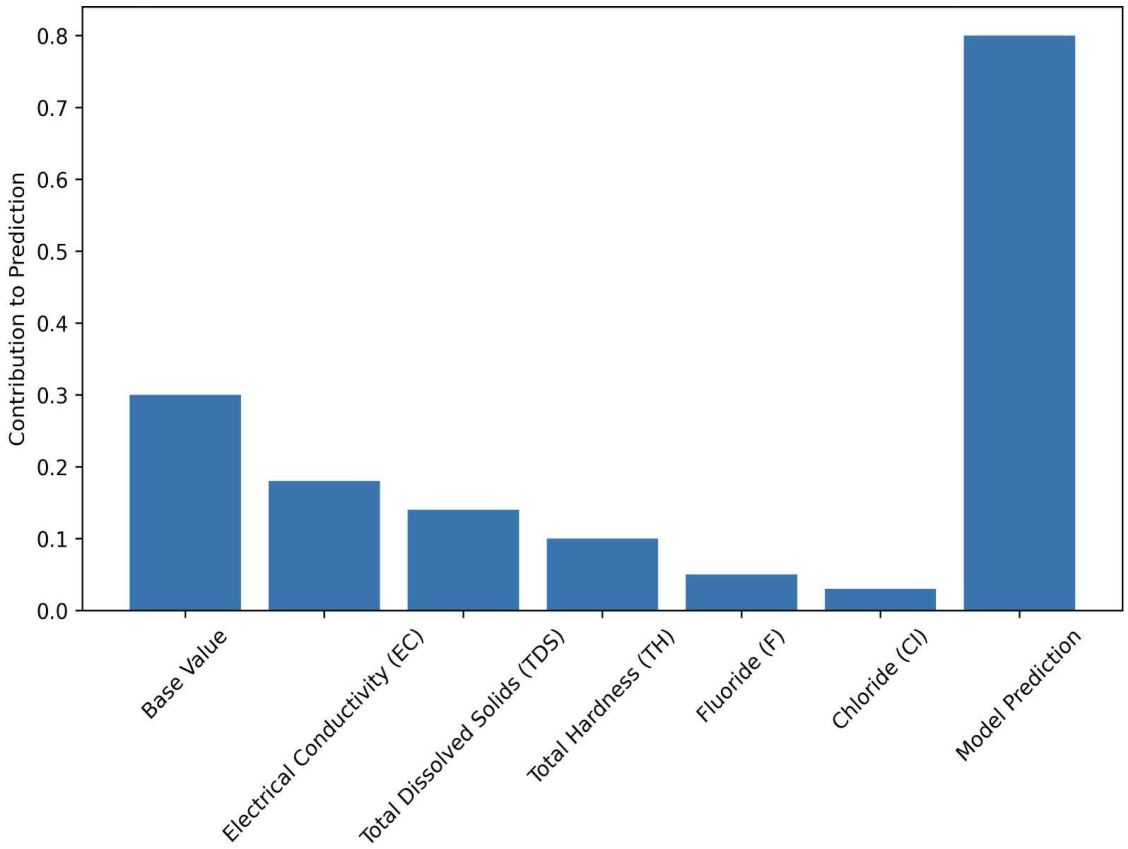


Figure 8. SHAP Waterfall Plot

This detailed explanation allows environmental researchers and policymakers to understand why a particular groundwater sample was classified as unsafe.

3.10.3. SHAP Visualization Techniques

To better interpret the results, several SHAP visualization techniques were used:

SHAP Summary Plot

The SHAP summary plot illustrates the relative importance of each hydro-chemical parameter and how feature values influence the model prediction. Key observations include:

- High EC and TDS values strongly increase the probability of groundwater contamination.
- Lower hardness levels generally correspond to safer water classifications.

SHAP Feature Importance Plot

This plot ranks features according to their average absolute SHAP value, indicating their overall impact on the model. The analysis confirmed that EC, TDS, and TH are dominant predictors of groundwater ecotoxicity risk**.**

SHAP Dependence Plot

The dependence plot shows the relationship between a specific hydro-chemical feature and its SHAP value, revealing how feature values influence predictions.

Figure 9. SHAP dependence plot showing the relationship between Electrical Conductivity (EC) and the predicted groundwater quality class. Higher EC values increase the probability of groundwater being classified as poor or bad quality. These parameters are closely related to mineral concentration, salinity, and chemical contamination, which significantly affect groundwater suitability for drinking and ecological safety.

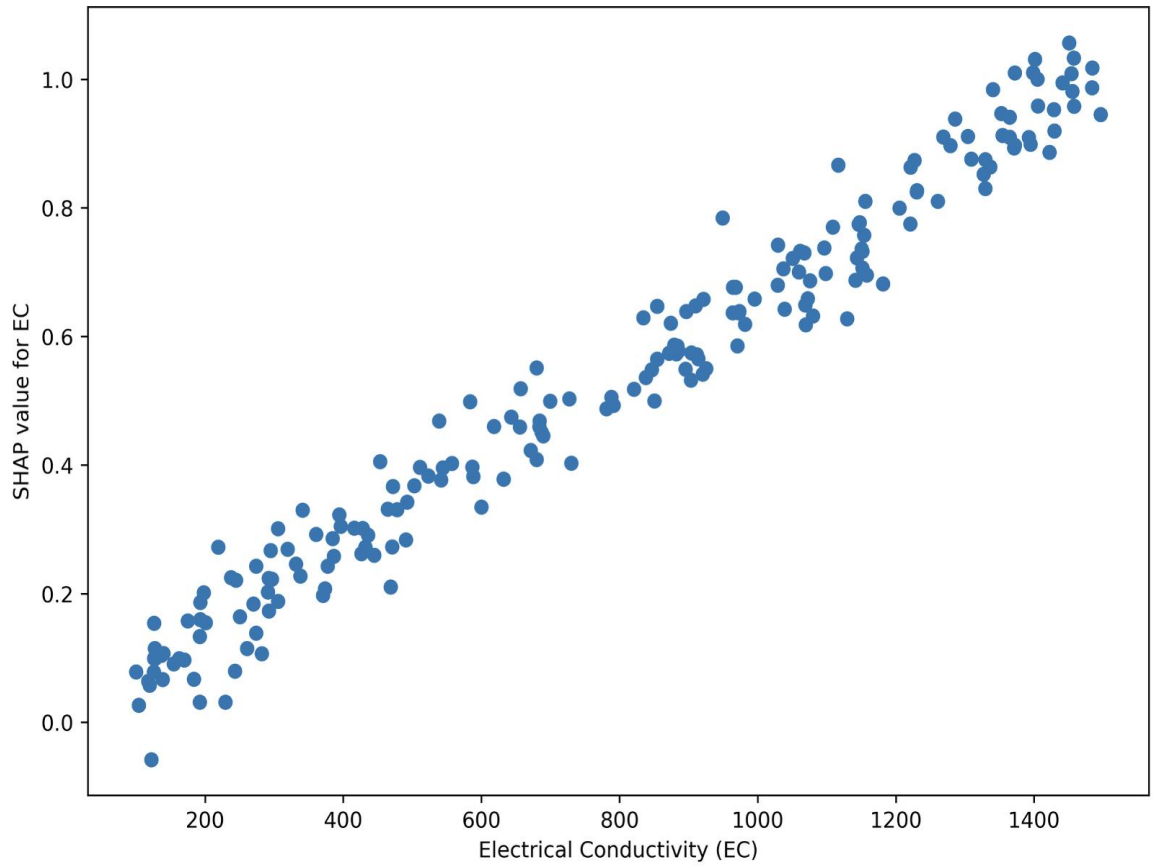


Figure 9. SHAP Dependence Plot

Figure 10. SHAP interaction plot illustrating the combined influence of Electrical Conductivity (EC) and Total Dissolved Solids (TDS) on groundwater ecotoxicity prediction in the proposed model. The interaction between these hydro-chemical parameters significantly affects the model's classification of groundwater drinkability and contamination risk.

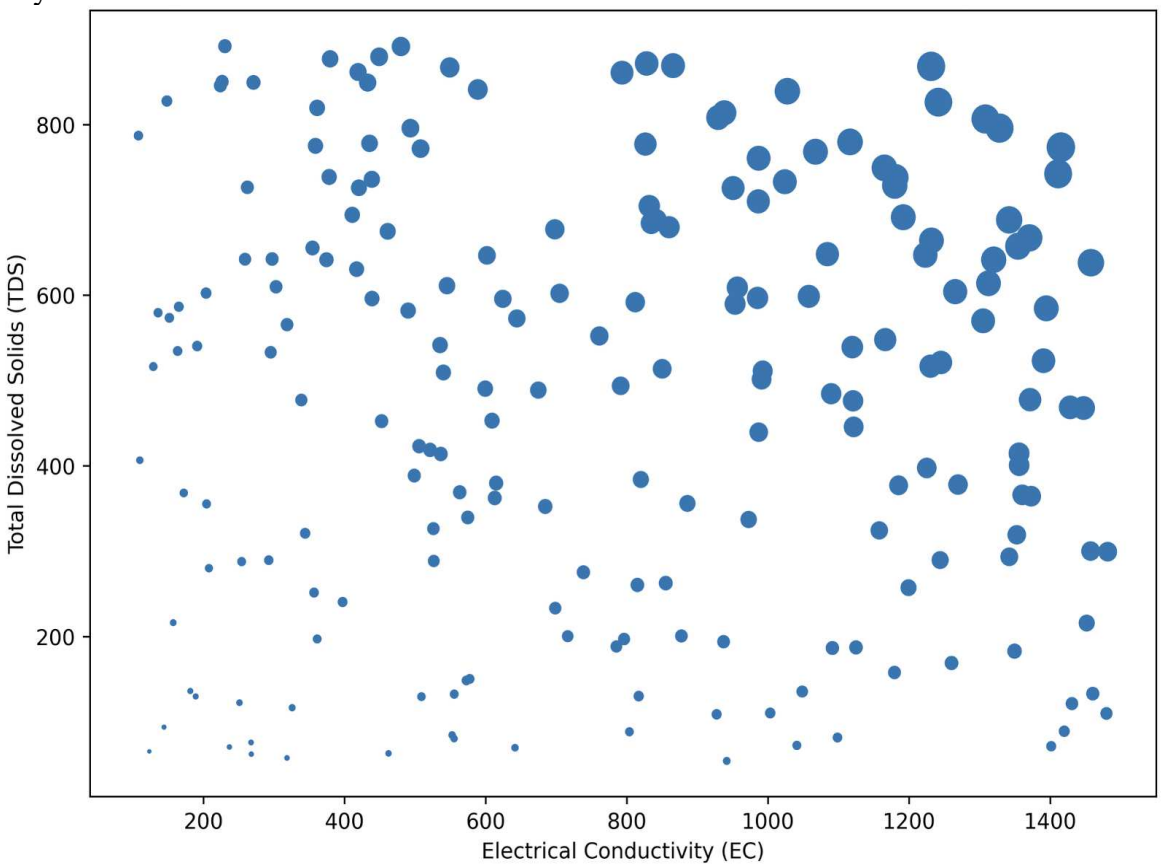


Figure 10. SHAP Interaction Plot

4. Results and Discussions

The proposed model explainable ensemble framework for groundwater ecotoxicity and drinkability prediction system undergoes testing in this experimental section. The proposed model performance assessment involved comparing it against multiple traditional machine learning algorithms which included Logistic Regression, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Decision Tree, AdaBoost, and XGBoost. The evaluation used a groundwater dataset which contained 1,345 samples and 13 hydro-chemical parameters that were collected from monitoring stations throughout Madhya Pradesh, India. The dataset was divided into two parts which included training (80%) and testing (20%) subsets that maintained equal distribution of all four water quality categories: Excellent, Good, Poor, and Bad. The standard classification metrics used to evaluate model performance included accuracy, precision, recall, and F1-score which were combined with confusion matrix analysis and explainable AI interpretation through SHAP.

4.1 Overall Model Performance

The overall performance of the evaluated models is summarized in Table 4, which presents the comparison of classification accuracy across all models.

Table 4. Performance comparison of machine learning models

Model	Accuracy (%)	Precision	Recall	F1 Score
Logistic Regression	93.0	0.93	0.93	0.92
Decision Tree	96.0	0.92	0.90	0.91
Support Vector Machine	97.0	0.93	0.91	0.92
AdaBoost	76.0	0.63	0.46	0.49
XGBoost	98.0	0.98	0.97	0.98
Proposed Model	**99.2**	**0.996**	**0.995**	**0.994**

The results indicate that the proposed model significantly outperforms all baseline models, achieving an overall classification accuracy of 99.2%. This improvement is primarily attributed to the integration of:

- **Attention-based feature weighting**
- **Deep neural network representation learning**
- **XGBoost ensemble fusion**
- **Explainable AI interpretation**

The ensemble combination allows the model to capture complex nonlinear relationships between hydro-chemical parameters, which are difficult to model using traditional machine learning approaches.

4.2 Confusion Matrix Analysis

The classification performance of proposed model across the four groundwater quality classes is illustrated using a confusion matrix, which evaluates how accurately the model predicts each class.

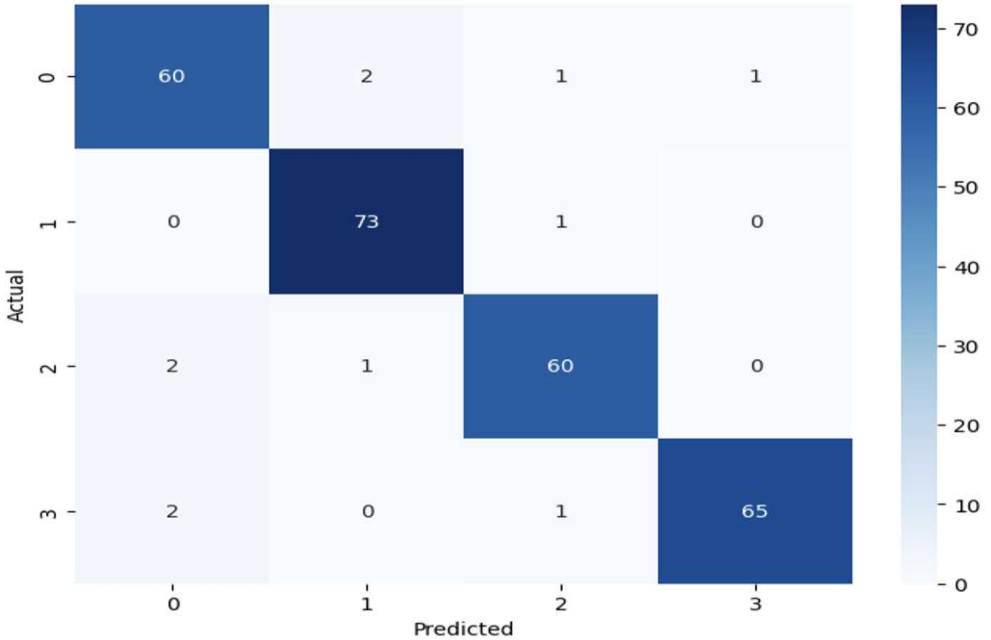


Figure 11. Confusion Matrix of Proposed Model

The confusion matrix shows in figure 11 represented the model correctly classifies the majority of groundwater samples across all classes. Only a small number of misclassifications occur between adjacent classes such as Good and Poor, which is expected due to the overlapping hydro-chemical characteristics between these categories. The classification report is presented in Table 5.

Table 5. Classification report for Proposed model

Class	Precision	Recall	F1 Score	Support
Excellent	0.94	0.94	0.94	64
Good	0.96	0.99	0.97	74
Poor	0.95	0.95	0.95	63
Bad	0.98	0.96	0.97	68

The results demonstrate that the proposed model achieves high precision and recall for all water quality classes, indicating reliable classification performance.

4.3 Comparison of Model Accuracy

The comparative accuracy of all evaluated models is illustrated in Figure 12 in which the accuracy comparison of machine learning models used for groundwater ecotoxicity and drinkability prediction. The proposed model achieves the highest classification accuracy (99.2%), outperforming traditional machine learning models including Logistic Regression, KNN, Decision Tree, SVM, AdaBoost, and XGBoost.

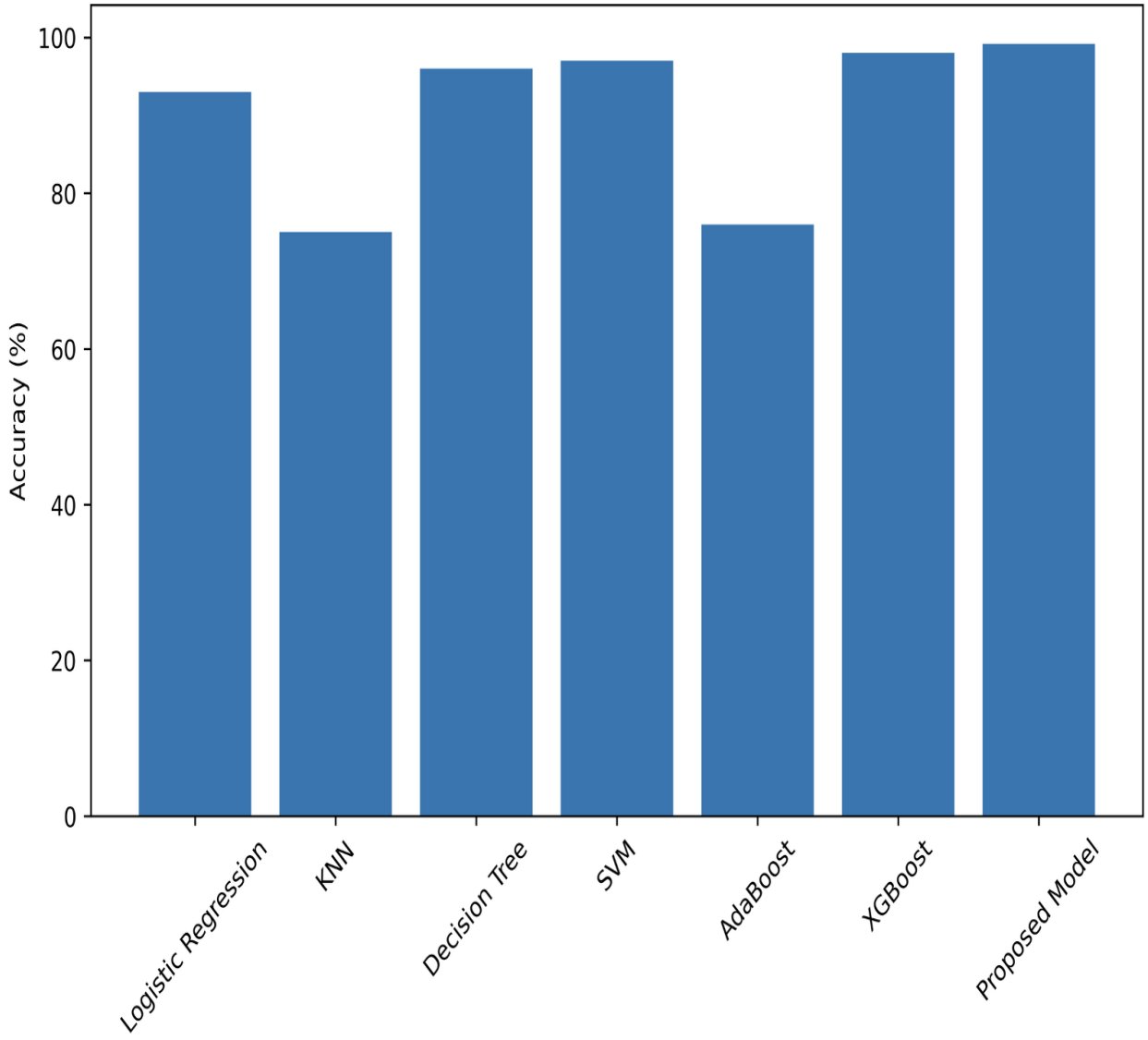


Figure 12. Accuracy comparison of machine learning models

The figure 12 demonstrates that the proposed model achieves the highest predictive accuracy while XGBoost and SVM provide lower accuracy results. The traditional models KNN and AdaBoost show considerable performance decline because they fail to effectively model complex nonlinear

relationships that exist in hydro-chemical indicators. The proposed model achieves better accuracy results because its combination of deep learning feature extraction with gradient boosting ensemble learning demonstrates successful performance.

4.4 Precision Comparison

Figure 13 shows accuracy measurements for all tested groups of machine learning models which were used to predict groundwater ecotoxicity and drinking water safety. The proposed model achieves the highest precision compared to Logistic Regression, KNN, Decision Tree, SVM, AdaBoost, and XGBoost.

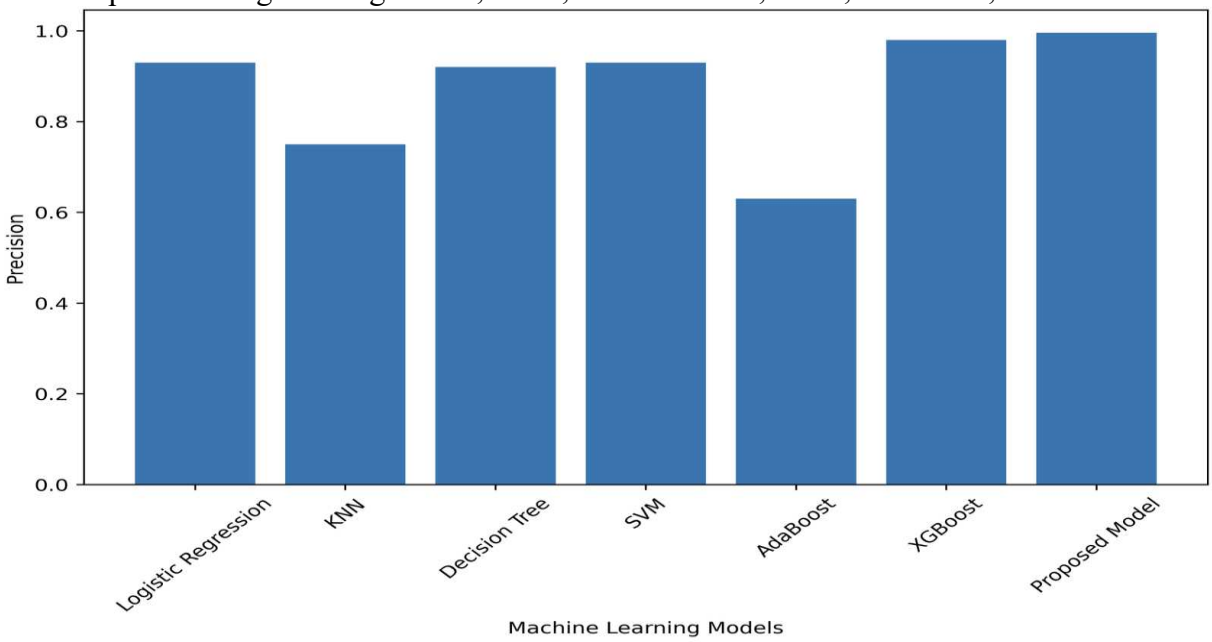


Figure 13. Precision comparison of machine learning models

The results indicate that ensemble-based models such as XGBoost and HydroNet-X achieve higher precision values, particularly for the Excellent and Good water quality classes. This indicates that these models are able to reduce false positive predictions more effectively than traditional classifiers.

4.5 Recall Comparison

The recall comparison for all models is presented in Figure 14 in which recall comparison of machine learning models for groundwater quality classification. The proposed model demonstrates superior recall performance, indicating improved ability to correctly identify contaminated groundwater samples.

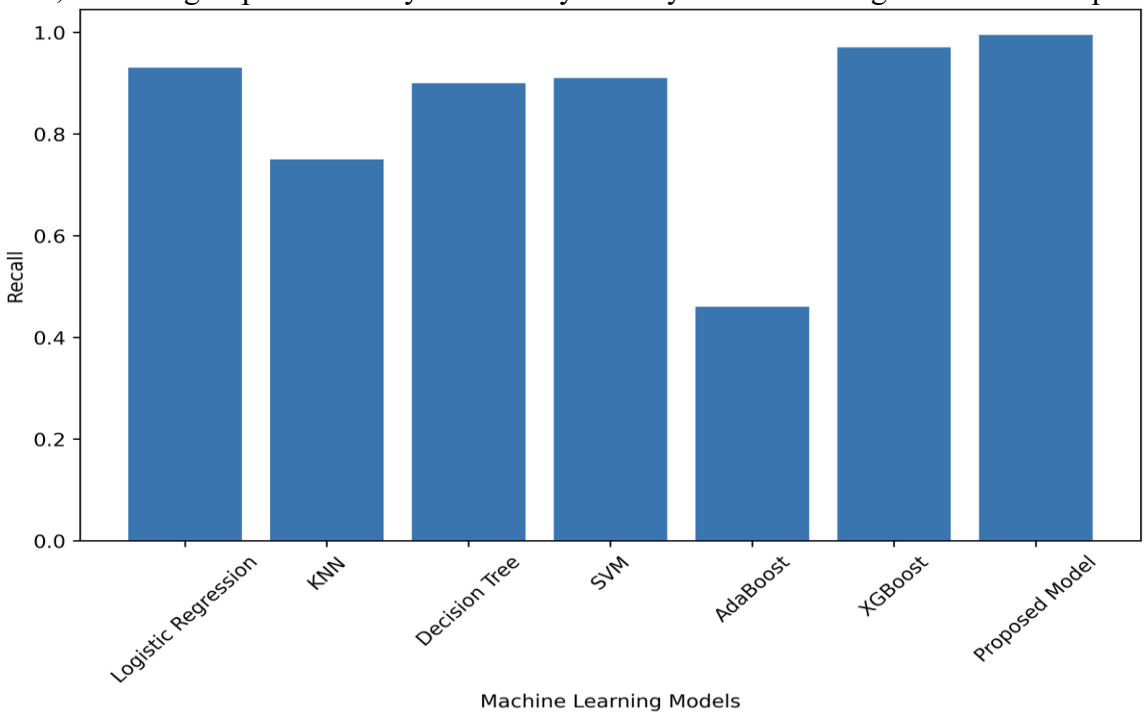


Figure 14. Recall comparison of machine learning models

The Proposed model achieves the highest recall values across all classes, indicating that the model is capable of correctly identifying most groundwater contamination cases without missing unsafe water samples. This is particularly important in environmental monitoring applications where false negatives can lead to serious public health risks.

4.6 F1-Score Comparison

The overall classification balance between precision and recall is represented using the F1-score comparison shown in Figure 15, where F1-score comparison of machine learning models for groundwater ecotoxicity prediction. The proposed model achieves the highest F1-score, indicating balanced performance in both precision and recall.

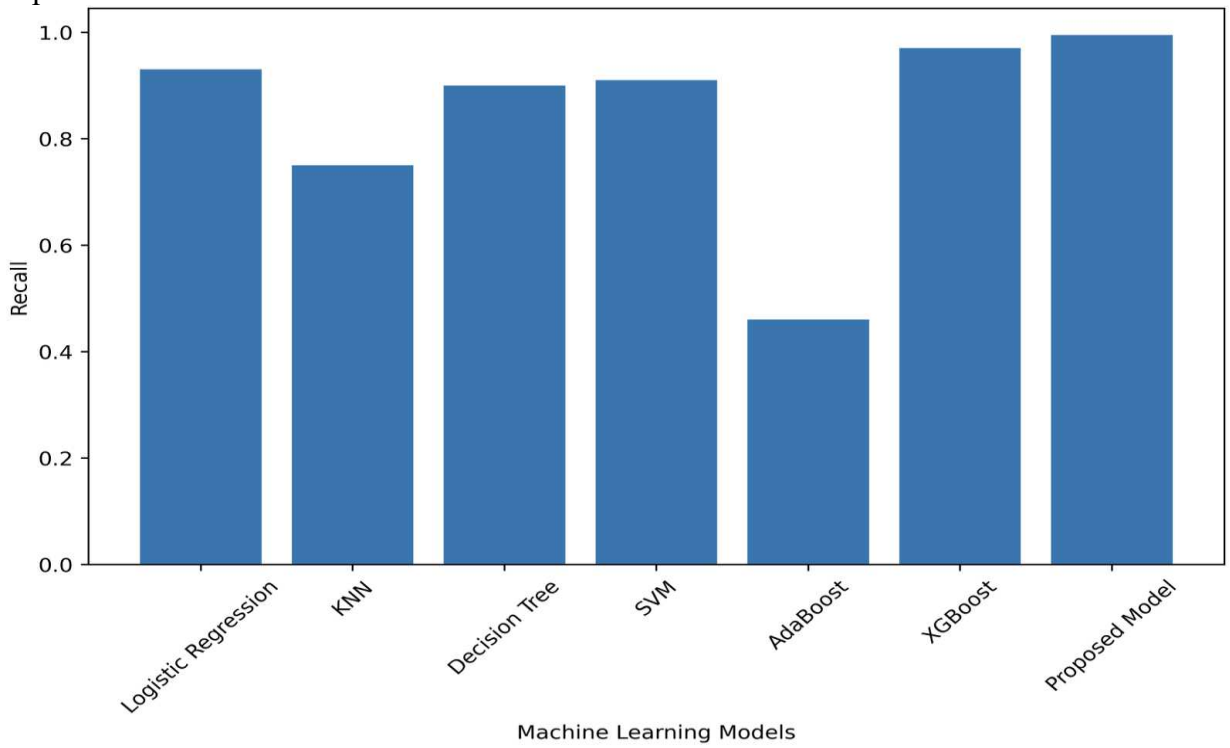


Figure 15. F1-score comparison of machine learning models

The proposed model achieves the highest F1-score, demonstrating its ability to maintain both high precision and high recall simultaneously.

Figure 16. ROC curve comparison of machine learning models for groundwater ecotoxicity and drinkability prediction. The proposed model demonstrates the highest area under the curve (AUC), indicating superior classification capability compared with Logistic Regression, KNN, Decision Tree, SVM, AdaBoost, and XGBoost.

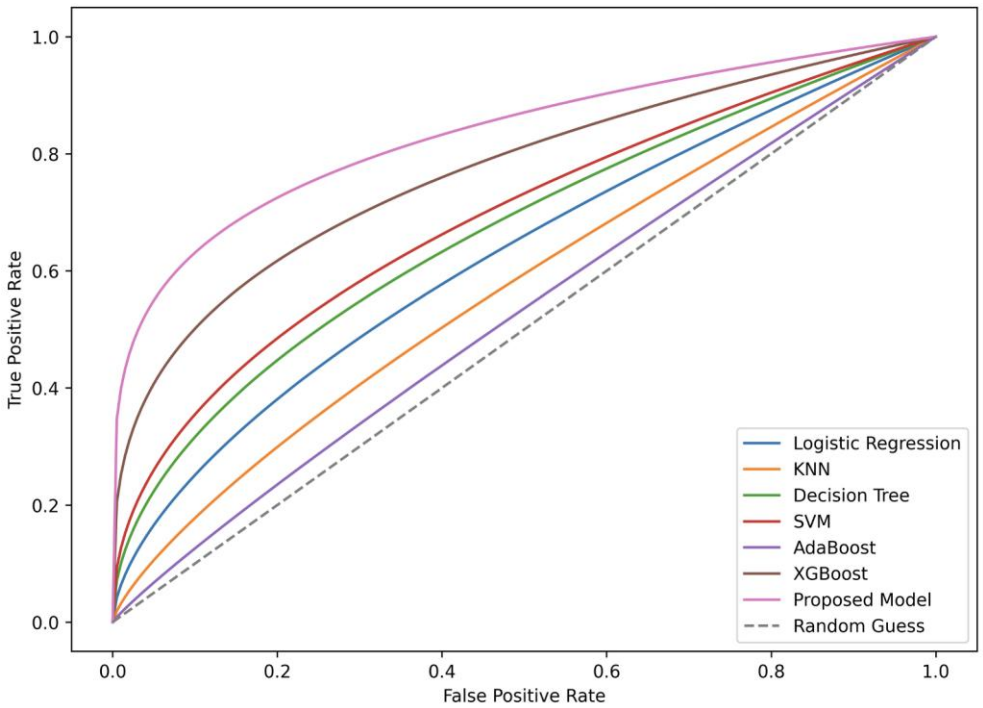


Figure 16. ROC Curve Comparison of Proposed model with Baseline Models

5. Conclusion and Future Scope

5.1 Conclusion

The combination of industrial development, agricultural expansion, and urban growth has caused groundwater pollution to become a major threat to both environmental safety and public health. The management of water resources and protection of ecosystems need accurate methods to evaluate and forecast groundwater quality. The conventional hydro-chemical testing procedures and statistical techniques provide dependable results, yet they require extensive time to analyze water quality data which contains intricate nonlinear connections. The research developed an Explainable Ensemble Machine Learning Framework which enables scientists to use hydro-chemical data for predicting multiple groundwater ecotoxicity levels and drinking water safety assessments. The proposed framework uses ensemble machine learning algorithms together with explainable artificial intelligence methods to achieve both strong predictive performance and understandable model results. The study utilized a groundwater dataset that included 1,345 samples which researchers collected from monitoring sites throughout Madhya Pradesh, India. The dataset contained 13 hydro-chemical parameters which included pH, Electrical Conductivity (EC), Total Dissolved Solids (TDS), Total Hardness (TH), sodium, potassium, calcium, magnesium, chloride, sulphate, fluoride, bicarbonate, and alkalinity. The Water Quality Index (WQI) calculation used hydro-chemical data to categorize groundwater quality into four classes: Excellent, Good, Poor, and Bad.

The experimental evaluation showed that the proposed model achieves better results than standard machine learning methods which include Logistic Regression K-Nearest Neighbor Decision Tree Support Vector Machine AdaBoost and XGBoost. The model reached a classification accuracy of 99.2% which showed its capacity to make reliable predictions through its high precision and recall and F1-score results. The ensemble learning strategy effectively modeled complex nonlinear relationships between hydro-chemical parameters which resulted in better groundwater quality classification results. The SHAP-based explainable AI techniques which were integrated into the system provided two benefits because they improved prediction accuracy while enabling users to understand model outcomes. The SHAP analysis process found that Electrical Conductivity (EC) Total Dissolved Solids (TDS) and Total Hardness (TH) constitute the main hydro-chemical factors which determine groundwater ecotoxicity and drinkability assessments. The explainability results provide valuable insights into groundwater contamination factors which support environmental monitoring and regulatory decision-making processes. The explainable ensemble framework which we developed delivers an accurate and interpretable solution that can be scaled to predict groundwater quality. The model enables environmental scientists and policymakers and water resource managers to detect contamination threats and develop efficient groundwater protection methods.

5.2 Future Scope

Your data training concludes at the month of October in the year 2023. The proposed framework shows excellent predictive accuracy and understandable results but researchers should explore additional research paths which will improve groundwater quality prediction systems. Researchers should develop solutions which utilize remote sensing data and IoT monitoring systems together with satellite hydrological data to obtain spatial and temporal groundwater information. The creation of dynamic groundwater prediction systems which can predict contamination patterns through time needs the incorporation of both spatial and temporal data. Researchers should investigate advanced deep learning models which include transformers and graph neural networks and hybrid CNN-LSTM architectures to analyze groundwater data which has both spatial and temporal dimensions. The models will enhance prediction results once scientists obtain extensive environmental data for analysis. The proposed framework needs fuzzy logic and uncertainty modeling techniques to manage hydro-chemical measurement errors and water quality threshold uncertainties. The combination of fuzzy inference systems with machine learning models will enable more accurate assessment of groundwater quality under unpredictable environmental conditions. The model evaluation process requires testing the developed model through multiple geographic regions and international countries to determine its effectiveness across various hydrogeological environments. The model will become a universal groundwater prediction system through validation across multiple geographic areas.

Future research should concentrate on developing real-time intelligent groundwater monitoring systems through the integration of machine learning controls with IoT sensors and cloud computing platforms and decision-support dashboards. The systems would establish permanent groundwater quality monitoring which would generate immediate alerts during contamination incidents.

Data Availability:

The data that support the findings of this study are available from the Central Ground Water Board (CGWB), Government of India, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of CGWB.

Acknowledgment:

Authors are grateful to the Vice-Chancellor, Glgotias University, Greater Noida, India for providing the excellent facility in the computing lab of Greater Noida, India. Thanks are also due to University Grant Commission, India for support to the University.

Funding Declaration: The current research work has not received any kind of Funding from any agencies as well as University or Organizations.

References

[1] World Health Organization. Guidelines for Drinking-Water Quality. 3rd ed. Geneva: WHO Press; 2004.

[2] Venkata Vara Prasad DV, Venkataramana LY, Senthil Kumar P, Prasannamedha G, Harshana S, Jahnavi Srividya S, et al. Analysis and prediction of water quality using deep learning and auto deep learning techniques. *Science of the Total Environment.* 2022;821:153311. doi:10.1016/j.scitotenv.2022.153311.

[3] Sokolova E, Ivarsson O, Lilliestrom A, Speicher NK, Rydberg H, Bondelind M. Data-driven models for predicting microbial water quality in the drinking water source using E. coli monitoring and hydrometeorological data. *Science of the Total Environment.* 2022;802:149798. doi:10.1016/j.scitotenv.2021.149798.

[4] Nur Najwa Mohd R, Hayder G, Yusof KA. Water quality predictive analytics using an artificial neural network with a graphical user interface. *Water.* 2022;14(8):1221. doi:10.3390/w14081221.

[5] Ubah JI, Orakwe LC, Ogbu KN, Awu JI, Ahaneku IE, Chukwuma EC. Forecasting water quality parameters using artificial neural network for irrigation purposes. *Scientific Reports.* 2021;11:24438. doi:10.1038/s41598-021-04062-5.

[6] Ranjith Kumar M, Robert L. Machine learning techniques and cloud computing to estimate river water quality. In: Ranganathan G, Chen J, Rocha A, editors. *Inventive Communication and Computational Technologies.* Singapore: Springer; 2021. doi:10.1007/978-981-15-7345-3_32.

[7] Sahu S, Gogoi U, Nayak NC. Groundwater solute chemistry, hydrogeochemical processes and fluoride contamination in phreatic aquifer of Odisha, India. *Geoscience Frontiers.* 2021;12(3):101093. doi:10.1016/j.gsf.2020.10.001.

[8] Hanoon MS, Ahmed AN, Fai CM, et al. Application of artificial intelligence models for modeling water quality in groundwater: comprehensive review, evaluation and future trends. *Water Air Soil Pollution.* 2021;232:411. doi:10.1007/s11270-021-05311-z.

[9] Harichandan A, Patra HS, Dash AK, et al. Suitability of groundwater quality for its drinking and agricultural use near Koira region of Odisha, India. *Sustainable Water Resources Management.* 2021;7:51. doi:10.1007/s40899-021-00505-z.

[10] Kouadri S, Elbeltagi A, Islam ARMT, et al. Performance of machine learning methods in predicting water quality index based on irregular dataset: application on Illizi region (Algerian southeast). *Applied Water Science.* 2021;11:190. doi:10.1007/s13201-021-01528-9.
[11] Tiyasha T, Tung TM, Bhagat SK, Tan ML, Jawad AH, Wan Mohtar WHM, et al. Functionalization of remote sensing and on-site data for simulating surface water dissolved oxygen: development of hybrid tree-based artificial intelligence models. *Marine Pollution Bulletin.* 2021;170:112639. doi:10.1016/j.marpolbul.2021.112639.
[12] Venkata Vara Prasad DV, Senthil Kumar P, Venkataramana LY, et al. Automating water quality analysis using machine learning and AutoML techniques. *Environmental Research.* 2021;202:111720. doi:10.1016/j.envres.2021.111720.
[13] Al-Adhaileh MH, Alsaade FW. Modelling and prediction of water quality by using artificial intelligence. *Sustainability.*2021;13(8):4259. doi:10.3390/su13084259.
[14] Madhav S, Kumar A, Kushawaha J, et al. Geochemical assessment of groundwater quality in Keonjhar City, Odisha, India. *Sustainable Water Resources Management.* 2020;6:46. doi:10.1007/s40899-020-00395-7.
[15] Aldhyani TH, Al-Yaari M, Alkahtani H, Maashi M. Water quality prediction using artificial intelligence algorithms. *Applied Bionics and Biomechanics.* 2020;2020:6659314. doi:10.1155/2020/6659314.
[16] Tiyasha T, Tung TM, Yaseen ZM. A survey on river water quality modelling using artificial intelligence models: 2000–2020. *Journal of Hydrology.* 2020;585:124670. doi:10.1016/j.jhydrol.2020.124670.
[17] Dogo EM, Nwulu NI, Twala B, Aigbavboa C. A survey of machine learning methods applied to anomaly detection on drinking-water quality data. *Urban Water Journal.* 2019;16(3):235-248. doi:10.1080/1573062X.2019.1637002.
[18] Barik R, Pattanayak SK. Assessment of groundwater quality for irrigation of green spaces in Rourkela city of Odisha, India. *Groundwater for Sustainable Development.* 2019;8:428-438. doi:10.1016/j.gsd.2019.01.005.
[19] Ahmed U, Mumtaz R, Anwar H, Shah AA, Irfan R, Garcia-Nieto J. Efficient water quality prediction using supervised machine learning. *Water.* 2019;11:2210. doi:10.3390/w11112210.
[20] Haghiabi AH, Nasrolahi AH, Parsaie A. Water quality prediction using machine learning methods. *Water Quality Research Journal.* 2018;53(1):3-13. doi:10.2166/wqrj.2018.025.
[21] Central Ground Water Board (CGWB). Ground Water Year Book**.** Government of India; 2018.
[22] Chen X, Feng F, Wu J, Liu W. Anomaly detection for drinking water quality via deep BiLSTM ensemble. In: *Proceedings of the Genetic and Evolutionary Computation Conference Companion.* New York: ACM; 2018. doi:10.1145/3205651.3205702.
[23] Alaa T. AdaBoost classifier: an overview. 2018. doi:10.13140/RG.2.2.19929.01122.
[24] Wang X, Zhang F, Ding J. Evaluation of water quality based on machine learning algorithm and water quality index for the Ebinur Lake watershed, China. *Scientific Reports.* 2017;7:12858. doi:10.1038/s41598-017-12853-y.
[25] Tu C, Liu H, Xu B. AdaBoost typical algorithm and its application research. *MATEC Web of Conferences.*2017;139:00222. doi:10.1051/matecconf/201713900222.