2025 Journal of Engineering, Technology and Social Sciences
Jurnal Kejuruteraan, Teknologi dan Sains Sosial
Volume 11 Special Issue: ICoSCiD
International Conference on Smart Cities Development
e-ISSN: 27166848

ARTIFICIAL NEUTRAL NETWORK MODEL IN ESTIMATING WATER QUALITY INDEX (WQI) FOR UPSTREAM WATER INTAKE IN MALAYSIA

Rufaizal Che Mamat[1], Azuin Ramli[2], Muhd. Khidir Irham[3], Muugesh Rao[4], Anum Juita[5] and Fatin Nur Aliyah[6]

[1,2,3,4,5]*Department of Civil Engineering, Politeknik Ungku Omar, Jalan Raja Musa Mahadi, 31400 Ipoh, Perak, Malaysia*

rufaizal.cm@gmail.com

ARTICLE INFO

Article history:
Received
14 July 2025
Received in revised form
18 Sept 2025
Accepted
3 Oct 2025
Published online
15 Oct 2025

***Keywords**:*
water pollution; water quality index (WQI); artificial neutral network

ABSTRACT

This study addresses water pollution in Malaysia, a critical issue impacting water resource viability, the economy, human health, and ecosystems. Accurate environmental quality predictions are essential for effective management and proactive risk mitigation. This paper investigates the use of Artificial Neural Network (ANN) models to predict the Water Quality Index (WQI) for Malaysian rivers. The ANN model utilizes Biochemical Oxygen Demand (BOD), Ammoniacal Nitrogen (AN), and Suspended Solids (SS) as input layers and WQI as the output layer. The model was trained and validated using 65 river sample data points collected from 2014 to 2021 by the Environmental Quality Data. Model 11 demonstrated the best performance, exhibiting a correlation coefficient of 1.0, low Mean Squared Error (MSE) across training, test, and validation subsets, and a Root Mean Squared Error (RMSE) of 0.00. The predicted WQI values also showed a strong agreement with actual data, with a percentage error of less than 4%. These results confirm the effectiveness and suitability of ANN models for forecasting WQI values for upstream water intake, making them appropriate for use by environmental monitoring units in simulations.

1. Introduction

Malaysia relies heavily on its rivers for 98% of its total water use, but industrial pollution and urban environmental deterioration pose significant challenges to water quality and sustainability. Recent data from 2023 indicates that a notable percentage of rivers remain either slightly polluted or polluted, particularly in industrially active regions. While the country has regulations like the Environmental Quality Act (EQA) 1974, even with its 2024 amendments that introduce stricter penalties and mandatory imprisonment for severe offenses, persistent issues with enforcement and a reactive rather than proactive approach by authorities continue to hinder effective pollution control. Groundwater, constituting a significant 1.2% of total water production in Peninsular Malaysia, is also a critical resource facing contamination risks, necessitating robust monitoring and management. A spatiotemporal analysis of Malaysia's Groundwater Quality Index (GWQI) from 2014 to 2022 revealed that while the majority of Peninsular Malaysia consistently showed good groundwater quality, medium-quality areas increased by 155.66%, and very good quality areas decreased by 98.49%, highlighting a concerning deterioration over time (Algailani & Hayder, 2024). This trend, coupled with the economic impact of water supply disruptions and the health risks associated with waterborne diseases and emerging pollutants like microplastics, underscores the urgent need for continuous monitoring, targeted interventions, and improved pollution load control rather than just concentration limits.

The severe degradation of global freshwater resources, intensified by anthropogenic activities, necessitates robust water quality assessment. The Water Quality Index (WQI) effectively quantifies water quality, but traditional methods are resource-intensive. Machine Learning (ML) algorithms, including Neural Network Models (NNM), Random Forest (RF), K-Nearest Neighbor (KNN), Artificial Neural Network (ANN) and Linear Regression (LR), offer highly accurate and efficient WQI prediction, even with reduced predictor sets (Walczak & Walczak, 2025; Nishat et al., 2025). These models, demonstrating strong performance (e.g., R^2 up to 0.999), enable rapid, cost-effective assessment, supporting proactive water management and early warning systems. However, model performance depends on indicator selection and requires periodic retraining for dynamic environmental changes.

A systematic understanding of modeling environmental parameters is an essential aspect of analyzing any environmental management system to maintain the health and well-being of all living things, including humans. Environmental quality predictions are crucial for identifying potential risks and hazards, enabling proactive mitigation measures. For instance, predicting high levels of air pollution in a specific area allows for timely steps to reduce emissions and protect public health. Accurate prediction of future environmental quality is fundamental to effective management. Recent evidence suggests that machine learning applications, particularly Artificial Neural Network (ANN) models, are capable of producing accurate predictions (Mamat et al., 2023) and are frequently utilized to address water quality management issues.

Accurate prediction and modeling of environmental parameters are thus essential for effective management, with the Water Quality Index (WQI) serving as a key tool for comprehensive water quality evaluation. Machine learning (ML) applications have gained significant traction to overcome traditional challenges in WQI assessment, producing accurate predictions even with fewer parameters. Previous research, including those on the Warta River in Poland and the Kinta and Klang rivers in Malaysia, consistently demonstrate the high performance of Artificial Neural Network (ANN) models in Water Quality Index (WQI) prediction, showcasing their adaptability and accuracy across diverse environmental contexts (Nishat et al., 2025). Other ensemble methods like Random Forest Regressor have also shown strong predictive capabilities in WQI forecasting. This growing body of literature underscores the vital role of advanced computational methods in providing robust, scalable, and accurate tools for water quality monitoring and management (Coibasic et al., 2023).

In this critical context, the present study endeavors to advance the frontier in environmental quality prediction by developing a novel prototype system grounded in Artificial Neural Network (ANN) models. This research specifically proposes an integrated model designed to predict crucial environmental quality parameters by synergistically combining various algorithms. The methodology involves a rigorous evaluation of each modeling approach utilizing seven years (2014–2021) of observed data, followed by a comprehensive comparative analysis of the prediction accuracy performance of each of the developed ANN models. Ultimately, this work culminates in the creation of a user-friendly, ANN-based estimation tool poised to empower environmental practitioners and key stakeholders with enhanced decision- making capabilities.

2. Methodology

To conduct this water quality assessment, the study focused on river systems across Malaysia. Malaysia, with an average annual rainfall of 3,000mm, boasts an estimated 900 billion cubic meters of water reserves. Surface water sources, predominantly rivers, are critical, supplying over 97% of the raw water utilized for household, industrial, and agricultural needs (Wan Jaafar, 2021). Data for this research was collected from a network of 65 continuous water quality stations and water intakes located throughout Malaysia. Specifically, 57 of these stations are situated in Peninsular Malaysia, 2 in Sabah, and 6 in Sarawak. This comprehensive geographical coverage ensures a representative dataset for analyzing water quality dynamics across the nation.

To enhance data accuracy and reporting, significant improvements were implemented by 2021. River water quality reporting, previously conducted on a per-river basis, transitioned to a station-based approach, encompassing all observed river stations across Malaysia. This ensures comprehensive monitoring and status reporting, extending beyond just IKA (likely referring to specific intake points) to include detailed classifications for each river station over a five-year period. Furthermore, heavy metal reporting is now disaggregated by state and includes sample counts, based on the number of river monitoring stations. In 2021 alone, river water quality was evaluated using 8,059 samples collected from 1,351 manual monitoring stations, covering 670 rivers nationwide. The assessment revealed that 489 (73%) of these rivers had clean water quality, 158 (24%) showed moderate pollution, and 23 (3%) were heavily polluted.

The selection of primary indicators for this study was driven by their direct and strong correlation with river water quality alterations caused by pollutant loads from both point and non-point sources. Specifically, Biochemical Oxygen Demand (BOD), Ammoniacal Nitrogen (AN), and Suspended Solids (SS) were identified as key parameters. BOD quantifies the amount of oxygen required by bacteria and other microbes to decompose organic materials, with continuously high values often indicative of industrial effluent discharge. AN is primarily linked to domestic sewage and cattle rearing, while SS originates from faulty earthworks and unregulated land clearance activities. Table 1 in the original document presents the average Sub-Index values for BOD, AN, SS, and the Water Quality Index (WQI) over a seven-year period, from 2014 to 2021.

The Department of Environment of Malaysia (DOE) has adopted the Department of Environment WQI (DOE-WQI) technique, also known as the Opinion Poll WQI (OP WQI), as its standard for calculating the Water Quality Index. This formula, derived from expert panel consultation on parameter selection and weighting, includes Biochemical Oxygen Demand (BOD), Ammoniacal Nitrogen (AN), and Suspended Solids (SS) as key parameters. The current DOE approach for WQI computation involves manual calculations and requires the conversion of raw data into sub-pollutant indices. In contrast, the proposed technique utilizes historical and current raw data to predict the WQI. WQI values are categorized into five classes (Class I to Class V) based on pollution levels. While the DOE provides equations for determining parameters, WQI computations are performed using the sub-indices (SI) rather than the raw parameters themselves, employing optimized equations for rating curve fitting. The method for calculating the WQI is described in Equation (1).

$$WQI = 0.35*SIBOD + 0.35*SISS + 0.30*SIAN$$

For Artificial Neural Network (ANN) modeling, input parameters were established through correlation analysis utilizing the Statistica 13 software. Water Quality Index (WQI) prediction was performed within the MATLAB and Simulink computing environments, specifically using their Neural Network library. The model's architecture designated consideration factors (input neurons) as input parameters, with WQI serving as the output neuron. The modeling process itself was executed using the Neural Network Fitting software. Networks were constructed with a single concealed layer, where the neuron count in the hidden layer was iteratively adjusted from 2 to 10. Various learning methods, including Levenberg-Marquardt, Bayesian regularization, and scaled conjugate gradient, were employed. The optimal network was selected based on minimizing the Mean Squared Error (MSE) and maximizing the regression (R) values, as network quality improved with a drop in MSE and an increase in R (Mamat et al., 2025). The dataset was partitioned into subsets for training (70%), testing (15%), and validation (15%).

3. Results and Discussions

The described methodology was applied to a dataset from completed prototype development projects. After 49 iterations, the most effective network was identified as having eight neurons in its hidden layer (Figure 1). The Mean Squared Error (MSE) validation process showed the best validation performance, reaching a value of 5.9145e-06 after 25 iterations. The rate of error reduction (gradient) for a specific validation set iteration, based on successive MSE increases and momentum (Mu), is also detailed. The network's learning process was programmed to cease after six consecutive increases in the MSE validation error. Table 1 presents the MSE and regression (R-value) for the training, testing, and validation sections of the network's learning process. Model 11 was determined to be the best model, achieving an R-value of 1.0 and an MSE of 0.00, as indicated in Table 2. This finding is consistent with Kulisz and Kujawska (2021), who found that ANN models are effective tools for predicting the surface water quality index and are suitable for environmental monitoring simulations.

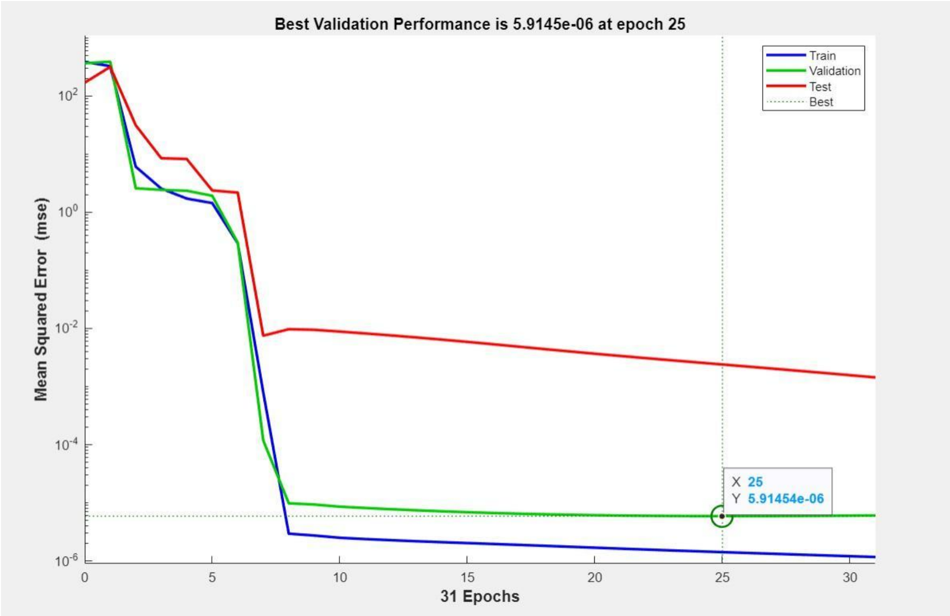

Figure 1. Variation of MSE during the training of the optimal ANN.

Table 1. Performance values of the best ANN architecture

Data Subset	Mean Square Error (MSE)	Regression (R) value
Training set (70%)	0.0000	1.0000
Validation set (15%)	0.0000	1.0000
Testing set (15%)	0.0024	1.0000

Table 2. List of models and the best ANN

Model	Hidden Layer Size	Mean Squared Error (MSE)	Regression (R)	MSE	R	MSE	R
1	10	75.0989	0.6906	27.6927	0.7222	63.2347	0.9898
2	8	8.1377E-10	1	3.42E-05	1	8.9366E-10	1
3	8	2.09E-11	1	1.28E-10	1	NaN	NaN
4	8	0.3032	0.997	4.0745	0.949	2.4605	0.9237
5	7	4.33E-11	1	1.27E-06	1	1.38E-10	1
6	7	3.12E-10	1	3.18E-10	1	NaN	NaN
7	7	0.6573	0.9937	0.4874	0.9894	0.1546	0.9984
8	6	6.89E-13	1	3.52E-10	1	5.25E-08	1
9	6	9.59E-11	1	8.70E-10	1	NaN	NaN
10	6	1.9629	0.9729	1.8814	0.9544	2.1891	0.9879
11 (best model)	8	0	1	0.0024	1	0	1

Figure 2 illustrates the model's exceptional regression performance with a consistent R-value of 1.0. The consistently high regression coefficient (R=1.0) for the training, validation, and testing sets, as well as for the overall dataset, signifies an exceptionally strong level of network matching and precise correspondence with measurement points (R>0.95), indicating excellent model fit.

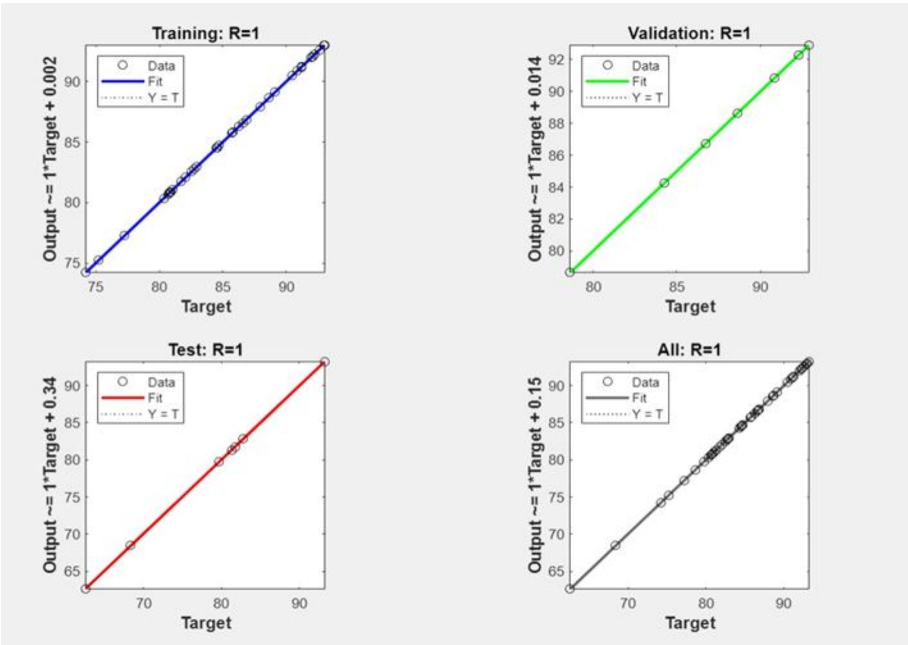

Figure 2. This is a figure. Schemes follow the same formatting

The Water Quality Index (WQI) forecast, derived from the Artificial Neural Network (ANN) modeling based on Biochemical Oxygen Demand (BOD), Ammoniacal Nitrogen (AN), and Suspended Solids (SS) as consideration factors, demonstrates remarkable accuracy. Overall, Figure 3 demonstrates a very close alignment between the WQI values measured in reality and those estimated by the model across various samples. This strong visual correlation underscores the excellent capability of the ANN model in accurately predicting water quality, particularly when considering input factors such as BOD, AN, and SS. The high fidelity of these predictions confirms the model's effectiveness in discerning and capturing complex patterns and relationships within the water quality data. This finding is consistent with Kulisz and Kujawska (2025), who established ANN models as effective tools for surface water quality index prediction and suitable for application in environmental monitoring simulations. Similarly,

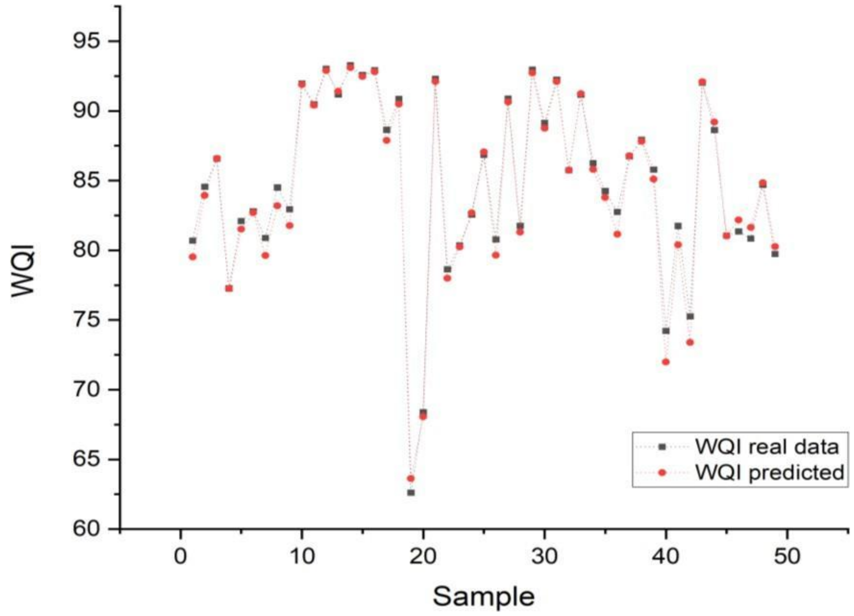

Figure 3. WQI Data Comparison Between Actual and Anticipated Values

4. Conclusion

highly effective tools for forecasting the Water Quality Index (WQI) for upstream water intake in Malaysia. The optimal ANN model, characterized by an eight-neuron hidden layer, achieved exceptional performance metrics, including a strong correlation coefficient of 1.0, a remarkably low Mean Squared Error (MSE) of 0.00 across training, test, and validation subsets, and a Root Mean Squared Error (RMSE) of 0.00. The predicted WQI values exhibited a minimal percentage error of less than 4%, confirming a robust agreement with actual data. These findings align with prior research. The flexibility of ANN architecture allows for accurate WQI prediction using a reduced set of critical physicochemical parameters (BOD, AN, SS) compared to more extensive analytical measurements, offering significant time and cost efficiencies in data acquisition. This study underscores the substantial potential of ANN models to streamline and accelerate WQI computation, thereby informing proactive environmental management and contributing to the long-term viability of Malaysia's water resources.

Acknowledgements

The authors wish to express their sincere gratitude to Politeknik Ungku Omar for their invaluable support and resources, which were instrumental in the successful completion of this research.

References

Algailani, E., & Hayder, G. (2024). Assessment and spatial mapping of water quality parameters using QGIS: creating a dynamic map. Water Quality Research Journal. https://doi.org/10.2166/wqrj.2024.036

Coibasic, S., Dmitrasinovic, S., Kostic, M., Turk Sekulic, M., Radonic, J., Dodig, A., & Stojkovic, M. (2023). Application of machine learning in river water quality management: a review. *Water Science & Technology, 88*(9), 2297.

Kulisz, M., & Kujawska, J. (2021). Application of artificial neural network (ANN) for water quality index (WQI) prediction for the river Warta, Poland. Journal of Physics: Conference Series.

Mamat, R. C., Ramli, A., & Rosyidi, S. a. P. (2023). Shear Strength Prediction of Treated Soft Clay with Sugarcane Bagasse Ash Using Artificial Intelligence Methods. Jurnal Kejuruteraan, 35(3), 597–605. https://doi.org/10.17576/jkukm-2023-35(3)-07

Mamat, R. C., Ramli, A., & Bawamohiddin, A. B. (2025). A comparative analysis of artificial intelligence techniques for carbon emission predictions in the construction industry. Mathematical Modeling and Computing, 12(2), 401–409. https://doi.org/10.23939/mmc2025.02.401

Nishat, M. H., Khan, M. H. R. B., Ahmed, T., Hossain, S. N., Ahsan, A., El-Sergany, M. M., Shafiquzzaman, M., Imteaz, M. A., & Alresheedi, M. T. (2025). Comparative analysis of machine learning models for predicting water quality index in Dhaka's rivers of Bangladesh. Environmental Sciences Europe, 37(1). https://doi.org/10.1186/s12302-025-01078-w

Niazkar, M., Piraei, R., Goodarzi, M. R., & Abedi, M. J. (2025). Comparative assessment of machine learning models for groundwater quality prediction using various parameters. Environmental Processes, 12(1). https://doi.org/10.1007/s40710-025-00751-9

Walczak, N., & Walczak, Z. (2025). Assessing the feasibility of using Machine learning algorithms to determine reservoir water quality based on a reduced set of predictors. Ecological Indicators, 175, 113556. https://doi.org/10.1016/j.ecolind.2025.113556

Wan Jaffar, W. A. (2021). Environmental Quality Report. Alamedia Sdn.Bhd., 38-100.