Process Safety and Environmental Protection 169 (2023) 808–828

Contents lists available at ScienceDirect

Process Safety and Environmental Protection

journal homepage: www.journals.elsevier.com/process-safety-and-environmental-protection

Performance analysis of the water quality index model for predicting water state using machine learning techniques

Md Galal Uddin [a,b,c,*], Stephen Nash [a,b,c], Azizur Rahman [d,e], Agnieszka I. Olbert [a,b,c]

[a] School of Engineering, University of Galway, Ireland
[b] Ryan Institute, University of Galway, Ireland
[c] MaREI Research Centre, University of Galway, Ireland
[d] School of Computing, Mathematics and Engineering, Charles Sturt University, Wagga Wagga, Australia
[e] The Gulbali Institute of Agriculture, Water and Environment, Charles Sturt University, Wagga Wagga, Australia

A R T I C L E I N F O

Keywords:
Water quality index
Coastal water quality classification
Model uncertainty
Classification algorithm
Cork Harbour

A B S T R A C T

Existing water quality index (WQI) models assess water quality using a range of classification schemes. Consequently, different methods provide a number of interpretations for the same water properties that contribute to a considerable amount of uncertainty in the correct classification of water quality. The aims of this study were to evaluate the performance of the water quality index (WQI) model in order to classify coastal water quality correctly using a completely new classification scheme. Cork Harbour water quality data was used in this study, which was collected by Ireland's environmental protection agency (EPA). In the present study, four machine-learning classifier algorithms, including support vector machines (SVM), Naïve Bayes (NB), random forest (RF), k-nearest neighbour (KNN), and gradient boosting (XGBoost), were utilized to identify the best classifier for predicting water quality classes using widely used seven WQI models, whereas three models are completely new and recently proposed by the authors. The KNN (100% correct and 0% wrong) and XGBoost (99.9% correct and 0.1% wrong) algorithms were outperformed in predicting the water quality accurately for seven WQI models. The model validation results indicate that the XGBoost classifier outperformed, including accuracy (1.0), precision (0.99), sensitivity (0.99), specificity (1.0), and F1 (0.99) score, in order to predict the correct classification of water quality. Moreover, compared to WQI models, higher prediction accuracy, precision, sensitivity, specificity, and F1 score were found for the weighted quadratic mean (WQM) and unweighted root mean square (RMS) WQI models, respectively, for each class. The findings of this study showed that the WQM and RMS models could be effective and reliable for assessing coastal water quality in terms of correct classification. Therefore, this study could be helpful in providing accurate water quality information to researchers, policymakers, and water research personnel for monitoring using the WQI model more effectively.

1. Introduction

The management policy of water resource is a critical and systematic process that is associated with diverse components like as policy, law and regulations, institutional framework, advanced analytical facilities, skilled labour, well organization infrastructures, financial freedom etc. A number of framework used to implement the management policy for restoring good water quality status. The monitoring programme is the most widely used approach for assessing water quality on a priority basis. Mainly, it aims is to obtain quantitative information on the physical, chemical and biological attributes of water statistical approaches (Strobl and Robillard, 2008). Whereas it has required specific resource requirements, it is particularly important to have access to technical and financial resources (Steele, 1987). The Water Framework Directive (WFD) is a useful tool that provides detailed guidelines for maintaining good water quality and a healthy aquatic ecosystem (Gikas et al., 2020). However, it has suggested a set of water quality measures for evaluating rivers and streams. In that case, it is frequently impractical and exceedingly costly for all parties involved, particularly those with minimal resources.

To date, many tools and technique have developed for assessing water quality. Water quality index model is one them, its allows converting a vast water quality information into single numerical values more simplified way than traditional approaches (Gupta and Gupta,

* Correspondence to: Civil Engineering, College of Science and Engineering, University of Galway, Ireland.
E-mail address: u.mdgalal1@nuigalway.ie (M.G. Uddin).

https://doi.org/10.1016/j.psep.2022.11.073
Received 14 October 2022; Received in revised form 14 November 2022; Accepted 25 November 2022
Available online 28 November 2022
0957-5820/

2021; Uddin et al., 2021). In recent, this tool widely used for assessing water quality (surface and ground) due to its easy mathematical operators (Gupta and Gupta, 2021; Uddin et al., 2020, 2017). A number of WQIs have established by various countries/organizations in order to specific goals such as ground water quality index, surface quality water index etc (Uddin et al., 2022c). This technique has been criticized for a number of reasons, including (i) uncertainty issues, (ii) model reliability, (iii) transparency, and (iv) model sensitivity. Recently, several studies have revealed that WQI model produced a considerable uncertainty to the final score due to its (Juwana et al., 2016; Sutadian et al., 2018; Uddin et al., 2021). Moreover, recently many researchers has revealed that the water quality index model does not express actual state of water quality due to the entire WQI index model's uses a variety of classification schemes (Uddin et al., 2021). Those are recommended for the interpretation of WQI score using many qualitative measures, including "excellent," "good," "bad," "very bad", "poor", marginal, higher, lower etc. Consequently, different methods provide a number of interpretations for the similar water properties that contribute to considerable uncertainty in the correct classification of water quality. These types of problems can be addressed as "metaphoring problems" of classification. Concerns regarding these problems are growing as the WQI model is becoming more widely used gradually. According to recent studies, the current WQI model provides ambiguous information to the water resources manager as a result of these issues, causing bodies to fail to respond as quickly as required (Uddin et al., 2022a). As a response of above circumstance, it should be refined into a unique scale for determining the proper water quality classification.

Therefore, in order to obtain the WQI values, the present study used seven WQI models, including four weighted (NSF, SRDD, WJ, and WQM) and three unweighted (RMS, Hanh, and AM). Details of the various WQI models can be found in Uddin et al. (2021) and (2022a). WQI values were obtained according to the improved WQI methodology of Uddin et al. (2022a). Details of the methodology can be found in Uddin et al. (2022a). As discussed earlier, the ultimate goal of the WQI model is to classify the water quality using a classification scheme (Najafzadeh et al., 2021; Uddin et al., 2021). Typically, the WQI model's final output is a numerical value that is well known as an index score, it ranges from 0 to 100, with 0 indicating "worst" water quality and 100 "good" (Najafzadeh et al., 2021; Uddin et al., 2021). Moreover, many recent studies have revealed that a significant amount of uncertainty has been produced in the WQI system due to inappropriate classification schemes. Consequently, the traditional classification techniques can provide inconsistent results in the final assessment of water quality for similar groups of water quality indicators (Uddin et al., 2022a, 2021). In addition, recently a number of studies have reported that the widely used classification scheme "One Out-All Out" of the water framework directive also received much more criticism for the same problem (Latinopoulos et al., 2021; Prato et al., 2014). Therefore, in order to avoid this inaccurate assessment and optimize the metaphoring problems of existing classification systems, the authors have proposed a universal classification scheme (see Table 3) for assessing coastal and transitional water quality in an earlier study (Uddin et al., 2022a). Uddin et al. (2022a) have revealed that the results of water quality using the universal scheme could be effective in reflecting accurate scenarios of water quality. Details of the classification scheme and developing methodology can be found in Uddin et al. (2022a). In this research, this classification scheme was utilized to obtain the state of water quality in Cork Harbour. After obtaining the water quality classes, the WQI models performance were evaluate utilizing various machine learning classifier algorithms. For the purposes of predicting classification of water quality, recently advanced machine-learning algorithms have widely used to reduce the model uncertainty (Islam Khan et al., 2021; Kaur et al., 2021; Malek et al., 2022; Najafzadeh et al., 2019; Najafzadeh and Ghaemi, 2019). Recently, a few studies have utilized the machine learning technique in order to assess the WQI model's reliability in terms of predicting the correct classification of water quality (Islam Khan et al., 2021; Najafzadeh et al., 2021). Up to date, most machine learning algorithms has developed for the solution of binary classification (Allwein, 2000; Babbar and Babbar, 2017). As a result, several studies have evaluated WQI models using binary classes (Islam Khan et al., 2021; Malek et al., 2022); even though most WQI models in the literature suggest using multiple classification schemes to evaluate water quality (Uddin et al., 2021, 2022e). In order to solve the multiclass problem, many researchers have developed a number of classifier algorithms using state-of-the-art machine learning technique (Bourel and Segura, 2018; Chamasemani, 2011; Tiyasha et al., 2021). For the purposes of the multiclass problem analysis, most commonly used algorithms are support vector machines (SVM), Naïve Bayes (NB), random forest, decision trees, logistic regression, k-nearest neighbour (KNN), and gradient boosting (XGBoost) classifiers (Uddin et al., 2022b). A few studies have reported that the multiple-kernel support vector regression algorithm and random forest could enhance the model's performance for predicting water quality using the WQI model (Najafzadeh and Ghaemi, 2019; Najafzadeh et al., 2021; Najafzadeh and Niazmardi, 2021). In order to predict water quality classification, recently, several studies have utilised machine learning techniques to assess the performance of water quality model in terms of binary classification of water quality (Asadollah et al., 2021; Cheryl A. Brown and Nelson, 2010; Danades et al., 2017; Savira and Suharsono, 2013). The present study utilized four classifier algorithms (including NB, SVM, KNN and XGBoost) for predicting multi-class classification of coastal water quality incorporating water quality index model. These classifiers were selected based on the initial assessment of six classifiers algorithms (see Table 4). Because, recently, several studies have revealed that the SVM, NB, random forest, KNN, and XGBoost classifiers are effective for the prediction of correct classification in assessing water quality (Danades et al., 2017; Dezfooli et al., 2018; Islam Khan et al., 2021; Kurt et al., 2008; Muhammad et al., 2015; Najafzadeh and Niazmardi, 2021; Prakash et al., 2018; Shakhari and Banerjee, 2019). To the best of the authors' knowledge, this study represents the first effort to evaluate the performance of a WQI model using a multi-class classifier algorithm. In order to evaluate prediction performance of classifier model(s), most studies have utilized widely the receiver operating characteristic (ROC) curves and confusion matrix to compare the model sensitivity, accuracy and efficiency in terms of multi-class classification [in that case it was considered the classification of water quality by using WQI model] (Morrison et al., 2003; Savira and Suharsono, 2013). For the development of the ROC curve(s), the present study was used four classifier algorithms (NB, SVM, KNN, and XGBoost), whereas the confusion matrix analysis technique was used for evaluating the performance of WQI model in terms of the correct classification of coastal water quality in Cork Harbour (Fawcett, 2006; Gonçalves et al., 2014).

As mentioned in earlier sections above, considering the limitations of exiting WQI model according to the findings in the literature (Uddin et al., 2021), in recent years, the authors have carried out several studies in terms of investigating the appropriate technique for selecting crucial indicators (Uddin et al., 2022a, 2022c, 2022f), developed three new sub-index functions for transferring various indicators concentration into the uniform scale (Uddin et al., 2022a), proposed a comprehensive weighting technique incorporating machine learning and statistical based rank order centroid approaches (Uddin et al., 2022a), proposed new two aggregation functions to reduce the model uncertainty (Uddin et al., 2022a), brand new classification scheme for assessing the state of coastal waters (Uddin et al., 2022a), sources of uncertainty and estimation them using machine learning approach (Uddin et al., 2022d). An effort to improve the method and develop a tool that can be used by environmental regulators to abate water pollution. After conducting the aforementioned studies, in this paper propose a more accurate algorithm for predicting and classifying coastal and transitional water quality using the brand new classification scheme in order to reduce the inconsistence of the final assessment of water quality. That could be effective for improving the WQI model performance in terms of model

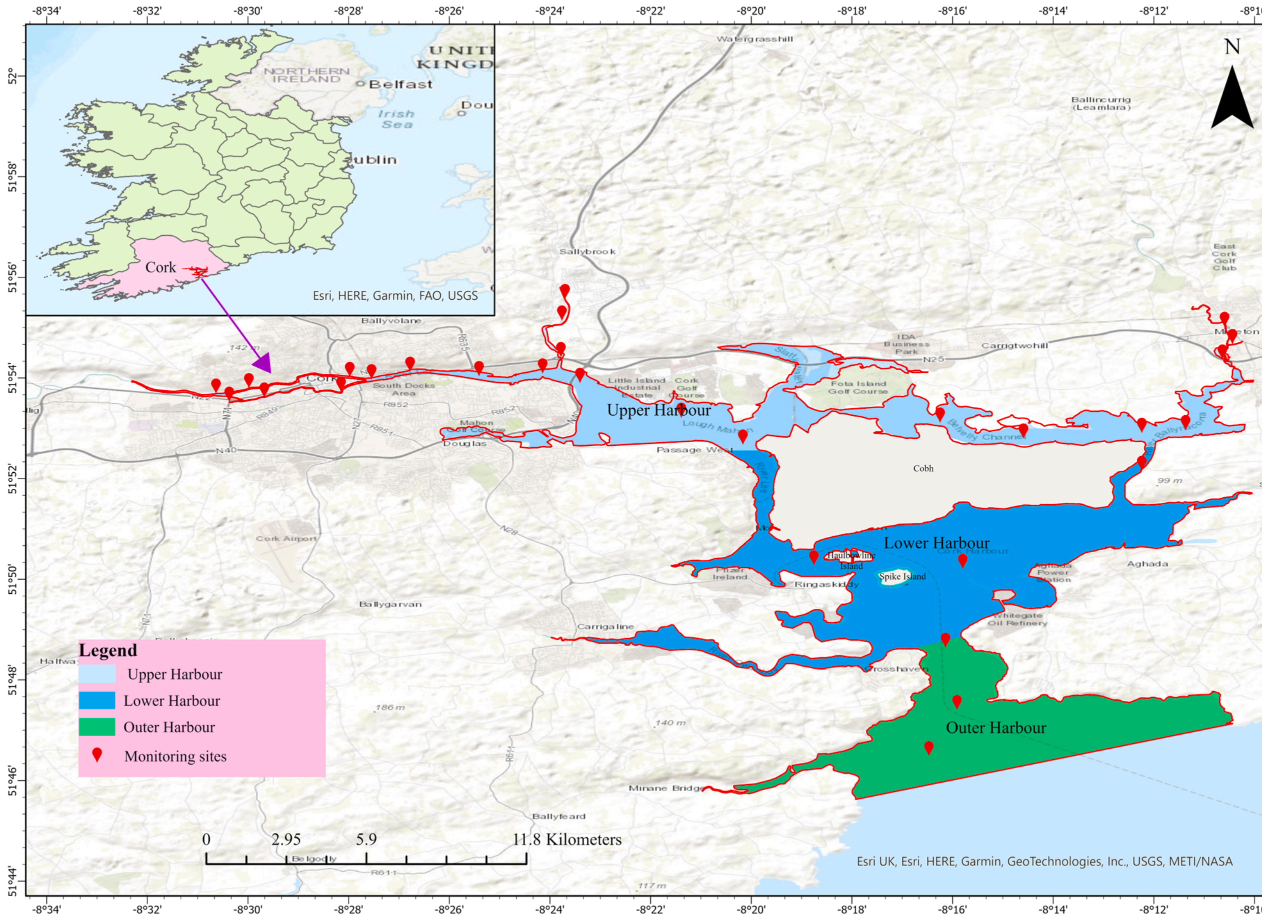

Fig. 1. Model application domain and water quality monitoring sites in Cork Harbour, Ireland.

reliability and consistence of the assessment results.

However, the aims of this study were to assess WQI model(s) performance in order to correct classification using machine learning approaches incorporating a brand new classification scheme for the assessment of coastal and transitional water quality. The goal(s) of the study were obtained by follows:

(i) Archiving WQI score for coastal water quality using recently developed improved WQI approaches (Uddin et al., 2022a) and then determine the water quality classes was utilized the coastal water quality classification scheme (Table 3).
(ii) Once the dataset was obtained, four commonly used predictive classifier models were utilized in this research to identify the best predictive model by comparing them incorporating seven WQI models.
(iii) After that, the best predictive model was applied to predict water quality class for each WQI model.
(iv) Finally, WQI models performance were evaluated using performance metrics (i.e., roc technique and confusion matrix) of machine learning predictive model.

This paper has been divided into five main sections. In the first section, a brief overview of this study are provided in this section. The section two discussed a new set of tools and techniques that are being applied to assess model performance. Results and findings are presented in the section three. Section four discussed how to identify the appropriate WQI model using model performance metrics, and section five presents the conclusions and implications of findings the research.

2. Materials and methods

2.1. Application domain: Cork Harbour

In this research, the proposed framework was employed in Cork Harbour a case study approaches for assessing the coastal water quality in order to correct classification. Cork Harbour located on the southwest coast of Ireland is the largest natural harbour in Ireland. Cork Harbour is heavily populated and industrialized. Cork City located at the mouth of the River Lee is home to a population of approximately 125,000. When its immediate suburbs are included, the population rises to approximately 200,000 (Hartnett and Nash, 2015). The city is the industrial hub of the Irish southwest region and the surrounding hinterlands are subject to relatively intense agricultural activities which impact water quality in the region. Additionally, effluent discharges (Fig. 1) from seven effluents treatment plants (ETPs) in the catchment area further impact water quality in the Harbour (EPA, 2016).

2.2. Data obtaining process

2.2.1. Description of water quality data

For the purposes of this study, the present study was used the water quality monitoring data in year 2019. Typically, the Irish Environmental Protection Agency (EPA) monitors the water quality of the Harbour at 32 monitoring stations. Water samples were taken from one-metre depth below water surface at approximately high and low tides over the year. In this research 29 monitoring sites were considered based on the indicators data availability and coverage of the full extents of the Cork Harbour. Details of monitoring sites and water quality indicators at each monitoring site are given, respectively Fig. 1 and Table S1 in

Table 1
A statistical summary and guideline values of water quality indicators for coastal water quality.

Parameter	unit	Standard threshold (Uddin et al., 2022a)		Statistical summary
		Lower	Upper	
CHL[(i)]	mg/m^3	0.0	14.2	5.32 ± 3.22
DOX[(i)]	% sat	72	128	107.90 ± 15.18
MRP[(i)]	µg/l as P	0.0	0.057	0.02 ± 0.01
DIN[(i)]	mg/l	0.0	1.208	1.54 ± 1.78
AMN[(ii)]	mg/l	0	1.5	0.07 ± 0.06
BOD5[(ii)]	mg/l	0	7	1.74 ± 0.87
pH[(iii)]	–	5	9	8 ± 0.23
TEMP[(ii)]	^{0}C	–	25	15.59 ± 0.74
TON[(iv)]	mg/l as N	0.0	2	1.48 ± 1.79
TRAN[(v)]	m/depth	> 1	–	1.57 ± 1.48

(i) *ATSEBI standards, determine the standard values based on median value of Salinity (see details* Toner et al., 2005*, pp. 72 – 76).*

(i) EPA, Ireland (2001), recommended values for the surface water/freshwater/river water/aquatic life.

(i) Estuary Monitoring Manual for pH and Alkalinity, EPA,USA

(i) The European Communities (Quality of surface water intended for the abstraction of drinking water) regulations, 1989 (S.I. No. 294/1989).

(a) Bathing Water Quality Regulations 2008, (S.I. No. 79/2008).

supplementary material 1. To perform this study, in total, average concentration of ten water quality indicators from the 2019 monitoring dataset were used: water temperature (TEMP), pH, dissolved oxygen (DOX), total organic nitrogen (TON), ammoniacal nitrogen (AMN), molybdate reactive phosphorus (MRP), biological oxygen demand (BOD5), transparency (TRAN), *Chlorophyll a* (CHL) (as a measure of algae), and dissolved inorganic nitrogen (DIN). Details of the water quality indicators data for Cork Harbour are available at https://www.catchments.ie/data/. Table 1 provides the details the statistical summary of water quality indicators and their guideline values for coastal water quality.

2.3. Importance of water quality indicators

Existing WQI models have utilized a range of statistical approaches for selecting the crucial water quality indicators (Uddin et al., 2021, 2022a, 2022b, 2022c). Recently, several studies have revealed that the existing techniques not effective in order to select important indicators (Sutadian et al., 2018; Uddin et al., 2022a). For the purposes of selecting relative importance indicators, authors have proposed an improved method in their earlier studies (Uddin et al., 2022a). Therefore, crucial water quality indicators were selected in this study according the methodology of Uddin et al. (2022a). Details the XGBoost can be found in Uddin et al. (2022a). Seven water quality indicators out of ten were found to be important for obtaining the goals of this research. Fig. 6 presents the important water quality indicators and their relationships.

2.4. Water quality index (WQI)

A range of techniques and tools are used to assess water quality for the management of water resources. The WQI model is one of them. This technique is widely used for the assessment of water quality, i.e., surface water, groundwater, etc. It allows converting huge amounts of water

Table 2
Selected seven WQI models aggregation functions and their properties according to Uddin et al. (2022a).

Types of functions	WQIs Models	Aggregation functions
(a) Weighted models	Weighted Quadratic Mean (WQM)	$$\mathrm{WQM} = \sqrt{\sum_{i=1}^{n} w_i s_i^2} \quad (1)$$
	NSF index [Weighted Arithmetic Mean (WAM)]	$$NSF = \sum_{i=1}^{n} s_i w_i \quad (2)$$
	SRDD index (modified additive function)	$$SRDD = \frac{1}{100}\left(\sum_{i=1}^{n} S_i W_i\right)^2 \quad (3)$$
	West Java WQI [Weighted Geometric Mean (WGM)]	$$WJ = \prod_{i=1}^{n} s_i^{w_i} \quad (4)$$ where s_i is the SI value for indicator i; w_i is weight value of respective variables and n is the number of indicators.
(b) unweighted models	Arithmetic Mean (AM)	$$AM = \frac{1}{n}\sum_{i=1}^{n} s_i \quad (5)$$
	Root Mean Squared (RMS)	$$RMS = \sqrt{\frac{1}{n}\sum_{i=1}^{n} s_i^2} \quad (6)$$
	Hanh index	$$WQI_b = \left[\frac{1}{n}\sum_{i=1}^{n} q_i \times \frac{1}{n} q_j \times q_k\right]^{1/3} \quad (7)$$ Where, WQI_b is the basic water quality index; q_i is the subindex value of the organic; q_j is the inorganic substance; q_k is the subindex value of the biological or bacterial groups components; n is the number of components each group

Table 3
Proposed classification for assessing coastal water quality by Uddin et al. (2022b).

Classifications scheme	Range of WQI score	Descriptions
Good	80–100	Unpolluted waterbodies are those that meet the guidelines' values. Water quality is maintained and is suitable for all uses.
Fair	50–79	Waterbodies that a few indicators meet the guidelines values; water quality is usually protected with a minor degree of impairment.
Marginal	30 – 49	The majority of water quality indicators failed to meet the criteria; water quality is unprotected, which may be posing a risk for aquatic life.
Poor	0–29	Eutrophic waterbodies are those that fail to meet all of the criteria. Water quality is completely unprotected and unsuitable for many specifics uses

Table 4
Model performance of different classifiers obtained from the trial and error approaches for selecting robust algorithm(s).

Classifier algorithms	Rank (based on the model accuracy)	Accuracy (%)	
		Training	Testing
XGBoost	1	98.0	100
KNN	2	93.1	94.0
SVM	3	89.8	92.0
NB	4	90.1	91.0
DT	5	81.1	85.3
ANN	6	87.93	88.2

quality information into a single numerical value that is well known as the index score (Parween et al., 2022; Uddin et al., 2021, 2022a). Since the development, its application has increased recently due to its ease of use and simple mathematical operators compared to other hydrological tools (Uddin et al., 2021, 2022a). In addition, this technique have received much criticism due to the model eclipsing and ambiguity problem that are described in detail in our recent study (Uddin et al., 2021). For the purposes of reducing model uncertainty, Uddin et al. (2022a) recently have proposed an enhanced and comprehensive WQI approach for computing WQI scores in order to assess the coastal and transitional water quality. In this research the WQI values were computed using the improved WQI methodology of Uddin et al. (2022a). The details of the procedure can be found in Uddin et al. (2022a). This approach is shown to be more reliable than that used in existing methods because it is the most up-to-date method for computing WQI, and it may be an effective tool to avoid model uncertainty and ambiguity (Uddin et al., 2022b). For the purposes of calculating WQI values, 10,000 random samples were generated using the Monte Carlo simulation technique. Details of the technique can be found in Ratick and Schwarz, (2009). Once the random samples were obtained, the WQI values were calculated for seven commonly used WQI models, including four weighted, inclusive of the national sanitation foundation (NSF), West Java (WJ), and weighted quadratic mean (WQM), the Scottish research development department (SRDD), and three unweighted with arithmetic mean (AM), root mean square (RMS), and Hanh models using the technique mentioned above. Table 2 provides the seven WQI models' functions and their properties that were used in this research for computing WQI scores. Details of the model components are described in detail in Uddin et al. (2021) and Uddin et al. (2022a). In supplementary material 2, details of the WQI model outcomes are given.

2.4.1. Interpretation of WQI model output

To date, a number of classification schemes have been proposed for assessing water quality for various purposes in the literature. Recently, several studies have revealed that the final score of the WQI model interpretation/evaluation is critical because, currently, entire WQI models use various evaluation schemes for assessing water quality. Consequently, the evaluation results of water quality varied for the unique range of scores. Therefore, the WQI model final score does not reflect the actual information of water quality. As a result, it is difficult to evaluate water quality using index scores. In our recent study, we proposed a new classification scheme for the assessment of coastal water quality based on the attributes of coastal water. It consists of four unique qualitative classes, including "good", "fair", "marginal", and "poor". A detailed classification scheme and their definitions are given in Table 3. The present study applied this classification to determining water quality classes.

2.5. Predicting classifier algorithms for classification of water quality

2.5.1. Predictive classifiers model

In order to select the classifier(s), an initial assessment of the six most widely used algorithms, including XGBoost, SVM, artificial neural network (ANN), NB, KNN, and decision tree (DT), was carried out. Table 4 provides the initial assessment results for six classifiers. It can be seen from the table below that most classifiers achieved higher prediction accuracy during both the training and testing periods, with the exception of the ANN and the DT, whereas the XGBoost obtained the highest prediction accuracy compared to other models (Table 4). Therefore, the present study utilized the top four ranked algorithms for the multiclass classification of coastal and transitional water quality in Cork Harbour. Details of the algorithm rank obtained based on the model prediction performance can be found in Table 4. The following is a brief overview of selected models that were used in this study.

(i) **NB classifier**

Naïve Bayes is an efficient supervised machine learning algorithms that is widely used for binary/multiclass classification (Aldhyani et al., 2020). Fundamentally, Naïve Bayes classifier developed based on the Bayes theorem (Radhakrishnan and Pillai, 2020). It is scalable, requiring a set of parameters proportional to the number of variables in a learning problem (Walley and Džeroski, 1996). Naïve Bayes makes a probability decision by comparing the likelihood of two features that are independent and of equal significance (Elmachtoub et al., 2020). In this study, the NB algorithm was utilized using the same approach that can be found in Radhakrishnan and Pillai (2020).

(ii) **KNN classifier**

For binary classification, the KNN is one of the most commonly used classifier in machine learning technique (Modaresi and Araghinejad, 2014). The classification of the KNN classifier is determined by measuring N number of nearest distance of neighbours (Ahmed et al., 2019). Since the KNN classifier is a more straightforward and reliable algorithm, it can be used to quickly evaluate unknown sample class data using previous learning systems. It can be easily integrated into any machine learning system without the need for prior data distribution knowledge (Modaresi and Araghinejad, 2014). For classification, the KNN classifier first calculated the distance between all sample points, then determined new classes based on the nearest sample categories. The new sample collection is used to determine the classification by taking into account the greatest number of samples' nearest neighbours (R. C. Chen et al., 2020). The KNN assumed that the distances between groups of nearest neighbours are identical (Awan et al., 2020). Usually, the distance between samples is measured using various distance metrics such as Euclidean distance, standardized euclidean distance, Minkowski distance, chebychev distance, correlation distance, city block distance, etc. In this study, the City block distance was used that is assumed by following:

Table 5
Optimized hyper-parameters of four predicting classifiers models.

Model hypermeters	Optimized value			
	NB	KNN	SVM	XGBoost
Distribution	Kernel	–	–	–
Kernel type	Gaussian	–	cubic	–
Kernel scale	–	–	1	–
Box constraint	–	–	4.86	–
Iterations	30	30	30	30
Support	–	–	–	
Standardized	true	false	true	true
Number of neighbours	–	34	–	–
Distance metrics	–	City block	–	–
Distance weight	–	Squared inverse	–	–
Learning rate	–	–	–	0.2
Number of learners	–	–	–	334
Ensemble method	–	–	–	XGBoost
Maximum depth	–	–	–	5
Subsample	–	–	–	1
Lambda	–	–	–	1
Multiclass method	–	–	One-vs-all	–

$$d_{st} = \sum_{i=1}^{n} |x_{si} - y_{ti}| \tag{8}$$

(iii) **SVM classifier**

The SVM classifier is another popular algorithm for the purposes of predicting binary/multi-class classification and regression in machine learning studies (Ahmed et al., 2019). It was first introduced by Boser and Guyon in 1992 (Modaresi and Araghinejad, 2014). Although it's basic architecture was proposed by Vapnik in 1995. The general idea underlying SVM is to construct a hyperplane that allows input data to be divided into different classes in a high-dimensional space (Mohammed et al., 2018). Since it uses a high-dimensional space to detect a hyperplane using a kernel function, the SVM generates the least error in the binary classification (R. C. Chen et al., 2020; Khullar and Singh, 2021). The right classification is given by the optimal hyperplane of SVM, which is optimized based on the minimum distance between all predictors (Haghiabi et al., 2018). The maximum predictor nearest to the hyperplane is used to calculate the best decision boundary (Khullar and Singh, 2021). The SVM classifier was implemented in this study using the same algorithm that was given in detail in Singh et al. (2011).

(iv) **XGBoost classifier**

Boosting is the most widely used algorithm of ensemble learning method that associates combining many weak classifier models (Bourel and Segura, 2018; Tanha et al., 2020). In ML studies, this technique has recently been extensively used to identify the potential models automatically in order to classification and regression (K. Chen et al., 2020). Moreover, this algorithm is a popular technique in a variety of machine learning problems, including feature selection, confidence estimation, missing features, incremental learning, error correction, and class-balanced data, among others (Polikar, 2012). Details algorithm of the XGBoost can be found in Tanha et al. (2020). To date, several boosting approaches have been developed including AdaBoost.MH, LogitBoost, GradientBoost, XGBoost, LightGBM, CatBoost, SMOTEBoost, RUSBoost, etc. The XGBoost is updated and optimized form of the gradientboost algorithm, which is first introduced and successfully implied by Chen in 2016 (Tanha et al., 2020). For performing water quality class prediction, the XGBoost algorithm was implemented in this research because this approach is effective, and versatile in real-world application (Bourel and Segura, 2018). The procedure of the XGBoost was applied in this study using the Uddin et al. (2022b) approaches.

2.5.2. *Input preparation and data pre-process for classifier*

Prior to commencing the application of classifiers algorithms, it is important to standardize predictors (water quality indicators) variables in order to optimize the model training errors. In this research, data standardized using the approach of Uddin et al. (2022b). Supplementary material 2 provides the standardize water quality indicators. For the purposes of obtaining better performance of classifiers, the present study was generated 10,000 random samples using Monte Carlo simulation technique according to the methodology of Uddin et al. (2022d). After obtaining WQI values for each WQI model, water quality classes were determined using the classification scheme in Table 3. Details of the WQI values for various models can be found in Table S2 (supplementary material 1). Once water quality classes were obtained, input dataset prepared for the prediction models, whereas it composes including seven predictors variables and four response attributes including “good”, “fair”, “marginal” and “poor”. Details input data can be found in the supplementary materials 2. When the input was prepared, four predictive models were developed for predicting water quality classes incorporating various WQI models.

2.5.3. *Models hyper-parameterization*

A number of parameters in the underlying algorithm influences the predictive model hyper-parameterization. Model performance and highest prediction accuracy of classifier(s) depends on best set of optimal parameters, that is called tuning parameters or hyperparameter (Qian et al., 2015; Thanh Noi and Kappas, 2017). Several techniques used for tuning the optimal parameters like random search (Bergstra et al., 2012; Florea and Andonie, 2020), Bayesian optimization (Victoria and Maragatham, 2021; Wang et al., 2018), genetic algorithm (Angelova and Pencheva, 2011; Yuan and Gallagher, 2005), Hybrid tuning technique (Serqueira et al., 2020; Szabo and Genge, 2020). Recently, a few studies have utilized the automated tuning technique for identifying the best parameters in order to obtain the higher accuracy of predicting models (Hamadi et al., 2013; Huang et al., 2020). Since the development of the automated tuning approaches, the application of this technique has increased because the typical tuning techniques may lead to set the wrong parameters, time consuming and computational coast is higher (Hamadi et al., 2013). For the purposes of the hyperparameterization, we used the classification learner app (CLA) with MATLAB R2021b in this research. The CLA is a compatible environment that allows users to optimize model hyperparameters using auto-tuning techniques. The dataset is divided into two groups: training (80%) and validation (20%). In details hyperparameters tuning process can be found in The MathWorks (1993). Table 5 provides the best-fitting predictive models and hyper-parameters for the four classifiers. Details of the settings parameters and procedures for four classifiers are described below:

For achieving the best performance of the XGBoost classifier, it's required to optimize a few crucial parameters (Islam Khan et al., 2021; Uddin et al., 2022a). A number of parameters, including the maximum depth of the tree, the learning rate, and the number of learners, are the key tuning parameters in XGBoost, which determine the model complexity for predicting more accurately (Kavzoglu and Teke, 2022). Commonly, a larger depth of the tree is recommended for achieving higher classification accuracy (Zhang et al., 2021). Recently, several studies have reported that too many nodes in a tree may also lead to an over-fitting problem in a classifier (Latha and Jeeva, 2019; Wu et al., 2020). Moreover, the learning rate is another crucial parameter that controls the step size at each iteration of an optimization algorithm as it advances toward a minimum of the loss function (Latha and Jeeva, 2019; Wu et al., 2020). In order to determine the best set of parameters, we tested the value of “maximum depth” from 1 to 50, the learning rate ranges between 0.1 and 1, and the number of learners was tested 1000

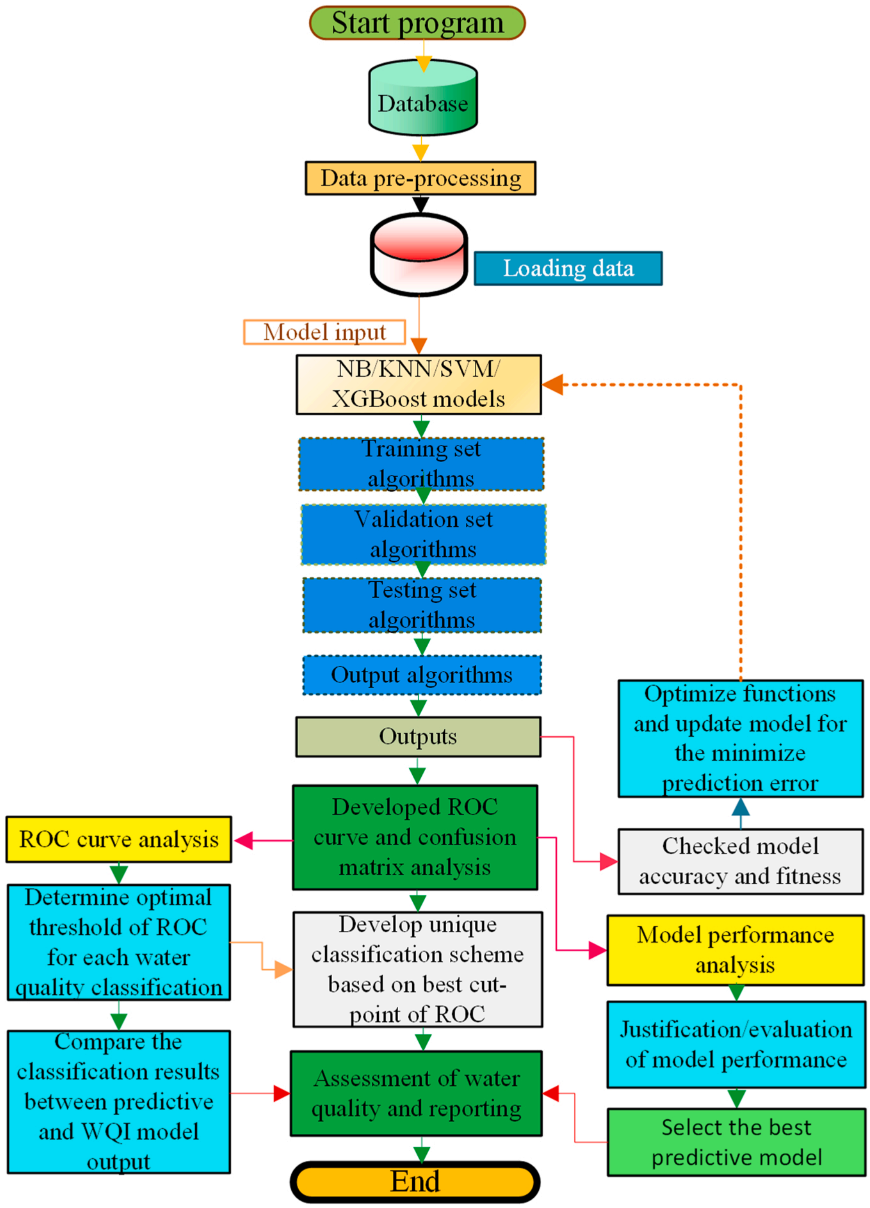

Fig. 2. A conceptual framework of this research.

for all seven training sample subsets whereas other parameters were set at the default value.

In the KNN classifier, the k plays a significant role in enhancing the prediction performance of the KNN (Akbulut et al., 2017; Thanh Noi and Kappas, 2017). To obtain a higher performance of the classifier, the present study used the automated Bayesian approach by implementing the CLA in MATLAB for the determination of the best value of k. In this research, the k values (k = 34) were obtained for all training sets, whereas other parameters were set as default.

On the other hand, the SVM classifier is another widely used algorithm for solving multiclass classification problems (Thanh Noi and Kappas, 2017). In terms of obtaining higher value of accuracy of the SVM, several identical parameters play a vital role in improving the SVM classifier's performance; the kernel function is one of them. Recently, different types of kernel have been widely utilized for predicting water quality using the SVM classifier in water research (Najafzadeh and Niazmardi, 2021). Commonly, four types of kernel functions, including linear, radial basis function (RBF), polynomial (cubic) and sigmoid kernels, are used most frequently in SVM algorithms (Kavzoglu & Colkesen, 2009; Talabani & Avci, 2019). In this study, the kernel function was selected using automatic tuning of hyperparameters using the Bayesian optimization approach (Victoria and Maragatham, 2021). For predicting multiclass water quality, the present study determined the cubic polynomial kernel function for all training sets. In addition, many studies have revealed that the cubic polynomial kernel function is more effective than others for predicting water quality (Chia et al., 2022; Hanoon et al., 2022; Leong et al., 2021). The box constraint is another crucial parameter in the SVM model that controls the maximum penalty rate, which may help to avoid the overfitting problem in the classifier (Kienzle & Schölkopf, 2005; Piccialli & Sciandrone, 2022). The present study optimized the box constraint using an automated Bayesian tuning technique, whereas the optimal value of the box constraint was set at 4.86 for all training sets in the SVM model (Table 5).

In the case of Naive Bayes (NB), it is a widely used probabilistic classifier that is driven by Bayesian statistics (Banchhor & Srinivasu, 2020). Recently, this classifier has been widely used for classifying the water quality and predicting the states in water resources management (Ali Haghpanah jahromi and Mohammad Taheri, 2017; Neha

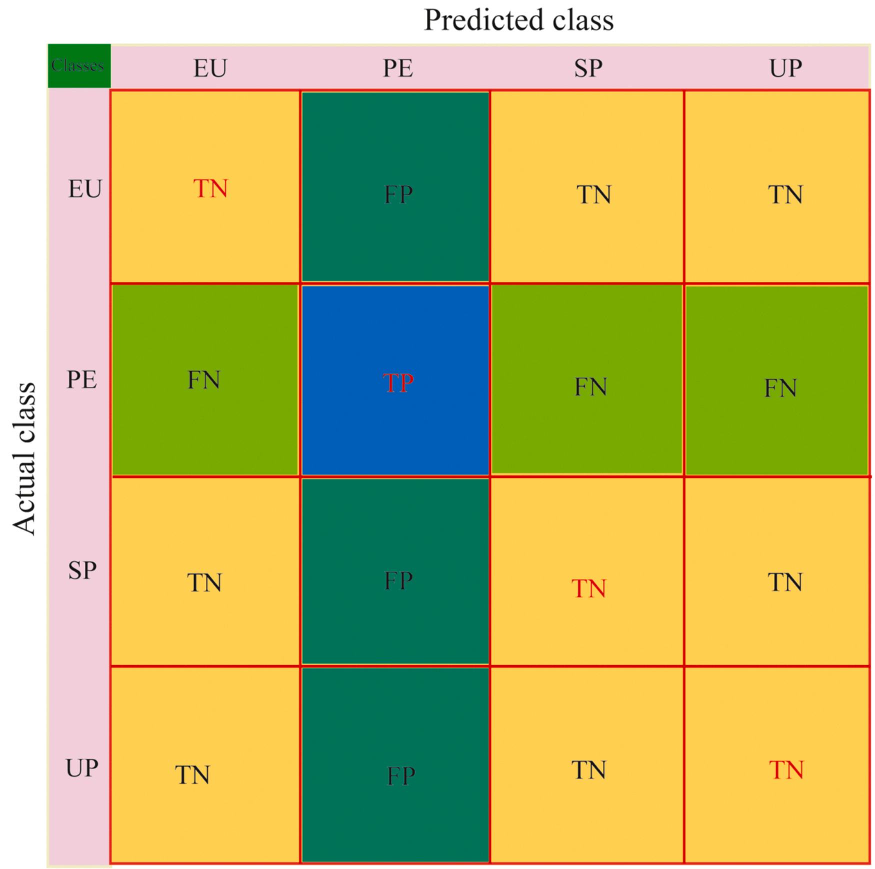

Fig. 3. Proposed architecture of the confusion matrices of the multi-class classification predictive model of WQI model.

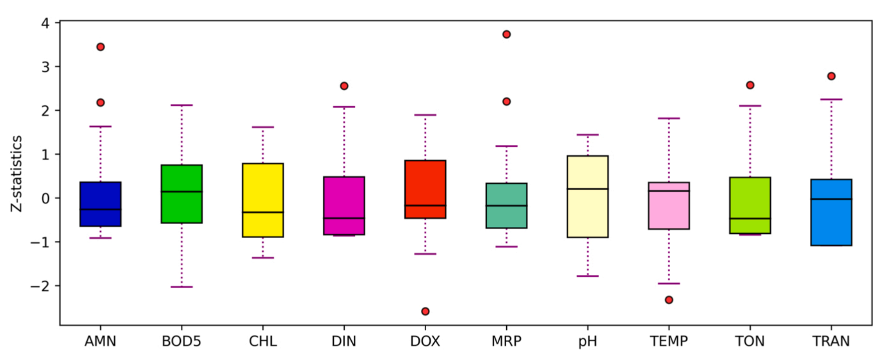

Fig. 4. Z statistics of water quality indicators.

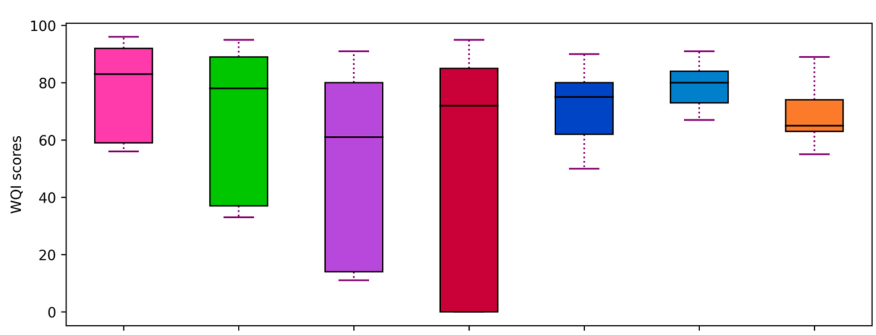

Fig. 5. Statistical overview of various WQI model outcomes.

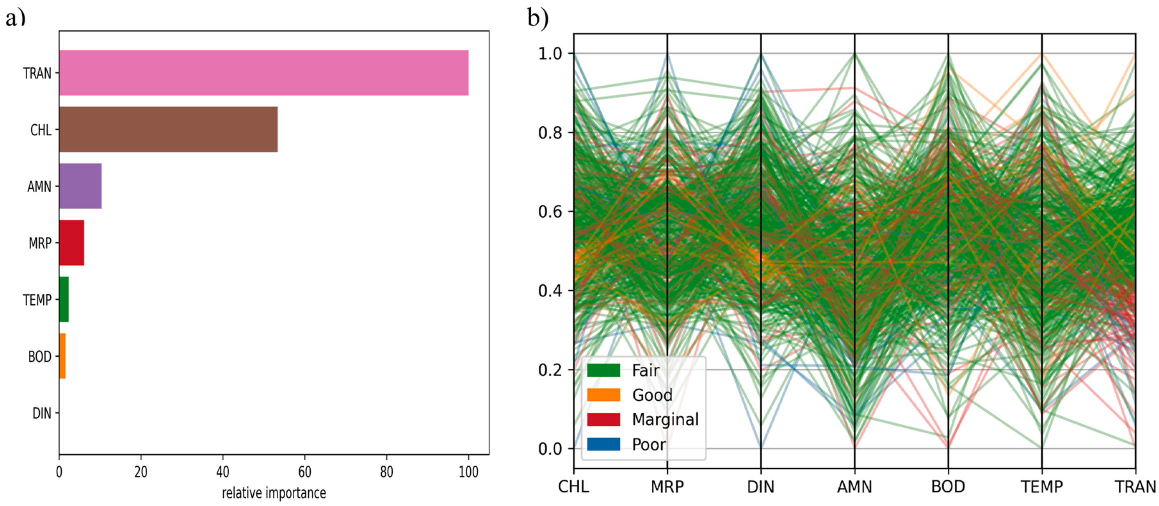

Fig. 6. Attributes of water quality indicators: (a) relative importance and (b) parallel coordinates for seven important water quality indicators.

Radhakrishnan and Anju S Pillai, 2020; Suwadi et al., 2022; Venkata Vara Prasad et al., 2021). Mainly, the NB classifier's performance depends on two identical parameters, including data distribution and kernel function. Commonly, the Gaussian data distribution function is widely used to obtain the highest performance of the classifier (Suwadi et al., 2022; Venkata Vara Prasad et al., 2021). (In contrast to other classifiers, the NB classifier does not require parameter optimization or the setting of any tuning parameters (s). That is one of the great benefits of the NB model in terms of cutting down on the time and cost of computation (Qian et al., 2015).

2.5.4. Development process of predictive model

The details of the process are presented in Fig. 2 below. After completing hyperparameterization, four selective models run in the Python 3.9 environment using the scikit-learn module, which is a built-in module with the capability of compiling several packages for machine learning. Moreover, for the purposes of ROC curve analysis, the Yellowbrick module was utilized in this study, which is specially designed for the analysis of multiclass classification and visualization of ROC and AUC of ROC. Finally, model results were evaluated and visualized using the Matplotlib library. All the analysis has been carried out in this research using the Python programming language.

2.5.5. Evaluation criteria of predictive model

In machine learning techniques, cross validation is widely used approach to evaluate the model's performance. In this research, 10-fold cross validation was applied to measure the accuracy of the predictive model in order to allow multi-class classification. Moreover, in binary or multiclass classification, another technique is the ROC approach that is most commonly used for selecting the best predictive classifier model. In addition, the confusion matrices of the ROC curve are utilized to evaluate the diagnostic results and assess the model performance. The confusion matrices are a cross-reference tables that store the number of occurrences between cells (Fig. 3). Fig. 3 presents the proposed architecture of the confusion matrices of the classifier model for the solution of multi-class classification of the WQI model(s).

For evaluating the classification performance of prediction models, the confusion matrices used to evaluate the performance of model accuracy and sensitivity in a machine learning technique. Commonly, a confusion matrices composes including four components (i) true positive (TP), (ii) false negative (FN), (iii) false positive (FP) and (iv) true negative (TN). The present study, the accuracy, precision, sensitivity, F1 score and area under the curve (AUC) of receiver operating characteristics (ROC) were used. Model evaluation metrics were computed using Eqs. (9) – (12) according to the methodology of Saberi-Movahed et al. (2022) and Mehrpooya et al. (2021).

A ROC curve associated with the average value of the sensitivity for all possible values of specificity within a predicted classifier model (Mandrekar, 2010). In this study, the ROC curve was obtained by considering the cumulative percentage of WQI probability scores (on the x axis) and the cumulative percentage of WQI values occurrence (on the y axis).

Accuracy is widely used criteria to evaluate the classification in machine learning approach (Mehrpooya et al., 2021; Saberi-Movahed et al., 2022). It is defined as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

Another important measures precision refers to how close the measurements between the algorithm predictions and observations of the same classification are to each other. The precision of the model is expressed as follows:

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

Usually, the true positive rate refers to the model sensitivity or recall. It measures how frequently the algorithm detects the correct classification from the given data whereas the actual correct classification has occurred in dataset. In particular, false negative are the classes that have labelled as negative by the classifier model whereas the observation classes are actually positive. The sensitivity is defined as follows:

$$Sensitivityorrecall = \frac{TP}{TP + FN} \tag{11}$$

The F1-score is another measure of a model's accuracy on a data set. It is used to evaluate multiclass classification. It is an approach to harmonizing the precision and recall of the predictive model. The F1-score is obtained following:

$$F_1 score = 2 \times \frac{Precision \times Sensitivity}{Precision + sensitivity} \tag{12}$$

where;.

(i) **TP:** the actual observation indicates that water quality classes has classified accurately and the model predicted correct classification of water quality from the given data.
(ii) **TN:** the actual observation indicates that water quality classes has classified accurately but the model detected incorrect classification of water quality from the given data.

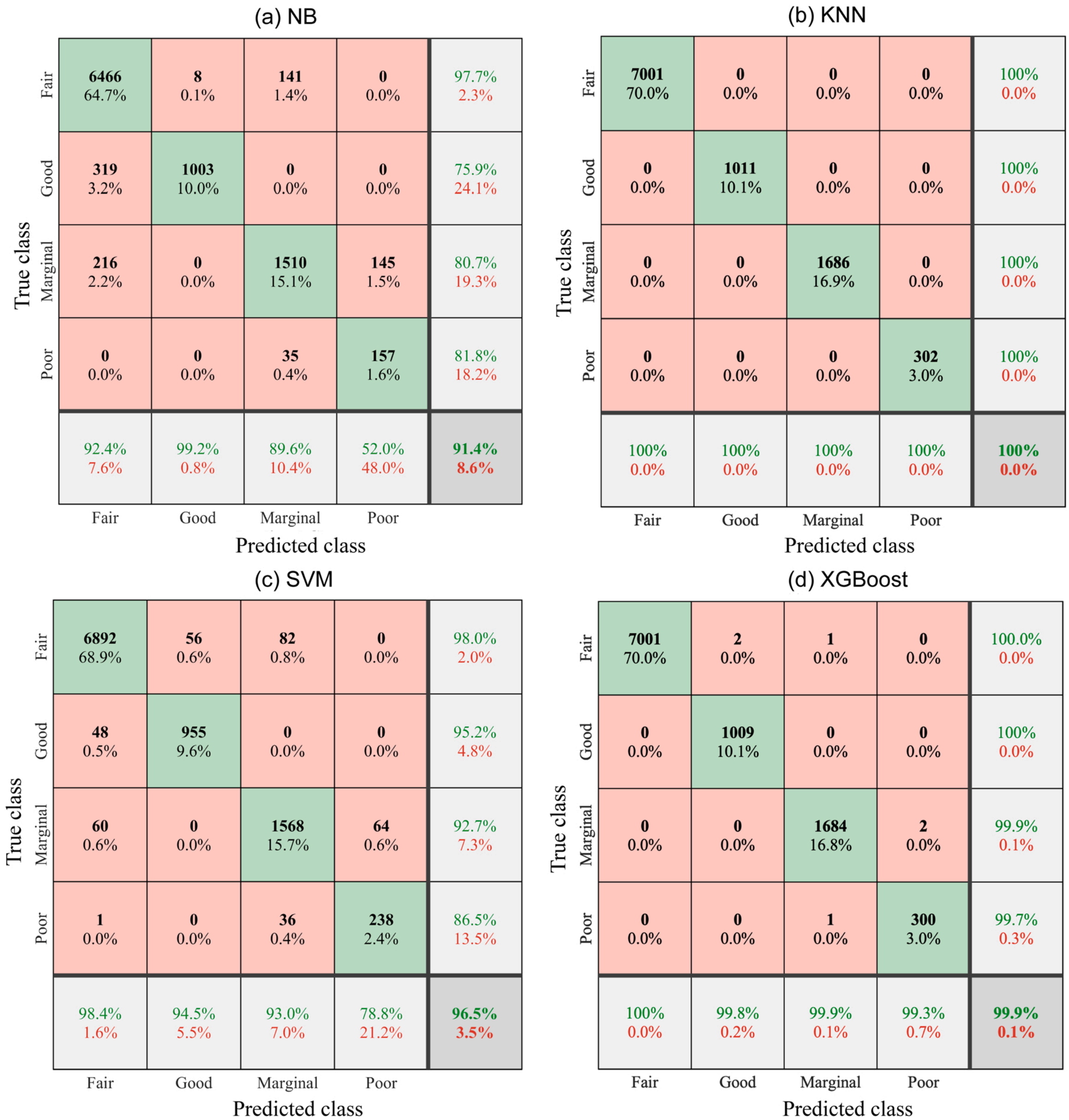

Fig. 7. Results of confusion matrices obtained from the tested four prediction models for the multi-class classification of water quality in Cork Harbour.

(iii) **FP:** the actual observation refers that water quality classes has not classified accurately whereas the model also detects the incorrect classification of water quality from the given data.

(iv) **FN:** the actual observation reveals that water quality classes has not classified accurately although the model predicted correct classification of water quality from the given dataset.

2.6. Developing the ROC curve

In general, the ROC curve computed the classification model that applies a probability, confidence interval or ranking to each prediction (Hamel, 2011). Several models, such as Nave Bayes (Fawcett, 2006), artificial neural networks (Hamel, 2011), and SVM, generate rankings as part of their algorithm (Cristianini & Shawe-Taylor, 2000). The prediction ranking is commonly employed in a ROC algorithm to achieve distinct decision thresholds in each prediction step, ranging from the highest to the minimum ranking value. Typically, prediction-rating values are used for classification to normalize decision threshold values between 0 and 1 where the default threshold is set to 0.5. In terms of model efficiency, the ROC curve was developed using the true positive and false positive rates at each threshold stage. Structurally, the traces a curve from lower left corner to upper right corner (diagonal) in the ROC curve. In order to model performance, the left part of the curve indicates the excellent performance thresholds (conservative) and the right part of the curve dealing with the poor decision thresholds (liberal) (Hamel, 2011). In this study, the ROC curves were obtained from four predictive classifier models. These are presented in Figs. 9 and 10 respectively for the weighted and unweighted WQI models. Commonly, a ROC curve allows examining two basic features for testing the efficiency of the model/method:

(i) To get the right classification, the ROC curve makes it possible to assess the overall performance of a predictor attribute. The significance of discrimination measures when analyzing an area under the ROC curve's (AUC);

(ii) It is to make it possible to compare efficiency between predictors in order to correct classification in terms of given classification schemes; and

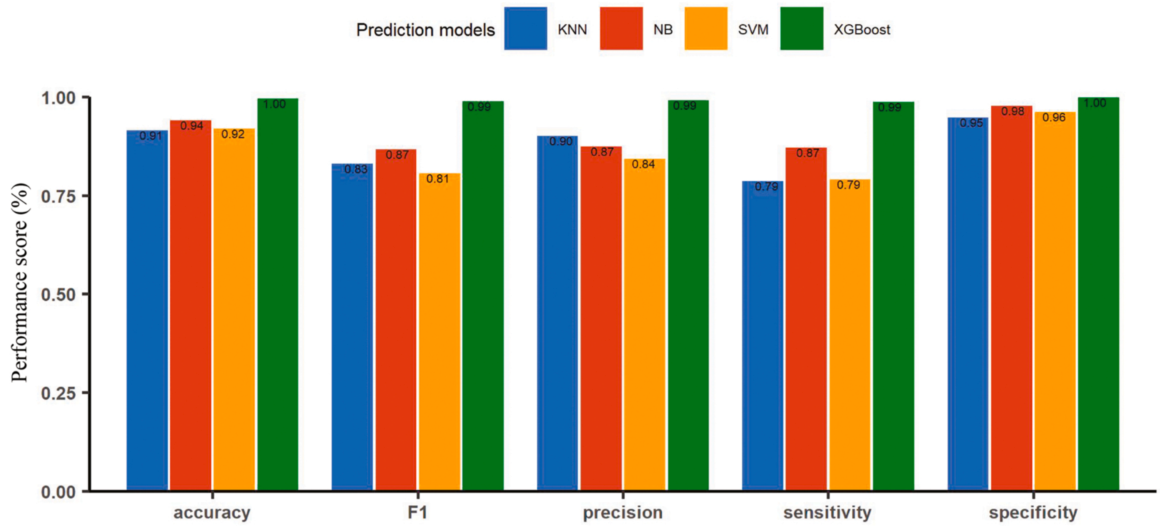

Fig. 8. Results of evaluation criteria for various classifiers.

It offers to identify the optimal threshold value of the predictor variable that refers to the optimal cut-point of ROC associated with true positive rate (TPR) and the lowest false positive rate (FPR).

2.7. *Justification of predicted classification*

The ROC curve technique used to evaluate the performance of predictive classifier model. Also, it is one of the most often used technique in machine learning approaches since it provides a variety of assessments in terms of model sensitivity and specificity (Morrison et al., 2003). The ROC analysis has been used in several recent studies to validate predictive models and assess the discrimination capabilities of a continuous variable as a classifier (Cheryl A. Brown and Nelson, 2010; Yin, 2017). This technique widely used by the biomedical field in the mid-nineteenth century to evaluate the reliability of diagnostic tests (Mandrekar, 2010). In 2003, Morrison et al. (2003) applied this technique to assess the beach water quality and its ability to identify water as suitable or unsuitable for swimming purposes. Cheryl A. Brown and Nelson (2010) provides in-depth analysis of this technique in order to assess water quality. In his analysis, Cheryl A. Brown and Nelson (2010) identify the optimal thresholds in the ROC curve associated with the excess water quality in the ocean. The present study has utilized this technique to evaluate the classification performance of WQI model. Details technique can be found in Brown and Nelson (2015) and Morrison et al. (2003). For the purpose of this analysis, three steps were followed in this study:

(i) A common comparison matrix was computed in order to assess the overall discrimination capacity of multiple random variables for classification into four groups. Fig. 7 provides the confusion matrix for four predictive classifiers.
(ii) Once the comparison matrix was obtained, it was used for the direct comparison of the abilities of different variables for the determination of how many classes were classified correctly (e.g., water quality as "good," "fair," "marginal" or "poor").
(iii) Finally, the optimal cut-off point of ROC was used to determine the best WQI model, which is associated with the optimum sensitivity and specificity trade-off of the model classification.

2.8. *Evaluation of classification performance of WQI model*

The AUC of the ROC curve is an identical indicator for evaluating the model performance using AUC value; it expresses the overall measure of test performance and allows to define the level of prediction accuracy of predicted classes (Morrison et al., 2003; Tesoriero et al., 2017). It also provides the likelihood of model classification (Hamel, 2011). In the present study, we utilized the AUC value for evaluating the predicting performance of classifiers. In order to determine the correct classification, the prediction was then compared to the actual classes of water quality. The AUC value ranges from 0 (no discrimination capability) to 1 (outstanding discrimination) (Walter, 2005). Five important interpretations could be useful to evaluate the classifier's performance. According to Hosmer and Lemeshow (2004), the following interpretations:

(i) Outstanding discrimination (AUC $\geq$ 0.9): the classifier model is capable to classify all classes correctly.
(ii) Excellent discrimination capability (AUC = 0.8 – 0.9): there is a high chance that the classifier is classified correctly.
(iii) Acceptable discrimination (AUC = 0.70 – 0.79): classifier performance is acceptable with a few misclassification
(iv) Poor discrimination (AUC = 0.5 – 0.7): classifier is classified features with higher wrong classification; and
(v) The model has no discrimination capability (AUC $\leq$ 0.5): the classifier is not able to distinguish among classes.

3. Results and discussion

3.1. *Statistical summary of WQI models input (indicators) and output (WQI scores)*

3.1.1. *A statistical summary of WQI models input (indicators)*

Fig. 4 presents the Z statistics of water quality indicators in Cork Harbour during study period. Most water quality indicators were found within the limit of coastal water quality guidelines, except for DOX, MRP, DIN, TON and TRAN. (Table S1). As shown in Fig. 4, normal data distribution was found for the BOD5, MRP, and pH, whereas most indicators showed positive skew data distribution pattern except for the TRAN and TEMP (Fig. 4). Most water quality indicators Z scores were found to be between − 2 and + 2 with the exception of MRP. The results of Z statistics revealed that most indicators had a significant impact on water quality in Cork Harbour without BOD5, CHL and pH. Except these indicators, all indicators had presence the data outliers through the study period.

3.1.2. *A statistical summary of WQI models output (WQI scores)*

Fig. 5 shows a statistical summary of different WQI scores in Cork Harbour through the study period. Supplementary material 2 has more

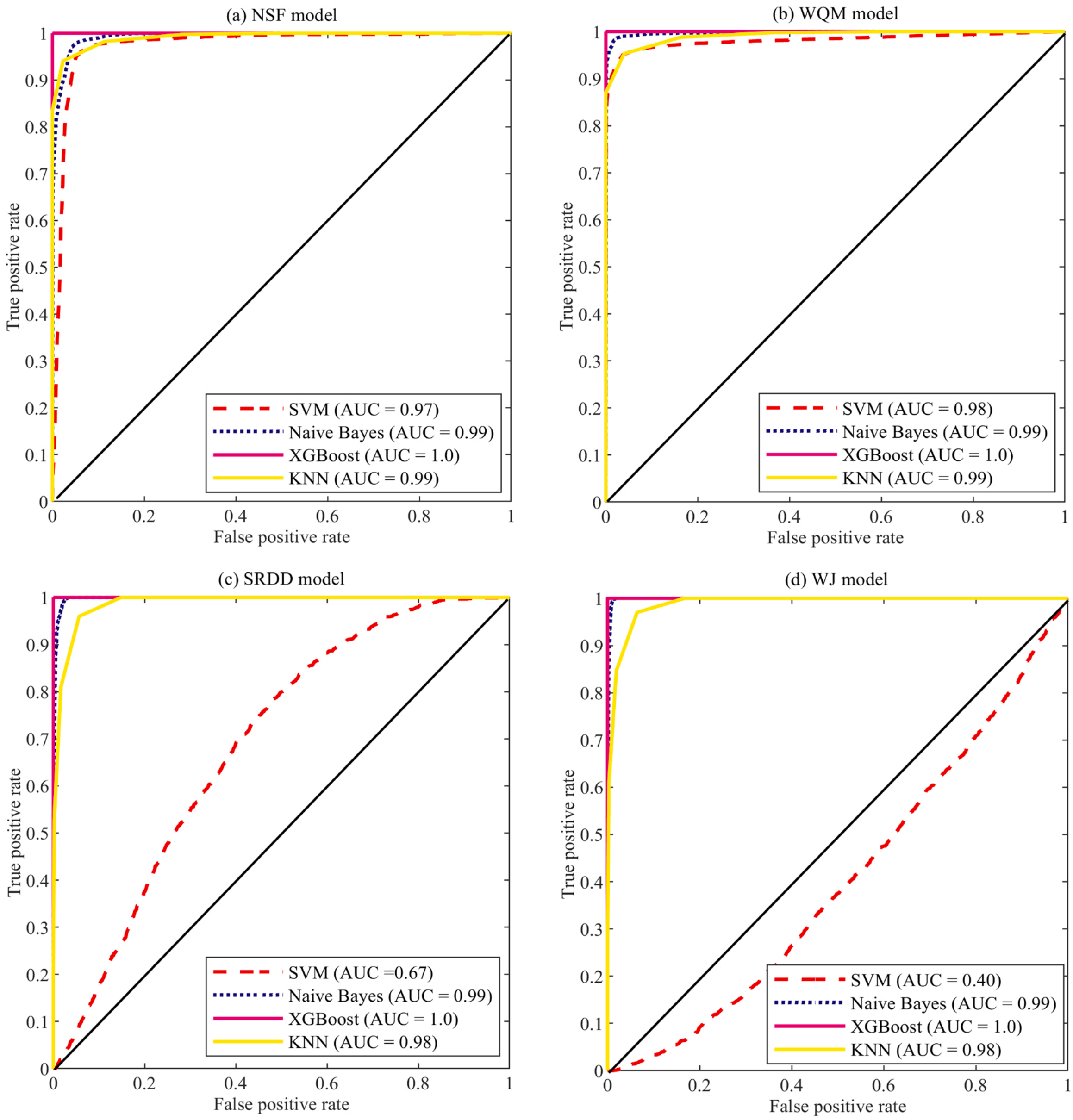

Fig. 9. Compare the ROC curves results among algorithms for various weighted WQI models.

information on the outcomes of different WQI models and their implications for the status of the water quality. Compared to the weighted WQI models, a significant difference was found among models, whereas higher index variation was calculated for the WJ and SRDD models, respectively. With the exception of the Hanh index, there were no discernible differences between the models when compared to unweighted models. However, a comparatively large index score variation was found in the weighted SRDD and WJ models in Cork Harbour over the study period. The results of the WQI scores were in line with the authors' earlier studies (Uddin et al., 2022a, 2022b, 2022f).

3.2. Selecting important predictors for the classifier model(s)

In order to determine the effect of water quality indicators on classification, the present study performed a relative importance analysis using the XGBoost algorithm. In this research, we found the outperforming impact of TRAN, CHL, AMN, and MRP on water quality in Cork Harbour. Fig. 6(a) shows the indicators of importance and their relative ranks. Moreover, in the present study, we utilized the parallel coordinates technique for the determination of relationships between water quality indicators and the tracing of the effectiveness association between them in various water quality classes. It was visualized in Fig. 6(b). It can be seen from Fig. 6b that the TRAN and CHL highly dominated on marginal water quality, whereas DIN showed a higher impact on the poor water quality in Cork Harbour.

3.3. Results of confusion matrix

The aim of the present study to compare the performance of the four machine learning classifiers in order to identify the best algorithms in terms of correct classification. The results of the classifiers were evaluated using five validation metrics (accuracy, precision, sensitivity, specificity, and F1 score) for the imbalanced dataset, the confusion matrices is one of them. Fig. 7 shows the confusion matrix for the four predictive classifiers models. In this analysis, 10,000 observations belonging to four classifications, including "good", "fair", "marginal," and "poor", were used to predict the classification.

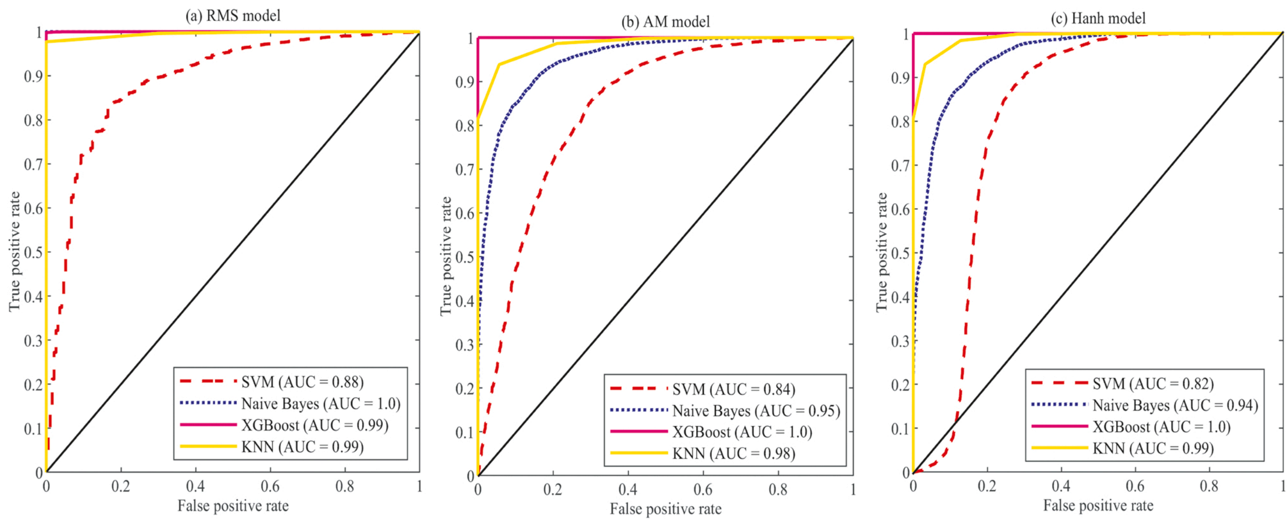

Fig. 10. Compare the ROC curves results among algorithms for various unweighted WQI models.

(i) Confusion matrix results of NB classifier

As shown in Fig. 7(a), good water quality is classified 99.2% correctly, whereas 0.8% is classified incorrectly. In contrast, the fair water quality class is correctly classified at 92.4% and wrongly classified at 7.6%, respectively. Whereas the marginal water quality is correctly classified at 89.6% and wrongly classified at 10.4%. Similarly, for poor water quality, 52.0% of observations are correctly classified, whereas 48.0% are incorrectly classified (Fig. 7a).

(ii) Confusion matrix results of KNN classifier

The results of KNN show that four water quality classes are 100% correctly classified. There was no prediction error in the classification (Fig. 7b). That means the KNN model had an overfitting problem, which may be due to the imbalanced dataset (Japkowicz, 2000).

(iii) Confusion matrix results of SVM classifier

In the SVM classifier, an average of 95% of the observations are classified correctly for all water quality classes except for poor water quality (Fig. 7c). Only 78.5% of the observations were correctly classified into the poor class, whereas the remaining observations were classified wrongly.

(iv) Confusion matrix results of XGBoost classifier

The XGBoost is classified water quality 99.5% correctly for all water quality classes in Cork Harbour (Fig. 7d).

Based on the confusion matrices results, in order to predict the correct classification of water quality, the XGBoost showed outperformance when compared to the confusion matrixes of four predictive classifiers.

3.4. Selection of the best predictive classifier

For the purposes of the performance analysis of classifiers, the present study compared the four classification predictive models using their accuracy, precision, sensitivity, or recall, and F1 scores. Fig. 8 shows the performance results of various predictive models. In this research, predictive accuracy was found to be 91%, 94%, 92%, and 100% for the NB, KNN, SVM, and XGBoost models, respectively. Compared to models, the excellent performance was found to be the XGBoost algorithm. The XGBoost model obtained the highest precision, sensitivity, and F1 scores, whereas the SVM model provided the lowest accuracy, precision, sensitivity and F1 scores (Fig. 8). The results of the performance metrics indicate that the XGBoost algorithm is effective for predicting the correct classification of water quality by incorporating the WQI model. The results of the predictive classifiers performance are line with those of earlier studies (Aldhyani et al., 2020; Islam Khan et al., 2021; Khoi et al., 2022; Malek et al., 2022; Nasir et al., 2022).

3.5. Performance analysis using ROC curve

The ROC curve is widely used to evaluate the classification abilities of the classifier predictive model (Hamel, 2011; Savira and Suharsono, 2013). Figs. 9 and 10 present the ROC curves for the various WQI models. They were obtained from the ROC curve technique using proposed multi-class classification schemes of coastal water quality as provided in Table 3 above. As seen in Figs. 9 and 10, excellent performance was found for the XGBoost and KNN classifiers for the prediction of water quality using weighted and unweighted WQI models. Whereas the SVM and NB classifiers showed the worst performance for the prediction of water quality using WQI models (Figs. 9 and 10). Compared to the all-predictive classifiers, the XGBoost model showed perfect performance for both weighted and unweighted WQI models. The results of ROC curves revealed that the XGBoost predictive classifier model could be effective and reliable to predict water quality class using WQI (Islam Khan et al., 2021; Khoi et al., 2022; Malek et al., 2022; Nasir et al., 2022).

Moreover, the AUC of the ROC curve is commonly used to measure the accuracy of the predictive model (Gonçalves et al., 2014; Savira and Suharsono, 2013). In this study, the AUC value was utilized to evaluate the overall performance of the predictive algorithm in order to classify water quality as good, fair, marginal, and poor. In this study, the AUC of ROC curves was calculated from the ROC curves of four predictive algorithms. According to Hosmer and Lemeshow (2004) classification based on the AUC value, the XGBoost model obtained the outstanding model discrimination for all weighted and unweighted WQI models. For all WQI models, the AUC of ROC was found to be 1, whereas the lowest values were computed for the SVM model.

3.6. Evaluation of classification scheme

The entire WQI index model final output composes several qualitative classes like as excellent, good, bad, very bad etc. A range of WQI models used various classification scheme to assess water quality. As a result, different method provides a number of interpretations for same water properties that contributes a considerable uncertainty to the correct classification of water quality. In that case, the present study proposed a universal classification scheme for coastal water quality that is given in Table 3. In order to evaluate the classification of water quality, the present study is utilized the ROC curve analysis technique.

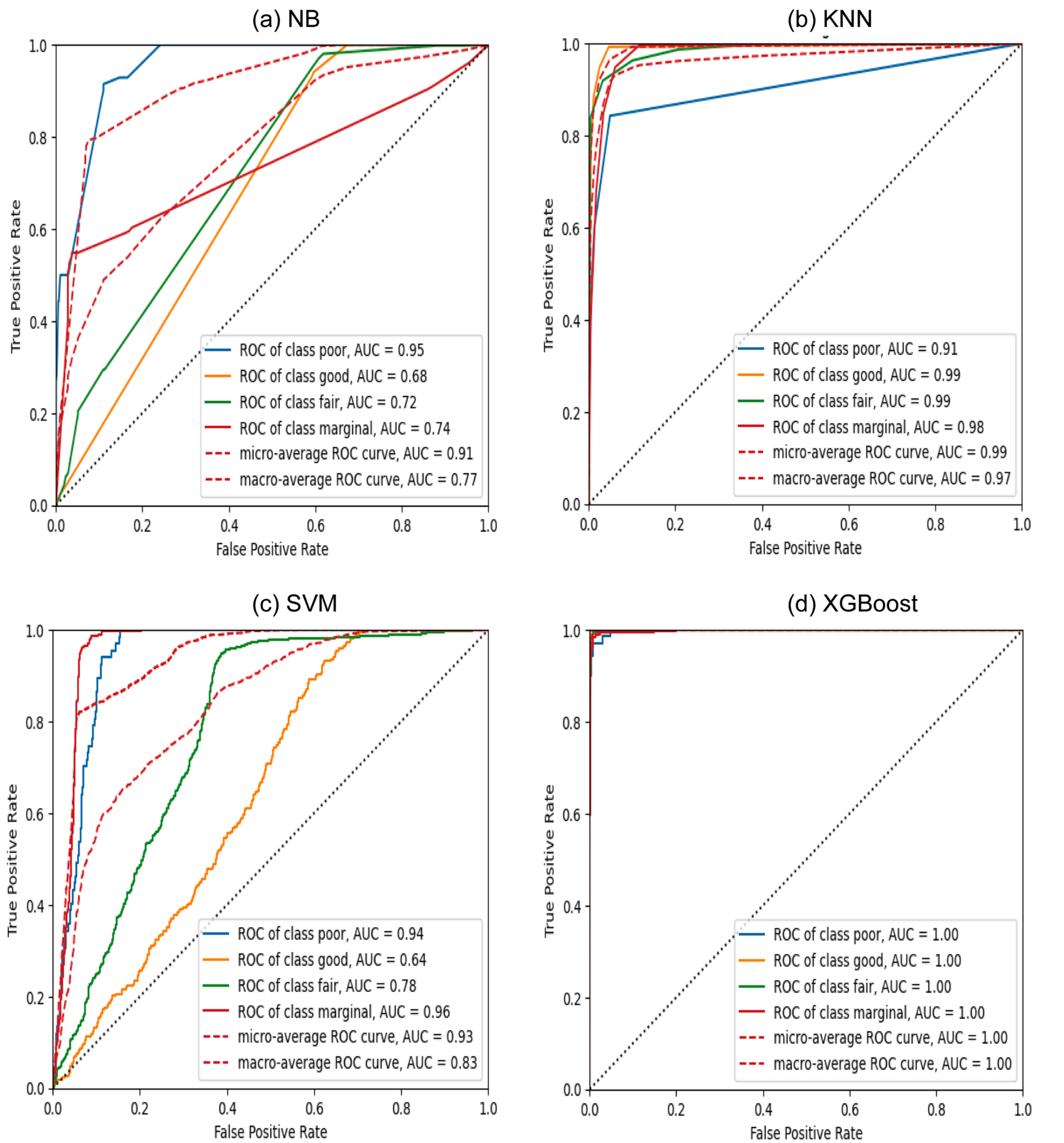

Fig. 11. ROC curves of four classifier predictive models for the multi-class classification of water quality using WQI.

Recently, several studies applied this method to rank the water quality based on AUC of ROC curve (Asadollah et al., 2021; Garabaghi, 2021; Islam Khan et al., 2021). This method is particularly useful in finding the best cut-point value of ROC curve in terms of correct classification (Cheryl A. Brown and Nelson, 2010). For the purpose of water quality classification using WQI, the ROC curves were developed using four water quality classes, including good, fair, marginal, and poor. Fig. 11 presents the ROC curves and AUC of the ROC curve for each water quality class. There was a significant statistical difference among the classifier models (at $p < 0.05$). As can be seen from the figure below, the XGBoost classifier correctly classifies all four water quality classes, whereas the remaining model shows excellent discrimination capability to distinguish between the four classes in accordance with Hosmer and Lemeshow (2004). The outstanding discrimination ability was found for the XGBoost model, whereas the AUC was obtained at 1.0 for each water quality class (Fig. 11d).

Fig. 12 shows the prediction accuracy, precision, sensitivity, specificity, and F1 score of various classifiers for predicting the water classes. As can be seen from the Fig. 11, the XGBoost classifier provided the highest accuracy for four water quality classes. The results of classifier evaluation metrics indicate that the XGBoost is the perfect classifier for predicting water quality using the WQI model.

Fig. 12 present the prediction performance of four selected algorithms. Findings reveal that XGBoost is over performing other methods

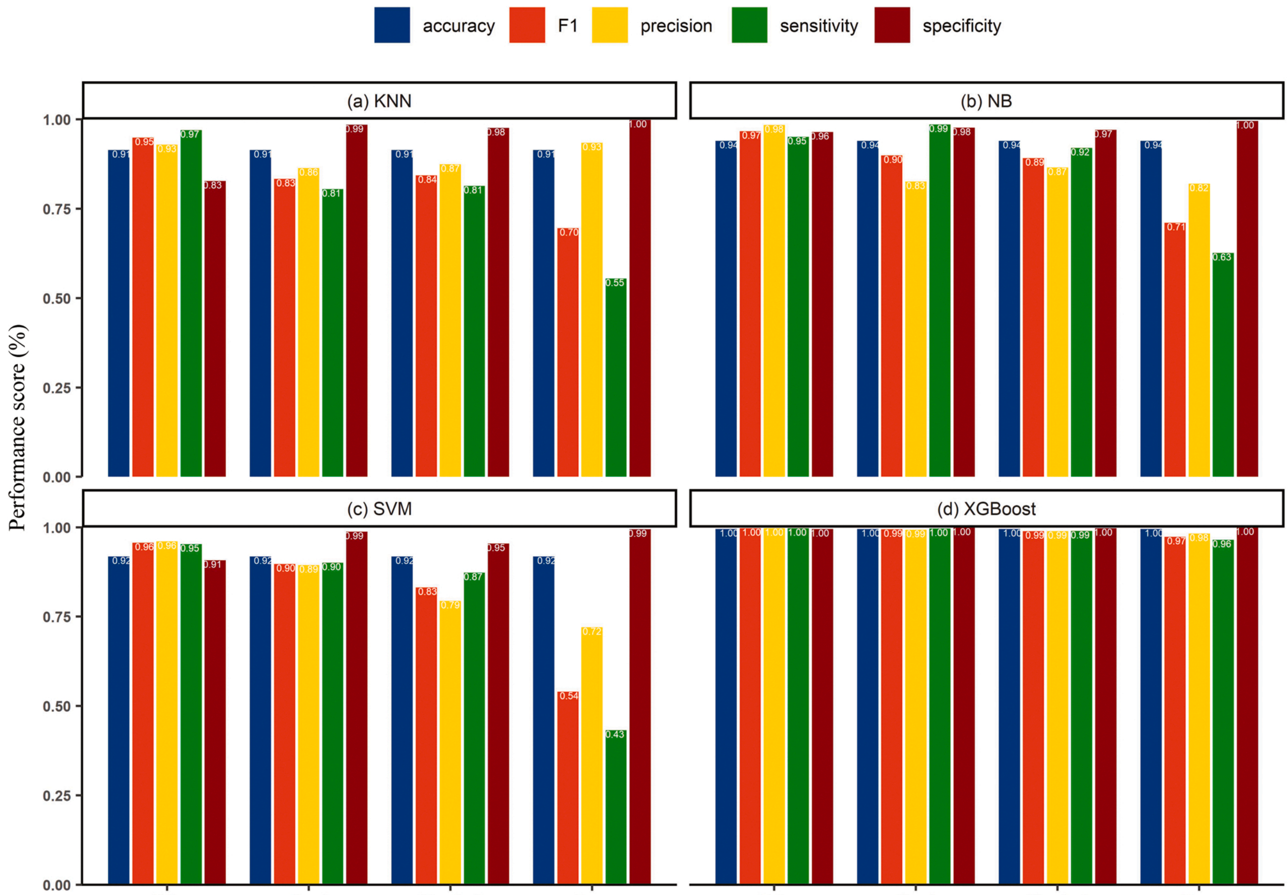

Fig. 12. Prediction performance of various algorithms for predicting water quality classification using the WQI model.

based on all five indicators. In particular, for the "good" to "marginal" classes it shows outstanding performance, while for the "poor" class it also shows nearly 96% performance score by F1, precision and sensitivity.

3.7. Class prediction error of classifier

Fig. 13 presents the results of class prediction error for four classifiers. The highest class prediction error was found for the NB and SVM prediction models, whereas relatively less error was found in the KNN model. Compared to four classifiers, 99.9% of water quality classes were classified correctly by the XGBoost model. The XGBoost classified "poor" and "good" water quality nearly 100% correctly, while less than 2% prediction error was detected for the "fair" and "marginal" classes' water quality (Fig. 13d). The results of the prediction error of classifiers consistent with those of earlier studies (Khoi et al., 2022; Malek et al., 2022; Nasir et al., 2022).

3.7.1. Optimization of class prediction error

The present study used the discrimination threshold technique to minimize the class prediction error of the water quality index model. Commonly, it is the probability or score of the ROC curve that obtained from the tuning of the normal threshold values. This technique is widely used in machine learning studies to optimize the classification error in order to correct classification by classifier algorithm(s) (Zou et al., 2016). Fig. 14 shows the discrimination threshold for the four predictive models. whereas on the x-axis presents the discrimination threshold level that is indicated by the FPR/FNR of classification; and the y-axis shows the percent of precision, recall, and F1 score of the predictive classifier(s) in terms of correct classification of water quality (Hossain et al., 2020; Zou et al., 2016). As shown in Fig. 14 below, the discrimination threshold can be found at 0.25, 0.67, 0.32, and 0.0 for the NB, KNN, SVM, and XGBoost classifiers, respectively. At a 0.0 discrimination threshold, compared to the four classifiers, the highest scores in precision, recall, and F1 were found for the XGBoost model (Fig. 14d). The XGBoost model classified water quality 100% correctly after tuning the normal threshold of the ROC curve. Whereas the remaining classifiers' precision, recall, and F1 scores were found between 0.80 and 0.95 (Fig. 14d). The results of the discrimination threshold also indicate that the XGBoost classifier predictive model is effective for the classification of water quality correctly using the WQI model.

As can be seen in Fig. 15 below, after tuning the normal threshold of ROC curve for the XGBoost model, the critical threshold values was associated at 0% false positive rate (FPR) and 100% true positive rate (TPR) of ROC curve for all ("good", "fair", "marginal" and "poor") water classes. The perfect classification performance was found for all classes. The excellent discrimination of the ROC curve indicates the optimum cut-point that is associated with the highest TPR (correctly classified) and the lowest FPR (wrongly classified) in the ROC curve (Hosmer and Lemeshow, 2004). In Fig. 15 below, in the top-left corner, the black point indicates the optimum threshold of ROC.

However, the critical threshold value of the ROC curve is an important measure for identifying the correct classification when a continuous variable is regarded as a discrete variable (classification). Therefore, the results of the critical threshold value of the ROC curve could help to improve the accuracy of water quality classification using the WQI model. Moreover, this approach might be useful for obtaining an accurate qualitative assessment of coastal water quality in order to reduce the uncertainty in the WQI model.

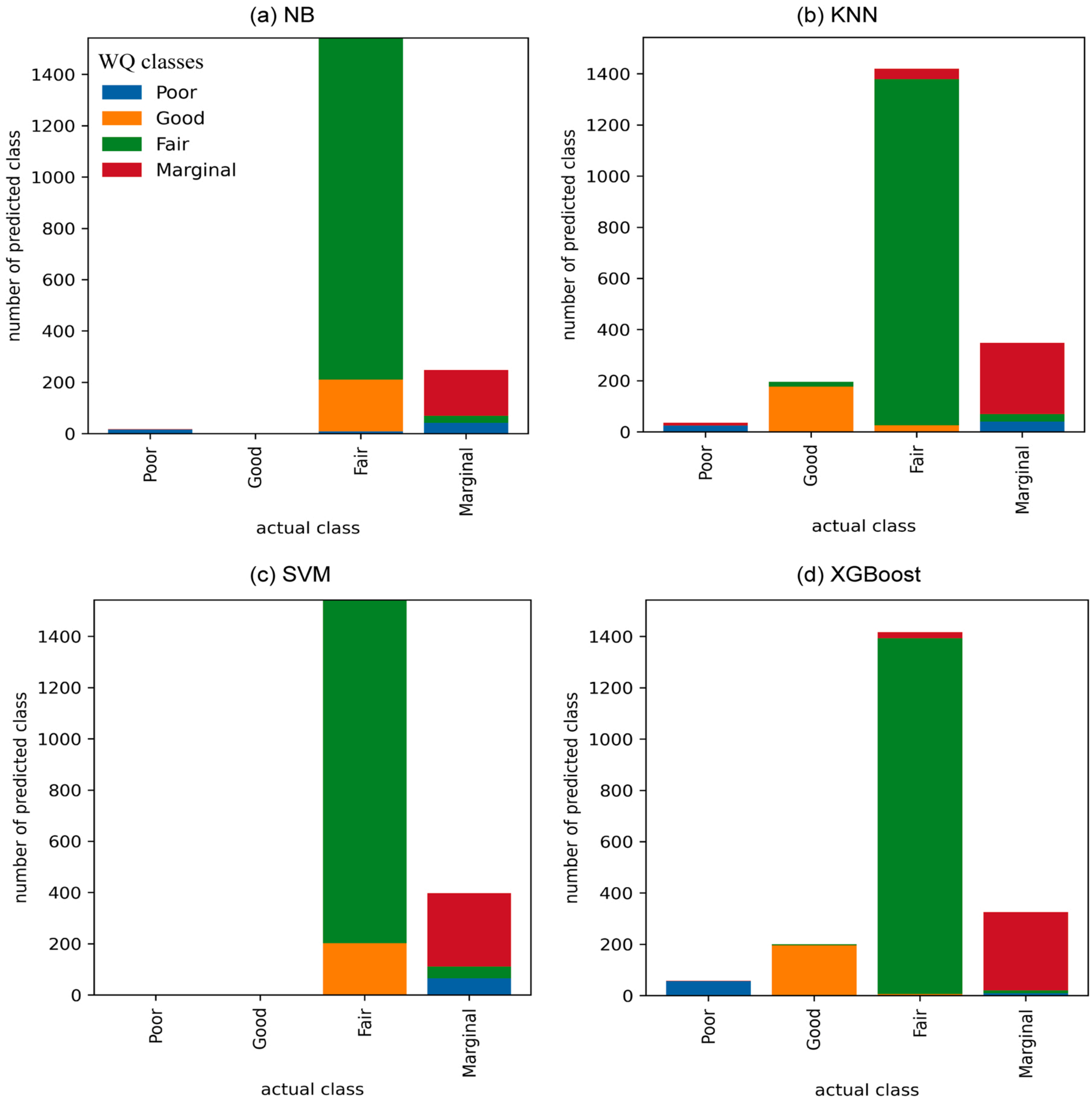

Fig. 13. Comparing the prediction errors of water quality classification between predicted and actual classes of the WQI model.

4. Evaluation of WQI model performance

Recently, several studies have applied the ROC curve technique to the selection of the best model or fitted dataset by comparing AUC-ROC values (Macskassy et al., 2005; Hamel, 2011; Walter, 2005). Moreover, the optimal threshold of the ROC curve allows determining the best performance point that is associated with model sensitivity and 1-specificity (Hong, 2009). Commonly, the optimal threshold indicates the highest TPR and the lowest FPR in the ROC curve where the sensitivity and specificity are most closely related to the value of the area under the ROC curve. It means the absolute difference between the sensitivity and specificity values is the smallest at the optimal point of ROC. (Gonçalves et al., 2014; Hamel, 2011; Unal, 2017).

Several studies utilized this technique to select the best model based on its classification performance (Zou et al., 2016). Commonly, the smallest cut-point value indicates the highest accuracy, and the largest value refers to the lowest accuracy of the classification. In this research, the best WQI model was selected using these approaches. The ROC curve with a pointwise 95% confidence interval was obtained from four weighted and three unweighted WQI models, respectively, using the best predictive classifier, the XGBoost model.

Fig. 16 shows the pointwise 95% confidence intervals that are associated with the vertical averaging values of the ROC curves for fixed false-positive rates and averages the corresponding true-positive rates at each point. In Fig. 16, the red circle indicates the critical threshold value for each WQI model. As can be seen from the figure below, outstanding performance showed by the WQM and the NSF models with the lowest critical threshold values (Fig. 16a). Whereas, the AUC of ROC curves was measured at 0.98 and 0.96, respectively, for the WQM and NSF models. Relatively, excellent and acceptable discrimination capabilities were found for the SRDD and WJ models with higher critical threshold values, respectively (Fig. 16a). Compared to unweighted models, outstanding performance was observed for the RMS model with lower critical threshold values, whereas excellent performance was found for the AM and Hanh WQI models (Fig. 16b). The results of the pointwise 95% confidence intervals and critical threshold values indicate that the WQM and the NSF models could be effective for evaluating coastal water quality in order to correct classification, whereas the unweighted RMS model showed similar results.

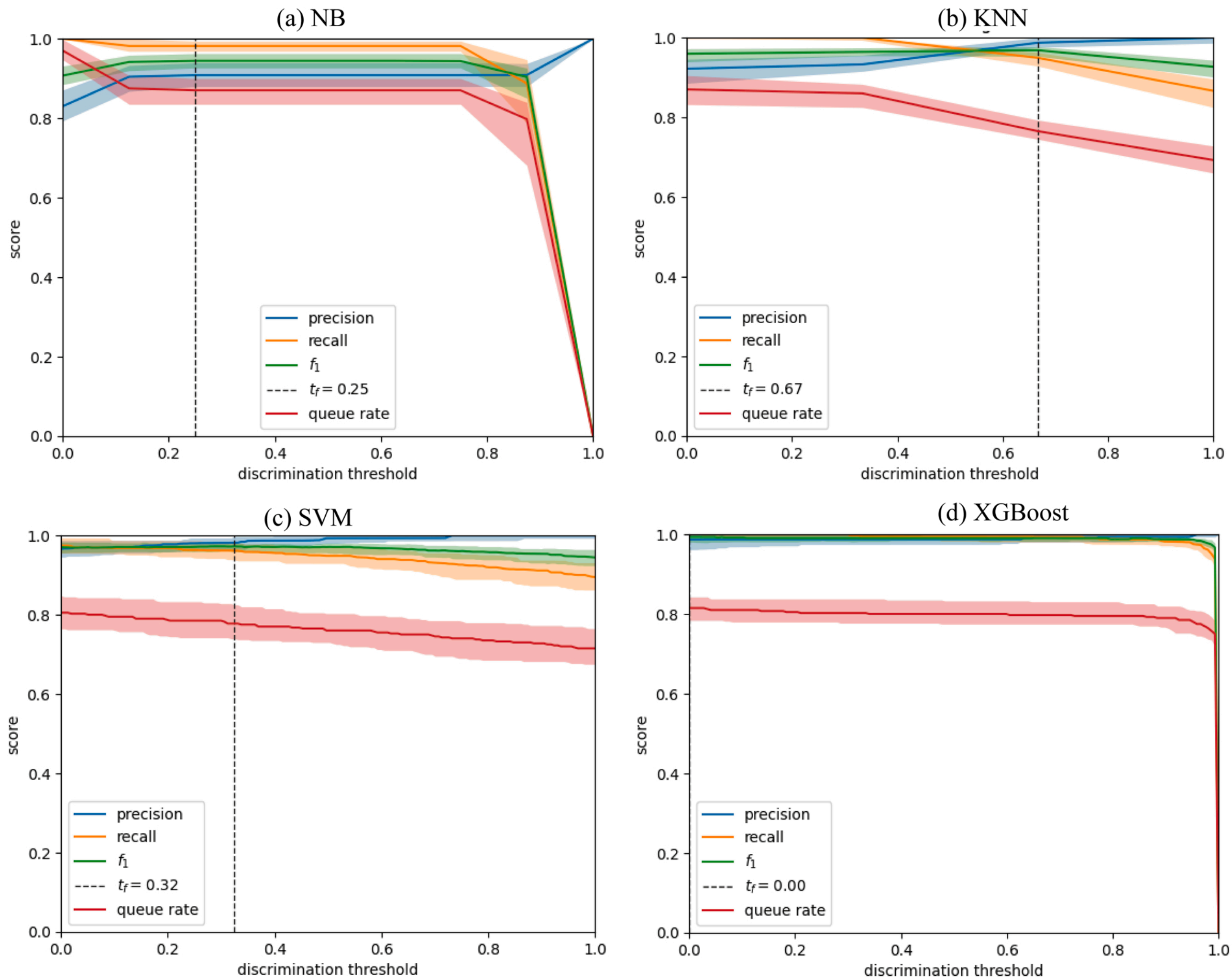

Fig. 14. Discrimination threshold plots for four predictive classifier models.

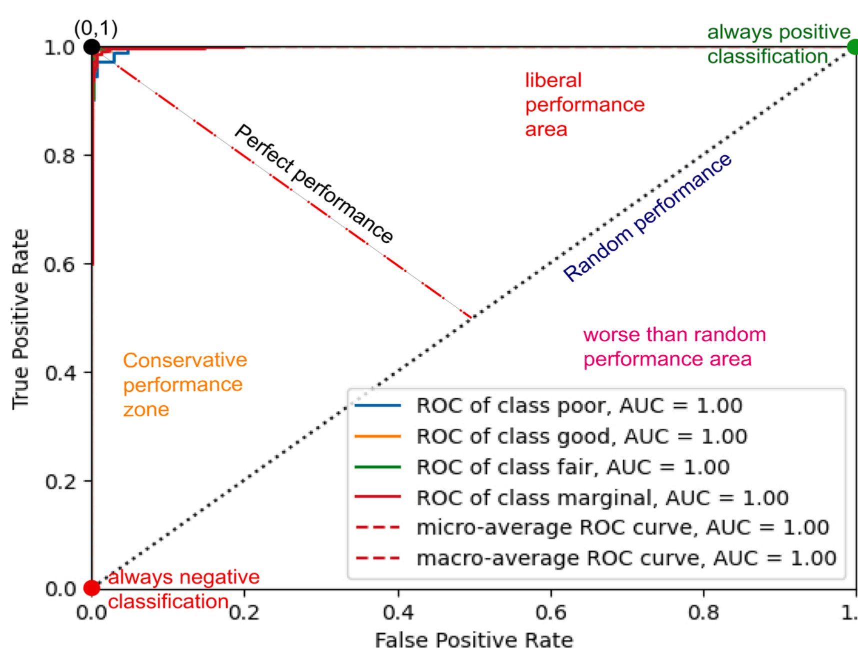

Fig. 15. Region of ROC curve and optimal thresholds of water quality classes for water quality index model [figure outlined, and concept developed according to (Hamel, 2011)].

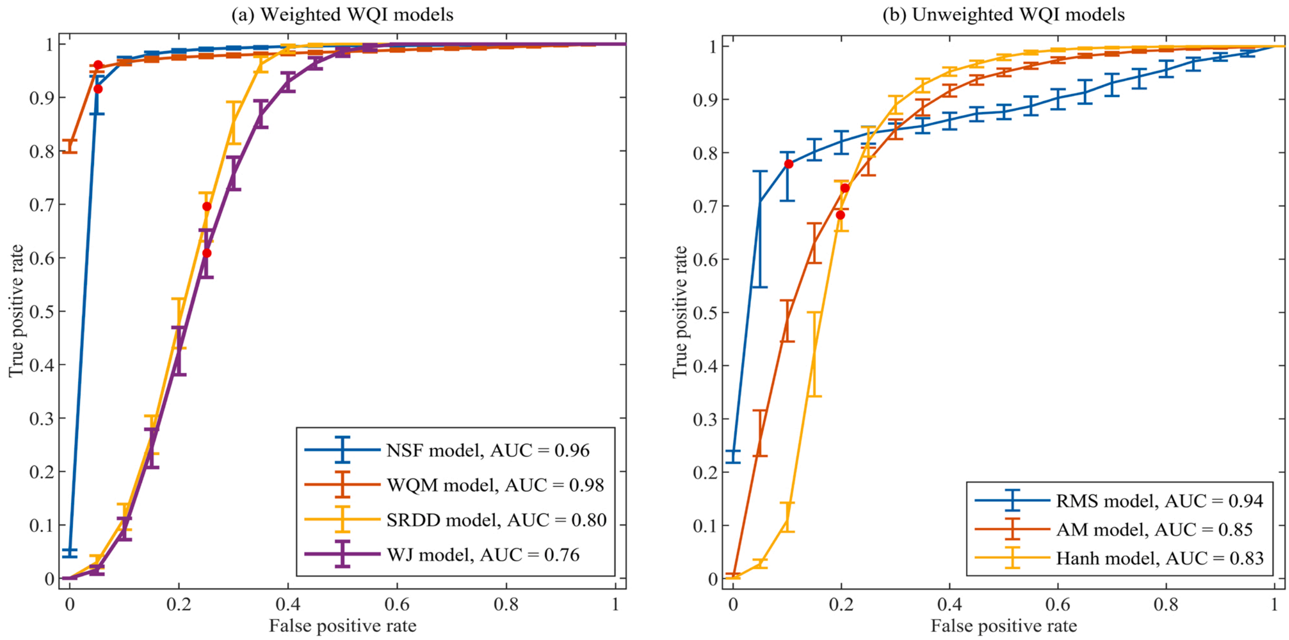

Fig. 16. ROC with point predictions of water quality classification using XGBoost of various WQI models with a 95% confidence interval at each observation. The red circle indicates the critical threshold values of ROC.

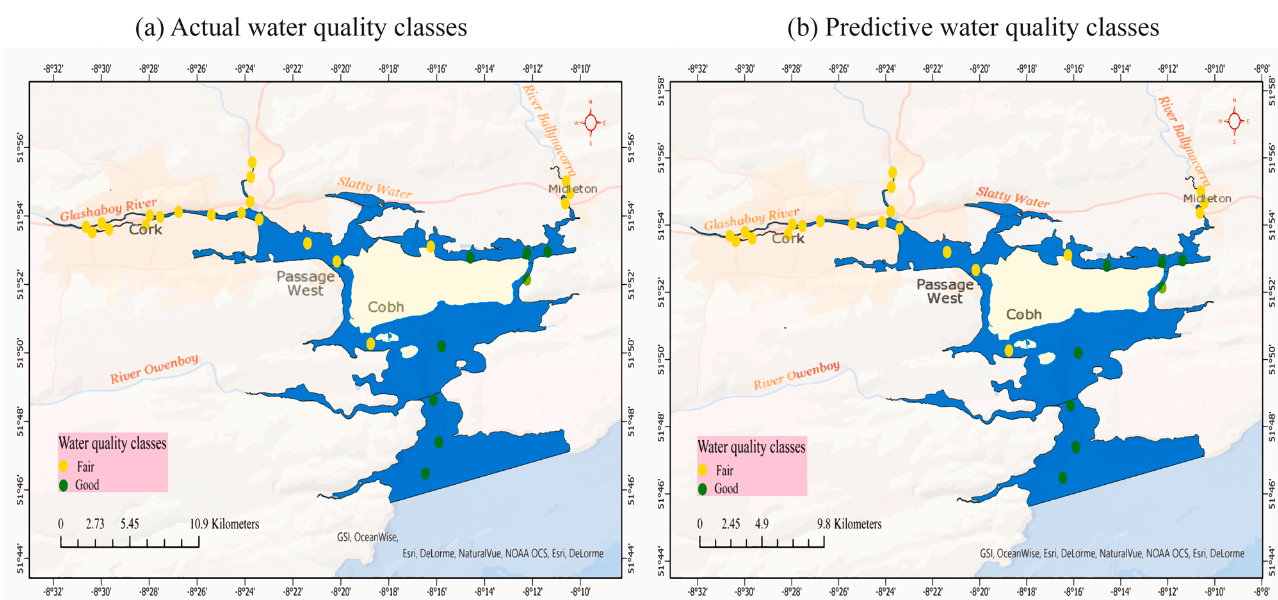

Fig. 17. Comparison between actual and predicted water quality classifications of WQM-WQI using the XGBoost predictive model.

4.1. Evaluation of water quality class in Cork Harbour using the XGBoost model

The ultimate goal of this study was to classify coastal water quality using the WQI model in terms of correct classification. Table S2 (supplementary material 1) provides the details of calculated WQI values and water quality classes for monitoring sites in Cork Harbour. In this study, the lowest error and higher predictive accuracy were found for the weighted WQM and unweighted RMS models. The results of predictive models recommended for the WQM and RMS index models could be classified effectively using the XGBoost model. In this section, Cork Harbour water quality from 29 monitoring sites was assessed using the WQM-WQI and RMS-WQI respectively, and then water quality classes were predicted using XGBoost. Fig. 17 and Fig. 18 shows the comparison results of water quality classes between WQI (actual class) and XGBoost models (predictive class) in Cork Harbour. The WQM model assessed "good" water quality at 27.6% (8) of the monitoring sites, whereas "fair" water quality was found for 72.4% (21) of the monitoring sites in the Harbour (Fig. 17a). On the other hand, the RMS model assessed "good" water quality at 52% (15) of the monitoring sites, and "fair" water quality was assessed at 42% (14) of the monitoring sites in the harbour (Fig. 18a). Whereas, the XGBoost model predicted 100% correct classification of water quality in the Harbour for both WQI models (Figs. 17b and 18b). The comparison results indicate that the weighted WQM and unweighted RMS models could be effective for assessing coastal water quality in order to correct classification.

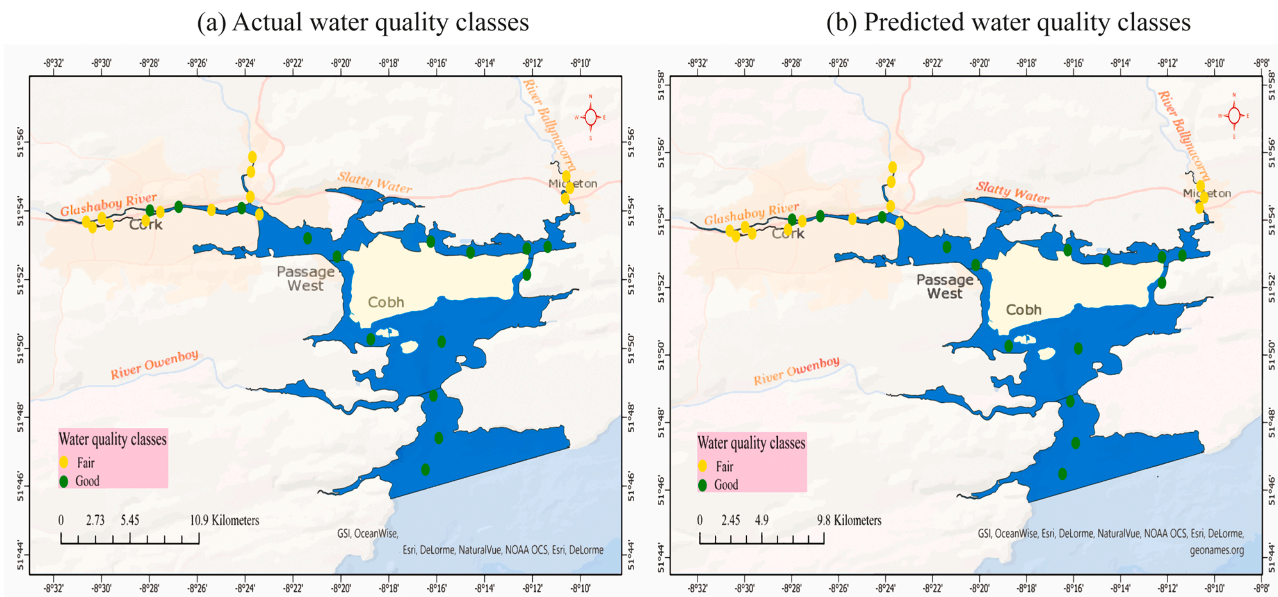

Fig. 18. Comparison between actual and predicted water quality classifications of RMS-WQI using the XGBoost predictive model.

5. Conclusion

The main aims of this study were to develop a framework for assessing performance of WQI model in order to correct classification of coastal water quality. Four machine-learning classifier algorithms were utilized to identify the best algorithm for predicting water quality class. The main summary of this study are as follows:

(i) Seven WQI models including four widely used and three newly proposed were assessed in this study
(ii) XGBoost algorithm and KNN were showed the outperformed in order to correct classification of water quality
(iii) XGBoost classified most water quality classes 100% correctly except "poor" classes.
(iv) In terms of WQI model(s) performance, the weighted WQM-WQI and unweighted RMS-WQI models could be effective for assessing coastal water quality status correctly.
(v) Both models were classified water quality into two classes including "Good" and "fair" water quality in Cork Harbour over the study period.

However, as best of our knowledge, this study provides the first comprehensive approach to evaluate the performance of WQI model(s) adopting new classification scheme for multi-class classification of coastal water quality. Moreover, the results of this study could be effective in obtaining the proper classification of water quality, which might be useful to improve the WQI model accuracy, transparency, and reliability in account of the correct classification of coastal water quality. The significant limitation of this research was that it did not consider the temporal variability of water quality indicators in Cork Harbour. Further studies should be carried out to assess WQI model(s) performance using temporal resolution of indicators, with other predictive classifier algorithm(s) included. However, in spite of its limitations, the findings of this study could useful for reducing the risk of model uncertainty due to inappropriate classification, which would provide insightful information to researchers, policymakers, and water research personnel.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors gratefully acknowledge the editor's and anonymous reviewers' contribution to the improvement of this paper. This research was funded by the Hardiman Research Scholarship of the University of Galway, which funded the first author as part of his PhD programme. The authors would like to acknowledge support from MaREI, the SFI Research Centre for Energy, Climate, and Marine research. The authors also would like to thank the Environmental Protection Agency of Ireland for providing water quality data. The authors sincerely acknowledge Charles Sturt University for providing all necessary supports to this PhD project through the excellent international co-supervision. Moreover, the authors also sincerely acknowledge the Eco HydroInformatics Research Group (EHIRG), School of Engineering, College of Science and Engineering, University of Galway, Ireland for providing computational laboratory facilities to complete this research.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.psep.2022.11.073.

References

Ahmed, U., Mumtaz, R., Anwar, H., Shah, A.A., Irfan, R., 2019, E ffi cient Water Quality Prediction Using Supervised 1–14.
Akbulut, Y., Sengur, A., Guo, Y., Smarandache, F., 2017. NS-k-NN: Neutrosophic set-based k-nearest neighbors classifier. Symmetry 9 (9). https://doi.org/10.3390/sym9090179.
Aldhyani, T.H.H., Al-Yaari, M., Alkahtani, H., Maashi, M., 2020. Water Quality Prediction Using Artificial Intelligence Algorithms. Appl. Bionics Biomech. 2020. https://doi.org/10.1155/2020/6659314.
Ali Haghpanah jahromi, & Mohammad Taheri. 2017. A non-parametric mixture of Gaussian naive Bayesclassifiers based on local independent features.
Allwein, E.L., 2000. Reducing multiclass to binary: a unifying approach for margin classifiers. J. Mach. Learn. Res. 1, 113–141.

Angelova, M., Pencheva, T., 2011. Tuning genetic algorithm parameters to improve convergence time. Int. J. Chem. Eng. https://doi.org/10.1155/2011/646917.
Asadollah, S.B.H.S., Sharafati, A., Motta, D., Yaseen, Z.M., 2021. River water quality index prediction and uncertainty analysis: A comparative study of machine learning models. J. Environ. Chem. Eng. 9, 104599 https://doi.org/10.1016/j.jece.2020.104599.
Awan, F.M., Saleem, Y., Minerva, R., Crespi, N., 2020. A comparative analysis of machine/deep learning models for parking space availability prediction. Sens. (Switz.) 20. https://doi.org/10.3390/s20010322.
Babbar, R., Babbar, S., 2017. Predicting river water quality index using data mining techniques. Environ. Earth Sci. 76, 1–15. https://doi.org/10.1007/s12665-017-6845-9.
Banchhor, C., Srinivasu, N., 2020. Integrating Cuckoo search-Grey wolf optimization and Correlative Naive Bayes classifier with Map Reduce model for big data classification. Data Knowl. Eng. 127 https://doi.org/10.1016/j.datak.2019.101788.
Bergstra, J., Ca, J.B., Ca, Y.B., 2012. Random search for hyper-parameter optimization Yoshua Bengio. J. Mach. Learn. Res. Vol. 13. ⟨http://scikit-learn.sourceforge.net⟩.
Bourel, M., Segura, A.M., 2018. Multiclass classification methods in ecology. Ecol. Indic. 85, 1012–1021. https://doi.org/10.1016/j.ecolind.2017.11.031.
Brown, C.A., Nelson, W.G., 2015. A method to identify estuarine water quality exceedances associated with ocean conditions. Environ. Monit. Assess. 187 https://doi.org/10.1007/s10661-015-4347-3.
Chamasemani, F.F., 2011, Multi-class Support Vector Machine (SVM) classifiers – An Application in Hypothyroid detection and Classification 353–358. ⟨https://doi.org/10.1109/BIC-TA.2011.51⟩.
Chen, K., Chen, H., Zhou, C., Huang, Y., Qi, X., Shen, R., Liu, F., Zuo, M., Zou, X., Wang, J., Zhang, Y., Chen, D., Chen, X., Deng, Y., Ren, H., 2020. Comparative analysis of surface water quality prediction performance and identification of key water parameters using different machine learning models based on big data. Water Res 171, 115454. https://doi.org/10.1016/j.watres.2019.115454.
Chen, R.C., Dewi, C., Huang, S.W., Caraka, R.E., 2020. Selecting critical features for data classification based on machine learning methods. J. Big Data 7. https://doi.org/10.1186/s40537-020-00327-4.
Cheryl A. Brown, Nelson, W.G., 2010, Approaches to Identify Exceedances of Water Quality Thresholds Associated with Ocean Conditions. Newport, USA.
Chia, S.L., Chia, M.Y., Koo, C.H., Huang, Y.F., 2022. Integration of advanced optimization algorithms into least-square support vector machine (LSSVM) for water quality index prediction. Water Supply 22 (2), 1951–1963. https://doi.org/10.2166/ws.2021.303.
Danades, A., Pratama, D., Anggraini, D., Anggriani, D., 2017, Comparison of accuracy level K-Nearest Neighbor algorithm and support vector machine algorithm in classification water quality status. Proc. 2016 6th Int. Conf. Syst. Eng. Technol. ICSET 2016 137–141. https://doi.org/10.1109/FIT.2016.7857553.
Dezfooli, D., Hosseini-Moghari, S.M., Ebrahimi, K., Araghinejad, S., 2018. Classification of water quality status based on minimum quality parameters: application of machine learning techniques. Model. Earth Syst. Environ. 4, 311–324. https://doi.org/10.1007/s40808-017-0406-9.
Elmachtoub, A.N., Liang, J.C.N., McNellis, R., 2020, Decision Trees for Decision-Making under the Predict-then-Optimize Framework. arXiv.
EPA, 2001. Parameters Of Water Quality: Interpretation and Standards. EnvironmentalProtection AGENCY. Ireland. The Environmental Protection Agency, Ireland,Ireland. 10.1017/CBO9781107415324.004.
EPA, 2016. Urban waste water treatment in 2015.
Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recognit. Lett. 27, 861–874. https://doi.org/10.1016/j.patrec.2005.10.010.
Florea, A.-C., Andonie, R., 2020. Weight. Random Search Hyperparameter Optim. https://doi.org/10.15837/ijccc.2019.2.3514.
Garabaghi, F.H., 2021, Performance Evaluation of Machine Learning Models with Ensemble Learning approach in Classi cation of Water Quality Indices Based on Different Subset of Features.
Gikas, G.D., Sylaios, G.K., Tsihrintzis, V.A., Konstantinou, I.K., Albanis, T., Boskidis, I., 2020. Comparative evaluation of river chemical status based on WFD methodology and CCME water quality index. Sci. Total Environ. 745, 140849 https://doi.org/10.1016/j.scitotenv.2020.140849.
Gonçalves, L., Subtil, A., Rosário Oliveira, M., De Zea Bermudez, P., 2014. ROC curve estimation: An overview. Revstat Stat. J. 12, 1–20.
Gupta, S., Gupta, S.K., 2021. A critical review on water quality index tool: Genesis, evolution and future directions. Ecol. Inform. 63, 101299 https://doi.org/10.1016/j.ecoinf.2021.101299.
Haghiabi, A.H., Nasrolahi, A.H., Parsaie, A., 2018. Water quality prediction using machine learning methods. Water Qual. Res. J. Can. 53, 3–13. https://doi.org/10.2166/wqrj.2018.025.
Hamadi, Y., Monfroy, E., Saubion, F., 2013. Autonomous search. In: Autonomous Search, Vol. 9783642214349. Springer-Verlag,, Berlin Heidelberg. https://doi.org/10.1007/978-3-642-21434-9.
Hamel, L., 2011. Model Assessment with ROC Curves. Encycl. Data Warehous. Min., Second Ed. 1316–1323. https://doi.org/10.4018/9781605660103.ch204.
Hanoon, M.S., Ammar, A.M., Ahmed, A.N., Razzaq, A., Birima, A.H., Kumar, P., Sherif, M., Sefelnasr, A., El-Shafie, A., 2022. Application of Soft Computing in Predicting Groundwater Quality Parameters. Front. Environ. Sci. 10 https://doi.org/10.3389/fenvs.2022.828251.
Hartnett, M., Nash, S., 2015. An integrated measurement and modeling methodology for estuarine water quality management. Water Sci. Eng. 8, 9–19. https://doi.org/10.1016/j.wse.2014.10.001.
Hong, C.S., 2009. Optimal threshold from ROC and CAP curves. Commun. Stat. Simul. Comput. 38, 2060–2072. https://doi.org/10.1080/03610910903243703.
Hosmer, D.W., Lemeshow, S., 2004. Applied Logistic Regression, Second Edition. Applied Logistic Regression.
Hossain, S., Sarma, D., Chakma, R.J., 2020. Machine learning-based phishing attack detection. Int. J. Adv. Comput. Sci. Appl. 11, 378–388. https://doi.org/10.14569/IJACSA.2020.0110945.
Huang, C., Li, Y., Yao, X., 2020. A survey of automatic parameter tuning methods for metaheuristics. In: IEEE Transactions on Evolutionary Computation, Vol. 24. Institute of Electrical and Electronics Engineers Inc.,, pp. 201–216. https://doi.org/10.1109/TEVC.2019.2921598.
Islam Khan, M.S., Islam, N., Uddin, J., Islam, S., Nasir, M.K., 2021. Water quality prediction and classification based on principal component regression and gradient boosting classifier approach. J. King Saud. Univ. - Comput. Inf. Sci. https://doi.org/10.1016/j.jksuci.2021.06.003.
Japkowicz, N., 2000, Learning from imbalanced data sets: a comparison of various strategies. AAAI Work. Learn. from Imbalanced Data Sets 0–5.
Juwana, I., Muttil, N., Perera, B.J.C., 2016. Uncertainty and sensitivity analysis of West Java Water Sustainability Index - A case study on Citarum catchment in Indonesia. Ecol. Indic. 61, 170–178. https://doi.org/10.1016/j.ecolind.2015.08.034.
Kaur, A., Khurana, M., Kaur, P., Kaur, M., 2021. Classification and Analysis of Water Quality Using Machine Learning Algorithms, Lecture Notes in Electrical Engineering. Springer,, Singapore. https://doi.org/10.1007/978-981-33-4866-0_48.
Kavzoglu, T., Colkesen, I., 2009. A kernel functions analysis for support vector machines for land cover classification. Int. J. Appl. Earth Obs. Geoinf. 11, 352–359. https://doi.org/10.1016/j.jag.2009.06.002.
Kavzoglu, T., Teke, A., 2022. Advanced hyperparameter optimization for improved spatial prediction of shallow landslides using extreme gradient boosting (XGBoost). Bull. Eng. Geol. Environ. 81 (5) https://doi.org/10.1007/s10064-022-02708-w.
Khoi, D.N., Quan, N.T., Linh, D.Q., Nhi, P.T.T., Thuy, N.T.D., 2022. Using Machine Learning Models for Predicting the Water Quality Index in the La Buong River. Vietnam. Water (Basel) 14, 1552. https://doi.org/10.3390/w14101552.
Khullar, S., Singh, N., 2021. Machine learning techniques in river water quality modelling: A research travelogue. Water Sci. Technol. Water Supply 21. https://doi.org/10.2166/ws.2020.277.
Kienzle, W., Schölkopf, B., 2005. LNAI 3720 - Training Support Vector Machines with Multiple Equality Constraints.
Kurt, I., Ture, M., Kurum, A.T., 2008. Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease. Expert Syst. Appl. 34, 366–374. https://doi.org/10.1016/j.eswa.2006.09.004.
Latha, C.B.C., Jeeva, S.C., 2019. Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. Inform. Med. Unlocked 16. https://doi.org/10.1016/j.imu.2019.100203.
Latinopoulos, D., Spiliotis, M., Ntislidou, C., Kagalou, I., Bobori, D., Tsiaoussi, V., Lazaridou, M., 2021. "One Out–All Out" Principle in the Water Framework Directive 2000—A New Approach with Fuzzy Method on an Example of Greek Lakes. Water (Switzerland) 13. https://doi.org/10.3390/w13131776.
Leong, W.C., Bahadori, A., Zhang, J., Ahmad, Z., 2021. Prediction of water quality index (WQI) using support vector machine (SVM) and least square-support vector machine (LS-SVM). Int. J. River Basin Manag. 19 (2), 149–156. https://doi.org/10.1080/15715124.2019.1628030.
Macskassy, S. a, Provost, F., Rosset, S., Heights, Y., 2005. Pointwise ROC Confidence Bounds : An Empirical Evaluation. New York.
Malek, N.H.A., Yaacob, W.F.W., Nasir, S.A.M., Shaadan, N., 2022. Prediction of Water Quality Classification of the Kelantan River Basin, Malaysia, Using Machine Learning Techniques. Water (Switz.) 14. https://doi.org/10.3390/w14071067.
Mandrekar, J.N., 2010. Receiver operating characteristic curve in diagnostic test assessment. J. Thorac. Oncol. 5, 1315–1316. https://doi.org/10.1097/JTO.0b013e3181ec173d.
Mehrpooya, A., Saberi-Movahed, Farid, Azizizadeh, N., Rezaei-Ravari, M., Saberi-Movahed, Farshad, Eftekhari, M., Tavassoly, I., 2021, High Dimensionality Reduction by Matrix Factorization for Systems Pharmacology. ⟨https://doi.org/10.1101/2021.05.30.446301⟩.
Modaresi, F., Araghinejad, S., 2014. A comparative assessment of support vector machines, probabilistic neural networks, and K-nearest neighbor algorithms for water quality classification. Water Resour. Manag 28, 4095–4111. https://doi.org/10.1007/s11269-014-0730-z.
Mohammed, H., Longva, A., Seidu, R., 2018. Predictive analysis of microbial water quality using machine-learning algorithms. Environ. Res. Eng. Manag. 74, 7–20. https://doi.org/10.5755/j01.erem.74.1.20083.
Morrison, A.M., Coughlin, K., Shine, J.P., Coull, B.A., Rex, A.C., 2003. Receiver operating characteristic curve analysis of beach water quality indicator variables. Appl. Environ. Microbiol. 69, 6405–6411. https://doi.org/10.1128/AEM.69.11.6405-6411.2003.
Muhammad, S.Y., Makhtar, M., Rozaimee, A., Aziz, A.A., Jamal, A.A., 2015. Classification model for water quality using machine learning techniques. Int. J. Softw. Eng. its Appl. 9, 45–52. https://doi.org/10.14257/ijseia.2015.9.6.05.
Najafzadeh, M., Ghaemi, A., 2019. Prediction of the five-day biochemical oxygen demand and chemical oxygen demand in natural streams using machine learning methods. Environ. Monit. Assess. 191 https://doi.org/10.1007/s10661-019-7446-8.
Najafzadeh, M., Niazmardi, S., 2021. A novel multiple-kernel support vector regression algorithm for estimation of water quality parameters. Nat. Resour. Res. 30, 3761–3775. https://doi.org/10.1007/s11053-021-09895-5.
Najafzadeh, M., Ghaemi, A., Emamgholizadeh, S., 2019. Prediction of water quality parameters using evolutionary computing-based formulations. Int. J. Environ. Sci. Technol. 16, 6377–6396. https://doi.org/10.1007/s13762-018-2049-4.

Najafzadeh, M., Homaei, F., Farhadi, H., 2021. Reliability assessment of water quality index based on guidelines of national sanitation foundation in natural streams: integration of remote sensing and data-driven models. Artif. Intell. Rev. 54, 4619–4651. https://doi.org/10.1007/s10462-021-10007-1.

Nasir, N., Kansal, A., Alshaltone, O., Barneih, F., Sameer, M., Shanableh, A., Al-Shamma'a, A., 2022. Water quality classification using machine learning algorithms. J. Water Process. Eng. 48, 102920. https://doi.org/10.1016/J.JWPE.2022.102920.

Neha Radhakrishnan, & Anju S. Pillai, 2020, Comparison of Water Quality Classification Models using Machine Learning.

Parween, S., Siddique, N.A., Mahammad Diganta, M.T., Olbert, A.I., Uddin, Md.G., 2022. Assessment of urban river water quality using modified NSF water quality index model at Siliguri city, West Bengal, India. Environ. Sustain. Indic. 16, 100202 https://doi.org/10.1016/j.indic.2022.100202.

Piccialli, V., Sciandrone, M., 2022. Nonlinear optimization and support vector machines. Ann Oper Res 314, 15–47. https://doi.org/10.1007/s10479-022-04655-x.

Polikar, R., 2012. Ensemble Machine Learning. Ensemble Mach. Learn. https://doi.org/10.1007/978-1-4419-9326-7.

Prakash, R., Tharun, V.P., Renuga Devi, S., 2018. A comparative study of various classification techniques to determine water quality. In: Proc. Int. Conf. Inven. Commun. Comput. Technol, 2018. ICICCT,, pp. 1501–1506. https://doi.org/10.1109/ICICCT.2018.8473168.

Prato, S., la Valle, P., de Luca, E., Lattanzi, L., Migliore, G., Morgana, J.G., Munari, C., Nicoletti, L., Izzo, G., Mistri, M., 2014. The "one-out, all-out" principle entails the risk of imposing unnecessary restoration costs: A study case in two Mediterranean coastal lakes. Mar Pollut Bull 80, 30–40.

Qian, Y., Zhou, W., Yan, J., Li, W., Han, L., 2015. Comparing machine learning classifiers for object-based land cover classification using very high resolution imagery. Remote Sens. 7 (1), 153–168. https://doi.org/10.3390/rs70100153.

Radhakrishnan, N., Pillai, A.S., 2020. Comparison of water quality classification Models using machine learning. Proc. 5th Int. Conf. Commun. Electron. Syst. ICCES 2020, 1183–1188. https://doi.org/10.1109/ICCES48766.2020.09137903.

Ratick, S., Schwarz, G., 2009. Monte Carlo Simulation. Int. Encycl. Hum. Geogr. 3, 175–184. https://doi.org/10.1016/B978-008044910-4.00476-4.

Saberi-Movahed, Farshad, Mohammadifard, Mahyar, Mehrpooya, A., Rezaei-Ravari, M., Berahmand, K., Rostami, M., Karami, S., Najafzadeh, M., Hajinezhad, D., Jamshidi, M., Abedi, F., Mohammadifard, Mahtab, Farbod, E., Safavi, F., Dorvash, M., Mottaghi-Dastjerdi, N., Vahedi, S., Eftekhari, M., Saberi-Movahed, Farid, Alinejad-Rokny, H., Band, S.S., Tavassoly, I., 2022. Decoding clinical biomarker space of COVID-19: Exploring matrix factorization-based feature selection methods. Comput. Biol. Med 146. https://doi.org/10.1016/j.compbiomed.2022.105426.

Savira, F., Suharsono, Y., 2013. A classification model for water quality analysis using decision tree. J. Chem. Inf. Model 01, 1689–1699.

Serqueira, M., González, P., & Bezerra, E., 2020, A Population-based Hybrid Approach to Hyperparameter Optimization for Neural Networks. ⟨http://arxiv.org/abs/2011.11062⟩.

Shakhari, S., Banerjee, I., 2019. A multi-class classification system for continuous water quality monitoring. Heliyon 5, e01822. https://doi.org/10.1016/j.heliyon.2019.e01822.

Singh, K.P., Basant, N., Gupta, S., 2011. Support vector machines in water quality management. Anal. Chim. Acta 703, 152–162. https://doi.org/10.1016/j.aca.2011.07.027.

Steele, T.D., 1987. Stratégies de la surveillance de la qualité des eaux. Hydrol. Sci. J. 32, 133–141. https://doi.org/10.1080/02626668709491178.

Strobl, R.O., Robillard, P.D., 2008. Network design for water quality monitoring of surface freshwaters: A review. J. Environ. Manag. 87, 639–648. https://doi.org/10.1016/j.jenvman.2007.03.001.

Sutadian, A.D., Muttil, N., Yilmaz, A.G., Perera, B.J.C., 2018. Development of a water quality index for rivers in West Java Province, Indonesia. Ecol. Indic. 85, 966–982. https://doi.org/10.1016/j.ecolind.2017.11.049.

Suwadi, N.A., Derbali, M., Sani, N.S., Lam, M.C., Arshad, H., Khan, I., Kim, K. il, 2022. An Optimized Approach for Predicting Water Quality Features Based on Machine Learning. Wirel. Commun. Mob. Comput. 2022. https://doi.org/10.1155/2022/3397972.

Szabo, P., Genge, B., 2020. Hybrid hyper-parameter optimization for collaborative filtering. In: Proceedings - 2020 22nd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, 2020. SYNASC,, pp. 210–217. https://doi.org/10.1109/SYNASC51798.2020.00042.

Talabani, H., Avci, E., 2019. Impact of Various Kernels on Support Vector Machine Classification Performance for Treating Wart Disease. In: International Conference on Artificial Intelligence and Data Processing, IDAP 2018. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/IDAP.2018.8620876.

Tanha, J., Abdi, Y., Samadi, N., Razzaghi, N., Asadpour, M., 2020. Boosting methods for multi-class imbalanced data classification: an experimental review. J. Big Data 7. https://doi.org/10.1186/s40537-020-00349-y.

Tesoriero, A.J., Gronberg, J.A., Juckem, P.F., Miller, M.P., Austin, B.P., 2017. Predicting redox-sensitive contaminant concentrations in groundwater using random forest classification. Water Resour. Res. https://doi.org/10.1002/2016WR020197.

Thanh Noi, P., Kappas, M., 2017. Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. Sens. (Basel, Switz.) 18 (1). https://doi.org/10.3390/s18010018.

The MathWorks, I., 1993, Statistics and Machine Learning ToolboxTM User's Guide R2021b, User's Guide.

Tiyasha, Tung, T.M., Yaseen, Z.M., 2021. Deep Learning for Prediction of Water Quality Index Classification: Tropical Catchment Environmental Assessment. Nat. Resour. Res. 30, 4235–4254. https://doi.org/10.1007/s11053-021-09922-5.

Toner, P., Bowman, J., Clabby, K., Lucey, J., Mc, McGarrigle, M., Concannon, C., Clenaghan, C., Cunningham, P.J.D., O'Boyle, S., MacCárthaigh, M., Craig, M., Quinn, R., 2005. Water Qual. Irel. 2001–2003.

Uddin, G., Nash, S., Rahman, A., Olbert, A.I., 2022f, Assessing optimization techniques for improving water quality model. Submitted for publication.

Uddin, Galal, Nash, S., Olbert, A.I., 2022c, Optimization of Parameters in a Water Quality Index Model using Principal Component Analysis, in: Proceedings of the 39th IAHR World Congress. Proceedings of the 39th IAHR World Congress.

Uddin, M.G., Moniruzzaman, M., Khan, M., 2017. Evaluation of Groundwater Quality Using CCME Water Quality Index in the Rooppur Nuclear Power Plant Area, Ishwardi, Pabna, Bangladesh. Am. J. Environ. Prot. https://doi.org/10.12691/env-5-2-2.

Uddin, M.G., Stephen, Nash, Olbert, A.I., 2020. Application of water quality index models to an Irish Estuary. Civ. Environ. Res. 576–581.

Uddin, M.G., Nash, S., Olbert, A.I., 2021. A review of water quality index models and their use for assessing surface water quality. Ecol. Indic. 122, 107218 https://doi.org/10.1016/j.ecolind.2020.107218.

Uddin, M.G., Nash, S., Mahammad Diganta, M.T., Rahman, A., Olbert, A.I., 2022b. Robust machine learning algorithms for predicting coastal water quality index. J. Environ. Manag. 321, 115923 https://doi.org/10.1016/j.jenvman.2022.115923.

Uddin, M.G., Nash, S., Diganta, M.T.M., Rahman, A., Olbert, A.I., 2022e. A comparison of geocomputational models for validating geospatial distribution of water quality index. In: Priyanka, H., Rahman, A., Agarwal, Basant, Tiwari, Binita (Eds.), Computational Statistical Methodologies and Modeling for Artificial Intelligence. CRC Press, Taylor & Francis Publisher, USA.

Uddin, Md. Galal, Nash, S., Rahman, A., Olbert, A.I., 2022a. A comprehensive method for improvement of water quality index (WQI) models for coastal water quality assessment. Water Res 219, 118532. https://doi.org/10.1016/j.watres.2022.118532.

Uddin, Md.G., Nash, S., Rahman, A., Olbert, A.I., 2022d. A novel approach for estimating and predicting uncertainty in water quality index model using machine learning approaches. Water Res., 119422 https://doi.org/10.1016/j.watres.2022.119422.

Unal, I., 2017. Defining an optimal cut-point value in ROC analysis: An alternative approach. Comput Math Methods Med 2017. https://doi.org/10.1155/2017/3762651.

Venkata Vara Prasad, D., Senthil Kumar, P., Venkataramana, L.Y., Prasannamedha, G., Harshana, S., Jahnavi Srividya, S., Harrinei, K., Indraganti, S., 2021. Automating water quality analysis using ML and auto ML techniques. Environ. Res. 202. https://doi.org/10.1016/j.envres.2021.111720.

Victoria, A.H., Maragatham, G., 2021. Automatic tuning of hyperparameters using Bayesian optimization. Evol. Syst. 12 (1), 217–223. https://doi.org/10.1007/s12530-020-09345-2.

Walley, W.J., Džeroski, S., 1996. Biological Monitoring: a Comparison between Bayesian, Neural and Machine Learning Methods of Water Quality Classification. Environ. Softw. Syst. 229–240. https://doi.org/10.1007/978-0-387-34951-0_20.

Walter, S.D., 2005. The partial area under the summary ROC curve. Stat. Med. 24, 2025–2040. https://doi.org/10.1002/sim.2103.

Wang, J., Xu, J., & Wang, X., 2018, Combination of Hyperband and Bayesian Optimization for Hyperparameter Optimization in Deep Learning. ⟨http://arxiv.org/abs/1801.01596⟩.

Wu, J., Chen, S.P., Liu, X.Y., 2020. Efficient hyperparameter optimization through model-based reinforcement learning. Neurocomputing 409, 381–393. https://doi.org/10.1016/j.neucom.2020.06.064.

Yin, J., 2017. Using the ROC Curve to Measure Association and Evaluate Prediction Accuracy for a Binary Outcome. Biom. Biostat. Int. J. 5, 1–10. https://doi.org/10.15406/bbij.2017.05.00134.

Yuan, B., & Gallagher, M., 2005, A hybrid approach to parameter tuning in genetic algorithms. 2005 IEEE Congress on Evolutionary Computation, IEEE CEC 2005. Proceedings, 2, 1096–1103. https://doi.org/10.1109/cec.2005.1554813.

Zhang, W., Wu, C., Zhong, H., Li, Y., Wang, L., 2021. Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization. Geosci. Front. 12 (1), 469–477. https://doi.org/10.1016/j.gsf.2020.03.007.

Zou, Q., Xie, S., Lin, Z., Wu, M., Ju, Y., 2016. Finding the Best Classification Threshold in Imbalanced Classification. J. Big Data 5, 2–8. https://doi.org/10.1016/j.bdr.2015.12.001.