Al-Farabi Journal of Engineering Sciences Volume (3), Issue (1) October (2024)

ISSN: 2617-5517 (Print)
Al-Farabi Journal of Engineering Sciences
https://www.iasj.net/iasj/journal/392/issues
Published by Al-Farabi University College

العـــراقيــة
المجلات الاكاديمية العلمية
IRAQI
Academic Scientific Journals

Predicting Monthly Water Quality Index Using Neural Network Analysis for Tigris River at Al-Rashediya Water Station in Baghdad City

Muna Yousif Abdul-Ahad[1]

[1]**Petroleum and Gas Refining Engineering, Al-Farabi University College, Baghdad, Iraq.**

[1] **E-mail: myabdulahad@gmail.com**

Abstract

In this study, the quality of TGRIS River is studied at the intake of Al-Rashediya Water Station using neural network analysis. 14 measured parameters of water quality, daily periods for 11 years (2013-2023), monthly mean averaged were studied which are: K^+, Na^+, T.S.S, T.D.S, $SO4^{2-}$, Cl^-, Mg^{2+}, Ca2+, T.H, Alk., E.C, pH, Turb, and Temp., from which WQI was calculated. In this study, a sophisticated artificial neural network (ANN) model. to predict water quality (WQIى). Neural network fitting app. is applied using multi-layer feed- forward neural network with back propagation algorithm. The data were randomly divided into three phases, training (70%), validation (15%), and testing (15%). Efficiency statistics were used to evaluate the model prediction abilities. The results showed that the model performed well with high predicting ability for the water quality index (WQI), and the model performed best with accuracy (R .9921, and MSE 221.7468) at the testing phase, which will help to enhance the WQ using cheap and valuable method. The Predicted WQI mathematical model is estimated by the equation:
Output WQI=0.99target WQI+1.1, which can be used for the becoming years.

Key words
Water quality index, Artificial neural network (ANN), Deep learning.

الخلاصة

الهدف من هذه الدراسة هو التنبؤ لحالة التلوث الحاصلة لمياه نهر دجلة عند مدخل محطة مياه الراشدية باستخدام تحليل الشبكات العصبية الصناعية. حيث تم دراسة 14 معلمة مقاسة لجودة المياه ، وعلى فترات يومية لمدة 11 سنة (2023-2013) ، وحيث تم اخذ المتوسط الشهري لهذه المعالم وهي: البوتاسيوم K^+, الصوديوم Na^+ , المواد الصلبة العالقة T.S.S , المواد الصلبة الذائبة T.D.S , الكبريتات SO_4^{2-}, الكلوريدات Cl^-, المغنيسيوم Mg^{2+}, الكالسيوم Ca^{2+}, العسرة الكلية T.H, القاعدية Alk, التوصيلية الكهربائية E.C , الدالة الحامضية PH, العكورة Turb, درجة الحرارة Temp. والتي تم من خلالها تم حساب WQI.
هذه الدراسة تستخدم شبكة عصبية صناعية ذات تغذية أمامية متعددة الطبقات مع خوارزمية الانتشار الخلفي. feed- forward neural network with back propagation algorithm كنموذج لتخمين دليل جودة المياه. تم استخدام هذا النموذج لتدريب وتجربة وفحص دليل جودة المياه. كما تم استخدام المعايير الاحصائية منها معدل الخطأ التربيعي MSE= 221.4768 ومعامل الارتباط R=0.9921 لتقييم اداء النموذج على ضوء نتائج هذه المعايير، والتي تعتبر قيم ذات دقة عالية والنموذج قادر تماما على التنبؤ بجودة المياه, وبناء عليه يمكن تقليص التكاليف والمخاطر الى حد كبير. كما اظهرت هذه الدراسة ان النموذج الرياضي لمعادلة التنبؤ للسنوات اللاحقة هي:
OutputWQI = 0.99targetWQI + 1.

Introduction

Water quality affects life for its direct impact on human health. salinity, urban and domestic wastewater entrance into surface streams, agricultural drainage, geological structures, ground water usage, and a wide range of chemical compounds throughputs [1],[2]. Different methods and approaches to investigate and predict the quality of water are used. Also, the majority of water software such as SWAT, QUAL2K MIKE-11, QGIS, SAGA GIS, HEC-RAS, iRIC, PRMS, SPSS, MATLAB, and Python, are used as tools to assess the quality of streams.

Applying artificial neural network analysis in water quality modeling and prediction are very useful methods for understanding and analyzing the process of different phenomena. It is also helpful in generating past observations for predicting the future values based on the past memory. The use of soft computing techniques such as ANN and ANFIS is applied in modelling the relationship between the dependent and independent variables [3],[4]. ANN is a modelling approach based on the brain's neural structure; they are generally used to model and optimize a complex process. [5],[6],[7]. Thus, it analyses complex variables to form deterministic equations [8].

Deep learning is an important component of data science, which also includes statistics and predictive modelling. It assists researchers who are tasked with collecting, analyzing, and interpreting large datasets [9]. Deep learning neural networks, which include artificial neural networks (ANNs), recurrent neural networks, and convolutional neural networks, have been used in the predictive modelling of water quality parameters [10],[11],[12],[13]. Moreover, other data-intelligent models like linear regression, multiple linear regression (MLR), support vector machines, amongst others, have also been utilized globally to predict various water quality parameters. The application of deep learning and data-intelligent models has significantly reduced the cost of monitoring and assessment of water quality. Studies conducted include the prediction of pH in water (Egbueri & Agbasi, 2022b; Huang et al., 2019; Son et al., 2021; Stackelberg et al., 2020), prediction of TDS in water [14],[15],[16],[17], prediction of TH in water [18],[19], prediction of anions in water [20], prediction of cations in water [21],[22], prediction of metals in water [23],[24],[25], and prediction of water quality [26],[27],[28],ANN is a powerful tool designed to mimic the neural functions of the human nervous system [29]. As a result, ANN has the ability to learn a dataset, and its learning ability aids in simulating complex nonlinear [30],[31],[32],[33], making it possible to produce meaning out of a dataset in a short period of time. The general structure of the ANN consists of an input layer, a hidden layer, and an output layer, each with numerous neurons [34]. ANN has been shown to be effective in the prediction of water quality parameters in many regions of the world [14],[35],[36]. This tool has also been undoubtedly valuable in studies related to other disciplines.

The objective of this study is predicting by using artificial neural network analysis (ANN modelling) for the becoming years (after2023) to provide information on the physicochemical characteristics of Tigris River water quality within Al-Rashediya station in Baghdad city, and the impacts of unregulated waste discharge on the quality of the river as well as to discuss its suitability for human consumption based on predicted and forecasted water quality index values (WQI),where a large data matrix, obtained during 11-years (2013-2023) monitoring program, is subjected to (ANN modeling analysis technique).

Water Quality Index Application and Formulation

In the formulation of WQI, the importance of various parameters depends on the intended use of water, and its suitability for human consumption. The standard permissible values of various parameters for the drinking water used in this study are those recommended by the Iraqi drinking water standards (Drinking-Water Standard IQS: 417,2001), and by the (World Health Organization WHO, 2004).

For calculating the Water Quality Index, a set of fourteen water quality parameters have been collected from Al-Rashediya station. The full calculations for the actual overall Water Quality Index (WQI) were calculated using the Weighted Arithmetic Index method which were given elsewhere [37].

The data used to calibrate and validate the ANN analysis were collected from Baghdad Water Governorate. In Al-Rashediya station, data were collected from raw water (river water near intakes of the water treatment plant). MATLAB (2022b), SPSS© v.26, and Microsoft Office Excel© 2021, software packages were used to implement all the mathematical and ANN analyses.

Applied Artificial Neural Network Model

(ANN analysis)

ANN is a mathematical programming model that mimics the functioning process of the human brain. An ANN method can perform brain processes, decide, arrive at a solution in the absence of sufficient data using current knowledge, absorb continuous data input, learn, and remember. The capability of a neural network to model complicated nonlinear relation sans making prior assumptions about the nature of the relation is its greatest advantage [38]. An ANN is comprised of multiple nodes that represent neurons. The independent variables are represented by the input nodes, while the dependent variables are represented by the output nodes [39]. The main purpose of the learning procedure is to identify the best set of weights that can give the best output for the given inputs. The network output is compared to the target answer to calculate the error [40]. Different structures can be found in neural networks. Feed forward and recurrent networks can be distinguished in principle. Only forward-directed information flows from the input nodes through hidden nodes to the output nodes in feed forward networks. There are links in recurrent networks where information can travel forwards and backwards through network node connections. Feedback networks are another name for the recurrent networks [Mijwel and Alsaadi, 2019]. Back Propagation Algorithm (BP) Back propagation (BP) is the most common and widely applied learning algorithm over all neural network models among the various learning existing algorithms. This algorithm is employed in supervised learning [41]. The primary training concept of BP is founded on gradient descent algorithm, which modifies weights to reduce Mean Square Error (MSE) [43]. The BP algorithm is divided in two phases: forward and backward phase. In the forward phase, the network input data is propagated to the following level and so forth. The network error is calculated after that. In the backward phase, the network error is propagated backwards, and the weights are adjusted accordingly [44]. As illustrated in Figure 1, the network structure is consisting of ten layers, each of which has n neurons. The number of input variables determines the number of neurons in the first layer (input layer). This layer takes the input from external world and transfers them without any alteration to the hidden layer. Since they are only indirectly related to the outside environment, intermediate layers are usually known as hidden. The individual values are summed together and transmitted to the output layer by the output layer activation function. If the output is acceptable up to a particular level of error, it is permitted; otherwise, it is returned to the input layer for more updating of the weights and biases. It is worth noting that there is no link among nodes within the same layer. This cycle will continue until all of the limitations have been met [45].

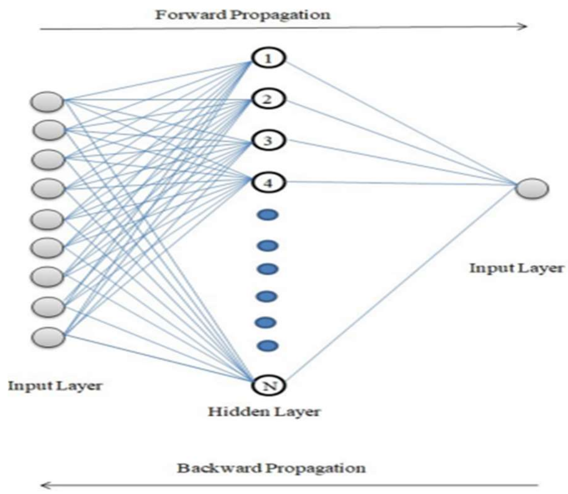

Figure 1 Multi- layer feed- forward neural network with BP algorithm.

Back Propagation Algorithm (BP)

Back propagation (BP) is the most common and widely applied learning algorithm over all neural network models among the various learning existing algorithms. This algorithm is employed in supervised learning [42]. The primary training concept of BP is founded on gradient descent algorithm, which modifies weights to reduce Mean Square Error (MSE) [43]. The BP algorithm is divided in two phases: forward and backward phase. In the forward phase, the network input data is propagated to the following level and so forth. The network error is calculated after that. In the backward phase, the network error is propagated backwards, and the weights are adjusted accordingly [44]. As illustrated in Figure 2, the network structure is consisting of three layers, each of which has n neurons. The number of input variables determines the number of neurons in the first layer (input layer). This layer takes the input from external world and transfers them without any alteration to the hidden layer. Since they are only indirectly related to the outside environment, intermediate layers are usually known as hidden. The individual values are summed together and transmitted to the output layer by the output layer activation function. If the output is acceptable up to a particular level of error, it is permitted; otherwise, it is returned to the input layer for more updating of the weights and biases. It is worth noting that there is no link among nodes within the same layer. This cycle will continue until all of the limitations have been met [45].

Performance criteria

The models were evaluated using Mean squared error (MSE) and Correlation Coefficient (R), as follows [46]:

$$MSE = \frac{1}{N}\sum_{j=1}^{N}(T_j - O_j)^2$$

$$R = \frac{\sum_{j=1}^{N}(T_j - \bar{T})(O_j - \bar{O})}{\left(\sqrt{\sum_{j=1}^{N}(T_j - \bar{T})^2 \sum_{j=1}^{N}(O_j - \bar{O})^2}\right)}$$

where: N is number of data, T is the target value, O is the output value of the network, $\bar{T}$ is the mean value of target data, and $\bar{O}$ is the mean value of network output.

Methodology

The data obtained from the Al-Rashediya station observed daily for years (2013-2023), which were transferred to an average monthly data, and their descriptive statistics will be presented.

The Neural Net Fitting app will be used to create, visualize, and train a two-layer feed-forward network to solve data fitting problems.

The procedure for the model is based on the following series of steps including to:

- Import data from file, to the MATLAB® workspace.
- Split data into training, validation, and test sets.
- Define and train a neural network.
- Evaluate network performance using mean squared error and regression analysis.
- Analyze results using visualization plots, such as regression fit or histogram of errors.
- Generate MATLAB scripts to reproduce results and customize the training process.
- Generate functions suitable for deployment with MATLAB Compiler™ and MATLAB Coder™ tools, and export to Simulink® for use with Simulink Coder.

To define the fitting (regression) problem for

the toolbox, a set of 14 input vectors (predictors) as columns in a matrix were set, which are the input variables K^+, Na^+, T.S.S, T.D.S, $SO4^{2-}$, Cl^-, Mg^{2+}, Ca^{2+}, T.H, Alk., E.C, pH, Turb, and Temp. (predictors). Then, the response, WQI (the correct output for each of the input vectors) is set into a second matrix vector, which are saved in the EXCEL sheets. These data sets were imported to MATLAB workspace.

To train the shallow neural network to fit the data using the **Neural Net Fitting** app, using nftool is opened, and Information about the imported data appears in the **Model Summary**.

the data were Split into training, validation, and test sets, Keeping the default settings. The data is split into:70% for training,15% to validate that the network is generalizing and to stop training before overfitting,15% to independently test the network generalization.

The network created is a two-layer feedforward network with a Sigmoid transfer function in the hidden layer and a linear transfer function in the output layer. The Layer size value defines the number of hidden neurons. Keeping the default layer size, 10. The network architecture is shown below Figure 2. The network plot updates to reflect the input data. In this research, the data has 14 inputs (features) and one output.

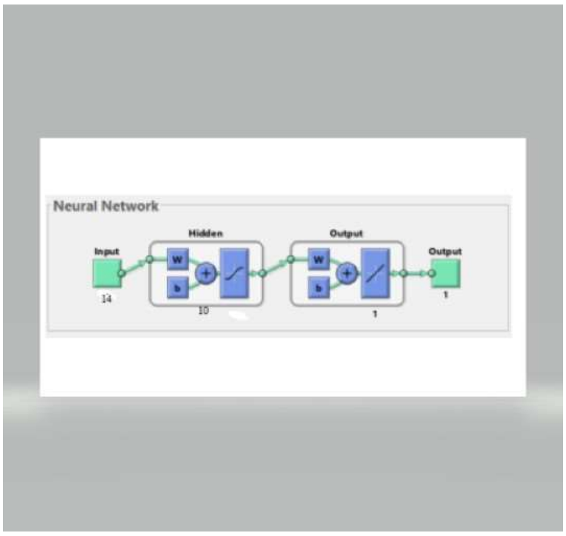

Figure 2 network architecture

Training with Levenberg-Marquardt (trainlm) is recommended for most problems. Training continues until one of the stopping criteria is met. In this research, training continues until the validation error increases consecutively for seven iterations ("Met validation criterion").

To analyze the results, The **Model Summary** contains information about the training algorithm and the training results for each data set.

For further analyses the results can be done by generating plots. The regression plot displays the network predictions (output WQI) with respect to responses (target variables (K^+, Na^+, T.S.S, T.D.S, $SO4^{2-}$, Cl^-, Mg^{2+}, Ca^{2+}, T.H, Alk., E.C, pH, Turb, and Temp.) for the training, validation, and test sets.

For a perfect fit, the data should fall along a 45-degree line, where the network outputs are equal to the responses. For more accurate results, the network can be retrained by training again. Each training will have different initial weights and biases of the network, and can produce an improved network after retraining.

To obtain additional verification of network performance, the error histogram can be viewed.

To increase the network performance the following steps can be performed: Training the network again, increasing the number of hidden neurons, and using a larger training data set. If performance on the training set is good but the test set performance is poor, this could indicate the model is overfitting. Reducing the number of neurons can reduce the overfitting.

For additional test data to evaluate the network. The **Model Summary** displays the additional test results and generate plots to analyze the additional test data results.

Trained network can be exported to the workspace or Simulink®. Also deploy the network with MATLAB Compiler™ tools and other MATLAB code generation tools.

Analysis and results

Descriptive Statistics & Line Charts

Table 1, presents the descriptive statistics for WQI. the data for each index consist of 132 months over a span of years (2013-2023). Figure3, illustrates the line chart where WQI is plotted against time. The line chart and the descriptive statistics show that the mean and variance are changing over time, i.e., they are non-stationary.

Modeling the WQI

The WQI variable was modeled by means of ANNs using 14 variables (K^+, Na^+, T.S.S, T.D.S, $SO4^{2-}$, Cl^-, Mg^{2+}, Ca^{2+}, T.H, Alk., E.C, pH, Turb, and Temp.) as input variables. In order to find the optimal number of nodes in the hidden layer, many ANN models were created and evaluated.

The effectiveness of the ANN models was assessed utilizing the coefficient of correlation

Table1 The Descriptive Statistics for Input Variables and Calculated WQI.

	N	Range	Minimum	Maximum	Mean		Std. Deviation	Variance	Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Statistic	Std. Error
K.1	132	1.75	1.05	2.80	1.9767	.03049	.35027	.123	.266	.419
Na.1	132	35.00	27.00	62.00	43.8648	.54221	6.22955	38.807	.343	.419
T.S.S.1	132	269.50	8.50	278.00	52.6248	3.04998	35.04165	1227.917	15.03	.419
T.D.S.1	132	339.00	311.00	650.00	436.3580	5.20017	59.74537	3569.510	.410	.419
SO4.1	132	153.00	76.00	229.00	136.2516	2.66818	30.65501	939.730	-.208	.419
Cl.1	132	49.25	31.00	80.25	54.8010	.85000	9.76575	95.370	.197	.419
Mg.1	132	26.00	20.00	46.00	26.7534	.37579	4.31749	18.641	2.670	.419
Ca.1	132	38.50	50.50	89.00	67.6145	.65302	7.50265	56.290	.052	.419
T.H.1	132	151.00	220.00	371.00	278.1411	2.50583	28.78983	828.854	-.029	.419
Alk.1	132	47.33	130.00	177.33	151.6411	.74853	8.59991	73.958	.497	.419
E.C.1	132	667.00	274.00	941.00	695.7411	8.98417	103.22031	10654.432	1.319	.419
PH.1	132	1.50	7.00	8.50	7.9187	.02783	.31978	.102	.154	.419
Turb.1	132	202.69	9.31	212.00	39.1412	2.58030	29.64535	878.847	11.95	.419
Temp.1	132	26.30	13.70	40.00	24.9683	.50147	5.76142	33.194	-.822	.419
WQI	132	810.84	59.00	869.84	180.0114	10.3486	118.89651	14136.381	11.71	.419
Valid N (listwise)	132									

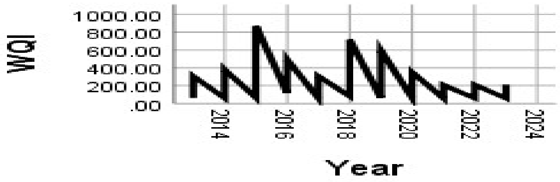

Fig.3 Line chart for WQI with time.

(R) and the mean squared error (MSE), as shown by the model summery in Figure 4.

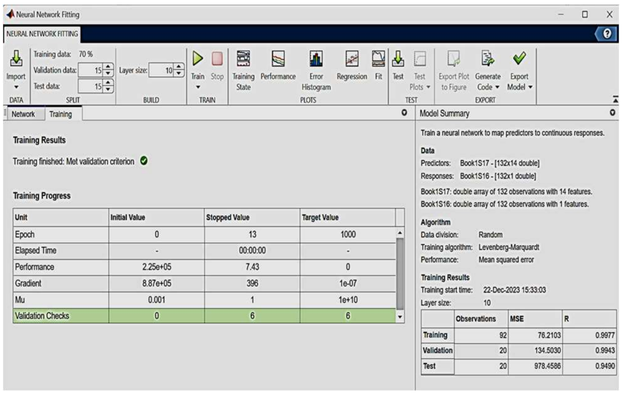

Figure4. Training ANN Results.

From the executed ANN modeling, the best training results were obtained for the network with ten neurons, after 13 iterations. The ANN performance was validated with MSE. The best validation result was exhibited by the network's seventh iteration Figure 5. In general, the error

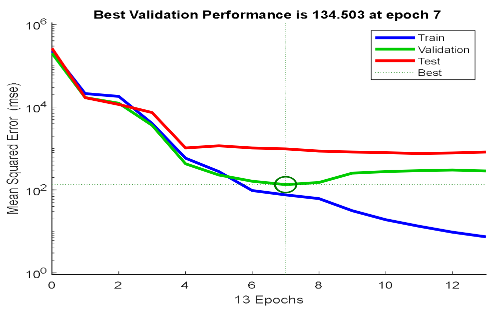

Figure 5. The best validation performance is 134.603 at epoch 7.

was found to decrease over successive training periods, but it is likely that it would start to increase in the validation dataset in the case when the network began to overfit the training data. The training was designed to stop after eleventh consecutive increases in the validation error (or no decrease), and the best results were obtained from the iteration with the lowest validation error.

In figure 6, the pace of the error decrease (gradient) for a particular iteration of the validation set depending on the number of consecutive increases of the MSE for this set, and the momentum (Mu). With seven consecutive increases of the MSE validation error, the learning process of the network is being stopped.

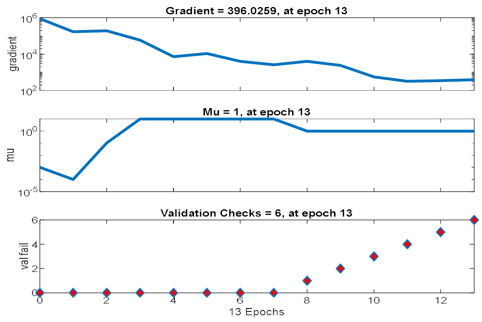

Figure 6. Gradient, Mu, and Validation Checks at epoch 13.

In the training stage, the predicted values histogram error is shown in Figure 7. Metrics such as the error histogram may be used to identify discrepancies between the expected and observation values. These error numbers might be negative, since they indicate how the prediction values differ from the training target values. An error of the model is 0.0005.

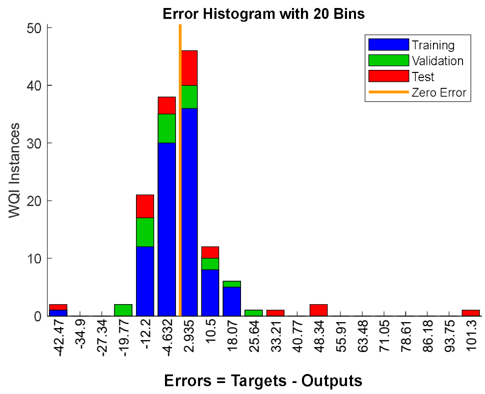

Figure7. Error Histogram With 20 Bins.

Figure 8, shows the regression statistics for the particular subsets, where the regression value (R) for training data is 0.9977, for validation data—0.9943, and the test data—0.9490, thus, in each case R > 0.95, which is typical of a good fit of the network. The overall regression coefficient was 0.99212 from the regression plot, which confirms the high degree of overlap between the measurement points and the fit line with the ideal Y = T prediction line. It demonstrates that there is a perfect match between the observed values and the prediction values of water quality.

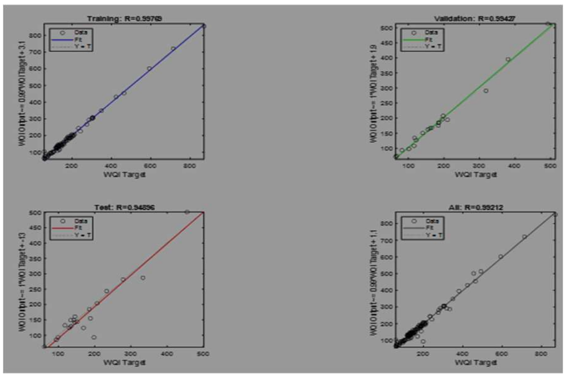

Figure8. Regression Statistics of the ANN.

Also, to evaluate the network performance on an additional test set. The **Model Summary** displays the additional test results, and generate plots to analyze the additional test data results as shown in Figures 9,10, and 11. Results show MSE is 221.4761, R is 0.99212, and error 0.0005, which are acceptable.

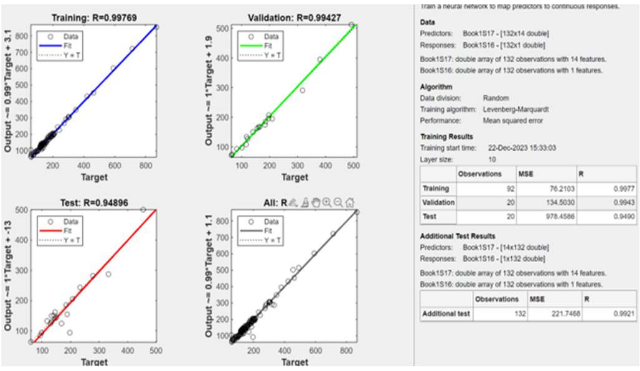

Figure9. Regression Plot and Model Summery for Additional Test.

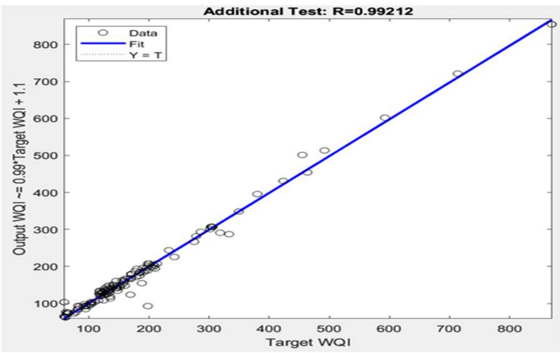

Figure 10. Test Regression Plot for Additional test.

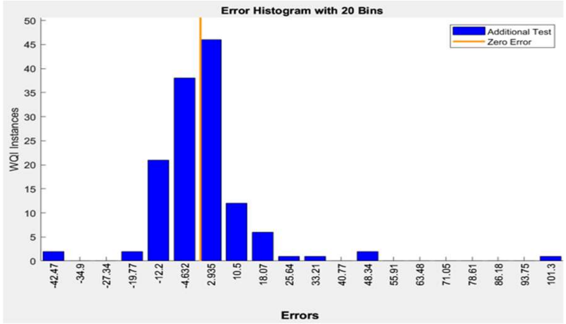

Figure 11. Error Histogram with 20 Bins for Additional Test.

Discussion

In the paper, an attempt was made to prove that the ANN model is an efficient tool for predicting the water quality, which can be employed for planning integrated water protection systems, while implementing good environmental management practices

The ANN method is much easier to apply an existing ANN model that uses raw data without the need for additional recalculation, which reduces the cost of water quality monitoring, and of network quality indicators R=0.9921, MSE=67.2103, 134.5030, and 978.4576 for the training, testing and validating set, respectively., and ten neurons in the hidden layer as the best performer, and with set division of , training (70%), validation (15%), testing (15%),and using the Neural Network Fitting app, and the Levenberg–Marquardt algorithm for training.

Some of the advantages using ANN models are, in cases of data deficiency, due to the unavailability, difficulty or high-cost expense of obtaining the actual figures, and allowing for the reconstruction of missing data, reducing costs and saving time for otherwise necessary analyses, which can be a strong economic factor in scientific researches, and should encourage authorities, and water quality management bodies to invest in the use of ANN modeling.

The advantage of using the ANN model proposed in this paper is the easy assessment of surface water pollution levels; in addition, it enables the avoidance of lengthy calculations involved in prevalent conventional WQI.

In future works, other machine learning methods for surface water quality assessment will be attempted to effectively forecast water quality, including: Multilayer preceptor (MLP), and Radial Basis Function (RBF). The results obtained using these methods will be compared and their impact on prediction quality will be investigated.

Acknowledgments

I would like to acknowledge and thank Baghdad Governorate (Baghdad water directory), for providing the required data for my research work, and for making it convenient in various aspects to carry out this research.

References

[1] G. a. D.Alexakis, "Water quality models: An overview," *European Water,* vol. 37, pp. 33-46, 2012.

[2] M. S. , S. P. a. D. A. George Tsakiris, "Assessing the water potential of karstic saline springs by applying a fuzzy approach: The case of Almyros (Heraklion, Crete)," *Desalination,* vol. 237, pp. 54-64, 2009.

[3] Nwosu-Obieogu K, Aguele F, Chiemenem LI," Soft computing prediction of oil extraction from huracrepitan seeds. Kem Ind 69(12):653–658, 2020.

[4] Ojediran OJ, Okonkwo CE, Adeyi AJ, Adeyi O, Olaniran FO, George NE, Olayanju AT "Drying characteristics of yam slices (Dioscorea rotundata) in a convective hot air dryer: application of ANFIS in the prediction of drying kinetics," Heliyon 6:e03555, 2020.

[5] Onoji S, Iyuke S, Igbafe IA," Daramola OM Hevea brasiliensis (rubber seed) oil: modelling and optimization of extraction process parameters using response surface methodology and artifcial neural network techniques". Biofuels 6:1–15, 2017.

[6] Sirohi R, Pandey JP, Singh A, Sindhu R, Lohan VC, Goel R, Kumar A "Acid hydrolysis of damaged wheat grains: modeling the formation of reducing sugars neural network approach", Ind Crops Prod 149:112351,2020.

[7] Rivera EC, Rabelo SC, Garcia DR, Filho RM, da Costa AC "Enzymatic hydrolysis of sugarcane bagasse for bioethanol production: determining optimal enzyme loading using neural networks". J Chem Technol Biotechnol 85:983–992, 2010.

[8] Kose E," Modelling of color perception of different age groups using artificial neural networks". Expert Syst Appl 34:2129–2139, (2008).

[9] Burns, E. & Brush, K.," What is deep learning and how does it work?", Retrieved Jan 2022 from https://www.techtarget.com/searchenterpriseai/defnition/deep-learning-deep-neural-network, (2021).

[10] Alizamir, M., & Sobhanardakani, S," A Comparison of performance of artificial neural networks for prediction of heavy metals concentration in groundwater resources of Toyserkan plain". Avicenna Journal of Environmental Health Engineering, 4(1), 11792. https://doi.org/10.5812/ajehe.11792 , 2017a.

[11] Alizamir, M., & Sobhanardakani, S, "Predicting arsenic and heavy metals contamination in groundwater resources of Ghahavand plain based on an artificial neural network optimized by imperialist competitive algorithm", Environmental Health Engineering and Management Journal, 4(4), 225–231, 2017b.

[12] Egbueri, J. C.," Prediction modeling of potentially toxic elements' hydro geo pollution using an integrated Q-mode HCs and ANNs machine learning approach in SE Nigeria", Environmental Science and Pollution Research, 28, 40938–40956. https://doi. org/10.1007/s11356-021-13678-z, 2021.

[13] Wagh, V. M., Panaskar, D. B., & Muley, A. "A., Prediction of groundwater suitability for irrigation using artificial neural network model: A case study of Nanded tehsil, Maharashtra, India", Modeling Earth Systems and Environment, 2, 1–10. https://doi. org/10.1007/s40808-016-0250-3, 2016.

[14] Egbueri, J. C., & Agbasi, J. C. "Data-driven soft computing modeling of groundwater quality parameters in southeast Nigeria: Comparing the performances of different algorithms", Environmental Science and Pollution Research, 29(38), 57147–57171. https://doi.org/10.1007/s11356-022-18520-8, 2022b.

[15] Huang, H., Feng, R., Zhu, J., & Li, P.," Prediction of pH value by multi-classification in the Weizhou Island area", Sensors, 19(18), 3875. https://doi.org/10.3390/s19183875, 2019.

[16] Son, M., Yoon, N., Jeong, K., Abass, A., Logan, B. E., & Cho, K. H." Deep learning for pH prediction in water desalination using membrane capacitive deionization", Desalination, https://doi.org/10.1016/j.desal.2021.115233, 2019.

[17] Stackelberg, P. E., Belitz, K., Brown, C. J., Erickson, M. L., Elliott, S. M., Kaufman, L. J., & Reddy, J. E.," Machine learning predictions of pH in the glacial aquifer system, northern USA. Groundwater", https://doi.org/10.1111/gwat.13063, 2020.

[18] Azad, A., Karami, H., & Farzin, S., "Prediction of water quality parameters using ANFIS optimized by

intelligence algorithms (case study: gorganrood river)", KSCE Journal of Civil Engineering, 22, 2206–2213. https://doi.org/10.1007/ s12205-017-1703-6, 2018.

[19] Roy, R., & Majumder, M.," A quick prediction of hardness from water quality parameters by artificial neural network", Int J Environ Sustain Dev, 17(2/3), 247–257, 2018.

[20] Zare, A. H., Bayat, V. M., & Daneshkare, A. P.," Forecasting nitrate concentration in groundwater using artificial neural network and linear regression models", International Agro Physics, 25(2), 187–192, 2011.

[21] Bondarev, N. V.," Artificial neural network and multiple linear regression for prediction and classification of sustainability of sodium and potassium coronates", Russian Journal of General Chemistry, 89, 1438–1446, https://doi.org/10.1134/S107036321 9070144, 2019.

[22] Subba Rao, N., Dinakar, A., & Sun, L.," Estimation of groundwater pollution levels and specific ionic sources in the groundwater, using a comprehensive approach of geochemical ratios, pollution index of groundwater, unmix model and land use/land cover—A case study", Journal of Contaminant Hydrology, 248, 103990, 2022b.

[23] Fard, Z. B., Ghadimi, F., & Fattahi, H.," Use of artificial intelligence techniques to predict distribution of heavy metals in groundwater of Lakan lead-zinc mine in Iran", Journal of Mining and Environment, 8(1), 35–48, 2017.

[24] Ozel, H. U., Gemici, B. T., & Gemici, E., "Application of artificial neural networks to predict the heavy metal contamination in 78 J. C. Agbasi, J. C. Egbueri 1 3 the Bartin River", Environmental Science and Pollution Research, 27, 42495–42512. https://doi.org/10.1007/s11356-020-10156-w, (2020).

[25] Rooki, R., Ardejani, F. D., Aryafar, A., & Asadi, A. B.," Prediction of heavy metals in acid mine drainage using artificial neural network from the Shur River of the Sarcheshmeh porphyry copper mine, Southeast Iran", Environment and Earth Science, 64, 1303–1316. https://doi.org/10.1007/s12665-011-0948-5, 2020.

[26] Chia, S. L., Chia, M. Y., Koo, C. H., & Huang, Y. F.," Integration of advanced optimization algorithms into least-square support vector machine (LSSVM) for water quality index prediction", Water Supply, 22(2), 1951–1963, 2022a.

[27] Egbueri, J. C.," Incorporation of information entropy theory, artificial neural network, and soft computing models in the development of integrated industrial water quality index, "Environmental Monitoring and Assessment, 194(10), 693. https://doi.org/10. 1007/s10661-022-10389-x, 2022.

[28] Egbueri, J. C.," Predicting and analyzing the quality of water resources for industrial purposes using integrated data-intelligent 76 J. C. Agbasi, J. C. Egbueri 1 3 algorithms", Groundwater for Sustainable Development, 18, 100794. https://doi.org/10.1016/j.gsd.2022.100794, 2022b.

[29] Wagh, V., Panaskar, D., & Muley, A., "Estimation of nitrate concentration in groundwater of Kadava river basin-Nashik district, Maharashtra, India by using artifcial neural network model", Modeling Earth Systems and Environment, 3, 36. https://doi.org/ 10.1007/s40808-017-0290-3, 2022b.

[30] Agatonovic-Kustrin, S., & Beresford, R.," Basic concepts of artifcial neural network (ANN) modeling and its application in pharmaceutical research", Journal of Pharmaceutical and Biomedical Analysis, 22(5), 717–727, 2000.

[31] Saljooghi, B. S., & Hezarkhani, A.," A new approach to improve permeability prediction of petroleum reservoirs using neural network adaptive wavelet (wave net)", Journal of Petroleum Science and Engineering, 133, 851–861, 2000.

[32] Song, H., Song, M., & Liu, X. ," Online autonomous calibration of digital twins using machine learning with application to nuclear power plants", Applied Energy, 326, 119995, 2000.

[33] Uncuoglu, E., Citakoglu, H., Latifoglu, L., Bayram, S., Laman, M., Ilkentapar, M., & Oner, A.A.," Comparison of neural network, Gaussian regression, support vector machine, long short term memory, multi-gene genetic programming, and M5 Trees methods for solving civil engineering problems", Applied Soft Computing, 129, 109623. https://doi.org/10.1016/j.asoc.2022. 109623, 2022.

[34] Pandey, D. S., Das, S., Pan, I., Leahy, J. J., & Kwapinski, W.," Artifcial neural network-based modelling approach for municipal solid waste gasification in a fluidized bed reactor", Waste Management, 58, 202–213, 2022.

[35] Irvan, R., Abdurohman, M., & Putrada, A.G.," Designing a Monitoring and Prediction System of Water Quality Pollution Using Artifcial Neural Networks for

Freshwater Fish Cultivation in Reservoirs.", In: Proceedings of Sixth International Congress on Information and Communication Technology (pp. 469–476). Springer, Singapore, 2022.

[36] Kouadri, S., Pande, C. B., Panneerselvam, B., Moharir, K. N., & Elbeltagi, A.," Prediction of irrigation groundwater quality parameters using ANN, LSTM, and MLR models", Environmental Science and Pollution Research, 29(14), 21067–21091, 2022.

[37] Abdul-Ahad M.Y, &, Subhee S.N," Forecasting Monthly Water Quality Index Using a Seasonal ARIMA Model for Tigris River at Al-Rashediya Water Station in Baghdad City", Journal of Al-Farabi for Engineering Sciences Vol. 1, No. 2, Des. 2022.

[38] Banejad H., Olyaie E.," Application of an artificial neural network model to rivers water quality indexes prediction–a case study", Journal of American science, 7, 60-65,2011.

[39] Nwobi-Okoye C.C., Ochieze B.Q.," Age hardening process modeling and optimization of aluminum alloy A356/Cow horn particulate composite for brake drum application using RSM, ANN and simulated annealing", Defense Technology, 14, 336-345, 2018.

[40] Najah A., El-Shafie A., Karim O.A., and El-Shafie A.H.," Application of artificial neural networks for water quality prediction", Neural Computing and Applications, 22, 187-201, 2013.

[41] Mijwel M.M., Alsaadi A.S., https://www.researchgate.net/publication/332655457," Overview_of_Neural_Networks", April, 2019.

[42] Banejad H., Olyaie E.," Application of an Artificial Neural Network Model to Rivers Water Quality Indexes Prediction – A Case Study", Journal of American Science, 7(1), 2011.

[43] Al Tobi M.A.S., Bevan G., Wallace P., Harrison D., Ramachandran K." A review on applications of genetic algorithm for artificial neural network", International Journal of Advanced Computational Engineering and Networking, 4, 50-54, 2016.

[44] Gallo C.," Artificial neural networks tutorial", Encyclopedia of Information Science and Technology, Third Edition, IGI Global, 179-189,2015.

[45] Chopra S., Yadav D., Chopra A.N.," Artificial neural networks based Indian stock market price prediction: before and after demonetization", International Journal of Swarm Intelligence and Evolutionary Computation, 8, 1-7, 2019.

[46] Kulisz M., Kujawska J., Przysucha B., Cel W.," Forecasting water quality index in groundwater using artificial neural network", Energies, 14, 2021.