**WATER QUALITY INDEXING AND ANALYSIS**

**A COMPREHENSIVE PROJECT BY**

VAIBHAV WAHIE S SHREYAS AAROHI P KOTHARI

Artificial Intelligence Artificial Intelligence Artificial Intelligence

SRM IST SRM IST SRM IST

Chennai, India Chennai, India Chennai, India

[Va4924@srmist.edu.in](mailto:Va4924@srmist.edu.in) [sk1169@srmist.edu.in](mailto:sk1169@srmist.edu.in) [ap4361@srmist.edu.in](mailto:ap4361@srmist.edu.in)

**ABSTRACT**

Estimating water quality has been one of the significant challenges faced by the world in recent decades. This paper presents a water quality prediction model utilizing the principal component regression technique. Firstly, the water quality index (WQI) is calculated using the weighted arithmetic index method. Secondly, the principal component analysis (PCA) is applied to the dataset, and the most dominant WQI parameters have been extracted. Thirdly, to predict the WQI, different regression algorithms are used to the PCA output. Finally, the Gradient Boosting Classifier is utilized to classify the water quality status. The proposed system is experimentally evaluated on a Gulshan Lake-related dataset. The results demonstrate 95% prediction accuracy for the principal component regression method and 100% classification accuracy for the Gradient Boosting Classifier method, which show credible performance compared with the state-of-art models.

**INTRODUCTION**

Water is the most significant resource of life, crucial for supporting the life of most existing creatures and human beings. Living organisms need water with enough quality to continue their lives. There are certain limits of pollutions that water species can tolerate. Exceeding these limits affects the existence of these creatures and threatens their lives.

Water quality for industrial uses also requires different properties based on the specific industrial processes. Some of the low-priced resources of fresh water, such as ground and surface water, are natural water resources. However, such resources can be polluted by human/industrial activities and other natural processes.

Hence, rapid industrial development has prompted the decay of water quality at a disturbing rate. Furthermore, infrastructures, with the absence of public awareness, and less hygienic qualities, significantly affect the quality of drinking water [1]. In fact, the consequences of polluted drinking water are so dangerous and can badly affect health, the environment, and infrastructures. As per the United Nations (UN) report, about 1.5 million people die each year because of contaminated water-driven diseases.

**LITERATURE SURVEY**

This section demonstrated the existing literature survey. The author took the most common approaches to detect and classify the water quality, including [deep neural network](https://www.sciencedirect.com/topics/computer-science/deep-neural-network), [recurrent neural network](https://www.sciencedirect.com/topics/computer-science/recurrent-neural-network), neuro-fuzzy inference, and [support vector regression](https://www.sciencedirect.com/topics/computer-science/support-vector-regression).

For example, [Barzegar et al. (2020)](https://www.sciencedirect.com/science/article/pii/S1319157821001361" \l "b0040), applied a CNN-LSTM amalgam model to predict two water quality variables, named Dissolved Oxygen (DO) and chlorophyll-a. Results indicated that the CNN-LSTM amalgam model outperformed both the individual [CNN](https://www.sciencedirect.com/topics/computer-science/convolutional-neural-network) and [LSTM](https://www.sciencedirect.com/topics/computer-science/long-short-term-memory-networks) model and the [machine learning](https://www.sciencedirect.com/topics/computer-science/machine-learning) models such as SVR, [Decision Tree](https://www.sciencedirect.com/topics/computer-science/decision-trees). [Oladipo et al. (2021)](https://www.sciencedirect.com/science/article/pii/S1319157821001361" \l "b0150), compared two statistical methods, including Fuzzy Logic Inference (FLI) and WQI methods, for evaluating the water quality in the Ikare community, Nigeria. They found moderate and poor water quality conditions using FLI and WQI methods, respectively. They also found that the FLI method is superior to the WQI method because of the relationship between measured values and WQI standard values. For the estimation of dissolved oxygen in aquaculture, [Li et al. (2018)](https://www.sciencedirect.com/science/article/pii/S1319157821001361" \l "b0125), suggested a synthetic model by combining Sparse-autoencoder and long short-term memory networks (LSTM). Although both CNN-LSTM and Sparse-autoencoder-LSTM models showed excellent performance since they predicted only DO and chlorophyll, it may be challenging to deal with more water quality variables using such models. In another research, [Asadollah et al. (2021)](https://www.sciencedirect.com/science/article/pii/S1319157821001361" \l "b0030), applied an ensemble machine learning method called Extra Tree Regression (ETR) which combines multiple week learners such as decision tree to predict WQI values in Tsuen River, Hong Kong. They applied the [ETR method](https://www.sciencedirect.com/topics/computer-science/regression-method) on ten water quality variables. Results indicated that the ETR method achieved 98% prediction accuracy, which outperformed the other state-of-the-art models such as support vector regression and decision tree. Further, [Hameed et al. (2017)](https://www.sciencedirect.com/science/article/pii/S1319157821001361" \l "b0075), developed two neural artificial network techniques: a [radial basis function neural network](https://www.sciencedirect.com/topics/computer-science/radial-base-function-neural-network) (RBFNN) and a backpropagation neural network (BNN) to predict the WQI in the tropical region of Malaysia. The WQI was measured using sub-indices equations in this study ([Agamuthu and Victor, 2011](https://www.sciencedirect.com/science/article/pii/S1319157821001361" \l "b0010)). In both RBFNN and BNN strategies, the training is faster, but the prediction takes a long time, making the model slow. [Bui et al. (2020)](https://www.sciencedirect.com/science/article/pii/S1319157821001361" \l "b0045), proposed a hybrid [machine learning algorithm](https://www.sciencedirect.com/topics/computer-science/machine-learning-algorithm) by combining the random tree and bagging (BA-RT) technique. The BA-RT method achieved 94% prediction accuracy using a 10–fold cross-validation technique, outdoing 15 standalone and hybrid algorithms. A more comprehensive study into the application of machine learning methods for modeling river water quality was performed by [Rajaee et al. (2020)](https://www.sciencedirect.com/science/article/pii/S1319157821001361" \l "b0160), where they reviewed a total of 51 articles published from 2000 to 2016. According to this study, artificial [neural networks](https://www.sciencedirect.com/topics/computer-science/neural-network) and wavelet-neural networks were the most widely used methods for predicting water quality. Furthermore, [Samsudin et al. (2019)](https://www.sciencedirect.com/science/article/pii/S1319157821001361" \l "b0175), developed an artificial neural network. For this study, the most significant water quality parameters were found through a spatially [discriminant analysis](https://www.sciencedirect.com/topics/computer-science/discriminant-analysis) (SDA). But these studies can barely show 71% accuracy. In another research, [Yilma et al. (2018)](https://www.sciencedirect.com/science/article/pii/S1319157821001361" \l "b0220), applied an artificial neural network for predicting WQI in Ethiopia’s Akaki River. In this analysis, an artificial neural network with eight hidden layers and 15 hidden neurons predict WQI with more than 90% accuracy. Also, [Imani et al. (2021)](https://www.sciencedirect.com/science/article/pii/S1319157821001361" \l "b0090), applied an artificial neural network with a single hidden layer for predicting water quality resilience in São Paulo, Brazil. Applying neural networks to predict WQI required lots of water quality data, which is expensive and time-consuming. [Ho et al. (2019)](https://www.sciencedirect.com/science/article/pii/S1319157821001361" \l "b0080), applied a decision tree for classifying water quality status in Klang River, Malaysia. They considered three scenarios where they used six water quality variables in the first scenario. After that, in each procedure, they removed water quality parameters such as NH3-N, pH, and SS to evaluate the decision tree algorithm’s ability in different situations. They achieved 84.09%, 81.82%, and 77.27% [classification accuracy](https://www.sciencedirect.com/topics/computer-science/classification-accuracy) in each scenario, which is higher than the 75% classification accuracy benchmark. Besides, to predict the WQI, [Ahmed et al. (2019)](https://www.sciencedirect.com/science/article/pii/S1319157821001361" \l "b0020), used several supervised machine learning methods. They conducted their model on four water quality parameters. They found that by using [gradient boosting](https://www.sciencedirect.com/topics/computer-science/gradient-boosting) and [polynomial regression](https://www.sciencedirect.com/topics/computer-science/polynomial-regression), the WQI is more successfully predicted where a [multilayer perceptron](https://www.sciencedirect.com/topics/computer-science/multilayer-perceptron) classifies the water quality category more effectively. However, this study worked with fewer water quality parameters, but both proposed prediction and [classification models](https://www.sciencedirect.com/topics/computer-science/classification-models) did not show more than 75% accuracy. On the other hand, [Wang et al. (2017)](https://www.sciencedirect.com/science/article/pii/S1319157821001361" \l "b0190), applied support vector regression to predict WQI. More than 90% of accuracy was achieved in this analysis. In this study, 22 specimens of water quality were used, which makes the model computationally costly. [Li et al. (2019)](https://www.sciencedirect.com/science/article/pii/S1319157821001361" \l "b0130), proposed an amalgam model for the study of time-series water quality data by integrating a recurrent neural network with the Dempster-Shafer Theory (DST), where the RNN is capable of analyzing time-series data effectively to predict WQI and DST, which is a probability method used to amalgamate the outcome of RNNs. It can be challenging to predict WQI using RNN and DST since specialized handling of the data is required when fitting and testing the model. Besides, [Ahmed et al. (2019)](https://www.sciencedirect.com/science/article/pii/S1319157821001361" \l "b0025), proposed a neuro-fuzzy inference method based on a wavelet-de-noise technique to predict water quality parameters. Results indicated that this model outperformed the other [neural network model](https://www.sciencedirect.com/topics/computer-science/neural-network-model), such as [RBF](https://www.sciencedirect.com/topics/computer-science/radial-basis-function) and MLP. But the neuro-fuzzy inference method causes a curse of [dimensionality problem](https://www.sciencedirect.com/topics/computer-science/dimensionality-problem), which occurs when [high dimensional data](https://www.sciencedirect.com/topics/computer-science/high-dimensional-data) is analyzed and classified.

In summary, from the above studies, most of the current approaches were based on a [predictive model](https://www.sciencedirect.com/topics/computer-science/predictive-model) but did not provide any classification model. Most of the models showed less accuracy and used many water quality specimens. The proposed method is applied to address the limitations described in the current approaches above. Also, the proposed model gives a dynamic approach to use any number of water quality specimens.

**PROPOSED ARCHITECTURE**

Prediction of Water Quality Classification

In this section, some machine learning algorithms, namely, support vector machine (SVM), K-nearest neighbor (KNN), and Naive Bayes, have been used to predict the water quality classification.

Support Vector Machine (SVM) Model The SVM model was developed in 1995 by Corinna Cortes and Vapnik. It has several unique benefits in solving small samples, and nonlinear and high-dimensional pattern recognition. It can be extended to function in the simulation of other machine learning problems. It uses the hyperplane to separate the points of the input vectors and finds the needed coefficients. The best hyperplane is the line with the largest margin, which is meant the distance between the hyperplane and the nearest input objects. The input points defined in the hyperplane are called support vectors. In this work, the linear SVM model along with the Gaussian radial basis function (equation (17)) is used to classify the tested water samples based on their quality.

K(X,X′)=exp(−||X−X′||22σ2) (17)

where X and X′ represent the feature vectors of the input dataset and the ‖X − X′‖2 is the squared Euclidean distance between the two feature inputs. The σ is a free parameter.

K-Nearest Neighbor (K-NN) Model The K-NN algorithm is a basic classification and regression method. It is used to find the K values that are close to values in the training dataset. Most of these values belong to a certain class, and thus, tested data can be classified. The K value is used to find the closest points in the feature vectors, and the value should be unique. The following expression of the Euclidean distance function (Di) can be used.

Di=(x1−x2)+(y1−y2)2,−−−−−−−−−−−−−−−−−√ (18)

where x1, x2, y1, and y2 are the variables for input data.

Naive Bayes Model The Bayesian method uses the knowledge of probability statistics to predict and classify datasets. The Bayesian algorithm combines prior and posterior probabilities to avoid the supervisor's bias and the overfitting phenomenon of using sample information alone.

This Naive Bayes is a type of classification algorithms based on Bayes' theorem and the assumption of the independence of characteristic conditions. Attributes are assumed to be conditionally independent of each other when the target value is given. This method greatly simplifies the complexity of the Bayesian method.

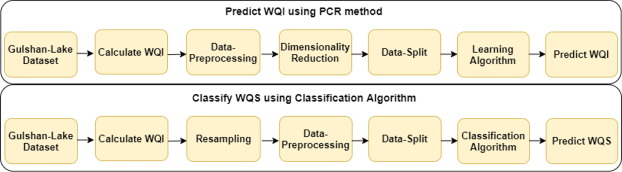
In Bayesian analysis, the probability of an event A given an event B is not the same as the probability of B given A as in equation (18).

P(A ∣ B)≠P(B ∣ A). (19)

Assuming that A1, A2 ⋯ .An and C are the feature vectors and the class of the WQC dataset, respectively, the Bayes equation can be expressed as follows:

P(C ∣ A)=P(C)×P(A ∣ C)P(A) (20)

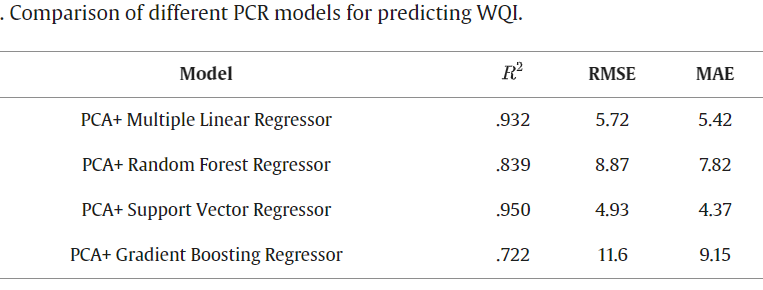
where the P(A) is a prior probability representing the feature vectors of the WQC dataset and P(A | C) is the prior probability of the class of the WQC dataset.



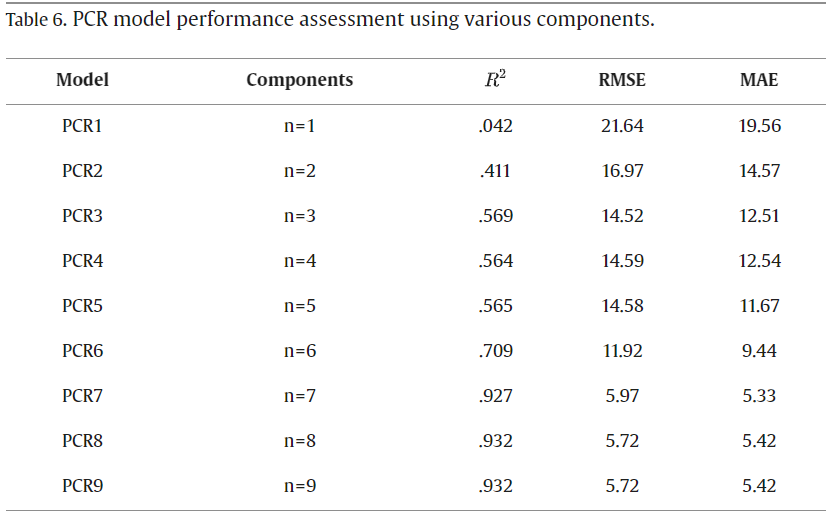
**RESULTS**

PCR model result assessment

The proposed PCR method implemented using python. The results of different PCR models showed in [Table 5](https://www.sciencedirect.com/science/article/pii/S1319157821001361" \l "t0025). From this table, PCA with Support Vector Regression has achieved the highest accuracy compared to the other PCR techniques. Although other PCR models also performed well, PCA with Gradient Boosting Regression proved to be a less useful model.

****

Since the PCR model provides to work with fewer parameters, so we reduced the number of components instead of taking all the features. The results of taking different features showed in [Table 6](https://www.sciencedirect.com/science/article/pii/S1319157821001361" \l "t0030). For this technique, PCA with Multiple linear regression is selected since PCA is mostly related to multiple linear regression to create new principal components. [Table 6](https://www.sciencedirect.com/science/article/pii/S1319157821001361#t0030) illustrated that, with nine and eight components, the PCR9 and PCR8 models showed the best performance, where PCR9 clarified all the variance. The PCR8 model gives the same result as the PCR9 model, and the number of parameters is also reduced. The R^2 value for the PCR8 model in testing steps is .932. If we reduce one more component from the PCR8 model, that model produced almost the same result as operating with all the components. The R^2 value in the PCR7 model is .927. After reducing one more component, the R^2 value reduced in the testing phases is .709. That shows less accuracy compared with the PCR7 and PCR8 models. Yet in the water samples, PCR6 still performed well. If we reduce more components from the PCR model, the R^2 value is barely 50 per cent, which shows low PCR model efficiency.

****

The accuracy comparison of the PCR model in each principal component showed in [Fig. 5](https://www.sciencedirect.com/science/article/pii/S1319157821001361" \l "f0025). It is evident from [Fig. 5](https://www.sciencedirect.com/science/article/pii/S1319157821001361#f0025) that the model performed well with six, seven and eight components. After then, it showed poor performance. Since PCR7 and PCR8 showed the same results as working with the PCR9, we could infer that the PCR method allows operating with fewer parameters instead of taking all the features.

[Fig. 4](https://www.sciencedirect.com/science/article/pii/S1319157821001361" \l "f0020) illustrated the plot between the observed and predicted WQI values for a better understanding of those models. Among them, the value appeared closer to the regression fit line in the PCA+ Support Vector Regression model because of the high training and testing accuracy.

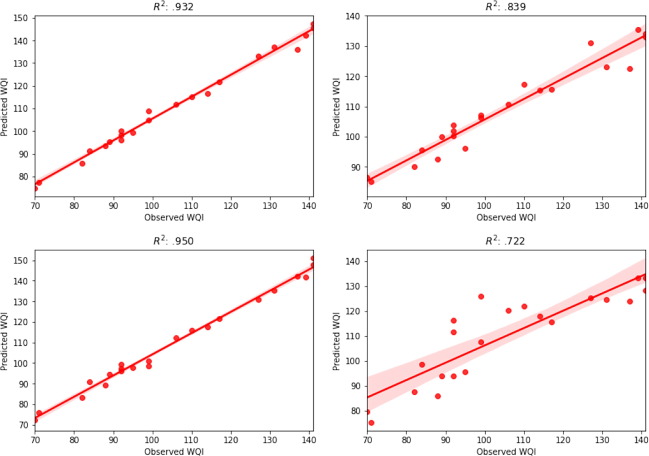
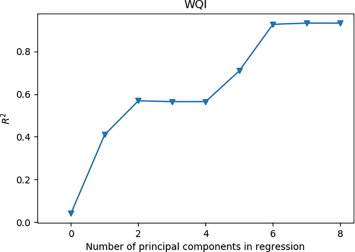
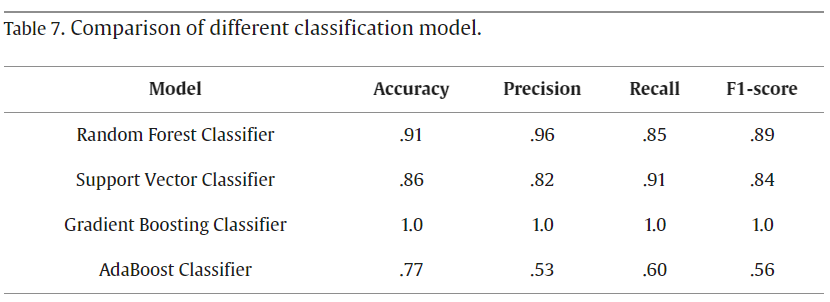


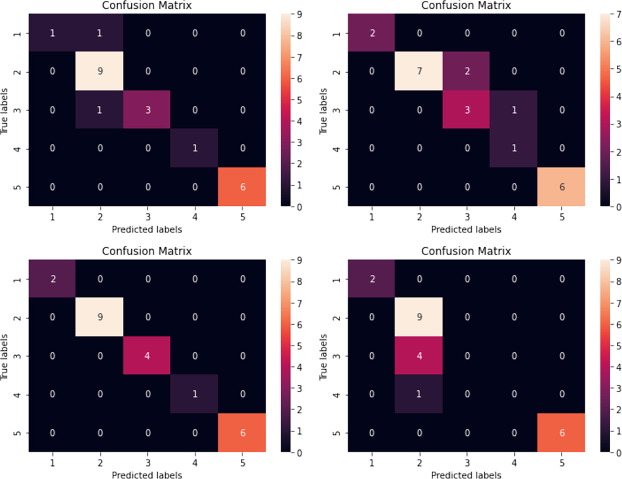
Fig. 4. Prediction and Classification of water quality indexPlot between observed and predicted WQI a. PCA+ Multiple Linear Regression model b. PCA+ [Random Forest](https://www.sciencedirect.com/topics/computer-science/random-decision-forest) Regression model c. PCA+ [Support Vector Regression](https://www.sciencedirect.com/topics/computer-science/support-vector-regression) model d. PCA+ [Gradient Boosting](https://www.sciencedirect.com/topics/computer-science/gradient-boosting) Regression model.



### Classification model result assessment

Different classification algorithms are implemented using python. The results of varying classification models are presented in [Table 7](https://www.sciencedirect.com/science/article/pii/S1319157821001361" \l "t0035). Among them, the Gradient Boosting Classifier has achieved the highest accuracy and proved to be an efficient model to predict water quality status. The second-best model is Random Forest Classifier, but to calculate recall, the Support Vector Classifier performs better than the Random Forest Classifier. Ada-Boost Classifier is found less effective model compared to the other techniques. The [confusion matrix](https://www.sciencedirect.com/topics/computer-science/confusion-matrix) for those models is presented in [Fig. 6](https://www.sciencedirect.com/science/article/pii/S1319157821001361" \l "f0030). From [Fig. 6](https://www.sciencedirect.com/science/article/pii/S1319157821001361#f0030), we can observe that the Gradient Boosting Classifier classify all the testing data according to the water quality level where other models misclassified some of the testing data.

****



**CONCLUSION**

This paper demonstrated a method for predicting and classifying the water quality using machine learning algorithms. The water metrics, including PH, DO, SS, EC, Turbidity, Chloride, COD, TDS, and Alkalinity, were used in this study. For [data preprocessing](https://www.sciencedirect.com/topics/computer-science/data-preprocessing), the median technique used to handle the null values and min–max scalar to scale the data. For the prediction purpose, we applied the principal component regression (PCR) method. After analyzing the performance of multiple PCR models, PCA with Support Vector Regression seems to be more effective with an accuracy of 95%. However, if the number of components reduced, then PCA with the Multiple Linear Regression model proved to be more effective. For the classification purpose, the Gradient Boosting classifier used to classify the water quality status. Besides, to check the performance of the model, the proposed model is compared with several state-of-art classifiers, including Ada-Boost Classifier, Support Vector Classifier, and Random Forest Classifier. Experimental results showed that the Gradient Boosting Classifier classified water quality status more efficiently. Despite the achievements outlined in this paper, some improvements are still possible, including we can collect more [training samples](https://www.sciencedirect.com/topics/computer-science/training-sample) to make the model more stable and more progress is possible on the prediction model. Those issues will be overcome in future research, perhaps by proper tuning of the PCR model and using deep neural network.

**REFERENCES**

1. [Adimalla, 2019](https://www.sciencedirect.com/science/article/pii/S1319157821001361#bb0005)

Narsimha Adimalla

Groundwater quality for drinking and irrigation purposes and potential health risks assessment: a case study from semi-arid region of South India

Exposure and Health, 11 (2) (2019), pp. 109-123

[View at publisher](javascript:;)

[CrossRef](https://doi.org/10.1007/s12403-018-0288-8)[View in Scopus](https://www.scopus.com/inward/record.url?eid=2-s2.0-85053729229&partnerID=10&rel=R3.0.0)[Google Scholar](https://scholar.google.com/scholar_lookup?title=Groundwater%20quality%20for%20drinking%20and%20irrigation%20purposes%20and%20potential%20health%20risks%20assessment%3A%20a%20case%20study%20from%20semi-arid%20region%20of%20South%20India&publication_year=2019&author=Narsimha%20Adimalla)

1. [Agamuthu and Victor, 2011](https://www.sciencedirect.com/science/article/pii/S1319157821001361#bb0010)

Pariatamby Agamuthu, Dennis Victor

Policy trends of extended producer responsibility in Malaysia

Waste Management & Research, 29 (9) (2011), pp. 945-953

[View in Scopus](https://www.scopus.com/inward/record.url?eid=2-s2.0-80052740384&partnerID=10&rel=R3.0.0)[Google Scholar](https://scholar.google.com/scholar_lookup?title=Policy%20trends%20of%20extended%20producer%20responsibility%20in%20Malaysia&publication_year=2011&author=Pariatamby%20Agamuthu&author=Dennis%20Victor)

1. [Aghel et al., 2019](https://www.sciencedirect.com/science/article/pii/S1319157821001361#bb0015)

B. Aghel, A. Rezaei, M. Mohadesi

Modeling and prediction of water quality parameters using a hybrid particle swarm optimization–neural fuzzy approach

International Journal of Environmental Science and Technology, 16 (8) (2019), pp. 4823-4832

[View at publisher](javascript:;)

[CrossRef](https://doi.org/10.1007/s13762-018-1896-3)[View in Scopus](https://www.scopus.com/inward/record.url?eid=2-s2.0-85049672455&partnerID=10&rel=R3.0.0)[Google Scholar](https://scholar.google.com/scholar_lookup?title=Modeling%20and%20prediction%20of%20water%20quality%20parameters%20using%20a%20hybrid%20particle%20swarm%20optimizationneural%20fuzzy%20approach&publication_year=2019&author=B.%20Aghel&author=A.%20Rezaei&author=M.%20Mohadesi)

1. [Ahmed et al., 2019](https://www.sciencedirect.com/science/article/pii/S1319157821001361#bb0020)

Ali Najah Ahmed, et al.

Machine learning methods for better water quality prediction

Journal of Hydrology, 578 (2019), Article 124084

[Google Scholar](https://scholar.google.com/scholar_lookup?title=Machine%20learning%20methods%20for%20better%20water%20quality%20prediction&publication_year=2019&author=Ali%20Najah%20Ahmed)

1. [Ahmed et al., 2019](https://www.sciencedirect.com/science/article/pii/S1319157821001361#bb0025)

Umair Ahmed, et al.

Efficient water quality prediction using supervised Machine Learning

Water, 11 (11) (2019), p. 2210

[View at publisher](javascript:;)

[CrossRef](https://doi.org/10.3390/w11112210)[View in Scopus](https://www.scopus.com/inward/record.url?eid=2-s2.0-85075581138&partnerID=10&rel=R3.0.0)[Google Scholar](https://scholar.google.com/scholar_lookup?title=Efficient%20water%20quality%20prediction%20using%20supervised%20Machine%20Learning&publication_year=2019&author=Umair%20Ahmed)

1. [Asadollah et al., 2021](https://www.sciencedirect.com/science/article/pii/S1319157821001361#bb0030)

Seyed Babak Asadollah, Haji Seyed, et al.

River water quality index prediction and uncertainty analysis: A comparative study of machine learning models

Journal of Environmental Chemical Engineering, 9 (1) (2021), Article 104599

[View PDF](https://www.sciencedirect.com/science/article/pii/S2213343720309489/pdfft?md5=e2568a568332fa9eb3a4472c2d55475e&pid=1-s2.0-S2213343720309489-main.pdf)[View article](https://www.sciencedirect.com/science/article/pii/S2213343720309489)[View in Scopus](https://www.scopus.com/inward/record.url?eid=2-s2.0-85097368119&partnerID=10&rel=R3.0.0)[Google Scholar](https://scholar.google.com/scholar_lookup?title=River%20water%20quality%20index%20prediction%20and%20uncertainty%20analysis%3A%20A%20comparative%20study%20of%20machine%20learning%20models&publication_year=2021&author=Seyed%20Babak%20Asadollah&author=Haji%20Seyed)

1. [Barzegar and Moghaddam, 2016](https://www.sciencedirect.com/science/article/pii/S1319157821001361#bb0035)

Rahim Barzegar, Asghar Asghari Moghaddam

Combining the advantages of neural networks using the concept of committee machine in the groundwater salinity prediction

Modeling Earth Systems and Environment, 2 (1) (2016), p. 26

[View at publisher](javascript:;)

This article is free to access.

[View in Scopus](https://www.scopus.com/inward/record.url?eid=2-s2.0-85018247043&partnerID=10&rel=R3.0.0)[Google Scholar](https://scholar.google.com/scholar_lookup?title=Combining%20the%20advantages%20of%20neural%20networks%20using%20the%20concept%20of%20committee%20machine%20in%20the%20groundwater%20salinity%20prediction&publication_year=2016&author=Rahim%20Barzegar&author=Asghar%20Asghari%20Moghaddam)

1. [Barzegar et al., 2020](https://www.sciencedirect.com/science/article/pii/S1319157821001361#bb0040)

Barzegar, Rahim, Mohammad Taghi, Aalami, Jan, Adamowski, 2020. Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model. Stochastic Environmental Research and Risk Assessment, pp. 1–19.

[Google Scholar](https://scholar.google.com/scholar?q=Barzegar%2C%20Rahim%2C%20Mohammad%20Taghi%2C%20Aalami%2C%20Jan%2C%20Adamowski%2C%202020.%20Short-term%20water%20quality%20variable%20prediction%20using%20a%20hybrid%20CNN%E2%80%93LSTM%20deep%20learning%20model.%20Stochastic%20Environmental%20Research%20and%20Risk%20Assessment%2C%20pp.%201%E2%80%9319.)

1. [Bui et al., 2020](https://www.sciencedirect.com/science/article/pii/S1319157821001361#bb0045)

Duie Tien Bui, et al.

Improving prediction of water quality indices using novel hybrid machine-learning algorithms

Science of The Total Environment, 721 (2020), Article 137612

[View PDF](https://www.sciencedirect.com/science/article/pii/S0048969720311232/pdfft?md5=760fa87cc95b3c18b9b2b16da5ef4870&pid=1-s2.0-S0048969720311232-mainext.pdf)[View article](https://www.sciencedirect.com/science/article/pii/S0048969720311232)[View in Scopus](https://www.scopus.com/inward/record.url?eid=2-s2.0-85080977331&partnerID=10&rel=R3.0.0)[Google Scholar](https://scholar.google.com/scholar_lookup?title=Improving%20prediction%20of%20water%20quality%20indices%20using%20novel%20hybrid%20machine-learning%20algorithms&publication_year=2020&author=Duie%20Tien%20Bui)

1. [Choubin, 2016](https://www.sciencedirect.com/science/article/pii/S1319157821001361#bb0050)

Bahram Choubin, et al.

Multiple linear regression, multi-layer perceptron network and adaptive neuro-fuzzy inference system for forecasting precipitation based on largescale climate signals

Hydrological Sciences Journal, 61 (6) (2016), pp. 1001-1009

[View at publisher](javascript:;)

[CrossRef](https://doi.org/10.1080/02626667.2014.966721)[View in Scopus](https://www.scopus.com/inward/record.url?eid=2-s2.0-84961390532&partnerID=10&rel=R3.0.0)[Google Scholar](https://scholar.google.com/scholar_lookup?title=Multiple%20linear%20regression%2C%20multi-layer%20perceptron%20network%20and%20adaptive%20neuro-fuzzy%20inference%20system%20for%20forecasting%20precipitation%20based%20on%20largescale%20climate%20signals&publication_year=2016&author=Bahram%20Choubin)

1. [Dataset, 2016](https://www.sciencedirect.com/science/article/pii/S1319157821001361#bb0055)

Dataset, Gulshan Lake, 2016. Published on May 20, 2018. URL: http://www.doe.gov.bd/site/publications/5132a8d7- 68e9 - 469d - a9af - 8981306b3b9f/ Surface - and - Ground - Water-Quality-Report-2016.

[Google Scholar](https://scholar.google.com/scholar?q=Dataset%2C%20Gulshan%20Lake%2C%202016.%20Published%20on%20May%2020%2C%202018.%20URL%3A%20http%3A%2F%2Fwww.doe.gov.bd%2Fsite%2Fpublications%2F5132a8d7-%2068e9%20-%20469d%20-%20a9af%20-%208981306b3b9f%2F%20Surface%20-%20and%20-%20Ground%20-%20Water-Quality-Report-2016.)

1. [Dzwairo et al., 2006](https://www.sciencedirect.com/science/article/pii/S1319157821001361#bb0060)

Bloodless Dzwairo, et al.

Assessment of the impacts of pit latrines on groundwater quality in rural areas: a case study from Marondera district, Zimbabwe

Physics and Chemistry of the Earth, Parts A/B/C, 31 (15-16) (2006), pp. 779-788

[View PDF](https://www.sciencedirect.com/science/article/pii/S1474706506001483/pdfft?md5=4bdc3171086e6598a9cb388083439166&pid=1-s2.0-S1474706506001483-main.pdf)[View article](https://www.sciencedirect.com/science/article/pii/S1474706506001483)[View in Scopus](https://www.scopus.com/inward/record.url?eid=2-s2.0-33749070374&partnerID=10&rel=R3.0.0)[Google Scholar](https://scholar.google.com/scholar_lookup?title=Assessment%20of%20the%20impacts%20of%20pit%20latrines%20on%20groundwater%20quality%20in%20rural%20areas%3A%20a%20case%20study%20from%20Marondera%20district%2C%20Zimbabwe&publication_year=2006&author=Bloodless%20Dzwairo)

1. [Ewaid et al., 2018](https://www.sciencedirect.com/science/article/pii/S1319157821001361#bb0065)

Salam Hussein Ewaid, Salwan Ali Abed, Safaa A. Kadhum

Predicting the Tigris River water quality within Baghdad, Iraq by using water quality index and regression analysis

Environmental Technology & Innovation, 11 (2018), pp. 390-398

[View PDF](https://www.sciencedirect.com/science/article/pii/S2352186418301548/pdfft?md5=5f357fd68156d02fef98af8ae49a87e6&pid=1-s2.0-S2352186418301548-main.pdf)[View article](https://www.sciencedirect.com/science/article/pii/S2352186418301548)[View in Scopus](https://www.scopus.com/inward/record.url?eid=2-s2.0-85050569161&partnerID=10&rel=R3.0.0)[Google Scholar](https://scholar.google.com/scholar_lookup?title=Predicting%20the%20Tigris%20River%20water%20quality%20within%20Baghdad%2C%20Iraq%20by%20using%20water%20quality%20index%20and%20regression%20analysis&publication_year=2018&author=Salam%20Hussein%20Ewaid&author=Salwan%20Ali%20Abed&author=Safaa%20A.%20Kadhum)

1. [Gaikwad et al., 2020](https://www.sciencedirect.com/science/article/pii/S1319157821001361#bb0070)

Satyajit Gaikwad, et al.

Geochemical mobility of ions in groundwater from the tropical western coast of Maharashtra, India: implication to groundwater quality

Environment, Development and Sustainability, 22 (3) (2020), pp. 2591-2624

[View at publisher](javascript:;)

[CrossRef](https://doi.org/10.1007/s10668-019-00312-9)[View in Scopus](https://www.scopus.com/inward/record.url?eid=2-s2.0-85060124719&partnerID=10&rel=R3.0.0)[Google Scholar](https://scholar.google.com/scholar_lookup?title=Geochemical%20mobility%20of%20ions%20in%20groundwater%20from%20the%20tropical%20western%20coast%20of%20Maharashtra%2C%20India%3A%20implication%20to%20groundwater%20quality&publication_year=2020&author=Satyajit%20Gaikwad)

1. [Hameed et al., 2017](https://www.sciencedirect.com/science/article/pii/S1319157821001361#bb0075)

Mohammed Hameed, et al.

Application of artificial intelligence (AI) techniques in water quality index prediction: a case study in tropical region, Malaysia

Neural Computing and Applications, 28 (1) (2017), pp. 893-905

[View at publisher](javascript:;)

[CrossRef](https://doi.org/10.1007/s00521-016-2404-7)[View in Scopus](https://www.scopus.com/inward/record.url?eid=2-s2.0-84976319262&partnerID=10&rel=R3.0.0)[Google Scholar](https://scholar.google.com/scholar_lookup?title=Application%20of%20artificial%20intelligence%20%20techniques%20in%20water%20quality%20index%20prediction%3A%20a%20case%20study%20in%20tropical%20region%2C%20Malaysia&publication_year=2017&author=Mohammed%20Hameed)

1. [Ho et al., 2019](https://www.sciencedirect.com/science/article/pii/S1319157821001361#bb0080)

Jun Yung Ho, et al.

Towards a time and cost effective approach to water quality index class prediction

Journal of Hydrology, 575 (2019), pp. 148-165

[View PDF](https://www.sciencedirect.com/science/article/pii/S0022169419304615/pdfft?md5=85d1c139c5023eaa93d0f7e8e7134a11&pid=1-s2.0-S0022169419304615-main.pdf)[View article](https://www.sciencedirect.com/science/article/pii/S0022169419304615)[View in Scopus](https://www.scopus.com/inward/record.url?eid=2-s2.0-85066089349&partnerID=10&rel=R3.0.0)[Google Scholar](https://scholar.google.com/scholar_lookup?title=Towards%20a%20time%20and%20cost%20effective%20approach%20to%20water%20quality%20index%20class%20prediction&publication_year=2019&author=Jun%20Yung%20Ho)

1. [Horton, 1965](https://www.sciencedirect.com/science/article/pii/S1319157821001361#bb0085)

Robert K. Horton

An index number system for rating water quality

Journal of Water Pollution Control Federation, 37 (3) (1965), pp. 300-306

[Google Scholar](https://scholar.google.com/scholar_lookup?title=An%20index%20number%20system%20for%20rating%20water%20quality&publication_year=1965&author=Robert%20K.%20Horton)

1. [Imani et al., 2021](https://www.sciencedirect.com/science/article/pii/S1319157821001361#bb0090)

Imani, Maryam, et al., 2021. A novel machine learning application: Water quality resilience prediction Model. Science of the Total Environment 768, 144459.

[Google Scholar](https://scholar.google.com/scholar?q=Imani%2C%20Maryam%2C%20et%20al.%2C%202021.%20A%20novel%20machine%20learning%20application%3A%20Water%20quality%20resilience%20prediction%20Model.%20Science%20of%20the%20Total%20Environment%20768%2C%20144459.)

1. [Kadam et al., 2019](https://www.sciencedirect.com/science/article/pii/S1319157821001361#bb0095)

A.K. Kadam, et al.

Prediction of water quality index using artificial neural network and multiple linear regression modelling approach in Shivganga River basin, India

Modeling Earth Systems and Environment, 5 (3) (2019), pp. 951-962

[View at publisher](javascript:;)

[CrossRef](https://doi.org/10.1007/s40808-019-00581-3)[View in Scopus](https://www.scopus.com/inward/record.url?eid=2-s2.0-85071007619&partnerID=10&rel=R3.0.0)[Google Scholar](https://scholar.google.com/scholar_lookup?title=Prediction%20of%20water%20quality%20index%20using%20artificial%20neural%20network%20and%20multiple%20linear%20regression%20modelling%20approach%20in%20Shivganga%20River%20basin%2C%20India&publication_year=2019&author=A.K.%20Kadam)

1. [Kar, 2013](https://www.sciencedirect.com/science/article/pii/S1319157821001361#bb0100)

Devashish Kar

Wetlands and Lakes of the World

Springer New Delhi, India (2013)

[Google Scholar](https://scholar.google.com/scholar_lookup?title=Wetlands%20and%20Lakes%20of%20the%20World&publication_year=2013&author=Devashish%20Kar)

1. [Kar, 2019](https://www.sciencedirect.com/science/article/pii/S1319157821001361#bb0105)

Kar, 2019. Wetlands and their Fish Diversity in Assam (India). Transylvanian Review of Systematical and Ecological Research 21 (3), 47–94.

[Google Scholar](https://scholar.google.com/scholar?q=Kar%2C%202019.%20Wetlands%20and%20their%20Fish%20Diversity%20in%20Assam%20(India).%20Transylvanian%20Review%20of%20Systematical%20and%20Ecological%20Research%2021%20(3)%2C%2047%E2%80%9394.)

1. [Khadr, 2017](https://www.sciencedirect.com/science/article/pii/S1319157821001361#bb0110)

Khadr, Mosaad, 2017. Modeling of water quality parameters in Manzala lake using adaptive neuro-fuzzy inference system and stochastic models. In: Egyptian Coastal Lakes and Wetlands: Part II. Springer, pp. 47–69.

[Google Scholar](https://scholar.google.com/scholar?q=Khadr%2C%20Mosaad%2C%202017.%20Modeling%20of%20water%20quality%20parameters%20in%20Manzala%20lake%20using%20adaptive%20neuro-fuzzy%20inference%20system%20and%20stochastic%20models.%20In%3A%20Egyptian%20Coastal%20Lakes%20and%20Wetlands%3A%20Part%20II.%20Springer%2C%20pp.%2047%E2%80%9369.)

1. [Kisi et al., 2019](https://www.sciencedirect.com/science/article/pii/S1319157821001361#bb0115)

Ozgur Kisi, et al.

Modeling groundwater quality parameters using hybrid neuro-fuzzy methods

Water Resources Management, 33 (2) (2019), pp. 847-861

[View at publisher](javascript:;)

[CrossRef](https://doi.org/10.1007/s11269-018-2147-6)[View in Scopus](https://www.scopus.com/inward/record.url?eid=2-s2.0-85058777342&partnerID=10&rel=R3.0.0)[Google Scholar](https://scholar.google.com/scholar_lookup?title=Modeling%20groundwater%20quality%20parameters%20using%20hybrid%20neuro-fuzzy%20methods&publication_year=2019&author=Ozgur%20Kisi)