M.KUMARASAMY
COLLEGE OF ENGINEERING
NAAC Accredited Autonomous Institution
Approved by AICTE & Affiliated to Anna University
ISO 9001:2015 & ISO 14001:2015 Certified Institution
Thalavapalayam, Karur – 639 113.

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

LITERATURE REVIEW OF NALAYATHIRAN PROJECT

TITLE : EFFICIENT WATER QUALITY ANALYSIS AND PREDICTION USING MACHINE LEARNING

DOMAIN : APPLIED DATA SCIENCE

FACULTY MEMBER : Dr. S. SUJANTHI M.E., PH.D.,

INDUSTRY MENTOR : SOWJANYA, SANDEEP DOODIGANI

PROJECT MEMBERS :

REGISTER NO	NAME	ROLE
19BCS4034	GOWRI J	LEADER
19BCS4023	DHANISH S	TEAM MEMBER
19BCS4024	DHARANEESH S K	TEAM MEMBER
19BCS4056	KAVIYA K K	TEAM MEMBER

ABSTRACT

Water is considered a vital resource that affects various aspects of human health and life. The quality of water is a major concern for people living in urban areas. The quality of water serves as a powerful environmental determinant and a foundation for the prevention and control of waterborne diseases. However, predicting the urban water quality is a challenging task since the water quality varies in urban spaces non-linearly and depends on multiple factors, such as meteorology, water usage patterns, and land uses, so this project aims at building a Machine Learning (ML) model to Predict Water Quality by considering all water quality standard indicators

LITERATURE SURVEY

1. Improving the robustness of beach water quality modeling using an ensemble machine learning approach

The best machine learning method varies between beaches and years, making method selection difficult. This study proposes an ensemble machine learning approach referred to as model stacking that has a two-layered learning structure, where the outputs of five widely-used individual machine learning models (multiple linear regression, partial least square, sparse partial least square, random forest, and Bayesian network) are taken as input features for another model that produces the final prediction. Applying this approach to three beaches along eastern Lake Erie, New York, USA, we show that generally, the model stacking approach was able to generate reliably good predictions compared to all of the five base models. The accuracy rankings of the stacking model consistently stayed 1st or 2nd every year, with yearly-average accuracy of 78%, 81%, and 82.3% at the three studied beaches, respectively. This study highlights the value of the model stacking approach in predicting beach water quality and solving other pressing environmental problems.

2. Application of artificial neural network in water quality index prediction: a case study in Little Akaki River, Addis Ababa, Ethiopia

The Little Akaki River is one of the most polluted Rivers in Ethiopia as reported in many studies. These studies, however, mainly used concentration measurements of certain constituents and compared them against local and international standards. The multitude of objectives and differences in selected constituents in the various reports require scientific knowledge to understand the River water quality status. This may hinder policymakers and the public from knowing the extent of the River's pollution. The water quality index is a useful method to get a summary of water bodies' pollution extent. To this end, the Canadian Council of Ministers of the Environment-water quality index approach was used. Furthermore, modeling of the index was performed using a trained and validated artificial neural network. Twelve water quality parameters from 27 sampling sites in the Dry season (January/February 2017) and Wet season (October/November 2015) were used for index determination. Results show that all sampling sites except one site upstream were under the poor water quality category. Afterward, the neural network model was trained and validated, for 12 inputs and one output, using several combinations of hidden layers (2–20), several neurons in the hidden layers (5, 10, 15, 20, 25), transfer, training, and learning functions. The most optimal model architecture was obtained with eight hidden layers and 15 hidden neurons which resulted in an R2 value of 0.93. This shows a good agreement between calculated and predicted index values. Hence, an artificial neural network can be successfully applied for modeling Little Akaki River's water quality index.

3. Improving prediction of water quality indices using novel hybrid machine-learning algorithms

River water quality assessment is one of the most important tasks to enhance water resources management plans. A water quality index (WQI) considers several water quality variables simultaneously. Traditionally WQI calculations consume time and are often fraught with errors during derivations of sub-indices. In this study, 4 standalone (random forest (RF), M5P, random tree (RT), and reduced error pruning tree (REPT)) and 12 hybrid data-mining algorithms (combinations of standalone with bagging (BA), CV parameter selection (CVPS) and randomizable filtered classification (RFC)) were used to create Iran WQI (IRWQIsc) predictions. Six years (2012 to 2018) of monthly data from two water quality monitoring stations within the Talar catchment were compiled. Using Pearson correlation coefficients, 10 different input combinations were constructed. The data were divided into two groups (ratio 70:30) for model building (training dataset) and model validation (testing dataset) using a 10-fold cross-validation technique. The models were evaluated using several statistical and visual evaluation metrics. The result shows that fecal coliform (FC) and total solids (TS) had the greatest and least effect on the prediction of IRWQIsc. The best input combinations varied among the algorithms; generally, variables with very low correlations displayed weaker performance. Hybrid algorithms improved the prediction power of several of the standalone models, but not all. Hybrid BA-RT outperformed the other models (R^2 = 0.941, RMSE = 2.71, MAE = 1.87, NSE = 0.941, PBIAS = 0.500). PBIAS indicated that all algorithms, with the exceptions of RT, BA-RT and CVPS-REPT, overestimated WQI values.

4. Prediction of Water Quality Parameters Using ANFIS Optimized by Intelligence Algorithms (Case Study: Gorganrood River)

Water quality management and control have high importance in the planning and development of water resources. This study investigated the application of Genetic Algorithm (GA), Ant Colony Optimization for Continuous Domains (ACOR), and Differential Evolution (DE) in improving the performance of adaptive neuro-fuzzy inference system (ANFIS), for evaluating the quality parameters of Gorganroud River water, such as Electrical Conductivity (EC), Sodium Absorption Ratio (SAR) and Total Hardness (TH). Accordingly, initially most suitable inputs were estimated for every model using sensitivity analysis and then all of the quality parameters were predicted using mentioned models. Investigations showed that for predicting EC and TH in the test stage, ANFIS-DE with R^2 values of 0.98 and 0.97, respectively and RMSE values of 73.03 and 49.55 and also MAPE values of 5.16 and 9.55, respectively were the most appropriate models. Also, ANFIS-DE and ANFIS-GA models had the best performance in prediction of SAR (R^2 = 0.95, 0.91; RMSE = 0.43, 0.37 and MAPE = 13.43, 13.72) in test stage. It is noteworthy that ANFIS showed the best performance in the prediction of all mentioned water quality parameters in training stage. The results indicated the ability of mentioned algorithms in improving the accuracy of ANFIS for predicting the quality parameters of river water.

5. Integrating water quality and operation into a prediction of water production in drinking water treatment plants by genetic algorithm enhanced artificial neural network

Stringent regulations and deteriorating source water quality could greatly influence the water production capacity of drinking water treatment plants (DWTPs). Using models to predict the performance of DWTPs under stress provides valuable information for decision making and future planning. A hybrid statistic model named HANN was established by combining an artificial neural network (ANN) with a genetic algorithm (GA) aiming at forecasting the overall performance of DWTPs nationwide in China. Monthly data from 45 DWTPs across China was employed. Water quality parameters like temperature and chemical oxygen demand (COD) and operational parameters like electricity consumption and chemical consumption were selected as input variables while drinking water production was employed as the output. Both preliminary data analysis and principal component analysis (PCA) suggested a clear non-linear relationship between the input and output variables. The structure of the HANN model was optimized by employing the lowest mean squared error (MSE) as the indicator. The resultant HANN model performed well when simulating the training datasets. Its predictive accuracy for the independent test datasets was enhanced when feeding more training datasets and the performance was constantly higher than the independent multi-layered ANN models using the coefficient of determination (R2) as the indicator, indicating the HANN model was capable of capturing complex non-linear relationship and extrapolation. Results from the Accuracy test, Garson sensitivity analysis, and Analysis of Variance (ANOVA) suggested the quantity of water produced by DWTPs was closely linked to water quality and operational parameters. The scenario analysis showed that the HANN model was capable of predicting water production variation based on the parameter variations, indicating that the HANN model could be a general management tool for decision-makers and DWTP managers to make plans in advance of regulatory changes, source water quality variations and market demand.

TABLE OF ARTICLES

S.NO	ARTICLE NAME	AUTHOR NAME	PUBLISHED YEAR	DRAWBACKS
1	Improving the robustness of beach water quality modeling using an ensemble machine learning approach	L Wang, Z Zhu, L Sassoubre, G Yu, C Liao, Q Hu, et al.	2021	the model is built only to predict beach water quality
2	Application of artificial neural network in water quality index prediction: a case study in Little Akaki River, Addis Ababa, Ethiopia	M Yilma, Z Kiflie, A Windsperger, and N Gessese	2018	Usage of ANN model which can be replaced by simple ML model with the same level of accuracy

3	Improving prediction of water quality indices using novel hybrid machine-learning algorithms	DT Bui, K Khosravi, J Tiefenbacher, H Nguyen, and N Kazakis	2020	The model is built with high complexity but gives the same accuracy level as simple models can.
4	Prediction of Water Quality Parameters Using ANFIS Optimized by Intelligence Algorithms (Case Study: Gorganrood River)	An Azad, H Karami, S Farzin, A Saeedian, H Kashi, and F Sayyahi	2018	It is specific to Gorganroud River water
5	Integrating water quality and operation into a prediction of water production in drinking water treatment plants by genetic algorithm enhanced artificial neural network	Y Zhang, X Gao, K Smith, G Inial, S Liu, LB Conil, et al.	2019	Usage of ANN model which can be replaced by simple ML model with the same level of accuracy

CONCLUSION

From this survey, we find that the specific web application for efficient water quality analysis & prediction using simple machine learning algorithm is not present as it deals with accuracy issues, also there are some limitations happened in the articles published related to water quality prediction as shown in the above table.

REFERENCE

1. L Wang, Z Zhu, L Sassoubre, G Yu, C Liao, Q Hu, et al., Improving the robustness of beach water quality modeling using an ensemble machine learning approach, Science of The Total Environment, Vol. 765, 2021, pp. 142760.

2. M Yilma, Z Kiflie, A Windsperger, and N Gessese, Application of artificial neural network in water quality index prediction: a case study in little Akaki River, Addis Ababa, Ethiopia, Modeling Earth Systems and Environment, Vol. 4, 2018, pp. 175-187

3. DT Bui, K Khosravi, J Tiefenbacher, H Nguyen, and N Kazakis, Improving prediction of water quality indices using novel hybrid machine-learning algorithms, Science ofThe Total Environment, Vol. 721, 2020, pp. 137612.

4. An Azad, H Karami, S Farzin, A Saeedian, H Kashi, and F Sayyahi, Prediction of water quality parameters using ANFIS optimized by intelligence algorithms (case study: Gorganrood river), KSCE Journal of Civil Engineering, Vol. 22, 2018, pp. 2206-2213.

5. Y Zhang, X Gao, K Smith, G Inial, S Liu, LB Conil, et al., Integrating water quality and operation into a prediction of water production in drinking water treatment plants by genetic algorithm enhanced artificial neural network, Water Research, Vol. 164, 2019, pp. 114888.