

Implementation of Long-Short Term Memory Neural Network (LSTM) for Predicting The Water Quality Parameters in Sungai Selangor

Nur Natasya Mohd Anuar¹, Nur Fatihah Fauzi^{2*}, Huda Zuhrah Ab Halim³, Nur Izzati Khairudin⁴,
Nurizatul Syarfinas Ahmad Bakhtiar⁵, Nor Hayati Shafii⁶

^{1,2,3,4,5,6} Faculty of Computer and Mathematical Sciences,
Universiti Teknologi MARA Perlis Branch, Arau Campus, 02600 Arau, Perlis, Malaysia

Corresponding author: *fatihah@uitm.edu.my

Received Date: 20 August 2021

Accepted Date: 6 September 2021

Published Date: 20 September 2021

HIGHLIGHTS

- Five (5) water quality parameters in the Selangor River considered in this study namely Biochemical Oxygen Demand (BOD), Ammonia Nitrogen (NH₃-N), Chemical Oxygen Demand (COD), pH, and Dissolved Oxygen (DO).
- The-water quality parameters data measured by the monitoring station of Sungai Selangor were utilized to analyse the water quality parameters in detail.
- The predicted values of the model and the actual values were in good agreement.
- For model prediction accuracy, RMSE was chosen as the unified metrics.

ABSTRACT

Predictions of future events must be factored into decision-making. Predictions of water quality are critical to assist authorities in making operational, management, and strategic decisions to keep the quality of water supply monitored under specific criteria. Taking advantage of the good performance of long short-term memory (LSTM) deep neural networks in time-series prediction, the purpose of this paper is to develop and train a Long-Short Term Memory (LSTM) Neural Network to predict water quality parameters in the Selangor River. The primary goal of this study is to predict five (5) water quality parameters in the Selangor River, namely Biochemical Oxygen Demand (BOD), Ammonia Nitrogen (NH₃-N), Chemical Oxygen Demand (COD), pH, and Dissolved Oxygen (DO), using secondary data from different monitoring stations along the river basin. The accuracy of this method was then measured using RMSE as the forecast measure. The results show that by using the Power of Hydrogen (pH), the dataset yielded the lowest RMSE value, with a minimum of 0.2106 at station 004 and a maximum of 1.2587 at station 001. The results of the study indicate that the predicted values of the model and the actual values were in good agreement and revealed the future developing trend of water quality parameters, showing the feasibility and effectiveness of using LSTM deep neural networks to predict the quality of water parameters.

Keywords: LSTM, water quality parameters, artificial neural network, monitoring stations, prediction model



INTRODUCTION

Today, the growing human population has increased the demand for forewater consumption and their anthropogenic activities, such as for land use, deforestation, industrialisation, transportation, solid waste generation and excess wastewater generation; cumulatively, changing the natural structure of the planet earth. According to the Department of Statistics in 2019, domestic and non-domestic metered water consumption in 2018 had risen by 6.2% and 14.3% respectively from 2014. The number of public sewage treatment plants in Malaysia between 2015 and 2018 had increased by 5.5%. Rapid development has created vast volumes of domestic, industrial, commercial and transportation waste, which eventually end up in water sources (Huang et al., 2015). The Selangor River Basin occupies an area of 2,200 km² or about 28% of Selangor, the most developed state in Malaysia (Santhi & Mustafa, 2012). Huge watersheds however pose many challenges to water quality monitoring and management, especially in multinational basins where regulatory mechanisms and goals for water resource management can vary (Bloesch et al., 2012). Leading onto effective river basin management requires consistent monitoring through the following four efforts: 1) identify patterns over time; 2) thoroughly consider the impacts of activities and their relationships in the watershed; 3) identify the impacts of downstream activities; and 4) the rest (Chapman et al., 2016).

Many technologies have been developed to consider the changes of water quality, such as Fuzzy Mathematics, 3S Engineering and ANN (Lee & Lee, 2018 and Maier et al., 2010). However, the ANN methodology is famous for its excellent applicability to unforeseen and non-linear circumstances for forecasting water quality (Liu et al., 2019). Artificial Neural Network (ANN) is one of the most reliable and commonly used forecasting models with effective applications in social, technological, engineering, foreign exchange, and stock problems (Khashei & Bijari, 2010). In the field of information, the neural network can overcome the conventional approach of processing information by offering fair recognition and judgement (Wu & Feng, 2017). Not only limited in the field of information technology, but they also noted that ANN was widely used in medical care due to the variability and unpredictability of the human body and health conditions. The complex non-linear interaction of biological information is worthy for the implementation of ANN. Apart from that, ANN is also popular in water quality analysis.

Three separate Artificial Neural Network (ANN) simulation techniques were used to identify the optimum forecast of water quality parameters by Najah et al. in 2012, which included the Logistic Regression Model (LRM), Multi-Layer Perceptron Neural Networks (MLP-NN) and Radial Basis Function Neural Network (RBF-NN). In their study, the RBF-NN Model was found to be the fastest computational model which increased the precision of predicting water quality parameters. The feed-forward ANN also facilitates fast simulation of the WQI and enables the recognition of the comparative significance to model predictions (Gazzaz et al., 2012). According to them, their analysis emphasised that ANN is an important water quality river evaluation instrument that simplifies the computation of WQI and saves significant effort and time by optimising the calculations. Based on Hayder et al. in 2020, with enough datapoints, a good prediction of WQP can be obtained by using three-layered Feedforward Neural Network. However, based on research from Zhou et al. in 2018, Long Short-Term Neural Network (LSTM) which is the new type of recurrent neural network is faster and easier to converge to the optimal solution when dealing with time series prediction. This is supported by a study published in 2017 by Wang et al., who concluded that the LSTM Neural Network is the best method for predicting water quality parameters when compared to the online sequential extreme learning method and the back propagation neural network method. The Root Mean Square Error (RMSE) values obtained from all three methods were compared in their study, and they discovered that the RMSE value for LSTM Neural Network consistently produces the lowest value for all time steps.

This paper proposes a water quality prediction model based on LSTM deep neural networks to predict water quality parameters data measured by the automatic monitoring station of the Sungai Selangor and then



compares the predicted results with the measured data. The results show the potential of application of LSTM and deep learning in predicting water quality parameters.

METHODOLOGY

This section focuses on data collection and data analysis. Steps in formulating and measuring the model validation will be explained concurrently.

Method of Data Collection

In this analysis, the researcher wants to scrutinize the water quality parameters in the Selangor River. Thus, the dataset used are the time series data of five criteria of water quality, which are DO, COD, pH and NH₃-N. The data for this study was obtained from the Department of Environmental Malaysia (DOE) and was collected from 10 monitoring stations along the Selangor River.

Method of Data Analysis

In data analysis, 3 phases area used in this research. The phases are the pre-processing data, formulation of the LSTM model to predict water quality parameters in Selangor River and measurement of model accuracy.

Pre-processing Data

The data size for each station varies depending on the data availability. The total number of data received in the first 4 stations out of 10, namely 2BSEL001, 2BSEL004, 2BSEL005 and 2BSEL010; is 24 data points measured every two months spanning over four years from January 2016 to November 2019. In the meantime, the comprehensive range of data received within the remaining 5 stations, specifically 2BSEL011, 2BSEL014, 2BSEL015, 2BSEL017 and 2BSEL018, is 15, spanning from July 2017 to November 2019. Finally, station 2BSEL023, the newest monitoring station along the river basin, has the smallest data available, with 13 data points recorded from November 2017 to November 2019. Figure 1 depicts the data size distribution for every station.

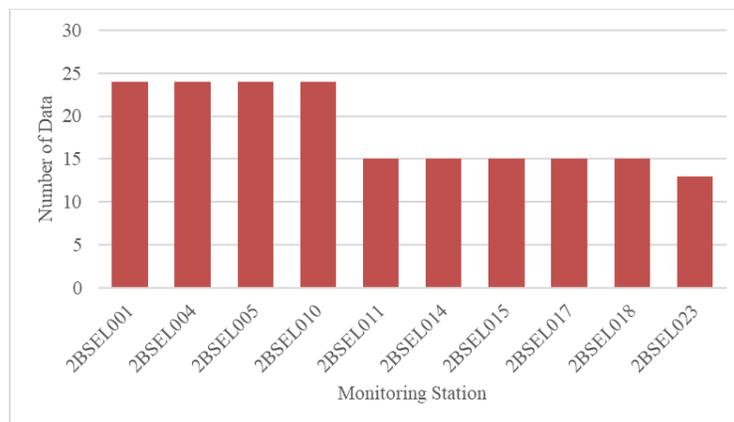


Figure 1: The data size for every station.

The dataset used in this study is the WQP from the first 4 stations, which are 2BSEL001, 2BSEL004, 2BSEL005 and 2BSEL010. This is due to the fact that the number of data points for the remaining 6 stations is insufficient for predictions because they are still considered new stations. To avoid inaccurate prediction,



the dataset trained will only include the 4 previously mentioned stations. Not only that, but shortage of time is also a factor contributing to the decrease in the number of stations trained in this study.

Based on the WQP dataset of the 4 stations, the linear interpolation technique was used to treat the missing value in the data using Microsoft Excel with NumXL function installed. Linear interpolation is a curve fitting method to generate new data points within the range of a discrete set of known data points. By implementing this method, the missing value at a particular time was fixed by taking the value before and after the time into account. Following the linear interpolation method, statistical data analysis was performed to analyse the data characteristics before proceeding with the prediction phases. The variabilities measured in this analysis are the mean, minimum, maximum, standard deviation, skewness, kurtosis, white noise, and stationarity of data based on each station.

LSTM Neural Network Model Formulation and Measurement of Accuracy

The Artificial Neural Network (ANN) is a technique that has been biologically influenced by the human brain and nervous system biology. It is a computational model composed of multiple computing components based on their predefined activation functions. For predicting the water quality parameters, the LSTM method is implemented. LSTM is a standard neural network consisting of an input layer that receives external data to perform pattern recognition, an output layer that solves the problem and a hidden fully connected intermediary layer that distinguishes the other layers (Wang et al., 2017). In this research, the network used consisted of 4 layers. Because the data used in the network was in time series, the first layer, which is the input layer, is called the Sequence Input Layer. The LSTM Layer emerges next, which discovers long-term correlations between time steps in a time series or data sequence. Several essential properties were determined in this layer, including the number of hidden units or hidden size, output format, and input size. The next important LSTM property is the activation function. The activation function is divided into two types: state activation function and gate activation function. The state activation function updates the cell and hidden state, whereas the gate activation function controls the gates in the LSTM. The third layer is the Fully Connected Layer which multiplies the input by a weight matrix and then adds a bias vector. Finally, the Regression Layer generates the output layer. The regression layer computes the half-mean-squared-errors loss for regression tasks. The network design used in this study is depicted in the Figure 2 below and taken from MatLab model's configuration.

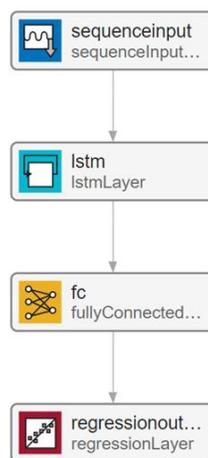


Figure 2: Network Design.



FINDINGS AND DISCUSSIONS

The results and discussion of the LSTM model will be explained in this section and in this section, all the results explained will be using the DO data from stations 001, 004, 005 and 010 considering the number of data points for these two stations is sufficient for predictions. In the meantime, the number of data points for the remaining six stations is insufficient for predictions because they are still considered new stations. To avoid inaccurate prediction, the dataset trained will only include the 4 previously mentioned stations. Not only that, but shortage of time is also a factor contributing to the decrease in the number of stations trained in this study. The five water quality parameters (WQP) considered in this study are Dissolved Oxygen (DO), Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), pH and Ammonia Nitrogen (NH₃-N) The model's ability to predict based on different water quality parameters will be discussed by examining the smallest error measurement.

Pre-Processing Data

The data were analysed by using Microsoft Excel with the NumXL function installed. The steps involved in this process are as follows:

Step 1: The data received from DOE were divided into 4 different stations and were arranged in ascending time format. The missing data was adjusted by using Linear Interpolation Method. Figure 3 shows the linear interpolation method performed in Microsoft Excel for each WQP value in Station 001.

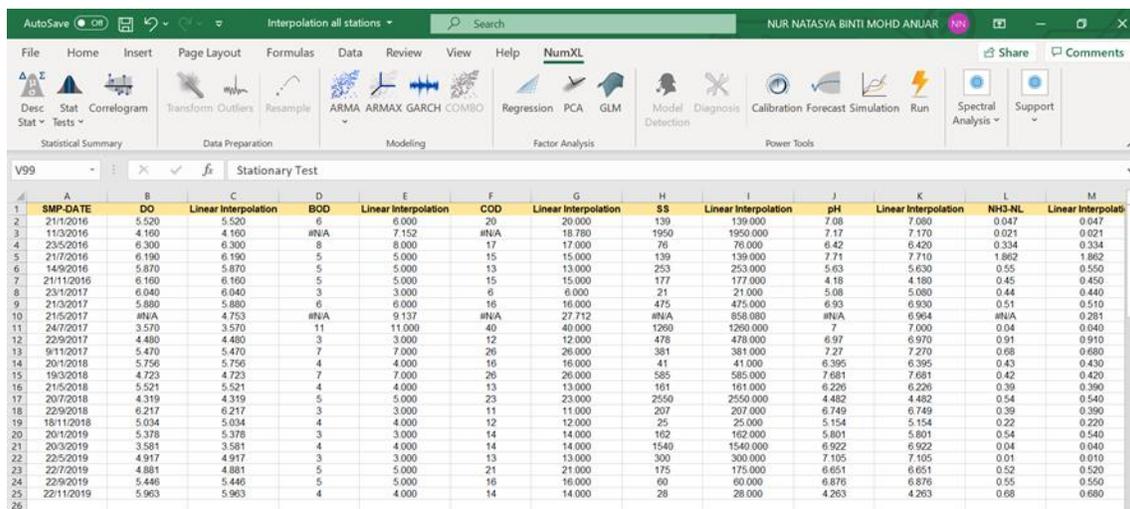


Figure 3: Linear Interpolation Method for WQP Data in Station 001

Step 2: Following the Linear Interpolation Method, the statistical analysis of data was performed. Table 1 and Table 2 provide a summary of the statistical analysis of water quality parameters based on station 001, 004, 005 and 010. The table describes the mean, minimum, maximum, standard deviation, skewness, kurtosis, white noise, and data stationarity.

Table 1: Statistical Summary of WQP Parameters at Station 001

Parameters (mg/l)	Mean	Minimum	Maximum	Standard Deviation	Skew	Excess Kurtosis	White Noise	Stationarity
DO	5.255	3.570	6.300	0.814	-0.65	-0.46	Yes	Yes
BOD	5.262	3.000	11.000	2.086	1.14	1.16	Yes	Yes



COD	17.479	6.000	40.000	7.094	1.52	3.31	Yes	Yes
pH	6.363	4.180	4.180	1.044	-0.95	-0.17	Yes	Yes
NH₃-N	0.452	0.010	1.862	0.381	2.17	7.73	Yes	Yes

As shown in Table 1, the standard deviation value for BOD and COD in Station 001 is high, indicating that the data are dispersed or less reliable. Meanwhile, the rest of the parameters have a low standard deviation, indicating that the values are spread out around the mean. As the value is less than -1 or greater than 1, all parameters except DO and pH are highly skewed. Meanwhile, the DO and pH are moderately skewed as the values lie around -1 to -0.5. Furthermore, the researcher discovered that the excess kurtosis values for DO and pH are negative, indicating that the distributions are less peaked. Meanwhile, the presence of outliers is indicated by the other parameters with positive excess kurtosis values making the prediction difficult.

Table 2: Statistical Summary of WQP Parameters at Station 004

Parameters (mg/l)	Mean	Minimum	Maximum	Standard Deviation	Skew	Excess Kurtosis	White Noise	Stationarity
DO	7.845	6.609	8.870	0.522	-0.31	0.82	Yes	Yes
BOD	4.125	3.000	7.000	1.116	0.76	0.18	No	Yes
COD	14.625	10.000	25.000	4.095	1.04	0.47	No	Yes
pH (unit)	7.438	6.830	8.485	0.354	0.93	2.06	Yes	Yes
NH₃-N	0.258	0.017	0.850	0.210	1.38	1.70	Yes	Yes

As shown in Table 2, the standard deviation value for BOD and COD in Station 004 is high, indicating that the data are dispersed or less reliable. Meanwhile, the rest of the parameters have a low standard deviation, indicating that the values are spread out around the mean. DO parameter for this station is fairly skewed, while the BOD and pH parameters are moderately skewed. Other than that, the parameters are highly skewed. All parameters have positive excess kurtosis indicating the presence of outliers. All the parameters except BOD and COD are white noise.

Table 3: Statistical Summary of WQP Parameters at Station 005

Parameters (mg/l)	Mean	Minimum	Maximum	Standard Deviation	Skew	Excess Kurtosis	White Noise	Stationarity
DO	7.658	6.573	8.790	0.474	-0.02	1.46	Yes	Yes
BOD	4.716	3.000	11.000	2.048	1.69	3.12	Yes	Yes
COD	15.550	7.000	33.000	5.648	1.41	2.85	Yes	Yes
pH (unit)	7.343	6.662	7.900	0.378	-0.11	-1.20	No	Yes
NH₃-N	0.216	0.012	2.310	0.502	3.76	14.46	Yes	Yes

As shown in Table 3, the standard deviation values for BOD and COD in Station 005 are high, indicating that the data are dispersed or less reliable. Meanwhile, the rest of the parameters have a low standard deviation, indicating that the values are spread out around the mean. Based on the skewness value, parameters of DO and pH are skewed, while the rest of the parameters are highly skewed because the values are less than -1 or greater than 1. Furthermore, the researcher found out that the excess kurtosis value for pH is negative, signalling that the distributions are less peaked. Meanwhile, the presence of outliers is indicated by the other parameters with positive excess kurtosis values making the prediction difficult.



Table 4: Statistical Summary of WQP Parameters at Station 010

Parameters (mg/l)	Mean	Minimum	Maximum	Standard Deviation	Skew	Excess Kurtosis	White Noise	Stationarity
DO	6.216	4.740	7.490	0.599	0.04	0.98	Yes	Yes
BOD	5.167	3.000	9.000	1.308	0.81	2.05	Yes	Yes
COD	17.691	8.000	32.000	4.776	0.71	2.96	Yes	Yes
pH (unit)	7.325	6.660	7.934	0.350	-0.02	-1.03	Yes	Yes
NH ₃ -N	0.633	0.010	1.673	0.368	0.78	1.56	Yes	Yes

As shown in Table 4, the standard deviation values for BOD and COD in Station 010 are high, representing that the data are dispersed or less reliable. Meanwhile, the rest of the parameters have a low standard deviation, signifying that the values are spread out around the mean. Based on the skewness value, parameters of DO and pH are skewed, while the parameter SS is highly skewed. Meanwhile, the rest of the parameters are moderately skewed because the values are between 0.5 and 1. Furthermore, the researcher discovered that the excess kurtosis value for pH is negative, indicating that the distributions are less peaked. Meanwhile, the presence of outliers is indicated by the other parameters with positive excess kurtosis values making the prediction difficult. Finally, all parameters in stations 001, 004, 005 and 010 are stationary.

Forecast Values Using the Trained LSTM

In this study, the data that had gone through the linear interpolation method were used to train the LSTM Neural Network using MATLAB software. After the LSTM model was trained, the value of all WQP were predicted using the model trained. The following table summarises the forecasted value of WQP in all stations.

Table 5: Actual and Forecasted Value for WQP in Station 001

Year	Month	DO	DO Forecast	BOD	BOD Forecast	COD	COD Forecast	pH	pH Forecast	NH ₃ -N	NH ₃ -N Forecast
2018	NOV	5.034	4.005	4.000	4.736	12.000	3.933	5.154	7.302	0.220	0.464
2019	JAN	5.378	6.555	3.000	4.203	14.000	2.820	5.801	6.354	0.540	0.478
	MAR	3.581	4.178	4.000	5.306	14.000	3.769	6.922	6.716	0.040	0.445
	MAY	4.917	6.401	3.000	7.738	13.000	5.825	7.105	6.855	0.010	0.446
	JULY	4.881	4.394	5.000	9.634	21.000	10.994	6.651	6.646	0.520	0.454
	SEPT	5.446	6.121	5.000	4.711	16.000	11.801	6.876	6.661	0.550	0.446
	NOV	5.963	4.625	4.000	6.188	14.000	2.237	4.263	6.716	0.680	0.442

Table 6: Actual and Forecasted Value for WQP in Station 004

Year	Month	DO	DO Forecast	BOD	BOD Forecast	COD	COD Forecast	pH	pH Forecast	NH ₃ -N	NH ₃ -N Forecast
2018	NOV	7.235	8.043	3.000	3.136	12.000	13.473	7.261	7.325	0.160	0.245
2019	JAN	8.165	7.995	3.000	3.191	11.000	11.176	7.383	7.430	0.090	0.208
	MAR	7.885	7.933	3.000	3.241	12.000	11.156	7.530	7.553	0.070	0.156
	MAY	6.609	7.884	3.000	3.275	11.000	10.721	7.480	7.645	0.150	0.105



	JULY	7.779	7.857	4.000	3.302	12.000	10.609	10.000	7.181	7.583	0.110	0.063
	SEPT	7.767	7.854	4.000	3.322	11.000	10.491	30.000	7.720	7.359	0.040	0.040
	NOV	7.375	7.874	3.000	3.338	12.000	10.430	8.000	7.201	7.208	0.240	0.057

Table 7: Actual and Forecasted Value for WQP in Station 005

Year	Month	DO	DO Forecast	BOD	BOD Forecast	COD	COD Forecast	pH	pH Forecast	NH ₃ -N	NH ₃ -N Forecast
2018	NOV	6.722	7.646	3.000	5.452	12.000	11.211	7.090	7.056	0.110	0.089
2019	JAN	7.901	7.701	4.000	6.369	13.000	11.868	7.307	7.342	0.100	0.089
	MAR	7.610	7.808	3.000	5.113	11.000	11.871	7.597	7.623	0.110	0.092
	MAY	6.573	7.849	3.000	3.476	12.000	11.592	7.545	7.814	0.070	0.110
	JULY	8.030	7.816	4.000	3.164	14.000	11.322	7.239	7.812	0.110	0.191
	SEPT	7.496	7.778	4.000	3.866	13.000	11.166	7.369	7.150	0.050	0.503
	NOV	7.778	7.770	3.000	4.359	14.000	11.109	7.228	6.715	0.030	1.440

Table 8: Actual and Forecasted Value for WQP in Station 010

Year	Month	DO	DO Forecast	BOD	BOD Forecast	COD	COD Forecast	pH	pH Forecast	NH ₃ -N	NH ₃ -N Forecast
2018	NOV	6.026	8.105	5.000	2.155	19.000	13.538	7.145	7.463	0.310	0.262
2019	JAN	5.996	9.109	7.000	2.301	20.000	11.197	7.043	7.484	0.570	0.182
	MAR	6.258	8.878	5.000	2.794	20.000	10.636	7.451	7.403	0.620	0.156
	MAY	5.466	8.043	5.000	3.569	18.000	10.136	7.677	7.307	0.380	0.132
	JULY	6.211	7.024	5.000	4.447	20.000	9.946	7.057	7.272	0.720	0.113
	SEPT	6.145	6.156	6.000	5.231	19.000	10.324	7.678	7.293	0.560	0.097
	NOV	6.676	5.885	6.000	5.773	17.000	11.334	7.217	7.323	0.320	0.084

Tables 5 until 8 depict the forecasted values of the WQP in Station 001, 004, 005 and 010, respectively from November 2018 to November 2019. The findings depict the relationship plot between the actual and forecasted values of all WQPs. We concluded that the predicted values had a good agreement with the effective values of the model, indicating that this model performed well in predicting the water quality parameters since the closeness of agreement between an actual value and a predicted value. Our result reveals the potential of applying LSTM and deep learning to predict drinking water quality, which can provide a reliable foundation for the formulation for water source protection policies and concrete measures.

RMSE for the LSTM Network

The following table summarises the results of the RMSE value obtained from LSTM predictions for all WQP in all stations. Table 9 shows the RMSE value obtained from the LSTM network using predicted values.



Table 9: RMSE Values Obtained from Each LSTM Network Using Predicted Values

Station	RMSE (Model using Predicted Values)				
	DO	BOD	COD	pH	NH ₃ -N
001	1.0340	2.7382	3.6519	1.2587	0.2644
004	0.6064	0.4225	1.0452	0.2160	0.2206
005	0.6197	1.6423	1.7616	0.3196	0.5608
010	2.0285	2.3309	8.1533	0.3028	0.3905

By comparing the RMSE, it is clear that the LSTM network trained using predicted values for NH₃-N consistently produces the lowest value among all WQP, with a minimum of 0.2206 at Station 004 and a maximum of 0.5608 at Station 005. The RMSE values obtained for all stations ranged from 0.6064 to 2.0285 based on parameter DO. Stations 004 and 010 yield the lowest and highest values, respectively. Meanwhile, the RMSE values obtained from parameter BOD ranges from 0.4225 to 2.7382 for all stations, with the lowest and highest values obtained from Stations 004 and 001, respectively. The RMSE values obtained from parameter COD ranges from 1.0452 to 8.1533 for all stations, with the lowest and highest values obtained from Stations 004 and 010, respectively. Finally, based on parameter pH in all stations, the RMSE value received ranges from 0.2160 to 1.2587, with the lowest and highest values obtained from stations 004 and 001, respectively. Based on this result, we can conclude that the LSTM method is efficient for predicting all parameters level in the Selangor River.

CONCLUSION AND RECOMMENDATIONS

Many important factors should be considered when developing a neural network, such as the ANN parameters, the design, which includes the number of layers, hidden numbers, epochs, and activation functions. In this study, the LSTM model was designed and trained to predict WQP in all four monitoring stations along the Selangor River using the appropriate parameters listed in the methodology. As a result, the model trained with the pH dataset consistently produced the lowest RMSE with a minimum of 0.2106 at Station 004. The established prediction model can be trained and learned automatically in the face of different water quality data samples and thus has broad application scenarios. The result shows that the built water quality model can predict the water quality parameters in the future, offering a feasible approach for water quality prediction.

Several other methods can be used to forecast the WQP in the Selangor River. Regression Analysis (RA), Grey Systems (GS), Support Vector Regression (SVR), and other ANN models such as Feedforward Neural Network, Backpropagation Neural Network, Non-Linear Input Variable Selection (IVS) algorithm, and Multi-Layer Perceptron Neural Network (MLP-NN) are examples of methods that can be used. Future researchers can use the suggested ways to compare two or more methods for forecasting WQP in river basins by using datasets in Selangor River and in any river basins worldwide to determine the best prediction method in different locations. Future researchers can also tweak the ANN parameters to achieve a more accurate model. For example, changing the ratio of training and test data, using a different activation function, increasing, or decreasing the number of epochs and hidden numbers

REFERENCES

Bloesch, J., Sandu, C., & Janning, J. (2012). Integrative water protection and river basin management policy: *The Danube case. River Systems*, 20(1), 129–144. <https://doi.org/10.1127/1868-5749/2011/0032>



- Chapman, D. V., Bradley, C., Gettel, G. M., Hatvani, I. G., Hein, T., Kovács, J., Liska, I., Oliver, D. M., Tanos, P., Trásy, B., & Várбірó, G. (2016). Developments in water quality monitoring and management in large river catchments using the Danube River as an example. *Environmental Science & Policy*, 64, 141–154. <https://doi.org/10.1016/j.envsci.2016.06.015>
- Gazzaz, N. M., Yusoff, M. K., Aris, A. Z., Juahir, H., & Ramli, M. F. (2012a). Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors. *Marine Pollution Bulletin*, 64(11), 2409–2420. <https://doi.org/10.1016/j.marpolbul.2012.08.005>
- Hayder, G., Kurniawan, I., & Mohamammed Mustafa, H. (2020). Implementation of Machine Learning Methods for Monitoring and Predicting Water Quality Parameters. *Biointerface Research in Applied Chemistry*, 11(2), 9285–9295. <https://doi.org/10.33263/briac112.92859295>
- Huang, Y. F., Ang, S. Y., Lee, K. M., & Lee, T. S. (2015). Quality of Water Resources in Malaysia. *Research and Practices in Water Quality*, 65–94. <https://doi.org/10.5772/58969>
- Khashei, M., & Bijari, M. (2010). An artificial neural network (p,d,q) model for timeseries forecasting. *Expert Systems with Applications*, 37(1), 479–489. <https://doi.org/10.1016/j.eswa.2009.05.044>
- Lee, S., & Lee, D. (2018). Improved Prediction of Harmful Algal Blooms in Four Major South Korea's Rivers Using Deep Learning Models. *International Journal of Environmental Research and Public Health*, 15(7), 1322. <https://doi.org/10.3390/ijerph15071322>
- Liu, P., Wang, J., Sangaiah, A., Xie, Y., & Yin, X. (2019). Analysis and Prediction of Water Quality Using LSTM Deep Neural Networks in IoT Environment. *Sustainability*, 11(7), 2058. <https://doi.org/10.3390/su11072058>
- Maier, H. R., Jain, A., Dandy, G. C., & Sudheer, K. P. (2010). Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. *Environmental Modelling & Software*, 25(8), 891–909. <https://doi.org/10.1016/j.envsoft.2010.02.003>
- Santhi, V. A., & Mustafa, A. M. (2012). Assessment of organochlorine pesticides and plasticisers in the Selangor River basin and possible pollution sources. *Environmental Monitoring and Assessment*, 185(2), 1541–1554. <https://doi.org/10.1007/s10661-012-2649-2>
- Wu, Y.-, & Feng, J.-. (2017). Development and Application of Artificial Neural Network. *Wireless Personal Communications*, 102(2), 1645–1656. <https://doi.org/10.1007/s11277-017-5224-x>
- Wang, Y., Zhou, J., Chen, K., Wang, Y., & Liu, L. (2017). Water quality prediction method based on LSTM neural network. *2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*. Published. <https://doi.org/10.1109/iske.2017.8258814>
- Zhou, J., Wang, Y., Xiao, F., Wang, Y., & Sun, L. (2018). Water Quality Prediction Method Based on IGRA and LSTM. *Water*, 10(9), 1148. <https://doi.org/10.3390/w10091148>

