

**EFFICIENT WATER QUALITY ANALYSIS AND PREDICTION USING
MACHINE LEARNING**

(APPLIED DATA SCIENCE)

IBM PROJECT-TEAM ID : PNT2022TMID50914

TEAM LEAD

ARULVASAN M 953419106006

TEAM MEMBERS

JOHNMATHEW S 953419106027

GIFTEN B 953419106020

BALAJI G 953419106011

Of

BACHELOR OF ENGINEERING

IN

ELECTRONICS AND COMMUNICATION ENGINEERING

V V COLLEGE OF ENGINEERING

ANNA UNIVERSITY - CHENNAI 600 025

LITERATURE REVIEW

PREDICTION OF GROUNDWATER QUALITY USING EFFICIENT MACHINE LEARNING TECHNIQUE :

Sudhakar singh et.al [1], a deep learning (DL) based model is proposed for predicting groundwater quality and compared with three other machine learning (ML) models, namely, random forest (RF), eXtreme gradient boosting (XGBoost), and artificial neural network (ANN). This study can further be enhanced by examining the DL model's prediction ability against other ML models by taking into account various possible hydrogeometeorological input.

DATA ANALYSIS, QUALITY INDEXING AND PREDICTION WATER QUALITY

Ali and Qamar et al.[2] used the unsupervised technique of the average linkage (within groups) method of hierarchical clustering to classify samples into water quality classes. However, they ignored the major parameters associated with WQI during the learning process and they did not use any standardized water quality index to evaluate their predictions.

PREDICTING DISSOLVED OXYGEN FLUCTUATIONS IN GOLDEN HORN BY FUZZY TIME SERIES:

A water quality data is a kind of time series dataset which is likely to have complicated linear and non linear relationships.[3]

The experimental results showed that their proposed model significantly improved the prediction accuracy. However, there were only 520 water quality samples available to build the cloud, and thus, the model was not reliable or robust.

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 time series data.

PREDICTION OF WATER QUALITY INDEX USING MULTIVARIATE LINEAR REGRESSION (MLR):

A.H.Zare proposed a kind of statistical analysis method that estimates the target values based

on set of independent variables[4] .This model is used to measure the Biological Oxygen Demand (BOD) and Chemical Oxygen Demand (COD) based on the following four factors namely temperature, pH, total suspended solids and suspended solids.

This approach proposed that deterministic and multivariate linear regression models were used to speed up the process of predicting the water quality, the dataset is likely to have a non-linear relationship.

RIVER WATER QUALITY MODELLING USING ARTIFICIAL NEURAL NETWORK TECHNIQUE:

This approach proposes that a time series prediction model was integrated with the ANN model to improve the prediction performance [5]. A comprehensive comparison between ANN and Multivariate Linear Regression models in Biological Oxygen Demand (BOD) and Chemical Oxygen Demand (COD) prediction has shown that the ANN model is a better option.

The major disadvantage in this proposed model is that the input parameters are ambiguous and neural network struggle to formulate a non linear relationship in some scenarios.

APPLICATION ADAPTIVE NEURO FIZZY INFERENCE SYSTEM (ANFIS) TO ESTIMATE THE BIOLOGICAL OXYGEN DEMAND (BOD) AND CHEMICAL OXYGEN DEMAND (COD):

ANFIS proposed approach is used in predicting the effluent water quality. An ANFIS model with eight input parameters is used to predict total phosphorus and total nitrogen, the experiment result based on 120 water samples shows the proposed model is reliable. The ANFIS model has also been applied to estimate the biochemical oxygen demand in the Surma River. The testing results from 36 water samples confirmed that the ANFIS model could accurately formulate the hidden relationship and correlation analysis can improve the prediction accuracy. ANFIS model works better than the Artificial Neural Network model in predicting the Dissolved Oxygen content in the water sample to be tested

The disadvantage in this proposed models shows the correlation between the data in the dataset are weak then it generates out of range errors.

IMPROVING WATER QUALITY INDEX PREDICTION THROUGH A COMBINATION OF MULTIPLE NEURAL NETWORKS:

Ahmad et al.[7] employed single feed forward neural networks and a combination of multiple neural networks to estimate the Water Quality Index using a combination of backward

elimination and forward selection selective combination methods, they achieved an R squared error (RSE) and Mean Square Error (MSE) of 0.9270, 0.9390 and 0.1200, 0.1158, respectively. The use of parameters like temperature, turbidity, pH and total dissolved solids makes their solution a little immoderate in terms of an inexpensive real time system, given the price of the parameter sensor.

SURFACE WATER POLLUTION DETECTION USING INTERNET OF THINGS:

Shafi et al. [8] 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. Using only three parameters and comparing them to standardized values is quite a limitation when predicting water quality.

EFFICIENT WATER QUALITY PREDICTION USING SUPERVISED MACHINE LEARNING:

In order to estimate the water quality index (WQI), a unique index to characterise the general quality of water, and the water quality class (WQC), a distinct class established on the basis of the Water Quality Index, this research investigates a number of supervised machine learning techniques. Ahmed et al suggested methodology that uses temperature, turbidity, pH, and total dissolved solids as its four input parameters. Out of all the algorithms used, polynomial regression with a degree of 2 and gradient boosting with a learning rate of 0.

REFERENCES

1. Sudhakar Singha, Srinivas Pasupuleti, Soumya S. Singha, Rambabu Singh, Suresh Kumar, Prediction groundwater quality using efficient machine learning technique, Chemosphere, Volume 276,.
2. Ali, M. Qamar, A.M. Data analysis, quality indexing and prediction of water quality for the management of rawal watershed in Pakistan. In Proceedings of the Eighth International Conference on Digital Information Management (ICDIM 2013), Islamabad, Pakistan, 10–12 September 2013; pp. 108–113.
3. X. Sun ,Y. Li “Predicting Dissolved Oxygen Fluctuations In Golden Horn By Fuzzy Time Series” International Journal of Nonlinear Science, vol. 17, no.3, pp. 234-240, Dec. 2013
4. A. H. Zare, "Evaluation of multivariate linear regression and artificial neural networks in prediction of water quality parameters," Journal of Environmental Health Science & Engineering, vol. 12, no. 1, pp. 1-8, Jan. 2014.
5. A. Sarkar and P. Pandey, "River Water Quality Modelling Using Artificial Neural Network Technique," Aquatic Procedia, vol. 4, pp. 1070-1077, 2015
6. A. A. M. Ahmed and S. M. A. Shah, "Application of adaptive neuro-fuzzy inference system (ANFIS) to estimate the biochemical oxygen demand (BOD) of Surma River," Journal of King Saud University - Engineering Sciences, vol. 29, no. 3, pp. 237-243, Jul. 2017.
7. Ahmad, Z.; Rahim, N.; Bahadori, A.; Zhang, J. Improving water quality index prediction in Perak River basin Malaysia through a combination of multiple neural networks. Int. J. River Basin Manag. 2017, 15, 79–87.
8. Shafi, U.; Mumtaz, R.; Anwar, H.; Qamar, A.M.; Khurshid, H. Surface Water Pollution Detection using Internet of Things. In Proceedings of the 2018 15th International Conference on Smart Cities: Improving Quality of Life Using ICT & IoT (HONET-ICT), Islamabad, Pakistan, 8–10 October 2018; pp. 92–96.
9. Ahmed, U.; Mumtaz, R.; Anwar, H.; Shah, A.A.; Irfan, R.; García-Nieto, J. Efficient Water Quality Prediction Using Supervised Machine Learning. Water 2019, 11, 2210. <https://doi.org/10.3390/w11112210>