Environmental Science

AIMS Environmental Science, 11(4): 649–664.
DOI: 10.3934/environsci.2024032
Received: 02 July 2024
Revised: 05 August 2024
Accepted: 07 August 2024
Published: 19 August 2024

https://www.aimspress.com/journal/environmental

Research article

Monitoring river water quality through predictive modeling using artificial neural networks backpropagation

Muhammad Andang Novianta[1,2], Syafrudin[1,3,*], Budi Warsito[1,4], Siti Rachmawati[5]

[1] Department of Doctoral Environmental Science, Faculty of Postgraduate, Universitas Diponegoro, Semarang 50275, Indonesia
[2] Department of Electrical Engineering, Faculty of Engineering, Universitas AKPRIND Indonesia, Yogyakarta 55222, Indonesia
[3] Department of Environmental Engineering, Faculty of Engineering, Universitas Diponegoro, Semarang 50275, Indonesia
[4] Department of Statistics, Faculty of Science and Mathematics, Universitas Diponegoro, Semarang 50275, Indonesia
[5] Department of Environmental Science, Faculty of Mathematics and Natural Sciences, Universitas Sebelas Maret, Surakarta 57126, Indonesia

* **Correspondence:** Email: syafrudin@lecturer.undip.ac.id; Tel: +628122849936.

Abstract: Predicting river water quality in the Special Region of Yogyakarta (DIY) is crucial. In this research, we modeled a river water quality prediction system using the artificial neural network (ANN) backpropagation method. Backpropagation is one of the developments of the multilayer perceptron (MLP) network, which can reduce the level of prediction error by adjusting the weights based on the difference in output and the desired target. Water quality parameters included biochemical oxygen demand (BOD), chemical oxygen demand (COD), total suspended solids (TSS), dissolved oxygen (DO), total phosphate, fecal coliforms, and total coliforms. The research object was the upstream, downstream, and middle parts of the Oya River. The data source was secondary data from the DIY Environment and Forestry Service. Data were in the form of time series data for 2013–2023. Descriptive data results showed that the water quality of the Oya River in 2020–2023 was better than in previous years. However, increasing community and industrial activities can reduce water quality. This was concluded based on the prediction results of the ANN backpropagation method with a hidden layer number of 4. The prediction results for period 3 in 2023 and period 1 in 2024 are that 1) the concentrations of BOD, fecal coli, and total coli will increase and exceed quality standards, 2) COD and TSS concentrations will increase but will still be below quality standards, 3) DO and total phosphate concentrations will remain constant and still on the threshold of quality standards. The possibility of several water quality parameters increasing above the quality standards remains, so the potential for contamination of the Oya River is still high. Therefore, early prevention of river water pollution is necessary.

Keywords: artificial neural network (ANN) backpropagation; system modeling; prediction

1. Introduction

One of the current damages to water resources is river pollution. Direct input of pollutants into rivers causes changes in physical, chemical, and biological factors, resulting in a decrease in river water quality and a low capacity to carry river pollution loads [1]. Many previous researchers have studied river pollution cases, including [2], who grouped rivers in several regions in the southeastern part of Central Java Province. In Banjarmasin, water quality research was also carried out [3], using K-means clustering. In the DIY province itself, research has also been carried out on river grouping based on water quality parameters using the K-means clustering method [4]. In China, research was conducted on the water quality of the Yangtze River [5]. Furthermore, groundwater quality using a combination of fuzzy techniques and K-means clustering has been reported by Mohammadrezapour et al. [6]. This proves that many researchers care about the environment, especially the cleanliness of river waters. River water is a source of life for most living creatures, and rivers can provide water intake for the surrounding environment.

Several rivers in the DIY region have several ecological functions. Variations in land use cause water quality to differ due to the discharge of waste directly into rivers, causing a decrease in river water quality. This will hurt rivers as one of the water bodies that receive wastewater discharge, pollution, and contamination originating from point sources and non-point sources, thereby reducing the ability of rivers to recover naturally (self-purification). The natural ability of river water against pollution needs to be maintained to minimize its decline in quality [7].

The analysis of pollutant load-carrying capacity is a complicated process due to continuous river flow and varying river water quality from upstream to downstream. The current stages of the Industrial Revolution 4.0 are based on information technology, such as the use of artificial intelligence (AI) and automation, which opens up wide opportunities for technological innovation to support efforts to reduce the risk of pollution through intelligent risk reduction systems [8].

Community and industrial activities are very dependent on water quality, especially river water. Therefore, those require water quality analysis in the form of predictions. Such prediction will be able to provide information on water quality levels in the future so that early treatment can be carried out. A simplified mathematical formulation for water quality index has been developed (WQI), using the weighted arithmetic index method [9].

Several statistical models can be used to predict future data, including the autoregressive integrated moving average (ARIMA) method, Fuzzy time series, and regression analysis. The ARIMA method is a combination of smoothing methods, regression analysis, and decomposition methods. It makes full use of past and present data to produce accurate predictions and is used when the data pattern forms a linear pattern or follows a straight-line pattern. However, this method will only work well if the data in the time series used are dependent or statistically related to each other. Another weakness of this method is that the prediction results will be poor when the data has a nonlinear pattern. Nonlinear patterns are data patterns that do not form a straight line of linear equations. This pattern is often found. Meanwhile, the machine learning (ML) model of Shapley Additive Explanations (SHAP) has been introduced in [10]. The authors stated that the SHAP was an invaluable addition to ML-based streamflow predictions and early warning systems, offering human-comprehensible interpretations. Another modified generative adversarial network with explainable AI was reported in [11]. The model was used to model the streamflow in ungauged basins with sparse data. The approach looks promising as it worked well with sparse data from an ungauged basin.

One prediction method that can be used to overcome nonlinear patterns is a neural network or artificial neural network (ANN). An artificial neural network (ANN) algorithm can facilitate more accurate prediction models with many input variables [12], which are used as input to form the network structure. This research uses six input variables and six hidden layers. This design can produce 92.79% accuracy, which shows that the resulting predictions are close to the actual data.

Many types of neural network architecture can be used, including single-layer network architecture, multilayer network, and competitive layer network. Backpropagation is one of the developments of the multilayer perceptron (MLP) network, which can reduce the level of prediction error by adjusting the weights based on the difference in output and the desired target.

The ANN backpropagation method plays a role in making complex problems simpler [13]. The thinking process is directed at decision-making with bounded rationality, the process of simplifying the model by taking the most essential core problems without involving all concrete problems. The ANN backpropagation method has been used to model an automatic temperature control system on a smart poultry farm [14]. The results showed that this method is better than ANN modeling with a single hidden layer. ANN backpropagation has the advantage of being able to approach the target output value.

Haekal and Wibowo's research applied the ANN backpropagation method to predict the water quality classification of the Ciliwung River [15]. The ANN backpropagation method performed better than the Naïve Bayes and SVM because of its high level of accuracy, namely 94.6%. Mustafa, Mustapha, Hayder, and Salisu used ANN backpropagation in the application of IoT and artificial intelligence in monitoring and predicting river water quality [16]. The results also prove that the method can evaluate historical data collected from various river stations and wastewater treatment plants with minimum errors in a short time. Based on a comparison of several research results, it can be seen that ANN backpropagation can be applied well in various cases, especially on river water quality. Therefore, this research includes modeling of a river water quality prediction system using the ANN backpropagation method. The research samples were taken from the upstream, middle, and downstream parts of the Oya River in the Special Region of Yogyakarta Province (DIY). The Oya River is a river that originates on the western slopes of the Mount Gajahmungkur Hills, Wonogiri Regency, Central Java Province. This river flows for around 106.75 km from northeast to southwest until it empties into the Opak River in Bantul Regency Village, DIY.

The choice of DIY as the location was due to the Yogyakarta Special Region (DIY) Environment and Forestry Service saying that river water pollution was one of the three main issues and that improving the quality of the environment in DIY was a priority; also, waste and land conversion did not comply with spatial planning [17]. The results of the research can provide predictions on river water quality and a system for controlling river water quality in DIY based on appropriate methods.

2. Materials and methods

2.1. Materials

The data source in this research is secondary data from the environmental service (DLHK) of Yogyakarta Special Region (DIY) in the DIY Environmental Management Performance Information Document book. Data are in the form of a time series for 2013–2023, where each year has three periods recorded. Data are the result of monitoring water quality in the upstream, middle, and downstream Oya River.

2.2. Data collection

The method used to predict river water quality data is the ANN backpropagation method. ANN, a method for predicting data, is a subset of machine learning, especially deep learning. This method imitates the way human nerves work, forming the structure of the human brain in solving a problem, namely by carrying out a learning process through a weighting value. Artificial neural networks can recognize activities based on past data. Past data will be studied by artificial neural networks so that they can make decisions on data that has never been studied.

Artificial neural networks have several architectures that are commonly applied, including single-layer networks, multilayer networks, and competitive layer networks. The most frequently used artificial neural network architecture is a multilayer network with a combination of backpropagation learning [18].

Table 1 shows the location of the Oya River water sampling locations, and a map of the research location is presented in Figure 1. Water quality parameters include biochemical oxygen demand (BOD), chemical oxygen demand (COD), total suspended solid (TSS), dissolved oxygen (DO), total phosphate, fecal coliforms, and total coliforms.

Table 1. Oya River research location points.

Point	Location	Coordinate	
		x	y
Upstream	Kedungwates Bridge, Semin, Gunungkidul	110.73750	–7.86222
Middle	Bunder Bridge, Patuk, Gunungkidul	110.54733	–7.89114
Downstream	Dogongan Bridge, Siluk, Imogiri, Bantul	110.38014	–7.95183

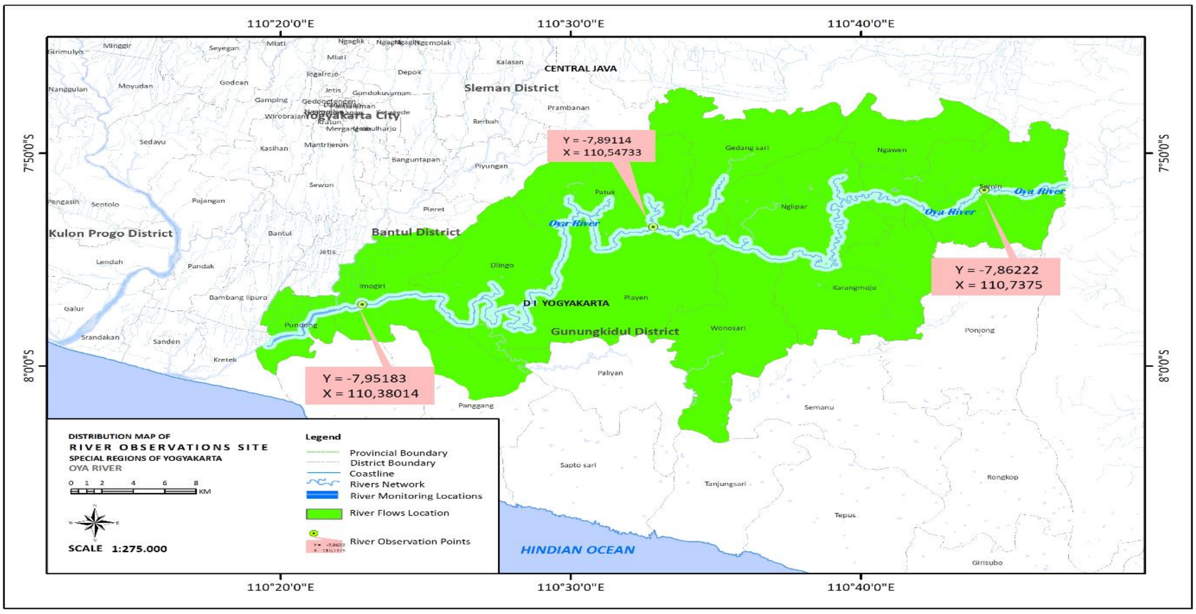

Figure 1. Map of the Oya River research location.

Figure 2 shows a network with one or more hidden layers. This multilayer network has an improved ability to solve problems than single-layer networks, so training may be more complicated. In some cases, training on these networks is better because it allows the network to solve problems that a single-layer network cannot solve.

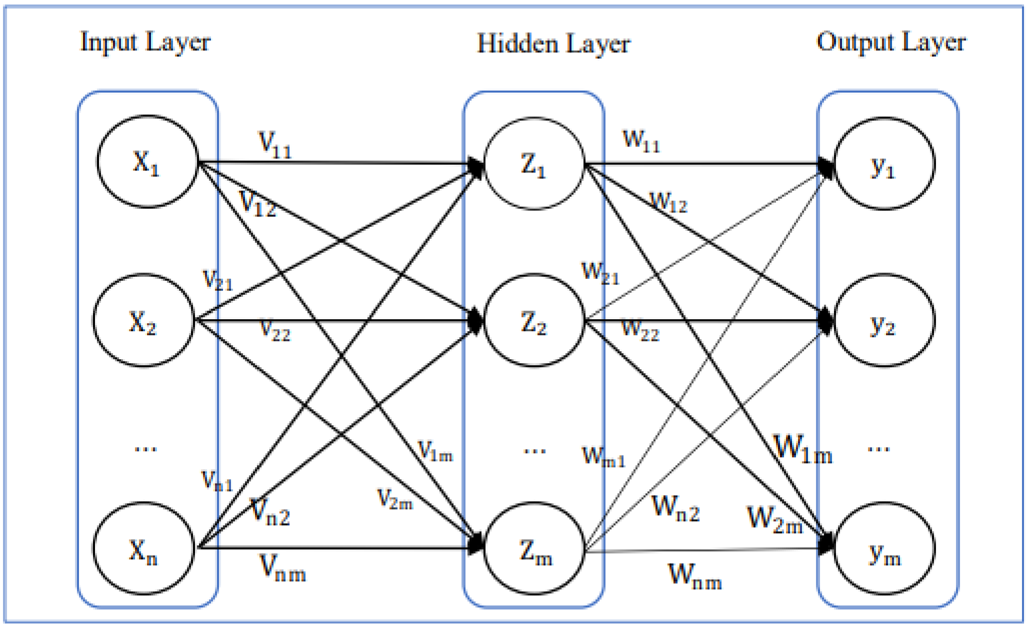

Figure 2. Multilayer network architecture.

Backpropagation is an ANN algorithm that reduces the level of prediction error by adjusting the weights based on the difference in output and the desired target. The advantage of the backpropagation method is that this method can formulate prediction experience and knowledge and is flexible in changing forecasting rules. Other advantages include having a good ability to recognize patterns in data so that it can produce an output with a high level of accuracy. Apart from that, backpropagation has good performance when applied to complex problems such as prediction and recognition of certain patterns in an image. This method can predict data using learning rules based on data that has never happened before [19]. ANN backpropagation is widely used in solving problems in various fields, including being used to predict the rate of population increase [20]. Backpropagation ANN has also been used to predict the availability of food commodities [21], to evaluate a computer project [22], and to predict floods using IoT technology and artificial neural networks [23]. The latter is relevant as river pollution also causes flooding, a natural disaster that is very detrimental to humans and that can damage natural ecosystems. Mitra et al. also forecasted or predicted floods using ANN and IoT [24]. Therefore, prediction algorithms applied to natural disasters, such as the widely used ANN backpropagation algorithm, are necessary and useful for preparing disaster mitigation activities. This was done by Alquisola et al., who used discrete wavelet transform (DWT), autoregressive integrated moving average (ARIMA), and artificial neural network (ANN) for predicting and visualizing disaster risks [25]. Meanwhile, Wu et al. conducted research using intelligent methods for preventing and mitigating disasters in river basins at high flood risk based on artificial neural networks [26]. The results of his research showed that natural ecological large sponges have a real influence on mitigating urban flood disasters.

In a backpropagation neural network, the network is given a pair of patterns: The input pattern and the desired pattern. When a pattern is given to the network, the weights are changed to minimize the difference between the output pattern and the desired pattern. This exercise is carried out repeatedly so that all the patterns produced by the network meet the desired pattern. This backpropagation neural network consists of three layers: The input, hidden, and output layers. The layer components are presented in Figure 3; Figure 4 is a flowchart of data prediction using the ANN backpropagation.

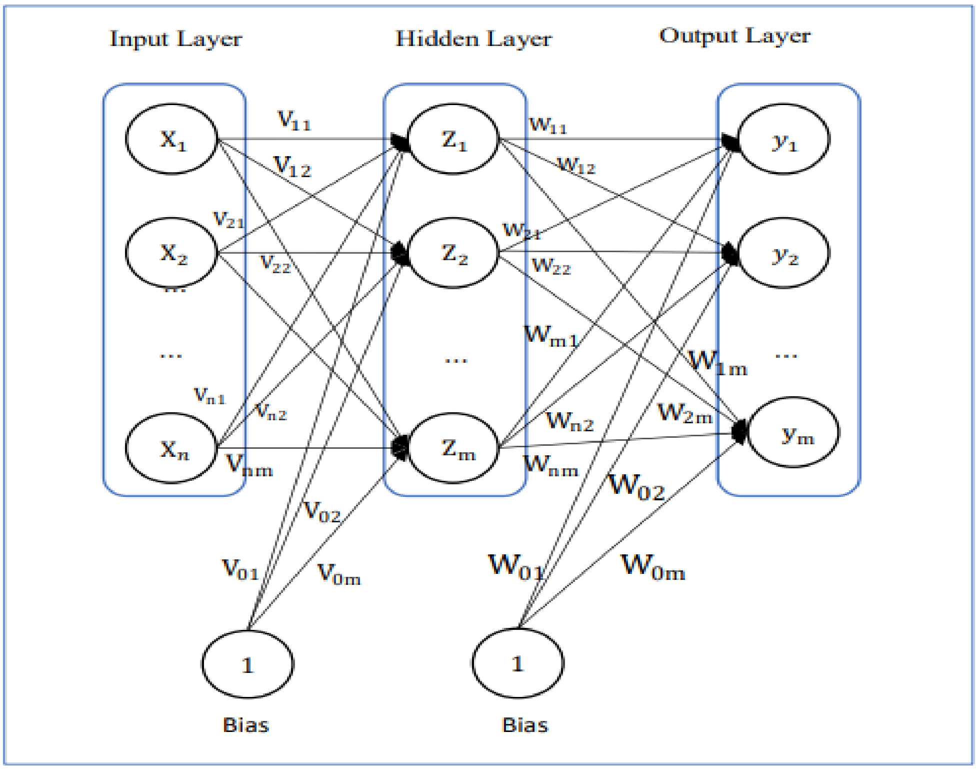

Figure 3. Backpropagation ANN network architecture with one hidden layer.

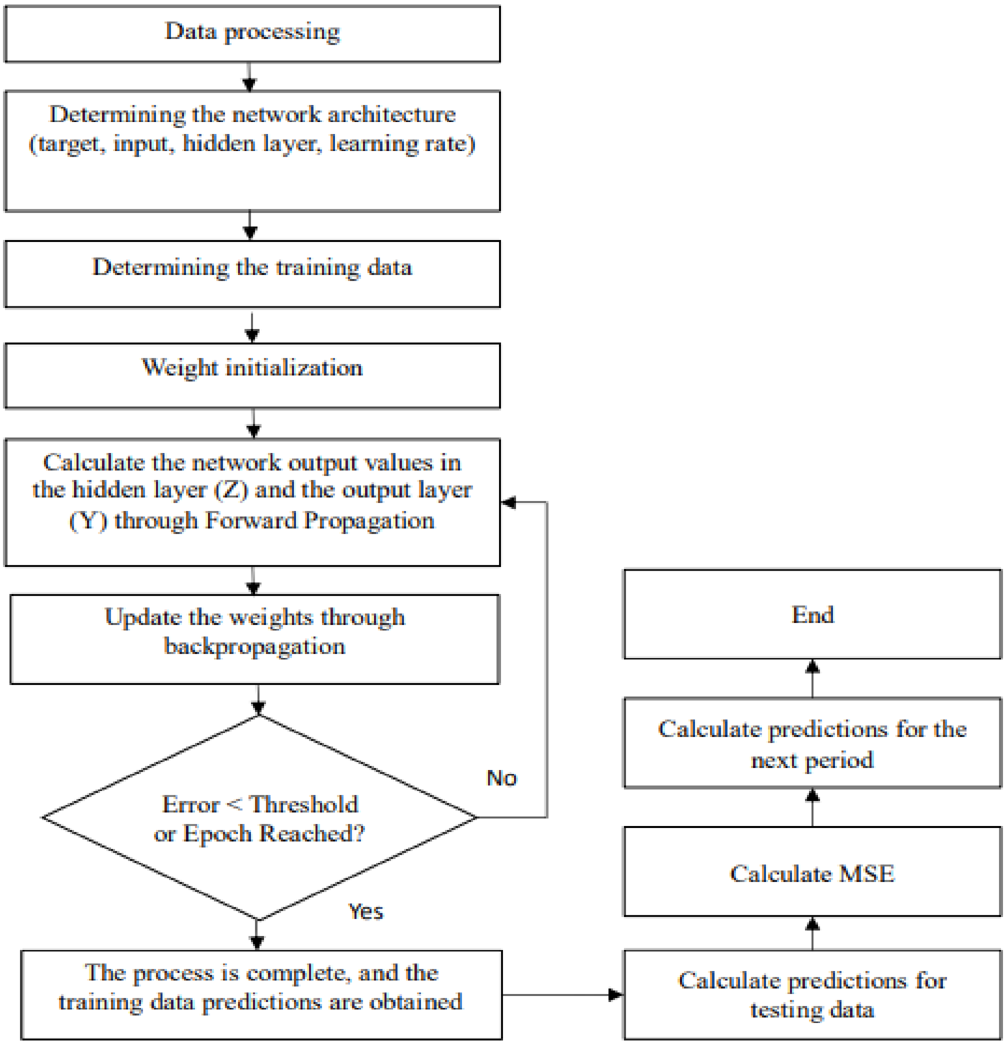

Figure 4. Flowchart of data prediction using the ANN backpropagation.

The steps in the research for predicting river water quality data using the ANN backpropagation method (Figure 4) are as follows:

(1) Prepare the river water quality data, including BOD, COD, TSS, DO, total phosphate, fecal coliforms, and total coliforms. The data is normalized using Eq 1.

Normalize the data to obtain conformity of the data values with the distance (coverage) of the activation function that has been used and defined in the network. This normalization calculation uses Eq 1:

$$z_i = (0.8(x_i - \min)/(\max\text{-}\min))+0.1 \tag{1}$$

with

z_i = Value of normalization calculation results

x_i = Value of river water quality data

Min = Minimum value in the dataset

Max = Maximum value in the dataset

(2) Determine the network architecture, where the target variable is the river water quality at time t. The input variables are the river water quality at time t–1 or one period prior, represented as X_1, and

the river water quality at time t–2 or two periods prior, represented as X_2. The number of hidden layers used is 1, 2, 3, 4, and 5, with a learning rate of 0.001.

(3) Determine the training data, which consists of river water quality data from the period 1 of 2013 to the period 3 of 2021, totaling 27 observations. This amount of data is sufficient for forming the network structure.

(4) Perform data predictions using ANN backpropagation, with the following steps:

a. Initialize parameters for the number of input neurons, hidden neurons, output layer neurons, learning rate, number of iterations, and error tolerance using Eq 2:

$$Value\ \ V = (n_{input} + 1) * n_{hidden}$$
$$Value\ \ W = (n_{hidden} + 1) * n_{ouput} \tag{2}$$

where V is the weight that connects the input layer and the hidden layer. The weight W is the weight of the hidden layer and output layer. The number of inputs is indicated by n_{input}, the number of hidden layers is indicated by n_{hidden}, and the number of outputs is indicated by n_{output}.

b. Calculate the value of the neurons in the hidden layer (Z value) including calculating the value of Z_i (Z_1, Z_2, Z_3, ... Z_n) using Eq 3:

$$Z\_in_j = V_{oj} + \sum_{i=0}^{n} X_i V_{ij}$$
$$Z\_in_j = f\ \ (Z_{in\ j}) \tag{3}$$

where Z_in_j is the value of the neuron in the j-th hidden layer, V_{oj} is the weight, and X_i is the input value of the i-th layer.

c. Calculate the value of neurons in the output layer (Y value), including calculating the value of Y_i (Y_1, Y_2, Y_3, ... Y_n) using the following Eq 4:

$$Y\_in_k = W_{ok} + \sum_{i=0}^{n} Z_j W_{jk}$$
$$Y_k = f(Y_{inj}) \tag{4}$$

where Y_in_k is the value of the neuron in the k-th wildness layer, W_{ok} is the weight, and Z_j is the value of the j-th hidden layer.

d. Perform backpropagation calculations from the output layer to the hidden layer to update the weight connecting the output layer and the hidden layer (weight W), using Eq 5 as follows:

$$W_{jk}(new) = W_{jk}(old) + \Delta W_{jk}$$
$$W_{ok}(new) = W_{ok}(old) + \Delta W_{ok} \tag{5}$$

where W_{jk} is the weight value that connects the k-th output layer and the j-th hidden layer.

e. Perform backpropagation calculations from the hidden layer to the input layer to update the weight V using the following Eq 6:

$$V_{ij}(new) = V_{ij}(old) + \Delta V_{ij}$$
$$V_{oj}(new) = V_{oj}(old) + \Delta V_{oj} \tag{6}$$

where V_{jk} is the weight value that connects the j-th hidden layer and the i-th input layer.

f. Calculate the error and evaluate whether it meets the specified threshold or if the epoch has been reached.
g. Determine the test data, which consists of river water quality data from period 1 of 2022 to period 2 of 2023, totaling 5 observations. This data is applied to the algorithm to calculate the mean squared error (MSE).
h. Calculate model accuracy by comparing prediction results and testing data using the mean square error (MSE) value in Eq 7:

$$MSE = \sum_{l=1}^{5} (z_l - \hat{z_l})^2 / l \tag{7}$$

with
z_l = first testing data
$\hat{z}_l$ = prediction results
l = 5 (amount of testing data)

i. Compare the MSE value from the prediction results of each hidden layer. The best model will be used to predict two more periods in the future.

3. Results and discussion

3.1. River water quality identification

Initial monitoring of river water quality is conducted by identifying river water quality parameters, examining data descriptions through minimum, average, and maximum values, and comparing them with quality standards. A description of the water quality parameter data in the Oya River is presented in Table 2. From 2013 to 2019, the average BOD figures in the upstream, middle, and downstream locations were 7.89, 6.19, and 7 mg/L, respectively. This is above the quality standard of 3 mg/L, meaning that many sample points contain high amounts of organic waste. Furthermore, in the 2020–2023 period, the average BOD value was 1.91 mg/L in the upstream zone, 0.94 mg/L in the middle zone, and 2.05 mg/L in the downstream zone: Decreasing from the upstream to the middle zone, and increasing in the downstream zone.

The COD and TSS variables behaved like BOD, with lower values in 2020–2023 than in 2013–2019, showing a better river water quality. The low COD and TSS indicate that the content of inorganic waste and suspended solids has decreased. Considering the COD and TSS quality standards of 25 mg/L and 50 mg/L, respectively, only a few sample points exceed those values.

The average DO in 2020–2023 was higher than in 2013–2019, namely 8.64, 2.7, and 8.82 mg/L in the upstream, middle, and downstream areas, respectively. This shows that the water quality of the Oya River in 2020–2023 was better as the amount of dissolved oxygen in the water was higher. The quality standard is 4 mg/L; many sample points met this quality standard. When comparing the DO in the upstream, middle, and downstream areas, it can be seen that the water quality in the upstream and downstream areas is better than in the middle zone.

Phosphate is a chemical compound in the form of ions that can reduce water quality; high amounts of phosphate will result in very large algae growth and result in less sunlight entering the water. The total phosphate quality standard is 0.2 mg/L. The total phosphate in 2020–2023 was higher than in 2013–2019.

Fecal and total coliform parameters decreased in 2020–2023 compared with 2013–2019. However, many sample points still presented numbers above the quality standard of 5000 mg/L, showing that the Oya River still contains a lot of domestic waste. Also, downstream locations presented higher values than upstream and middle areas.

Table 2. Descriptive data.

Parameter	Period	Characteristic	Upstream	Middle	Downstream
BOD (mg/L)	2013–2019	Minimum	1.11	0.61	1.21
		Average	7.89	6.19	7.00
		Maximum	15.10	11.45	20.90
	2020–2023	Minimum	0.07	0.10	0.34
		Average	1.91	0.94	2.05
		Maximum	3.82	4.02	3.92
COD (mg/L)	2013–2019	Minimum	6.60	3.66	4.15
		Average	20.48	14.27	15.85
		Maximum	61.03	22.60	42.40
	2020–2023	Minimum	1.39	0.10	3.18
		Average	12.19	3.14	11.52
		Maximum	25.93	22.48	19.80
TSS (mg/L)	2013–2019	Minimum	2.80	3.90	12.00
		Average	31.50	50.68	48.21
		Maximum	147.00	455.00	304.00
	2020–2023	Minimum	2.60	0.10	3.40
		Average	15.87	6.25	25.75
		Maximum	51.60	36.80	47.20
DO (mg/L)	2013–2019	Minimum	2.94	3.12	3.31
		Average	7.30	7.06	7.24
		Maximum	12.44	9.29	9.60
	2020–2023	Minimum	7.27	0.10	7.88
		Average	8.64	2.70	8.82
		Maximum	9.54	9.90	9.90
Total phosphate (mg/L)	2013–2019	Minimum	0.00	0.00	0.00
		Average	0.18	0.14	0.15
		Maximum	0.78	0.90	0.96
	2020–2023	Minimum	0.03	0.01	0.03
		Average	0.23	0.21	0.21
		Maximum	0.99	0.23	0.57

Continued on next page

Parameter	Period	Characteristic	Upstream	Middle	Downstream
Felic coliforms (MPN/100 mL)	2013–2019	Minimum	180.00	180.00	450.00
		Average	1490.95	12186.19	34973.81
		Maximum	93000.00	43000.00	93000.00
	2020–2023	Minimum	180.00	0.10	3400.00
		Average	8158.18	3589.55	25100.00
		Maximum	43000.00	9300.00	79000.00
Total coliforms (MPN/100 mL)	2013–2019	Minimum	400.00	180.00	780.00
		Average	59885.71	36410.00	106084.76
		Maximum	460000.00	150.000.00	460000.00
	2020–2023	Minimum	180.00	0.10	3400.00
		Average	38832.73	6752.04	49727.27
		Maximum	92000.00	9400.00	240000.00

Based on Table 2, it is seen that the river had poor water quality with high parameter values; from 2020 onward, the water quality improved. According to the DLHK DIY Water Quality Report, the river was categorized as lightly to moderately polluted in 2021 due to government cleanliness programs. However, increasing population and industrial activities may impact future pollution levels. To predict future water quality, we propose using statistical methods and ANN backpropagation, which simulate brain functions. We also used the R Shiny interface to process data and provide future water quality predictions in a user-friendly format.

3.2. Results of the ANN backpropagation algorithm

This paper used ANN backpropagation with several hidden layer values, namely 1, 2, 3, 4, and 5, and a learning rate of 0.001. Each parameter produces different predictions. To find out which algorithm is the best, a comparison of the MSE values is carried out. It can be seen that the smallest MSE value in all parameters and locations is when the number of hidden layers is 4, as shown in Table 3. Therefore, the ANN algorithm with 4 hidden layers is the best.

Table 3. Comparison of MSE values from water quality predictions in the Oya River.

Location	Hidden layer	BOD (mg/L)	COD (mg/L)	TSS (mg/L)	DO (mg/L)	Total Phosphate (mg/L)	Fecal coliforms (MPN/100 mL)	Total coliforms (MPN/100 mL)
Upstream	1	0.079	0.027	0.002	0.012	0.090	0.004	0.006
	2	0.097	0.032	0.002	0.013	0.091	0.003	0.006
	3	0.042	0.047	0.002	0.019	0.092	0.003	0.008
	4	0.033	0.018	0.001	0.008	0.101	0.005	0.005
	5	0.045	0.024	0.001	0.013	0.095	0.005	0.005
Middle	1	0.093	0.128	0.002	0.012	0.006	0.016	0.013
	2	0.108	0.137	0.002	0.016	0.005	0.021	0.015
	3	0.050	0.147	0.003	0.024	0.004	0.016	0.014
	4	0.045	0.082	0.000	0.006	0.008	0.006	0.008
	5	0.060	0.137	0.001	0.011	0.007	0.013	0.011
Downstream	1	0.041	0.041	0.002	0.006	0.035	0.002	0.007
	2	0.049	0.046	0.002	0.008	0.035	0.003	0.009
	3	0.046	0.060	0.004	0.010	0.035	0.004	0.014
	4	0.020	0.025	0.001	0.005	0.038	0.001	0.003
	5	0.038	0.043	0.001	0.006	0.039	0.002	0.005

3.3. Prediction results

The prediction results with 4 hidden layers are presented in Figure 5. Predictions were carried out in period 3 of 2023 and period 1 of 2024. The predicted BOD values at the upstream location were 4.537 mg/L in both periods. Meanwhile, the predicted BOD value at the middle location is 4.522 mg/L in the first period and 4.990 mg/L in the second period; in the downstream location, it was 2.523 mg/L in the first period and 4.474 mg/L in the second period. Based on Figure 5, almost all predicted BOD values still exceed the quality standard of 3 mg/L. This is supported by historical data: BOD concentrations tend to increase from period 2 in 2022 due to industrial and community activity starting to increase after the COVID-19 pandemic.

The overall prediction values show a COD concentration still below the quality standard of 25 mg/L. However, the predicted values show an increasing trend. Overall predicted values also show a TSS concentration still below the quality standard of 50 mg/L, also with an increasing trend. Predicted DO values are on the threshold of the quality standard of 4 mg/L, while predicted total phosphate concentration is below the quality standard of 0.2 mg/L. Predicted fecal and total coliform concentration values are above the quality standard of 5000 mg/L, the latter showing an increasing trend. This shows that there is a high chance of contamination of the Oya River from rubbish and anthropological waste.

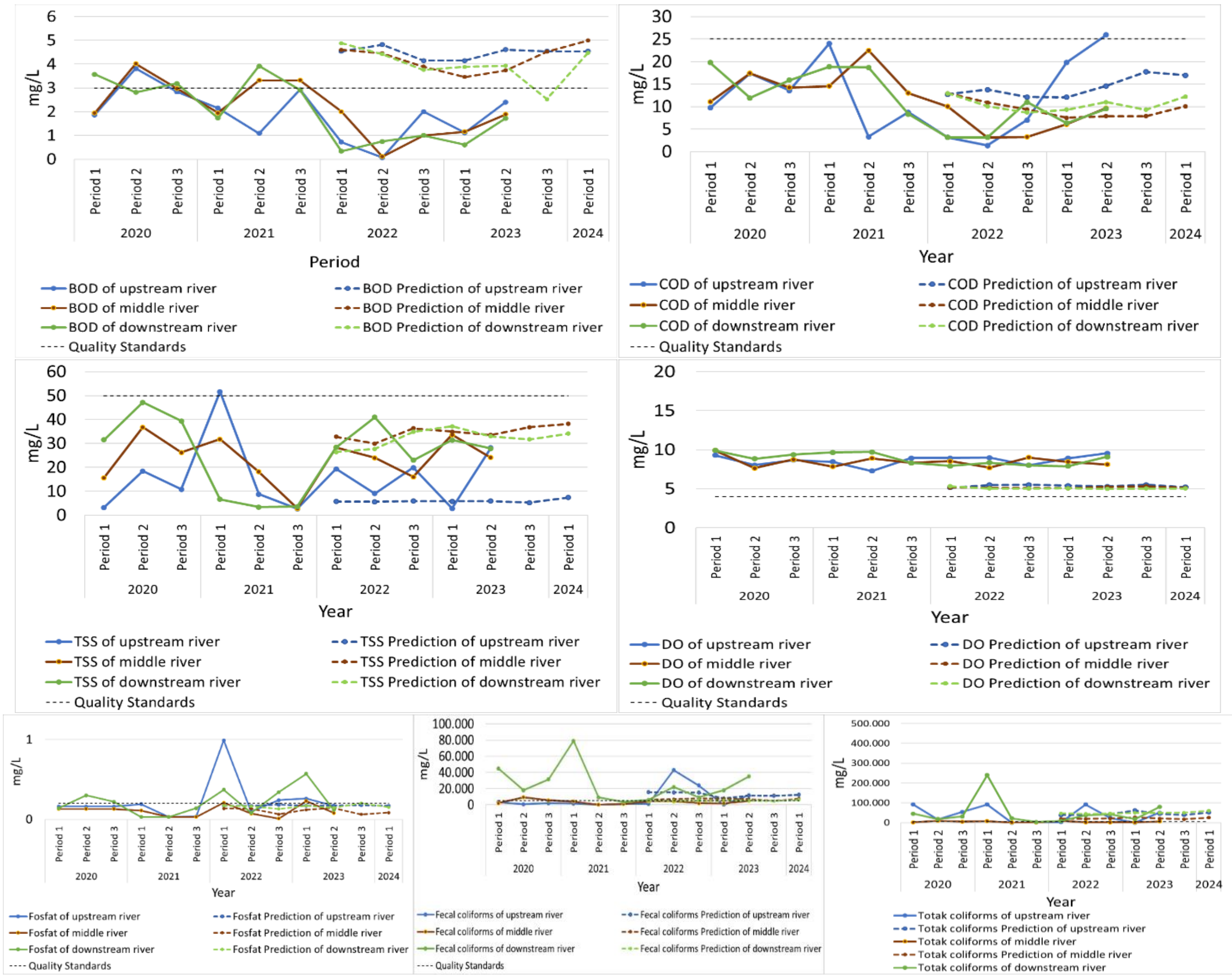

Figure 5. Comparison of actual and predicted data on concentrations.

This research used the R Shiny application, which implements the ANN backpropagation algorithm to predict river water quality. With this application, users can monitor water quality data in several locations and several periods.

The R Shiny interface as a result of the development of this research can be accessed via the link https://muhammadandangnovianta.shinyapps.io/RiverWaterQuality/.

4. Conclusions

Based on the data, it can be concluded that the water quality of the Oya River from 2020 to 2023 has improved compared with previous years. The BOD, COD, TSS, total phosphate, fecal coliforms, and total coliforms values are higher upstream than in the middle section, indicating poorer water quality upstream. The ANN backpropagation model using the R Shiny interface successfully predicted all parameters, with optimal results being achieved using 4 hidden layers. Predictions suggest that BOD and fecal and total coliform concentrations will increase and exceed quality standards in the future. Therefore, both the public and government need to be vigilant about increasing river pollution, which may come from residential waste, SME waste, and other sources. The research provides predictions for future pollution levels, highlighting the need for preventative actions and collaboration between the public and government to improve and maintain water quality, particularly for the Oya River.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in writing this article.

Acknowledgments

This research was supported by the Directorate of Research and Community Service of the Ministry of Education, Culture, Research, and Technology with contract No: 449A-06/UN7.D2/PP/VI/2023.

Conflict of interest

The authors declare no conflict of interest.

References

1. Utama JP, Syafrudin, Nugraha WD (2015) Penentuan Daya tampung beban pencemaran BOD dan Fecal Coliform Sungai Plumbon Kota Semarang Dengan software QUAL2E. *J Teknik Lingkungan* 4: 1–9.
2. Warsito B, Sumiyati S, Yasin H, et al. (2021) Evaluation of river water quality by using hierarchical clustering analysis, IOP Conference Series: Earth and Environmental Science. https://doi.org/10.1088/1755-1315/896/1/012072
3. Zubaidah T, Karnaningroem N, Slamet A (2018) K-means method for clustering water quality status on the rivers of Banjarmasin, Indonesia. *ARPN J Eng Appl Sci* 13. https://doi.org/10.31227/osf.io/s9n2u
4. Novianta MA, Syafrudin S, Warsito B (2023) K-means clustering for Grouping Rivers in DIY based on water quality parameters. *JUITA J Inform* 11: 155–163. https://doi.org/10.30595/juita.v11i1.16986
5. Di Z, Chang M, Guo P (2019) Water quality evaluation of the Yangtze River in China using machine learning techniques and data monitoring on different time scales. *Water* 11: 339. https://doi.org/10.3390/w11020339
6. Mohammadrezapour O, Kisi O, Pourahmad F (2020) Fuzzy c-means and K-means clustering with genetic algorithm for identification of homogeneous regions of groundwater quality. *Neural Comput Appl* 32: 3763–3775. https://doi.org/10.1007/s00521-018-3768-7
7. Widyastuti M, Marfai MA (2004) Kajian daya tampung sungai gajahwong thdp beban pencemaran. *Majalah Geografi Indonesia*. 8: 81–97.
8. Riza H, Santoso EW, Tejakusuma IG, et al. (2020) Utilization of artificial intelligence to improve flood disaster mitigation. *Jurnal Sains dan Teknologi Mitigasi Bencana* 30: 1–11. https://doi.org/10.29122/jstmb.v15i1.4145
9. Makubura R, Meddage DPP, Azamathulla HM, et al. (2022) A simplified mathematical formulation for water quality index (WQI): A case study in the Kelani River Basin, Sri Lanka. *Fluids* 7: 147. https://doi.org/10.3390/fluids7050147

10. Madhushani C, Dananjaya K, Ekanayake IU, et al. (2024) Modeling streamflow in non-gauged watersheds with sparse data considering physiographic, dynamic climate, and anthropogenic factors using explainable soft computing techniques. *J Hydrol* 631: 130846. https://doi.org/10.1016/j.jhydrol.2024.130846
11. Perera UAKK, Coralage DTS, Ekanayake IU, et al. (2024) A new frontier in streamflow modeling in ungauged basins with sparse data: A modified generative adversarial network with explainable AI. *Results Eng* 21: 101920. https://doi.org/10.1016/j.rineng.2024.101920
12. Pohan S, Warsito B, Suryono S (2020) Backpropagation artificial neural network for prediction plant seedling growth. *J Phys* 1524: 012147. https://doi.org/10.1088/1742-6596/1524/1/012147
13. Purba RA, Samsir S, Siddik M, et al. (2020), The optimization of backpropagation neural networks to simplify decision making, IOP Conference Series: Materials Science and Engineering, 830: 022091. https://doi.org/10.1088/1757-899X/830/2/022091
14. Nurjaya IK, Estananto E, Murti A (2022) Pemodelan sistem kendali suhu otomatis pada Smart poultry farm menggunakan metode jaringan saraf tiruan. *eProc Eng* 9.
15. Haekal M, Wibowo WC (2023) Prediksi kualitas air sungai menggunakan metode pembelajaran mesin: Studi kasus Sungai Ciliwung: Prediction of river water quality using machine learning methods: Ciliwung River case study. *Jurnal Teknologi Lingkungan* 24: 273–282. https://doi.org/10.55981/jtl.2023.795
16. Mustafa HM, Mustapha A, Hayder G, et al. (2021) Applications of iot and artificial intelligence in water quality monitoring and prediction: A review, In 2021 6th International Conference on inventive computation technologies (ICICT), 968–975. https://doi.org/10.1109/ICICT50816.2021.9358675
17. DLHK DIY (2021) Dokumen Informasi Kinerja Pengelolaan Lingkungan Hidup Daerah (DIKPLHD) Provinsi DIY Tahun 2021, Yogyakarta.
18. Masruroh M (2020) Perbandingan metode regresi linear dan neural network backpropagation dalam prediksi nilai Ujian nasional siswa smp menggunakan software R. *Joutica* 5: 331–336. https://doi.org/10.30736/jti.v5i1.347
19. Wanto A (2019) Prediksi produktivitas jagung di Indonesia sebagai upaya antisipasi impor menggunakan jaringan saraf tiruan backpropagation, *SINTECH* 2: 53–62. https://doi.org/10.31598/sintechjournal.v2i1.355
20. Sudarsono A (2016) Jaringan syaraf tiruan untuk memprediksi laju pertambahan penduduk menggunakan metode backpropagation (studi Kasus di Kota Bengkulu). *Jurnal Media Infotama* 12. https://doi.org/10.37676/jmi.v12i1.273
21. Cynthia EP, Ismanto E (2017) Jaringan syaraf tiruan algoritma backpropagation dalam memprediksi ketersediaan komoditi pangan provinsi riau. *RABIT Jurnal Teknologi Dan Sistem Informasi Univrab* 2. https://doi.org/10.36341/rabit.v2i2.152
22. Aggarwal L, Sahoo BM (2018) Back propagation algorithm for computerized paper evaluation using neural network. *IJSRST* 4.
23. Bande S, Shete VV (2017) Smart flood disaster prediction system using IoT & neural networks, 2017 International Conference on Smart Technologies For Smart Nation (SmartTechCon), 189–194. https://doi.org/10.1109/SmartTechCon.2017.8358367
24. Mitra P, Ray R, Chatterjee R, et al. (2016) Flood forecasting using Internet of things and artificial neural networks, 2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 1–5. https://doi.org/10.1109/IEMCON.2016.7746363

25. Alquisola GLV, Coronel DJA, Reolope BMF, et al. (2018) Prediction and visualization of the disaster risks in the Philippines using discrete wavelet transform (DWT), autoregressive integrated moving average (ARIMA), and artificial neural network (ANN), 2018 3rd International Conference on Computer and Communication Systems (ICCCS), 146–149. https://doi.org/10.1109/CCOMS.2018.8463238
26. Wu J, Ming H, Xu J, (2021) Research on intelligent disaster prevention and mitigation method in high flood risk area of river basin based on artificial neural network, 2021 7th International Conference on Hydraulic and Civil Engineering & Smart Water Conservancy and Intelligent Disaster Reduction Forum (ICHCE & SWIDR), 943–952. https://doi.org/10.1109/ICHCESWIDR54323.2021.9656352