MODELLING OF WATER QUALITY PARAMETERS OF LOWER USUMA DAM RESERVOIR, ABUJA, USING ARTIFICIAL NEURAL NETWORK (ANN)

Haruna Garba[1*] and Francis Peter[2]

Abstract

The water quality parameters of Lower Usuma Dam, Abuja, were analysed and modelled using an artificial neural network (ANN). Monthly water quality parameters of pH, turbidity, electrical conductivity, total dissolved solids, and total hardness for a duration of 6 years (2017–2021) were obtained from the Water Laboratory Department of the Federal Capital Territory of Nigeria (F.C.T.) Water Board, Abuja. Microsoft Excel was used to analyse the trends of these parameters. The Artificial Neural Network (ANN) was used to develop three model equations for the prediction of electrical conductivity, total dissolved solids, and total hardness, respectively, with pH and turbidity as input parameters. F-tests and t-tests were used to validate each model using Microsoft Excel. The error analysis and performance evaluation of the applied models were also done to evaluate the goodness and suitability of each of the models. The coefficients of determination (R2) between the parameters were 0.89085, 0.83156, and 0.86931 for testing, training, and validation, respectively. A very strong relationship between the predictors (pH and turbidity) and the response variables (electrical conductivity, total dissolved solids, and total hardness) was established. The root mean square errors were 11.2, 13.8, and 5.54. Thus, the total hardness model is the best among them because it has the lowest predictive error. The model validation carried out for the F-test and t-test for electrical conductivity, total dissolved solids, and total hardness, respectively, shows that F critical is greater than F, as well as t critical is greater than t-stat. This further shows that the ANN model is fit for the prediction of water quality.

Received: *21 February, 2024*
Revised: *11 June, 2024*
Accepted: *15 August, 2024*

[1,2]*Department of Civil Engineering, Nigerian Defence Academy, PMB 2109 Kaduna, Nigeria.*

****Corresponding author:***
hgarba@nda.edu.ng

Keywords:
Artificial neural network, Dam, Modelling, Parameters, Water quality

1.0 INTRODUCTION

All living things depend on water as a vital natural resource for survival. Water is mostly needed by humans for domestic, industrial, and agricultural purposes (Ehya and Saeedi, 2019). According to current data on water consumption worldwide, ten percent (10%) of available fresh water is used for home purposes (drinking, cooking, bathing, etc.), seventy percent (70%) is used for agriculture, mostly irrigation, and twenty percent (20%) is used for industrial purposes (Boretti and Rosa, 2019). According to Boretti and Rosa (2019), domestic, agricultural, and industrial purposes are expected to increase dramatically in the years to come. By 2050, there will likely be a 20–30% increase in the demand for water for a variety of uses. The human population is rapidly increasing around the world. Hence, there is an increased need for clean water to carry out human activities. Even though there is a greater need for water, there is less freshwater available because of pollution and fewer sources. Contaminated water resources have negative health effects and have a significant negative impact on both the environment and overall human wellbeing; ensuring water quality for domestic, drinking, and agricultural purposes is crucial and desirable (Egbueri *et al.,* 2019). According to Mrunmayee (2014), contamination of surface water by chemical, physical, and microbiological contaminants is a global epidemic. The physical, chemical, microbial, and biological conditions in the water courses and subsurface aquifers have an impact on fish survival and growth, biodiversity, conservation efforts, leisure activities like swimming and boating, industrial and municipal water supply, agricultural uses like irrigation and livestock watering, waste disposal, and all other water uses (Singh *et al.,* 2005).

In 2021, Marian *et al.* (2021) explore the use of ANN to predict the monthly values of dissolved oxygen (DO) and electrical conductivity (EC) to analyse the water quality parameters of four variables and discharges. The correlation coefficient, root-mean-square error, and mean absolute error were the statistical criteria explored for evaluating the model's performance. The potential of ANN for simulating relevance between water quality parameters indicates that ANN can discern the pattern of water quality to offer an appropriate prediction of changes in water quality data. In forecasting water parameters, use ANN for irrigation purposes. Uba *et al.* (2021) analysed the water quality index of four parameters [PH, Total Dissolved Solids (TDS), Electrical Conductivity (EC), and Sodium (Na) of the Ele river at different locations using ANN. Results from the analysis showed that the PH ranged from 6.01 to 6.87, while the TDS ranged from 3.01 to 5.76 and 40.42 to 73.45, respectively. Findings from the study showed that the R2 values range from 0.956 to 0.967, 0.953 to 0.970, 0.951 to 0.967, and 0.953 to 0.968 for each of pH, TDS, EC, and Na, while the forecast performance evaluation showed R^2 values of 0,022 to 0.088, 0.12 to 0.087, 0.015 to 0.085, and 0.014 to 0.084.

The efficiency of hybrid deep neural networks and the multivariate water quality forecasting model in the aquaculture ecosystem was examined by Elias *et al.* (2023) by developing a novel hybrid deep learning neural network multivariate water quality parameters forecasting model with the aid of the ensemble empirical mode decomposition (EEMD) method, deep learning long-short term memory (LSTM), neural network (NN), and multivariate linear regression (MLR) method. The performance of the novel hybrid water quality forecasting model is validated by comparing the forecasted results with water quality parameter data. The forecast accuracy of the result suggested that the novel hybrid water quality forecasting model can be used as a valuable support tool for water quality management in aquaculture industries.

An efficient river water quality indicator prediction model was designed and built by Jitha in 2023. Data were collected from eleven sampling stations at different points on the Bhavani River in Kerala and Tamil Nadu, India. The water quality index was computed using twenty-eight different parameters that affect water quality. Feature selection and data normalization are applied to develop an efficient river water quality database. The water quality index (WQI) prediction model was built using deep learning architectures. The performance of the deep learning-based WQI prediction model is compared with that of traditional learning-based models. The performance analysis indicates that the GRU-based prediction model shows promising results in predicting water quality.

A simple architecture consisting of an artificial neural network model for water quality and water consumption prediction was proposed by Furaun *et al.* (2022). An artificial neural network (ANN) consisting of one hidden layer and a couple of dropout and activation layers is utilised. The approach is tested using two data sets for predicting water quality and water consumption. Results show a 0.96 accuracy for water quality prediction. A 0.99 score is obtained for water consumption prediction.

The cost of labour and materials for many chemical tests, as observed by Mrunmayee (2014), can be somewhat reduced by using water quality models as effective tools to forecast and simulate contaminant transport in aquatic environments. The aim of this research work, therefore, is to leverage the application of ANN as a cost-effective means and approach for analysing and predicting water quality parameters. The ANN tool was used to develop model equations for the prediction of electrical conductivity, total dissolved solids, and total hardness for the Lower Usuma Dam reservoir. Though ANNs have been extensively used in modelling various environmental parameters, including water quality parameters, however, they come with their own set of limitations in the following contexts: complexity requirements, complexity and interpretability, sensitivity of input data, and limited transferability. To address these limitations and pave the way for future research into using ANNs for modelling water quality parameters, the following suggestions can be considered: Integration with other models, Data quality improvement, integration techniques, Feature selection and dimensionality reduction, Models explain ability.

2.0 MATERIALS AND METHODS

2.1 Description of Study Area

The lower Usuma dam is located at latitude 7o 25' 16" east and longitude 9o 01' 12" north. The dam is constructed across the River Usuma and is situated 10 kilometres from Bwari and 26 kilometres from the heart of Abuja city (Figure 1).The lower Usuma dam has an installed capacity of about 120 million m3 of untreated, raw water in its reservoir. The main dam and the saddle dam are the two sides of the dam. The main dam embankment is 10 meters in height, 1.3 kilometres long, and 47 meters high, while the saddle dam is 470 meters long, 15 meters high, and 10 meters wide. The dam occupies a total area of 2,500,000 m^2.

Figure 1: Location map of Lower Usuma Dam treatment plant

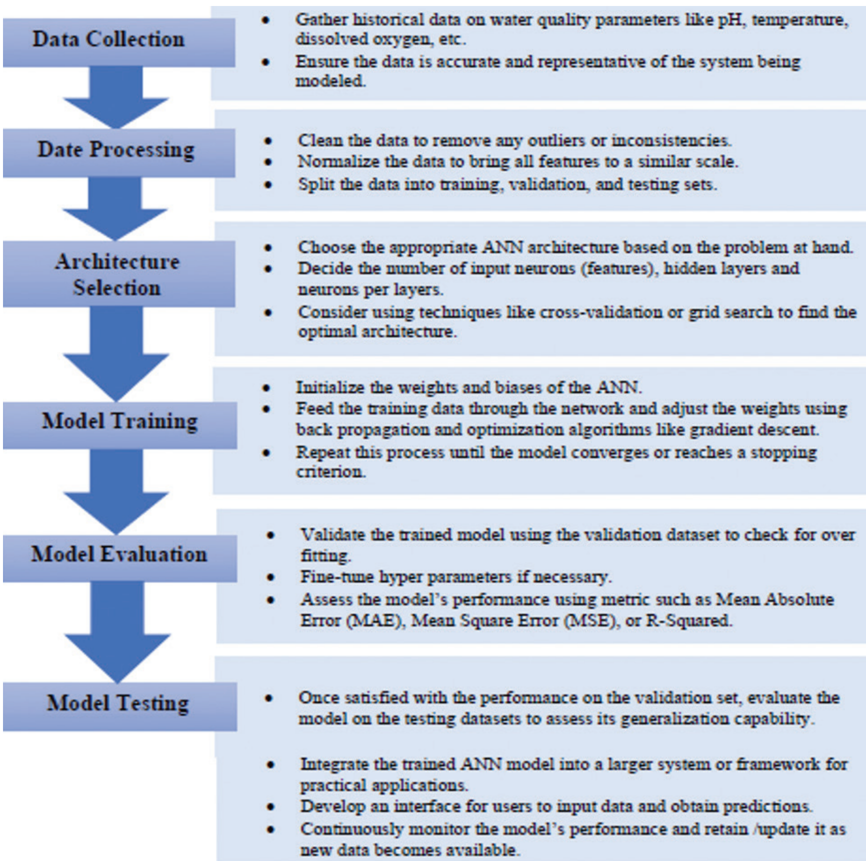

Figure 2: Flow chart of methodology used in the study

2.2 Data Collection

The monthly water quality samples were collected and analysed for the months of January to December from 2017 to 2021. Five physical and chemical water quality parameters were selected for the analysis. The parameters were pH, turbidity, electrical conductivity, total dissolved solids, and total hardness as $CaCO_3$. The flow chart methodology used in the study is presented in Figure 2.

2.3 Method of Analysis

Spearman's correlation analysis, trend analysis, and summary statistics were used to investigate the temporal and spatial variations and to interpret the large and complex water quality data sets that were collected. The data set was divided into three, namely, training, testing, and validation data sets (Nguyen *et al.,* 2018). The Spearman's rank correlation coefficient was estimated temporally for each parameter from 2017 to 2021 to know the positive and negative trends.

2.4 Tools for Analysis

The most effective learning method for multilayer neural network topologies was the back propagation algorithm. The feed-forward-back propagation neural network (BPNN) always has three layers: an input layer, a hidden layer, and an output layer (Figure 3). Before analysing fresh data for the following process, a network first needs to be trained. The neurons in each layer, known as layers, were connected to one another by weights (Figure 3); these neurons were known as neurons in the input layer, which delivered their output to neurons in the hidden layer as input, and similar connections existed between the hidden and output layers. Depending on the issue at hand, the number of hidden layers and the density of their neurons were altered. As with input and output variables, there were an equal number of input and output neurons, (ASCE Task Committee, 2000).

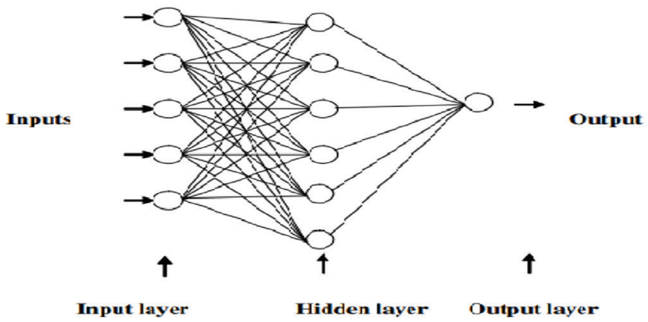

Figure 3: Structure of a multi-layer feed forward artificial neural network model

Values known as biases (Figure 4) were introduced in the transfer functions and were referred to as the temperature of a neuron in order to distinguish between the various processing units. The transfer function filtered the summed signals that were received from this neuron, but the bias behaved like a weight and had an input of 1 (Figure 3). The transfer functions were straightforward step functions, either linear or non-linear, that were intended to convert neurons' or layers' net output to their actual output. All the neurons in the BPNN were connected to a bias neuron and a transfer function, except for the input layer. Transfer function use was dependent on the neural network's intended use. The output layer was created, and the solution-related vectors were calculated (Archana and Prashant, 2015).

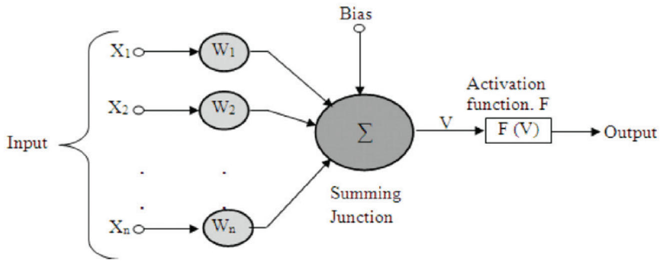

Figure 4: Basic elements of Artificial Neuron

Similarly, the architecture of the ANN model, which uses the non-associated flow rule for its analysis, consists of ten (10) neurons in the hidden layer, three (3) neurons in the output layer, and two (2) neurons in the input layer. Furthermore, the tan-sigmoid function was used as the nonlinear activation function (transfer function) for the hidden layer, while for the output layer, a pure linear function was used as the activation function. The feed-forward neural network trained by the back propagation algorithm was used (Ali, *et al,* 2009).

In this work, two approaches for selecting data were used to build and evaluate the models. In the first approach, the water quality data were divided into two sets. The first set contained 70% of the records and was used as a training set; the second set contained 30% of the records and was used as an over fitting test (Roza and Mohsen, 2021), with 15% for validation and 15% for testing. The training set is used to train the model, and the validation set is used to tune hyper parameters to prevent overfitting. An ANN architecture (feed-forward neuron network) was considered appropriate and used for the study. The number of input neurons (features), hidden layers, and the number of neurons in each layer were determined. Four input neurons corresponding to the number of water quality parameters were used.

2.4.1 Performance Evaluation

There are numerous statistical metrics that can be used to evaluate the suitability or goodness of any given model. In the current study, root-mean-square error (RMSE), coefficient of correlation (R), and coefficient of determination (R^2) performance evaluation statistics were used for ANN training. The (R^2) values were calculated analytically by calculating the square of the correlation coefficient (R), whilst the RMSE and R values were taken from the ANN (Thair and Abdul, 2014). The RMSE is calculated using Equation 1.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{1}$$

Where n is the number of observations of the data points, represent the actual n values of observations and represent the predicted or estimated values of observations. Microsoft Excel was used to carry out F- and T-tests for model validation and plot line charts in order to analyse the trend of each of the water quality parameters over time. The monthly water quality parameters (pH, turbidity, electrical conductivity, total dissolved solids, and total hardness), obtained from the F.C.T. Water Board, Abuja, were used for the model validation. Both the measured and the ANN-predicted outcomes for each water quality parameter mentioned above were used for the validation of the electrical conductivity, total dissolved solids, and total hardness models; pH and turbidity were used as input parameters. The t-test and F-test are two commonly used statistical methods in hypothesis testing and analysis of variance (ANOVA). The t-test is used to determine if there is a significant difference between the means of the observed and predicted values of model parameters. The significance of the test provides insight into whether the difference observed between the modelled and predicted parameters is likely due to chance variation or is statistically significant. The F-test, on the other hand, is used in the analysis of variance to compare the variance between the two groups. It is significant in determining whether the differences between multiple groups are due to actual differences between the groups, or could occur due to chance.

3.0 RESULTS AND DISCUSSIONS

3.1 Water Quality Parameters

The descriptive statistics and summary statistics for each of the five (5) raw water quality parameters under consideration were carried out using Microsoft Excel, as well as the various line charts showing data trends for each of the parameters from 2017 to 2022. These results are shown in Figure 5. A summary of the mean, standard error, median, standard deviation, and sample variance is presented in Table 1.

The analysis of the pH shows a range of 6.45–9.61, which is slightly above that of the Nigeria Standard for Drinking Water Quality (NSDWQ) range of 6.5–8.5. The lowest and highest values were recorded in December 2017 and February 2022, respectively, as shown in Figure 5 and Table 1. However, the mean pH value is 7.00.

The turbidity is within the range of 1.6–36.21 NTU, which is by far higher than the NSDWQ standard of 5.0 NTU. The mean value of turbidity is 7.11. The lowest and highest values were recorded in January/February 2017 and October 2022, respectively, as shown in Figure 5 and Table 1.

The electrical conductivity falls within the range of 16.8–89.9 µS/cm, which is within the NSDWQ acceptable limit of 1000 µS/cm. The lowest and highest values were recorded in November 2019 and April 2019, respectively, as shown in Figure 5 and Table 1. However, the mean EC value is 74.35 µS/cm.

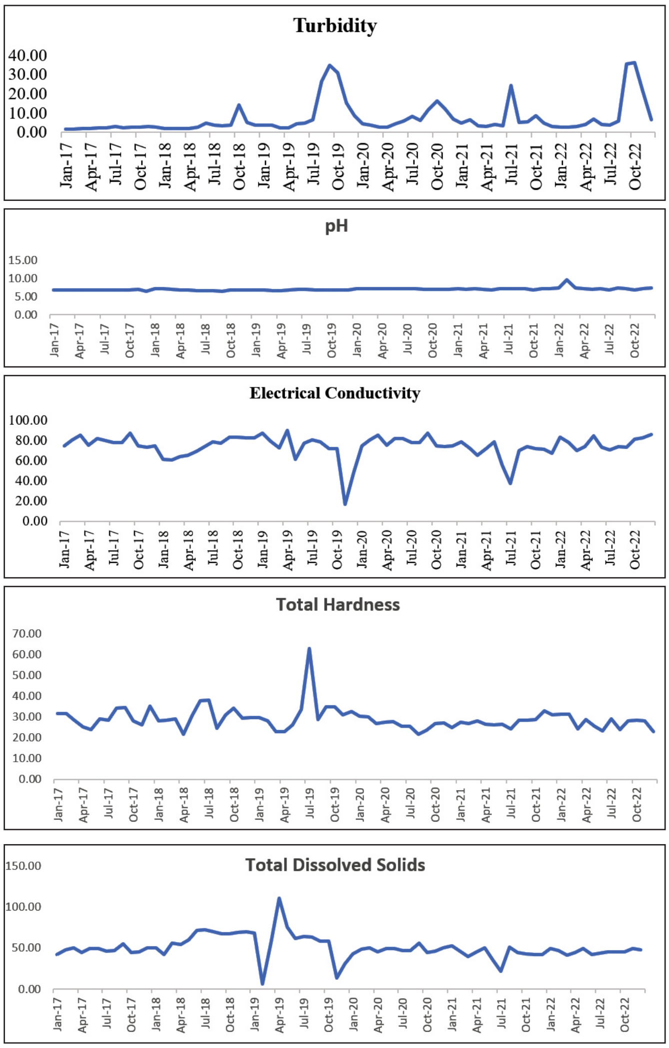

Figure 5: Line chart showing turbidity, pH, electrical conductivity, total hardness and total dissolved solids

The total dissolved solids (TDS) fall within the range of 6.5–110.7 mg/l and are within the NSDWQ permissible limit of 500 mg/l. The lowest and highest values were recorded in February 2019 and April 2019, respectively, as shown in Figure 4 and Table 6. However, the mean dissolved solid TDS value is 50.68 mg/l.

The total hardness is within a range of 21.5–63.0 mg/l and is within the NSDQW permissible limit of 150 mg/l. The lowest and highest values were recorded in August 2020 and July 2019, respectively, as shown in Figure 7 and Table 5. However, the mean total hardness value is 28.91 mg/l.

Table 1: Summary statistics of pH, turbidity, electrical conductivity, total hardness and total dissolved solids

Mean	Standard Error	Median	Mode	Standard Deviation	Sample Variance	Minimum	Maximum
7.0039	0.0432	7	6.9	0.3668	0.1346	6.45	9.61
7.1069	0.9832	3.85	2.7	8.3430	69.6067	1.62	36.21
74.3496	1.3177	75.05	78.2	11.1808	125.0111	16.8	89.9
50.6786	1.6483	49.155	44.9	13.9863	195.6188	6.5	110.7
28.9095	0.6468	28.38	28.9	5.4883	30.1222	21.5	63

3.2 Artificial Neural Analysis (ANN) Analysis

Water quality indicators (electrical conductivity, total dissolved solids, and total hardness) were modelled using the ANN model, and the Feed Forward Multilayer Perceptron (FFMLP), which is a learning algorithm and a feed forward neural network as described earlier, was executed in MATLAB (R2016a).

The ANN model performed excellently in both the training, testing, and validation data sets, based on the R (correlation coefficient) and R^2 (coefficient of determination) values in Table 2. The R values are 0.94385, 0.9119, and 0.93237 for the testing, training, and validation datasets, respectively. Consequently, the R^2 values for the testing, training, and validation datasets are 0.89085, 0.83156, and 0.86931, respectively. This is similar to the result obtained by (Khandelwal and Singh, 2005). A statistical instrument that assesses the strength of the linear relationship between experimental and expected values is the correlation coefficient (R).

The results of the root-mean-square error (RMSE) are displayed in Table 3 and are also consistent with those of the R, showing a close range for all the training, testing, and validation data sets. RMSE revealed that the performance for electrical conductivity and total hardness was lower when compared with total dissolved solids. In other words, the electrical conductivity and total hardness models have the lowest prediction error for the training and validation data sets. The total hardness model is the best among them because it has the lowest RMSE value of 5.54. The ANN model performed very well, as their coefficient of multiple determinations, R^2, was very close to 1, which is in agreement with the studies of (Awu, *et al,* 2017) and (Abrahart *et al,* 2005).

Table 2: R and R^2 values for the ANN Model

Data Set	R-Value	R^2 Value
Testing	0.94385	0.89085
Training	0.9119	0.83156
Validation	0.93237	0.86931

Table 3: Water quality parameters error indices for ANN model

S/N	Water Quality Parameter	RMSE
1.	Electrical Conductivity	11.2
2.	Total Dissolved Solids	13.8
3.	Total Hardness	5.54

3.3 Multilinear Regression (MLR) Analysis

The MLR model was used as the standard approach to simulate the system's linear interactions. It frequently serves as the non-linear models' benchmark comparison model. In the ANN model calibration, a computer program of multiple regressions is used to obtain a set of coefficients for a linear model and determine how well the linear model represents the observed data (Muhammad *et al,* 2020).

Multiple linear regression (MLR) was applied in this work to justify the relationship between the water quality parameters. An MLR model takes the form:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{p-1} x_{p-1} + \varepsilon \quad (2)$$

Where Y is the response variable, and there is p −1 explanatory variable, with p parameters (regression coefficients) $\beta_0, \beta_1, \beta_2 \ldots, \beta_{p-1}$,

This section deals with the development and results of electrical conductivity, total dissolved solids, and total hardness prediction models using MLR techniques using the best input combination based on the MATLAB (2016a) environment. The input parameters include pH and turbidity. The following regression models were derived, and the regression model equations are shown explicitly in Equations 3–5. The models are for the prediction of electrical conductivity (EC), total dissolved solids (TDS), and total hardness (TH).

$$Q_{(EC)} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + 11.2 \quad (3)$$

$$Q_{(TDS)} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + 13.8 \quad (4)$$

$$Q_{(TH)} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + 5.54 \quad (5)$$

Where x_1 and x_2 are input (independent) variables of pH and turbidity respectively. The other parameters are various estimated constants (β_0, β_1, β_2) generated by the ANN, as presented in Table 4 below. Thus, equations 6-8 above can be re-written as follows:

$$Q_{(EC)} = 84.266 + 0.38268 x_1 - 0.19656 x_2 \quad (6)$$

$$Q_{(TDS)} = 126.2 - 8.673 x_1 - 0.13822 x_2 \quad (7)$$

$$Q_{(TH)} = 41.98 - 1.1229 x_1 + 0.047042 x_2 \quad (8)$$

Table 4: Estimated constants for MLR model electrical conductivity, total dissolved solids and total hardness

	Estimate	SE	t-Stat	p-value	RMSE
β_0	73.066	25.54	2.8608	0.0055874	11.2
β_1	0.38268	3.6319	0.10537	0.91639	
β_2	-0.19656	0.15968	-1.231	0.22252	
β_0	112.41	31.368	3.5835	0.00062736	13.8
β_1	-8.673	4.4606	-1.9444	0.055926	
β_2	-0.13822	0.19611	-0.70478	0.48332	
β_0	36.44	12.605	2.8908	0.005133	5.54
β_1	-1.1229	1.7925	-0.62643	0.5331	
β_2	0.047042	0.078809	0.59691	0.55252	

3.4 Model Validation

To compare the goodness of fit of the ANN model, some representative hypothesis tests were conducted for the model construction process. These tests are the t-test to test the means and the F-test for variance. A paired t-test is used to compare the means of two sample populations, in which observations in one sample can be paired with observations in the other sample. The F-test is used to test if the variances of two populations are equal (Snedecor and Cochran, 1989), to further ascertain the performance efficiency of the predictive regression models, the R^2 and RMSE values for each were generated using Microsoft Excel.

3.4.1 F-Statistical Test

The F-test with two samples for variance was used to compare both the measured and predicted results. If F > Fcrit, the null hypothesis is rejected. Otherwise, it is accepted. Table 5 and 6 presents the results of the three analyses: For electrical conductivity, total dissolved solids, and total hardness as shown in Table 5, F is equal to 1.017605 and F critical is equal to 2.81793, suggesting that F crit is greater than F. Thus, the null hypothesis was accepted.

F value for total dissolved solids, as shown in Table 5, is equal to 1.670802 and Fcritical is equal to 2.81793, which clearly indicates that Fcritical is greater than F. Similarly, for total hardness, as shown in Table 5, F is equal to 2.307812 and Fcritical is equal to 2.81793, which shows that Fcritical is greater than F.

As a consequence, the null hypothesis is not rejected in any case since Fcritical is greater than F in all three analyses above. This means, however, that there is no significant difference between the measured and predicted model outcomes. Thus, the ANN model is valid and can be effectively used for the prediction of water quality parameters.

Table 5: F-Test two-sample for variances for electrical conductivity (Q_{EC}), total dissolved solids

	Measured	Predicted
Mean	77.53833	78.30528
Variance	32.64358	32.07883
Observations	12	12
Df	11	11
F	1.017605	
P(F<=f) one-tail	0.488715	
F Critical one-tail	2.81793	
Mean	46.2275	48.80959
Variance	7.379566	4.416782
Observations	12	12
Df	11	11
F	1.670802	
P(F<=f) one-tail	0.203914	
F Critical one-tail	2.81793	
Mean	27.02	29.84466
Variance	9.195982	3.984718
Observations	12	12
Df	11	11
F	2.307812	
P(F<=f) one-tail	0.090585	
F Critical one-tail	2.81793	

3.4.2 t-Test

When a t-test is performed, the null hypothesis is rejected if t stat > t critical two tails. Otherwise, it is accepted. Here, the paired-two sample for means was used. The results of the t-tests for electrical conductivity, total dissolved solids, and total hardness are shown in Table 6.

The result for electrical conductivity, as shown in Table 7, indicates that the value of the critical two tails is 2.200985 and the t start is -0.94542. The one for TDS, as shown in Table 5, indicates that t critical two tails has a value of 2.200985 and t stat has a value of -2.42495. Also, the total hardness as shown in Table 6 indicates that t critical two tails has a value of 2.200985 and t stat has a value of -5.15467. The outcomes of all three analyses, as clearly seen in the respective tables, show that t critical is greater than t stat. Thus, the null hypothesis is accepted. Hence, the model can be effectively used to predict the water quality parameters stated above.

Table 6: t-Test: Paired two sample for means for electrical conductivity (Q_{EC}) total dissolved solids

	Measured	Predicted
Mean	77.53833	78.30528
Variance	32.64358	32.07883
Observations	12	12
Pearson Correlation	0.87802	
Hypothesized Mean Difference	0	
Df	11	
t Stat	-0.94542	
P(T<=t) one-tail	0.182374	
t Critical one-tail	1.795885	
P(T<=t) two-tail	0.364749	
t Critical two-tail	2.200985	
Mean	46.2275	48.80959
Variance	7.379566	4.416782
Observations	12	12
Pearson Correlation	-0.15846	
Hypothesized Mean Difference	0	
Df	11	
t Stat	-2.42495	
P(T<=t) one-tail	0.016855	
t Critical one-tail	1/795885	
P(T<=t) two-tail	0.03371	
t Critical two-tail	2.200985	
Mean	27.02	20.84466
Variance	3.984713	3.984718
Observations	12	12
Pearson Correlation	0.791071	
Hypothesized Mean Difference	0	
Df	11	
t Stat	5.15467	
P(T<=t) one-tail	0.000158	
t Critical one-tail	1.795885	
P(T<=t) two-tail	0.000316	
t Critical two-tail	2.200985	

3.4.3 RMSE and R^2 Values

The RMSE and R^2 values for each of the linear regression models as shown in Table 7 indicate that electrical conductivity, total dissolved solids, and total hardness have RMSE values of 5.36, 6.94, and 0.08, respectively. This clearly implies that the total hardness model has the lowest value and thus has the best predictive performance. Consequently, the R^2 values for electrical conductivity, total dissolved solids, and total hardness are 0.9925, 0.9462, and 0.8876, respectively, for the model output parameters. In each case, the R^2 was very close to 1, thereby indicating excellent predictive performance.

Table 7: RMSE and R^2 values of the MLR models

Model	RMSE	R^2
Electrical Conductivity	5.36	0.9925
Total Dissolved Solids	6.94	0.9462
Total Hardness	0.08	0.8876

3.5 Water Quality Parameters Prediction (Forecast)

The predictions for electrical conductivity, total dissolved solids, and total hardness for a period of 2023–2029 are presented in figures 6–8. The model, which is a general representation of a system, is implored in the study for explanation of the water quality parameters being modelled, while the forecast model was used to predict future outcomes based on the current data. The R^2 values for the forecast are 0.999, 0.9487, and 0.9998 for electrical conductivity, total dissolved solids, and total hardness, respectively. All the R^2 values are very close to 1, indicating 99.9%, 94.8%, and 99.9% of the variance in the dependent variable is predictable from the independent variable. Thus, the forecasts and predictions for all three models are very effective and reliable.

4.0 CONCLUSION

Results obtained indicate that pH, which has a range of 6.45–8.5, and turbidity, which has a range of 1.6–36.21 NTU, were above the NSDWQ standard. TDS, EC, and total hardness, which have a range of 6.5–110.7 mg/l, 16.8–89.9 µS/cm, and 21.5–63.0 mg/l, respectively, were within the NSDWQ permissible standard. The performance of ANN was tested using RMSE, R, and R^2. The R^2 values obtained from the water quality parameters (electrical conductivity, total dissolved solids, and total hardness) and prediction were very close to 1, indicating a good model and prediction. The R values for the ANN testing, training, and validation data sets are 0.94385, 0.9119, and 0.93237, respectively. Consequently, the corresponding R^2 values are 0.89085, 0.83156, and 0.86931. Furthermore, the ANN RMSE values for EC, TDS, and TH are 11.2, 13.8, and 5.54, respectively. Multilinear Regression Model equations were obtained for predicting electrical conductivity, total dissolved solids, and total hardness. The R^2 values obtained for EC, TDS, and TH are 0.9925, 0.9462, and 0.8876, respectively. The corresponding RMSE values are 5.36, 6.94, and 0.08. Each of the mathematical equations is very effective for prediction. The model validation carried out through the F-test and t-test for each of electrical conductivity, total dissolved solids, and total hardness, respectively, shows that F critical is greater than F, as well as t critical is greater than t-stat. This further shows that the ANN model is fit for prediction. ■

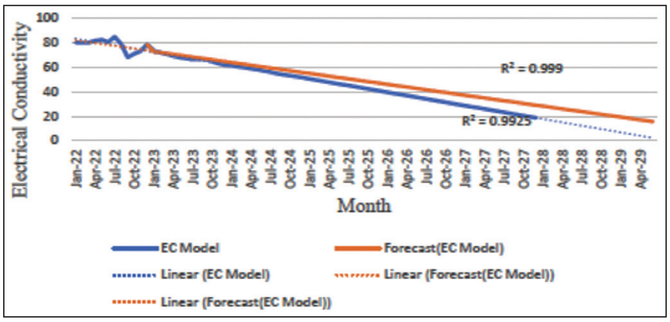

Figure 6: Prediction for EC model (2023-2029)

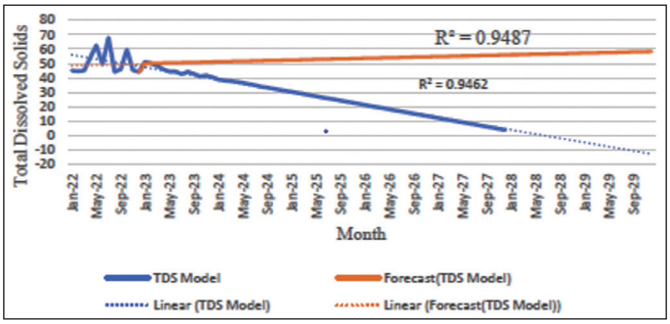

Figure 7: Prediction for TDS model (2023-2029)

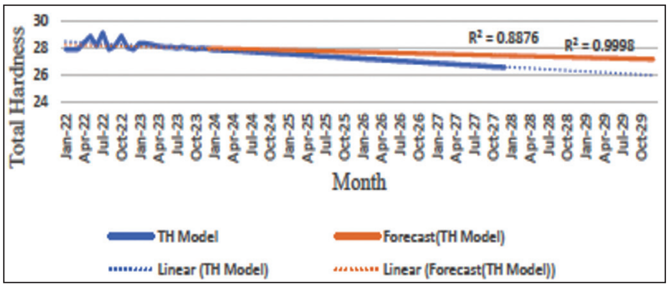

Figure 8: Prediction for total hardness model (2029-2029)

REFERENCES

[1] Ehya, F., & Saeedi, F. (2019). Assessment of groundwater quality in the Garmez area (Southeastern Khuzestan Province, SW Iran) for drinking and irrigation uses. Carbonates and Evaporites, 34(4), 1443-1454.

[2] Boretti, A., & Rosa, L. (2019). Reassessing the projections of the World Water Development Report. NPJ Clean Water, 2(15), 1-6.

[3] Egbueri, J. C., Mgbenu, C. N., & Chukwu, C. N. (2019). Investigating the hydrogeochemical processes and quality of water resources in Ojoto and environs using integrated classical methods. Modeling Earth Systems and Environment, 5(4), 1443-1461.

[4] Mrunmayee, M. S. (2014). Analysis and modelling of surface water quality in river basins (MTech Research Thesis, NIT Rourkela).

[5] Singh, K. P., Malik, A., & Sinha, S. (2005). Water quality assessment and apportionment of pollution sources of Gomti River (India) using multivariate statistical techniques: A case study. Analytica Chimica Acta, 538(1-2), 355-374.

[6] Marjan, S., Ehsan, T., & Zeinab, N. (2021). Application of an artificial neural networks model estimating of water quality parameters in Karun River Iran. Journal of Environmental Treatment Techniques, 9(4), 1710-1719. https://doi.org/10.4277/JETT/9(40)727

[7] Uba, J. I., Orakwe, L. C., Ogbu, K. N., Awu, J. I., Ahaneku, I. E., & Chukwuma, E. C. (2021). Forecasting water quality parameters using artificial neural network for irrigation purposes. Scientific Reports, 11, Article 24438. https://doi.org/10.1038/s41598-021-03435-6

[8] Elias, E., Sam, K., & Tahima, A. (2023). Aquaculture 4.0: Hybrid neural network multivariate water quality parameter forecasting model. Scientific Reports, 13, Article 16129. https://doi.org/10.1038/s41598-023-45678-y

[9] Jitha, P. N. (2023). Design and development of efficient water quality prediction models using variants of recurrent neural networks. NR Chem. Bull., 12(Special Issue 5), 1210-1223. https://doi.org/10.31838/ecb/2023.12.5.5.014

[10] Furaun, R., Abid, I., Sayyidi, T. K., Isabel, de la T. D., Juan, L. V. M., Carmen, L. R., & Imran, A. (2022). An artificial neural network for water quality and water consumption prediction. Water, 14(2), 3359. https://doi.org/10.33901/w/42133591

[11] Nguyen, T. H., Phan, D. T., & Dinh, T. N. (2018). Water quality evaluation of the Tien River by means of Water Quality Index (WQI) and statistical techniques. Vietnam Journal of Science and Technology, 56(2), 141-148.

[12] ASCE Task Committee. (2000). Artificial neural network in hydrology. I: Preliminary concepts. Journal of Hydrologic Engineering, 5(2), 115-123.

[13] Archana, S., & Prashant, P. (2015). River water quality modelling using artificial neural network technique. Aquatic Procedia, 4, 1070-1077.

[14] Ali, N., Ahmed, E., Othman, A. K., & Othman, J. (2009). Prediction of Johor River water quality parameters using artificial neural networks. European Journal of Scientific Research, 28(3), 422-435.

[15] Roza, D., & Mohsen, S. (2021). Artificial neural network system. International Journal of Imaging and Robotics, 21(2), 14-25.

[16] Khandelwal, M., & Singh, T. (2005). Prediction of mine water quality by physical parameters. Journal of Scientific and Industrial Research, 64(7), 564-570.

[17] Awu, J. I., Mbajiorgu, C. C., Ogunlela, A. O., Kasali, M. Y., Ademiluyi, Y. S., & James, D. D. (2017). Optimization of neural network architecture and transfer functions for rainfall river flow modeling. Journal of Environmental Hydrology, 25(8), 1-15.

[18] Abrahart, R. J., Kneale, P. E., & See, L. M. (2005). Neural networks for hydrological modelling. In R. J. Abrahart, P. E. Kneale, and L. M. See (Eds.), Hydroinformatics in hydrology, hydrogeology and water resources (pp. 223-250). Taylor and Francis Group.

[19] Muhammad, S. G., Sani, I. A., Aliyu, M. A., Abubakar, I. T., Mubarak, A. S., Parvaneh, E., & Norhaliza, A. (2020). Estimation of water quality index using artificial intelligence approaches and multi-linear regression. IAES International Journal of Artificial Intelligence (IJ-AI), 9(1), 126-134.

[20] Snedecor, G. W., and Cochran, W. G. (1989). Statistical methods (8th ed.). Iowa State University Press.

PROFILES

HARUNA GARBA is a senior lecturer in the Department of Civil Engineering, Nigerian Defence Academy (NDA) Kaduna, an Engineer with The Council for Regulation of Engineering in Nigeria and a member of Nigerian Society of Engineers. Area of specialisation is Hydrology and River Flow Dynamics. Area of research interest include Surface and Ground Water Hydrology and Climate Change.
Email address: hgarba@nda.edu.ng

FRANCIS PETER is a distinct and dynamic individual with a great passion for research, academic excellence, and knowledge impartation, aiming to build a robust career in the engineering profession and beyond. He is currently the Vice Principal Administration of Standard Foundation Schools, Barnawa, as well as a chief consultant in health, safety, and environmental management with Zeteo Citadel, Kaduna State, Nigeria.
Email address: edokapeter2014@gmail.com