LITERATURE SURVEY

1. Machine learning algorithms for efficient water quality prediction: A Review of literature Jamal Mabrouki, Ghizlane Fattah, Azedine Guezzaz, Faissal Aziz, 2021

Water is an essential resource for human existence. In fact, more than 60% of the human body is made up of water. Our bodies consume water in every cell, in the different organisms and in the tissues. Therefore, to design a model that predicts water quality is nowadays very important to control water pollution, as well as to alert users in case of poor quality detection. 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. Water quality predicting is a key and primary task in the context of the environmental control strategy. Certainly, the accuracy of predictions will surely contribute significantly to more appropriate conservation of water resources. In this research paper, our goal is to suggest a model for prediction water quality based on machine learning algorithms and with minimal parameters. Machine learning is an analytical approach of data expected to make the analysis model more automatic.

2. Evaluation of *E. coli* in sediment for assessing irrigation water quality using machine learning

Fresh produce irrigated with contaminated water poses a substantial risk to human health. This study evaluated the impact of incorporating sediment information on improving the performance of machine learning models to quantify *E. coli* level in irrigation water. Field samples were collected from irrigation canals in the Southwest U.S., for which meteorological, chemical, and physical water quality variables as well as three additional flow and sediment properties: the concentration of *E. coli* in sediment, sediment median size, and bed shear stress. Water quality was classified based on *E. coli* concentration exceeding two standard levels: 1 *E. coli* and 126 *E. coli* colony forming units (CFU) per 100 ml of irrigation water. Two series of features, including (FIS) and excluding (FES) sediment features, were selected using multi-variant filter feature selection. The correlation analysis revealed the inclusion of sediment features improves the correlation with the target standards for *E. coli* compared to the models excluding these features. Support vector machine, logistic regression, and ridge classifier were tested in this study. The support vector machine model performed the best for both targeted standards. Besides, incorporating sediment features improved all models' performance.

3. machine learning in water quality evaluation : A Review of literature MengyuanZhu,JiaweiWang,XiaoYang,2022.

With the rapid increase in the volume of data on the aquatic environment, machine learning has become an important tool for data analysis, classification, and prediction. Unlike traditional models used in water-related research, data-driven models based on machine learning can efficiently solve more complex nonlinear problems. In water environment research, models and conclusions derived from machine learning have been applied to the construction, monitoring, simulation, evaluation, and optimization of various water treatment and management systems. Additionally, machine learning can provide solutions for water pollution control, water quality improvement, and watershed ecosystem security management. In this review, we describe the cases in which machine learning algorithms have been applied to evaluate the water quality in different water environments, such as surface water, groundwater, drinking water, sewage, and seawater. Furthermore, we propose possible future applications of machine learning approaches to water environments.

4. Efficient water quality prediction models based on machine learning: A Review of literature Nainital Lake, Uttarakhand

Water quality deterioration increases day by day in hilly areas due to increasing tourism activity, unplanned construction, disposal of solid waste, improper sewage management. With this idea, the work investigates different machine learning algorithms to evaluate the water quality index (WQI) and the water quality class (WQC). This paper utilizes Nainital Lake as a study area. The models used for testing and training comprise algorithms of machine learning for both binary and multiclass classification. In this paper, eight machine learning algorithms were employed for regression analysis, and nine machine learning algorithms were used for classification analysis. The result demonstrates that in regression analysis, the Random Forest algorithm comes out to be the most efficient Machine Learning algorithm. However, in the case of classification analysis, no single algorithm is good enough for prediction, three algorithms Stochastic Gradient Descent, Random Forest, and Support Vector Machine with the same accuracy proved to be efficient to predict water quality.

5.Prediction of groundwater quality using efficient machine learning technique

To ensure safe drinking water sources in the future, it is imperative to understand the quality and pollution level of existing groundwater. The prediction of water quality with high accuracy is the key to control water pollution and the improvement of water management. In this study, 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 (XG Boost), and artificial neural network (ANN). A total of 226 groundwater samples are collected from an agriculturally intensive area Arang of Raipur district, Chhattisgarh, India, and various physicochemical parameters are measured to compute entropy weight-based groundwater quality index (EWQI). Prediction performances of models are determined by introducing five error metrics. Results showed that DL model is the best prediction model with the highest accuracy in terms of R^2, i.e., R^2 = 0996 against the RF (R^2 = 0.886), XG Boost (R^2 = 0.0.927), and ANN (R^2 = 0.917). The uncertainty of the DL model output is cross-verified by running the proposed algorithm with newly randomized dataset for ten times, where minor deviations in the mean value of performance metrics are observed. Moreover, input variable importance computed by prediction models highlights that DL model is the most realistic and accurate approach in the prediction of groundwater quality.