Use of Artificial Neural Network method to Predict the Amount of Oxygen in the Tigris River

Faiez Musa L. Alrufaye¹, Firas H. Neamah², Hakeem I. Mhaibes¹, Marwah Th. Ali³

¹ Middle Technical University, Baghdad, Iraq.
² Basrah University, Basrah, Iraq.
³ University of Wasit, Wasit, Iraq.

E-mail: Faiez.Alrufaye@mtu.edu.iq

Abstract. The neural network algorithm is one of the most important algorithms used to predict results in artificial intelligence. It provides a set of important predictions in this field, given the importance of the Tigris River in the lives of the Iraqi people, and especially the people in the countries that it passes through. We developed an algorithm using neural networks where we sought to measure the proportion of oxygen in the water, the extent of the proportion’s impact on the quality and purity of water and the water’s suitability for human use. In our research we built a system to predict Chemical Oxygen Demand, Biochemical Oxygen Demand and Dissolved Oxygen by inputting a set of parameters as the Input Layer. Weights will be found for each input and, by using the neural network function, the output layer will be generated, where the output layer is Chemical Oxygen Demand, Biochemical Oxygen Demand and Dissolved Oxygen as single output. To test the results and performance of the system, there is an interface in the application which evaluates it by calculating the following statistical parameter: correlation coefficient (r), root mean square error (RMSE) and mean absolute percentage error (MAPE).

1. Introduction

Artificial intelligence is one of the most important computer sciences that have entered into various areas of life. It has helped a great deal in solving many complex problems, especially when predicting future results then comparing them with actual results in order to reach a future solution to the problem. The neural network algorithm is one of the most important algorithms used to predict the results in artificial intelligence. It provides a set of important predictions in this field, given the importance of the Tigris River in the lives of the Iraqi people, and especially the people in the countries that it passes through. We developed an algorithm using neural networks where we sought to measure the proportion of oxygen in the water, the extent of the proportion’s impact on the quality and purity of water and the water’s suitability for human use. The oxygen in water can generally be divided into three basic types: Chemical Oxygen Demand (COD), Biochemical Oxygen Demand (BOD) and Dissolved Oxygen (DO). We are not going to discuss extensively the details of oxygen in the water because there are many sources to clarify (and clarify the importance of knowledge). But we'll talk about it briefly before focusing on the algorithm that has been adopted for the prediction of results.
COD is the amount of chemically consumed oxygen (in milligrams / Liter) required to oxidize organic matter into inorganic substances under certain conditions of time and temperature and in the presence of an oxidizing agent. BOD is the amount of oxygen biologically consumed by microorganisms during their biological activity at a constant temperature. During a specific period of time - called the incubation period - the larger the amount of bio-oxygen consumed, the more contaminated the water. The amount of bio-oxygen consumed depends on the following factors: Water and, prone to degradation, the amount of oxygen dissolved in the water; the amount of food ingredient in water; water temperature; water pH - preferably between (6-8) - as well to the possibility of inhibiting the process of decomposition of materials. Note that the greater the amount of BOD the greater the decrease in the amount of COD, as the ability of bacteria to oxidize decreases the amount of organic matter available and therefore the amount of chemical oxygen required to oxidize it [1].

2. Related Works

There are numerous studies and research projects that have used neural networks in the analysis of river and sea water to obtain accurate guess results. In 2012, Nasr, Moustafa, Seif, and El Kobro designed and implemented the ANN system to predict the performance of the El-Agamy WWTP-Alexandria in terms of Chemical Oxygen Demand (COD), Biochemical Oxygen Demand (BOD) and Total Suspended Solids (TSSs). The ANN can predict the plant performance with a correlation coefficient between the observed and predicted output variables reaching 0.90 [2].

Multilayer Perceptron (MLP), Radial Basis Network and Adaptive Neuro-fuzzy Inference System (ANFIS) are models developed by Emamgholizadeh, Kashi, Marofpoor and Zalaghi, using artificial intelligence techniques, for the Karoon River in Iran, to compute dissolved oxygen (DO), biochemical oxygen demand (BOD) and chemical oxygen demand (COD) [3]. In 2015, researchers from Szént István University, in collaboration with a researcher from Eötvös Loránd University in Hungary, predicted the quality of the Danube water - the second largest river in Europe - using inputs such as PH, runoff, temperature, and electrical conductivity data, using neural networks [4].

The researchers Harun Türkmenler and Murat Pala from Adıyaman University in Turkey developed an application to predict the performance of a wastewater treatment plant, which used a back-propagation learning algorithm. They depended on the Mean Absolute Percentage Error (MAPE), the sum of the Squares Error (SSE), the absolute fraction of variance (R2), the Root-Mean-Square (RMS), the Coefficient of Variation in percent (COV) values, and ANN-models-predicted effluent BOD concentration. The R2 values were found to be 94.13% and 93.18% for the training and test sets of the treatment plant process, respectively. It was found that the ANN model could be employed successfully in estimating the daily BOD in the effluent of wastewater biological treatment plants that was in 2017 [5].

In our research we use a neural network to Predict BOD, COD and DO for The Tigris River - THE important river in Iraq.

3. Artificial Neural Networks

The human mind was the basis for building societies. Deliberately, in order to benefit from its way of thinking, the human mind’s way of working was simulated when building neural networks. The neuron network is similar to the human neuron in its work. An ANN is a collection of connected nodes called artificial neurons, which loosely model the neurons in a biological brain. Each connection, like the synapses in a biological brain, can transmit a signal to other neurons. An artificial neuron is one that receives a signal then processes it and can signal to neurons connected to it [6]. Neural networks go through two stages: the first stage is the training stage, and the second stage is the testing stage. In the first stage the neurons are trained by giving them inputs and outputs as well as the way of working to reach those outputs by using a set of weights. It then matches between the outputs.

In the testing stage, we give only the inputs and weights. The system gives the outputs based on the knowledge base formed through the training stage. Figure 1 illustrates the training stage and Figure 2 illustrates the testing stage.
Neural networks are composed of three layers: The Input layer, which consists of a set of nodes, each node representing an input to the system. The second layer is the hidden layer, which includes the function of neural networks and the activation function. The outputs of the data processing in the hidden layer, are outputs within the Output layer [7].

4. The Proposed System
In our research we built a system to predict COD, BOD, and DO by entering a set of parameters, which may be properties of water or materials present in it. These parameters are represented as the first layer of the system (the input layer). A weight will be generated for each input and the values for this weight are between (0 and 1), and by using the neural network function, an output layer will be generated, with the output layer being COD, BOD and DO as a single output. Note that by increasing...
the neurons in the hidden layer, the results will be more accurate. In a neural network function, the inputs and their weights are produced, and results are summed according to the following function:

\[ y = \sum_{i=1}^{n}(x_i \cdot w_i) + b_i \]  

(1)

When:

- \( x_i \) is a parameter to compute COD, BOD and DO,
- \( w_i \) is a weight which computes as a random number between (0,1) by random function in Visual basic (RandomNumber(X)),
- \( b_i \) is a threshold.

**Figure 4.** Arithmetic neural network algorithm.

To get the final value, we apply the activation function. The purpose of the activation function is to transform the input signal into an output signal and is necessary for neural networks to model complex non-linear patterns that simpler models might miss. There are many types of activation functions—linear, sigmoid, hyperbolic tangent, even step-wise, Figure 5 shows types of activation functions.

**Figure 5.** Activation functions.

For our example, let’s use the sigmoid function for activation, which between (0 to 1), so the main reason why we use it. Therefore, it is used particularly for models where the system has to anticipate the probability as an output. The sigmoid is the correct choice, since the probability of anything only exists between the range 0 and 1.
\[ F(x) = \frac{1}{1 + e^{-x}} \]  

The sigmoid function looks like this, graphically:

![Sigmoid function graph](image)

**Figure 6.** Sigmoid function graph.

### 5. Implementation And Results

As it is known in neural networks, it goes through two stages of work, in the first stage the cells are trained by providing them with inputs and outputs, and then the hidden layer is based on the light of those inputs and outputs, applying the equation of the neural networks to obtain results close to the outputs.

#### 5.1. Training Stage

The application is designed and implemented using Visual Basic version 2013, where the parameters are imported by Excel file and this is considered as input to the system - except COD, BOD and DO which are outputs for the system.

The neural network algorithm works to apply the neural network equation on the inputs to produce new values for BOD, COD and DO - then the system matches the new values with our target values. This is called the training stage. Figure 7 illustrates the training stage implementation interface.

The number of nodes in the input layer is on the number of entries plus one node of the basic term. As for the number of nodes in the hidden layer, this depends on the user's input, as the greater the number of nodes the more accurate the results are, in the training phase we entered 50 as the number of nodes in the hidden layer and the results were good but not close as we want, so we entered the number 100 for the number of nodes Hidden, so we came to very good results. The result nodes are only 3, and they are COD, BOD and DO. Figure 7 shows the number of nodes in the hidden layer interface.

![Number of nodes in the hidden layer interface](image)

**Figure 7.** The number of nodes in the hidden layer interface.

The inputs were a set of parameters collected and measured for the waters of the Tigris River, where 500 meters was adopted as the standard distance for measuring these parameters. The data collection process took 6 months. In our current research, many of the inputs that we found do not affect the results either negatively or positively, such as the time of sampling from the river, water flow, and the distance between the sampling point to another point, were neglected, and the inputs in the following table were relied upon for their significant impact on the results.
Table 1. The system parameters (Inputs)

| Parameters | Meaning |
|------------|---------|
| PH         | PH is a measure of acidity or alkalinity of water. |
| TDS        | Total Dissolved Solids causes unacceptable taste in drinking water. |
| EC         | Electrical Conductivity is the measure of the amount of electrical current a material can carry or its ability to carry a current. |
| TN         | Total Nitrogen: It is an important nutrient for aquatic plants. |
| TP         | Total Phosphorus: It is an important nutrient for aquatic plants. |
| CL         | Chlorine is added for the purpose of sterilization and killing bacteria and microbes in water. |

Table 2 illustrates the system parameters (Inputs) and results (Outputs) in training stage.

Table 2. The system parameters (Inputs) and results (Outputs) in training stage.

| PH  | TDS  | EC    | TN    | TP    | CL    | DO    | COD   | BOD   | NN-BOD | NN-DO  | NN-COD |
|-----|------|-------|-------|-------|-------|-------|-------|-------|--------|--------|--------|
| 7.4 | 76   | 1663  | 1130  | 0.033 | 120   | 7.98  | 55    | 14.5  | 14.51954438 | 7.21   | 54     |
| 10.5| 476  | 2461  | 2183  | 0.101 | 299   | 3.39  | 318   | 335.1 | 335.5329975 | 3.2    | 317    |
| 10.3| 465  | 2445  | 2133  | 0.098 | 284   | 3.78  | 315   | 329.7 | 330.1327406 | 3.28   | 314    |
| 10.2| 455  | 2425  | 2084  | 0.086 | 279   | 4.36  | 310   | 324.3 | 324.7205859 | 3.86   | 309    |
| 10   | 447  | 2409  | 2009  | 0.082 | 274   | 4.45  | 300   | 316.7 | 317.1068224 | 3.95   | 299    |
| 9.8  | 428  | 2398  | 1989  | 0.08  | 267   | 4.6   | 298   | 311.3 | 311.7002058 | 4.1    | 297    |
| 9.5  | 422  | 2379  | 1964  | 0.077 | 262   | 4.79  | 273   | 297.3 | 297.6827666 | 4.29   | 272    |
| 9.2  | 401  | 2365  | 1926  | 0.074 | 253   | 4.89  | 265   | 271.9 | 272.250271   | 4.39   | 264    |
| 9    | 358  | 2349  | 1912  | 0.069 | 249   | 4.99  | 254   | 260.6 | 260.9464215 | 4.49   | 253    |
| 8.9  | 339  | 2334  | 1899  | 0.066 | 242   | 5.18  | 237   | 237.8 | 238.107275  | 4.68   | 236    |
| 8.9  | 319  | 2324  | 1857  | 0.064 | 245   | 5.37  | 225   | 217.1 | 217.3804302 | 4.87   | 224    |
| 8.7  | 300  | 2313  | 1843  | 0.062 | 240   | 5.51  | 218   | 193.4 | 193.6497724 | 5.01   | 217    |
| 8.6  | 262  | 2294  | 1818  | 0.059 | 231   | 5.56  | 198   | 170.6 | 170.8198143 | 5.06   | 197    |
| 8.5  | 220  | 2283  | 1785  | 0.056 | 209   | 5.94  | 176   | 147.9 | 148.0919554 | 5.44   | 175    |
| 8.4  | 184  | 2233  | 1725  | 0.053 | 207   | 5.99  | 164   | 136.5 | 136.6771585 | 5.49   | 163    |
| 8.1  | 168  | 1970  | 1697  | 0.051 | 200   | 6.04  | 150   | 108.6 | 108.7403202 | 5.54   | 149    |
| 7.9  | 131  | 1934  | 1645  | 0.045 | 192   | 6.52  | 138   | 95.1  | 95.22198429 | 6.02   | 137    |
| 7.7  | 79   | 1868  | 1563  | 0.044 | 185   | 6.62  | 121   | 87.9  | 88.01275003 | 6.12   | 120    |
| 7.7  | 100  | 1855  | 1425  | 0.042 | 170   | 6.76  | 106   | 77.6  | 77.69959012 | 6.26   | 105    |
| 7.6  | 79   | 1833  | 1325  | 0.04  | 163   | 6.81  | 78    | 64.1  | 64.1824324  | 6.31   | 77     |
| 7.6  | 68   | 1812  | 1223  | 0.036 | 160   | 6.93  | 53    | 44.5  | 44.5572276  | 6.43   | 52     |
| 7.3  | 47   | 1798  | 1190  | 0.034 | 154   | 7.19  | 34    | 25.9  | 25.9330837  | 6.69   | 33     |
| 7.6  | 77   | 1704  | 1164  | 0.033 | 122   | 8.18  | 57    | 14.8  | 14.8208732  | 7.68   | 56     |
| 10.7 | 485  | 2522  | 2248  | 0.102 | 305   | 3.47  | 334   | 341.8 | 342.2373292 | 2.97   | 333    |
| 10.5 | 475  | 2507  | 2197  | 0.098 | 290   | 3.87  | 330   | 336.3 | 336.7295896 | 3.37   | 329    |
| 10.4 | 464  | 2486  | 2146  | 0.087 | 285   | 4.47  | 325   | 330.8 | 331.2203571 | 3.97   | 324    |
| 10.2 | 456  | 2469  | 2069  | 0.083 | 280   | 4.57  | 315   | 323.1 | 323.5084249 | 4.07   | 314    |
5.2. Testing Stage
In the testing stage, only inputs without outputs are used to produce new outputs. At this stage, we enter the inputs into the system (which are the same inputs in the training phase), after that the system applies the function of neural networks to it, which is within the hidden layer, where the user enters the same number of nodes for the hidden layer in the training phase, and on the basis of training for neurons. In the training phase, we will get the results of COD, BOD and DO. Table 3 shows the system parameters (Inputs) and results (Outputs) in testing stage.

| PH | TDS | EC | TN | TP | CL | NN-BOD | NN-DO | NN-COD |
|----|-----|----|----|----|----|--------|-------|--------|
| 7.4| 76  | 1663| 1130| 0.033| 120| 14.3855555| 7.58| 54     |
| 10.5| 476| 2461| 2183| 0.101| 299| 335.1022222| 3.6| 317    |
| 10.3| 465| 2445| 2133| 0.098| 284| 330.7014412| 3.52| 314    |
| 10.2| 455| 2425| 2084| 0.086| 279| 324.8755555| 3.86| 309    |
| 10  | 447| 2409| 2009| 0.082| 274| 317.8744444| 3.95| 299    |
| 9.8 | 428| 2398| 1989| 0.08 | 267| 311.5010215| 4.1 | 297    |
| 9.5 | 422| 2379| 1964| 0.077| 262| 297.8575246| 4.29| 272    |
| 9.2 | 401| 2365| 1926| 0.074| 253| 272.0125555| 4.39| 264    |
| 9   | 358| 2349| 1912| 0.069| 249| 260.728963 | 4.49| 253    |
| 8.9 | 339| 2334| 1899| 0.066| 242| 238.1107275| 4.68| 236    |
| 8.9 | 319| 2324| 1857| 0.064| 245| 217.8022545| 4.87| 224    |
| 8.7 | 300| 2313| 1843| 0.062| 237| 193.3254855| 5.01| 217    |
| 8.6 | 262| 2294| 1818| 0.059| 231| 170.6254555| 5.06| 197    |
| 8.5 | 220| 2283| 1785| 0.056| 209| 148.2547896| 5.44| 175    |
| 8.4 | 184| 2233| 1725| 0.053| 207| 136.2547855| 5.49| 163    |
| 8.1 | 168| 1970| 1697| 0.051| 200| 108.7403202| 5.54| 149    |
| 7.9 | 131| 1934| 1645| 0.045| 192| 95.22198429| 6.02| 137    |
| 7.7 | 79 | 1868| 1563| 0.044| 185| 88.01275003| 6.12| 120    |
| 7.7 | 100| 1855| 1425| 0.04 | 170| 77.69959012 | 6.26| 105    |
| 7.6 | 79 | 1833| 1325| 0.04 | 163| 64.1824324 | 6.31| 77     |
| 7.6 | 68 | 1812| 1223| 0.036| 160| 44.5572276 | 6.43| 52     |
| 7.3 | 47 | 1798| 1190| 0.034| 154| 25.93330837 | 6.69| 33     |
| 7.6 | 77 | 1704| 1164| 0.033| 122| 14.82087327 | 6.78| 56     |
| 10.7| 485| 2522| 2248| 0.102| 305| 342.2373292 | 2.97| 333    |
| 10.5| 475| 2507| 2197| 0.098| 290| 336.7295896 | 3.37| 329    |
| 10.4| 464| 2486| 2146| 0.087| 285| 331.2203571 | 3.97| 324    |
| 10.2| 456| 2469| 2069| 0.083| 280| 323.5084249 | 4.07| 314    |

To test the results and performance of the system, an interface in the application, evaluates by calculating the following statistical parameter: correlation coefficient (r), root mean square error (RMSE) and mean absolute percentage error (MAPE) defined by Eqs. (3-5), respectively [8]:

$$R = \frac{\sum (Q^o - M^o)(Q^p - M^p)}{\sqrt{\sum (Q^o - M^o)^2 \sum (Q^p - M^p)^2}}$$  \hspace{1cm} (3)

$$RMSE = \frac{1}{N} \sum (Q^o - Q^p)$$  \hspace{1cm} (4)
MAPE = \frac{1}{N} \sum |Q^o - Q^p| \tag{5}

Where $Q^o$ and $Q^p$ are the observed and estimated concentrations at the time steps, $Mo$ and $Mp$ are the mean of the observed and estimated concentrations, respectively, and $N$ is the total number of observations of the data set. The RMSE and MAPE measure the errors –however, RMSE is the most popular measure of errors which receives much greater attention than small errors [8], Figure 8 illustrates the system evaluation interface.

![Figure 8. System evaluation interface.](image)

6. Conclusion
When designing and implementing a neural network system, we observed that, there are many constraints, the most important of which is the collection of large amounts of data for the system, because it depends largely on the amount of data entered to the testing algorithm to obtain results comparable to the target results. The system depends very heavily on the data entered. The number of neurons in the hidden layer also has a large role in the accuracy of the results, where increasing the number of neurons in the hidden layer leads to an increase in the weights and is thus reflected positively on the system output. The process of evaluating the system and its results was excellent, as the system was evaluated using correlation coefficient ($r$), root mean square error (RMSE) and mean absolute percentage error (MAPE).

References
[1] Ruth E Weiner, Robin A. Matthews, "ENVIRONMENTAL ENGINEERING: Fourth Edition ", Elsevier Science (USA), 2013.
[2] Mahmoud S. Nasr, Medhat A.E. Moustafa, Hamdy A.E. Seif, Galal El Kobrosy, "Application of Artificial Neural Network (ANN) for the prediction of EL-AGAMY wastewater treatment plant performance-EGYPT", Alexandria Engineering Journal 51, Pages 37–43, 2012.
[3] S. Emamgholizadeh, H. Kashi, I. Marofpoor, E. Zalaghi, "Prediction of water quality parameters of Karoon River (Iran) by artificial intelligence-based models", Int. J. Environ. Sci. Technol. 11: Pages 645–656, 2014.
[4] A. Csábrági, S. Molnár, P. Tanos, J. Kovács, "FORECASTING OF Dissolved Oxygen In The River Danube Using Neural Networks", Hungarian Agricultural Engineering, Pages 38-41, 2015.
[5] Harun Türkmenler, Murat Pala, "Performance Assessment of Advanced Biological Wastewater Treatment Plants Using Artificial Neural Networks ", INTERNATIONAL JOURNAL OF ENGINEERING TECHNOLOGIES-IJET Harun Türkmenler et al., Vol.3, No.3, Pages 151-156, 2017.
[6] V.N. Delgrange, N. Cabassud, M. Cabassud, L. Durand Bourlier, J.M. Laine, "Neural networks for prediction of ultrafiltration transmembrane pressure: application to drinking water
production", Journal of Membrane Science 150, Pages 111–123, 1998.

[7] R. A. Naji, E. J. Khadum, "Efficiency of Using Artificial Neural Networks to Predict Child Mortality in Najaf Province and Neuroshell Predictor Program", AL GHAREE for Economics and Administration Sciences, Vol.14, No.38 Pages 311–320, 2016.

[8] Ayad Sleibi Mustafa, "Artificial Neural Networks Modeling of Total Dissolved Solid in the Selected Locations on Tigris River, Iraq", Journal of Engineering, College of Engineering University of Baghdad, Vol.21, No.6 Pages 162–179, 2015.