1.Water Quality Prediction Using Artificial Intelligence Algorithms:

Author: Theyazn H. H Aldhyani,1**Mohammed Al-Yaari**,2Hasan Alkahtani,3and Mashael Maashi

#### Abstract

During the last years, water quality has been threatened by various pollutants. Therefore, modeling and predicting water quality have become very important in controlling water pollution. In this work, advanced artificial intelligence (AI) algorithms are developed to predict water quality index (WQI) and water quality classification (WQC). For the WQI prediction, artificial neural network models, namely nonlinear autoregressive neural network (NARNET) and long short-term memory (LSTM) deep learning algorithm, have been developed. In addition, three machine learning algorithms, namely, support vector machine (SVM), -nearest neighbor (K-NN), and Naive Bayes, have been used for the WQC forecasting. The used dataset has 7 significant parameters, and the developed models were evaluated based on some statistical parameters. The results revealed that the proposed models can accurately predict WQI and classify the water quality according to superior robustness. Prediction results demonstrated that the NARNET model performed slightly better than the LSTM for the prediction of the WQI values and the SVM algorithm has achieved the highest accuracy (97.01%) for the WQC prediction. Furthermore, the NARNET and LSTM models have achieved similar accuracy for the testing phase with a slight difference in the regression coefficient ( RNARNET = 96.17% and RLSTM = 94.27% ). This kind of promising research can contribute significantly to water management.

ADVANTAGES:

For validating the developed model, the dataset has been divided into 70% training and 30% testing subsets. While the ANN and LSTM models were used to predict the WQI, the SVM, KNN, and Naive Bayes were utilized for the water quality classification prediction. Modeling and prediction of water quality are very important for the protection of the environment. Developing a model by using advanced artificial intelligence algorithms can be used to measure the future water quality. In this proposed methodology, the advanced artificial intelligence algorithms, namely, NARNET and LSTM models were used to predict the WQI. Moreover, machine learning algorithms such as SVM, KNN, and Naive Bayes were used to classify the WQI data. The proposed models were evaluated and examined by some statistical parameters. For the WQI prediction, the result has revealed that the performance of the NARNET model is slightly better than the LSTM model based on the obtained R value. However, the SVM algorithm has achieved the highest accuracy of the prediction of the WQC as compared with KNN and Naive Bayes algorithms. After examining the robustness and efficiency of the proposed model for predicting the WQI, in future work, the developed models will be implemented to predict the water quality in Saudi Arabia for different types of water.

# 2. Research on the Application of Multimodal-Based Machine Learning Algorithms to Water Quality Classification

Authors: **Lei Xin**and Tianyu Mou

#### Abstract

With the development of society and the accelerated industrialization, the problem of water pollution has become increasingly prominent. In order to stop the gathering and diffusion of harmful substances in water bodies, leading to further deterioration of water quality and more serious environmental problems, environmental management departments have developed a series of pollutant discharge standards to prevent water pollution in real time. Common testing methods are the colorimetric method and TDS (total dissolved solids) value testing method, which are mostly through water bodies that contain acid, alkali, salt, and other indicators of the concentration test, to produce an assessment of water quality. However, the traditional methods of water quality testing, whether in the measurement time or in the accuracy of the test, are certain defects. In order to be able to quickly detect the concentration of water quality indicators in water bodies, timely response and treatment of highly polluted water bodies are urgently needed. In this paper, we propose a water quality detection classification model based on multimodal machine learning algorithm. Firstly, we preprocessed and analyzed the collected water quality dataset and determined the reasonable and perfect water quality classification influencing factors. Then, we successively built 15 kinds of classification models based on machine learning algorithms for water quality detection. At the same time, we evaluated the performance of each model. From the four evaluation indexes of precision, recall rate, F1 value, and accuracy, respectively, the real value is compared with the predicted value of each model. The experimental results show that sulfate, pH, solids, and hardness are the important influencing factors to perform water quality testing. And the three models XGBoost (Extreme Gradient Boosting), CatBoost (Categorical Boosting), and LGBM (Light Gradient Boosting Machine) have better performances in conducting water quality testing. Finally, we further optimized the classification models based on XGBoost, CatBoost, and LGBM by using two major tools: cross-validation and hyperparameter tuning.

ADVANTAGES:

In this paper, the open-source water quality test dataset with the attributes of pH value, hardness, total dissolved solids (TDS), chloramine content, sulfate content, conductivity, organic carbon, trihalomethane (THM), turbidity, and drinkability is studied. 15 mainstream machine learning water quality detection algorithms based on XGBoost, CatBoost, RF, Naive Bayes, and LGBM are designed, respectively, and the real values are compared with the predicted values of various models from the four evaluation indexes of accuracy, recall, F1 value, and accuracy. The experimental results show that sulfite, pH, solids, and hardness are important influencing factors for water quality detection. And XGBoost, CatBoost, and LGBM have good performance in water quality detection. Finally, we further optimize the classification model of water quality detection based on XGBoost, CatBoost, and LGBM by means of cross-validation and super parameter optimization. The widely used models have proved that XGBoost, CatBoost, and LGBM models have good treatment effects on water quality indicators and can be applied to water quality detection on a large scale.

# 3. Novel PSO Optimized Voting Classifier Approach for Predicting Water Quality

AUTHOR: **Shweta Agrawal**,1Sanjiv Kumar Jain,2Ajay Khatri,3Mohit Agarwal,4Anshul Tripathi,5and **Yu-Chen Hu**

#### Abstract

Over the last few years, different contaminants have posed a danger to the quality of the water. Hence modelling and forecasting water quality are very important in the management of water contamination. The paper proposes an ensemble machine learning-based model for assessing water quality. The results of the proposed model are compared with several machine learning models, including k-nearest neighbour, Naïve Bayes, support vector machine, and decision tree. The considered dataset contains seven statistically important parameters: pH, conductivity, dissolved oxygen, Biochemical Oxygen Demand, nitrate, total coliform, and fecal coliform. The water quality index is calculated for assessing water quality. To utilize an ensemble approach, a voting classifier has been designed with hard voting. The highest prediction accuracy of 99.5% of the water quality index is presented by the voting classifier as compared to the prediction accuracy of 99.2%, 90%, 79%, and 99% presented through k-nearest neighbour, Naïve Bayes, support vector machine, and decision tree, respectively. This was further enhanced to 99.74% using particle swarm based optimization.

ADVANTAGES:

Assessing water quality through machine learning will play a very important role in policymaking and water management. The paper presented an ensemble learning-based machine learning model for predicting water quality. For quality prediction, the water quality index has been calculated by considering many standard parameters. The performance of the ensemble-based voting classifier has been compared with other machine learning models like KNN, Naïve Bayes, SVM, and decision tree. Experimental results reveal that the highest presented accuracy is 99.5% which is through the voting classifier. The voting classifier model further improved using particle swarm optimization and this enhanced the accuracy to 99.74%.

The main contributions in this study are that a novel method is adopted to find the best set of parameters for the voting classifier, which is an ensemble of SVM, k-NN, DT, and Naïve Bayes. The test accuracy is improved to 99.74% using PSO. The limitation of this study is that the proposed algorithms are applied on fixed dataset and not considering the stochastic patterns. In terms of the improvement in this work, which also reflects the future work, it is using the some more latest metaheuristic optimization based ML predictions for different water quality indexes. Also, the present work may be utilized as the reference for further improvements in the water quality predictions utilizing the newest machine learning algorithms.

# 4. Performance Evaluation of Two ANFIS Models for Predicting Water Quality Index of River Satluj (India)

#### Abstract

Water quality index is the most convenient way of communicating water quality status of water bodies, but its evaluation requires subjectivity in terms of user involvement and dealing with uncertainty. Recently, artificial intelligence algorithms that are appropriate for nonlinear forecasting and also dealing with uncertainties have been applied to various domains of water quality forecasting. This paper focuses on development of a data-driven adaptive neurofuzzy system for the water quality index using a real data set obtained from eight different monitoring stations across River Satluj in northern India. Novelty in the paper lies in the estimation of water quality index using two different clustering techniques: fuzzy C-means and subtractive clustering-based ANFIS and assessing their predictive accuracy. Each model was used to train, validate, and test the index that was obtained from seven water quality parameters including pH, conductivity, chlorides, nitrates, ammonia, and fecal coliforms. The models were evaluated on the basis of statistical performance criteria. Based on the evaluations, it was found that the SC-ANFIS method gave more accurate result as compared to the FCM-ANFIS. The tested model, SC-ANFIS model, was further used to identify those sensitive parameters across various monitoring stations that were capable of causing change in the existing water quality index value.

ADVANTAGES:

In this study, two different clustering algorithms were used to develop the ANFIS model for water quality index prediction of River Satluj in northern India. The formulation for WQI of River Satluj is established in literature, and the index was computed by considering seven water quality parameters, crucial for municipal use. The data set for 16 years across eight monitoring stations on the river was used.

The two ANFIS models that were based on subtractive clustering and fuzzy c-means methods were trained, validated, and tested for modeling WQI. Based on the statistical evaluations, it was found that SC-ANFIS model predictions at training and testing stages were very close to the experimental value when compared to the FCM-ANFIS model.

The SC-ANFIS model, because of its good predictive capability over the FCM-ANFIS model, was further used to perform sensitivity analysis. The effect of perturbation in each water quality parameter was modeled and analyzed. The analyses showed that ammonia, chlorides, and fecal coliform were the most sensitive parameters that were capable of switching the existing water quality class to poorer class and hence warrant more attention in terms of their analysis for WQI computation.

The study reveals that ANFIS modeling with SC-ANFIS can be a useful approach to characterization of water quality in the form of water quality index. Since the approach obviates the otherwise lengthy computations of WQI, the present study holds its importance in developing a model and employing it for faster dissemination of information as well identifying the critical water quality parameters affecting WQI. The future scope of the work lies in the usage of a combination of hybrid SC-FCM and ANN (neurodynamic fuzzy expert system) to evaluate water quality.