Determine Rate of Contamination in Tigris River in River Baghdad city by Artificial Intelligence

Farah F. Ghazi[1], Luma N. M. Tawfiq[2*]

[1,2] Department of Mathematics, College of Education for Pure Science Ibn Al-Haitham, University of Baghdad, Iraq

*Corresponding author's email: luma.n.m@ihcoedu.uobaghdad.edu.iq

Abstract

In this article, we used type of artificial intelligence that is suggest efficient design of artificial neural network ANN of type Feed Forward Neural Network (FFNN) based on new LM training algorithm. Then we used to determine the rate of contamination in Tigris River in Baghdad city. Architectural of design consist 4 layers: input layer contains 8 nodes represent: (PH, Total dissolve solids (T.D.S) Electrical Connection (EC), Total suspended solids (T.S.S), Chemical oxygen demanding (COD), Oil, Nephelometric Turbidity Unit) NTU (, Nitrate (NO)), 1st hidden layer contain 17 nodes, 2nd hidden layer contain 8 nodes with tansig. transfer function for each hidden layers and output layer contains 5 nodes for which are (Cd, Mg, pb, Fe, Cr), with linear transfer function. We used 100 sample distributed as 65 sample for training FFNN, 20 sample for testing and 15 sample for validation and the proposed FFNN achieved high predictive accuracy, with a minimum mean square error (MSE) of 2.47×10^{-5} and a final performance gradient of 1.57×10^{-5}, indicating excellent convergence. The model successfully predicted the concentrations of heavy metals with an error less than 0.0012, which is considered scientifically acceptable for environmental assessments. Comparison between laboratory results and ANN outputs showed strong agreement, confirming the reliability of the model.

Keywords: Tigris River, Contamination, Heavy metals, Artificial Intelligence, Artificial neural networks.

تحديد معدل التلوث في نهر دجلة داخل مدينة بغداد باستخدام الذكاء الاصطناعي

فرح فيصل غازي[1]، لمى ناجي محمد توفيق*[2]

[1،2] قسم الرياضيات، كلية التربية للعلوم الصرفة/ابن الهيثم، جامعة بغداد، العراق

*البريد الالكتروني للمؤلف المراسل: luma.n.m@ihcoedu.uobaghdad.edu.iq

الملخص

في هذا البحث، تم استخدام نوع من تقنيات الذكاء الاصطناعي يتمثل في تصميم كفء للشبكة العصبية الاصطناعية (ANN)من نوع الشبكة العصبية المتقدمة ذات الانتشار الأمامي - Feed Forward Neural Network) (FFNN)اعتمادًا على خوارزمية تدريب جديدة من نوع .LMثم تم توظيف هذا النموذج لتحديد معدل التلوث في نهر دجلة ضمن مدينة بغداد. يتألف التصميم المعماري للشبكة من أربع طبقات: الطبقة الإدخالية تحتوي على 8 عقد تمثل المتغيرات التالية :الأس الهيدروجيني(pH) ، المواد الصلبة الذائبة الكلية(T.D.S) ، التوصيلية الكهربائية (EC)، المواد الصلبة العالقة الكلية(T.S.S) ، الطلب الكيميائي للأوكسجين(COD) ، الزيت، وحدة العكارة النيفيلومترية(NTU) ، والنترات(NO_3). أما الطبقة المخفية الأولى فتحتوي على 17عقدة، والطبقة المخفية الثانية على 8عقد، مع استخدام دالة النقل tansigفي كل طبقة مخفية، في حين تحتوي طبقة الإخراج على 5عقد تمثل العناصر الثقيلة :الكادميوم(Cd) ، المغنيسيوم(Mg) ، الرصاص(Pb) ، الحديد(Fe) ، والكروم(Cr) ، وذلك باستخدام دالة نقل خطية. تم استخدام 100عينة موزعة كما يلي 65 :عينة للتدريب، 20عينة للاختبار، و 15عينة للتحقق من صحة النموذج(validation) . و أظهرت نتائج التدريب كفاءة عالية للنموذج، حيث بلغ متوسط مربع الخطأ (MSE) الأدنى 2.47×10^{-5}، كما وصل منحدر الأداء النهائي إلى 1.57×10^{-5} مما يدل على سرعة التقارب ودقته. وأثبت النموذج قدرته على تقدير تراكيز المعادن الثقيلة بخطأ يقل عن 0.0012، وهو ضمن الحدود المقبولة علمياً وبيئياً. كما أظهرت المقارنة بين النتائج المخبرية ومخرجات النموذج توافقاً كبيراً، مما يؤكد موثوقية الشبكة العصبية في التنبؤ بمستوى التلوث.

الكلمات المفتاحية : نهر دجلة، التلوث، العناصر الثقيلة، الذكاء الاصطناعي، الشبكات العصبية الاصطناعية.

Introduction

Contamination is one of the most serious environmental problems that human face and it is considered one of the problems of the age due to its danger to human, other living organisms and the elements of the environment (Ghazi, 2020).

Water is affected by many contaminates, the most important of these contaminates of which are: pesticides and chemicals from agricultural lands, radioactive materials and oil leaks, toxic organic materials, inorganic chemicals, disease-causing organisms, factory waste and other contaminants (Isaev & Mikhailova, 2009).

The Tigris River is one of the seven major sources of potable water in Iraq. The length of the Tigris River is about 1718 km. It originates from the Taurus Mountains in eastern Turkey and ends in the Shatt al-Arab. The length of the river in Iraq is 1,400 km. In the Baghdad region specifically, it is about 50 km. It divides the Baghdad region into two parts, Karkh. and Rusafa (Isaev & Mikhailova, 2009) (Al-Ani, et al 2014).

In recent years, there has been an increased interest in heavy metals (HMs) contamination due to its toxicity, and it is considered among the inorganic contaminate. Therefore, many researchers dealt with the issue of environmental contamination in their research to reduce contamination because of its impact on humans and living organisms, and try to find appropriate solutions to reduce this contamination, and among these researches (Ghazi, 2020)., (Allawi, et al 2022) (Huynh, et al 2022) (Abdulkareem, et al 2022) (Ismael, et al 2020) (Ibrahim, et al 2020) (Tawfiq & Ghazi, 2015) (Ghazi, 2018).

The study (Jasim, et al 2025) focused on the Abu Gharq water project, the Al-Tayara water project, and the unified Hilla water project in the city of Hilla, and the contamination with heavy elements aluminum, copper, and zinc.

Machine learning techniques (ANN, GEP, M5P, RF, PINN) was used (Al-layla & Al-Ogaidi 2025) for allowed them to represent field data and handle soil complexities with greater accuracy compared to traditional models. Also, they provided greater predictive accuracy and flexibility in dealing with nonlinearity and different soil types.

In this paper, we suggest an efficient approach to estimate the rate of contamination by heavy metals of the Tigris River water by designing a neural network to estimate the heavy metal river contamination in Baghdad using MATLAB version 2022a.

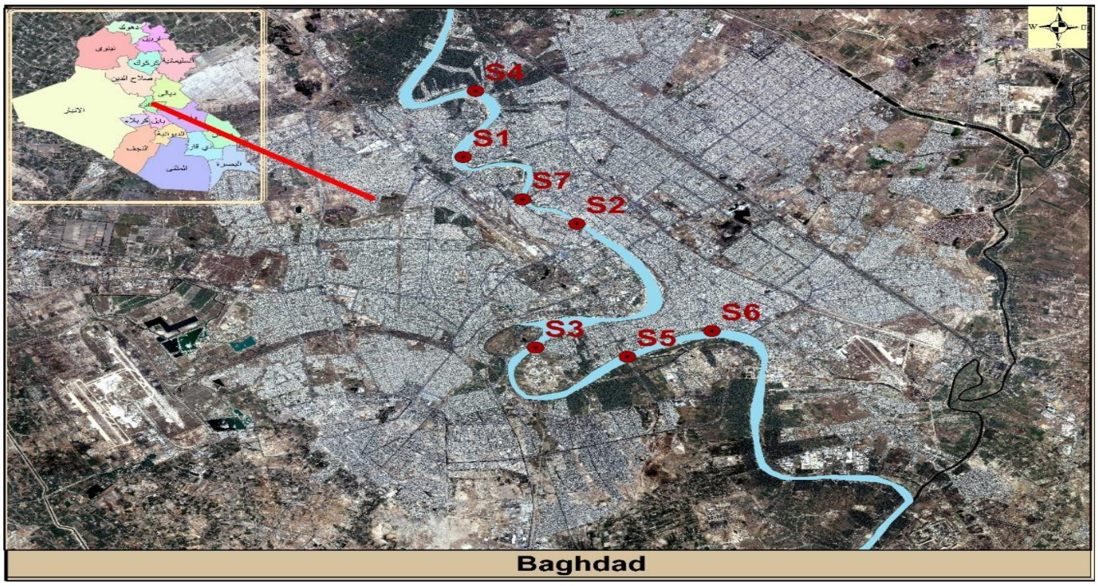
1. Research Methodology

1.1. Samples

The capital of Iraq is Baghdad City it was (44°31‘-44°17'E, 33°14'-33°25'N), the climate with dry hot in summers and cold winters; the mean rainfall is about 151.8 mm (Tawfiq & Ghazi, 2015).

The samples were collected from seven regions of Baghdad, from Ś1: Al-Adhamiya, Ś2: Al-Shuhada, Ś3: Al-Jadriya, Ś4: East Tigris Refining Unit (It is located in the Al-Krayat region on the Rusafa side), Ś5: Dora station (It is located on the Karkh side), Ś6: Al-Wahda station (It is located on the Rusafa side in the Karrada), Ś7: Al-Wathba Unit (It is located in the center of Baghdad, on the Rusafa side). The results in the regions Ś4, Ś5, Ś6 & Ś7 of the tests show the raw water withdrawn from the river into the stations, which is separated only without chemical additives, it meaning represents the specifications of the river water see Table 1.

The collection samples were taken in Tigris River from surface of river about (20-30) cm below the surface then the samples put in the ice-cool box to present the laboratory and Metals determinations were done by Inductively coupled Plasma (ICP) (The Iraqi Ministry of Health and Environment). See map below explain the samples by used (GIS) Arcmap 10.9 program.



Map showing the location of samples.

Table 1: The results of heavy metals for different regions.

	المحددات البيئية وفق قانون 25 لسنة 1967	Ś1	Ś2	Ś3	Ś4	Ś5	Ś6	Ś7
PH	6.5-8.5	7.3	7.4	7.3	7.1	7.2	7.4	7.3
TDS (ppm)	-	330	314	305	327	304	300	305
Ec	-	530	490	474	546	505	464	507
TSS (ppm)	-	388	310	200	388	310	94	78
COD (ppm)	-	20	40	20	20	40	20	40
Oil (ppm)	-	11	20.2	14.6	-	-	-	-
NTU (ppm)	-	415	370	250	418	339	405	416
NO3 (ppm)	15	1	1.5	3.1	0.96	1.6	3.36	3.6
Output layer which consisted the continuation of heavy metals in Tigris River in different location of Baghdad city								
Cd (ppm)	0.005	0.048	0.027	0.032	0.01415	0.01271	0.0124	0.01069
Mg (ppm)	0.1	34.54	33.22	31.416	34.095	33.132	32.316	31.787
Pb (ppm)	0.05	0.01078	00.0279	0.01078	0.01078	0	0.0279	0.0285
Fe (ppm)	0.3	0.0222	0.0142	0.0267	0.0222	0.0142	0.0267	0.0195
Cr (ppm)	0.05	0.039	0.024	0.025	0.01597	0.01184	0.01195	0.0124

1.2. Artificial neural networks (ANNs):

Artificial neural networks are very similar to the human brain in processing data and uses fully interconnected nodes or neurons in a layered structure similar to a neuron. Artificial neural networks began from 1940s it was used in many applications in engineering and science and used to solve complex problems, such as summarizing documents or recognizing faces, with greater accuracy.

Networks especially FFNN can be help computers make intelligent decisions with limited human assistance. This is because of its ability to learn and model the relationships between complex and nonlinear input-output data (Yadav, 2015).

In this paper, we suggest artificial neural network (ANN) of type Feed Forward Neural Network (FFNN) supervised learning with Levenberg-Marquardt learning algorithm (LM) used to approximate the concentration of HMs in Tigris River in Baghdad city.

In this article we proposed a multilayer FFNN which consist: input layer consist 8 nodes which are: (PH, Total dissolve solids (T.D.S) Electrical Connection (EC), Total suspended solids (T.S.S), Chemical oxygen demanding (COD), Oil, Nephelometric Turbidity Unit) NTU(, Nitrate (NO)), Two hidden layers first layer consist of 17 nodes and the second consist of 8 nodes with tansig. activation function (transfer function) and five nodes for output layer which are (Cd, Mg, pb, Fe, Cr) with linear Transfer function. See Figure 1.

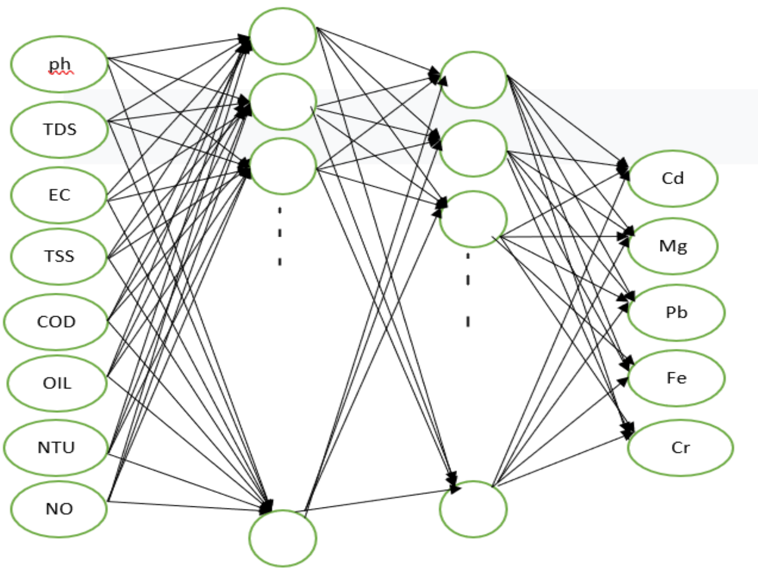


Figure1: Architecture of suggestion feed forward neural network.

1.3. Mathematical Model of ANN:

A neuron x_i considered an input layer which x_i = {PH, TDS, EC, TSS, COD, Oil, NTU, NO} where i= 1, 2, …, 8 nodes in input layer each one of those nodes connected with the next layer by weight factor W_{ik} for k=1,2,3, …,17 and b_0 is the bias. Activation function used tansig function

$$S_i = f\left[\mathrm{b}_0 + \sum_{k=1}^{17} \mathrm{W}_{\mathrm{ik}}\, x_i\right] \quad (1)$$

Where $f = \frac{2}{e^{-2x_i+1}} - 1$ (2)

So, in hidden layer, we suggest two layers first content 17 nodes and the second 8 nodes with tansig function

$$q_k = h\left[\sum_{j=1}^{8}\left[f\left[\mathrm{b}_0 + \sum_{k=1}^{17} \mathrm{W}_{\mathrm{ik}}\, x_i\right]\right]_j v_{kj} + b1\right] = h\left[\sum_{j=1}^{8}[S_i]_j v_{kj} + b1\right] \quad (3)$$

where $h = f$ in eq (2)

the output layer consists of five nodes with linear function which consider of the contamination by heavy metals for Tigris River in Baghdad city like (Cd, Mg, pb, Fe, Cr)

$$o_l = T\left[\sum_{l=1}^{5}\left[h\left[\sum_{j=1}^{8}\left[f\left[\mathrm{b}_0 + \sum_{k=1}^{17} \mathrm{W}_{\mathrm{ik}}\, x_i\right]\right]_j v_{kj} + b1\right]\right]_l r_{jl} + b2\right]$$

$$= T\left[\sum_{l=1}^{5} [q_k]_l r_{jl} + b2\right] = \sum_{l=1}^{5} [q_k]_l r_{jl} + b2 \quad (4)$$

Where $l = 1, 2, \ldots, 5$.

1.4. LM Training Algorithm:

Levenberg-Marquardt method is approach used to training ANNs with 2^{nd} order training speed without having to calculate the Hessian matrix, when the performance function (energy function) has the form of a sum of squares, then the Hessian matrix can be approximated as $H=J^TJ$ and the gradient can be computed as g= $J^T E(w_k)$, where J is the Jacobian matrix, which contains first derivatives of the error function (performance function) with respect to the weights, and $E(w_k)$, is a vector of network errors. The LM method use the following approximation to the Hessian matrix:

$$H \approx J^T J + \mu I \tag{5}$$

μ: is positive always, called the combination of coefficient,

I: identity matrix.

So, we used that method to update the weight to find the appropriate and efficient design that describes the problem in all its aspects:

$$w_{k+1} = w_k - (J^T J + \mu I)^{-1} J_k^T e \tag{6}$$

To motivate this, recall that LM algorithm can be interpreted as approximating $\mathrm{E(x + p)}$ by a quadratic form:

$$\mathrm{E(x + p) \approx E(x) + p^T g(x) + \frac{1}{2} p^T (J^T J + \mu I) p} \tag{7}$$

An alternative view is to minimize the quadratic as a function of p. If $J^T J + \mu I$ is positive definite, then the minimum is obtained by setting the derivative equal to zero. If $J^T J + \mu I$ is indefinite, then the quadratic function does not have a finite minimum. Additionally, it has a streamlined implementation in MATLAB software, as the matrix equation solution is built-in. These attributes make it particularly effective in a MATLAB environment (Mark et al 2017) (Marquardt, 1963) (Tawfiq & Hassan 2018) (Oraibi, 2013) (Mohammed & Hasan 2019).

Resalt & Desiccation:

The performance of the results of network (FFNN) is measured by the computational in training, testing and validation case which equal to perf = 2.466568e-05.

The mean square error (MSE) will be computed to check the accuracy of the approximation of resultss that obtained in these cases for different values of the epochs, presented in Figure 2. Table 2 and 3, The accuracy of the training for epoch and time. Moreover, the target of output in train, test and validation case shown in Figure 3. Also,

Figure 4 represent the behavior of gradient in validation case at epoch 100000 which stopped in epoch 14 because it reached to minimum values.

Finally, when comparing the laboratory valued for heavy metals for Tigris River with the value of output for suggestion neural network, we see a small error shown by Figure 5.

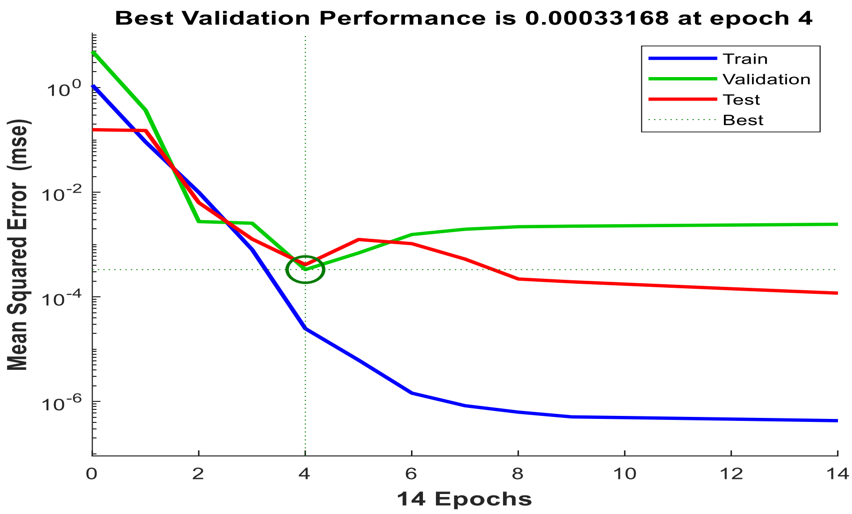


Figure 2: Comparison between Mean square error for training, testing and validation.

Table 2: The accuracy of the train for epoch and time.

Algorithms	**Levenberg-Marquardt**		
Training results	**Reached minimum gradient**		
Unit	**Value of Initial**	**Value of Stopped**	**Value of Target**
Epoch	0	14	1000
Elapsed time	-	00:00:00	-
performance	1.12	4.3 e-7	0
Gradient	2.51	1.57e-05	1e-20
Mu	0.001	1e-05	1e+308
Validation checks	0	10	10

Table 3: Mean squer error for training, volidation and testing.

Type	**Target values**	**MSE**
Training	65	2.466568e-05
Validation	15	3.31682267e-04
Testing	20	4.118766753e-04

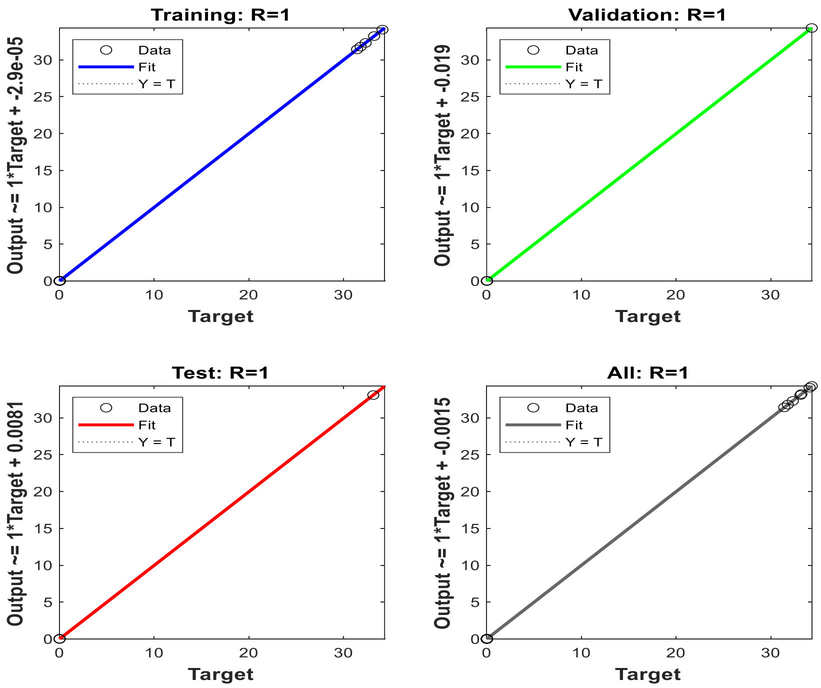


Figure 3: Comparison between the target of training, testing and validation.

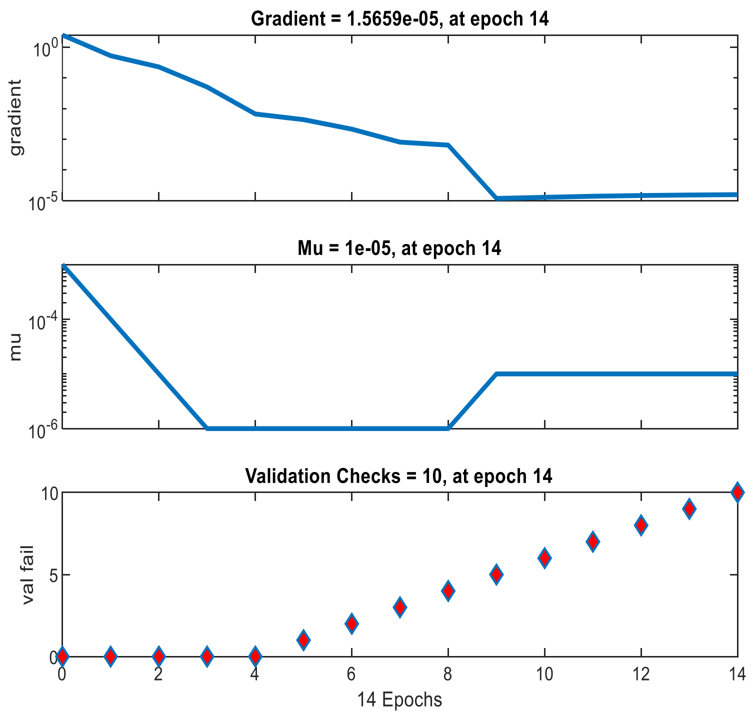


Figure 4: Validation case: behavior of gradient depending on epoch.

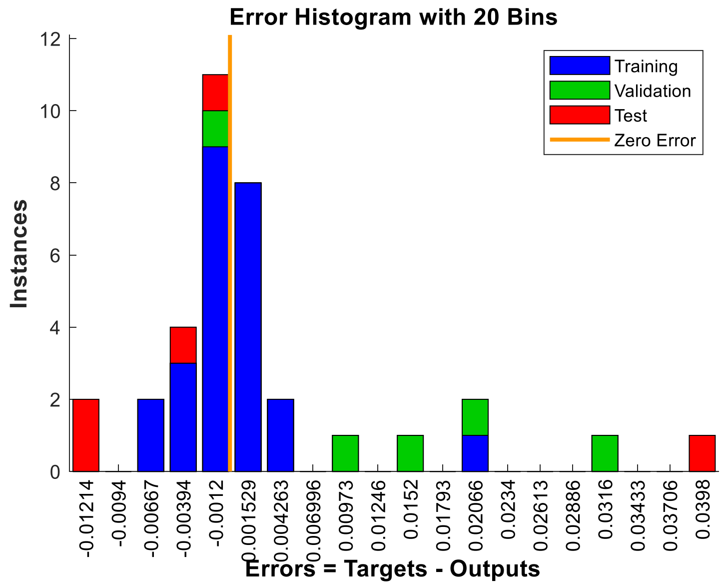


Figure 5: The error between laboratory values and the neural network values.

Conclusions

Used neural network for consider the contamination of Tigris River in Baghdad city by used the Chemical and physical properties of the river as input and the output represent some of heavy metal like (Cd, Mg, pb, Fe, Cr), The error between laboratory values and the neural network values is too small about less than (0.0012) which represent a good tool to estimate the heavy metals in river.

We conclude that artificial intelligence is a powerful technique can be used to estimate the rate of Contamination or can be used for any estimation with high accuracy and easy implementation

References

Abdulkareem, I. A., Abbas, A. A., & Dawood, A. S. (2022). Modeling pollution index using artificial neural network and multiple linear regression coupled with genetic algorithm. *Journal of Ecological Engineering, 23*(3), 236–250.

Al-Ani, R. R., Abdul Hameed, M., Al-Obaidy, J., & Badri, R. M. (2014). Assessment of Tigris water quality in the Baghdad–Iraq. *International Journal of Advanced Research, 2*(5), 1125–1131.

Al-layla, S. Z. M., Al-Ogaidi, A. A. M. (2025). *A Comprehensive and Extensive Review of the Process of Water Infiltration in Soil.* Journal of Water Resources and Geosciences. 4(2), 109 – 158. https://jwrg.gov.iq/index.php/jwrg/article/view/138/82

Allawi, M. F., Salih, S. Q., Kassim, M., Ramal, M. M., Mohammed, A. S., & Yaseen, Z. M. (2022). Application of computational model based probabilistic neural network for surface water quality prediction. *Mathematics, 10*(3960). https://doi.org/10.3390/math10213960.

Ghazi, F. F. (2018). Estimation of heavy metals contamination in soil of Zaafaraniya City using the neural network. *Journal of Physics: Conference Series, 1003*(1), 012058.

Ghazi, F. F. (2020). Modeling the contamination of soil adjacent to Mohammed ALQassim Highway in Baghdad. *Iraqi Journal of Science, 61*(10), 2663–2670. https://doi.org/10.24996/ijs.2020.61.10.23 .

Huynh, T.-M.-T., Ni, C.-F., Su, Y.-S., Nguyen, V.-C.-N., Lee, I.-H., Lin, C.-P., & Nguyen, H.-H. (2022). Predicting heavy metal concentrations in shallow aquifer systems based on low-cost physiochemical parameters using machine learning techniques. *International Journal of Environmental Research and Public Health, 19*(12180). https://doi.org/10.3390/ijerph1912180 .

Ibrahim, M. A., Mohammed-Ridha, M. J., Hussein, H. A., & Faisal, A. A. H. (2020). Artificial neural network modeling of the water quality index for the Euphrates River in Iraq. *Iraqi Journal of Agricultural Sciences, 51*(6), 1572–1580.

Isaev, V. A., & Mikhailova, M. V. (2009). The hydrology, evolution, and hydrological regime of the mouth area of the Shatt al-Arab River. *Water Resources, 36*(4), 380–395.

Ismael, A. N., Abed, H. A., & Abed, M. A. (2020). Classification of groundwater quality using artificial neural networks in Safwan and Al-Zubayr in Basra. *Advances in Computer, Signals and Systems, 4*(1), 25–35. Clausius Scientific Press. https://doi.org/10.23977/acss.2020.040105 .

Jasim, H. K., Tayyeh, H. K., Hussein, A. M. (2025). *Studying the Effects of Heavy Metals Concentrations on selected Water Treatment Plants in Babylon Governorate*. Journal of Water Resources and Geosciences. 4(1), 28 – 44. https://jwrg.gov.iq/index.php/jwrg/article/view/116/62

Mark, H. B., Martin, T. H., & Howard, B. D. (2017). *Neural Network Toolbox™ user's guide R2017b.* The MathWorks, Inc.

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics, 11*(2), 431–441.

Mohammed, M. N., & Tawfiq, L. N. M. (2010). The effect of number of training samples for artificial neural network. *Ibn Al-Haitham Journal for Pure and Applied Science, 23*(3), 1–7.

Oraibi, Y. A. (2013). Fast training algorithms for feed forward neural networks. *Ibn Al-Haitham Journal for Pure and Applied Science, 26*(1), 275–280.

Tawfiq, L. N. M., & Ghazi, F. F. (2015). Using artificial neural network technique for the estimation of Cd concentration in contaminated soils. *International Journal of Innovations in Scientific Engineering, 1*(1), 58–68.

Tawfiq, L. N. M., & Hassan, M. A. (2018). Estimate the effect of rainwaters in contaminated soil by using Simulink technique. *Journal of Physics: Conference Series, 1003*(1), 012057.

Tawfiq, L. N. M., & Hasan, M. A. (2019). Evaluate the rate of pollution in soil using Simulink environment. *Ibn Al-Haitham Journal for Pure and Applied Science, 32*(1), 132–138.

Tawfiq, L. N. M., & Khamas, A. H. (2023). New approach for calculate exponential integral function. *Iraqi Journal of Science, 64*(8), 4034–4042. https://doi.org/10.24996/ijs.2023.64.8.27 .

The Iraqi Ministry of Health and Environment. (n.d.). *Results of laboratory tests performed by the Iraqi Ministry of Health and Environment for the Tigris River.*

Yadav, N., Yadav, A., & Kumar, M. (2015). *An introduction to neural network methods for differential equations.* Springer. https://doi.org/10.1007/978-94-017-9816-7 .