Literature Survey

[1] Ahmed, R Mumtaz, H Anwar, AA Shah, R Irfan, and J García-Nieto,Efficient water quality prediction using supervised machine learning, Water, Vol. 11, 2019, pp. 2210.

Ahmed et al. used supervised machine learning methods to evaluate the water quality index (WQI), where a single index was used to sum up the general water quality and water quality class (WQC). The WQI was then estimated with a mean absolute error (MAE) of 1.9642 and 2.7273, and it was discovered that the gradient boosting with a learning rate of 0.1 and polynomial regression with a degree of 2 had predicted the WQI most accurately. The MLP in this situation has the maximum classification accuracy of 85.07% and has the configuration of (3, 7).

[2] 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.

Wang et al. have suggested a two-layered model stacking method for forecasting the water quality of beaches. The final forecast is created by integrating the five most popular techniques (partial least square, sparse partial least square, random forest, bayesian network, and akhand linear regression) into a machine learning model. In this instance, three separate beaches along eastern Lake Erie in New York, USA, were subjected to the model stacking technique, and all five foundation models were contrasted. Following examination, the model stacking approach outperformed each of the basis models. Stacking model accuracy results were consistently among the highest, with an average accuracy of 78%, 81%, and 82.3% at the three examined beaches year over year.

[3] CV Sillberg, P Kullavanijaya, and O Chavalparit, Water quality classification by integration of attribute-realization and support vector machine for the chao phraya river, Journal of Ecological Engineering, Vol. 22, 2021, pp. 70-86.

To categorise the water quality of the Chao Phraya River, Sillberg et al have developed a machine learning-based method that combines the attribute-realization (AR) and support vector machine (SVM) algorithms. The linear function used by the AR to determine the most important parameters to improve the condition of the river. The parameters that contributed the most to the categorization were NH3-N, TCB, FCB, BOD, DO, and Sal, increasing the contribution values from 0.80 to 0.98, compared to 0.25 to 0.64 for TDS, Turb, TN, SS, NO3-N, and Cond. The accuracy of 0.94, the precision average of 0.84, the recall average of 0.84, and the F1-score average of 0.84 are the best classification results made possible using the SVM linear approach. The validation revealed that AR-SVM was an effective technique to determine river water quality with an accuracy of 0.86 to 0.95.

[4] 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.

Yilma et al. simulated the WQI of the Akaki River using an artificial neural network. The index was calculated using the twelve water quality indicators from 27 sample sites collected during the dry and wet seasons. All forecasted findings have shown poor water quality, with the exception of one area upstream. Here, the neural network model was trained and verified using 12 inputs and 1 output while using hidden layer neurons (5, 10, 15, 20, 25) and the number of hidden layers (2–20). According to their research, an artificial neural network with 15 hidden neurons and eight hidden layers was able to predict the WQI with a 0.93 accuracy rate.

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

A hybrid random tree and bagging (BA-RT) machine learning technique has been developed by Bui et al. In their study, 12 hybrid data-mining algorithms (hybrids of the standalone with bagging, CVPS, and RFC) and four standalone data-mining algorithms (RF, M5P, RT, and REPT) for monthly WQI forecasting in a humid climate in northern Iran were examined. They discovered that total solids and faecal coliform had the greatest and least impact on forecasting IRAQIs, respectively. Although several algorithms have produced different ideal input combinations in this case, the variables with low correlations have fared worse. Although not all standalone models have shown an increase in prediction power from hybrid algorithms, the hybrid BA-RT has beaten the others by obtaining R2 0.941 using a 10-fold cross-validation technique, beating out 15 standalone and hybrid algorithm.

[6] YR Ding, YJ Cai, PD Sun, and B Chen, The use of combined neural networks and genetic algorithms for prediction of river water quality, Journal of Applied Research and Technology, Vol. 12, 2014, pp. 493-499.

For the purpose of forecasting river water quality, Ding et al. have developed a hybrid intelligent system that incorporates Principal Component Analysis (PCA), Genetic Algorithm (GA), and Back Propagation Neural Network (BPNN) methodologies. Each of the 23 different water quality indicator variables used in this study has a complex non-linear relationship with water quality. In this instance, PCA has substantially sped up follow-up algorithm training, while GA has optimised the BPNN's parameters. According to the results, the average prediction rates for non-polluted and polluted water quality were 88.9% and 93.1%, respectively, while the global prediction rate was close to 91%.

[7] A 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.

In order to enhance the performance of the adaptive neuro-fuzzy inference system (ANFIS) for the prediction of water quality measures, Azad et al. have used the three evolutionary algorithms GA, DE, and ACOR. To forecast the EC, SAR, and THE water quality measures, these algorithms have been combined with the ANFIS. According to their research, the ANFIS-DE model had the best accuracy in predicting EC and TH in the test stage, with an R2 of 0.98, an RMSE of 73.03, and a MAPE of 5.16. Additionally, in a test stage, the ANFIS-DE and ANFIS-GA models performed the best for SAR prediction (R2 = 0.95, 0.91; RMSE = 0.43, 0.37; MAPE = 13.43, 13.72). It has been demonstrated that ANFIS can create the best results.

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

A hybrid artificial neural network (HANN) model for the prediction of drinking water treatment facilities in China has been enhanced by Zhang et al. using a genetic algorithm (GA). The model has been trained, validated, and validated continuously using monthly data from 45 DWTPs located around China, which includes eleven input variables for operational performance and water quality. When combined with operational and water quality criteria, the HANN model has demonstrated superior ability and consistency in predicting the total water production of DWTPs. By increasing the training data supplied, the HANN model's performance increased from 0.71 to 0.93 (R2), as demonstrated by their prediction and the model's capacity to develop to the highest degree of performance.

[9] Ali, M.; Qamar, A.M. Data analysis, quality indexing and prediction of water quality for the management of rawal watershed in Pakistan. In Proceedings of the Eighth International Conference on Digital Information Management (ICDIM 2013), Islamabad, Pakistan, 10–12 September 2013; pp. 108–113.

To categorise samples into classes of water quality, Ali and Qamar employed the average linkage (within groups) approach of hierarchical clustering, an unsupervised methodology. They did not employ a standardised water quality indicator to assess their predictions, and they overlooked the key WQI-related characteristics during the learning process.

[10] Gazzaz, N.M.; Yusoff, M.K.; Aris, A.Z.; Juahir, H.; Ramli, M.F. Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors. Mar. Pollut. Bull. 2012, 64, 2409–2420.

With a model that accounts for over 99.5% of the volatility in the data, Gazzaz et al. employed ANN to predict the WQI. Given the cost of the sensors, they employed 23 characteristics to forecast the WQI, which turned out to be rather expensive if used for an IoT system.