



(REVIEW ARTICLE)



A novel approach to predict Water Quality Index using machine learning models: A review of the methods employed and future possibilities

Ismail I Aminu *

Department of Civil Engineering, Faculty of Engineering, Bayero University, Kano, Nigeria.

Global Journal of Engineering and Technology Advances, 2022, 13(02), 026–037

Publication history: Received on 26 September 2022; revised on 06 November 2022; accepted on 09 November 2022

Article DOI: <https://doi.org/10.30574/gjeta.2022.13.2.0184>

Abstract

The development of computer models for water quality index forecasting has been a leading research topic worldwide which has been considerably recognized over the last two decades; the balance between efficient water quality requires a good water management technique. The balance is said to be achieved through various procedures many of which require the application of computer-aided forecasting tools. In this paper, a decade research review on the water quality index in the field of artificial intelligence was carried out with the aim to present the most viable or most suitable methods and models to be adopted for future researchers in the field of water quality. The review incorporates the developed models such as ANN, ANFIS, SVM, other regression or time-series, and other soft computing models. This research shows that the study focused on a decade review of the methods and models, and also, there is room for long-term forecasts. It also shows that there is no single AI model that outperforms all the remaining AI models but It is necessary to evaluate the strength of each model combination for each region thus to know what type of method or model that works best for the country or region. The use of AI has grown significantly in recent decades however there is enough room for researchers to duel in and improve in the field of water quality index.

Keywords: Artificial intelligence; WQI; Machine learning; Mini review

1. Introduction

The concern for clean and drinkable water is quite essential for health, water resources and environmental purposes [1]–[7]. The demand by billions of people and numerous industries for clean, safe and adequate freshwater on the planet encouraged the practitioners and research communities to be much engaged in prediction and monitoring of water quality and to address this universal concern [8]. However, statistics as in 2015, that 29% of the world population does not have access to safe drinking water (WHO/IWA, 2017). As Groundwater being one of the main sources of water worldwide (65% of the groundwater is used for drinking purposes, 20% of it is used for irrigation and livestock, and 15% for industry and mining), the quality of the groundwater is dependent on the interactions within the aquifer between chemical constituents, soil, rock, and gases [9], [10].

The Water Quality Index (WQI) is a classification indicator that has been widely used for both the ground and surface water in determining the quality of the water which is essential and also plays a significant role in water resource management [11]–[13]. WQI is a certified method to determine the quality of water by a simple way, that can respond to the various changes in the basic characteristics of water, using the relative weight for each parameters and classify the waters according to the quality categories levels [14]–[17]. The Application of the WQI was first presented by (Brown et al., 1970; Horton, 1965) And adopted and modified by countless researchers (Cude, 2001; Pesce and Wunderlin, 2000; Said et al., 2004). All these applications came from the general concept of Water Quality Index

* Corresponding author: Ismail I. Aminu

Department of Civil Engineering, Faculty of Engineering, Bayero University, Kano, Nigeria.

proposed by the National Sanitation Foundation (NSF) of the United States, and the NSFQI [18] is the world's most commonly adopted method.

Artificial Neural Network (ANN) is an AI-based approach that not only proved to be effective in handling large amount of dataset, complex nonlinear input and output relationship but also flexible and powerful computational tool [19], [20]. Over the past years, AI in the form of deep learning and machine learning models, have been increasingly applied to solve various environmental engineering problems and has been proved to be very effective, including river water quality modeling [21]–[23].

The main objective of this paper is to evaluate and review the application of Artificial Intelligence in determining water quality index which will be able to provide some guidance to the upcoming researchers. A decade review will be carried out to provide the best model to be used globally or in a particular country. The performance of the models using performance criteria's will be compared to other models (single or hybrid) and outperforming model will be suggested in the particular region, state or country.

2. Machine learning advancements

2.1. Artificial Neural Networks

An artificial neural network (ANN) is a model that comprises of interconnected mathematical nodes or neurons similar to the biological neurons to form a structure and can also relate input and output parameters [24]. Usually, signals are fed through the neuron to form a net input into another neuron, and the output layer which equals the number of desired output is determined by weight and bias associated with links among the neurons [25] (see, Fig. 1). Because of a range of variables, ANN is an exceptional predicting tool. The ANN, according to Adamowski et al. (2002), allows the application of information to predict future variables of possibly noisy multivariate time series [26]–[31].

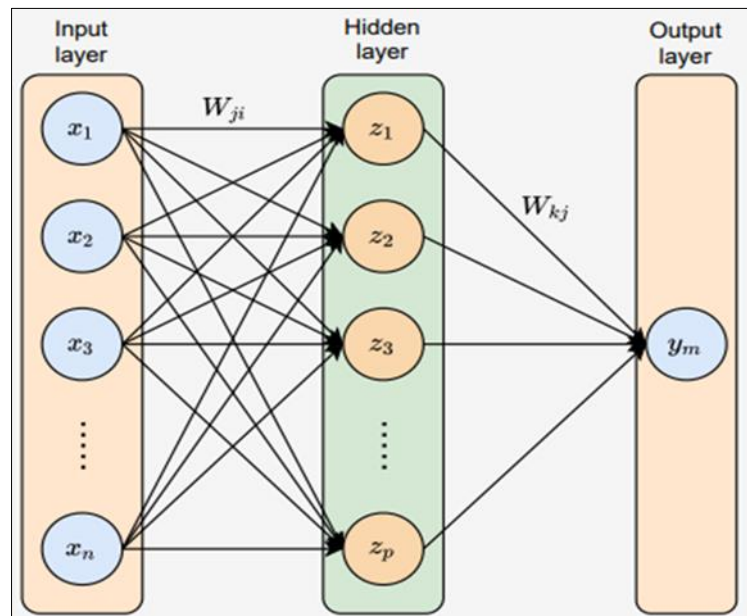


Figure 1 A three-layer ANN structure

The ANNs have been proved to be effective model in predicting highly complicated relationships, processes, and phenomena. They have been successfully employed in a multitude of applications including river flow forecasting and also water quality index (WQI) [32]. However, the use of ANNs to develop accurate models to predict WQI for rivers is an application which has yet to be investigated as evidenced by search in literature which lead to only two such publications: Juahir et al., 2004 and Khuan et al., 2002. [33][34][35] developed a receptor model and then compared it to ANN and MLR in order to predict the WQI of a river in Malaysia due to agricultural practices along the sea which suggests that APCS-ANN can be used for environmental monitoring agencies in Malaysia. [36] Developed an ANN model in order to determine the WQI in Euphrates River which has six sampling stations and then compared it to MLR model, comparison of the results showed that the ANN has a higher degree of accuracy than the MLR.

2.2. Adaptive Neuro-Fuzzy Interference System

The adaptive neuro-fuzzy interference system (ANFIS), was formally referred to as the Takagi–Sugeno–Kang system, was created by Jang in 1993 and is the widely used fuzzy inference system [37]. It comprises five components; input(s), a fuzzy system generator, a fuzzy inference system (FIS), an adaptive neural network, and an output [38]. This technique uses the hybrid learning or back-propagation algorithm for training and combines the features of neural networks and the capabilities of fuzzy logic [39].

ANFIS hybrid model was employed to predict the performance of WQI namely ANFIS-FCM (Fuzzy C-Means data clustering), ANFIS-GP (Grid Partition), and ANFIS-SC (Subtractive Clustering); they observed ANFIS-SC to be the best performing model. [40]. [41] developed a data driven ANFIS models for WQI in India, two different techniques were employed which are fuzzy C-means and subtractive clustering based ANFIS. Based on the performance of the models it was found that SC-ANFIS gave more accurate result than FCM-ANFIS. [42] Developed MLR, ANN, and ANFIS models were developed to predict the WQI for river Yamuna in India. The obtained results showed that the ANN and ANFIS outperformed the MLR model by 10% in the verification phase. Although the performance of ANN is slightly better than the ANFIS, but the ANN and ANFIS models outperformed MLR model in estimating the WQI.

2.3. Support Vector Machines/Regression

Support vector machines (SVM) are supervised machine learning models created which possess as powerful regression tool that has found applications in numerous prediction problems in various fields. In SVM, the statistical learning theory and structural error minimization principle are employed to map the initial training samples into a higher dimensional feature space through nonlinear kernel functions and the optimal solution is obtained by converting the problem to linear from nonlinear [43][44].

The Support Vector Machine (SVM) model is also believed to be a very powerful machine learning technique for both linear and nonlinear regression problems and has been used in various scientific issues with high prediction accuracy [40]. [45] Studied the optimization of the SVM model to identify the major parameters that significantly affect the WQI and found that Nitrate is the major parameter for WQI prediction. In a study of prediction WQI in constructed wetlands, SVM and two other AI method were used and the results had shown that the SVM result predicted WQI with high accuracy than the two other models [26]–[31], [46]–[54].

2.4. Linear and Statistical Methods

Multilinear regression (MLR) is a statistical model that determines the relationship between a dependent variable and at least one independent variable [55]–[57]. The main idea of the research by Fullerton Jr et al. (2016) was to determine the dynamics of water necessity for the city of El Paso (Texas, USA) using several predicting techniques, comprising the Linear Transfer Function (LTF) by Box and Jenkins (1976) [14], [57]–[60]. [61] Adopted ANN and MLR techniques to predict WQI in Shivganga River basin India. The parameters such as pH, EC, TDS, TH, Ca, Mg, Na, K, Cl, HCO₃, SO₄, NO₃ and PO₄ Were considered for computing WQI. Based on the results obtained from the models, ANN model would become more beneficial in the prediction of water quality index in future. [62] Developed an MLR model to compare it to WQI of Tigris river, therefore the developed model of MLR can be used to predict and monitor the water quality of Tigris with reasonable precision. (A. P. Kogekar et al. 2021) Developed time series models Auto-Regressive Integrated Moving Average (ARIMA), Seasonal ARIMA (SARIMA), to predict the water quality index of the river Ganga, Further, only two important water parameters such as dissolved oxygen and biochemical oxygen demand, are considered for prediction and subsequently for the forecasting of the WQI. The result concludes that SARIMA predicts the water quality parameters as well as Water Quality Index (WQI) more accurately than ARIMA.

3. Evaluation of performance

For the development of any model, goodness-of-fit, error-of-fit, and biases are very crucial for evaluating the accuracy and precision of the computing approach. The metrics for appraising the model accuracy were the mean absolute deviation (MAD), mean square error (MSE), root mean square error (RMSE), mean absolute error, coefficient of determination (R^2), and correlation coefficient (R). The choice of these parameters stemmed from their application in numerous related studies as effective means of establishing the accuracy of a prediction model [26]–[31], [46]–[50], [52]–[54], [60], [63]–[67].

Some of the most important evaluation criteria metrics that were used for evaluating the performance in this review are explained below;

RMSE (RMSE (Root Mean Squared Error) is the error rate by the square root of MSE.

RMSE has been used as a standard statistical metric to measure model performance in various fields the formula for determining the RMSE is given below

$$\text{RMSE} = [\sum_{i=1}^N (z_{fi} - z_{oi})^2 / N]^{1/2}$$

3.1. R-SQUARED (Coefficient of determination)

R-Squared represents the coefficient of how well the values fit compared to the original values. The value from 0 to 1 interpreted as percentages. The higher the value is, the better the model is

$$R^2 = \frac{1 - \sum(O - P)^2}{\sum(O - \theta)^2}$$

3.2. MSE (Mean Squared Error)

MSE represents the difference between the original and predicted values extracted by squared the average difference over the data set it is also a measure of closeness of the estimator to the true value.

$$\text{MSE} = N^{-1} \sum_{i=1}^N (O_t - P_t)^2$$

4. Discussion

It is important to mention that the reviewed study indicates that artificial intelligence has gotten a lot of attention in the field of water quality index recently. However, several AI models and optimization techniques has not been explored yet which leave a gap for researchers to explore in order to contribute to the development of water quality index. The study shows that majority of the researchers employed ANN model in their study and it also showed that various types of ANN models were used such as BPNN, FFNN etc. other types of ANN classification such as ENN, RBFEL received little or no attention in this field. There are also some advancements which showed that researchers began developing hybrid models to increase the prediction accuracy. In addition, ANFIS and SVM also showed a considerable attention in water quality index simulation. But in most cases the studies using ANFIS did not consider the various membership functions (MFs) which justifies the precision of the model. It is also important to mention that the classical approach no longer produce the best results hence researchers tend to start producing integrated methods or hybrid models which tends to give some higher degree of accuracy than the classical models. Table 1 shows the decade review for water quality index modelling and forecasting.

Table 1 A decade water quality index forecasting methods according to the referenced literature

Author	Country	Performance Metrix	Remarks
[68]	India	R^2 ,	Adaptive network fuzzy inference system (ANFIS) was used for the prediction of water quality of groundwater near mines due to the contamination by heavily acids, alkaline, toxicity, heavy mineral, and microbes. The model performed quite satisfactorily with the actual and predicted water quality.
[32]	Malaysia	R^2 ,AAE,R,P,N	ANN was applied to predict problem entailing use of archival measurements on WQVs of a surface water body for construction of a model capable of forecasting and calculating WQI. This approach can be used to any aquatic system worldwide. Therefore, empirical data analysis techniques such as the ANNs are recommended for analysis of long-term environmental monitoring records.
[33]	Malaysia	R^2 ,RMSE	A receptor model was developed then compared to ANN and MLR to a dataset from Kuantan river Malaysia in order to predict the water quality due to agricultural practices along the area. From the results it showed that the prediction of WQI values using APCS-ANN model can be used for environmental monitoring agencies in Malaysia.

[69]	Pakistan	R^2 ,	ANN techniques were used to predict the water quality index in Pakistan, several models were developed and compared and the results showed that MLP had a high degree of accuracy compared to its counterparts
[70]	Mexico	NSE	Water Quality Index for monitoring and controlling shrimp culture systems using an analytical hierarchical process was developed which detects poor water. Based on the results it conclude that priority parameter assignment provides a more +effective water quality assessment than traditional approaches
[71]	Malaysia	R^2 ,RMSE,MAE	Support vector machine (SVM) and two methods of artificial neural networks (ANNs), namely feed forward back propagation (FFBP) and radial basis function (RBF), were used to predict the water quality index (WQI) in a free constructed wetland. Based on the results obtained, the SVM and FFBP can be successfully employed for the prediction of water quality in a free surface constructed wetland environment
[36]	Iraq	R, RMSE, MAPE	ANN was employed to predict water quality in Euphrates River, MLR was used to obtain a set of coefficients for a linear model, Six sampling stations along the river were chosen. Comparison of the results showed that the ANN has a higher degree of prediction than the MLR.
[72]	Malaysia	R^2 , RMSE	ANN technique was adopted to predict water quality index problem. Two methods were examined and compared, a WAMF and USHM. Models were trained and tested on land use and water quality. USHM model performs somewhat better and is more generalizable than the WAMF WQI model.
[73]	Iran	MSE	Bayesian regularization and Ensemble averaging models were developed to predict WQI with respect to the concentrations of 16 groundwater quality variables collected from 47 wells and springs in a city in Iran. Comparison among the performance of the models shows that the Bayesian regularization performed better.
[74]	Iran	R^2 , RMSE	DELPHI and CCME methods were applied to calculate the average WQI data of every month of a year. The WQI was formulated by both DELPHI and CCME techniques and the RMSE, R2 and ADD were used together to compare the water quality performance of the CCME and DELPHI models. The DELPHI method was found to have higher predictive capability than the CCME method.
[75]	Malaysia	R^2 , MAE	Gene expression programming and (ANNs) were employed to predict WQI in free surface constructed wetlands.17 points were selected and were monitored for over 14 months and an extensive data set was collected for 11 water quality variables. The GEP was able to successfully predict the WQI with high accuracy than the ANN.
[76]	China	RMSE,	A novel data-driven approach to forecast the water quality index of a station by fusing multiple sources of urban data. stMTMV model was compared to ANN, RC decay model, ARMA, Kalman filter, LR, LASSO. The experiments have shown that stMTMV has the best predictive accuracy than the other models.
[77]	Malaysia	R^2 ,MSE	FFANN was developed to predict WQI by excluding the biological oxygen demand and chemical oxygen demand as they cannot be measured in real time. From the results it showed that FFANN will be able to predict WQI with the exclusion of BOD and COD from the model input variables.
[78]	Malaysia	R^2 ,RMSE,NE	Radial basis function neural network (RBFNN) and back propagation neural networks models, have been applied to examine and mimic the relationship of WQI with the other water quality variables in a tropical environment in Malaysia. The results achieved are positively promising with RBFNN showing high degree accuracy.

[41]	India	$R^2, RMSE, MSE$	A data-driven adaptive neurofuzzy system for the water quality index was developed from eight different monitoring stations in India. Two different techniques, fuzzy C-means and subtractive clustering-based ANFIS were adopted and their performance were compared. Based on the evaluations, it was found that the SC-ANFIS method gave more accurate result as compared to the FCM-ANFIS.
[79]	Ethiopia	$R^2, RMSE, NSE$	ANN was used to predict the WQI in the most polluted Rivers in Ethiopia from 27 sampling sites. In addition, to minimize time and effort burden of repeated WQI determination, a modelling approach based on ANN can be employed successfully for the determination of the water quality index WQI.
[23]	India	RMSE	The Cascade and the Feed Neural network were adopted to predict WQI in India. The Cascade model have shown its ability to predict the WQI when the five parameters defined by WHO are employed which gives it more degree of accuracy than the Feed neural network.
[61]	India	R^2	ANN and MLR techniques were adopted to predict WQI in Shivganga River basin India. The parameters such as pH, EC, TDS, TH, Ca, Mg, Na, K, Cl, HCO ₃ , SO ₄ , NO ₃ and PO ₄ Were considered for computing WQI. Based on the results obtained from the models, ANN model would become more beneficial in the prediction of water quality index in future
[42]	India	$R^2, RMSE, NSE$	MLR, ANN, and ANFIS models were developed to predict the WQI for river Yamuna in India. The obtained results showed that the ANN and ANFIS outperformed the MLR model by 10% in the verification phase. Although the performance of ANN is slightly better than the ANFIS, but the ANN and ANFIS models outperformed MLR model in estimating the WQI.
[29]	Malaysia	MAPE, RMSE, R^2	Two evolutionary nonlinear input variable selection techniques (GP and XGB) coupled with (ELM) were developed for the determination of WQI in Kinta River Malaysia. Different trials were conducted and the best combination was chosen. The proposed hybrid GPELM and XGBELM models have improved the prediction accuracy with minimum number of input variables and can therefore serve as reliable predictive tools for WQI.
[80]	Sudan	RMSE	ANN was applied to predict the WQI by using the algorithm of feed forward back propagation artificial neural networks (BP ANN) for optimization in red sea state, Sudan. ANN can recognize the relation between the input and output parameters without explicit physical consideration The ANN model showed much high prediction accuracy of WQI modeling
[81]	India	RMSE, MSE, MAE	WQI prediction was carried out from samples collected from 1944 wells surrounding the Vellore district. ANN was implemented with three different kinds of activation functions namely Tanh, Maxout and rectifier for the classification and prediction of WQI. The novelty of this paper is that WQI is determined by combining the three activation functions like Tanh, Maxout, and rectifier.

Abbreviations

ANFIS: Adaptive Neuro-Fuzzy Inference System

ANN: Artificial Neural Networks

ARIMA: Autoregressive Integrated Moving Averages

B: Bootstrap

B-ANN: Bootstrap Artificial Neural Networks

B-ELM: Bootstrap Extreme Learning Machine

CNN or ConvNet: Convolutional Neural Network

DFS: Demand Forecasting System
DMA: District Metered Areas
E-ANN: Evolutionary Artificial Neural Networks
EC: Evolutionary Computation
EDBD: Extended Delta bar Delta
EKF: Extended Kalman Filter
EKF-PG: Extended Kalman Filter with Programming
ELM: Extreme Learning Machine
EMD: Empirical Mode Decomposition
FFNN: Feedforward Neural Networks
FIS: Fuzzy Inference Systems
FL: Fuzzy Logic
FLR: Fuzzy Linear Regression
GA: Genetic Algorithms
K-NN: K-Nearest Neighbors
LS-SVM: Least-Squares Support Vector Machine
LSTM: Long Short-Term Memory
LTF: Linear Transfer Function
MAPE: Mean Absolute Percentage Error
MARS: Multivariate Adaptive Regression Splines
ML: Machine Learning
MLP: Multilayer Perceptron
MLR: Multiple Linear Regression
MNLR: Multiple Non-Linear Regression
MSE: Mean Square Error
RBNN: Radial Basis Neural Networks
RF: Random Forest
SARMA: Spatial Autoregressive Moving Average
SC: Soft Computing
SEM: Spatial Error Model
SVR: Support Vector Regression
W-ANN: Wavelet Artificial Neural Networks
WB-ANN: Wavelet-Bootstrap Artificial Neural Networks
WDF: Water Demand Forecasting Module
W-ELM: Wavelet Extreme Learning Machine
RRMSE: Relative Root Mean Square Error
NRMSE: Normalized Root Mean Square Error
RMS: Absolute Error Measure
NMSE: Normal Mean Square Error

5. Conclusion

This review has presented an overview of the water quality index forecasting literature appearing from 2011 to 2021, in an attempt to provide some guidance, mostly for researchers and practitioners seeking to adopt methods and models suitable for addressing planning-related decisions that are dependent on future levels of water quality, it should also be used to address short, medium and long-term planning and decisions for researchers seeking to improve the predictive models.

Up till the present time there is no global model that outperforms all the models in water quality index forecasting, but each country has their own model that performs better than other models. Another point concerns the performance of hybrid models which shows more excellent result when compared to the classical models.

Although major advances in AI methods have been made recently, but there is still no new method, such as deep neural networks, has emerged as the best forecasting model. Therefore, water quality index forecasting still remains a research problem, which makes room for researchers to develop hybrid or specific methods for specific applications.

Compliance with ethical standards

Acknowledgments

The authors would like to acknowledge all the cited references for their previous effort toward water quality index modeling and prediction.

References

- [1] D. Sharma and A. Kansal, "Water quality analysis of River Yamuna using water quality index in the national capital territory, India (2000 – 2009)," vol. 0, pp. 147–157, 2011, doi: 10.1007/s13201-011-0011-4.
- [2] G. Elkiran, V. Nourani, and S. I. Abba, "Multi-step ahead modelling of river water quality parameters using ensemble artificial intelligence-based approach," *J. Hydrol.*, vol. 577, no. July, p. 123962, 2019, doi: 10.1016/j.jhydrol.2019.123962.
- [3] S. I. Abba *et al.*, "Comparative implementation between neuro-emotional genetic algorithm and novel ensemble computing techniques for modelling dissolved oxygen concentration," *Hydrol. Sci. J.*, vol. 66, no. 10, pp. 1584–1596, 2021, doi: 10.1080/02626667.2021.1937179.
- [4] B. Mohammadi *et al.*, "Adaptive neuro-fuzzy inference system coupled with shuffled frog leaping algorithm for predicting river streamflow time series," *Hydrol. Sci. J.*, vol. 00, no. 00, pp. 1–14, 2020, doi: 10.1080/02626667.2020.1758703.
- [5] Q. B. Pham *et al.*, "Potential of Hybrid Data-Intelligence Algorithms for Multi-Station Modelling of Rainfall," *Water Resour. Manag.*, vol. 33, no. 15, pp. 5067–5087, 2019, doi: 10.1007/s11269-019-02408-3.
- [6] S. I. Malami, F. H. Anwar, S. Abdulrahman, S. I. Haruna, S. I. A. Ali, and S. I. Abba, "Implementation of hybrid neuro-fuzzy and self-turning predictive model for the prediction of concrete carbonation depth: A soft computing technique," *Results Eng.*, vol. 10, no. May, p. 100228, 2021, doi: 10.1016/j.rineng.2021.100228.
- [7] M. A. S. S.I. Abba, A.S. Maihula, M.B. Jibril, A.M. Sunusi, M.A. Ahmad, "Application of data driven algorithms for the forecasting of non-linear parameter.," *Int. J. Recent Eng. Sci.*, vol. 6, no. 2, pp. 15–20, 2019.
- [8] G. Elkiran and M. Ergil, "The Water Budget Analysis of Girne Region, North Cyprus," no. January, 2016.
- [9] S. Selvakumar, N. Chandrasekar, and G. Kumar, "Hydrogeochemical characteristics and groundwater contamination in the rapid urban development areas of Coimbatore, India," *Water Resour. Ind.*, vol. 17, no. December 2016, pp. 26–33, 2017, doi: 10.1016/j.wri.2017.02.002.
- [10] N. Adimalla and A. K. Taloor, "Hydrogeochemical investigation of groundwater quality in the hard rock terrain of South India using Geographic Information System (GIS) and groundwater quality index (GWQI) techniques," *Groundw. Sustain. Dev.*, p. 100288, 2019, doi: 10.1016/j.gsd.2019.100288.
- [11] M. Reza *et al.*, "Assessment of water quality in groundwater resources of Iran using a modified drinking water quality index (DWQI)," *Ecol. Indic.*, vol. 30, pp. 28–34, 2013, doi: 10.1016/j.ecolind.2013.02.008.
- [12] P. Debels, R. Figueroa, R. Urrutia, R. Barra, and X. Niell, "“ RIVER AND A MODIFIED WATER QUALITY INDEX,” pp. 301–322, 2005, doi: 10.1007/s10661-005-8064-1.
- [13] A. Lumb and T. C. S. J. Bibeault, "A Review of Genesis and Evolution of Water Quality Index (WQI) and Some Future Directions," pp. 11–24, 2011, doi: 10.1007/s12403-011-0040-0.
- [14] S. I. Abba and A. G. Usman, "Modeling of Water Treatment Plant Performance using Artificial Neural Network: Case Study Tamburawa Kano-Nigeria," *Dutse J. Pure Appl. Sci. (DUJOPAS), Vol.*, vol. 6, no. 3, pp. 135–14, 2020.
- [15] S. I. Abba, S. J. Hadi, and J. Abdullahi, "River water modelling prediction using multi-linear regression, artificial neural network, and adaptive neuro-fuzzy inference system techniques," in *Procedia Computer Science*, 2017, vol. 120, pp. 75–82, doi: 10.1016/j.procs.2017.11.212.
- [16] N. Serre and S. Karuppanan, "Journal of African Earth Sciences Groundwater quality assessment using water quality index and GIS technique in Modjo River Basin, central Ethiopia," *J. African Earth Sci.*, vol. 147, no. June, pp. 300–311, 2018, doi: 10.1016/j.jafrearsci.2018.06.034.
- [17] R. E. Rabeiy, "Assessment and modeling of groundwater quality using WQI and GIS in Upper Egypt area," 2017, doi: 10.1007/s11356-017-8617-1.

- [18] S. Africa *et al.*, “No No Covariance Structure Analysis of Health-Related Indices in the Elderly at Home with a Focus on Subjective Feelings of HealthTitle,” *Hum. Relations*, vol. 3, no. 1, pp. 1–8, 2020, [Online]. Available: http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,shib&db=bth&AN=92948285&site=eds-live&scope=site%0Ahttp://bimimpactassessment.net/sites/all/themes/bcorp_impact/pdfs/em_stakeholder_engagement.pdf%0Ahttps://www.glo-bus.com/help/helpFiles/CDJ-Pa.
- [19] H. Tao, “Groundwater Level Prediction using Machine Learning Models : A,” no. March, 2022.
- [20] A. S I, M. A S, J. M B, S. A M, A. M A, and S. M A, “Application of data-driven algorithms for the forecasting of non-linear parameter,” *Int. J. Recent Eng. Sci.*, vol. 6, no. 2, pp. 14–19, 2019, doi: 10.14445/23497157/ijres-v6i2p103.
- [21] M. Dehghani, B. Saghafian, and N. Saleh, “Uncertainty analysis of streamflow drought forecast using artificial neural networks and Monte-Carlo simulation,” 2013, doi: 10.1002/joc.3754.
- [22] S. Mandal, S. S. Mahapatra, S. Adhikari, and R. K. Patel, “Modeling of Arsenic (III) Removal by Evolutionary Genetic Programming and Least Square Support Vector Machine Models,” no. Iii, 2014, doi: 10.1007/s40710-014-0050-6.
- [23] R. Gupta, A. N. Singh, and A. Singhal, “Application of ANN for Water Quality Index,” vol. 9, no. 5, pp. 688–693, 2019, doi: 10.18178/ijmlc.2019.9.5.859.
- [24] U. Ozveren, “An artificial intelligence approach to predict a lower heating value of municipal solid waste,” *Energy Sources, Part A Recover. Util. Environ. Eff.*, vol. 38, no. 19, pp. 2906–2913, 2016, doi: 10.1080/15567036.2015.1107864.
- [25] H. Uzun, Z. Yıldız, J. L. Goldfarb, and S. Ceylan, “Improved prediction of higher heating value of biomass using an artificial neural network model based on proximate analysis,” *Bioresour. Technol.*, vol. 234, pp. 122–130, 2017, doi: 10.1016/j.biortech.2017.03.015.
- [26] S. I. Abba, A. M. Sunusi, and Y. S. Yakasai, “APPRAISAL OF YEARLY AND SEASONAL WATER QUALITY VARIATION AT CHAMBAL RIVER USING CANADIAN COUNSEL OF MINISTER OF ENVIRONMENT MODEL (CCME).”
- [27] S. I. Abba, G. Saini, and R. A. Abdulkadir, “Evaluation of Yearly Water quality index and Seasonal Variation of Yamuna River at Hathikund Station Using Canadian Council of Ministers of Environment Water Quality Index (CCME) Model,” *IOSR J. Eng. www.iosrjen.org ISSN*, vol. 07, no. 01, pp. 2250–3021, 2017.
- [28] V. Nourani, G. Elkiran, and S. I. Abba, “Wastewater treatment plant performance analysis using artificial intelligence – an ensemble approach,” *Water Sci. Technol.*, 2018, doi: 10.2166/wst.2018.477.
- [29] S. I. Abba, S. Jasim, S. Sh, S. Q. Salih, and R. A. Abdulkadir, “Evolutionary computational intelligence algorithm coupled with self-tuning predictive model for water quality index determination,” *J. Hydrol.*, vol. 587, no. April, p. 124974, 2020, doi: 10.1016/j.jhydrol.2020.124974.
- [30] S. I. Abba *et al.*, “Emerging evolutionary algorithm integrated with kernel principal component analysis for modeling the performance of a water treatment plant,” *J. Water Process Eng.*, vol. 33, no. December 2019, p. 101081, 2020, doi: 10.1016/j.jwpe.2019.101081.
- [31] M. S. Gaya *et al.*, “Estimation of water quality index using artificial intelligence approaches and multi-linear regression,” vol. 9, no. 1, pp. 126–134, 2020, doi: 10.11591/ijai.v9.i1.pp126-134.
- [32] N. M. Gazzaz, M. Kamil, A. Zaharin, and H. Juahir, “Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors,” *Mar. Pollut. Bull.*, vol. 64, no. 11, pp. 2409–2420, 2012, doi: 10.1016/j.marpolbul.2012.08.005.
- [33] M. Fahmi, M. Nasir, M. A. Zali, H. Juahir, H. Hussain, and S. M. Zain, “ENVIRONMENTAL HEALTH Application of receptor models on water quality data in source apportionment in Kuantan River Basin,” pp. 1–12, 2012.
- [34] O. P. Okeke *et al.*, “Performance analysis and control of wastewater treatment plant using Adaptive Neuro-Fuzzy Inference System (ANFIS) and Multi-Linear Regression (MLR) techniques,” pp. 1–16, 2022.
- [35] M. M. Jibril *et al.*, “An overview of streamflow prediction using random forest algorithm,” pp. 0–7, 2022.
- [36] T. Khayyun, A. Hameed, and M. J. Al, “PREDICTION OF WATER QUALITY OF EUPHRATES RIVER BY USING ARTIFICIAL NEURAL NETWORK MODEL (SPATIAL AND TEMPORAL STUDY),” no. November 2018, 2014.
- [37] A. Baghban and T. Ebadi, “ANFIS modeling of higher heating value of wastes : Application to fuel upgrading,” *Energy Sources, Part A Recover. Util. Environ. Eff.*, pp. 1–7, 2018, doi: 10.1080/15567036.2017.1344746.

- [38] M. Onifade, A. I. Lawal, A. E. Aladejare, S. Bada, and M. A. Idris, "Prediction of gross calorific value of solid fuels from their proximate analysis using soft computing and regression analysis," *Int. J. Coal Prep. Util.*, pp. 1–15, 2019, doi: 10.1080/19392699.2019.1695605.
- [39] A. Dashti, A. Sajadi, M. Raji, A. Razmi, and S. Ceylan, "Estimation of biomass higher heating value (HHV) based on the proximate analysis: Smart modeling and correlation," *Fuel*, vol. 257, pp. 1–11, 2019, doi: 10.1016/j.fuel.2019.115931.
- [40] S. Babak, H. Seyed, A. Sharafati, D. Motta, and Z. M. Yaseen, "River Water Quality Index prediction and uncertainty analysis: a comparative study of machine learning models," *Biochem. Pharmacol.*, no. September, p. 104599, 2020, doi: 10.1016/j.jece.2020.104599.
- [41] S. Tiwari, R. Babbar, and G. Kaur, "Performance Evaluation of Two ANFIS Models for Predicting Water Quality Index of River Satluj (India)," vol. 2018, 2018.
- [42] M. S. Gaya, S. I. Abba, A. M. Abdu, and A. I. Tukur, "Estimation of water quality index using artificial intelligence approaches and multi-linear regression," vol. 9, no. 1, pp. 126–134, 2020, doi: 10.11591/ijai.v9.i1.pp126-134.
- [43] Q. Feng, J. Zhang, X. Zhang, and S. Wen, "Proximate analysis based prediction of gross calorific value of coals : A comparison of support vector machine, alternating conditional expectation and artificial neural network," *Fuel Process. Technol.*, vol. 129, pp. 120–129, 2015, doi: 10.1016/j.fuproc.2014.09.001.
- [44] J. Xing, K. Luo, H. Wang, Z. Gao, and J. Fan, "A comprehensive study on estimating higher heating value of biomass from proximate and ultimate analysis with machine learning approaches," *Energy*, vol. 188, pp. 1–11, 2019, doi: 10.1016/j.energy.2019.116077.
- [45] F. Kamyab-talesh, S. Mousavi, and M. Khaledian, "WATER QUALITY AND PROTECTION : Prediction of Water Quality Index by Support Vector Machine : a Case Study in the Sefidrud Basin, Northern Iran 1," vol. 46, no. 1, pp. 112–116, 2019, doi: 10.1134/S0097807819010056.
- [46] S. G. Meshram, H. R. Pourghasemi, S. I. Abba, E. Alvandi, C. Meshram, and K. M. Khedher, "A comparative study between dynamic and soft computing models for sediment forecasting," *Soft Comput.*, vol. 0123456789, 2021, doi: 10.1007/s00500-021-05834-x.
- [47] M. Benaafi, M. A. Yassin, A. G. Usman, and S. I. Abba, "Neurocomputing Modelling of Hydrochemical and Physical Properties of Groundwater Coupled with Spatial Clustering, GIS, and Statistical Techniques," 2022.
- [48] M. Saood *et al.*, "New generation neurocomputing learning coupled with a hybrid neuro-fuzzy model for quantifying water quality index variable : A case study from Saudi Arabia," *Ecol. Inform.*, vol. 70, no. May, p. 101696, 2022, doi: 10.1016/j.ecoinf.2022.101696.
- [49] S. I. Abba, G. Elkiran, and V. Nourani, "Improving novel extreme learning machine using PCA algorithms for multi-parametric modeling of the municipal wastewater treatment plant," vol. 215, no. October 2019, pp. 414–426, 2021, doi: 10.5004/dwt.2021.26903.
- [50] S. I. Abba *et al.*, "Emerging evolutionary algorithm integrated with kernel principal component analysis for modeling the performance of a water treatment plant," *J. Water Process Eng.*, vol. 33, no. February, p. 101081, 2020, doi: 10.1016/j.jwpe.2019.101081.
- [51] W. C. Leong, A. Bahadori, J. Zhang, and Z. Ahmad, "Prediction of water quality index (WQI) using support vector machine (SVM) and least square- support vector machine (LS-SVM)," *Intl. J. River Basin Manag.*, vol. 0, no. 0, pp. 1–8, 2019, doi: 10.1080/15715124.2019.1628030.
- [52] R. A. Abdulkadir, S. I. A. Ali, S. I. Abba, and P. Esmaili, "Forecasting of daily rainfall at Ercan Airport Northern Cyprus: a comparison of linear and non-linear models," *Desalin. Water Treat.*, vol. 177, no. May 2019, pp. 297–305, 2020, doi: 10.5004/dwt.2020.25321.
- [53] S. I. Abba, S. J. Kawu, H. S. Maccido, S. M. Lawan, G. Najashi, and A. Y. Sada, "Short-term load demand forecasting using nonlinear dynamic grey-black-box and kernel optimization models: a new generation learning algorithm," pp. 1–6, 2022, doi: 10.1109/icmeas52683.2021.9692314.
- [54] S. I. Abba, A. G. Usman, and S. Işik, "Simulation for response surface in the HPLC optimization method development using artificial intelligence models: A data-driven approach," *Chemom. Intell. Lab. Syst.*, p. 104007, 2020, doi: 10.1016/j.chemolab.2020.104007.

- [55] A. G. Usman, M. H. Ahmad, R. N. Danraka, and S. I. Abba, "The effect of ethanolic leaves extract of *Hymenodictyon floribundum* on inflammatory biomarkers: a data-driven approach," *Bull. Natl. Res. Cent.*, vol. 45, no. 1, 2021, doi: 10.1186/s42269-021-00586-y.
- [56] A. G. Usman, S. Isik, S. I. Abba, and F. Mericli, "Artificial intelligence based models for the qualitative and quantitative prediction of a phytochemical compound using HPLC method," *Turkish J. Chem.*, vol. 44, no. 5, 2021, doi: 10.3906/kim-2003-6.
- [57] A. G. USMAN, S. İŞİK, S. I. ABBA, and F. MERİÇLİ, "Artificial intelligence-based models for the qualitative and quantitative prediction of a phytochemical compound using HPLC method," *Turkish J. Chem.*, vol. 44, no. 5, pp. 1339–1351, 2020, doi: 10.3906/kim-2003-6.
- [58] S. I. Abba, M. Benaafi, A. G. Usman, and I. H. Aljundi, "Sandstone groundwater salinization modelling using physicochemical variables in Southern Saudi Arabia : Application of novel data intelligent algorithms," *Ain Shams Eng. J.*, no. xxxx, p. 101894, 2022, doi: 10.1016/j.asej.2022.101894.
- [59] A. G. Usman, S. İşik, and S. I. Abba, "A Novel Multi-model Data-Driven Ensemble Technique for the Prediction of Retention Factor in HPLC Method Development," *Chromatographia*, 2020, doi: 10.1007/s10337-020-03912-0.
- [60] M. H. Ahmad, A. G. Usman, and S. I. Abba, "Comparative performance of extreme learning machine and Hammerstein–Weiner models for modelling the intestinal hyper-motility and secretory inhibitory effects of methanolic leaf extract of *Combretum hypopilinum* Diels (Combretaceae)," *Silico Pharmacol.*, vol. 9, no. 1, 2021, doi: 10.1007/s40203-021-00090-1.
- [61] A. K. K. V. M. Wagh and A. A. M. B. N. Umrikar, "Prediction of water quality index using artificial neural network and multiple linear regression modelling approach in Shivganga River basin, India," *Model. Earth Syst. Environ.*, vol. 0, no. 0, p. 0, 2019, doi: 10.1007/s40808-019-00581-3.
- [62] S. H. Ewaid, S. A. Abed, and S. A. Kadhum, "Prediction the Tigris River Water Quality within Baghdad, Iraq by Using Water," *Environ. Technol. Innov.*, 2018, doi: 10.1016/j.eti.2018.06.013.
- [63] K. Chapi *et al.*, "Flash-flood susceptibility assessment using multi-criteria decision making and machine learning supported by remote sensing and GIS techniques," *Remote Sens.*, vol. 12, no. 1, pp. 1–30, 2020, doi: 10.3390/w10111536.
- [64] A. G. Usman, S. İşik, and S. I. Abba, "Hybrid data-intelligence algorithms for the simulation of thymoquinone in Hybrid data - intelligence algorithms for the simulation of thymoquinone in HPLC method development," *J. Iran. Chem. Soc.*, no. January, 2021, doi: 10.1007/s13738-020-02124-5.
- [65] * Author For Correspondence and S. I. Abba, "Modeling of Water Treatment Plant Performance using Artificial Neural Network: Case Study Tamburawa Kano-Nigeria," *Dutse J. Pure Appl. Sci.*, vol. 6, no. 3, pp. 135–144, 2020.
- [66] S. I. Abba *et al.*, "Hybrid machine learning ensemble techniques for modeling dissolved oxygen concentration," *IEEE Access*, vol. 8, no. September, pp. 157218–157237, 2020, doi: 10.1109/ACCESS.2020.3017743.
- [67] Z. S. A. A.S. Mubarak, Parvaneh Esmaili, M. S. G., R.A. Abdulkadir, M. Ozsoz, and S. I. A., Gaurav Saini, "Metro-environmental data approach for the prediction of chemical oxygen demand in new Nicosia wastewater treatment plant," vol. 27049, pp. 1–10, 2021, doi: 10.5004/dwt.2021.27049.
- [68] R. K. Patel, "Prediction of Water Quality Index Using Neuro Fuzzy Inference System," pp. 175–191, 2011, doi: 10.1007/s12403-011-0054-7.
- [69] M. Ali and A. M. Qamar, "Data Analysis, Quality Indexing and Prediction of Water Quality for the Management of Rawal Watershed in Pakistan," pp. 108–113, 2013.
- [70] J. J. Carbajal-hernández, L. P. Sánchez-fernández, L. A. Villa-vargas, J. A. Carrasco-ochoa, and J. F. Martínez-trinidad, "Water quality assessment in shrimp culture using an analytical hierarchical process," *Ecol. Indic.*, vol. 29, pp. 148–158, 2013, doi: 10.1016/j.ecolind.2012.12.017.
- [71] R. Mohammadpour, S. Shaharuddin, and C. K. Chang, "Prediction of water quality index in constructed wetlands using support vector machine," 2014, doi: 10.1007/s11356-014-3806-7.
- [72] N. M. Gazzaz, M. K. Y, M. F. Ramli, H. Juahir, and A. Z. Aris, "Artificial Neural Network Modeling of the Water Quality Index Using Land Use Area s as Predictors," 2011, doi: 10.2175/106143014X14062131179276.
- [73] M. Sakizadeh, "Artificial intelligence for the prediction of water quality index in groundwater systems," *Model. Earth Syst. Environ.*, vol. 2, no. 1, pp. 1–9, 2016, doi: 10.1007/s40808-015-0063-9.

- [74] P. Taylor, K. Sinha, and P. Das Saha, “Desalination and Water Treatment Assessment of water quality index using cluster analysis and artificial neural network modeling : a case study of the Hooghly River basin, West Bengal, India,” no. April 2014, pp. 37–41, doi: 10.1080/19443994.2014.880379.
- [75] R. Mohammadpour, S. Shaharuddin, and N. Azazi, “Prediction of water quality index in free surface constructed wetlands,” *Environ. Earth Sci.*, 2016, doi: 10.1007/s12665-015-4905-6.
- [76] Y. Liu, Y. Liang, S. Liu, D. S. Rosenblum, Y. Zheng, and S. Member, “Predicting Urban Water Quality with Ubiquitous Data,” pp. 1–14.
- [77] Z. Ahmad, N. A. Rahim, A. Bahadori, and J. Zhang, “Improving water quality index prediction in Perak River basin Malaysia through a combination of multiple neural networks Improving water quality index prediction in Perak River basin Malaysia through a combination of multiple neural networks,” *Intl. J. River Basin Manag.*, vol. 0, no. 0, pp. 1–9, 2016, doi: 10.1080/15715124.2016.1256297.
- [78] M. Hameed, S. Shartoo, S. Zaher, and M. Yaseen, “Application of artificial intelligence (AI) techniques in water quality index prediction : a case study in tropical region, Malaysia,” *Neural Comput. Appl.*, 2016, doi: 10.1007/s00521-016-2404-7.
- [79] 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,” *Model. Earth Syst. Environ.*, vol. 0, no. 0, p. 0, 2018, doi: 10.1007/s40808-018-0437-x.
- [80] M. Ismael, A. Mokhtar, M. Farooq, and X. Lü, “Groundwater for Sustainable Development Assessing drinking water quality based on physical, chemical and microbial parameters in the Red Sea State, Sudan using a combination of water quality index and artificial neural network model,” *Groundw. Sustain. Dev.*, vol. 14, no. May, p. 100612, 2021, doi: 10.1016/j.gsd.2021.100612.
- [81] S. Vijay and K. Kamaraj, “Prediction of Water Quality Index in Drinking Water Distribution System Using Activation Functions Based Ann,” pp. 535–553, 2021