

Article

Machine Learning Classification of High-Turbidity Exceedance Using Water Quality and Meteorological Predictors

Saleh H. Alhathloul *  and Yazeed Algurainy 

Department of Civil Engineering, College of Engineering, King Saud University, P.O. Box 800, Riyadh 11421, Saudi Arabia

* Correspondence: saalhathloul@ksu.edu.sa

Abstract

High-turbidity exceedance events are important indicators of abnormal coastal water quality conditions, but their relatively low frequency compared with normal observations makes reliable classification challenging. This study developed a scenario-based machine learning framework to classify high-turbidity exceedance using water quality and meteorological predictors from an automated seawater monitoring station along the Arabian Gulf. After quality control screening and temporal matching with NASA POWER meteorological variables, turbidity was converted into a binary target using a 2 NTU threshold, resulting in 4408 low-turbidity observations (85.3%) and 758 high-turbidity observations (14.7%). Four predictor scenarios were evaluated: meteorological-only selected variables, combined selected variables, water-quality-only all variables, and combined all variables. Random Forest, XGBoost, a Support Vector Machine, and K-Nearest Neighbors were tested under three class-imbalance treatments: undersampling, oversampling, and SMOTE. The meteorological-only scenario showed moderate classification ability, with ROC-AUC values of 0.81–0.86, but relatively low precision, indicating limited control of false alarms. Adding water quality predictors substantially improved classification performance, with the combined selected-variable scenario achieving ROC-AUC values of 0.96–0.98. Under oversampling, Random Forest achieved an accuracy = 0.95, precision = 0.83, specificity = 0.97, weighted F1-score = 0.95, and ROC-AUC = 0.98. The water-quality-only scenario also performed strongly, with accuracy values of 0.90–0.95 and ROC-AUC values of 0.95–0.96, confirming that direct water quality measurements carried most of the predictive signal. The combined all-variable scenario produced similarly high performance, with ROC-AUC values of 0.96–0.98, while variable selection reduced redundancy without loss of predictive skill. Computational analysis showed that the top-ranked XGBoost model under the combined all-variable oversampling scenario required only 0.08 s for training and 0.003 ms per sample for prediction. Overall, the results demonstrate that integrating selected water quality and meteorological predictors provides an accurate, efficient, and operationally practical framework for high-turbidity early-warning classification.



Academic Editor: Athanasios Sfetsos

Received: 8 July 2026

Revised: 18 July 2026

Accepted: 29 July 2026

Published: 30 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: turbidity exceedance; water quality; machine learning; binary classification; class imbalance; SMOTE; XGBoost; Random Forest

1. Introduction

Turbidity is a key water quality indicator because it reflects the presence of suspended particles, including sediments, organic matter, plankton, and colloidal materials within the water column [1–3]. Elevated turbidity can reduce light penetration, degrade aquatic

habitats, transport particle-associated pollutants, and indicate runoff, sediment resuspension, or pollution-related disturbances [4,5]. In drinking-water and operational monitoring systems, high particulate loads can also reduce treatment efficiency, increase filtration demand, interfere with disinfection processes, and signal abnormal water quality conditions requiring operational attention [6,7]. Therefore, reliable turbidity monitoring is essential for aquatic ecosystem protection, water treatment operation, and early detection of sediment- or contamination-related events [8,9].

High-turbidity exceedance events are often less frequent than normal water quality conditions, but they are operationally important because they may represent short-term disturbances requiring rapid response. In coastal and marine environments, turbidity variability can be influenced by both internal water quality conditions and external meteorological drivers, including wind, precipitation, temperature, humidity, radiation, and atmospheric pressure [10–12]. Recent remote sensing and water-quality monitoring studies also show that turbidity is strongly affected by spatially and temporally variable surface processes, which supports the need for flexible data-driven monitoring frameworks [13–16]. For this reason, threshold-based turbidity classification can be more useful for early-warning applications than continuous turbidity prediction alone, particularly when the main objective is to identify whether turbidity exceeds a predefined management or operational threshold [17,18].

Recent advances in data-driven environmental modeling have increased the use of machine learning methods for water quality prediction, classification, anomaly detection, and decision support [19–21]. Recent reviews show that algorithms such as Random Forest, gradient boosting, Support Vector Machines, artificial neural networks, long short-term memory networks, convolutional neural networks, and hybrid deep learning architectures have been widely applied to water quality forecasting, index estimation, real-time monitoring, and remote sensing-based retrieval [16,22]. Machine learning models are well suited for turbidity applications because they can capture nonlinear relationships and interactions among physicochemical, hydrological, meteorological, and operational predictors [13,23,24]. However, many previous turbidity studies have focused on continuous prediction or remote sensing retrieval rather than binary exceedance classification using integrated water quality and meteorological predictors. This leaves a need for classification frameworks that directly address high-turbidity events as a minority class and evaluate model performance under class-imbalance conditions.

Several machine learning classifiers are commonly used for environmental classification problems because they represent different learning mechanisms. Random Forest (RF) can model nonlinear predictor interactions and provide feature-importance information [25]. XGBoost is effective for structured tabular datasets because it combines sequential decision tree boosting with regularization [26]. Support Vector Machines (SVMs) can construct nonlinear decision boundaries using kernel functions, while K-Nearest Neighbors (KNN) provides a non-parametric distance-based classification benchmark [27,28]. Comparing these models under consistent predictor scenarios can clarify which learning strategy is most suitable for turbidity exceedance classification and whether ensemble, kernel-based, or distance-based approaches provide better discrimination of minority exceedance events.

A key challenge in exceedance classification is class imbalance because high-turbidity events usually occur less frequently than normal observations. Under imbalanced conditions, accuracy alone can be misleading, as a model may perform well for the majority class while failing to identify the minority exceedance class [18,29]. Class-imbalance treatments, including random undersampling, random oversampling, class weighting, and the Synthetic Minority Oversampling Technique (SMOTE), can improve learning from

minority-class observations [30–32]. Nevertheless, because resampling methods may affect models differently depending on data structure, noise, and class overlap, their influence should be evaluated empirically rather than assumed [33,34].

Predictor selection is also important because water quality and meteorological datasets often include strongly correlated variables. For example, temperature-related atmospheric variables, humidity-related variables, and salinity-related seawater indicators may contain overlapping information. Although correlated predictors can introduce redundancy and reduce interpretability, they may still contribute useful information to nonlinear classifiers [35–37]. Therefore, variable selection should consider both statistical redundancy and predictive relevance. Selected-variable models may be preferable when they achieve comparable performance to all-variable models while improving interpretability and reducing unnecessary predictor complexity.

Meteorological predictors derived from NASA POWER provide an accessible source of atmospheric and surface-weather information for environmental modeling, especially in regions where in situ meteorological data are limited [12,38,39]. These variables can support turbidity classification by representing external forcing conditions related to wind-driven mixing, rainfall, heat exchange, and sediment transport. However, meteorological data alone may not fully capture internal seawater conditions such as pH, conductivity, alkalinity, chloride, chlorine, and water temperature. Therefore, comparing meteorological-only, water-quality-only, and combined predictor scenarios is necessary to determine the added value of direct water quality measurements for high-turbidity classification.

The regional context of the Arabian Gulf and the wider Middle East further supports the need for localized turbidity-exceedance classification frameworks. Recent regional studies have shown that coastal water quality in Bahrain, Qatar, Kuwait, the UAE, Saudi Arabia, and the wider Arabian Gulf exhibits strong spatial and temporal variability in turbidity, chlorophyll-a, total suspended solids, salinity, nutrients, dissolved oxygen, Secchi depth, and other optical or physicochemical indicators [40–44]. These variations are shaped by shallow bathymetry, restricted water exchange, high evaporation, extreme temperature regimes, urbanization, desalination and industrial activities, sediment resuspension, and episodic biological or particulate-loading events [45–47]. However, most regional studies have focused on water quality assessment, standards development, long-term trend analysis, or satellite-based retrieval of optical water quality variables rather than operational classification of short-term high-turbidity exceedance events [48,49]. This creates a clear need for site-specific machine learning frameworks that combine automated in situ seawater observations with meteorological predictors to support early warning in Gulf coastal monitoring systems.

Accordingly, this study develops a scenario-based machine learning framework for classifying high-turbidity exceedance using water quality and meteorological predictors. The specific objectives are to (i) evaluate whether selected meteorological variables alone can classify turbidity exceedance; (ii) determine whether adding water quality variables improves classification performance; (iii) compare selected-variable scenarios with all-variable scenarios; (iv) assess the influence of imbalance treatments on minority-class detection; and (v) rank model performance using multiple classification metrics and computational-time considerations. We used an automated seawater monitoring dataset from the Arabian Gulf, where turbidity was converted into a binary target using a 2 NTU operational threshold and where water quality records were temporally matched with NASA POWER meteorological variables. This threshold was not intended as a universal regulatory limit but was adopted as a conservative exceedance criterion representing the upper-bound baseline of relatively clear open and coastal marine waters. Previous studies on open-coastal and reef-associated waters have shown that turbidity commonly remains within the <1–2 NTU range under

non-disturbed conditions, whereas higher values are often associated with sediment resuspension, river plume influence, runoff, wind-driven mixing, or other particulate-loading events [8,9,50]. In addition, long-term turbidity above approximately 3 NTU has been associated with sublethal coral stress, supporting the use of 2 NTU as a conservative early-warning threshold below reported ecological stress levels [51,52]. The final dataset included 5166 valid observations after quality control screening and temporal matching, and four predictor scenarios were evaluated: meteorological-only selected variables, combined selected variables, water-quality-only all variables, and combined all variables. Class-imbalance treatments, including oversampling and SMOTE, were also assessed as part of the modeling framework. This framework supports the development of practical early-warning tools for high-turbidity events in coastal water-quality monitoring systems.

2. Materials and Methods

2.1. Data Acquisition and Study Framework

The data used in this study were obtained from a seawater monitoring station along the eastern coast of Saudi Arabia in the Arabian Gulf. This station was selected because it was the only site accessible to the authors with a sufficiently long and complete record containing all required water quality variables. The exact station coordinates are not reported because the site is associated with confidential operational coastal water-quality monitoring infrastructure. The analyzed water quality variables were recorded using an automated monitoring system installed at the station. The original dataset comprised 9644 monitoring records collected between June 2017 and December 2023, including both daily and sub-daily measurements. Before model development, the raw records were subjected to quality control screening to remove missing, incomplete, and inconsistent observations. Following quality control screening and temporal matching with NASA POWER meteorological variables, 5166 valid observations were retained from the original 9644 records for analysis. The excluded records contained missing, incomplete, or inconsistent water quality measurements, resulting in an irregular monitoring record with temporal gaps rather than a fully continuous daily time series. Because these gaps limited the application of conventional chronological folds and high-turbidity observations represented a minority class, a stratified random split was used to preserve the class proportions in the training and testing subsets.

The water quality variables included turbidity (Turb), water temperature (Temp), pH, conductivity (Cond), chlorides (Cl^-), alkalinity (Alk), and chlorine (Chlor). Turbidity was used to construct the binary classification target (Y), while the remaining water quality variables were used as predictors in the combined modeling scenario. To complement the water quality observations, meteorological variables were obtained from the NASA POWER database for the same study period and location. The meteorological predictors included air temperature (T), relative humidity (RH), wind speed (WS), wind direction (WD), dew-point temperature (DEW), wet-bulb temperature (WET), specific humidity (QV), surface pressure (PS), shortwave radiation (SW), longwave radiation (LW), precipitation (Precip), cloud amount (CC), and surface temperature (TS).

The water quality and meteorological datasets were temporally aligned using the date and time information available in the records. This matching ensured that each water quality observation was paired with the corresponding meteorological conditions. After temporal matching, the integrated dataset was used to develop a binary classification framework for predicting turbidity exceedance.

Turbidity was transformed from a continuous variable into a binary response using a threshold value of 2 NTU. The binary target variable was defined as given by Equation (1), where T_i is the measured turbidity value for observation i , $Y_i = 1$ represents high-turbidity

conditions, and $Y_i = 0$ represents low- or normal-turbidity conditions. Accordingly, the classification objective was to estimate the probability of turbidity exceedance given the available predictor variables as expressed in Equation (2), where X_i represents the predictor variables for observation i .

$$Y_i = \begin{cases} 1, & \text{if } T_i > 2 \\ 0, & \text{if } T_i \leq 2 \end{cases} \quad (1)$$

$$P(Y_i = 1 \mid X_i) = P(T_i > 2 \mid X_i) \quad (2)$$

Four predictor scenarios were developed to evaluate the predictive contribution of meteorological variables, water quality variables, and their combined use for turbidity exceedance classification. In Scenario 1, only selected NASA-derived meteorological predictors were used to determine whether atmospheric and surface-weather conditions alone could classify high-turbidity exceedance. This scenario represents a weather-driven early-warning model. In Scenario 2, selected NASA meteorological predictors were combined with selected measured seawater quality variables to assess whether internal physicochemical water conditions improve classification performance beyond meteorological forcing alone. In Scenario 3, all available water quality variables were used without meteorological predictors to evaluate the standalone predictive value of direct seawater measurements. In Scenario 4, all available water quality and meteorological variables were combined to determine whether retaining the full predictor set provides additional improvement compared with the selected-variable scenarios. Comparing these four scenarios allowed the study to assess the relative importance of meteorological forcing, direct water quality measurements, variable selection, and predictor redundancy. The variables used in the modeling framework, along with their categories, units, and descriptions, are summarized in Table 1, while their descriptive statistics, including the minimum, first quartile, median, mean, third quartile, and maximum values, are presented in Figure 1.

Table 1. Description of variables used in the turbidity classification models.

Variable	Category	Unit	Definition/Description
Turb	Water quality	NTU	Seawater turbidity
Y	Binary target	Unitless	($Y = 1$) if turbidity (>2); ($Y = 0$) if turbidity (≤ 2)
Temp	Water quality	°C	Seawater temperature
pH	Water quality	Unitless	Seawater acidity/alkalinity
Cond	Water quality	µS/cm	Electrical conductivity
CL [−]	Water quality	mg/L	Chloride concentration
Alk	Water quality	mg/L as CaCO ₃	Seawater alkalinity
Chlor	Water quality	mg/L	Residual chlorine
T	Meteorological	°C	Air temperature at 2 m
RH	Meteorological	%	Relative humidity at 2 m
WS	Meteorological	m/s	Wind speed at 2 m
WD	Meteorological	degrees	Wind direction at 2 m
DEW	Meteorological	°C	Dew-point temperature at 2 m
WET	Meteorological	°C	Wet-bulb temperature at 2 m
QV	Meteorological	g/kg	Specific humidity at 2 m
PS	Meteorological	kPa	Surface pressure
SW	Meteorological	W/m ²	Shortwave radiation
LW	Meteorological	W/m ²	Longwave radiation
Precip	Meteorological	mm/hour	Corrected precipitation
CC	Meteorological	%	Cloud cover
TS	Meteorological	°C	Surface temperature

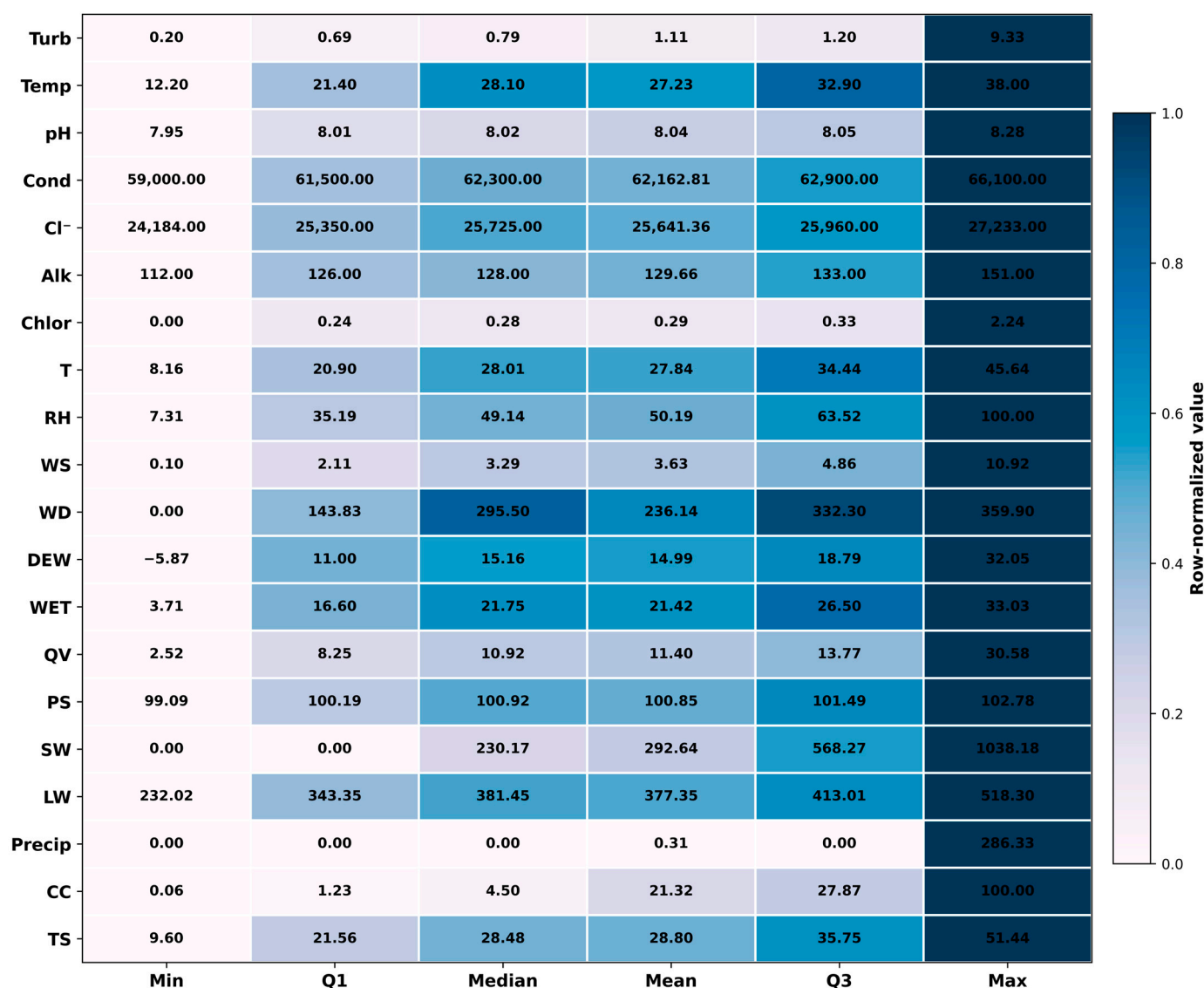


Figure 1. Descriptive statistics heat table showing the minimum, first quartile (Q1), median, mean, third quartile (Q3), and maximum values of the water quality and NASA meteorological variables used in the turbidity classification models.

2.2. Data Preprocessing and Class-Imbalance Treatment

After temporal matching of the water quality records with the NASA meteorological variables, the integrated dataset was prepared for binary classification modeling. Records with missing turbidity values were excluded because turbidity was required to define the response variable. Records with missing or incomplete predictor values were also removed during the data screening process. Therefore, the final reduced dataset used for model development contained only complete valid records, and no missing-value replacement was applied.

Feature scaling was applied according to the requirements of each model. Standardization was performed for scale-sensitive classifiers, including the SVM and KNN, because their learning processes depend on distances, margins, or similarity measures among observations. These predictors were standardized as given by Equation (3), where X_i^* is the standardized value of predictor X_i , μ is the mean of the training data, and σ is the standard deviation of the training data. In contrast, tree-based models, including RF and XGBoost,

do not require feature scaling because their learning process is based on decision-tree splitting rules rather than distance calculations or coefficient estimation [25].

$$X_i^* = \frac{X_i - \mu}{\sigma} \quad (3)$$

The dataset was divided into training and testing subsets using a stratified 80:20 split. Stratification was applied to preserve the relative proportions of low- and high-turbidity observations in both subsets, which was important because high-turbidity events represented the minority class. The training subset, representing 80% of the data, was used for model development, cross-validation, and hyperparameter tuning, whereas the remaining 20% was held out as an independent testing subset and used only for final model evaluation. This separation ensured that model tuning was performed without using the independent test data, thereby reducing optimistic bias and providing a more reliable assessment of model generalization to unseen observations.

Within the training subset, stratified five-fold cross-validation was applied for model tuning and validation. In this procedure, the training data were divided into five approximately equal folds while maintaining the proportion of low- and high-turbidity observations in each fold. During each cross-validation iteration, four folds were used for model training, and the remaining fold was used for validation. This process was repeated five times so that each fold served once as the validation fold. Hyperparameters were selected based on the average cross-validation performance across the five folds. A systematic Grid Search procedure was used to evaluate predefined combinations of model hyperparameters within the stratified five-fold cross-validation framework. For each model, the hyperparameter combination achieving the best mean validation performance was selected and subsequently fitted to the full training subset before evaluation on the independent testing subset.

Because high-turbidity observations represented the minority class, class imbalance was addressed using class weighting and three resampling strategies: random undersampling, random oversampling, and the Synthetic Minority Oversampling Technique (SMOTE). Random undersampling reduces the number of majority-class observations to balance the class distribution, whereas random oversampling balances the dataset by duplicating minority-class observations. In contrast, SMOTE generates synthetic minority-class observations based on nearest-neighbor relationships rather than simply duplicating existing samples, making it suitable for improving model learning from imbalanced datasets [53]. These imbalance treatments were considered for the machine learning classifiers to evaluate the sensitivity of model performance to different sampling strategies. Class weighting increases the penalty assigned to misclassification of the minority high-turbidity class. For XGBoost, imbalance was handled using the scale-positive-weight parameter, which was calculated as expressed in Equation (4), where $N_{negative}$ is the number of low-turbidity observations, and $N_{positive}$ is the number of high-turbidity observations.

$$scale_pos_weight = \frac{N_{negative}}{N_{positive}} \quad (4)$$

To avoid information leakage, all preprocessing steps that learned information from the data were applied only within the training workflow. Feature scaling was fitted using the training folds and then applied to the corresponding validation fold. Similarly, when undersampling, oversampling, or SMOTE was used, the resampling procedure was applied only to the training folds during cross-validation and was not applied before the train–test split, to the validation folds, or to the independent testing subset. After hyperparameter tuning, the final optimized models were evaluated on the independent

testing subset using the original class distribution to assess their ability to classify unseen observations [54]. All data preprocessing, statistical analyses, machine learning modeling, hyperparameter tuning, and performance evaluations were implemented in Python 3.13.9. Meteorological variables from NASA POWER were retrieved through the NASA POWER API and processed in Python, where temporal matching, data cleaning, and integration with the water quality dataset were performed using standard scientific libraries, including pandas, NumPy, and scikit-learn.

2.3. Variable Selection and Classification Model

Predictor correlation analysis was first performed to examine pairwise linear associations among the input variables before model development. Pearson's correlation coefficient was used because it provides a direct measure of the strength and direction of the linear association between two continuous variables [55]. Pearson's correlation coefficient was calculated using Equation (5), where r is the Pearson correlation coefficient, X and Y are two predictor variables, and $\bar{X}$ and $\bar{Y}$ are their mean values. The resulting correlation coefficients were arranged in a correlation matrix to visually assess the relationships among all predictors. Values of r range from -1 to $+1$, where values close to $+1$ indicate strong positive correlation, values close to -1 indicate strong negative correlation, and values close to 0 indicate weak or no linear correlation. Predictor pairs with an absolute correlation coefficient greater than 0.80 were considered strongly correlated.

$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}} \quad (5)$$

Feature-importance analysis was then applied to evaluate the relative contribution of each predictor to turbidity classification. RF was used for feature-importance estimation because it can evaluate predictor relevance based on the reduction in node impurity achieved by each variable across multiple decision trees [25,56]. The feature importance of predictor j was expressed in Equation (6), where FI_j is the feature-importance score of predictor j , and ΔI_j represents the reduction in impurity associated with splits using that predictor.

$$FI_j = \sum \Delta I_j \quad (6)$$

For binary classification, node impurity was measured using the Gini impurity (Equation (7)), where G is the Gini impurity, and p_c is the proportion of samples belonging to class c . A higher feature-importance value indicates that the predictor contributes more strongly to separating low- and high-turbidity observations.

$$G = 1 - \sum p_c^2 \quad (7)$$

The correlation matrix and feature-importance results were used together to guide variable selection. The correlation matrix was used to identify strongly correlated predictor pairs and potential multicollinearity, whereas feature-importance analysis was used to assess the predictive relevance of each variable. Variables were not removed based on correlation alone because highly correlated predictors may still provide useful information for nonlinear machine learning classifiers. Instead, when two predictors were strongly correlated, the variable with lower feature importance, weaker physical interpretability, or lower relevance to turbidity classification was considered for removal. This combined approach follows the principle that feature selection should account for both statistical redundancy and predictive relevance [57].

Accordingly, the variable-selection process was performed in three steps. First, the correlation matrix was examined to identify predictor pairs with strong correlation ($|r| > 0.80$). Second, feature-importance scores were used to determine which variables contributed most to classification performance. Third, variables showing both strong correlation and low feature importance were considered redundant and were excluded only when their removal did not reduce model performance during cross-validation or independent testing. This procedure reduced predictor redundancy while preserving variables that contributed useful information for classifying turbidity exceedance.

To assess the robustness of the variable-selection procedure, model performance was evaluated under four predictor scenarios, as summarized in Table 2. The selected-variable meteorological-only and combined scenarios were used as the main modeling configurations, whereas the water-quality-only and combined all-variable scenarios were used as sensitivity analyses. In all scenarios, three sampling strategies were evaluated: random undersampling, random oversampling, and SMOTE. This comparison allowed the study to determine whether variable selection improved classification performance, whether retaining all available water quality and meteorological predictors provided additional predictive information, and whether model performance was sensitive to the class-balancing strategy.

Table 2. Summary of predictor scenarios and sampling strategies evaluated in the turbidity classification framework.

Scenario	Predictor Set	Purpose
Scenario 1	Meteorological only, selected variables	Evaluate whether selected meteorological variables alone can classify turbidity exceedance.
Scenario 2	Combined, selected variables	Evaluate whether adding selected water quality variables improves classification performance.
Scenario 3	Water quality only, all variables	Evaluate whether measured water quality variables alone can classify turbidity exceedance.
Scenario 4	Combined, all variables	Test whether retaining all available water quality and meteorological predictors improves performance.

2.4. Classification Models

Four machine learning classifiers were used to predict turbidity exceedance: Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN). These models were selected to represent different classification strategies, including bagging-based ensemble learning, optimized gradient boosting, kernel-based classification, and distance-based classification. The models were trained and evaluated under the four predictor scenarios described in Table 2: meteorological-only selected variables, combined selected variables, water-quality-only all variables, and combined all variables.

2.4.1. Random Forest

RF was used as a nonlinear ensemble machine learning classifier. The method constructs multiple randomized decision trees and aggregates their predictions to improve predictive performance and reduce the instability of individual trees [25]. RF was included

because it can model nonlinear relationships, handle predictor interactions, and provide measures of predictor importance [56].

For binary classification, the predicted class is obtained by aggregating the predictions from all trees. The final class prediction can be calculated using Equation (8), where $\hat{Y}_i$ is the predicted class for observation i , $h_b(X_i)$ is the prediction from tree b , and B is the total number of trees in the forest.

$$\hat{Y}_i = \text{mode}\{h_b(X_i)\}_{b=1}^B \quad (8)$$

The predicted probability of turbidity exceedance can be estimated as the average proportion of trees voting for the high-turbidity class (Equation (9)), where $I(\cdot)$ is an indicator function that equals 1 if tree b predicts the high-turbidity class and 0 otherwise. The main tuned RF hyperparameters included the number of trees, maximum tree depth, minimum number of samples required to split an internal node, minimum number of samples required at a leaf node, maximum number of predictors considered at each split, and class-weighting option.

$$\hat{P}(Y_i = 1|X_i) = \frac{1}{B} \sum_{b=1}^B I(h_b(X_i) = 1) \quad (9)$$

2.4.2. XGBoost

XGBoost was used as an optimized gradient-boosting classifier. Unlike RF, which builds trees independently, XGBoost builds trees sequentially, with each new tree improving the previous ensemble by reducing classification errors. XGBoost was selected because it is effective for structured tabular datasets and can model nonlinear relationships and complex interactions among meteorological and water quality predictors [26].

The XGBoost model output is based on an additive ensemble of decision trees as expressed in Equation (10), where $\hat{y}_i$ is the model output for observation i , X_i is the predictor vector for observation i , f_k is the k -th decision tree, K is the total number of trees, and F is the space of regression trees.

$$\hat{y}_i = \sum_{k=1}^K f_k(X_i) \quad (10)$$

For binary classification, the predicted probability of turbidity exceedance is obtained using the logistic transformation as expressed in Equation (11). XGBoost optimizes a regularized objective function as expressed in Equation (12), where $l(y_i, \hat{y}_i)$ is the loss function that measures the difference between the observed and predicted values, and $\Omega(f_k)$ is the regularization term used to penalize model complexity [26]. The tuned hyperparameters included the number of estimators, maximum tree depth, learning rate, subsample ratio, column-sampling ratio, and scale-positive-weight parameter for class imbalance.

$$\hat{P}(Y_i = 1|X_i) = \frac{1}{1 + \exp(-\hat{y}_i)} \quad (11)$$

$$L = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (12)$$

2.4.3. Support Vector Machine

A Support Vector Machine was used as a kernel-based classifier. The SVM aims to identify an optimal decision boundary that separates the two turbidity classes with the maximum possible margin. Because the relationship between turbidity exceedance and the

predictor variables may be nonlinear, the kernel formulation of the SVM was used to allow nonlinear separation in a transformed feature space [27]. The SVM decision function can be calculated using Equation (13), where $f(X_i)$ is the decision function for observation i , w is the weight vector, $\phi(X_i)$ is the transformed feature vector, and b is the bias term. The predicted class is assigned according to the sign of the decision function as expressed in Equation (14).

$$f(X_i) = w^T \phi(X_i) + b \quad (13)$$

$$\hat{Y}_i = \begin{cases} 1, & f(X_i) > 0 \\ 0, & f(X_i) \leq 0 \end{cases} \quad (14)$$

An SVM was included because it can construct nonlinear decision boundaries and is suitable for binary classification problems with complex predictor relationships. Since SVM is sensitive to predictor scaling, standardized input variables were used. The main tuned SVM hyperparameters included the regularization parameter C , kernel type, kernel coefficient γ , and class-weighting option.

2.4.4. K-Nearest Neighbors

K-Nearest Neighbors was used as a distance-based classifier. KNN assigns a class to a new observation based on the classes of the nearest observations in the training data. It is a non-parametric method that does not assume a specific functional relationship between predictors and the response variable [28]. For a given observation X_i , the predicted class is determined using majority voting among its K nearest neighbors as expressed in Equation (15), where $N_K(X_i)$ is the set of the K nearest neighbors of observation X_i , and Y_j is the observed class of neighbor j .

$$\hat{Y}_i = \text{mode}\{Y_j : j \in N_K(X_i)\} \quad (15)$$

The predicted probability of high-turbidity conditions can be computed using Equation (16), where $I(Y_j = 1)$ is an indicator function equal to 1 when neighbor j belongs to the high-turbidity class and 0 otherwise. KNN was included to assess distance-based classification performance. Because KNN depends directly on distances among observations, standardized predictor variables were used. The tuned KNN hyperparameters included the number of neighbors, distance metric, weighting method, and class-imbalance treatment where applicable.

$$\hat{P}(Y_i = 1|X_i) = \frac{1}{K} \sum_{j \in N_K(X_i)} I(Y_j = 1) \quad (16)$$

2.5. Model Performance Evaluation

The performance of the turbidity classification models was evaluated using metrics derived from the confusion matrix. In this study, high-turbidity observations were treated as the positive class, while low- or normal-turbidity observations were treated as the negative class. The confusion matrix includes true positives (TPs), true negatives (TNs), false positives (FPs), and false negatives (FNs). TPs represent high-turbidity observations correctly classified as high turbidity, TNs represent low-turbidity observations correctly classified as low turbidity, FPs represent low-turbidity observations incorrectly classified as high turbidity, and FNs represent high-turbidity observations incorrectly classified as low turbidity.

Accuracy was first used to measure the overall proportion of correctly classified observations as expressed in Equation (17). Precision was calculated to evaluate the reliability of high-turbidity predictions using Equation (18).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (17)$$

$$Precision = \frac{TP}{TP + FP} \quad (18)$$

Recall, also known as sensitivity, was used to measure the ability of the model to correctly identify actual high-turbidity observations as expressed in Equation (19). Specificity, as expressed in Equation (20), was used to evaluate the ability of the model to correctly identify low- or normal-turbidity observations.

$$Recall = \frac{TP}{TP + FN} \quad (19)$$

$$Specificity = \frac{TN}{TN + FP} \quad (20)$$

The F1-score was calculated as the harmonic mean of precision and recall using Equation (21). Because the dataset was imbalanced and high-turbidity observations represented the minority class, the weighted F1-score was also used. The weighted F1-score calculates the F1-score for each class and weights it according to the number of observations in that class as expressed in Equation (22), where C is the number of classes, n_c is the number of observations in class c , n is the total number of observations, and $F1_c$ is the F1-score of class c . This metric was included because it provides a more balanced assessment than accuracy alone when class frequencies are unequal. The false-positive rate (FPR) was also calculated using Equation (23).

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (21)$$

$$Weighted\ F1 = \sum_{c=1}^C \frac{n_c}{n} F1_c \quad (22)$$

$$FPR = \frac{FP}{FP + TN} \quad (23)$$

In addition, the area under the receiver operating characteristic curve (ROC-AUC) was used to assess the ability of each model to distinguish between low- and high-turbidity classes across different classification thresholds as expressed in Equation (24), where (TPR) represents the true-positive rate, which is equivalent to recall. Higher ROC-AUC values indicate stronger separability between the two turbidity classes.

$$AUC = \int_0^1 TPR(FPR)d(FPR) \quad (24)$$

Model performance was evaluated using accuracy, precision, recall, specificity, F1-score, weighted F1-score, ROC-AUC, and computational time. These metrics were calculated using the independent testing subset after model training and hyperparameter tuning. Because high-turbidity observations represented the minority class, performance interpretation focused on precision, recall, F1-score, and ROC-AUC rather than accuracy alone. The weighted F1-score was included as an additional summary metric but was not used as the only performance criterion because it can be influenced by the majority class.

Computational time was also recorded to assess the practical efficiency of each classifier. Training time was measured as the time required to fit each optimized model, while prediction time was measured as the time required to generate class predictions for the testing subset. This allowed for the comparison of both predictive performance and computational efficiency across the RF, XGBoost, SVM, and KNN models under the four evaluated predictor scenarios. The results were summarized using confusion matrices, grouped bar plots of the classification metrics, ROC curves, and computational-time plots. These graphical outputs were used to compare model performance and efficiency across the evaluated classification scenarios.

Model ranking was performed using a multi-metric evaluation approach rather than relying on accuracy alone. Because the dataset was imbalanced and high-turbidity observations represented the minority class, priority was given to metrics that reflect the model's ability to identify exceedance events while maintaining reliable overall classification. Therefore, the ranking considered the ROC-AUC, F1-score, precision, recall, specificity, and weighted F1-score. To calculate the overall ranking, each model was first evaluated using these metrics, and the metric values were then combined into a single overall score. This score represents the average multi-metric performance of each model, giving a more balanced comparison than any individual metric. Models with higher overall scores were ranked higher because they showed stronger and more consistent performance across the full set of evaluation criteria.

The ROC-AUC was used to assess the overall separability between low- and high-turbidity classes, while the F1-score was used to evaluate the balance between precision and recall for the high-turbidity class. Recall was considered important because missing high-turbidity events may reduce the usefulness of the model for early-warning applications, whereas precision and specificity were used to assess false-positive control. The weighted F1-score was included to account for the performance across both classes while considering their different sample sizes. When models showed similar classification performance, the selected-variable scenario was considered preferable because it achieved comparable accuracy with fewer predictors and lower redundancy.

3. Results

3.1. Descriptive Analysis and Class Distribution

The descriptive distribution of turbidity values indicates a strongly imbalanced binary classification problem, with most observations concentrated under low-turbidity conditions. As shown in Figure 2, 4408 observations, representing 85.3% of the dataset, were classified as low turbidity (≤ 2 NTU), whereas only 758 observations, representing 14.7%, exceeded the 2 NTU threshold and were classified as high turbidity. The histogram and density curve show that turbidity values are highly right-skewed, with a dominant concentration of observations below approximately 1 NTU and a smaller secondary cluster above the exceedance threshold. This pattern suggests that high-turbidity events occurred less frequently but formed a distinct minority class that is important for water-quality warning and classification purposes. Therefore, the observed class imbalance justified the use of stratified data splitting and imbalance-treatment strategies during model development to ensure that the minority high-turbidity class was adequately represented during training and evaluation. The final dataset used for modeling included 5166 valid observations after quality control screening and temporal matching with meteorological variables.

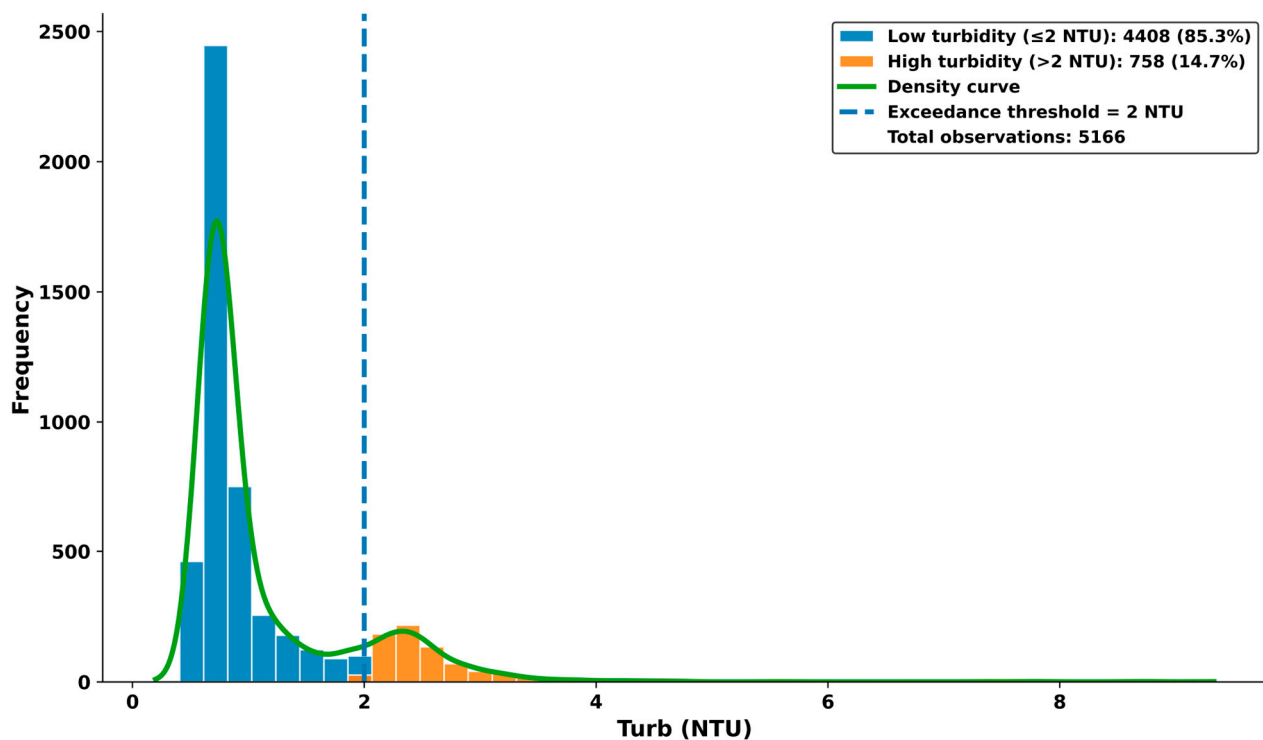


Figure 2. Distribution of turbidity observations and binary class separation using the 2 NTU exceedance threshold. Blue bars represent low-turbidity observations (≤ 2 NTU), orange bars represent high-turbidity observations (> 2 NTU), and the green curve shows the density distribution.

3.2. Predictor Correlation and Variable Selection

As shown in Figure 3, the predictor correlation analysis revealed several strong associations among the water quality and meteorological variables, indicating the presence of potential redundancy within the full predictor set. Among the water quality variables, conductivity and chloride showed a very strong positive correlation ($r = 0.99$), suggesting that both variables represented closely related ionic-strength information. Temperature-related meteorological variables also showed strong intercorrelations, including air temperature with surface temperature ($r = 0.97$), wet-bulb temperature ($r = 0.92$), and longwave radiation ($r = 0.88$). In addition, dew-point temperature and specific humidity were strongly correlated ($r = 0.98$), while surface pressure showed strong negative correlations with air temperature, wet-bulb temperature, longwave radiation, and surface temperature. Although these correlations indicated possible multicollinearity, variables were not removed based on correlation alone because nonlinear classifiers may still benefit from partially redundant predictors.

Therefore, the correlation results were interpreted together with the RF feature-importance rankings shown in Figure 4. The feature-importance analysis indicated that pH was the most influential predictor for turbidity classification, followed by alkalinity, short-wave radiation, wind speed, conductivity, chlorine, and chloride. This ranking shows that both internal water quality conditions and selected meteorological drivers contributed to distinguishing low- and high-turbidity observations. Based on the combined interpretation of Figures 3 and 4, variables with strong redundancy, low feature importance, and limited additional physical contribution were excluded, whereas variables with high importance or clear relevance to turbidity processes were retained.

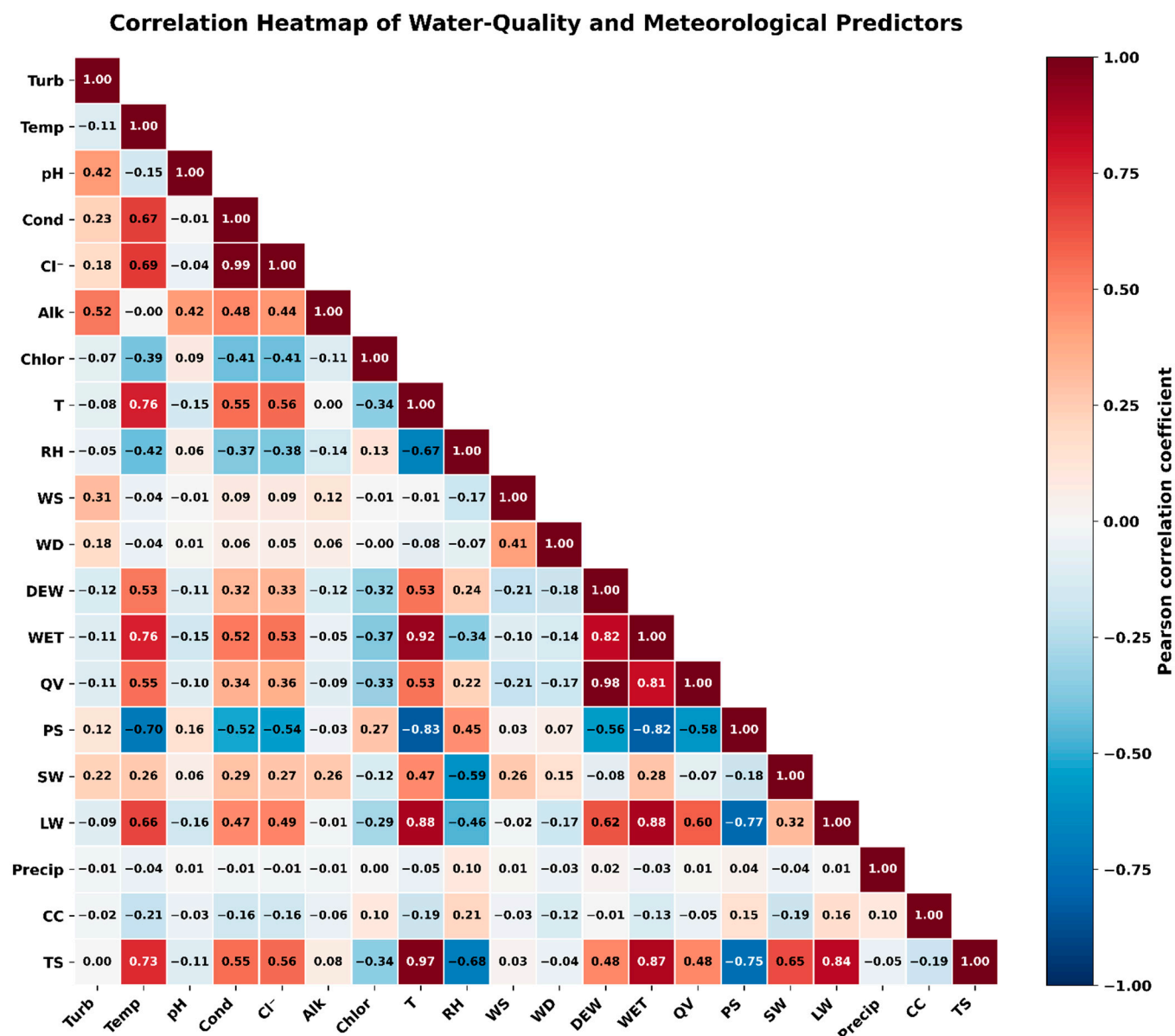


Figure 3. Correlation heatmap of water quality and meteorological predictors used for turbidity classification.

Accordingly, the selected meteorological-only scenario included T, RH, WS, WD, QV, SW, Precip, and CC, whereas the selected combined scenario included Temp, pH, Cond, Alk, Chlor, T, RH, WS, WD, DEW, SW, Precip, and CC. These selected predictor sets were used in the subsequent classification modeling to reduce redundancy while preserving the variables most relevant to turbidity exceedance prediction. In addition to the two selected-variable scenarios, the water-quality-only all-variable scenario (S3) and the combined all-variable scenario (S4) were also retained as sensitivity scenarios. Including S3 and S4 allowed the analysis to compare the selected predictor sets against broader predictor configurations and to determine whether variable selection improved classification performance relative to using all available water quality and meteorological variables.

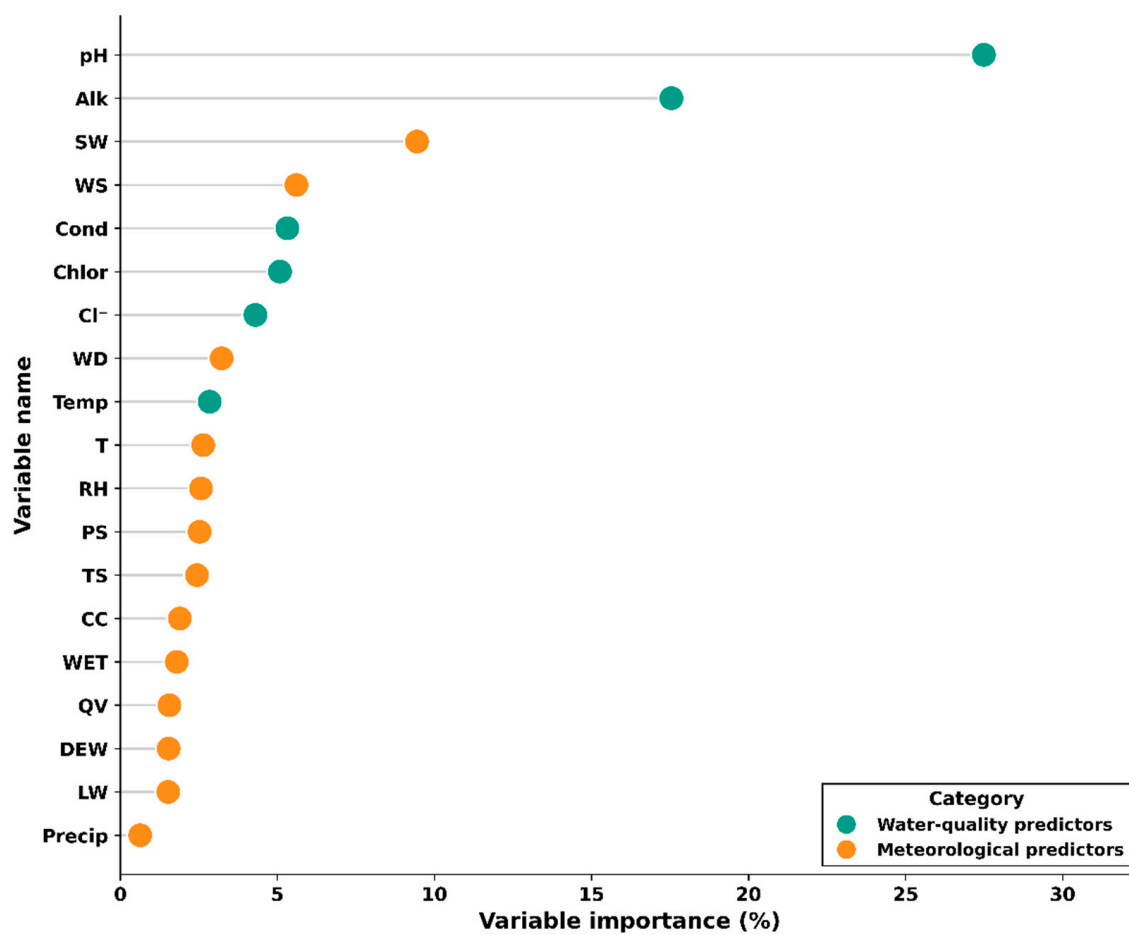


Figure 4. RF feature-importance ranking of water quality and meteorological predictors.

3.3. Model Performance

3.3.1. Meteorological-Only Model Using Selected Variables

The meteorological-only selected-variable scenario demonstrated moderate ability to classify turbidity exceedance under the three class-balancing strategies: undersampling, oversampling, and SMOTE (Figure 5). Overall, the ROC-AUC values ranged from 0.81 to 0.86 across the evaluated classifiers, indicating that selected meteorological predictors contained useful but incomplete information for distinguishing low- and high-turbidity conditions. Under undersampling, the models achieved high recall values, particularly the SVM (0.91), showing strong sensitivity to high-turbidity events. However, this was accompanied by low precision values of approximately 0.31–0.32, suggesting that many low-turbidity observations were incorrectly classified as high turbidity. Oversampling provided a more balanced performance, with RF achieving the highest accuracy (0.85), specificity (0.91), weighted F1-score (0.85), and ROC-AUC (0.85) among the oversampling models. SMOTE also produced stable results, where RF and XGBoost achieved the highest ROC-AUC value of 0.86 and weighted F1-scores of 0.83 and 0.81, respectively. Across all sampling strategies, precision remained relatively lower than recall and specificity, highlighting the difficulty of accurately identifying the minority high-turbidity class using meteorological variables alone. These findings suggest that selected meteorological variables can support early-warning classification, but their limited precision indicates that meteorological forcing alone does not fully capture the internal water quality processes controlling turbidity exceedance.

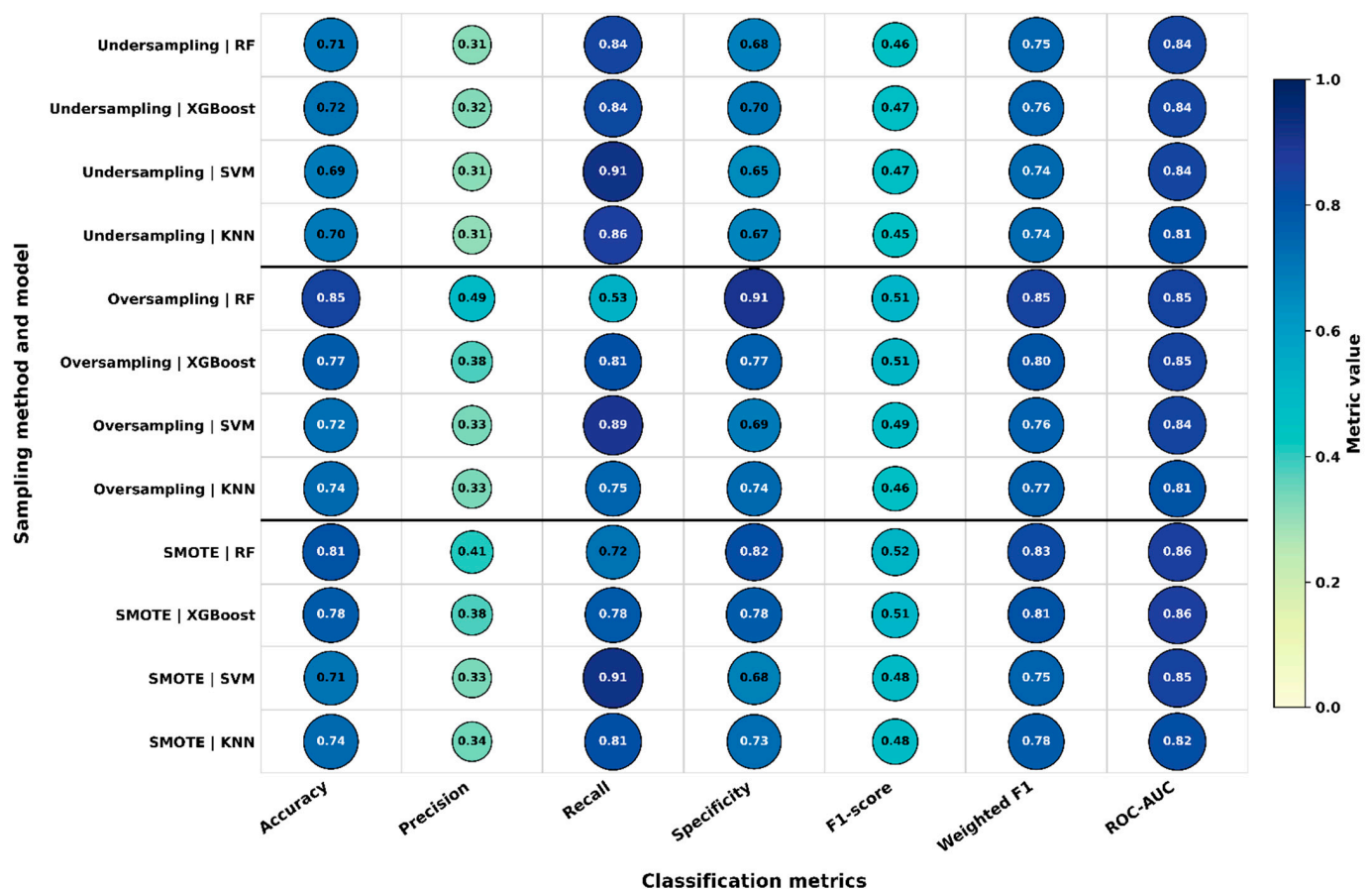


Figure 5. Performance of meteorological-only selected-variable models under different sampling strategies.

3.3.2. Combined Model Using Selected Variables

The combined selected-variable scenario showed a clear improvement in turbidity classification performance compared with the meteorological-only scenario, as shown by the higher metric values in Figure 6 relative to Figure 5. Across all sampling strategies, the models achieved high accuracy values of 0.89–0.95 and ROC-AUC values of 0.96–0.98, indicating strong discrimination between low- and high-turbidity observations. Under undersampling, recall remained high for all classifiers (0.88–0.91), but precision was lower (0.59–0.63), suggesting a greater tendency toward false-positive predictions. Oversampling produced the most balanced results, with RF achieving the highest accuracy (0.95), precision (0.83), specificity (0.97), weighted F1-score (0.95), and ROC-AUC (0.98). SMOTE also provided stable performance, with both RF and XGBoost achieving an ROC-AUC value of 0.97 and weighted F1-score of 0.94. The higher precision achieved after adding water quality variables indicates better identification of actual high-turbidity events with fewer false alarms. Overall, these results confirm that adding water quality variables improved model reliability and provided a better-balanced classification of turbidity exceedance than meteorological predictors alone.

3.3.3. Water-Quality-Only Model Using All Variables

The water-quality-only all-variable scenario achieved strong and consistent classification performance across the three sampling strategies, confirming that measured water quality variables alone provided substantial information for identifying turbidity exceedance, as shown in Figure 7. Overall, the models produced high accuracy values ranging from 0.90 to 0.95 and ROC-AUC values ranging from 0.95 to 0.96, indicating strong separa-

tion between low- and high-turbidity observations. Under undersampling, all classifiers showed balanced performance, with accuracy close to 0.91–0.92, recall between 0.86 and 0.89, and specificity between 0.91 and 0.93. Oversampling improved precision and the F1-score, particularly for RF and XGBoost, where both models achieved an F1-score of 0.81. SMOTE produced the best overall result for XGBoost, with the highest accuracy, precision, specificity, and weighted F1-score and an ROC-AUC of 0.95. Compared with the meteorological-only scenario, the higher precision and F1-score values indicate that water quality predictors provided more direct information for reducing false alarms and improving classification balance. These findings show that water quality variables are highly effective for turbidity exceedance prediction, even without meteorological predictors.

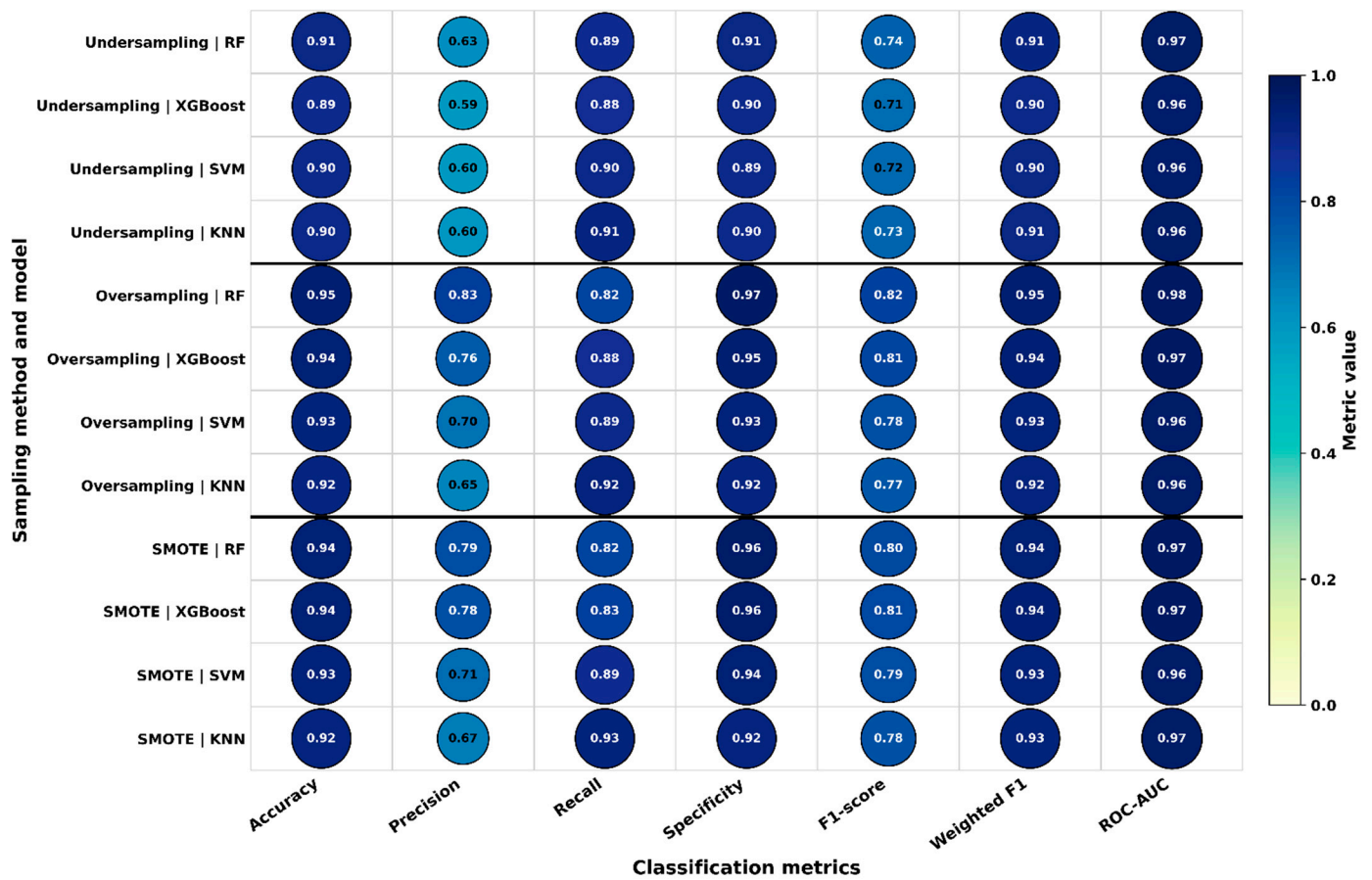


Figure 6. Classification performance of the combined selected-variable scenario across sampling strategies.

3.3.4. Combined Model Using All Variables

The combined all-variable scenario also achieved strong classification performance, as shown in Figure 8, confirming that the full set of water quality and meteorological predictors provided robust information for turbidity exceedance prediction. Overall, accuracy ranged from 0.89 to 0.95, and the ROC-AUC ranged from 0.96 to 0.98, indicating very good discrimination between low- and high-turbidity observations. Under undersampling, recall remained high at 0.88–0.91, although precision was lower at 0.58–0.67, indicating a greater tendency toward false-positive predictions. Oversampling produced the most balanced performance, with RF achieving the highest precision and F1-score of 0.83 and the highest ROC-AUC of 0.98, while XGBoost and the SVM also showed strong and stable results. SMOTE likewise yielded consistently high performance, with XGBoost reaching an accuracy of 0.95 and both RF, XGBoost, and KNN achieving ROC-AUC values of 0.98.

Compared with the water-quality-only scenario, the inclusion of meteorological variables provided only a modest overall improvement, suggesting that water quality predictors carried most of the predictive signal, while meteorological inputs offered supplementary information. These findings indicate that the combined all-variable scenario was highly effective, but its performance was broadly comparable to that of the combined selected-variable scenario, showing that variable selection reduced redundancy without sacrificing predictive skill.

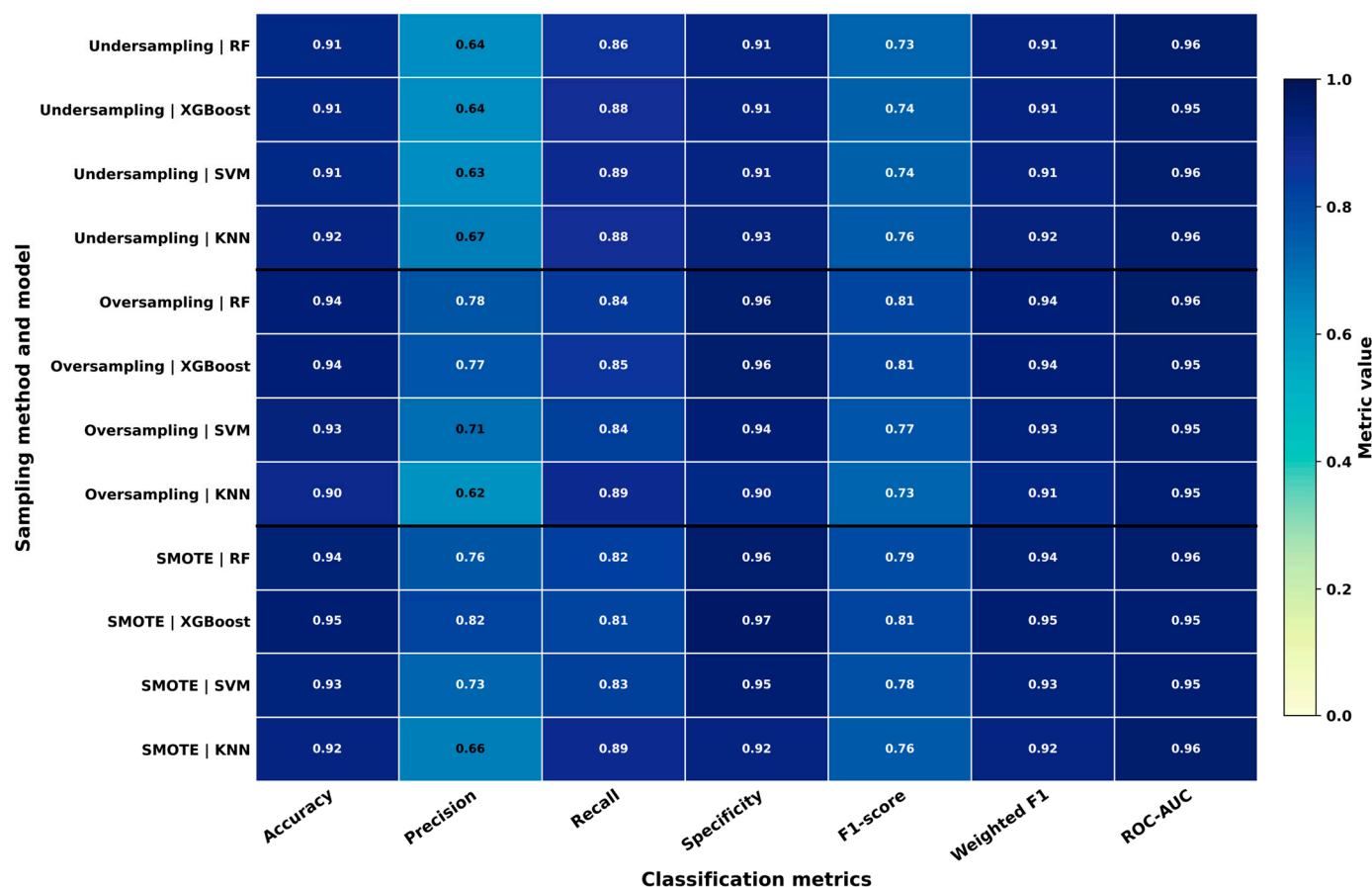


Figure 7. Classification performance of water-quality-only all-variable models across sampling strategies.

3.4. Ranking of Scenario-Based Model Performance

The scenario-based models were ranked using a multi-metric comparison that emphasized balanced detection of the minority high-turbidity class rather than accuracy alone. As shown in Table 3, the ranking considered the ROC-AUC, F1-score, precision, recall, specificity, and weighted F1-score, and these metrics were summarized using an overall score. The top-ranked models were mainly from the combined-predictor scenarios, indicating that integrating water quality and meteorological variables improved turbidity exceedance classification. The highest-ranked model was XGBoost under the combined all-variable scenario with oversampling, which achieved the highest overall score of 0.905, along with strong accuracy, high recall, a high F1-score, and an ROC-AUC of 0.973. Although RF under the same scenario showed slightly higher precision, specificity, and ROC-AUC, it was ranked second with an overall score of 0.904 because its recall was lower, meaning that it missed more high-turbidity observations. The third-ranked model was XGBoost with SMOTE under the combined all-variable scenario, which achieved the highest ROC-AUC of 0.977 but ranked third with an overall score of 0.902. The fourth-ranked model

was RF under the combined selected-variable scenario with oversampling, also with an overall score of 0.902, showing performance close to the all-variable models while using fewer predictors.

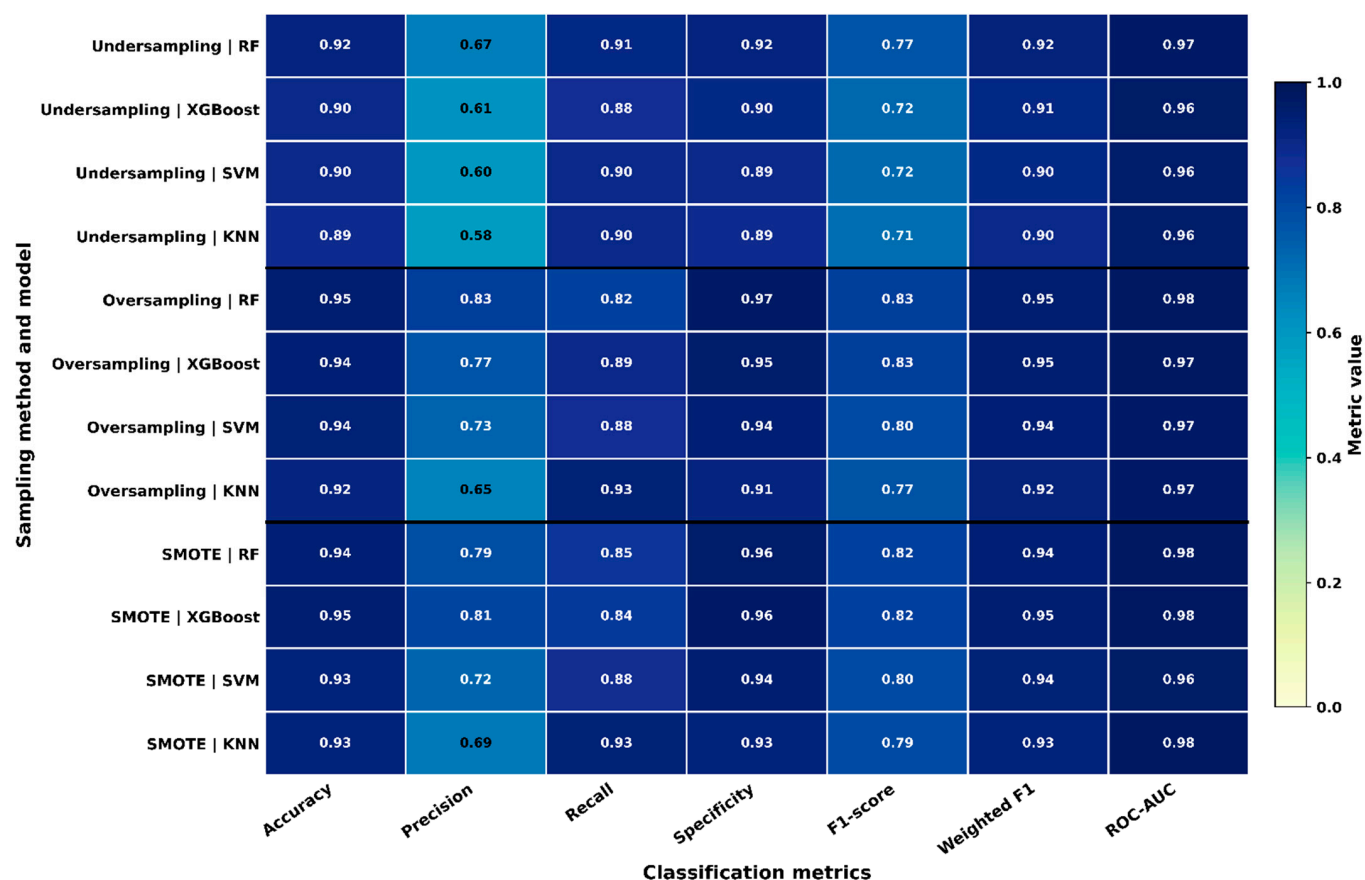


Figure 8. Classification performance of the combined all-variable scenario across sampling strategies.

Table 3. Ranking of the top-performing scenario-based turbidity classification models.

Rank	Scenario	Sampling/Model	Accuracy	Precision	Recall	Specificity	F1-Score	Weighted F1	ROC-AUC	Overall Score
1	S4	Oversampling/XGBoost	0.945	0.768	0.895	0.954	0.827	0.947	0.973	0.905
2	S4	Oversampling/RF	0.950	0.833	0.822	0.972	0.828	0.950	0.976	0.904
3	S4	SMOTE/XGBoost	0.947	0.805	0.842	0.965	0.823	0.947	0.977	0.902
4	S2	Oversampling/RF	0.949	0.832	0.816	0.972	0.824	0.949	0.976	0.902

The confusion matrices in Figure 9 provide a more detailed explanation of the ranking by showing how each top model balanced correct high-turbidity detection against false alarms. The first-ranked model correctly detected the largest number of high-turbidity observations, with 136 true positives and only 16 false negatives. This explains why it obtained the highest overall score in Table 3 despite not having the highest precision among the four models, as the main objective was to improve the detection of exceedance events. The second-ranked model produced fewer false positives, with only 25 low-turbidity observations incorrectly classified as high turbidity, but it missed more high-turbidity cases than the first-ranked model. The third-ranked model showed strong overall discrimination and the highest ROC-AUC, but its slightly lower recall and higher number of missed high-turbidity cases reduced its overall rank. The fourth-ranked model had a confusion-matrix pattern very close to the second-ranked model, indicating that the combined selected-variable scenario retained nearly the same predictive capability as the combined all-variable scenario while using fewer predictors.

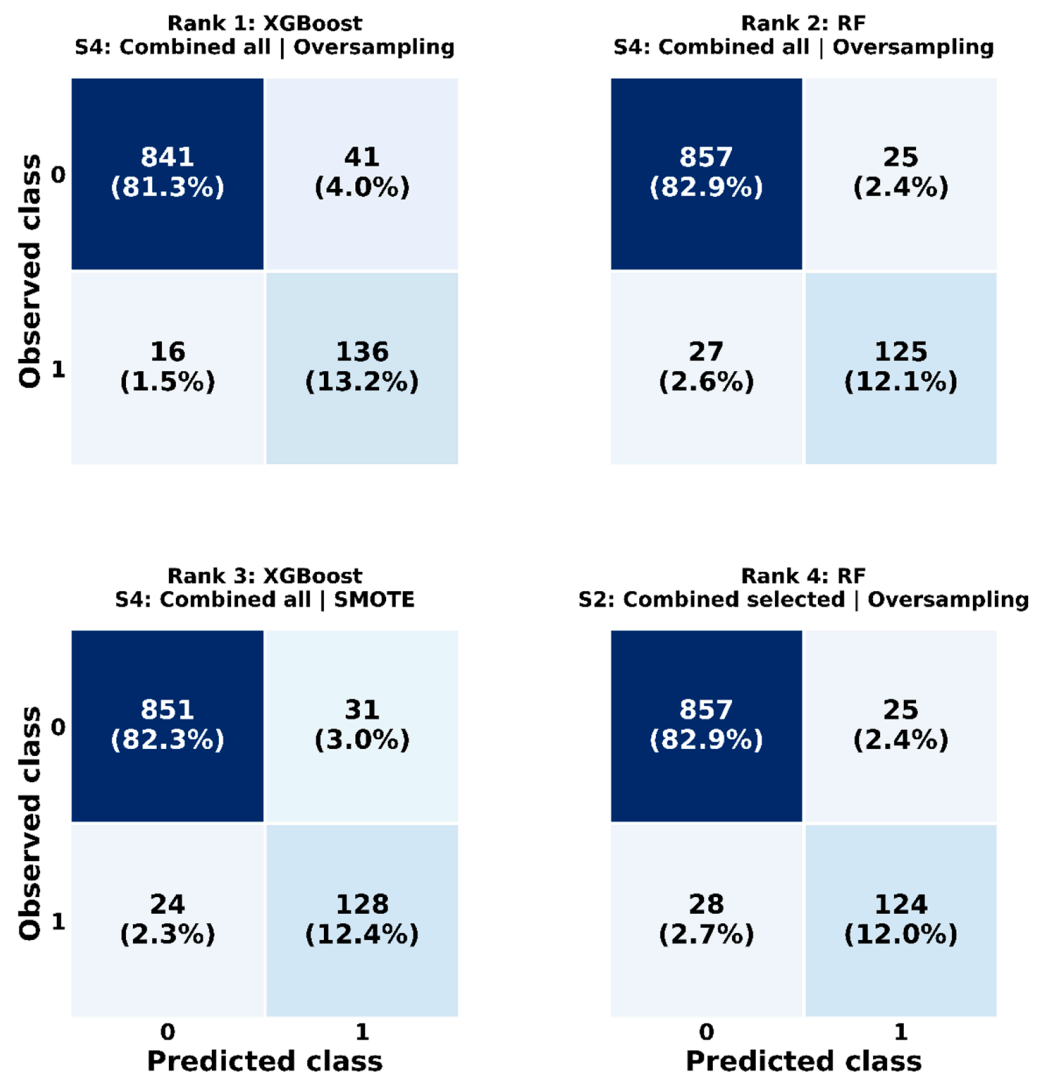


Figure 9. Confusion-matrix comparison of the four top-ranked turbidity classification models.

Figure 10 further confirms that all four top-ranked models had excellent class separation ability, with ROC curves well above the random-classifier line and AUC values close to 0.98. The small differences among the ROC curves indicate that all selected models had strong discrimination ability; however, the final ranking followed the overall score reported in Table 3, which reflects the combined balance between recall, precision, F1-score, specificity, weighted F1-score, and ROC-AUC. Overall, oversampling-based XGBoost under the combined all-variable scenario was selected as the best model because it achieved the highest overall score (0.905) and provided the strongest balance between high-turbidity detection, classification reliability, and overall discrimination.

3.5. Optimized Hyperparameters of the Ranked Models

The optimized hyperparameters of the four highest-ranked classification models are summarized in Table 4. These tuned settings were used to train the final ranked models and ensure that the comparison was based on optimized configurations rather than default parameters. For example, the first-ranked XGBoost with oversampling used 100 estimators, a maximum depth of 5, a learning rate of 0.10, a subsampling ratio of 0.80, a column-sampling ratio of 0.90, and a minimum child weight of 3, reflecting a moderately complex model with controlled tree growth. In comparison, the second-ranked RF with oversampling used 200 estimators, no maximum-depth restriction, log2 as the maximum-feature criterion,

5 minimum samples for splitting, and 1 minimum sample per leaf, allowing the ensemble to capture nonlinear relationships while reducing excessive feature use at each split. Overall, reporting these hyperparameters improves the reproducibility of the classification analysis and supports the reliability of the final scenario-based model ranking.

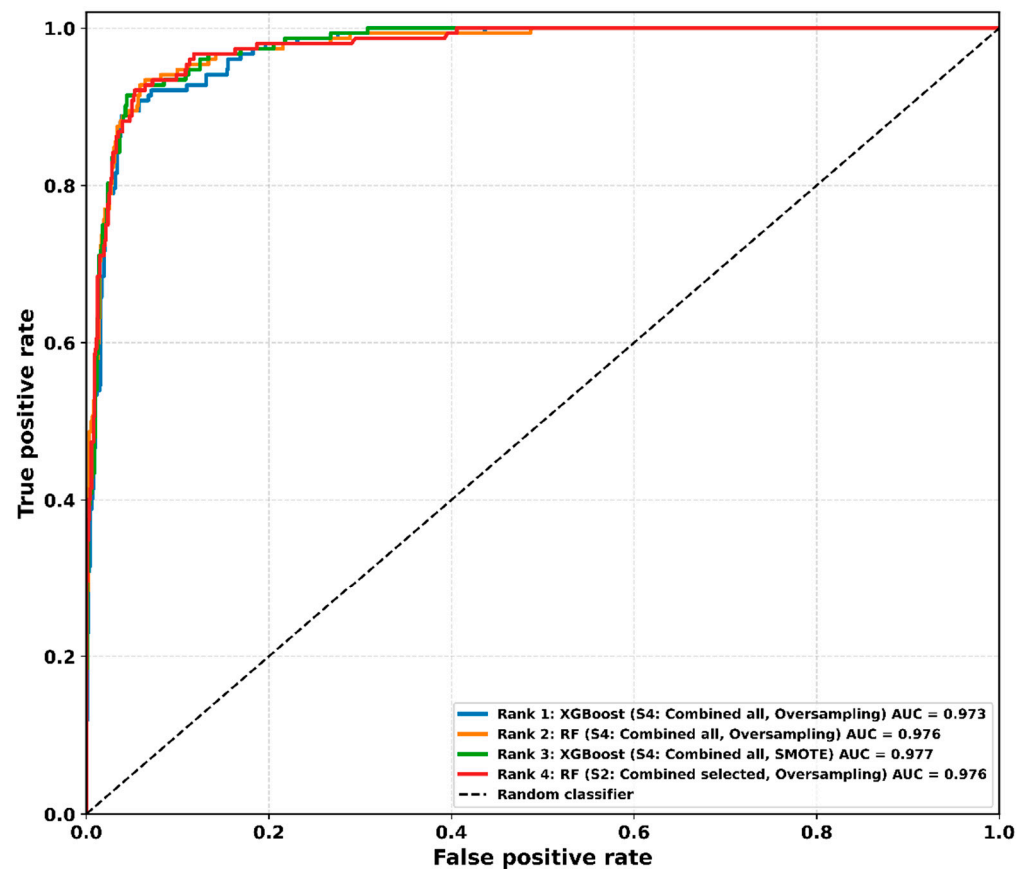


Figure 10. ROC curves of the four highest-ranked turbidity classification models.

Table 4. Optimized hyperparameters of the four highest-ranked classification models.

Rank	Scenario	Sampling/Model	Optimized Parameters
1	S4	Oversampling/XGBoost	n_estimators = 100, max_depth = 5, learning_rate = 0.10, subsample = 0.80, colsample_bytree = 0.90, and min_child_weight = 3.
2	S4	Oversampling/RF	n_estimators = 200, max_depth = None, max_features = log2, min_samples_split = 5, and min_samples_leaf = 1.
3	S4	SMOTE/XGBoost	n_estimators = 300, max_depth = 5, learning_rate = 0.10, subsample = 0.90, colsample_bytree = 1.00, and min_child_weight = 1.
4	S2	Oversampling/RF	n_estimators = 300, max_depth = 15, max_features = log2, min_samples_split = 2, and min_samples_leaf = 1.

3.6. Computational-Time Comparison

The overall computational-time comparison of the tuned classification models is shown in Figure 11. Clear differences were observed between training and prediction time across both model types and scenarios. At the model level, the SVM required the longest average training time, reaching about 2.2 s, followed by RF at approximately 0.4 s, while XGBoost and KNN showed much shorter training times. For example, KNN had almost

negligible training time because it mainly stores the training data rather than building a complex fitted model. However, its prediction time was higher than XGBoost because each new sample must be compared with stored observations during classification. At the scenario level, the meteorological-only selected-variable scenario required the highest average training time, exceeding 1.0 s, and also showed the highest prediction time of about 0.07 ms/sample. In contrast, the combined all-variable and water-quality-only scenarios remained computationally efficient, indicating that computational cost was influenced by both model structure and predictor interactions.

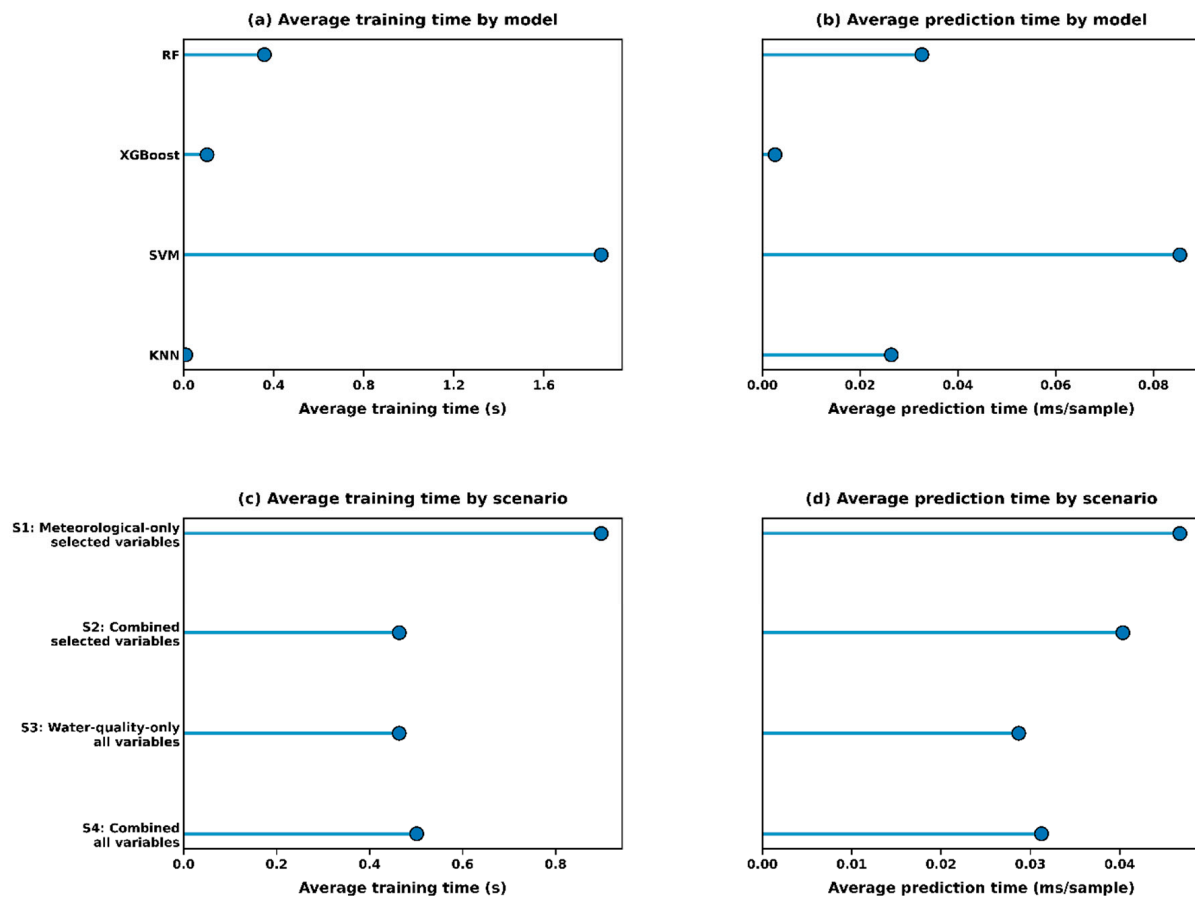


Figure 11. Computational-time comparison of the tuned classification models across model types and predictor scenarios, showing (a) average training time by model type, (b) average prediction time by model type, (c) average training time by predictor scenario, and (d) average prediction time by predictor scenario.

The computational-time comparison of the four highest-ranked classification models is presented in Figure 12. Among these models, the first-ranked XGBoost model under S4 with oversampling was the most efficient, requiring only about 0.08 s for training and approximately 0.003 ms/sample for prediction. This is important because the same model also achieved the highest overall score in Table 3, showing that strong classification performance was obtained with low computational cost. In comparison, the second-ranked RF model under S4 with oversampling required about 0.42 s for training and 0.028 ms/sample for prediction, while the third-ranked XGBoost model under S4 with SMOTE remained efficient, with about 0.21 s of training time and 0.003 ms/sample prediction time. The fourth-ranked RF model under S2 with oversampling had the highest cost among the top four models, requiring approximately 0.53 s for training and 0.039 ms/sample for prediction. Overall, these results support selecting the first-ranked XGBoost model because

it provided the best balance between classification performance, high-turbidity detection, and computational efficiency.

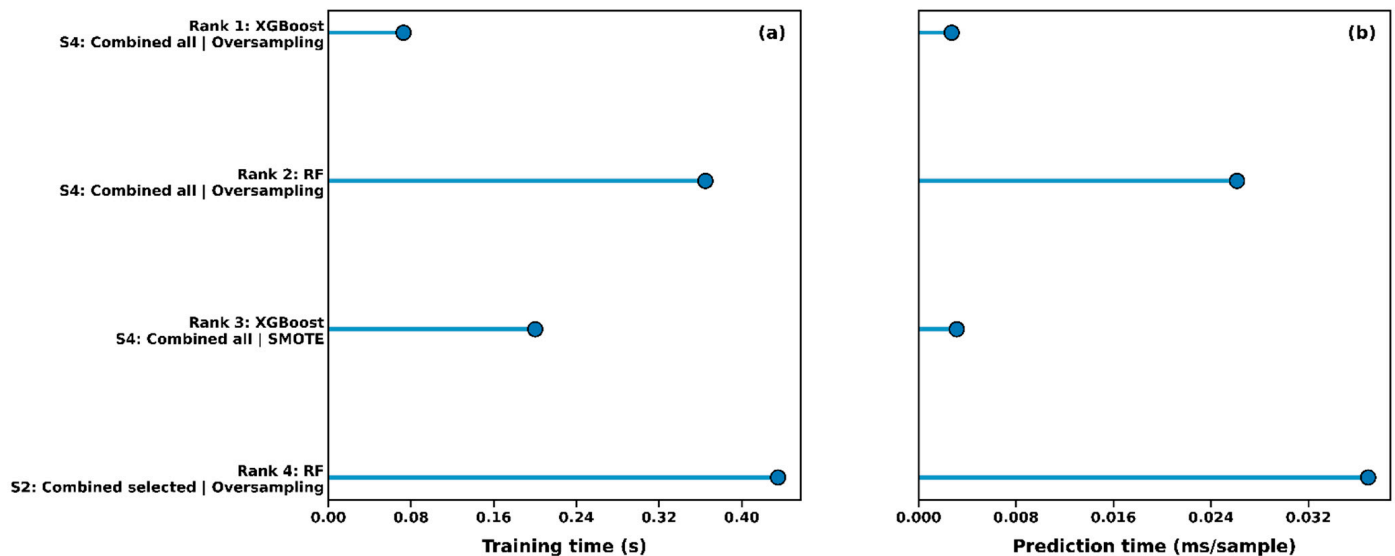


Figure 12. Computational efficiency of the four highest-ranked classification models using tuned hyperparameters, showing (a) training time and (b) prediction time for the top-ranked scenario–sampling–model combinations.

4. Discussion

4.1. Role of Water Quality and Meteorological Predictors in High-Turbidity Classification

The comparison among the evaluated predictor scenarios demonstrates that high-turbidity exceedance classification depends on both external meteorological forcing and internal water quality conditions. The meteorological-only selected-variable scenario showed moderate classification ability, with ROC-AUC values ranging from 0.81 to 0.86, indicating that atmospheric and surface-weather variables contain useful information for identifying turbidity exceedance. This result is physically reasonable because wind speed, wind direction, precipitation, radiation, and temperature-related variables can influence coastal mixing, runoff inputs, sediment resuspension, and surface-water conditions [10–14]. However, the relatively low precision observed in the meteorological-only scenario suggests that meteorological variables alone were not sufficient to distinguish actual high-turbidity events from normal conditions with high reliability. This indicates that weather forcing can support early-warning classification, but it does not fully represent the internal physicochemical processes controlling turbidity variability.

Adding water quality predictors substantially improved classification performance, confirming the importance of direct seawater measurements for high-turbidity detection. In the combined selected-variable scenario, the ROC-AUC increased to approximately 0.96–0.98, and the models achieved higher precision, specificity, F1-score, and weighted F1-score compared with the meteorological-only scenario. This improvement suggests that variables such as pH, alkalinity, conductivity, chlorine, chloride, and water temperature provided more direct information on the conditions associated with turbidity exceedance. The feature-importance analysis also supported this interpretation, as pH and alkalinity were among the most influential predictors, followed by shortwave radiation, wind speed, conductivity, chlorine, and chloride. These results are consistent with previous studies showing that machine learning models can capture nonlinear interactions among physicochemical, hydrological, and meteorological predictors in water quality applications [19–21,23,24].

The strong performance of the water-quality-only and combined scenarios indicates that internal water quality variables were the main contributors to exceedance classification, while meteorological variables provided complementary information. This supports an integrated monitoring framework in which meteorological predictors represent external forcing and water quality predictors capture the immediate physicochemical state of seawater. Operationally, the combined selected-variable scenario is useful because it achieved strong classification performance while reducing predictor redundancy. Therefore, combining selected water quality and meteorological predictors provides a more reliable basis for detecting high-turbidity events and reducing false alarms.

4.2. Influence of Predictor Selection and Redundancy

Predictor selection was important in this study because the full water quality and meteorological dataset contained several strongly related variables that could introduce redundancy without necessarily improving classification performance. The correlation analysis showed strong associations among conductivity and chloride, as well as among temperature-related and humidity-related meteorological variables, indicating that some predictors represented overlapping physical or atmospheric information. In environmental modeling, such redundancy is common because many monitored variables respond to shared forcing processes or describe similar system states [35,37,58]. However, correlated variables should not be removed automatically, especially in nonlinear machine learning models, because they may still contain complementary information useful for classification [25,26,59].

The combined use of correlation analysis and feature-importance ranking provided a more balanced basis for variable selection than relying on correlation alone. In this study, the RF feature-importance results identified pH, alkalinity, shortwave radiation, wind speed, conductivity, chlorine, and chloride as influential predictors, showing that both water quality conditions and selected meteorological drivers contributed to separating low- and high-turbidity observations. This approach is consistent with feature-selection principles, where predictor reduction should consider statistical redundancy, predictive contribution, model interpretability, and physical relevance [36,60]. Therefore, the selected-variable scenarios were designed to reduce unnecessary predictor complexity while retaining variables that were physically meaningful and useful for turbidity exceedance classification.

The comparison between selected-variable and all-variable scenarios further showed that retaining all available predictors did not necessarily provide a clear practical advantage. Although the all-variable scenarios achieved strong performance, the selected combined scenario provided similarly high classification ability with fewer predictors and lower redundancy. This is important for operational early-warning systems because simpler models are easier to interpret, maintain, and implement when sensor availability, data quality, and computational efficiency are considered [37,59]. Overall, the results indicate that predictor selection improved the practical usefulness of the classification framework by preserving the most informative water quality and meteorological variables while avoiding unnecessary duplication in the predictor set. This interpretation is consistent with your correlation and feature-importance results, where selected variables were retained based on both statistical and physical relevance.

4.3. Effect of Class-Imbalance Treatment on Minority Exceedance Detection

Class imbalance strongly influenced model interpretation because high-turbidity observations represented only 14.7% of the dataset, compared with 85.3% low-turbidity observations. Under such conditions, accuracy alone can be misleading because a model may appear to perform well by correctly classifying the majority class while still missing

many minority exceedance events [18,29]. Therefore, recall, precision, F1-score, specificity, and ROC-AUC were more appropriate for evaluating model performance. Recall was particularly important because it measured the ability to detect actual high-turbidity events, whereas precision and specificity helped assess whether the model produced excessive false alarms [17,18].

The imbalance-treatment strategy affected the trade-off between detecting high-turbidity events and controlling false positives. Undersampling generally improved recall, indicating greater sensitivity to exceedance events, but it also reduced precision, meaning that more normal observations were incorrectly classified as high turbidity. This behavior is expected because undersampling reduces the dominance of the majority class but may also remove useful low-turbidity information needed to define normal conditions [30,31]. In contrast, oversampling produced more balanced performance in the combined selected-variable scenario, particularly for RF, which achieved strong accuracy, precision, specificity, weighted F1-score, and ROC-AUC. This suggests that increasing the representation of high-turbidity observations during training helped the models learn the minority class more effectively while maintaining good separation between the two classes.

SMOTE also provided stable classification performance by generating synthetic minority-class samples instead of simply duplicating existing high-turbidity observations [53]. This can help expand the minority-class decision region and improve learning when exceedance observations are limited. However, SMOTE is not always superior because its performance depends on class overlap, noise, and the structure of minority observations [33,61]. Overall, the results indicate that imbalance treatment was necessary for improving minority exceedance detection, but the most suitable method depended on the balance between recall and false-alarm control. In the strongest scenarios, oversampling and SMOTE provided a better overall balance than undersampling, supporting their use for high-turbidity early-warning classification.

4.4. Model Ranking, Practical Efficiency, and Early-Warning Implications

Model ranking in this study was based on multiple performance metrics rather than accuracy alone because high-turbidity observations represented the minority class. In imbalanced classification problems, accuracy can overestimate model reliability when the majority class dominates the dataset [18,29]. Therefore, the ROC-AUC, F1-score, precision, recall, specificity, and weighted F1-score were considered together to evaluate both discrimination ability and the balance between exceedance detection and false-alarm control. The ROC-AUC was particularly useful for assessing the ability of each model to separate low- and high-turbidity classes across thresholds, while the F1-score summarized the balance between precision and recall for the high-turbidity class [17,18].

Across the evaluated scenarios, the strongest models were generally obtained when water quality predictors were included, especially in the combined selected-variable and combined all-variable scenarios. This indicates that direct physicochemical measurements improved the ability of the classifiers to identify high-turbidity exceedance compared with meteorological predictors alone. Among the evaluated classifiers, RF and XGBoost showed strong and stable performance because tree-based ensemble methods can capture nonlinear relationships and interactions among predictors [25,26]. However, model selection should also consider practical efficiency. A model with slightly higher accuracy may not always be preferable if it requires more predictors, has greater computational cost, or provides limited interpretability for operational monitoring.

From an early-warning perspective, the preferred model should provide high recall without producing excessive false alarms. High recall is important because missing true high-turbidity events may reduce the usefulness of the warning system, whereas high

precision and specificity are needed to avoid unnecessary operational responses. The selected-variable scenarios are therefore practically valuable because they reduce predictor redundancy while maintaining strong classification performance. This makes the framework more suitable for real-time or near-real-time coastal monitoring, where rapid prediction, sensor availability, and simple model interpretation are important. Overall, the ranking results support the use of a balanced model selection strategy that combines predictive performance, computational efficiency, and operational reliability for high-turbidity early-warning applications.

4.5. Limitations and Future Research

Although the proposed framework provided strong classification performance, several limitations should be acknowledged. First, the dataset was obtained from a single automated seawater monitoring station along the eastern coast of Saudi Arabia, and the exact site coordinates were not reported because of operational confidentiality. Therefore, the results reflect the monitored coastal conditions and should be validated at additional sites before broader regional generalization. Second, the final dataset consisted of 5166 valid observations after quality control screening and temporal matching, but the record was irregular rather than a fully continuous time series. This may limit the ability to fully capture event persistence, seasonal patterns, and temporal dependence in turbidity exceedance behavior. Future studies should therefore test time-aware validation strategies and compare them with random stratified splitting, particularly because temporal and spatial dependence can affect model generalization in environmental datasets [62].

Another limitation is that the present modeling framework focused on water quality and NASA POWER meteorological predictors, while other potentially important coastal drivers were not included. Variables such as tides, currents, waves, bathymetry, coastal circulation, and nearby discharge or operational activities may influence sediment resuspension and turbidity dynamics, especially in shallow or semi-enclosed coastal environments. Future research should integrate hydrodynamic and oceanographic predictors with water quality and meteorological data to improve the physical representation of turbidity exceedance mechanisms. In addition, although NASA POWER variables provide useful meteorological information in data-limited settings, future studies should compare them with local in situ meteorological measurements when available to evaluate whether local observations improve classification reliability [12,38].

Although Random Forest and XGBoost are ensemble learning algorithms, higher-level ensemble approaches such as voting and stacking were not evaluated. Future research should examine whether combining classifiers through these approaches can further improve predictive stability and performance. Future research should also examine model explainability and transferability. While RF and XGBoost showed strong performance, operational early-warning systems require not only high accuracy but also transparent interpretation of why exceedance warnings are produced. Explainable machine learning methods such as local explanation approaches and feature attribution techniques could help identify the conditions leading to high-turbidity classifications and improve user confidence in model outputs [30]. In addition, future work could test hybrid modeling frameworks that combine data-driven learning with process knowledge because purely empirical models may perform less reliably under conditions not represented in the training data [63,64]. Overall, future studies should focus on multi-site validation, time-aware testing, integration of hydrodynamic predictors, and explainable modeling to improve the robustness and operational applicability of high-turbidity early-warning systems.

Although the selected models represented different learning strategies, other recent algorithms, such as LightGBM, CatBoost, and neural networks, were not evaluated. There-

fore, the best-performing model should be interpreted as the best among the models examined in this study rather than as universally optimal. Future research should compare these additional algorithms to determine whether they provide further improvements in predictive performance and stability.

5. Conclusions

This study developed a scenario-based machine learning framework for classifying high-turbidity exceedance using integrated water quality and meteorological predictors from an automated coastal seawater monitoring station in the Arabian Gulf. After quality control screening and temporal matching with NASA POWER meteorological variables, 5166 valid observations were retained from 9644 original records. Using a 2 NTU turbidity threshold, the dataset was converted into a binary classification problem, with high-turbidity exceedance representing 14.7% of the observations.

The results showed that meteorological predictors alone provided useful but incomplete information for high-turbidity classification, with the meteorological-only selected-variable scenario achieving ROC-AUC values of 0.81–0.86. In contrast, adding water quality variables substantially improved classification performance, with the combined selected-variable scenario achieving ROC-AUC values of 0.96–0.98. Overall, water quality variables carried most of the predictive signal, while meteorological variables provided complementary information. The water-quality-only all-variable scenario also performed strongly, whereas the selected combined scenario maintained high performance while reducing predictor redundancy.

Class-imbalance treatment influenced the balance between detecting high-turbidity events and controlling false alarms. Undersampling improved recall but reduced precision, whereas oversampling and SMOTE produced more balanced performance. Overall, the findings demonstrate that combining selected water quality and meteorological predictors provides an accurate and practical approach for high-turbidity early-warning classification. Future research should validate the framework at additional coastal sites, apply time-aware validation strategies, compare NASA POWER data with local meteorological observations, and include hydrodynamic variables such as tides, currents, waves, and coastal circulation.

Author Contributions: S.H.A.: formal analysis, investigation, and writing—original draft. Y.A.: writing—reviewing and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ongoing Research Funding program (ORF-2026-1807), King Saud University, Riyadh, Saudi Arabia.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are available from the authors upon reasonable request due to governmental privacy and data-sharing regulations imposed by the original data-providing agencies.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Bilotta, G.S.; Brazier, R.E. Understanding the Influence of Suspended Solids on Water Quality and Aquatic Biota. *Water Res.* **2008**, *42*, 2849–2861. [[CrossRef](#)] [[PubMed](#)]
2. Davies-Colley, R.J.; Smith, D.G. Turbidity Suspended Sediment, and Water Clarity: A Review. *JAWRA J. Am. Water Resour. Assoc.* **2001**, *37*, 1085–1101. [[CrossRef](#)]
3. Gholizadeh, M.H.; Melesse, A.M.; Reddi, L. A Comprehensive Review on Water Quality Parameters Estimation Using Remote Sensing Techniques. *Sensors* **2016**, *16*, 1298. [[CrossRef](#)] [[PubMed](#)]

4. Behmel, S.; Damour, M.; Ludwig, R.; Rodriguez, M.J. Water Quality Monitoring Strategies—A Review and Future Perspectives. *Sci. Total Environ.* **2016**, *571*, 1312–1329. [[CrossRef](#)] [[PubMed](#)]
5. Henley, W.F.; Patterson, M.A.; Neves, R.J.; Lemly, A.D. Effects of Sedimentation and Turbidity on Lotic Food Webs: A Concise Review for Natural Resource Managers. *Rev. Fish. Sci.* **2000**, *8*, 125–139. [[CrossRef](#)]
6. Betancourt, W.Q.; Rose, J.B. Drinking Water Treatment Processes for Removal of *Cryptosporidium* and *Giardia*. *Vet. Parasitol.* **2004**, *126*, 219–234. [[CrossRef](#)] [[PubMed](#)]
7. Hijnen, W.A.M.; Beerendonk, E.F.; Medema, G.J. Inactivation Credit of UV Radiation for Viruses, Bacteria and Protozoan (Oo)Cysts in Water: A Review. *Water Res.* **2006**, *40*, 3–22. [[CrossRef](#)] [[PubMed](#)]
8. Abushawish, A.; Bouaziz, I.; Almanassra, I.W.; AL-Rajabi, M.M.; Jaber, L.; Khalil, A.K.A.; Takriff, M.S.; Laoui, T.; Shanableh, A.; Atieh, M.A.; et al. Desalination Pretreatment Technologies: Current Status and Future Developments. *Water* **2023**, *15*, 1572. [[CrossRef](#)]
9. Macdonald, R.K.; Ridd, P.V.; Whinney, J.C.; Larcombe, P.; Neil, D.T. Towards Environmental Management of Water Turbidity within Open Coastal Waters of the Great Barrier Reef. *Mar. Pollut. Bull.* **2013**, *74*, 82–94. [[CrossRef](#)] [[PubMed](#)]
10. Deng, T.; Chau, K.-W.; Duan, H.-F. Machine Learning Based Marine Water Quality Prediction for Coastal Hydro-Environment Management. *J. Environ. Manag.* **2021**, *284*, 112051. [[CrossRef](#)] [[PubMed](#)]
11. Karthick, M.; Shanmugam, P.; Saravana Kumar, G. Long-Term Water Quality Assessment in Coastal and Inland Waters: An Ensemble Machine-Learning Approach Using Satellite Data. *Mar. Pollut. Bull.* **2024**, *209*, 117036. [[CrossRef](#)] [[PubMed](#)]
12. Sparks, A.H. Nasapower: A NASA POWER Global Meteorology, Surface Solar Energy and Climatology Data Client for R. *J. Open Source Softw.* **2018**, *3*, 1035. [[CrossRef](#)]
13. Kong, Y.; Jimenez, K.; Lee, C.M.; Winter, S.; Summers-Evans, J.; Cao, A.; Menczer, M.; Han, R.; Mills, C.; McCarthy, S.; et al. Monitoring Coastal Water Turbidity Using Sentinel2—A Case Study in Los Angeles. *Remote Sens.* **2025**, *17*, 201. [[CrossRef](#)]
14. Wu, Z.; Pang, J.; Li, J.; Wang, Y.; Ruan, J.; Zhang, X.; Yang, L.; Pang, Y.; Gao, Y. A Review of Remote Sensing-Based Water Quality Monitoring in Turbid Coastal Waters. *Intell. Mar. Technol. Syst.* **2025**, *3*, 24. [[CrossRef](#)]
15. Sun, Y.; Wang, D.; Li, L.; Ning, R.; Yu, S.; Gao, N. Application of Remote Sensing Technology in Water Quality Monitoring: From Traditional Approaches to Artificial Intelligence. *Water Res.* **2024**, *267*, 122546. [[CrossRef](#)] [[PubMed](#)]
16. Mohan, S.; Kumar, B.; Nejadhashemi, A.P. Integration of Machine Learning and Remote Sensing for Water Quality Monitoring and Prediction: A Review. *Sustainability* **2025**, *17*, 998. [[CrossRef](#)]
17. Fawcett, T. An Introduction to ROC Analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [[CrossRef](#)]
18. Saito, T.; Rehmsmeier, M. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLoS ONE* **2015**, *10*, e0118432. [[CrossRef](#)] [[PubMed](#)]
19. Dharmarathne, G.; Abekoon, A.M.S.R.; Bogahawaththa, M.; Alawatugoda, J.; Meddage, D.P.P. A Review of Machine Learning and Internet-of-Things on the Water Quality Assessment: Methods, Applications and Future Trends. *Results Eng.* **2025**, *26*, 105182. [[CrossRef](#)]
20. Yan, X.; Zhang, T.; Du, W.; Meng, Q.; Xu, X.; Zhao, X. A Comprehensive Review of Machine Learning for Water Quality Prediction over the Past Five Years. *J. Mar. Sci. Eng.* **2024**, *12*, 159. [[CrossRef](#)]
21. Zhu, M.; Wang, J.; Yang, X.; Zhang, Y.; Zhang, L.; Ren, H.; Wu, B.; Ye, L. A Review of the Application of Machine Learning in Water Quality Evaluation. *Eco-Environ. Health* **2022**, *1*, 107–116. [[CrossRef](#)] [[PubMed](#)]
22. Pang, Z.; Zhou, Z.; Fu, J.; Jiang, W.; Qin, X.; Sun, M. Deep Learning-Based Remote Sensing Retrieval of Inland Water Quality: A Review. *J. Hydrol. Reg. Stud.* **2025**, *61*, 102759. [[CrossRef](#)]
23. Devi, P.D.; Mamatha, G. Machine Learning Approach to Predict the Turbidity of Saki Lake, Telangana, India, Using Remote Sensing Data. *Meas. Sens.* **2024**, *33*, 101139. [[CrossRef](#)]
24. Kwarko-Kyei, J.; Tornyeviadzi, H.M.; Seidu, R. A Machine Learning Approach to Predicting the Turbidity from Filters in a Water Treatment Plant. *Water* **2025**, *17*, 2938. [[CrossRef](#)]
25. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
26. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016*; ACM: New York, NY, USA, 2016; pp. 785–794.
27. Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
28. Cover, T.; Hart, P. Nearest Neighbor Pattern Classification. *IEEE Trans. Inf. Theory* **1967**, *13*, 21–27. [[CrossRef](#)]
29. He, H.; Garcia, E.A. Learning from Imbalanced Data. *IEEE Trans. Knowl. Data Eng.* **2009**, *21*, 1263–1284. [[CrossRef](#)]
30. Branco, P.; Torgo, L.; Ribeiro, R.P. A Survey of Predictive Modeling on Imbalanced Domains. *ACM Comput. Surv.* **2016**, *49*, 31:1–31:50. [[CrossRef](#)]
31. Fernández, A.; García, S.; Galar, M.; Prati, R.C.; Krawczyk, B.; Herrera, F. *Learning from Imbalanced Data Sets*; Springer International Publishing: Cham, Switzerland, 2018; ISBN 978-3-319-98073-7.

32. Gelaro, R.; McCarty, W.; Suárez, M.J.; Todling, R.; Molod, A.; Takacs, L.; Randles, C.A.; Darmenov, A.; Bosilovich, M.G.; Reichle, R.; et al. The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *J. Clim.* **2017**, *30*, 5419–5454. [[CrossRef](#)] [[PubMed](#)]
33. Elreedy, D.; Atiya, A.F.; Kamalov, F. A Theoretical Distribution Analysis of Synthetic Minority Oversampling Technique (SMOTE) for Imbalanced Learning. *Mach. Learn.* **2024**, *113*, 4903–4923. [[CrossRef](#)]
34. Yang, Y.; Khorshidi, H.A.; Aickelin, U. A Review on Over-Sampling Techniques in Classification of Multi-Class Imbalanced Datasets: Insights for Medical Problems. *Front. Digit. Health* **2024**, *6*, 1430245. [[CrossRef](#)] [[PubMed](#)]
35. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*; Springer Series in Statistics; Springer: New York, NY, USA, 2009; ISBN 978-0-387-84857-0.
36. Kuhn, M. Building Predictive Models in R Using the Caret Package. *J. Stat. Softw.* **2008**, *28*, 1–26. [[CrossRef](#)]
37. Kuhn, M.; Johnson, K. *Applied Predictive Modeling*; Springer: New York, NY, USA, 2013; ISBN 978-1-4614-6848-6.
38. Monteiro, L.A.; Sentelhas, P.C.; Pedra, G.U. Assessment of NASA/POWER Satellite-Based Weather System for Brazilian Conditions and Its Impact on Sugarcane Yield Simulation. *Int. J. Climatol.* **2018**, *38*, 1571–1581. [[CrossRef](#)]
39. Rodrigues, G.C.; Braga, R.P. Evaluation of NASA POWER Reanalysis Products to Estimate Daily Weather Variables in a Hot Summer Mediterranean Climate. *Agronomy* **2021**, *11*, 1207. [[CrossRef](#)]
40. Painting, S.J.; Smith, A.J.; Khamis, A.S.; Abdulla, K.H.; Le Quesne, W.J.F.; Lyons, B.P.; Devlin, M.J.; Garcia, L. Development of Standards for Assessing Water Quality in Marine Coastal Waters of Bahrain. *Mar. Pollut. Bull.* **2023**, *196*, 115560. [[CrossRef](#)] [[PubMed](#)]
41. Xue, C.; Hu, C.; Cannizzaro, J.P.; Barnes, B.B.; Qi, L.; Shi, J.; Xie, Y.; Jaffe, B.D.; Palandro, D.A. Remotely Sensed Water Quality in Qatari Coastal Waters Between 2002 and 2022. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 16944–16960. [[CrossRef](#)]
42. Barkhordar, H.; Mohammadpour, G.; Hassanzadeh, S.; Karemi, H. Investigation of Water Quality in the Shallow Coastal Waters of the Persian Gulf. *Mar. Pollut. Bull.* **2024**, *209*, 117263. [[CrossRef](#)] [[PubMed](#)]
43. Albanai, J. Determining Seasonal Spatial and Temporal Distribution of Chlorophyll-a, and Its Concentration in Kuwait and the Arabian Gulf Using Satellite and In-Situ Data. *Trans. Marit. Sci.* **2024**, *13*. [[CrossRef](#)]
44. Ahmed, N.; Mohammad, A.; Knuteson, S.L.; Samara, F. Evaluating Temporal Changes in Water Quality Due to Urbanization: A Multi-Year Observational Study in Khalid Khor, Sharjah, UAE. *Front. Mar. Sci.* **2025**, *12*, 1538897. [[CrossRef](#)]
45. Al-Mur, B.A. Environmental Assessment Using Phytoplankton Diversity, Nutrients, Chlorophyll-a, and Trophic Status Along Southern Coast of Jeddah, Red Sea. *J. Mar. Sci. Eng.* **2024**, *13*, 29. [[CrossRef](#)]
46. Shafeeque, M.; Saleem, A.A.; Chaitanya, A.V.S.; Abdulla, C.P.; Lateef, S.; Albert, J. Chlorophyll-a Modulation in the Arabian Gulf Using Two Decades of Merged Ocean-Color Data. *Front. Mar. Sci.* **2025**, *12*, 1619821. [[CrossRef](#)]
47. Ibrahim, A.; Alkarbi, N.; Alsaadi, M.; Almazrouei, A.; Alshamsi, S.; Hamouda, M.A. Satellite-Based Machine Learning Models for Chlorophyll-a and TSS Retrieval in Abu Dhabi's Coastal Waters. *Front. Mar. Sci.* **2026**, *13*, 1787597. [[CrossRef](#)]
48. Algurainy, Y.; Refaat, A.; Alrehaili, O. Developing Insights into Pretreatment Optimization: Effects of Eliminating Lime and Soda Ash in Groundwater RO Desalination. *Water* **2025**, *17*, 3186. [[CrossRef](#)]
49. Alhathloul, S.H.; Algurainy, Y. Wavelet-Enhanced Machine Learning for Seawater Alkalinity Prediction in the Arabian Gulf Using Monitored Water-Quality Variables. *Water* **2026**, *18*, 1578. [[CrossRef](#)]
50. Wattelez, G.; Dupouy, C.; Lefèvre, J.; Ouillon, S.; Fernandez, J.-M.; Juillot, F. Application of the Support Vector Regression Method for Turbidity Assessment with MODIS on a Shallow Coral Reef Lagoon (Voh-Koné-Pouembout, New Caledonia). *Water* **2017**, *9*, 737. [[CrossRef](#)]
51. Petus, C.; Waterhouse, J.; Tracey, D.; Wolanski, E.; Brodie, J. Using Optical Water-Type Classification in Data-Poor Water Quality Assessment: A Case Study in the Torres Strait. *Remote Sens.* **2022**, *14*, 2212. [[CrossRef](#)]
52. Cooper, T.F.; Ridd, P.V.; Ulstrup, K.E.; Humphrey, C.; Slivkoff, M.; Fabricius, K.E. Temporal Dynamics in Coral Bioindicators for Water Quality on Coastal Coral Reefs of the Great Barrier Reef. *Mar. Freshw. Res.* **2008**, *59*, 703–716. [[CrossRef](#)]
53. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-Sampling Technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [[CrossRef](#)]
54. Demircioğlu, A. Applying Oversampling before Cross-Validation Will Lead to High Bias in Radiomics. *Sci. Rep.* **2024**, *14*, 11563. [[CrossRef](#)] [[PubMed](#)]
55. Pearson, K., VII. Note on Regression and Inheritance in the Case of Two Parents. *Proc. R. Soc. Lond.* **1895**, *58*, 240–242. [[CrossRef](#)]
56. Biau, G.; Scornet, E. A Random Forest Guided Tour. *Test* **2016**, *25*, 197–227. [[CrossRef](#)]
57. Guyon, I.; Elisseeff, A. An Introduction to Variable and Feature Selection. *J. Mach. Learn. Res.* **2003**, *3*, 1157–1182. [[CrossRef](#)]
58. Dormann, C.F.; Elith, J.; Bacher, S.; Buchmann, C.; Carl, G.; Carré, G.; Marquéz, J.R.G.; Gruber, B.; Lafourcade, B.; Leitão, P.J.; et al. Collinearity: A Review of Methods to Deal with It and a Simulation Study Evaluating Their Performance. *Ecography* **2013**, *36*, 27–46. [[CrossRef](#)]
59. Bommert, A.; Sun, X.; Bischl, B.; Rahnenführer, J.; Lang, M. Benchmark for Filter Methods for Feature Selection in High-Dimensional Classification Data. *Comput. Stat. Data Anal.* **2020**, *143*, 106839. [[CrossRef](#)]

60. Chandrashekar, G.; Sahin, F. A Survey on Feature Selection Methods. *Comput. Electr. Eng.* **2014**, *40*, 16–28. [[CrossRef](#)]
61. Matharaarachchi, S.; Domaratzki, M.; Muthukumarana, S. Enhancing SMOTE for Imbalanced Data with Abnormal Minority Instances. *Mach. Learn. Appl.* **2024**, *18*, 100597. [[CrossRef](#)]
62. Roberts, D.R.; Bahn, V.; Ciuti, S.; Boyce, M.S.; Elith, J.; Guillerá-Arroita, G.; Hauenstein, S.; Lahoz-Monfort, J.J.; Schröder, B.; Thuiller, W.; et al. Cross-Validation Strategies for Data with Temporal, Spatial, Hierarchical, or Phylogenetic Structure. *Ecography* **2017**, *40*, 913–929. [[CrossRef](#)]
63. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N.; Prabhat, P. Deep Learning and Process Understanding for Data-Driven Earth System Science. *Nature* **2019**, *566*, 195–204. [[CrossRef](#)] [[PubMed](#)]
64. Shen, C. A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists. *Water Resour. Res.* **2018**, *54*, 8558–8593. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.