Efficient Water Quality Analysis and Prediction using Machine Learning

(Efficient Water Quality Analysis and Prediction using Machine Learning is used to find the purity of the water)

MOHAMED ILIYAZ S , ASWIN KUMAR I , DHAYANIDHI V , SAMRAJ S .

LITERATURE REVIEW

The basic idea of this research is to devise a comprehensive methodology that analyzes and predicts the water quality of particular regions with the help of certain water quality parameters. These parameters include physical, biological, or chemical factors which influence water quality. There are certain quality standards set up by international organizations like the World Health Organization (WHO) and the Environmental Protection Agency (EPA), which serve as a benchmark for determining the quality of water. In its document "Efficient Water Quality Analysis and Prediction using Machine Learning", EPA mentions a total of 101 parameters thataffect water quality in one way or another. However, some parameters have a greater and more visible effect on water quality than others.

TITLE: IMPROVING THE ROBUSTNESS OF BEACH WATER QUALITY MODELING USING AN ENSEMBLE MACHINE LEARNING

AUTHOR: Wang et al (2021)

This study demonstrates the utility of using a model stacking approach for predictive modeling of beach water quality. Since model stacking averages out noise from its base models, it is theoretically more promising than individual models in generating predictions with greater accuracy and robustness. The results from this study suggest that the model stacking algorithm has promise for improving the reliability of predictive modeling for beach microbial water quality of other sites with similar hydrogeological and environmental conditions such as other beaches along the Great Lakes.A comprehensive test needs to be done to understand the strength and weaknesses of individual base models and the stacking approach. This study indicated that the model stacking approach may improve the robustness of beach water quality modeling.

TITLE:ACCURATE PREDICTION SCHEME OF WATER QUALITY IN SMARTMARICULTURE WITH A DEEP BI-S-SRU LEARNING NETWORK

AUTHOR: J. Liu, C. Yu, Z. Hu et al (2020)

This paper proposed the process and model for the accurate prediction of key water quality parameters (pH, water temperature, and dissolved oxygen). Firstly, the collected water quality data is repaired and corrected by the improved preprocessing method, and then the data is filtered and denoised by the wavelet transform method. After preprocessing, the data received by remote transmission can be recovered well. Next, we construct the Bi-S-SRU (Bi-directional Stacked SRU) deep learning prediction model by importing a pretreated dataset weighted with the discovered correlation coefficients. The experimental results demonstrate that our proposed prediction model can achieve higher prediction accuracy and stability compared with RNN-based and SRU-based prediction models. The experimental results also show that the Bi-S-SRU-based prediction method is only slightly higher in time complexity than the traditional RNN-based or LSTM-based prediction method.

TITLE: ASSESSMENT OF SURFACE WATER QUALITY BY USING SATELLITE IMAGES FUSION BASED ON PCA METHOD IN THE LAKE GALA, TURKEY

AUTHOR: E. Batur and D. Makita (2019)

In this paper, the PCA model is presented to integrate surface water reflectance values from satellite images to monitor Gala Lake's surface water quality. The values of Chl-a, DO, TSS, SDD, TDS, and pH values calculated by the PCA method were found to be highly correlated with the measured water quality parameters. The results obtained were found to be directly proportional to the number of sensors. L8 OLI and S2A have higher spectral resolution than GK2 images. However, the high temporal resolution of GK2 allows the desired region to be displayed at more frequent intervals, allowing for better monitoring of the instantaneous changes in surface water quality. Therefore, longer measurements should be made and analyzed for a model covering all periods.

TITLE:SURFACE WATER POLLUTION DETECTION USING THE INTERNET OF THINGS

AUTHOR: Shafi et al (2018)

In this paper, theproposed an IoT-based solution to monitor water quality in real-time. The proposed system provides remote monitoring of water quality assessment along with water flow control via a mobile app. Four machine learning algorithms including Support Vector Machine (SVM), k Nearest Neighbor (kNN), single layer neural network, and deep neural network have been applied for the classification of water quality and experimental results revealed that deep neural network outperforms all other algorithms with an accuracy of 93. This system has the potential to effectively utilize to overcome the challenges of water quality in the agriculture sector and various industries.

TITLE:IMPROVING WATER QUALITY INDEX PREDICTION IN PERAK RIVER BASIN MALAYSIA THROUGH A COMBINATION OF MULTIPLE NEURAL NETWORKS

AUTHOR: Ahmad et al (2017)

In this paper, they proposed a reliable real-time prediction model for WQI developed through a

selective combination of multiple neural networks by excluding COD and BOD from model inputs as they cannot be measured in real-time. Single and multiple FANN are used in this paper to model the WQI in the Perak River basin. The selective combination schemes provide models with better generalization capability compared to combining all neural networks. The bootstrap aggregated models with selective combination provide a real-time WQI prediction tool without delay as only real-time measurements are used as model inputs.

TITLE:ARTIFICIAL INTELLIGENCE FOR THE PREDICTION OF WATER QUALITY INDEX IN GROUNDWATER SYSTEMS AUTHOR: Mohamad Sakizadeh (2016)

One of the problems of ANN's modeling in environmental studies which suffers from the problem of the small data records is the danger of over-fitting the model to the training data resulting in poor generalization of the model for the data out-of-the training data range.This study's results proved that this problem can be obviated by using some algorithms like Bayesian regularization and Ensemble methods. The prediction of water quality index (WQI) was successfully implemented by Bayesian regularization and Ensemble averaging methods, though the performance of Bayesian regularization was roughly better, with minimum test error indicating the good generalization ability of these methods in this field. The poor generalization ability is a problem that has been overlooked by most of the research all around the world although it is an important issue that should be taken into account.

TITLE:THE USE OF COMBINED NEURAL NETWORKS AND GENETIC ALGORITHMS FOR THE PREDICTION OF RIVER WATER QUALITY

AUTHOR: Ding et al (2014)

In this paper, they proposea water quality prediction model that combines PCA, BPNN, and GA. Using the BPNN model to study water classification and prediction can overcome disadvantages including the large workload of traditional evaluation methods and strong subjectivity. This model possesses objectivity, universality, and practicality. PCA converts the multi-indices into a few aggregative indices with little original data information loss and reduces the input data to speed the training process. Using GA to optimize network parameters can effectively prevent the search process from converging to local optimum solutions, optimize global optimal network parameters, and significantly improve the accuracy of water quality prediction. This model can obtain high training speed and good prediction rate and can be extended to other classification problems.

Our Ideology

The estimated water quality in our work is based on nine parameters: ph, Hardness, Solids, Chloramines, Sulfate, Conductivity, Organic carbon, Trihalomethanes, Turbidity, and pH, which are tested according to World Health Organization (WHO) standards.

The proposed methodology improves on these notions and the methodology being followed is depicted in Figure 1.

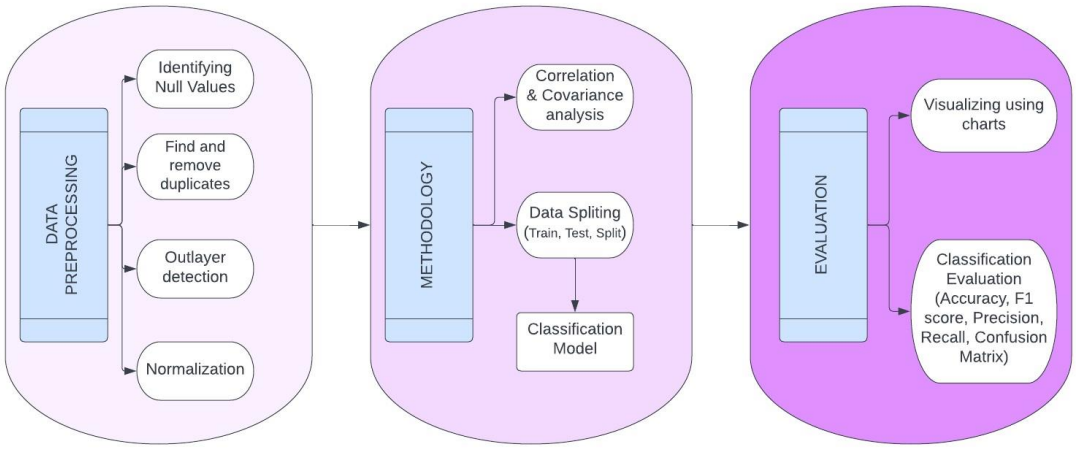

Figure 1

COMPARATIVE ANALYSIS OF LITERATURE SURVEY:

S.No	Year	Researcher	Title	Parameters	Algorithm	Remarks
01	2021	Wang et al	Improving the robustness of beach water quality modeling using an ensemble machine learning	turbidity, temperature, Culturable fecal indicator bacteriasuch as Escherichia coli (E. coli)	Partial least square, sparse partial least square, random forest, Bayesian network, Akhand linear regression	Highest accuracy of 82.3% with ensemble machine learning algorithm
02	2020	J. Liu, C. Yu, Z. Hu et al	Accurate prediction scheme of water quality in smart mariculture with a deep Bi-S-SRU learning network	Salinity, chlorophyll,turbidity,Water Temperature,PH,Dissolved Oxygen(DO)	LSTM, SRU, RNN, LSTM, SRU and Bi-SSRU	Highest accuracy of 94.42% using a Bi-S-SRU
03	2019	E. Batur and D. Makita	Assessment of surface water qualityby using satellite images fusion based on PCA method in theLake Gala, Turkey	DO, SDD, TDS, and pH Chl-a and TSS	MLR, SVM, ANN, AND PCA	Highest accuracy of 92% using a PCA-based RSR model
04	2018	Shafi et al	Surface Water Pollution Detection usingthe Internet of Things	turbidity, temperature and Ph	Support Vector Machines (SVM), Neural Networks (NN), Deep NeuralNetworks (Deep NN), and k Nearest Neighbors (kNN)	Highest accuracy of 93% with Deep NN
05	2017	Ahmad et al	Improving water quality index prediction in Perak River basin Malaysia through a combination of multiple neural networks	Nitrate, PH, Electrical conductivity, Dissolved oxygen, total coliform, Biochemical Oxygen Demand	feedforward artificial neural network; forward selection; backward elimination; artificial neural network; multiple neural networks	Highest accuracy of 92.7% using a selective combination methods
06	2016	Sakizadeh	Artificial intelligence for the prediction of water quality index in groundwater systems	EC, TDS, Mn, Cu, Cr(VI),Turbidity, pH, Ca, Mg, Totalhardness, Sulfate, Fe, Fluoride Phosphate, Nitrate, Nitrite	ANN with Bayesian regularization	Highest accuracy of 80% using an Artificial Neural Network

07	2014	Ding et al	The Use of Combined Neural Networks and Genetic Algorithms for Prediction of River Water Quality	pH, NH3-N, TN, Cr6+, TP, CODMn, BOD5, TCN, COD, Cd, Cu, Zn, Pb, Hg, As, Se, F-, sulfide, dissolved oxygen,and LAS, etc.	Genetic Algorithm (GA), and Back Propagation Neural Network (BPNN)	The highest accuracy of Non polluted and polluted of 88.9% and 93.1% with PCA technique