Journal of AL-Rafidain University College for Sciences (2023); Issue 54; 518 - 530

AL- Rafidain University College

PISSN: (1681-6870); EISSN: (2790-2293)

Journal of AL-Rafidain University College for Sciences

Available online at: https://www.jrucs.iq

JRUCS

Journal of AL-Rafidain University College for Sciences

Using Neural Networks (Back Propagation Network) to Determine the Variables Affecting Water Pollution in the Euphrates River

Ons E. Musa	Sabah M. Ridha
ons.edin1001a@coadec.uobaghdad.edu.iq	drsabah@coadec.uobaghdad.edu.iq
College of Physical Education and Sport Sciences - Mustansiriyah University, Baghdad, Iraq	Department of Statistics - College of Administration and Economic - Baghdad University, Baghdad, Iraq

Article Information

Article History:
Received: December, 24, 2022
Accepted: March, 3, 2023
Available Online: December, 31, 2023

Keywords:
principal component analysis, neural networks, back propagation, Euphrates River, Water Pollution.

Correspondence:
Ons E. Musa
ons.edin1001a@coadec.uobaghdad.edu.iq

https://doi.org/10.55562/jrucs.v54i1.619

Abstract

With the economy's steady growth and rapid development of modern industrial technology, environmental problems have increased. Water pollution is one of the complex problems that most countries of the world suffer from, especially developing countries, because it contains various pollutants, some of which can decompose as organic pollutants, and some of which are difficult to decompose, such as toxic heavy metals. In Iraq, according to the Iraqi Ministry of Environment - Water Pollution Division statistics for the year 2018, specifically the waters of the Euphrates River, a group of variables affecting the water quality of this river has been identified. The scarcity of water and the accumulation of various wastes in the water networks and on the banks of the Euphrates River also contributed greatly to the increase in the problem of pollution. The continued poor quality of the water has led to severe environmental health concerns. The research aims to shed light on the most important variables affecting the pollution of the waters of the Euphrates River and the water quality and quality in this river. The principal components analysis *and the neural network (*backpropagation algorithm) *were used to determine the variables that affect the pollution of the Euphrates River water. They were applied to real data from the Ministry of Environment for the Euphrates water pollution. One of the most important conclusions we reached when using statistical methods was that the variable (DO2) is not significant and does not affect the pollution of the Euphrates River either. Still, When Using the algorithm, we get that all these variables (TH, TDS, EC) do not affect the water pollution of the Euphrates River.*

1. Introduction

Artificial intelligence (AI) is a scientific field that is developing rapidly, as it has many important applications in practical life and various natural, scientific, and human sciences. Artificial intelligence includes thinking, knowledge, planning, learning, communication, and the ability to deal with the target. Artificial intelligence algorithms are based on the principle and concept of artificial intelligence. These algorithms are characterized by their ability to devise dynamic methods commensurate with the nature of the problem to be studied and to determine a suitable practical method to find an optimal solution among the set of solutions to the problem. Also, these algorithms improve the solution's value according to the obstacles and variables.

Large datasets are increasingly widespread in many disciplines. To interpret such datasets, methods are required to drastically reduce their dimensionality in an interpretable way, such that most of the information in the data is preserved.

In this research, we will use principal components analysis to determine the variables affecting the Euphrates River's water pollution. We will also use the neural network algorithm, specifically the backpropagation algorithm, to determine the variables affecting the pollution of the Euphrates River.

We will mention some authors who wrote on the topic of water pollution using a neural network algorithm. (Zhao, K., He, T., Wu, S., Wang, S., Dai, B., Yang, Q. and Lei, Y., 2019) They analyzed the results of various parameters through the experiments, and continuously modified the network parameters to improve the accuracy of the evaluation results. Taking the regional environment as an example, they proposed two monitoring methods, and a variety of neural network models were used to analyze each prediction method [1], (Lungu, E., Oprea, M. and Dunea, D., 2008) they presented an application of short-time feed-forward artificial neural networks for air pollution prediction. The time series used in the experiments are measurements of some air pollutants specific to urban areas (such as NO2, SO2, PM10, and TSP). They focused on the comparisons made between the RPop and Quickproptraining algorithms concerning the MSE obtained after a fixed number of epochs [2], (Jiang, X., Zhang, P. and Huang, J., 2022) They presented a method for processing environmental pollution data and designing experiments. The first experiment is to compare the model in this paper with the other three neural network models, and it is found that the model in this paper is smaller than the other models in terms of ARE and MAE; The second experiment is to design an environmental pollution prediction system based on the model in this paper and use the system to predict the value of sulfur dioxide in two provinces [3], (Hemavani, B. and Rao, G.S., 2020) They used principal component analysis and cluster analysis to understand the associations between different air pollutants and meteorological parameters in Hyderabad, India. The analysis is done using long data, and correlations are obtained seasonally as the results of both analyzes show similar trends, and the adequacy of KMOtest sampling and Bartlett test significance values prove satisfactory results of the analysis [4], (Nasir, M.F.M., Samsudin, M.S., Mohamad, I., Awaluddin, M.R.A., Mansor, M.A., Juahir, H. and Ramli, N., 2011) They developed accurate multiple linear regression techniques as an advanced tool for surface water modeling and prediction and used principal component analysis (PCA) to simplify and understand the complex relationship between water quality parameters. The result showed that using PCA as input improved the prediction of the MLR model by reducing its complexity and eliminating the collinearity of the data [5].

2. Research goal

1. Determination of the variables affecting the pollution of the Euphrates River water using the principal component analysis (PCA).
2. Determining the variables affecting the pollution of the Euphrates River water using the backpropagation algorithm.

3. Theoretical side

3.1. Principal Component Analysis (PCA)

Principal component analysis (PCA) has become one of the most useful tools for data modeling, compression, and visualization. It solves a dimensional fit problem for a set of data points in a high-dimensional space [6].

Principal component analysis refers to the problem of fitting a low-dimensional subspace S of dimension d "D into a set of points $(X_1, X_2, \dots, X_n)$ in a high-dimensional space R^D [6].

The central idea of Principal Component Analysis (PCA) is to reduce the Dimensions of a data set consisting of a large number of interrelated variables while retaining the greatest amount of Variance found in the data set [7]. This is done by converting to a new set of variables, The principal components, which are not linked, retain the first most of the Variance found throughout original variants [7].

It does so by creating new uncorrelated variables that successively maximize Variance. Finding such new variables, the principal components, reduces to solving an eigenvalue/eigenvector problem, and the new variables are defined by the dataset at hand, not a priori, hence making PCA an adaptive data analysis technique [8].

3.2. Definition and Derivation of Principal Component Analysis (PCA) :

Suppose that X is a vector of P random variables and that the variances of the P random variables and the structure of the covariances or correlations between the P variables are of interest. Unless P is small or the structure is very simple, it will often not be very helpful to look at the P variances and all of the $\frac{1}{2}P(P-1)$ correlations or covariances. An alternative approach is to look for a few "P derived variables that preserve most of the information given by these variances and correlations or covariances [7].

While PCA does not ignore covariances, nor does it ignore correlations, it does focus on variances. The first step is to find a linear function $\acute{\alpha}_1 x$ of the elements of x have maximum Variance, where α_1 is a vector of P constants $\alpha_{11}, \alpha_{12}, \dots, \alpha_{1P}$, and refers to transposing so that [7]

$$\acute{\alpha}_1 x = \alpha_{11}x_1 + \alpha_{12}x_2 + \cdots + \alpha_{1P}x_P = \sum_{j=1}^{P} \alpha_{1j}x_j \tag{1}$$

Whereas $\acute{\alpha}_2 x$ uncorrelated with $\acute{\alpha}_1 x$ having a maximum variance, as well as the rest of the linear functions, so that the kth a linear function $\acute{\alpha}_k x$ is found to have a maximum variance of being uncorrelated with $\acute{\alpha}_1 x, \acute{\alpha}_2 x, \dots, \acute{\alpha}_{k-1} x$. The kth derived variable, $\acute{\alpha}_k x$ is the kth PC. Most of the variation x will be accounted for by m PCs, where m "P [7].

To drive the form of PCs, let first $\acute{\alpha}_1 x$; the vector α_1 maximizes $var[\acute{\alpha}_1 x] = \acute{\alpha}_1 \sum \alpha_1$. To maximize $\acute{\alpha}_1 \sum \alpha_1$ subject to $\acute{\alpha}_1 \alpha_1 = 1$, the standard approach uses the Lagrange multipliers technique. Maximize [7]

$$\acute{\alpha}_1 \sum \alpha_1 - \lambda(\acute{\alpha}_1 \alpha_1 - 1) \tag{2}$$

where λ is a Lagrange multiplier. Differentiation concerning α_1 gives

$$\sum \alpha_1 - \lambda \alpha_1 = 0 \tag{3}$$

$$\left(\sum - \lambda I_P\right) \alpha_1 = 0 \tag{4}$$

Where

I_P : is $P * P$ identity matrix.

λ : is eigenvalue of $\sum$.

α_1: is the corresponding eigenvector.

Decide the P eigenvectors gives $\acute{\alpha}_1 x$ with maximum Variance, the quantity to be maximized is

$$\acute{\alpha}_1 \sum \alpha_1 = \acute{\alpha}_1 \lambda \alpha_1 = \lambda \acute{\alpha}_1 \alpha_1 = \lambda \qquad (5)$$

λ must be as large as possible.
α_1 is the eigenvector corresponding to the largest eigenvalue of $\sum$.
$var(\acute{\alpha}_1 x) = \acute{\alpha}_1 \sum \alpha_1 = \lambda_1$ the largest eigenvalue. In general, the kth PC of x is $\acute{\alpha}_k x$ and $var(\acute{\alpha}_k x) = \lambda_k$.
λ_k : is the kth largest eigenvalue of $\sum$.
α_k : is the corresponding eigenvector.[7]

3.3. Neural Network algorithm :

Interest in neural networks has increased in the past years. It is successfully applied to solve problems in various scientific fields, such as environment, finance, medicine, engineering, geology, and physics [9].

The main goals of the algorithm of artificial neural networks are to build artificial intelligence that has the ability and flexibility to teach how the brain and nervous system perform. Humans and their intersections with the rest of the cells regarding their foundations and the desire to build and design machines capable of solving complex problems that operate sequentially. Computers cannot solve them [9].

3.4. Multilayer neural networks :

The multi-layered neural network algorithm eliminates many limitations of single-layer network calculations. The sigmoid function is used because it is suitable for use as a transformation function in multilayer networks.

$$F(X) = \frac{1}{1 + e^{(-\beta X)}} \quad , \quad \beta > 0 \qquad (6)$$

The best method for these types of neural networks is the Back Propagation Error method, based on generating the Gradient Method [the sum of the differences between the actual output and the desired output and applying the Minimization Method].
A backpropagation network is one of the most important and widely used neural networks.

3.5. Back propagation algorithm

The backpropagation algorithm is the most common type of NN algorithm. The Back Propagation Algorithm was used as having the following properties:

- Type of learning strategy: Supervised learning strategy
- Type of learning: Error Correction
- Architecture type: Multilayer Feed Forward type
- Type of application: Pattern Recognition.

The following steps can describe the appurtenant backpropagation algorithm [10]:

a) Initialization.
b) Pattern submitting.
c) Comparison.
d) Back propagation of an error and weight modification.
e) The values termination of pattern selection from the training set.
f) Termination of the learning process.

The algorithm is stopped when the value of the error function has become sufficiently small [9]. The general idea of the Back Propagation algorithm can be explained as follows:

1. Select the next test pair from the test set and apply the input vector to the network to input and customize the desired output vector.
2. Calculation of the actual output of the network.
3. Calculating the error between the real output and the desired output of the neural network.
4. Adjust the weights so that the error becomes very small.

5. Return to step (2) and repeat the process for each output vector in the experimental group until the error becomes less than the permissible limit [11].

3.6. Flowchart of the Back propagation algorithm

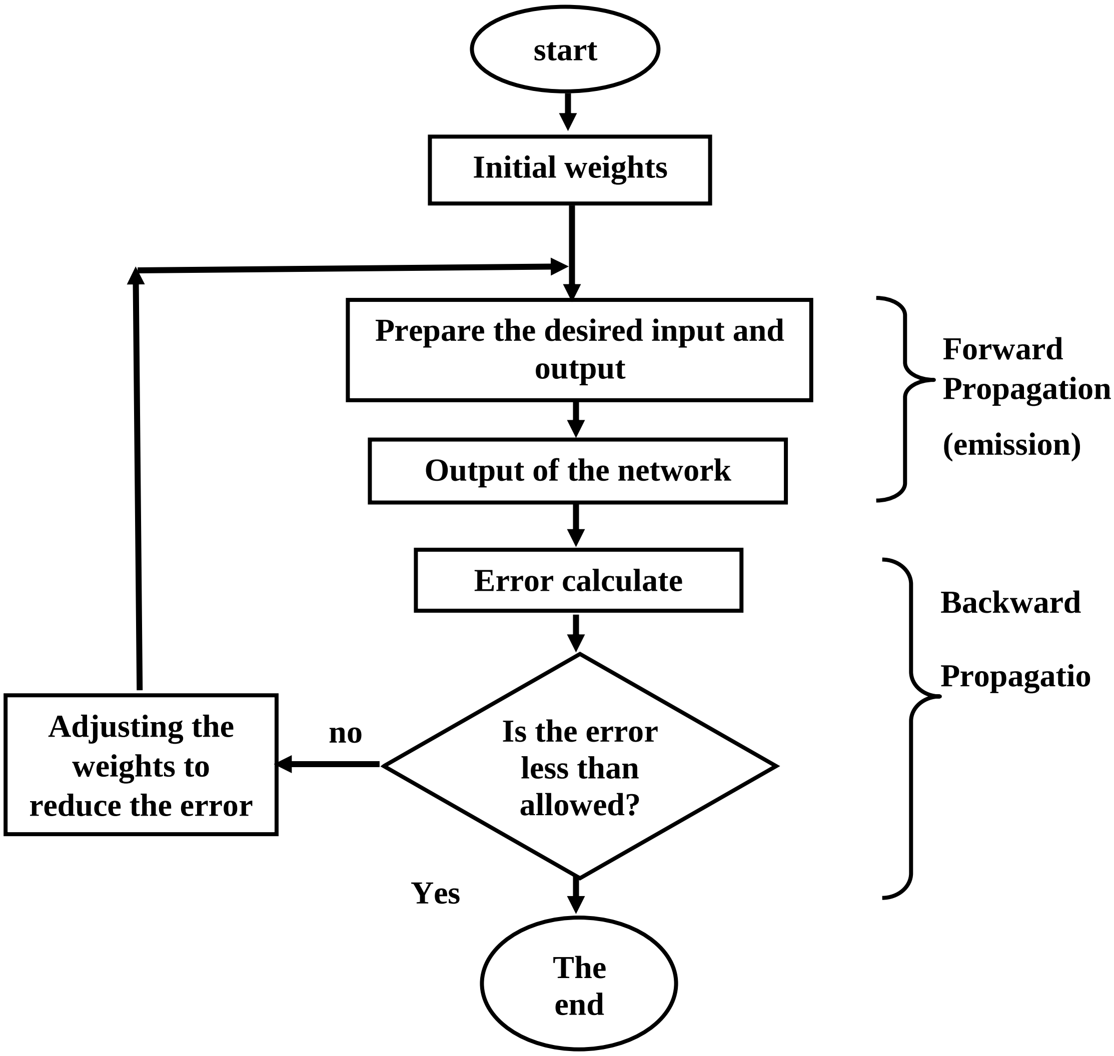

Figure (1): Flowchart of the Backpropagation algorithm

3.7. Procedures Back propagation algorithm

1. Write the necessary office functions in the program.
2. Write the program start (main) statement.
3. Write the parenthesis at the beginning of the program.
4. Define the necessary variables in work.
5. Give random values for w-weights using the random function.
6. Define the learning coefficient and give it a value of a=1.
7. Enter a matrix representing the inputs and outputs of the logical gate trained on it.
8. Matrix rotary writing.
9. Passing the inputs on the network to the output layer.
10. Get the sum of the inputs entering the output layer.
11. Extract the value of the network's actual output (output).
12. Comparing the actual output (output) with the required output when an error is detected Between the two comparisons, in this case, we need to adjust the weights and determine the amount of The error.
13. Modify the weights that link the output layer and the hidden layer
14. We continue in the same way, but this time between the hidden layer and the input layer.

4. Application side

Environmental pollution, especially water pollution, has become an increasingly significant and common problem in various countries.

In this aspect, what was stated in the theoretical aspect will be studied and applied to real data related to the environmental aspect (water pollution/pollution of the Euphrates River) for 2018.

The Euphrates River water pollution was studied, and what are the factors affecting it, where the influencing factors were: pH function (PH), temperature (Temp), dissolved oxygen (Do.), phosphates (PO_4), nitrates (NO_3), calcium (Ca), magnesium (Mg), total hardness (TH), potassium (K), sodium (Na), sulfates (SO_4), chlorides (Cl), dissolved solid salts (TDS), conductivity (EC), banality (AlK), turbidity (Turb). This data was taken from the Iraqi Ministry of Environment / Technical Department Water Pollution Division.

Stations for monitoring and measuring pollutants in the water of the Euphrates River:

Twenty-two monitoring stations have been installed and distributed in Iraqi governorates: Anbar, Babel, Karbala, Najaf, Qadisiyah, Muthanna, Dhi Qar, and finally, Basra.

5. Results and discussion

The real data results were obtained using MATLAB 2016a.

- **Results of the factorial analysis :**

Factorial analysis was used to determine the variables affecting the pollution of the Euphrates River water. Principal components were used before recycling, and the Varimax method after recycling. To judge the importance of the variables within the factors, loadings (saturation ratios) were used. More significant than (0.50), and the results were as follows:

1. In the beginning, (KMO and Bartlett's) were tested to judge the adequacy of the sample and to try to determine whether the correlation matrix is monolithic or not, as it is required in the factorial analysis that the correlation matrix is not monolithic, as from the following table

Table (1): Shows KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.558
Bartlett's Test of Sphericity	**Approx. Chi-Square**	676.428
	Df	105
	Sig.	0.000

We note that the value of (Kaiser-Meyer) amounted to (0.558), which is greater than the value of the minimum limit for this test, which is (0.50), and this means that we can judge the adequacy of the sample size in this analysis.

As for the Bartlett test (to prove that the correlation matrix is a non-monometric matrix), the following hypothesis must be tested :

- Null hypothesis: The correlation matrix is unimodal
- Alternative hypothesis: The correlation matrix is not monadic

Through the Bartlett value of (676.428) and a level of significance (0.000), which is much smaller than the level of significance (0.05), this means rejecting the null hypothesis and accepting the alternative hypothesis that the correlation matrix is not monolithic.

2. Explained total variance table: From the table (2):

Three factors appeared, distributed among the research paragraphs, with a rate of (80.6%), which is an acceptable percentage. The remaining percentage (19.4%) is due to other variables that may be variables that the researcher did not consider and which are essential or uncontrolled. By observing the differences for each Factor before recycling, we found that it differed after recycling, especially the first Factor, as its percentage before recycling was (57.292%), while after recycling, it became (37.726). This difference means that the difference is distributed among the rest of the factors after recycling, and thus we can say that the Results after recycling were the best.

Table (2): Shows Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	8.594	57.292	57.292	8.594	57.292	57.292	5.659	37.726	37.726
2	1.817	12.112	69.404	1.817	12.112	69.404	4.735	31.564	69.291
3	1.686	11.239	80.643	1.686	11.239	80.643	1.703	11.352	80.643
4	.989	6.591	87.233						
5	.804	5.357	92.590						
6	.482	3.213	95.803						
7	.318	2.122	97.925						
8	.268	1.789	99.714						
9	.027	.180	99.894						
10	.008	.053	99.947						
11	.006	.038	99.985						
12	.002	.010	99.995						
13	.001	.004	99.999						
14	9.675E-5	.001	100.000						
15	1.903E-5	.000	100.000						

Extraction Method: Principal Component Analysis.

3. Components Matrix Table: Here, two tables will be presented, one of which is the results before the Rotation, and the other is the results after the Rotation. Only the essential variables whose saturation ratios are greater than (0.50) will be displayed, i.e., the variables affecting the study and according to their importance in the Factor, then after that naming these factors, the results are as follows :

❖ **Pre-roll results :**

Table (3): Shows the saturation percentages for the variables before Rotation

	1	2	3
TH	.986		
Mg	.974		
Ca	.960		
EC	.924		
TDS	.897		
Na	.888		
K	.869		
SO4	.869		
CL	.839		
PO4	.564	.558	
Alk	.556		
PH		.717	
DO2			
Temp			.793
NO3			.773
Extraction Method: Principal Component Analysis.			
a. three components extracted.			

From the above table, which shows the results before Rotation, we note that most variables were distributed among the first Factor. The rest of the variables were scattered among the rest of the factors, as one or two variables appeared in some factors. The Factor must contain at least three variables to be taken as an influencing factor. For this reason, we will rely on the results after the Rotation, which distributed the factors better.

- **Results after roll**

Table (4): Shows the saturation percentages for the variables after Rotation

	1	2	3
K	.881		
Ca	.873		
SO4	.847		
Na	.844		
PH	.788		
TH	.757		
Mg	.700		
Alk	.577		
TDS		.849	
CL		.847	
EC		.829	
PO4			.770
DO2			
Temp			.801
NO3			.772
Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.			
a. Rotation converged in 6 iterations.			

Through the above table, the paragraphs were distributed among the factors according to their importance as follows :

1. The variables according to their importance (AIK, Mg, TH, PH, Na, SO_4, Ca, K) in the first Factor, with a variance rate for this Factor (37.726%).
2. The variables according to their importance (EC, CL, TDS) in the second Factor, with a variance rate for this Factor (31.564%).
3. The variables according to their importance (NO3, Temp, PO4) in the third Factor, with a variance rate for this Factor (11.352%).

We also note that the variable (DO2) did not appear, which means this variable is unimportant.

- **Results of Artificial Neural Networks (ANN) :**

To obtain the best variables and according to their order in their effect on the water pollution of the Euphrates River, we will use artificial neural networks (ANNs) by showing the errors for each variable and comparing them with the general error.

The models of neural networks are nonlinear functions that gain more community confidence, while the general form of the function of artificial neural networks is as in the following model:

Y=F[H[suB] 1[/suB](X),H[suB]2[/suB](X), . . . , HN(X)] + U (7)

Where as:
Y : Dependent Variable
X : represents the independent or explanatory variables
$F - H$: define the functions of neural networks (activation - transmission)
U : Error term in the Error Term function
(1000) repetitions were used during network training, and the number of hidden layers was used according to the number of variables entered.
The network inputs were fifteen variables: PH, Temp, Do_2, PO_4, NO_3, Ca, Mg, TH, K, Na, SO_4, Cl, TDS, EC, and AlK.
The data included the governorates located on the Euphrates River, and in each governorate, there are several stations. Hidden nodes (15) were identified, and the following results were obtained:

Table (5): ANN application results

	average errors	order of importance
PH	0.0131	3
Temp	0.0412	6
DO2	0.0137	4
PO4	0.0002	1
NO3	0.0066	2
Ca	0.2404	8
Mg	0.1345	7
TH	**1.1294**	**Not important**
K	0.0187	5
Na	0.4403	10
SO4	0.6622	11
CL	0.6790	12
TDS	**2.7009**	**Not important**
EC	**4.2129**	**Not important**
Alk	0.3091	9
PH	0.7201	general average errors

By table (5) and compared with the general average errors, It was found that there are variables with errors that are greater than the general average of the errors, and this means that these variables are not important. They did not appear to affect water pollution in the Euphrates River. These variables are (TH, TDS, EC), and Figure (2), which shows the architecture of the neural network, where we note the highest frequency in training the network was (57).

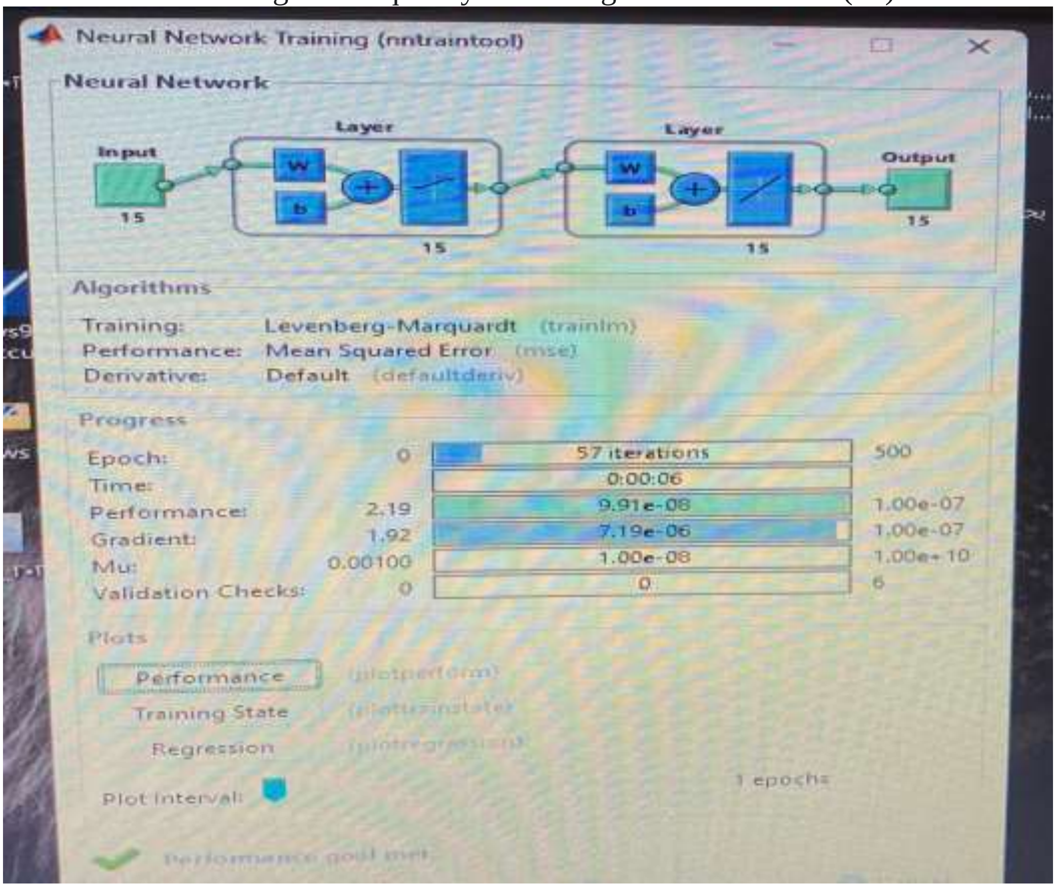

Figure (2): Shows the neural network architecture

The performance of the network is shown in Figure (3), which shows the regression line of the neural network, where we notice that the value of the coefficient of determination during training is very large, reaching (R = 0.99997), and this indicates that the performance of the network is high.

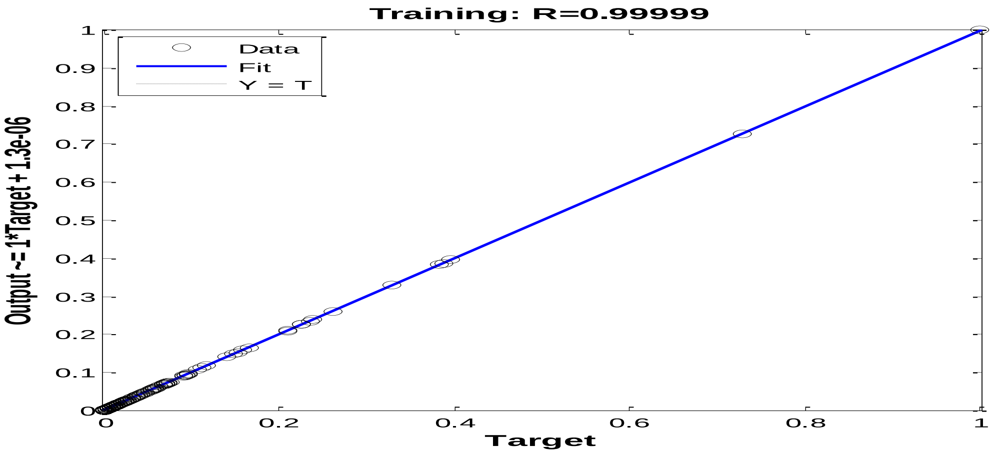

Figure (3): Shows the regression line equation of the neural network

As we can see in Figure (3), which shows the stability of the neural network through the stability of the mean square error (MSE) values when training.

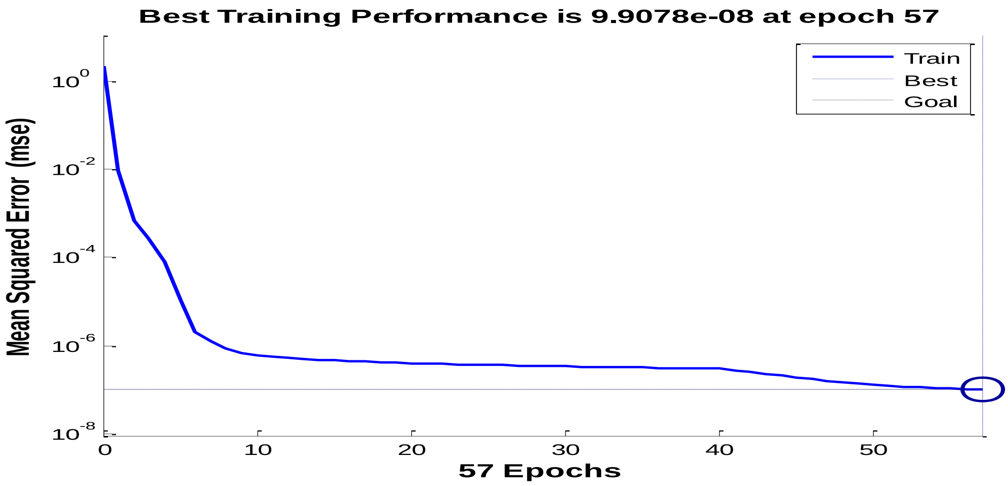

Figure (4): Shows the stability of mean square error values when training the neural network

We also notice in Figure (5) the stability states that the network undergoes when training

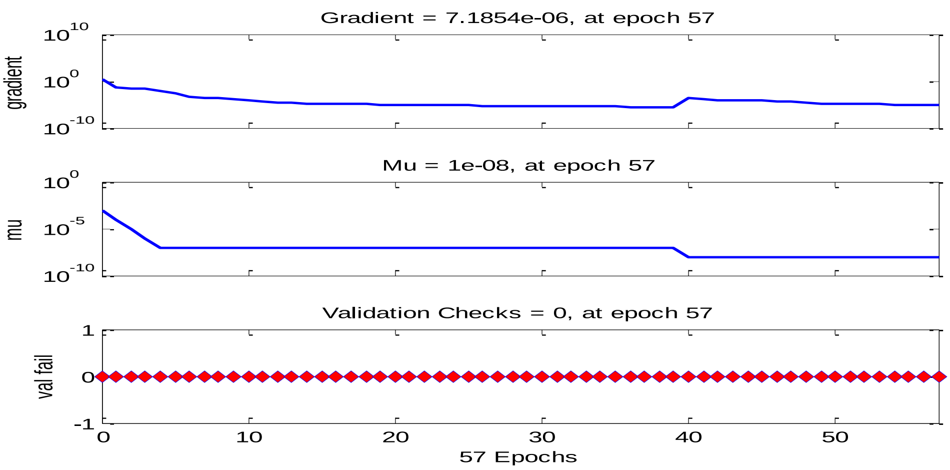

Figure (5): It shows the states that the network goes through when training the neural network

6. Conclusions and Recommendations :

Through the results of the theoretical side, the following conclusions and recommendations were reached:

6.1.Conclusions

After experimenting with the Euphrates River water pollution data, the two researchers concluded the following:

1. When using the factorial analysis (principal components analysis), we found that the results of the analysis after recycling were better than before the Rotation, as the results of the important variables after recycling were as follows: The variables according to their importance (AIK, Mg, TH, PH, Na, SO_4, Ca, K) in the first Factor, with a variance rate for this Factor (37.726%), the variables according to their importance (EC, CL, TDS) in the second Factor, with a variance rate for this Factor (31.564%), the variables according to their importance (NO_3, Temp, PO_4) in the third Factor, with a variance rate for this Factor (11.352%), we also note that the variable (DO2) did not appear, which means this variable is unimportant.
2. When using neural networks (Backpropagation algorithm), we found that there are variables that have errors greater than the general average of errors, which means that these variables have no effect on water pollution in the Euphrates River and these variables are (TH, TDS, EC).
3. When comparing the results of the statistical method using the principal component analysis with the results of the Backpropagation algorithm, we note the superiority of the algorithm in determining the variables affecting the pollution of the Euphrates River. In contrast, when using the statistical method, we found that the variable (DO2) is not significant and does not affect the pollution of the Euphrates River either. When Using the algorithm, we found that all these variables (TH, TDS, EC) do not affect the water pollution of the Euphrates River.

6.2.Recommendations

From our conclusions, the following recommendations can be summarized:

1. Conducting more studies to determine the most important variables of water pollution in Iraq and their treatment and to determine in which governorates the percentage of pollution is greater, to be addressed by the competent departments.
2. The broader use of artificial intelligence algorithms in the environmental aspect. Examples of proposed algorithms include the hill-climbing algorithm, the ant colony algorithm, and the swarm algorithm.
3. Conduct a study to determine the variables affecting water pollution according to the governorates through which the Euphrates River passes, determining the percentage of pollution. Is it more than the permissible limit?
4. Our recommendations to the Iraqi Ministry of Health and Environment are: We recommend conducting environmental studies to measure environmental pollution in various departments (water pollution, soil pollution, chemical pollution, air pollution), and care must be taken to conduct daily measurements of the percentage of pollution in the water and all stations.

References

[1] Zhao, K., He, T., Wu, S., Wang, S., Dai, B., Yang, Q. and Lei, Y., 2019. "Application Research of Artificial Neural Network in Environmental Quality Monitoring". International Journal of Pattern Recognition and Artificial Intelligence, 33(12), p.1959039.

[2] Lungu, E., Oprea, M. and Dunea, D., 2008, "An application of artificial neural networks in environmental pollution forecasting". In Proceedings of the 26th IASTED International Conference on Artificial Intelligence and Applications, Acta Press, Innsbruck, Austria, February (pp. 11-13).

[3] Jiang, X., Zhang, P. and Huang, J., 2022. "Prediction method of environmental pollution in smart city based on neural network technology". Sustainable Computing: Informatics and Systems, Vol. 36, p.100799.

[4] Hemavani, B. and Rao, G.S., 2020. “A study on air pollution trends in Sanathnagar, Hyderabad using principle component and cluster analysis”. International Journal of Science and Research, Vol. 9 Issue 2.

[5] Nasir, M.F.M., Samsudin, M.S., Mohamad, I., Awaluddin, M.R.A., Mansor, M.A., Juahir, H. and Ramli, N., 2011. “River water quality modeling using combined principle component analysis (PCA) and multiple linear regressions (MLR): a case study at Klang River, Malaysia”. World Applied Sciences Journal, 14, pp.73-82.

[6] Vidal, R., Ma, Y. and Sastry, S., 2005. “Generalized principal component analysis (GPCA)”, IEEE transactions on pattern analysis and machine intelligence, 27(12), pp.1945-1959.

[7] Jolliffe, I., 2005. Principal component analysis. Encyclopedia of statistics in behavioral science., 1st edition, Wiley online Library.

[8] Jolliffe, I.T. and Cadima, J., 2016. “Principal component analysis: a review and recent developments. Philosophical transactions of the royal society A: Mathematical”, Physical and Engineering Sciences, 374(2065), p.20150202.

[9] Cilimkovic, M., 2015. “Neural networks and Back propagation algorithm. Institute of Technology Blanchardstown, Blanchardstown Road North Dublin, 15(1).

[10] Skorpil, V. and Stastny, J., 2006. Neural networks and Back propagation algorithm. Electron Bulg Sozopol, 20(2).

[11] Li, J., Cheng, J.H., Shi, J.Y. and Huang, F., 2012. “Brief introduction of Back propagation (BP) neural network algorithm and its improvement”. In Advances in Computer Science and Information Engineering: Volume 2 (pp. 553-558). Springer Berlin Heidelberg.

Journal of AL-Rafidain University College for Sciences (2023); Issue 54 EISSN (2790-2293) PISSN (1681-6870)

مجلة كلية الرافدين الجامعة للعلوم (2023)؛ العدد 54؛ 518 - 530

AL- Rafidain University College

PISSN: (1681-6870); EISSN: (2790-2293)
مجلة كلية الرافدين الجامعة للعلوم

Available online at: https://www.jrucs.iq

JRUCS
Journal of AL-Rafidain University College for Sciences

استخدام الشبكات العصبية (خوارزمية الانبعاث الخلفي) لتحديد المتغيرات المؤثرة في تلوث مياه نهر الفرات

أنــس ايدن موسى	صباح منفي رضا
ons.edin1001a@coadec.uobaghdad.edu.iq	drsabah@coadec.uobaghdad.edu.iq
كلية التربية البدنية وعلوم الرياضة - الجامعة المستنصرية، بغداد، العراق	قسم الاحصاء - كلية الادارة والاقتصاد - جامعة بغداد، بغداد، العراق

المستخلص

مع النمو الكبير للاقتصاد والتطور السريع للتكنولوجيا الصناعية الحديثة، زادت المشاكل البيئية، حيث يعد تلوث المياه من المشاكل المعقدة التي تعاني منها معظم دول العالم وخاصة الدول النامية لاحتوائها على ملوثات مختلفة بعضها يمكن أن يتحلل كملوثات عضوية وبعضها يصعب تحللها، مثل المعادن الثقيلة السامة... في العراق وبحسب إحصائيات وزارة البيئة العراقية - قسم تلوث المياه لعام 2018 وتحديداً مياه نهر الفرات ، تم تحديد مجموعة من المتغيرات التي تؤثر على جودة مياه هذا النهر، كما ساهمت ندرة المياه وتراكم المخلفات المختلفة في شبكات المياه وعلى ضفاف نهر الفرات بشكل كبير في زيادة مشكلة التلوث. أدى استمرار سوء نوعية المياه إلى مخاوف صحية وبيئية شديدة.

يهدف البحث إلى إلقاء الضوء على أهم المتغيرات المؤثرة على تلوث مياه نهر الفرات ونوعية وجودة المياه في هذا النهر. استخدمنا في بحثنا طريقة المركبات الرئيسية وخوارزمية الشبكة العصبية (خوارزمية الانبعاث العكسي) لتحديد المتغيرات التي تؤثر على تلوث مياه نهر الفرات. تم تطبيقها على بيانات حقيقية من وزارة البيئة – شعبة تلوث المياه – تلوث مياه نهر الفرات.

من أهم الاستنتاجات التي توصلنا إليها عند استخدام الأساليب الإحصائية أن المتغير (DO2) والذي يمثل الاوكسجين المذاب في المياه ليس مهمًا ولا تؤثر نسبته على تلوث مياه نهر الفرات، اما عند استخدام خوارزمية الذكاء الاصطناعي توصلنا إلى أن هذه المتغيرات (TH، TDS، EC) والتي تمثل (العسرة الكلية، الاملاح الصلبة الذائبة الكلية والتوصيلية على التوالي) لا يؤثرون على تلوث مياه نهر الفرات.

معلومات البحث

تواريخ البحث:
تاريخ تقديم البحث: 2022/12/24
تاريخ قبول البحث: 2023/3/3
تاريخ رفع البحث على الموقع: 2023/12/31

الكلمات المفتاحية:
تحليل المركبات الرئيسية، الشبكات العصبية، خوارزمية الانبعاث الخلفي، نهر الفرات، تلوث المياه

للمراسلة:
انــس ايدن موسى
ons.edin1001a@coadec.uobaghdad.edu.iq

https://doi.org/10.55562/jrucs.v54i1.619