

# **LITERATURE REVIEW**

## **EFFICIENT WATER QUALITY ANALYSIS AND PREDICTION**

### **1.EFFICIENT WATER QUALITY PREDICTION USING SUPERVISED MACHINE LEARNING**

Umair Ahmed, Rafia Mumtaz , Hirra Anwar, Asad A. Shah, Rabia Irfan and José García-Nieto (School of Electrical Engineering and Computer Science (SEECS), National University of Sciences and Technology (NUST), Islamabad 44000, Pakistan; Department of Languages and Computer Sciences, Ada Byron Research Building, University of Málaga)

**PUBLISHED: 24 OCTOBER 2019**

This research explores the methodologies that have been employed to help solve problems related to water quality. Typically, conventional lab analysis and statistical analysis are used in research to aid in determining water quality, while some analyses employ machine learning methodologies to assist in finding an optimized solution for the water quality problem. Local research employing lab analysis helped us gain a greater insight into the water quality problem in Pakistan. In one such research study, Daud et al. gathered water samples from different areas of Pakistan and tested them against different parameters using a manual lab analysis and found a high presence of E. coli and fecal coliform due to industrial and sewerage waste. Alamgir et al. tested 46 different samples from Orangi town, Karachi, using manual lab analysis and found them to be high in sulphates and total fecal coliform count. After getting familiar with the water quality research concerning Pakistan, we explored research employing machine learning methodologies in the realm of water quality. When it comes to estimating water quality using machine learning, Shafi et al. estimated water quality using classical machine learning algorithms namely, Support Vector Machines (SVM), Neural Networks (NN), Deep Neural Networks (Deep NN) and k Nearest Neighbors (kNN), with the highest accuracy of 93% with Deep NN. The estimated water quality in their work is based on only three parameters: turbidity, temperature and pH, which are tested according to World Health Organization (WHO) standards.

### **2. MACHINE LEARNING ALGORITHMS FOR EFFICIENT WATER QUALITY PREDICTION**

Mourade Azrour(Université Moulay Ismail), Ghizlane Fattah(Mohammadia School of Engineers), Azidine Guezaz(Cadi Ayyad University Morocco), Jamal Mabrouki(Mohammed V University of Rabat)

**PUBLISHED : 26 AUGUST 2021**

In this study, the advantages of machine learning algorithms are used to develop a model that is capable of predicting the water quality index and then the water quality class. The method we propose is based on four water parameters: temperature, pH, turbidity and coliforms. The use of the multiple regression algorithms has proven to be important and effective in predicting the water quality index. In addition, the adoption of the artificial neural network provides the most highly efficient way to classify the water quality.

### **3.WATER QUALITY PREDICTION BASED ON MACHINE LEARNING TECHNIQUES**

Zhao Fu(Bachelor of Information Security China University of Geosciences (Wuhan))

**PUBLISHED : 8 JANUARY 2020**

In this dissertation, several methods have been proposed to improve the performance of ANFIS-based water quality prediction models. Stratified sampling is employed to cover different kinds of data distribution in the training and testing datasets. The wavelet denoising technique is used to remove the noise hidden in the dataset. A deep prediction performance comparison between MLR, ANN, and ANFIS model is presented after stratified sampling and wavelet denoising techniques are applied. Because water quality data can be thought as a time series dataset, a time series analysis method is integrated with the ANFIS model to improve prediction performance. Lastly, intelligence algorithms are used to optimize the parameters of membership functions in the ANFIS model to promote the prediction accuracy. Experiments based on water quality datasets collected from Las Vegas Wash since 2007 and Boulder Basin of Lake Mead, Nevada, between 2011 and 2016 are used to evaluate the proposed models. Various ANN models have been designed to predict water and wastewater discharge quality based on previous existing datasets. A two-layer ANN model has been applied to predict the DO concentration in the Mathura River, and the experimental result showed that the ANN model worked well. An ANFIS model with eight input parameters is used to predict total phosphorus and total nitrogen, the experiment result based on 120 water samples. Time series analysis is also proposed to address dissolve oxygen prediction, and the experimental results show that the proposed analysis method can find out valuable knowledge from water quality historical timeseries data. In this , MLR, ANN, ANFIS, and FTS models are integrated with statistical analysis, wavelet denoising, and intelligence algorithm to explore the prediction of water quality.

### **4.GROUND WATER QUALITY PREDICTION USING MACHINE LEARNING ALGORITHMS IN R**

S.Vijay & 2Dr.K.Kamaraj Assistant Professor,Assistant ProfessorDepartment of Computer Science, 2Department of Computer Science Vivekanandha College of Arts and Sciences for Women, Tiruchengode, Namakkal Dt, SSM College of Arts and Science, Komarapalayam, Namakkal.

**PUBLISHED : MARCH 2019**

Sundarambal Palani proposed ANN models to predict water quality parameters whereas salinity, temperature, dissolved oxygen and Chl-a concentrations using continuous weekly measurements at different locations. Changjun Zhu proposed fuzzy neural network (FNN) model to evaluate and classify outer water quality in Suzhou. The FNN model is reliable and effective and can deal with the problem of solitary elements which reflects the water quality at current stage. Yafra Khan has developed a water quality forecast model using the support of water quality components applying Artificial Neural Network (ANN) and time-series analysis with ANN-NAR. The performance measures such as Regression, Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) indicated the best prediction accuracy results with ANN-NAR time series algorithm.

## **5. WATER QUALITY ANALYSIS USING MACHINE LEARNING ALGORITHMS**

Aleksei Shkurin (Bachelor's Thesis in Environmental Engineering)

**PUBLISHED : 29 MARCH 2016**

Random forests (RF) show the best performance and are advised for scientists and engineers working with environmental data. Artificial neural networks (ANN) are another alternative, though their performance is inferior and they are prone to overfitting. Support vector machines (SVM) are the good example for the cases where a baseline model is needed, being one of the basic algorithms. K-nearest neighbours (KNN) model was successfully used for data imputation and is also suggested for this task for other researchers. Clusterization techniques, such as k-means clustering, may assist data scientist with possible algorithms to classify given data, for example defining good, average and bad conditions of the water based on various chemical, biological and physical parameter models generated during this research may be used by IT students for producing software meant to help environmental specialists in analysing collected water quality data.

## **6. GROUND WATER QUALITY PREDICTION USING MACHINE LEARNING ALGORITHMS IN R**

S. Vijay (Assistant Professor, Department of Computer Science, Vivekanandha College of Arts and Sciences for Women, Tiruchengode, Namakkal Dt ) and Dr. K. Kamaraj (Assistant Professor, Department of Computer Science, SSM College of Arts and Science, Komarapalayam, Namakkal)

**PUBLISHED : 01 MARCH 2019**

Water plays a dominant role in the growth of the country's economy and essential for all the activities. The present study deals with the physico-chemical characteristics of ground water quality in Ranipet, Arcot, Walljah pet, towns in vellore district. Such a water samples were collected from different identified bore wells for the purpose of studying the quality of groundwater . The bore wells from which the samples were collected are extensively used for drinking purpose. The water quality parameters such as TDS, EC, Chloride, Sulphate, Nitrate, Carbonate, Bicarbonate, metal ions, trace elements have been estimated. There are two major classifications like High , Low level of water contamination observed in Vellore district. This paper focus on predicting water quality by using Machine Learning classifier algorithm C5.0, Naïve Bayes and Random forest as leaner for water quality prediction with high accuracy anf efficiency.

## **7.EFFICIENT PREDICTION OF WATER QUALITY INDEX (WQI) USING MACHINE LEARNING ALGORITHMS**

Md. Mehedi Hassan (Computer Science and Engineering, North Western University, Khulna, Bangladesh), Md. Mahedi Hassan(Computer Science and Engineering, Bangladesh University of Business and Technology, Dhaka, Bangladesh), Laboni Akter(Biomedical Engineering, Khulna University of Engineering & Technology, Khulna, Bangladesh), Md. Mushfiquir Rahman, Sadika Zaman, Khan Md. Hasib, Nusrat Jahan, Raisun Nasa Smrity, Jerin Farhana, M. Raihan, Swarnali Mollick

**PUBLISHED : 22 November 2021**

The quality of water has a direct influence on both human health and the environment. Water is utilized for a variety of purposes, including drinking, agriculture, and industrial use. The water quality index (WQI) is a critical indication for proper water management. The purpose of this work was to use machine learning techniques such as RF, NN, MLR, SVM, and BTM to categorize a dataset of water quality in various places across India. Water quality is dictated by features such as dissolved oxygen (DO), total coliform (TC), biological oxygen demand (BOD), Nitrate, pH, and electric conductivity (EC). These features are handled in five steps: data pre-processing using min-max normalization and missing data management using RF, feature correlation, applied machine learning classification, and model's feature importance. The highest accuracy Kappa, Accuracy Lower, and Accuracy Upper findings in this research are 99.83, 99.17, 99.07, and 99.99, respectively. The finding showed that Nitrate, PH, conductivity, DO, TC, and BOD are the key qualities that contribute to the orderly classification of water quality, with Variable Importance values of 74.78, 36.805, 81.494, 105.770, 105.166, and 130.173, respectively.

## **8.WATER QUALITY FACTOR PREDICTION USING SUPERVISED MACHINE LEARNINGMACHINE LEARNING**

Kathleen Joslyn(Portland State University)

**PUBLISHED : 2018**

The objective of this research is to explore prediction accuracy of water quality factors with techniques and algorithms in machine learning consisting of a variation of support vector machines - Support Vector Regression (SVR) and the gradient boosting algorithm Extreme Gradient Boosting (XGBoost). Both the XGBoost and SVR algorithms were used to predict nine different factors with success rates ranging from 79% to 99%. Parameters of these algorithms were also explored to test the prediction accuracy levels of individual water quality factors. These parameters included normalizing the data, filling missing data points, and training and testing on a large set of data.

## **9.WATER QUALITY PREDICTION AND CLASSIFICATION BASED ON PRINCIPAL COMPONENT REGRESSION AND GRADIENT BOOSTING CLASSIFIER APPROACH**

Md. SaikatIslam Khan(Department of Computer Science and Engineering, Santosh, Tangail-1902, Bangladesh) , NazrulIslam(Department of Information and Communication and Technology, Santosh, Tangail-1902, Bangladesh) , SifatulIslam Department of Technology Studies, Endicott College, Woosong University, Daejeon, South Korea , Mostofa KamalNasir(Mawlana Bhashani Science and Technology University, Santosh, Tangail 1902, Bangladesh)

**PUBLISHED : 3 JUNE 2021**

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.

# **10.PERFORMANCE OF MACHINE LEARNING METHODS IN PREDICTING WATER QUALITY INDEX BASED ON IRREGULAR DATA SET: APPLICATION ON ILLIZI REGION (Algerian southeast)**

Saber Kouadri, Ahmed Elbeltagi, Abu Reza Md. Towfiqul Islam & Samir Kateb

**PUBLISHED : 06 NOVEMBER 2021**

Groundwater quality appraisal is one of the most crucial tasks to ensure safe drinking water sources. Concurrently, a water quality index (WQI) requires some water quality parameters. Conventionally, WQI computation consumes time and is often found with various errors during subindex calculation. To this end, 8 artificial intelligence algorithms, e.g., multilinear regression (MLR), random forest (RF), M5P tree (M5P), random subspace (RSS), additive regression (AR), artificial neural network (ANN), support vector regression (SVR), and locally weighted linear regression (LWLR), were employed to generate WQI prediction in Illizi region, southeast Algeria. Using the best subset regression, 12 different input combinations were developed and the strategy of work was based on two scenarios. The first scenario aims to reduce the time consumption in WQI computation, where all parameters were used as inputs. The second scenario intends to show the water quality variation in the critical cases when the necessary analyses are unavailable, whereas all inputs were reduced based on sensitivity analysis. The models were appraised using several statistical metrics including correlation coefficient (R), mean absolute error (MAE), root mean square error (RMSE), relative absolute error (RAE), and root relative square error (RRSE). The results reveal that TDS and TH are the key drivers influencing WQI in the study area. The comparison of performance evaluation metric shows that the MLR model has the higher accuracy compared to other models in the first scenario in terms of 1,  $1.4572 \times 10^{-08}$ ,  $2.1418 \times 10^{-08}$ ,  $1.2573 \times 10^{-10\%}$ , and  $3.1708 \times 10^{-08\%}$  for R, MAE, RMSE, RAE, and RRSE, respectively. The second scenario was executed with less error rate by using the RF model with 0.9984, 1.9942, 3.2488, 4.693, and 5.9642 for R, MAE, RMSE, RAE, and RRSE, respectively. The outcomes of this paper would be of interest to water planners in terms of WQI for improving sustainable management plans of groundwater resources.