EFFICIENT WATER QUALITY ANALYSIS AND PREDICTION USING MACHINE LEARNING

LITERATURE SURVEY

Modeling the quality of water resources is vitally important for water scheduling and management. In the past, scientists regularly sampled the water in water quality monitoring stations and assessed the components in the water sample in a lab. However, this process takes a long time, and thus, the detected results are not timely. With the emergence of artificial intelligence (AI) techniques since the last decade, researchers have begun to adopt multivariate linear regression (MLR), artificial neural networks (ANN), adaptive neuro-fuzzy inference system (ANFIS), and Fuzzy time series (FTS) model to predict water quality by exploring the linear and non-linear relationships residing in water quality datasets. In addition, the wavelet denoising method and intelligent algorithms are also proposed to combine with machine learning techniques to enhance the prediction accuracy. In the following, we will review these related work in four categories of machine learning methods. MLR is a kind of statistical analysis method which is used to estimate the target value based on given values collected from a set of independent variables the MLR model is used to predict biochemical oxygen demand (BOD) and chemical oxygen demand base on four independent variables, temperature, pH, total suspended solid, and total suspended.

ANN 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. Various neural network types are compared in predicting water temperatures in streams. A radial basis function neural network has also been proposed to describe the water quality parameters in . The summary of the experiment result shows the model outperforms the linear regression model in conductivity, turbidity, and total dissolved solids prediction.. Additionally, a comprehensive comparison between ANN and MLR models in biochemical oxygen demand and chemical oxygen demand prediction has been performed . The experimental results show that a three-layer neural network model outperforms an MLR model.

ANFIS Many studies have proven that ANFIS, which can integrate linear and non-linear relationships hidden in the dataset, is a better option in this scenario. The experimental results show that an ANFIS model works much better than an ANN model in predicting dissolved oxygen, even though there are only 45 data samples available. 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. Two different kinds of ANFIS model, fuzzy c-means and subtractive clustering-based was compared in , the experiment result shows the ANFIS model built by fuzzy c-means provides more accurate prediction result. In , the ensemble models of wavelet ANNs are found to be superior to the best single model for forecasting chlorophyll and salinity concentrations in coastal water. . l. A few out-of-range errors can cause a very large prediction error, even though the model can accurately predict most of the data samples. In a dataset collected from 122 wells in Mashhad plain (Iran) is used to investigate the performance of ANFIS, ANN, and geostatistical models in groundwater quality prediction.

FTS A water quality data is a kind of time series dataset which is likely to have complicated linear and nonlinear relationships. The Fuzzy time series (FTS) model was first proposed by Song and Chissom in 1993 to address an enrollment prediction problem . A Heuristic Gaussian cloud transformation was integrated with an FTS model to forecast water quality. The experimental results showed that the 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 12 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 dissertation, MLR, ANN, ANFIS, and FTS models are integrated with statistical analysis, wavelet denoising, and intelligence algorithm to explore the prediction of water quality.

REFERENCE

Ahmed U, Mumtaz R, Anwar H et al (2019) Efficient water quality prediction using supervised machine learning. Water 11(11):2210. https://doi.org/10.3390/w11112210

Article **Google Scholar**

Aldhyani THH, Al-Yaari M, Alkahtani H, Maashi M (2020) Water quality prediction using artificial intelligence algorithms. Appl Bionics Biomech. https://doi.org/10.1155/2020/6659314

Article **Google Scholar**

Asadollah SBHS, Sharafati A, Motta D, Yaseen ZM (2021) River water quality index prediction and uncertainty analysis: a comparative study of machine learning models. J Environ Chem Eng 9(1):104599. https://doi.org/10.1016/j.jece.2020.104599

Article **Google Scholar**

Azrour M, Farhaoui Y, Ouanan M, Guezzaz A (2019) SPIT detection in telephony over IP using K-means algorithm. Proc Comput Sci 148:542–551. https://doi.org/10.1016/j.procs.2019.01.027

Article **Google Scholar**

Bekesiene S, Meidute-Kavaliauskiene I, Vasiliauskiene V (2021) Accurate prediction of concentration changes in ozone as an air pollutant by multiple linear regression and artificial neural networks. Mathematics 9(4):356. https://doi.org/10.3390/math9040356

Article **Google Scholar**

Ciulla G, D'Amico A (2019) Building energy performance forecasting: a multiple linear regression approach. Appl Energy 253:113500. https://doi.org/10.1016/j.apenergy.2019.113500

Article **Google Scholar**

Deng T, Chau K-W, Duan H-F (2021) Machine learning based marine water quality prediction for coastal hydro-environment management. J EnviroManag 284:112051. https://doi.org/10.1016/j.jenvman.2021.112051

Article **Google Scholar**

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. https://doi.org/10.1007/s40808-017-0406-9

Article **Google Scholar**

El Bilali A, Taleb A (2020) Prediction of irrigation water quality parameters using machine learning models in a semi-arid environment. J Saudi Soc Agric Sci 19(7):439–451. https://doi.org/10.1016/j.jssas.2020.08.001

Article **Google Scholar**

Ewaid SH (2017) Water quality evaluation of Al-Gharraf river by two water quality indices. Appl Water Sci 7(7):3759–3765

Article **Google Scholar**

Griffiths O, Henderson H, Simpson M (2010) Environmental Health Practitioner Manual: (Common wealth of Australia). https://ww2.health.wa.gov.au/~/media/Files/Corporate/general-documents/Environmental-health/Aboriginal-environmental-health/Environmental-Health-Practitioner-Manual.pdf. Accessed 10 Aug 2021