**CHAPTER 2**

**LITERATURE SURVEY**

**TITLE:** Prediction of groundwater quality using efficient machine learning technique (2021)

**AUTHORS:** Lingbo Li, Jundong Qiao, Guan Yu, Leizhi Wang, Hong-Yi Li, Chen Liao, Zhenduo Zhu

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 (XGBoost), 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 R2, i.e., R2 = 0996 against the RF (R2 = 0.886), XGBoost (R2 = 0.0.927), and ANN (R2 = 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.

**TITLE:** Proposed formulation of surface water quality and modelling using gene expression, machine learning, and regression techniques(2021)

**AUTHORS:** M. I. Shah, M. F. Javed, and T. Abunama

Water pollution is an international concern that impedes human health, ecological sustainability, and agricultural output. This study focuses on the distinguishing characteristics of an evolutionary and ensemble machine learning (ML) based modelling to provide an in-depth insight of escalating water quality problems. The 360 temporal readings of electric conductivity (EC) and total dissolved solids (TDS) with several input variables are used to establish multi-expression programming (MEP) model and random forest (RF) regression model for the assessment of water quality at Indus River. New hydrological insight for the region: The developed models were evaluated using several statistical metrics. The findings reveal that the determination coefficient (R2) in the testing phase (subject to unseen data) for the all the developed models is more than 0.95, indicating the accurateness of the developed models. Furthermore, the error measurements are much lesser with root mean square logarithmic error (RMSLE) nearly equals to zero for each developed model. The mean absolute percent error (MAPE) of MEP models and RF models falls below 10% and 5%, respectively, in all three phases (training, validation and testing). According to the sensitivity study of generated MEP models about the relevance of inputs on the predicted EC and TDS, shows that bi-carbonates and chlorine content have significant influence with a sensitiveness score more than 0.90, whereas the impact of sodium content is less pronounced. All the models (RF and MEP) have lower uncertainty based on the prediction interval coverage probability (PICP) calculated using the quartile regression (QR) approach. The PICP% of each model is greater than 85% in all three stages. Thus, the findings of the study indicate that developing intelligent models for water quality parameter is cost effective and feasible for monitoring and analyzing the Indus River water quality.

**TITLE:** Interpretable tree-based ensemble model for predicting beach

water quality (2022)

**AUTHORS:** L. Li et al

Tree-based machine learning models based on environmental features offer low-cost and timely solutions for predicting microbial fecal contamination in beach water to inform the public of the health risk. However, many of these models are black boxes that are difficult for humans to understand, which may cause severe consequences such as unexplained decisions and failure in accountability. To develop interpretable predictive models for beach water quality, we evaluate five tree-based models, namely classification tree, random forest, CatBoost, XGBoost, and LightGBM, and employ a state-of-the-art explanation method SHAP to explain the models. When tested on the Escherichia coli (E. coli) concentration data collected from three beach sites along Lake Erie shores, LightGBM, followed by XGBoost, achieves the highest averaged precision and recall scores. For all three sites, both models suggest lake turbidity as the most important predictor, and elucidate the crucial role of accurate local data of wave height and rainfall in the model development. Local SHAP values further reveal the robustness of the importance of lake turbidity as its SHAP value increases nearly monotonically with its value and is minimally affected by other environmental factors. Moreover, we found an intriguing interaction between lake turbidity and day-of-year. This work suggests that the combination of LightGBM and SHAP has a promising potential to develop interpretable models for predicting microbial water quality in freshwater lakes.

**TITLE:** A Comparative Analysis of the Weighted Arithmetic and Canadian Council of Ministers of the Environment Water Quality Indices for Water Sources in Ohaozara, Ebonyi State, Nigeria (2021)

**AUTHORS:**  Chijioke K. Ojukwu, Gift O. Chukwu Okeah, Prince C. Mmom Department of Geography and Environmental Management University of Port Harcourt

This study compared the Weighted Arithmetic and Canadian Council of Ministers of the Environment water quality indices on water sources in Ohaozara, Ebonyi State, Nigeria. Water samples were collected from two communities of Uburu and Okposi in Ohaozara Local Government Area. Borehole sample was collected from Uburu while hand dug well sample was collected from Okposi. These samples were collected in the months of May and June 2021, presenting a total of four samples. The Weighted Arithmetic Water Quality Index (WAWQI) and the Canadian Council of Ministers of the Environment Water Quality Index (CCME-WQI) analytical methods were used in the determination of the water quality index (WQI). The results revealed that the WAWQI value for hand dug well in the month of May and June were 10.379 and 9.445 respectively, representing a water quality status of excellent for each month. However, the WAWQI value for the borehole sample in the month of May and June indicated 89.371 and 100.19 respectively, representing a water quality status of very poor and unfit for drinking in that order. The high WQI values of the borehole sample revealed the presence of toxic heavy metal contaminants of lead and arsenic in the water source as observed in the physicochemical analysis. The CCME-WQI value for the hand dug well sample indicated 53.513 while the CCME-WQI value for the borehole sample showed 49.668 with both water sources having a water quality ranking of marginal. This meant both water sources were frequently threatened and their conditions far from desired levels. The study revealed that while both indices provided a single number to describe water quality status and are widely utilised globally, the WAWQI is more sensitive to toxic heavy metal contaminants due to the use of unit weight calculation in the WQI determination. It showed that the WAWQI is better suited to be applied in the study area to alert consumers due to the abundance of solid mineral deposits in the study area and the potential for water pollution. Comparatively, the study also revealed that CCME-WQI is more moderate to all contaminants due to the use of scope (F1), Frequency (F2) and Amplitude (F3) calculation in the determination of the WQI which resulted to the marginal water quality ranking obtained for both the borehole and the hand dug well water samples.

**TITLE:** Water quality assessment in terms of water quality index (WQI) (2021)

**AUTHORS:** Wisam Thamer Al-Mayah1, Sattar Obaid Maiws Al-Mayyahi2 and Sarteel Hamid Al- Shammary2

This work deals with the monitoring and assessment of water quality of the Tigris River within Baghdad. Samples were taken monthly from September 2018 till August 2019 for a year, from eleven sites in Baghdad city. The National Sanitation Foundation Index (NSFWQI) values of river water deteriorated from “medium” to “bad” to “very bad” in almost all the eleven sampling sites. The water quality is found to be most deteriorate during the summer season with an average NSF-WQI value of 34.9 as compared to spring, winter and autumn seasons, having an average NSF-WQI value of 40.8,43.1and 44, respectively. Out of the eleven sampling sites, Al-Wathba site (S7) and Al-Rasheed site (S11) is observed to be the most polluted sites. The metal pollution index (MI) model is categorized the water quality of the Tigris as seriously affected where the Iron (Fe) and Lead (Pb), are prominent parameters and most deteriorated in this model. Based on these indices, it is concluded that industrial facilities, city wastewater and intensives communities that living along the river bank are negatively

affecting the water quality of the Tigris River.

**TITLE:** Modelling and Prediction of Water Quality by Using Artificial

Intelligence (2021)

**AUTHORS:** Mosleh Hmoud Al-Adhaileh,

With fast economic growth and increased urbanization, water pollution has become grimmer. Understanding the issues and patterns of water quality is also critical for water pollution reduction and regulation. Most countries around the world have started to develop environmental water management schemes to truly understand the quality of the marine ecosystem. Water is life’s most important substance. Although 71% of the Earth’s surface is covered with water, the vast majority of it (95%) is salt water. Thus, conserving the quality of fresh water is essential. Almost one billion people do not have access to adequate drinking water sources, and two million people die every year from contaminated water and poor sanitation and hygiene. With advanced computing using artificial intelligence (AI) techniques, the modelling of water quality has been developed to resolve water quality issues. Artificial neural networks (ANNs) have aided in the monitoring of water quality systems by predicting changes in water quality. They can immensely improve the efficiency of aquaculture. The simulation of water quality conditions has difficulties and challenges regarding the use of the hydrodynamic and water quality model, a relatively novel computational approach. ANNs have been widely established in many disciplines and provide an alternative technique for understanding and monitoring water quality in reservoirs. ANNs have been successfully applied to simulate and forecast water quality in water bodies. Numerous ANN methods, such as feed-forward neural networks, have been used in various applications. The fuzzy logic system has been developed to solve complex nonlinear systems. ANN applications have been successfully used as tools to compute and predict the quality of water bodies. ANN models require parameter values for designing predictions. ANNs have numerous advantages, including their ability to learn, manage very complex nonlinear systems, and work with parallel processing.

**TITLE:** A survey on applications of machine learning algorithms in water quality assessment and water supply and management (2023)

**AUTHORS:** Abdulhalık Oğuz, Ömer Faruk Ertuğrula

Managing water resources and determining the quality of surface and groundwater is one of the most significant issues fundamental to human and societal well-being. The process of maintaining water quality and managing water resources well involves complications due to human-induced errors. Therefore, applications that facilitate and enhance these processes have gained importance. In recent years, machine learning techniques have been applied successfully in the preservation of water quality and the management and planning of water resources. Water researchers have effectively used these techniques to integrate them into public management systems. In this study, data sources, pre-processing, and machine learning methods used in water research are briefly mentioned, and algorithms are categorized. Then, a general summary of the literature is presented on water quality determination and applications in water resources management. Finally, the study was detailed using machine learning investigations on two publicly shared datasets.