LITERATURE SURVEY

S.NO	RESEARCH PAPER
1.	Efficient Water Quality Prediction Using Supervised Machine Learning
2.	The Application of Machine Learning in Water Quality Evaluation
3.	Comparative Assessment of Individual and Ensemble Machine Learning Models for Efficient Analysis of River Water Quality

Survey Done By:

Ananth S S

Ambareeshkumar K

Manikandan P

Karthikeyan S

TITLE: Efficient Water Quality Prediction Using Supervised Machine Learning

AUTHOR: Umair Ahmed , Rafia Mumtaz, Hirra Anwar, Asad A. Shah, Rabia Irfan 1 and José García-Nieto

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. [5] 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. [6] 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. [7] 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 (Available online at URL https://www.who.int/airpollution/guidelines/en/). Using only three parameters and comparing them to standardized values is quite a limitation when predicting water quality

LINK:https://www.researchgate.net/publication/336808732_Efficient_Water_Quality_Prediction_Using_Supervised_Machine_Learning

TITLE: The Application of Machine Learning in Water Quality Evaluation

AUTHOR: Mengyuan ZhuJiawei WangXiao YangYu ZhangLinyu ZhangHongqiang RenBing WuLin Ye

Municipal and industrial wastewater generated by human activities has become the main factor in deteriorating water quality in urban areas [15]. The application of machine learning in surface water quality research has become a hotspot [16,17]. A series of surface water quality prediction and analysis methods have been developed Many efforts have been devoted to optimizing machine learning models and improving their prediction accuracy. Data acquisition is a fundamental step in developing machine learning models. Both integrated and periodic water quality monitoring results can be used as benchmarks in water system management. Traditional environmental monitoring methods are widely applied by environmental agencies. However, for in situ monitoring, the traditional methods are limited by realistic difficulties [34]. Remote sensing technologies can meet the needs of real-time and large-scale water quality monitoring, and can also be used to reveal the migration and distribution characteristics of pollutants that are difficult to detect using conventional methods. Sagan et al. [29] found that experiment-based machine learning allowed for sophisticated optimization based on the combination of real-time monitoring sensor data and satellite data, and the accuracies of the partial least squares (PLS) regression, support vector regression (SVR), and deep neural network (DNN) models were all higher than those of traditional models. However, some water quality variables, such as the concentration of pathogens, cannot be directly measured by remote sensing, as they are not optically active or lack high-spatial-resolution hyperspectral data, but can be estimated indirectly using other measurable data [29]. Wu et al. [30] developed an attentional neural network based on a convolutional neural network (CNN) to identify clean and polluted water on the basis of water images. They conducted several comparison experiments on a water surface image dataset and verified the effectiveness of this attentional neural network.

LINK: https://www.sciencedirect.com/science/article/pii/S2772985022000163

TITLE: Comparative Assessment of Individual and Ensemble Machine Learning Models for Efficient Analysis of River Water Quality

AUTHOR: Abdulaziz Alqahtani, Muhammad Izhar Shah, Ali Aldrees, and Muhammad Faisal Javed

The outcomes of the parametric analysis demonstrated that the modelling process had accounted for the effects of every input element. The external assessment criteria confirmed the overall result and reliability of the offered approaches. The results of this study demonstrated that creating AI-based models for river water quality assessment, management, and policy making is both feasible and advantageous.

Seven input parameters were chosen for the dataset of the projected models' training and testing on the basis of their significant association. The ensemble RF model was optimised by creating 20 sub-models and selecting the most accurate one. Known statistical measures including the coefficient of determination (R2), mean absolute error (MAE), root mean squared error (RMSE), and Nash-Sutcliffe efficiency were used to evaluate the models' goodness-of-fit (NSE). The R2 value for the GEP, RF, and ANN models, respectively, was found to be 0.96, 0.98, and 0.92, demonstrating a significant correlation between inputs and modelling outputs. The comparative effectiveness of the suggested methodologies demonstrated the RF's relative superiority over GEP and ANN. The most accurate model among the 20 RF sub-models produced R2 values of 0.941 and 0.938, with 70 and 160 numbers of corresponding estimators. On training and testing data, respectively, the ensemble RF model produced the lowest RMSE values of 1.37 and 3.1.

The generalised outcomes of all the aforementioned procedures were ensured by the models' evaluation on outside criteria. As a result of the current study, it was concluded that the RF model with a few essential parameters might be prioritised for water quality assessment and management.

LINK: https://www.mdpi.com/2071-1050/14/3/1183