water

MDPI

Article

Advanced Machine Learning and Water Quality Index (WQI) Assessment: Evaluating Groundwater Quality at the Yopurga Landfill

Hongmei Zheng [1], Shiwei Hou [2], Jing Liu [3], Yanna Xiong [3,*] and Yuxin Wang [3,*]

1 HUAZE EcoEnviron Technology & Engineering Institute (Beijing) Co., Ltd., Beijing 101400, China; zhm0218@126.com
2 School of Civil Engineering, Shenyang Jianzhu University, Shenyang 110168, China; hsw1375@126.com
3 Technical Centre for Soil, Agriculture and Rural Ecology and Environment, Ministry of Ecology and Environment, Beijing 100012, China; 15690861161@163.com
* Correspondence: xiongyanna@tcare-mee.cn (Y.X.); 18935039901@163.com (Y.W.); Tel.: +86-186-0119-1331 (Y.X.); +86-189-3503-9901 (Y.W.)

Abstract: As industrial development and population growth continue, water pollution has become increasingly severe, particularly in rapidly industrializing regions like the area surrounding the Yopurga landfill. Ensuring water resource safety and environmental protection necessitates effective water quality monitoring and assessment. This paper explores the application of advanced machine learning technologies and the Water Quality Index (WQI) model as a comprehensive method for accurately assessing groundwater quality near the Yopurga landfill. The methodology involves selecting water quality indicators based on available data and the hydrochemical characteristics of the study area, comparing the performance of Decision Trees, Random Forest, and Xgboost algorithms in predicting water quality, and identifying the optimal algorithm to determine indicator weights. Indicators are scored using appropriate sub-index (SI) functions, and six different aggregation functions are compared to find the most suitable one. The study reveals that the Xgboost model surpasses Decision Trees and Random Forest models in water quality prediction. The top three indicator weights identified are pH, Manganese (Mn), and Nickel (Ni). The SWM model, with a 0% overestimation eclipsing rate and a 34% underestimation eclipsing rate, is chosen as the most appropriate WQI model for evaluating groundwater quality at the Yopurga landfill. According to the WQI results from the SWM aggregation function, the overall water quality in the area ranges from moderately polluted to slightly polluted. These assessment results provide a scientific basis for regional water environment protection.

Keywords: WQI; machine learning; groundwater; landfill; Yopurga

check for updates

Citation: Zheng, H.; Hou, S.; Liu, J.; Xiong, Y.; Wang, Y. Advanced Machine Learning and Water Quality Index (WQI) Assessment: Evaluating Groundwater Quality at the Yopurga Landfill. *Water* **2024**, *16*, 1666. https://doi.org/10.3390/w16121666

Academic Editor: Nicolò Colombani

Received: 8 April 2024
Revised: 1 June 2024
Accepted: 5 June 2024
Published: 12 June 2024

1. Introduction

As industrialization and urbanization rapidly progress, environmental pollution has emerged as a global challenge, with water pollution attracting particularly serious concern. Water, being an indispensable resource for maintaining ecological balance and supporting human life, demands the utmost protection of its quality [1]. In the Yopurga region, groundwater pollution resulting from the Yopurga landfill and surrounding activities underscores the urgent need for stringent assessment and management of water resource quality in the area. Groundwater serves not only as a critical freshwater resource but also as a primary source for agricultural irrigation. Thus, assessing the groundwater quality near the Yopurga landfill is vital for ensuring its safe use in agricultural fields and minimizing potential risks to the environment and public health [2].

While studies have explored various water quality indices (WQI) and their predictive capabilities, there is a dearth of research integrating sophisticated ML algorithms to enhance the objectivity and precision of these assessments. Specifically, the integration of ML with WQI in the context of landfill impact has not been extensively explored, presenting an opportunity for this study to contribute original insights. Various techniques are employed for water quality assessment, including Single Factor Evaluation (SF) [3], Fuzzy Comprehensive Evaluation (FCE) [4], Principal Component Analysis (PCA) [5], and the WQI [6]. Each method has its limitations when assessing water quality [7,8]. Among these, the WQI model is widely applied and studied due to its comprehensiveness and intuitiveness. The WQI model typically involves four steps: selection of water quality indicators, scoring of indicators, determination of indicator weights, and aggregation of results. Its simplicity and flexibility have earned widespread recognition [9–11].

In the traditional application framework of the WQI model, the assignment of indicator weights often suffers from subjective judgment, significantly impacting the precision of assessment outcomes. Although the Delphi method, a mainstream strategy for determining weights, is widely used, its reliance on expert knowledge and judgment can introduce subjective bias, affecting the objectivity and reliability of the assessment. Machine learning algorithms excel at handling large datasets and their complex, high-dimensional features, offering a data-driven, objective mechanism for determining weights. Research by Shah et al. and Taromideh et al. has demonstrated that the application of machine learning can significantly enhance the objectivity of the weight determination process [12,13]. Further, Uddin et al. confirmed the efficacy of machine learning techniques in accurately identifying key water quality indicators from complex datasets [14], further validating the practicality and value of machine learning in water quality assessment. Additionally, aggregation functions, as a critical component of the WQI model, have a decisive impact on the final water quality assessment outcomes [15–17]. Based on their different focuses, aggregation functions can be categorized into weighted and unweighted types [18].

Previous studies in China have identified various water quality parameters essential for calculating the WQI. For instance, research in Shenzhen has highlighted parameters such as pH, Fe, and Mn as critical indicators of water quality [19]. The application of WQI as a tool for water quality assessment has been explored in several studies in China. For example, in a study conducted in Xianyang City, groundwater quality evolution was assessed and the WQI was predicted [20]. Given the complexity of water quality data and the need for objective, data-driven assessments, the use of advanced machine learning technologies becomes justified. Key findings from recent studies have informed the methodology and underscored the potential of ML in enhancing environmental assessments [20–22]. These studies have demonstrated the efficacy of ML in predicting complex environmental patterns and managing large datasets, which is crucial for application in evaluating groundwater quality.

The core of this study is the application of machine learning algorithms to predict groundwater quality, selecting algorithms with superior performance and integrating them with the WQI model for a comprehensive evaluation of groundwater quality at the Yopurga landfill. The main contributions of this research include utilizing machine learning algorithms to determine indicator weights, significantly reducing subjective interference in evaluation outcomes, and comparing the effectiveness of different weighted aggregation functions, identifying the best aggregation function suitable for the groundwater of the Yopurga landfill.

2. Study Area

The Yopurga landfill is situated in Yopurga County within the Kashgar region of Xinjiang, delimited between longitudes 76°25′ to 77°25′ E and latitudes 38°46′ to 39°22′ N, As shown in Figure 1. This area lies at the eastern end of the Kashgar Plain and the western edge of the Taklamakan Desert, near the lower reaches of the Gez River. The topography of this region is characterized by slightly higher elevations around the perimeter with a flatter central area, greater north–south relief, minor slope gradients, and more significant east–west slope gradients. The landfill covers an area of 40,000 square meters, including a 4000 square meter management area.

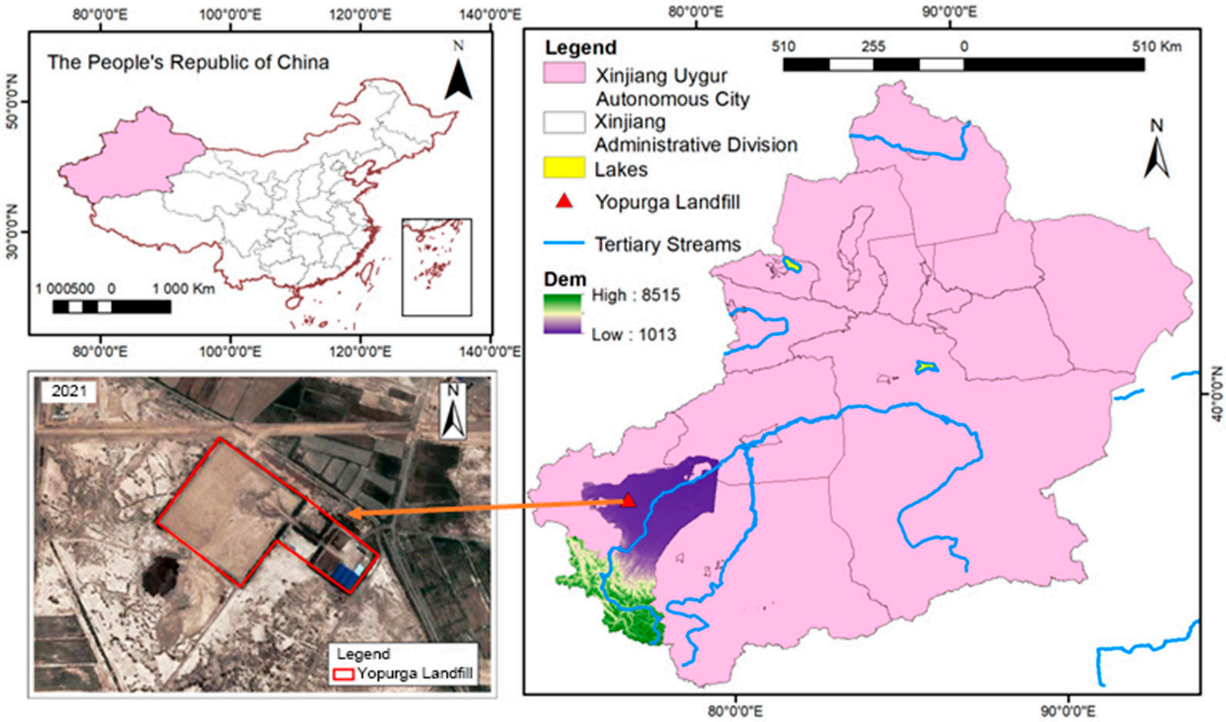

Figure 1. Map of the study area.

This region experiences a warm temperate continental arid climate, with an average annual temperature of 11.8 °C and extreme temperature fluctuations ranging from −23.4 °C to 41.5 °C. It receives an average of 2780.3 h of sunshine per year, with an average frost-free period of approximately 232 days. Precipitation is scarce, totaling only 66.4 mm annually. The area is frequently affected by strong winds and dust storms. Its surface water primarily derives from the Gez River, the Yarkand River, and the Kizil River, which show significant seasonal variations between their dry and flood periods, with the majority of the annual water volume concentrated from June to August.

Groundwater resources are relatively abundant in this region, mainly replenished through vertical infiltration from channels, reservoirs, and irrigation, with groundwater levels typically ranging from 2 to 4 m deep and a salinity of about 2 to 3 g per liter. The soil organic matter content is low, generally not exceeding 1%. Additionally, Yopurga County is endowed with rich mineral resources, including petroleum and natural gas [23].

3. Materials and Methods

3.1. Data Collection

The monitoring site for this study was located at the Yopurga landfill. The depth of groundwater extraction was determined based on the actual occurrence of groundwater in the area and the hydrogeological characteristics. A total drilling depth of 553 m was achieved, with the deepest individual monitoring well reaching 67 m. The sampling depths were chosen to cover the groundwater aquifers likely to be contaminated, ensuring that the sampling sites representatively reflect the landfill's impact on the groundwater environment.

In accordance with the "Guidelines for Groundwater Environmental Condition Survey and Assessment" (2019) and considering the specific environmental conditions of the survey area, multiple factors were taken into account, including the direction of groundwater flow, distribution of pollution sources, geological structure, and hydrogeological conditions. A total of 28 monitoring wells were established, comprising 23 newly constructed wells and 5 existing wells. All groundwater samples were tested in the laboratory of the Nonferrous Geology Exploration Bureau Testing Center in the Xinjiang Uygur Autonomous Region. This study selected 29 sets of data, encompassing 10 indicators: pH, Ammonium

Nitrogen (NH_3-N), Mn, Ni, Boron (B), Lead (Pb), Zinc (Zn), Fluoride (F^-), Chemical Oxygen Demand (COD), and Iron (Fe). NH_3-N was determined by Nessler's Reagent Spectrophotometry (HJ535-2009), COD by GB11892-1989, pH by the Standard Examination Methods for Drinking Water (GB/T5750.4-2006), and Fe, Mn, and B by Inductively Coupled Plasma Optical Emission Spectrometry (HJ776-2015). Zn, Ni, and Pb were measured using Inductively Coupled Plasma Mass Spectrometry (HJ700-2014), and the inorganic anion F^- was determined by Ion Chromatography (HJ84-2016). As the groundwater around the Yopurga landfill and its vicinity currently has no drinking function and is solely used for agricultural irrigation, the standards applied were the Class IV water limits according to the "Groundwater Quality Standards" (GB/T 14848-2017) (Table 1).

Table 1. Indicator thresholds under Class IV water quality criteria.

Parameter	Unit	Standard for Groundwater Quality (GB/T 14848-93)	Standard for Groundwater Quality (GB/T 14848-2017)
		IV	IV
pH	-	5.5~6.5/8.5~9	$5.5 \leq pH < 6.5/8.5 < pH \leq 9.0$
Ammonium Nitrogen	mg/L	0.5	1.5
Manganese	mg/L	1	1.5
Nickel	mg/L	0.1	0.1
Boron	mg/L	-	2
Lead	mg/L	0.1	0.1
Zinc	mg/L	5	5
Fluoride	mg/L	2	2
Chemical Oxygen Demand	mg/L	10	-
Iron	mg/L	1.5	2

3.2. Machine Learning Algorithms

Machine learning algorithms are extensively utilized for classification, regression, and clustering issues. Among these, classification algorithms have been widely applied in the field of water quality prediction, achieving notable outcomes. By analyzing historical data on water quality parameters, classification algorithms can effectively predict water quality categories and assess the health status of water bodies. This study employed three machine learning algorithms—Decision Trees, Random Forest, and Xgboost—to predict water quality status.

The Decision Tree model [24] simulates the decision-making process by utilizing a series of judgment logics based on water quality feature attributes, refining the data progressively until reaching a node that allows for clear classification. This model's significant advantages include its intuitiveness and interpretability, offering a clear visualization of decision paths. However, it may suffer from overfitting to training data, reducing its generalization capability.

The Random Forest model [25,26], a representative of ensemble learning, enhances prediction accuracy by constructing multiple decision trees and aggregating their predictions. This algorithm randomly selects data sets and features during the training of each tree, effectively reducing the risk of overfitting and enhancing the model's stability and accuracy. Random Forest can handle complex, high-dimensional data and assess the importance of different features.

Xgboost [27], an advanced version of gradient boosting decision trees, is an efficient and flexible ensemble learning algorithm. It introduces regularization terms to reduce model complexity and employs sophisticated algorithms for finding split points, thus improving training efficiency and avoiding overfitting. Xgboost, known for its excellent performance, can process various types of data and is widely used in a broad range of prediction and classification issues.

This study compares the performance of Decision Trees, Random Forest, and Xgboost in water quality prediction to select the best model for calculating the WQI.

3.2.1. Input Data

In this study, a total of 29 sets of water quality sample data were compiled and analyzed, each including ten different water quality indicators. The evaluation of water quality status was benchmarked against the national Class IV groundwater quality standards. Comparing the standard limit values and actual measurements of each water quality indicator, the condition of water quality was categorized into two groups: samples meeting all the Class IV water quality standards were labeled as 0, and those with any indicator exceeding the Class IV standards were labeled as 1. Based on this classification criterion, a binary classification machine learning model was constructed to accurately predict the quality status of water bodies.

3.2.2. Data Preprocessing

To train an efficient and accurate machine learning model, this study conducted data preprocessing, including normalization and splitting. The data were transformed into a standard normal distribution via the Z-score normalization method, as shown in Equation (1) [28]. After data normalization, the dataset was randomly divided into a training set and a test set, with proportions of 70% and 30%, respectively, to ensure the model's generalizability and prevent overfitting.

$$x_2 = \frac{(x_1 - \mu)}{\sigma} \tag{1}$$

where x_2 is the normalized value; x_1 is the original feature value; μ is the mean of the feature; σ is the standard deviation of the feature.

3.2.3. Model Training

This study employed three different algorithms—Decision Trees, Random Forest, and Xgboost—to analyze the preprocessed data. To further optimize model performance, five-fold cross-validation was introduced. This technique accounts for the potential high overlap between test and training sets, enhancing the representativeness and reliability of the evaluation results.

3.2.4. Hyperparameter Optimization

In the field of machine learning, adjusting the model's hyperparameters is a key step in enhancing model performance. Research indicates that grid search and random search are two mainstream methods widely applied for hyperparameter optimization of models [29,30]. Comparing these methods, grid search provides a more comprehensive means of hyperparameter adjustment through a systematic exploration of the parameter space and is thus generally considered a more effective optimization strategy. Therefore, this study chose a grid search technique for the hyperparameter tuning of the model [30].

3.2.5. Model Performance Comparison

In the process of evaluating the performance of machine learning models, it is crucial to utilize multiple metrics to comprehensively understand the model's predictive capabilities. This study employed a range of metrics, including Accuracy, Precision, Recall, F1 Score, Root Mean Square Error (RMSE), and the Area Under the Curve (AUC), for the assessment of model performance.

Accuracy is defined as the ratio of the number of correctly predicted samples to the total number of samples. It serves as a fundamental indicator for evaluating the overall performance of classification models, as shown in Equation (2) [31].

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \tag{2}$$

Precision reflects the proportion of samples that are accurately identified as positive out of all samples predicted as positive by the model. It is a key metric for measuring the accuracy of a classifier's predictions, as shown in Equation (3) [32].

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3}$$

Recall describes the proportion of actual positive samples that are correctly predicted as positive by the model, marking an important metric for evaluating the completeness of a model's predictions, as shown in Equation (4) [33].

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4}$$

F1 Score, as the harmonic mean of Precision and Recall, provides a composite reflection of a classifier's overall performance, ideally representing the average of Precision and Recall, as shown in Equation (5) [34].

$$\text{F1 Score} = 2 \times \frac{\frac{\text{TP}}{\text{TP+FP}} \times \frac{\text{TP}}{\text{TP+FN}}}{\frac{\text{TP}}{\text{TP+FP}} + \frac{\text{TP}}{\text{TP+FN}}} \tag{5}$$

where TP denotes True Positive, TN denotes True Negative, FP denotes False Positive and FN denotes False Negative.

RMSE is a commonly used metric to measure the variance between model predictions and actual values, represented as the square root of the average of the squares of the differences between predicted and actual values, as shown in Equation (6) [35].

$$\text{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y})} \tag{6}$$

where y_i is the actual value for the ith sample, $\hat{y}$ is the predicted value for the ith sample, and m is the total number of samples.

AUC, a key metric for evaluating the overall performance of binary classification models, represents the proportion of positive-negative sample pairs correctly discriminated by the model, as shown in Equation (7) [36]. An AUC value closer to 1 indicates superior model performance. Since AUC measures the entire area under the ROC curve, it remains stable even if the model's classification threshold changes, showcasing high reliability.

$$\text{AUC} = \frac{1}{m}\sum_{i=1}^{m}\max(0, r_i - r_{i-1}) \tag{7}$$

where r_i denotes the cumulative recall of the ith sample after sorting all samples in descending order of predictive probability. m denotes the total number of samples.

Each metric reflects different aspects of model performance: Accuracy provides an intuitive perception of the model's overall efficacy; Precision and Recall focus on the quality of predictions for specific categories; F1 Score balances the importance of Precision and Recall; RMSE measures the accuracy of predictions; and the AUC value collectively reflects the efficiency of a classification model in distinguishing between positive and negative samples. Through the comprehensive evaluation of these metrics, one can fully understand and optimize the predictive performance of the model.

3.3. Water Quality Index Model

3.3.1. Indicator Selection

In the process of constructing the WQI model, this study considered a comprehensive set of indicators commonly used in water quality assessment and the availability of existing data. Given the specific anticipated uses of the water bodies near the Yopurga landfill, the following ten indicators were selected for analysis: This included several general chemical indicators such as pH, Fe, Mn, Zn, NH_3-N, and COD, as well as toxicological indicators including F-, B, Ni, and Pb.

3.3.2. Sub-Index Functions

To provide a standardized framework for the importance measure in the WQI calculation, each indicator's measured value was converted into a dimensionless score ranging from 0 to 100 through SI functions [16,17], where 0 represents the lowest level, and 100 represents the optimal level. Following Uddin's method, this study calculated the SI for each water quality indicator [11]. The standard values for all water quality indicators were based on the national groundwater quality standards. The specific formulas for each indicator are shown in Table 2. The *SI* calculations are given by

$$SI = (SI_u - SI_l) - \frac{(SI_u \times WQ_{im})}{(STD_u - STD_l)} \tag{8}$$

$$SI = \frac{(WQ_{im} - STD_l)}{(STD_u - STD_l)} \times SI_u \tag{9}$$

$$SI = (SI_u - SI_l) - \frac{(WQ_{im} - STD_l)}{(STD_u - STD_l)} \tag{10}$$

where SI_l and SI_u are the lower limit 0 and upper limit 100 of the SI value, respectively; STD_l is the lower threshold, and WQ_{im} is the actual measured value of each indicator.

Table 2. Calculation formula for each indicator sub-indicator function.

Indicators	Conditions	Sub-Index Functions
Ammonium Nitrogen Manganese Nickel Boron Lead Zinc Fluoride Chemical Oxygen Demand Iron	-	Equation (8)
pH	(i) If pH $\geq$ 5.5 and pH < 6.5 (ii) If pH > 8.5 and pH $\leq$ 9.0 (iii) If pH $\geq$ 6.5 and pH $\leq$ 8.5	Equation (9) Equation (10) 100

3.3.3. Calculation of Weights

In the assessment of aquatic environments, the impact weights of various water quality indicators significantly differ. This study employs the Xgboost algorithm as a data-driven method to determine the weights of each indicator. Unlike traditional expert judgment, this method identifies the impact weights of each indicator by analyzing a large volume of water quality monitoring data, thereby ranking their importance. This process effectively reduces the potential biases that subjective judgment might introduce in weight allocation, objectively reflecting the relative importance of each indicator in the overall water quality assessment [37,38].

3.3.4. Aggregation Functions

In the WQI model, the final step to form a single composite index value involves synthesizing the SI functions and weights using a weighted aggregation function. To find the method most suitable for the water quality assessment of the Yopurga landfill, this study compared six common weighted aggregation functions: (i) National Sanitation Foundation (NSF) index [39], (ii) Scottish Research Development Department (SRDD) index [40], (iii) West Java (WJ) index [41], (iv) weighted quadratic mean (WQM) [16], (v) Log-weighted Quadratic Mean (LQM) [42], and (vi) Sinusoidal Weighted Mean (SWM). Detailed formulas for each aggregation function can be found in Table 3 within the document [42].

Table 3. Formulas for each aggregation function.

Aggregate Function	Calculation Formula
NSF index (Weighted Arithmetic Mean (WAM))	$NSF = \sum_{i=1}^{n} s_i w_i$
SRDD index (Modified Additive Function)	$SRDD = \frac{1}{100}\left(\sum_{i=1}^{n} s_i w_i\right)^2$
West Java WQI	$WJ = \prod_{i=1}^{n} {s_i}^{w_i}$
Weighted Quadratic Mean (WQM)	$WQM = \sqrt{\sum_{i=1}^{n} w_i {s_i}^2}$
Log-weighted Quadratic Mean (LQM)	$LQM = 50\sqrt{\sum_{i=1}^{n} w_i [lg(s_i+1)]^2}$
Sinusoidal Weighted Mean (SWM)	$SWM = 100\sum_{i=1}^{n} w_i \sin(s_i)$

3.4. *Evaluation of Water Quality Index Model Scores*

Through the calculations of the aggregation functions, each sampling point yields a WQI score ranging from 0 to 100. This study employed various classification schemes to interpret the WQI scores and assess the condition of the water quality. These classification schemes categorize water quality from excellent to poor, with higher scores indicating better water quality conditions and 0 representing the worst condition. The classification schemes corresponding to each aggregation function are presented in Table 4, providing clear criteria for judging different water quality states.

Table 4. Classification scheme corresponding to each aggregation function.

Aggregate Function	Classification Scheme
NSF index	Five categories ① Excellent (90–100) ② Good (70–89) ③ Medium (50–69) ④ Bad (25–49) ⑤ Very bad (0–24)
SRDD index	Seven categories ① Clean (90–100) ② Good (80–89) ③ Good with treatment (70–79) ④ Tolerable (40–69) ⑤ Polluted (30–39) ⑥ Several polluted (20–29) ⑦ Piggery waste (0–19)
West Java WQI	Five categories ① Excellent (90–100) ② good (90–75) ③ Fair (75–50) ④ Marginal (50–25) ⑤ Poor (25–5)

Table 4. *Cont.*

Aggregate Function	Classification Scheme
Weighted Quadratic Mean	Four categories ① Good (80–100) ② Fair 50–79) ③ Marginal (30–49) ④ Poor (0–29)
Log-weighted Quadratic Mean	Five categories ① Clean (90–100) ② Slightly polluted (75–90) ③ Moderately polluted (50–75) ④ Heavily polluted (25–50) ⑤ Seriously polluted (0–25)
Sinusoidal Weighted Mean	Five categories ① Clean (90–100) ② Slightly polluted (75–90) ③ Moderately polluted (50–75) ④ Heavily polluted (25–50) ⑤ Seriously polluted (0–25)

4. Results and Discussion

4.1. Model Validation

Hyperparameter optimization for three machine learning algorithms, including Decision Trees, Random Forest, and Xgboost, was conducted using grid search, yielding the following optimized parameters:

(Decision Tree): min_samples_leaf: 2; min_samples_split: 2.

(Random Forest): min_samples_leaf: 1, min_samples_split: 5, n_estimators: 50.

(Xgboost): colsample_bytree: 0.8, learning_rate: 0.01, max_depth: 3, n_estimators: 100, subsample: 0.9.

Under the best hyperparameter settings, Decision Trees, Random Forest, and Xgboost models were validated using five-fold cross-validation. The models' performances were compared across six evaluation metrics: Accuracy, Precision, Recall, F1, AUC, and RMSE, as illustrated in Figure 2. To mitigate the influence of overfitting and other phenomena on the validation results, the average of five runs was taken as the final outcome for each metric. For Accuracy, Precision, Recall, and F1, values closer to 1 indicate better performance, while for RMSE, values closer to 0 are preferable. The use of multiple validation metrics ensures the reliability of the Xgboost algorithm's ranking of water quality indicator importance, significantly reducing the subjectivity and uncertainty of water quality evaluation results. The Xgboost algorithm's low RMSE and high Accuracy, Precision, Recall, and F1 scores demonstrate its superiority over other algorithms. Due to its excellent performance and adaptability to the dataset, the Xgboost model is considered the preferred tool for determining key indicator weights in water quality assessment, as also supported by research from Dao Nguyen Khoi et al. and Md Galal Uddin et al. [16,43].

After five iterations of the three algorithm models, it is evident from Figure 3 that although the Xgboost model does not always achieve the best prediction accuracy, it demonstrates stable and superior performance in the latter three iterations for metrics other than RMSE. Random Forest shows considerable fluctuation and the weakest performance, slightly inferior to Decision Trees. Overall, in the construction of this water quality evaluation model, the ranking is Xgboost > Decision Trees > Random Forest. Further analysis will also be conducted on the Xgboost model for the relative importance calculation.

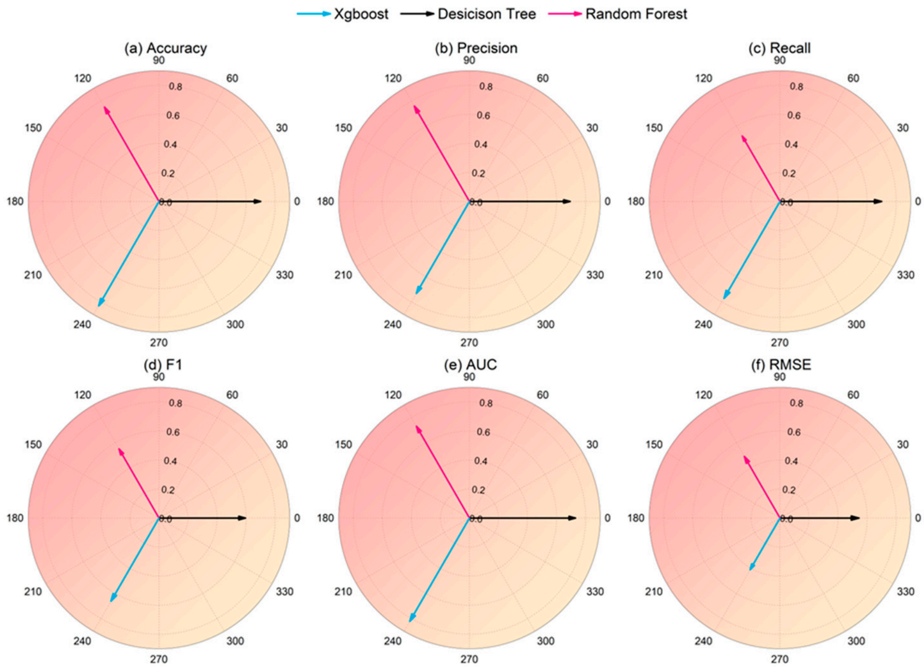

Figure 2. Comparison of six evaluation metrics for three algorithmic models.

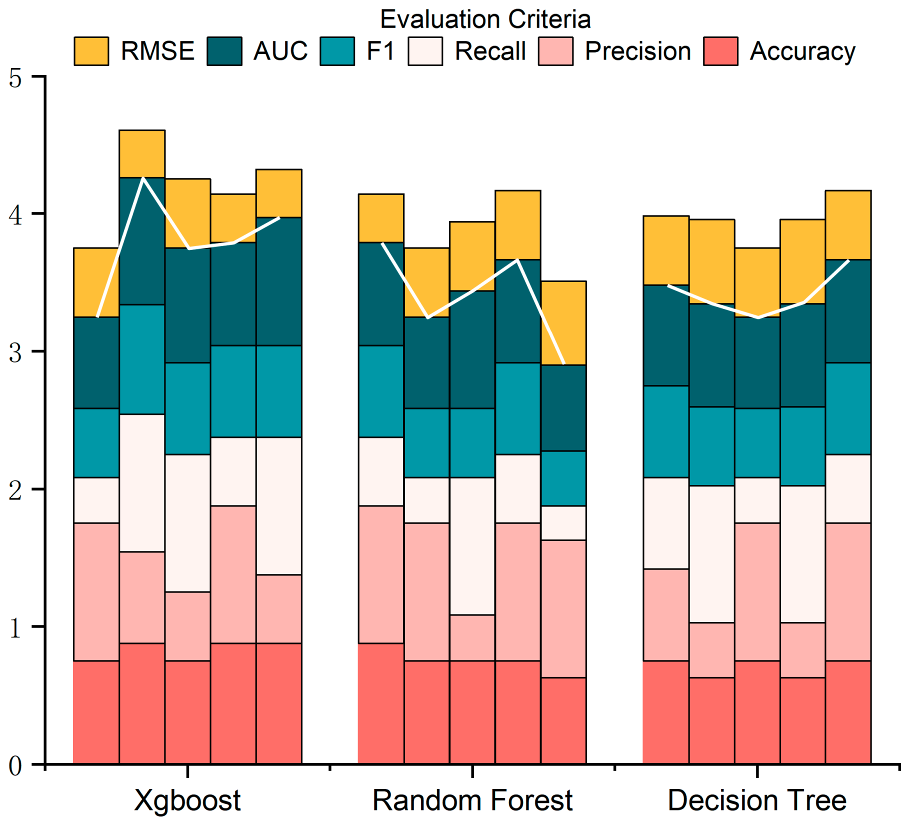

Figure 3. Comparison of the results of five iterations of the three algorithmic models. The white lines shows the combined trend for each evaluation indicator except RMSE.

4.2. Machine Learning Algorithms

4.2.1. The Relationship between Actual Values and Predicted Values

As shown in Figure 4, the Random Forest model has a high true positive rate but one false negative. Xgboost model performs perfectly within this dataset, with no misclassifications. The Decision Tree model also performs well, correctly identifying most positive and negative cases without any misclassifications. From the confusion matrices, it can be observed that the Xgboost model provides the best performance for this dataset, with no false positives or false negatives. Both the Random Forest and Decision Tree models perform well but have minor misclassifications compared to the Xgboost model. However, due to the limited data used for training and prediction, there may be a risk of overfitting, leading to some degree of uncertainty.

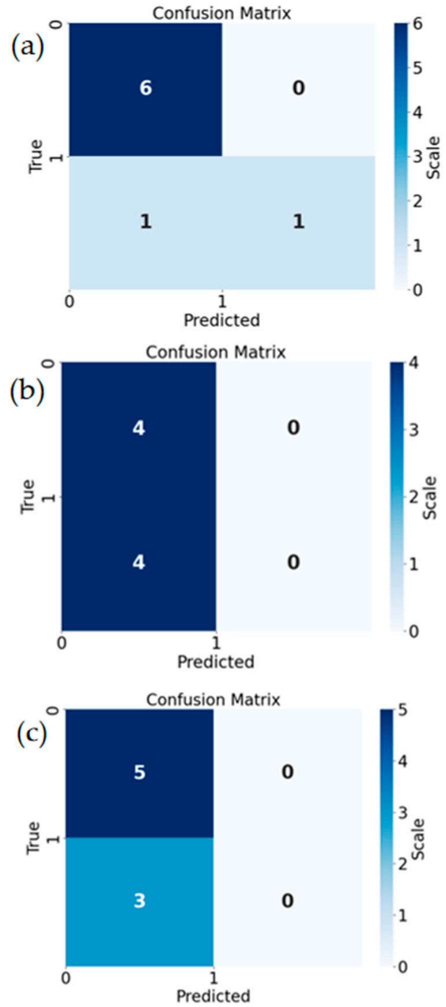

Figure 4. Confusion Matrix Diagrams for Three Machine Learning Models. ((**a**): Random Forest; (**b**): Xgboost; (**c**): Decision Tree).

4.2.2. Advantages and Limitations of Machine Learning Algorithms

In the application of ML algorithms for groundwater quality assessment, several advantages were observed, such as high predictive accuracy and the ability to handle complex, nonlinear relationships. However, these benefits come with certain limitations, including the computational expense and time required for training sophisticated models like Xgboost and Random Forest. The trade-off between model complexity and computational efficiency is an important consideration for practical implementation, particularly in resource-constrained settings.

4.2.3. Scientific and Industrial Background of Machine Learning Algorithms

ML technologies have a solid foundation in both scientific theory and industrial application. In the scientific domain, ML has been extensively studied for its ability to model and predict complex systems. Industrially, ML is used to optimize processes, automate decision-making, and enhance efficiency in various sectors, including water management. The technical implementation involves data preprocessing, feature selection, model training, and hyperparameter tuning. The selected algorithms are known for their balance between accuracy and interpretability to ensure that the results are not only effective but also understandable.

4.2.4. The Practical Application of Machine Learning-Based Approach in Water Resource Management

The models were trained on historical water quality data and used for the prediction of water quality changes. The results from ML models provided actionable insights for the local environmental agency. For instance, the identification of key water quality indicators (e.g., pH, Manganese, Nickel) allowed for targeted monitoring and mitigation efforts. The use of a machine learning-based approach led to the development of a more effective water quality monitoring program and contributed to the improvement of water resource management in the Yopurga landfill area.

The case study demonstrates the potential of an approach to address complex water quality issues and provides a template for its application in other regions facing similar challenges.

4.2.5. Scalability and Transferability of the Machine Learning Framework

The scalability and transferability of the ML framework are key to its broader application. Designing the framework to be modular and adaptable allowed it to be applied to different geographical locations and environmental monitoring scenarios. Abbas et al. utilized six algorithms, including Random Forest, Xgboost, and Decision Trees, for water quality prediction. They demonstrated the effectiveness of machine learning models in predicting the WQI in the Mirpurkhas area of Sindh province, Pakistan [44]. While the framework is inherently transferable, the model parameters may require recalibration for new environments.

4.2.6. Future Applications and Studies

The application of machine learning approaches in this study has opened avenues for future research and practical applications in water environment protection. Machine learning models offer a promising alternative to traditional modeling techniques due to their adaptability, accuracy, and ability to analyze large datasets.

Future studies should focus on the following areas:

(1) Cross-validation of ML models with different datasets to ensure robustness and reliability.

(2) Integration of ML models with real-time monitoring systems for continuous water quality assessment.

(3) Exploration of hybrid models that combine ML with other techniques, such as geographical information systems (GIS), for more comprehensive assessments.

4.3. Water Quality Index Model Component Analysis

4.3.1. Results of Indicator Selection

The selection of groundwater quality evaluation indicators in this study was based on the hydrochemical characteristics of the groundwater in the study area and the available datasets. Thallium was excluded from the initial list of 11 indicators because it only existed in trace amounts in the samples and did not exceed safe thresholds, thereby reducing excessive concern over potential exceedances of this indicator [45].

Ultimately, 10 water quality indicators were selected for evaluation: pH, NH_3-N, Mn, Ni, B, Pb, Zn, Fluoride, COD, and Fe. The selection of water quality indicators was made independently of subjective expert judgment, and reducing the number of indicators helped avoid the uncertainty introduced by too many dimensions, which could interfere with water quality evaluation results. By carefully selecting key indicators, the study enhanced the accuracy of water quality status judgments.

4.3.2. Sub-Index Functions

As shown in Figure 5, this study differentiated individual water quality parameters through SI scoring, mitigating the inherent ambiguity in overall water quality assessments. A scatter plot generated for the Yopurga landfill illustrates the individual SI values for ten scrutinized parameters, with red markers indicating the SI scores at each monitoring station. A score of 100 represents the best water quality, while a score of 0 indicates the worst. Multiple studies have shown that the SI method is one of the main sources of uncertainty in the WQI model [10,15,46,47]. Uncertainty typically arises when the SI method estimates small values without any input indicators exceeding critical threshold values. Another complexity occurs when SI functions estimate higher values for input indicators that exceed critical threshold values, known as ambiguity in the WQI model. Evidence suggests that while sub-indices reduce ambiguity, they cannot completely eliminate uncertainty within the WQI model. The overall condition of pH values appears relatively good, with data points all within the threshold range. For other parameters, there is an inverse relationship between parameter values and their corresponding SI scores when values are below standard thresholds. Specifically, SI values for Mn, B, and NH_3-N reached 0 at some measurement points, particularly Mn, which had SI values of 0 at multiple points, possibly indicating significant negative impacts on water quality in the area due to Mn content exceeding standards. Continuous exceedances of these thresholds pose potential long-term risks to ecosystems and public health, even if the water is used for irrigation rather than for drinking.

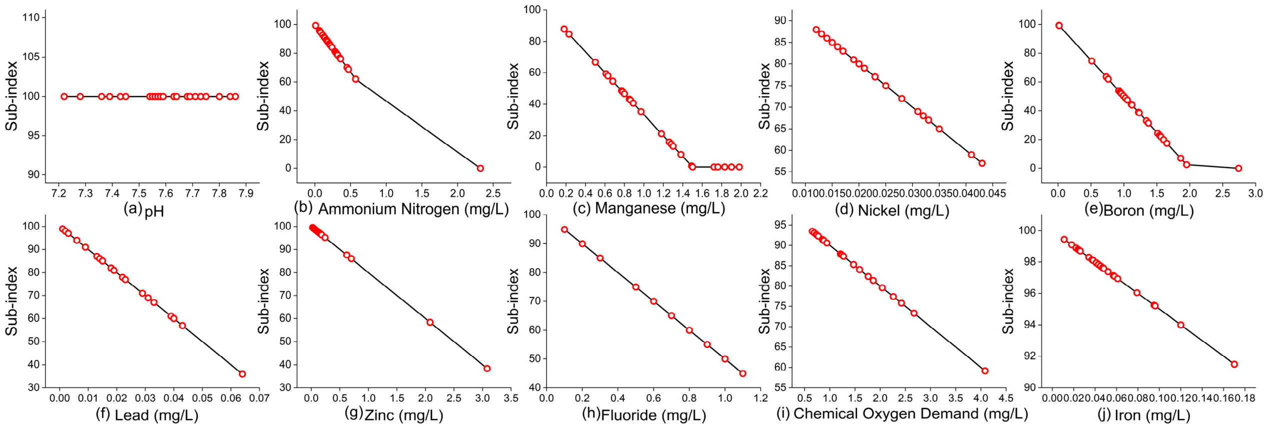

Figure 5. SI values versus corresponding indicator concentration values at each point for each water quality indicator. The red circle represents the SI value calculated for a specific indicator at a point location.

Analysis in Figure 6 shows significant heterogeneity in the function values of different water quality indicators. Normal distribution is a fundamental assumption for data symmetry and central tendency, and many statistical tests are predicated on this basis. From the perspective of individual indicators, Mn and B exhibited the widest range of variation, with function values ranging from 0 to 48.7 and 23 to 52.5, respectively. Mn displayed a certain degree of skewness with several outliers, indicating its distribution was not normal, while the variability in SI function values for pH and Fe was comparatively lower. Specifically, pH values consistently scored full marks across all sub-indices, indicating uniformity and excellent performance throughout the dataset. Meanwhile, the SI function values for Fe content fluctuated within a narrower range of 97.1 to 98.7, reflecting its stability. These observations reveal that the quality indicators for Mn and B generally performed poorly across the sampled locations, significantly below satisfactory quality thresholds.

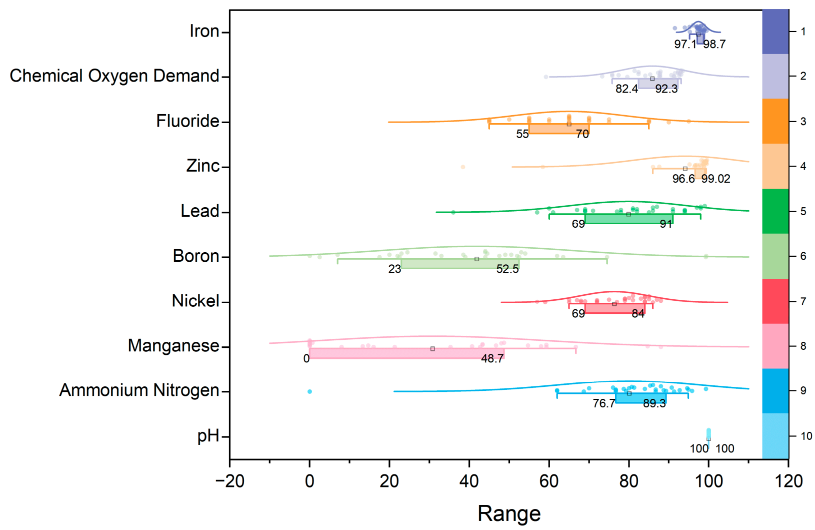

Figure 6. Distribution of SI values for different water quality indicators.

In summary, based on the output of the SI functions, it can be concluded that the presented results accurately reflect the state of various water quality indicators. Particularly in the case of the Yopurga landfill, the poor performance of Mn and B could be attributed to a combination of factors, including the mix of pollutants and environmental effects of leachate. These findings highlight the critical need for continuous monitoring and management strategies to address the potential adverse effects of exceeding quality thresholds for specific indicators, ensuring the protection of ecosystems and public health.

4.3.3. Weight Function

The validated Xgboost algorithm was applied in this study to quantitatively rank water quality classification indicators according to their importance. As illustrated in Figure 7, the results reveal that pH, Mn, and Ni have relative importance weights of 0.16, 0.14, and 0.13, respectively, indicating their significant impact on the model's predictive performance. In contrast, F-, NH_3-N, and COD have lower importance weights of 0.05, 0.07, and 0.07, respectively. This weight allocation method based on the Xgboost algorithm significantly reduces the model's prediction uncertainty compared to traditional models that rely on expert judgment. Xgboost assigns weights to the ten water quality indicators by constructing decision trees and evaluating the importance of features. Within the studied groundwater area, pH, Mn, and Ni were identified as particularly critical indicators for water quality classification predictions due to their high weights.

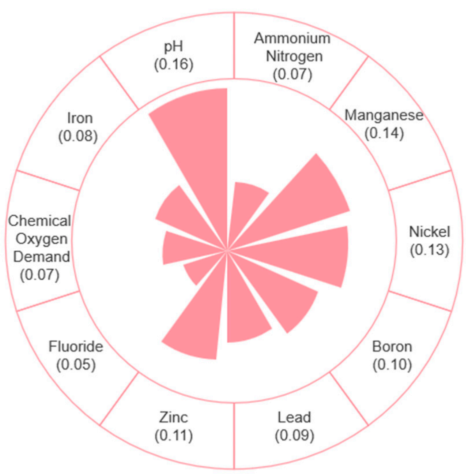

Figure 7. Weighting diagram for different water quality indicators.

4.3.4. Aggregation Functions

In this research, the WQI for the Yopurga landfill was calculated using six different aggregation functions: NSF, SRDD, WJ, WQM, LQM, and SWM. The boxplot in Figure 7 displays the distribution of WQI scores calculated by each aggregation function.

From Figure 8, it can be observed that the WQI distribution from the NSF aggregation function primarily ranges between medium and good quality, indicating relatively satisfactory water quality results. Conversely, scores derived from the SRDD aggregation function are concentrated in the polluted to tolerable range, suggesting poorer water quality conditions. Scores obtained with the WJ aggregation function are mainly distributed from margined to fair, indicating some fluctuation in water quality. Scores from the WQM aggregation function are centered around the fair category, while those from the LQM function are mostly labeled as slightly polluted. The SWM aggregation function produces WQI distributions ranging from moderately polluted to slightly polluted, potentially indicating problematic areas of water quality.

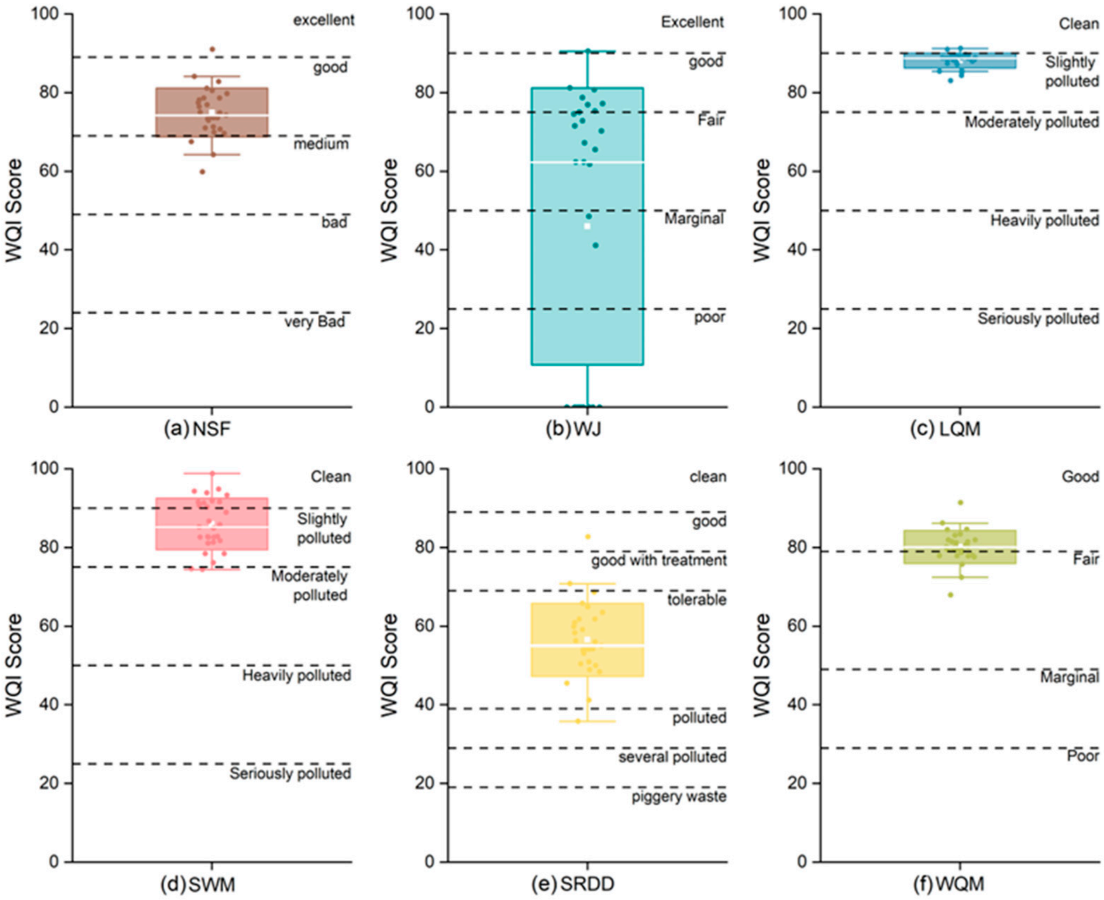

Figure 8. Range of WQI scores calculated for six different aggregation functions at the Yopurga landfill. (Horizontal dashed lines indicate the scores used for water quality classification. White boxes indicate average scores and the solid white line in each box is the median WQI score).

The WJ model's output exhibits noticeable score variability, suggesting that this model may be more sensitive to changes in water quality indicator variables. In contrast, outputs from the LQM model show minimal score variation, implying it may not be sufficiently sensitive to changes in water quality data. Observing the outputs of the WQM, LQM, and SWM models reveals they operate within a consistent scoring range, possibly reflecting their consistency in water quality assessment or similar responses to certain water quality indicators.

The variation in WQI scores across different aggregation functions may significantly be influenced by their respective computational approaches, emphasizing the importance of selecting suitable aggregation functions in water quality assessments. This also reveals the sensitivity of different aggregation methods to interpret and respond to changes in water quality. To ensure the accuracy and reliability of water quality assessments, further research is needed on how these aggregation functions reflect actual water quality conditions.

4.4. Comparison of Different Aggregation Functions

The WQI results calculated through various weighted aggregation functions demonstrated significant variations across different monitoring points at the Yopurga landfill. Solid lines represent the thresholds between different water quality grades, ascending from poor to good quality.

In Figure 9a, the NSF-WQI results showed the following: Class I water samples constituted 1, Class II samples 25, and Class III samples 3, accounting for 3.45%, 86.21%, and 10.35% of the total samples, respectively. In Figure 9b, the SRDD-WQI results indicated the following: no Class I samples, 1 Class II sample, 1 Class III sample, 26 Class IV samples, and 1 Class V sample, representing 0.00%, 3.45%, 3.45%, 89.65%, and 3.45% of the total samples, respectively. In Figure 9c, the WQM-WQI results revealed the following: 16 Class I samples and 13 Class II samples, making up 55.17% and 44.83% of the total, respectively. In Figure 9d, the LQM-WQI results showed the following: 2 Class I samples and 27 Class II samples, accounting for 6.90% and 93.10% of the total, respectively. In Figure 9e, the SWM-WQI results indicated the following: 9 Class I samples and 20 Class II samples, comprising 31.03% and 68.97% of the total, respectively. In Figure 9f, the WJ-WQI results showed the following: no Class I samples, 6 Class II samples, 11 Class III samples, 2 Class IV samples, and 10 Class V samples, accounting for 0.00%, 20.69%, 37.93%, 6.90%, and 34.48% of the total, respectively. It can be seen from Figure 8 that WQI results calculated using the WJ model at multiple monitoring points scored zero, indicating extremely poor water quality at these points under this model. However, in the other five models, most monitoring points scored above the passing line, showing relatively satisfactory water quality. The WQM, SWM, and LQM models especially displayed relatively good water quality performance at most monitoring points. Within the WQI series, WQM and SWM evaluated the water quality of the Yopurga landfill as relatively good, with WJ being the poorest. Overall, excluding WJ, the WQI results suggest nearly all monitoring points' water quality exceeds Class IV standards, implying overall water quality ranges from average to good. Different WQI models may significantly vary in sensitivity to input data, with the WJ model showing zero scores at multiple points due to its susceptibility to the concentration of disqualifying indicators. Significant differences between models also reflect variations in evaluation standards and actual water quality conditions at each monitoring point. Further analysis of the correlation among various WQI models will follow.

Figure 10 displays the Pearson correlation coefficients between the results of different WQI models. There exists a significant positive correlation among most WQI model results, particularly between NSF, SRDD, WQM, and SWM models, with correlation coefficients exceeding 0.9. This indicates that when one model suggests an improvement in water quality, other models tend to show similar improvement trends. Conversely, the correlation between the WJ model and WQM LQM models is weaker, with coefficients of 0.26 and 0.39, respectively, indicating the WJ model might employ a different calculation method or evaluation standard, thus showing lower consistency with other models. The stronger correlation among the five models, excluding WJ, indicates a consistent trend in water quality assessments, with the anomaly in the WJ model likely due to its sensitivity to specific indicators or differences in computational methods. Additionally, the WQI of this study is compared with other studies using different WQI evaluation methods in similar contexts, as shown in Table 5.

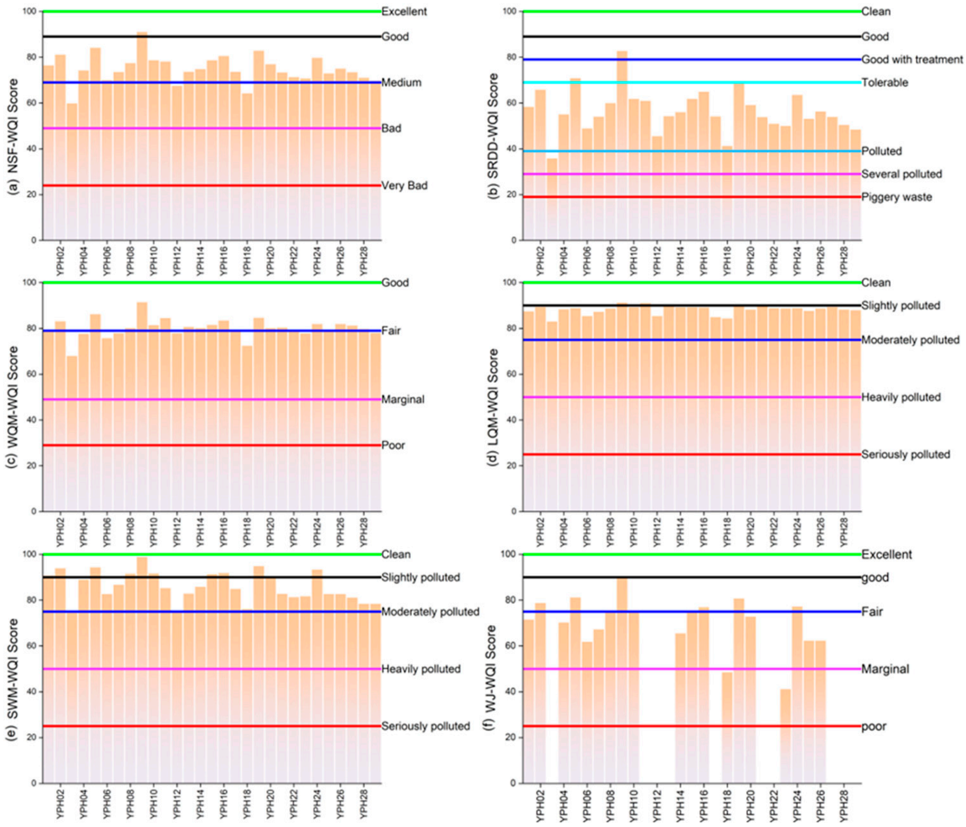

Figure 9. WQI results for different WQI models based on different points: solid lines represent different levels of water quality.

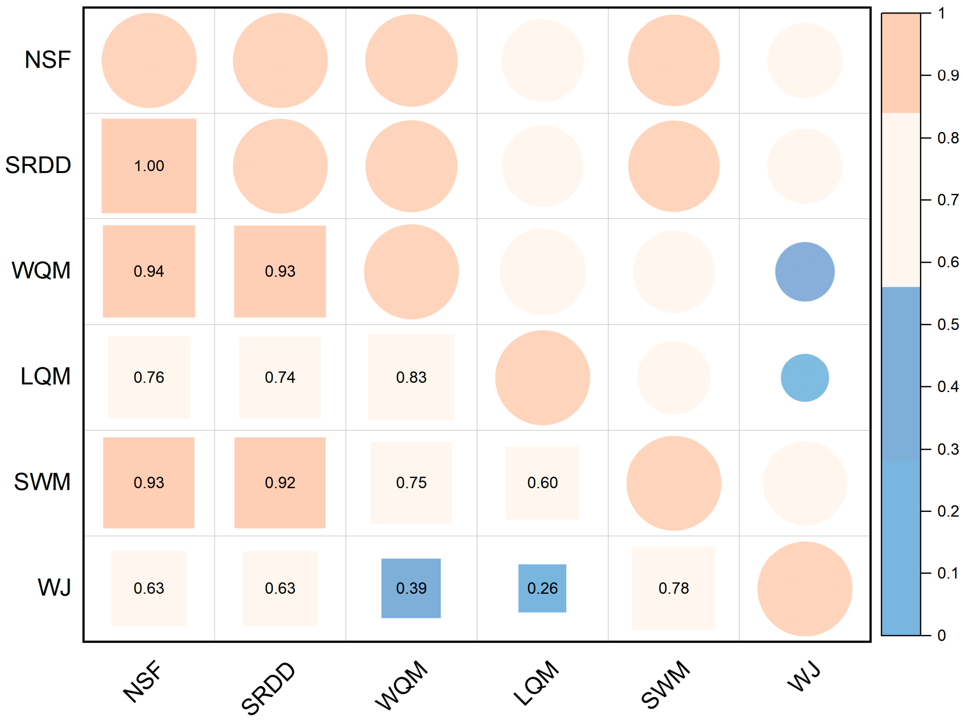

Figure 10. Correlation coefficients between different WQI models.

Table 5. Comparison table of studies evaluating WQI methods.

Study	Region	Evaluation Method	WQI Range	Key Indicators	Remarks
The Study	the Yopurga Landfill	Machine Learning Optimized WQI	Moderate Pollution to Slight Pollution	pH, Mn, Ni	Xgboost algorithm used to determine indicator weights
Use of Principal Component Analysis for parameter selection for development of a novel Water Quality Index: A case study of river Ganga India	Ganges Basin, India	PCA	---------	Dissolved Oxygen (DO), pH, Conductivity, Biochemical Oxygen Demand (BOD), Total Coliform (TC), Chlorides, Magnesium, Sulfates, and Total Dissolved Solids (TDS)	PCA analysis reduced the number of parameters from 28 to 9
Assessment of groundwater quality in a highly urbanized coastal city using water quality index model and Bayesian model averaging	Shenzhen	Machine Learning Optimized WQI	In the marginal to good level	NH3-N, Mn, pH	Xgboost algorithm and ROC weight method used to determine indicator weights

4.5. Eclipsing Error Analysis in Water Quality Models

Figure 11 shows the eclipsing distribution when assessing water quality using different WQI models. Eclipsing is categorized as overestimation or underestimation, representing the model's deviation from actual conditions [16,48]. Results suggest that underestimation eclipsing issues (where assessed water quality is worse than the actual condition) appear more common across all models. The LQM and WQM models displayed relatively fewer overestimation eclipsing errors, with only 1 and 2 occurrences among 21 and 9 misjudgments, respectively. The WQM model had an overestimation eclipsing rate of 6.9% and an underestimation eclipsing rate of 24%, while the SWM model showed a 0% overestimation eclipsing rate and a 34% underestimation rate.

In studies by Ding Fei et al. [42], using the SWM-WQI model to assess surface water quality, a 25.49% bias occurrence was found, with a higher overestimation eclipsing rate of 19.61% and an underestimation eclipsing rate of 5.88%. The findings in this study are similar, indicating that the SWM model tends to have a significant issue with overestimation eclipsing when evaluating both surface and groundwater quality. Overestimation eclipsing can lead to overly optimistic water quality assessments, potentially resulting in serious consequences, especially in the realms of water quality regulation and environmental protection. In contrast, underestimating water quality status is usually considered preferable over overestimation because it prompts more cautious water quality improvement measures. Therefore, selecting a WQI model that tends toward underestimation rather than overestimation may be safer, avoiding the risks associated with overly optimistic water quality evaluations.

In summary, based on the observations and to reduce the risk of overestimation, the SWM model emerges as a more suitable choice because it did not show cases of overestimation. Choosing such a model in decision-making processes can lower the potential negative impacts of overrated water quality assessments while also encouraging relevant managers to remain vigilant and take necessary measures for water quality improvement.

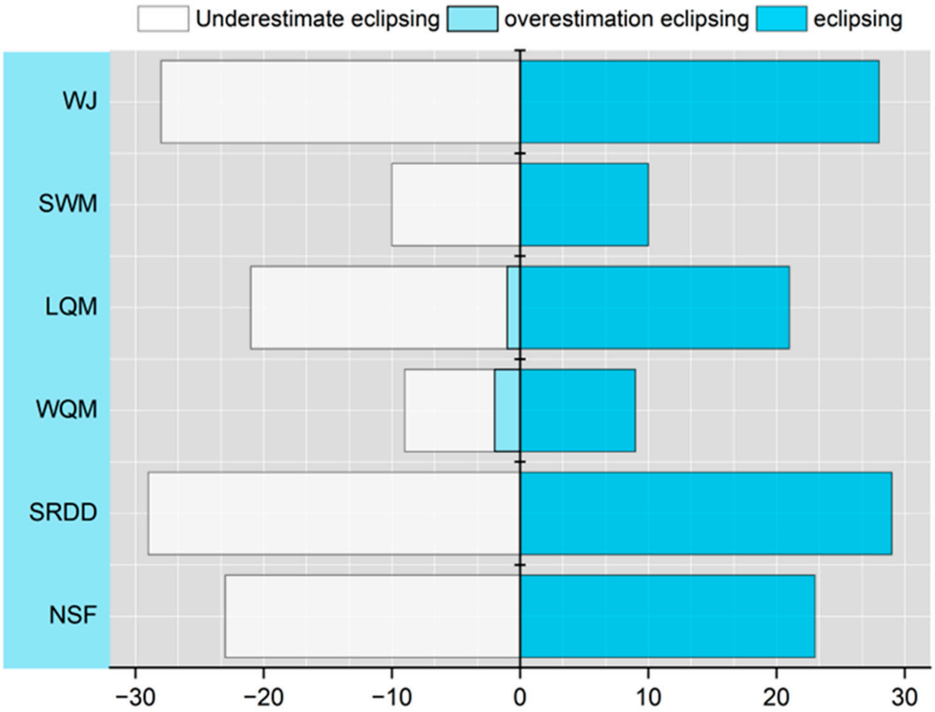

Figure 11. Eclipsing error analysis of different WQI models.

5. Conclusions

This study primarily selected ten indicators based on existing data and groundwater chemical characteristics to predict water quality using Decision Trees, Random Forest, and Xgboost algorithms. The main conclusions drawn are as follows:

(1) The Xgboost model was preferred for determining key indicator weights due to its outstanding performance and adaptability to the dataset. Using a data-driven machine learning approach helps to overcome the subjectivity and uncertainty inherent in traditiome the subjectivity and uncertainty inherent in traditional WQI methods.

(2) The key indicators selected were pH, Mn, and Ni. In the calculation of SI functions, it was found that the indicators for Mn and B performed poorly.

(3) The indicator weights, combined with SI functions, were used to calculate aggregation functions, and upon comparing the results of six weighted aggregation functions, significant inconsistencies were found between the WJ aggregation function and others. Moreover, the SWM model, due to its lower rate of eclipsing, particularly in underestimation eclipsing, became the most suitable aggregation function for this study, ensuring the accuracy and practicality of the evaluation results.

(4) The WQI results from the SWM aggregation function indicate that the water quality at the Yopurga landfill ranges from moderately polluted to slightly polluted.

This study enhances the objectivity and accuracy of the assessment method by combining advanced machine learning technology with a comprehensive evaluation approach. Additionally, it offers new directions for future research, particularly in analyzing long-term time series data and considering water quality assessments under different geographical and environmental conditions.

Author Contributions: Conceptualization, H.Z.; methodology, S.H.; software, J.L.; validation, Y.X.; formal analysis, Y.W.; investigation, S.H.; resources, J.L.; data curation, Y.X.; writing—original draft preparation, J.L.; writing—review and editing, H.Z.; visualization, Y.W.; supervision, S.H.; project administration, J.L.; funding acquisition, J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Detailed Investigation and Risk Assessment of the Environmental Conditions of Groundwater at the Yopurga Landfill Project (Grant No: 202112).

Data Availability Statement: The original contributions presented in the study are included in the article material, further inquiries can be directed to the corresponding authors.

Conflicts of Interest: Author Hongmei Zheng was employed by HUAZE EcoEnviron Technology & Engineering Institute (Beijing) Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Kumar, P. Simulation of Gomti River (Lucknow City, India) Future Water Quality under Different Mitigation Strategies. *Heliyon* **2018**, *4*, e01074. [CrossRef] [PubMed]
2. Han, Z.; Ma, H.; Shi, G.; He, L.; Wei, L.; Shi, Q. A Review of Groundwater Contamination near Municipal Solid Waste Landfill Sites in China. *Sci. Total Environ.* **2016**, *569–570*, 1255–1264. [CrossRef]
3. He, N.; Liu, L.; Wei, R.; Sun, K. Heavy Metal Pollution and Potential Ecological Risk Assessment in a Typical Mariculture Area in Western Guangdong. *Int. J. Environ. Res. Public Health* **2021**, *18*, 11245. [CrossRef] [PubMed]
4. Yao, X.; Cao, Y.; Zheng, G.; Devlin, A.T.; Yu, B.; Hou, X.; Tang, S.; Xu, L.; Lu, Y. Use of Life Cycle Assessment and Water Quality Analysis to Evaluate the Environmental Impacts of the Bioremediation of Polluted Water. *Sci. Total Environ.* **2021**, *761*, 143260. [CrossRef] [PubMed]
5. Obiri, S.; Addico, G.; Mohammed, S.; Anku, W.W.; Darko, H.; Collins, O. Water Quality Assessment of the Tano Basin in Ghana: A Multivariate Statistical Approach. *Appl. Water Sci.* **2021**, *11*, 49. [CrossRef]
6. Sim, S.F.; Ling, T.Y.; Lau, S.; Jaafar, M.Z. A Novel Computer-Aided Multivariate Water Quality Index. *Environ. Monit. Assess.* **2015**, *187*, 181. [CrossRef] [PubMed]
7. Ding, F.; Chen, L.; Sun, C.; Zhang, W.; Yue, H.; Na, S. An Upgraded Groundwater Quality Evaluation Based on Hasse Diagram Technique & Game Theory. *Ecol. Indic.* **2022**, *140*, 109024. [CrossRef]
8. Ding, F.; Zhang, W.; Chen, L.; Sun, Z.; Li, W.; Li, C.; Jiang, M. Water Quality Assessment Using Optimized CWQII in Taihu Lake. *Environ. Res.* **2022**, *214*, 113713. [CrossRef] [PubMed]
9. Lumb, A.; Sharma, T.C.; Bibeault, J.-F. A Review of Genesis and Evolution of Water Quality Index (WQI) and Some Future Directions. *Water Qual. Expo. Health* **2011**, *3*, 11–24. [CrossRef]
10. Uddin, M.G.; Nash, S.; Olbert, A.I. A Review of Water Quality Index Models and Their Use for Assessing Surface Water Quality. *Ecol. Indic.* **2021**, *122*, 107218. [CrossRef]
11. Uddin, M.G.; Nash, S.; Rahman, A.; Olbert, A.I. A Sophisticated Model for Rating Water Quality. *Sci. Total Environ.* **2023**, *868*, 161614. [CrossRef] [PubMed]
12. Shah, M.I.; Javed, M.F.; Alqahtani, A.; Aldrees, A. Environmental Assessment Based Surface Water Quality Prediction Using Hyper-Parameter Optimized Machine Learning Models Based on Consistent Big Data. *Process Saf. Environ. Prot.* **2021**, *151*, 324–340. [CrossRef]
13. Taromideh, F.; Fazloula, R.; Choubin, B.; Emadi, A.; Berndtsson, R. Urban Flood-Risk Assessment: Integration of Decision-Making and Machine Learning. *Sustainability* **2022**, *14*, 4483. [CrossRef]
14. Uddin, M.G.; Nash, S.; Rahman, A.; Olbert, A.I. Assessing Optimization Techniques for Improving Water Quality Model. *J. Clean. Prod.* **2023**, *385*, 135671. [CrossRef]
15. Sutadian, A.D.; Muttil, N.; Yilmaz, A.G.; Perera, B.J.C. Development of a Water Quality Index for Rivers in West Java Province, Indonesia. *Ecol. Indic.* **2018**, *85*, 966–982. [CrossRef]
16. Uddin, M.G.; Nash, S.; Rahman, A.; Olbert, A.I. A Comprehensive Method for Improvement of Water Quality Index (WQI) Models for Coastal Water Quality Assessment. *Water Res.* **2022**, *219*, 118532. [CrossRef] [PubMed]
17. Uddin, M.G.; Nash, S.; Rahman, A.; Olbert, A.I. A Novel Approach for Estimating and Predicting Uncertainty in Water Quality Index Model Using Machine Learning Approaches. *Water Res.* **2023**, *229*, 119422. [CrossRef] [PubMed]
18. Gazzaz, N.M.; Yusoff, M.K.; Aris, A.Z.; Juahir, H.; Ramli, M.F. Artificial Neural Network Modeling of the Water Quality Index for Kinta River (Malaysia) Using Water Quality Variables as Predictors. *Mar. Pollut. Bull.* **2012**, *64*, 2409–2420. [CrossRef]
19. Wang, X.; Tian, Y.; Liu, C. Assessment of Groundwater Quality in a Highly Urbanized Coastal City Using Water Quality Index Model and Bayesian Model Averaging. *Front. Environ. Sci.* **2023**, *11*, 1086300. Available online: https://www.frontiersin.org/articles/10.3389/fenvs.2023.1086300/full (accessed on 1 June 2024).
20. Forecasting Groundwater Quality Using Automatic Exponential Smoothing Model (AESM) in Xianyang City, China-Web of Science Core Collection. Available online: http://webofscience-clarivate-cn-s.libziyuan.bjut.edu.cn:8118/wos/woscc/full-record/WOS:000812667000001 (accessed on 27 May 2024).
21. Aghamohammadghasem, M.; Azucena, J.; Hashemian, F.; Liao, H.; Zhang, S.; Nachtmann, H. System Simulation And Machine Learning-Based Maintenance Optimization for an Inland Waterway Transportation System. In Proceedings of the 2023 Winter Simulation Conference (WSC), San Antonio, TX, USA, 10–13 December 2023; Available online: https://ieeexplore.ieee.org/document/10408112 (accessed on 27 May 2024).

22. Li, Q.; Lu, L.; Zhao, Q.; Hu, S. Impact of Inorganic Solutes' Release in Groundwater during Oil Shale In Situ Exploitation. *Water* **2022**, *15*, 172. Available online: https://www.mdpi.com/2073-4441/15/1/172 (accessed on 1 June 2024).
23. Zhang, J.; Zhou, J.; Zhou, Y.; Zeng, Y.; Ji, Y.; Sun, Y.; Lei, M. Hydrogeochemical Characteristics and Groundwater Quality Assessment in the Plain Area of Yarkant River Basin in Xinjiang, P.R. China. *Environ. Sci. Pollut. Res.* **2021**, *28*, 31704–31716. [CrossRef]
24. Radhakrishnan, N.; Pillai, A.S. Comparison of Water Quality Classification Models Using Machine Learning. In Proceedings of the 2020 5th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 10–12 June 2020; pp. 1183–1188.
25. El Bilali, A.; Taleb, A.; Brouziyne, Y. Groundwater Quality Forecasting Using Machine Learning Algorithms for Irrigation Purposes. *Agric. Water Manag.* **2021**, *245*, 106625. [CrossRef]
26. Naloufi, M.; Lucas, F.S.; Souihi, S.; Servais, P.; Janne, A.; Wanderley Matos De Abreu, T. Evaluating the Performance of Machine Learning Approaches to Predict the Microbial Quality of Surface Waters and to Optimize the Sampling Effort. *Water* **2021**, *13*, 2457. [CrossRef]
27. Bedi, S.; Samal, A.; Ray, C.; Snow, D. Comparative Evaluation of Machine Learning Models for Groundwater Quality Assessment. *Environ. Monit. Assess.* **2020**, *192*, 776. [CrossRef]
28. Nasif, A.; Othman, Z.A.; Sani, N.S. The Deep Learning Solutions on Lossless Compression Methods for Alleviating Data Load on IoT Nodes in Smart Cities. *Sensors* **2021**, *21*, 4223. [CrossRef]
29. Elgeldawi, E.; Sayed, A.; Galal, A.R.; Zaki, A.M. Hyperparameter Tuning for Machine Learning Algorithms Used for Arabic Sentiment Analysis. *Informatics* **2021**, *8*, 79. [CrossRef]
30. Shekar, B.H.; Dagnew, G. Grid Search-Based Hyperparameter Tuning and Classification of Microarray Cancer Data. In Proceedings of the 2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP), Gangtok, India, 25–28 February 2019; pp. 1–8.
31. Powers, D.M. Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation. *arXiv* **2020**. Available online: https://arxiv.longhoe.net/abs/2010.16061 (accessed on 31 May 2024).
32. The Relationship between Precision-Recall and ROC Curves | Proceedings of the 23rd International Conference on Machine Learning. Available online: https://dl.acm.org/doi/abs/10.1145/1143844.1143874 (accessed on 31 May 2024).
33. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets-Web of Science Core Collection. Available online: http://webofscience-clarivate-cn-s.libziyuan.bjut.edu.cn:8118/wos/woscc/full-record/WOS:000350685900033 (accessed on 31 May 2024).
34. A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation | SpringerLink. Available online: https://springer.longhoe.net/chapter/10.1007/978-3-540-31865-1_25 (accessed on 31 May 2024).
35. Root Mean Square Error (RMSE) or Mean Absolute Error (MAE)?—Arguments against Avoiding RMSE in the Literature-Web of Science Core Collection. Available online: http://webofscience-clarivate-cn-s.libziyuan.bjut.edu.cn:8118/wos/woscc/full-record/WOS:000341600100030 (accessed on 31 May 2024).
36. The Use of the Area under the Roc Curve in the Evaluation of Machine Learning Algorithms-Web of Science Core Collection. Available online: http://webofscience-clarivate-cn-s.libziyuan.bjut.edu.cn:8118/wos/woscc/full-record/WOS:A1997XE56500009 (accessed on 31 May 2024).
37. Li, Q.; Liu, A.; Yu, K.; Yu, K.F. A Weighted Rank-Sum Procedure for Comparing Samples with Multiple Endpoints. *Stat. Interface* **2009**, *2*, 197–201. [CrossRef] [PubMed]
38. Odu, G.O. Weighting Methods for Multi-Criteria Decision Making Technique. *J. Appl. Sci. Environ. Manag.* **2019**, *23*, 1449–1457. [CrossRef]
39. Hamlat, A.; Guidoum, A.; Koulala, I. Status and Trends of Water Quality in the Tafna Catchment: A Comparative Study Using Water Quality Indices. *J. Water Reuse Desalination* **2017**, *7*, 228–245. [CrossRef]
40. Bordalo, A.A.; Nilsumranchit, W.; Chalermwat, K. Water Quality and Uses of the Bangpakong River (Eastern Thailand). *Water Res.* **2001**, *35*, 3635–3642. [CrossRef] [PubMed]
41. Sutadian, A.D.; Muttil, N.; Yilmaz, A.G.; Perera, B.J.C. Using the Analytic Hierarchy Process to Identify Parameter Weights for Developing a Water Quality Index. *Ecol. Indic.* **2017**, *75*, 220–233. [CrossRef]
42. Ding, F.; Zhang, W.; Cao, S.; Hao, S.; Chen, L.; Xie, X.; Li, W.; Jiang, M. Optimization of Water Quality Index Models Using Machine Learning Approaches. *Water Res.* **2023**, *243*, 120337. [CrossRef]
43. Khoi, D.N.; Quan, N.T.; Linh, D.Q.; Nhi, P.T.T.; Thuy, N.T.D. Using Machine Learning Models for Predicting the Water Quality Index in the La Buong River, Vietnam. *Water* **2022**, *14*, 1552. [CrossRef]
44. Abbas, F.; Cai, Z.; Shoaib, M.; Iqbal, J.; Ismail, M.; Alrefaei, A.F.; Albeshr, M.F. Machine Learning Models for Water Quality Prediction: A Comprehensive Analysis and Uncertainty Assessment in Mirpurkhas, Sindh, Pakistan. *Water* **2024**, *16*, 941. Available online: https://www.mdpi.com/2073-4441/16/7/941?utm_campaign=releaseissue_waterutm_medium=emailutm_source=releaseissueutm_term=doilink133 (accessed on 1 June 2024).
45. Xiong, Y.; Zhang, T.; Sun, X.; Yuan, W.; Gao, M.; Wu, J.; Han, Z. Groundwater Quality Assessment Based on the Random Forest Water Quality Index—Taking Karamay City as an Example. *Sustainability* **2023**, *15*, 14477. [CrossRef]
46. Abbasi, T.; Abbasi, S.A. *Water Quality Indices*; Elsevier: Amsterdam, The Netherlands, 2012; ISBN 978-0-444-54304-2.

47. Juwana, I.; Muttil, N.; Perera, B.J.C. Uncertainty and Sensitivity Analysis of West Java Water Sustainability Index—A Case Study on Citarum Catchment in Indonesia. *Ecol. Indic.* **2016**, *61*, 170–178. [CrossRef]
48. Akhtar, N.; Ishak, M.I.S.; Ahmad, M.I.; Umar, K.; Md Yusuff, M.S.; Anees, M.T.; Qadir, A.; Ali Almanasir, Y.K. Modification of the Water Quality Index (WQI) Process for Simple Calculation Using the Multi-Criteria Decision-Making (MCDM) Method: A Review. *Water* **2021**, *13*, 905. [CrossRef]