



River Water Quality Prediction and Analysis–Deep Learning Predictive Models Approach

Nur Najwa Mohd Rizal, Gasim Hayder, and Salman Yussof

Abstract

In depth research about river water qualities are no more outlandish nowadays due to river water pollutions and contaminations. In order to have an accurate and precise measurement taken towards these river water pollution, advanced and new technologies need to be applied rather than old technique of everyday lab testing. Therefore, with the usage of deep learning predictive models approach, the decision makers able to provide immediate response and give precautionary measures to prevent a disastrous event. In the current research, Adaptive Neuro-fuzzy Inference System (ANFIS) has been used to predict six different types of river water quality parameters at Langat River, Malaysia. Root mean square error (RMSE) and determination of coefficient (R^2) were used to assess the performances of the models. The results have been proven that ANFIS able to predict the parameters of river water quality as ANFIS Model 5 has achieved the highest value of R^2 (0.9712). It also obtained the low values of RMSE which were 0.0028, 0.0144 and 0.0924 for training, testing and checking data sets, respectively. Overall, the six ANFIS models have successfully predict six different water quality parameters.

Keywords

River water quality prediction • River • Water quality • Deep learning • ANFIS

1 Introduction

The prediction of river water quality is no more out of the ordinary today especially in the field of hydrology and environmental science. Clean water is extremely important for all. Therefore, there is no living beings including animals, plants and humans can survive in this world without clean water. Besides drinking, numerous sectors of economy, viz. manufacturing and commercial, agriculture, hydroelectric power supply, fisheries and even animal husbandry depends on the clean river water supply. Thus it shows that water, particularly river water plays an important role (Tyagi et al. 2013).

As urbanization and population growth increased, it has caused the needs of fresh water increased to the very great extent over the past several decades (Al-Badaii et al. 2013). As stated by Abba et al. (2020), water pollution is known as the existence of toxic or harmful substances in water that may results in disadvantageous to living beings at a certain level (Abba et al. 2020). The chances of rivers to be polluted by heavy metals and other contaminants that results from human activities are high. Therefore, this has placed the river system at high risk due to the disadvantageous of environment pollution since the river can be effortless accessed for waste disposal and also because of the dynamic nature of the river itself (Ahmed et al. 2019).

These contaminants and pollutions occur in rivers have deteriorated the river water quality. There are two main factors that will affect the water quality which are the natural factors (hydrological, climate and geological factors) and the human factors (Sami et al. 2021). Human factors usually are the contaminants and pollutions that results from rapid

N. N. M. Rizal (✉)

College of Graduate Studies, Universiti Tenaga Nasional (UNITEN), 43000 Kajang, Selangor Darul Ehsan, Malaysia
e-mail: nur.najwa@uniten.edu.my

G. Hayder

Department of Civil Engineering, College of Engineering, Universiti Tenaga Nasional (UNITEN), 43000 Kajang, Selangor Darul Ehsan, Malaysia

S. Yussof

Institute of Informatics and Computing in Energy, Universiti Tenaga Nasional (UNITEN), 43000 Kajang, Selangor Darul Ehsan, Malaysia

urbanization, agriculture and livestock farming. Thus, a suitable measure need to be done in order to maintain the management of the river water quality from the river water pollution.

Artificial Intelligence (AI) have been widely used by most scientist and investigators from all over the world to predict the parameters of the river water quality. Lafdani et al. (2013) have stated that nowadays, the growth in AI gives a difference in prediction as an estimator used for hydrological phenomenon (Lafdani et al. 2013). When the hydrological data is introduced to the model, it able to learn or discover the system behaviour which gives the main advantage of the AI models.

There are abundance of AI modelling that have been developed to predict river water quality. Adaptive Neuro-fuzzy Inference System (ANFIS), Fuzzy Logic (FL), Support Vector Machine (SVM), Artificial Neural Network (ANN), Radial Basis Neural Network (RBNN) and Multi-layer Perceptron (MLP) are the examples of AI models that able to be applied to predict time series related modelling based on historical data (Rizal 2020). ANFIS on the other hand is a machine learning that has a feed-forward multi-layer neural network composed of fuzzy logic and ANN. In order to produce the input–output relationship with the nonlinear depiction, ANFIS uses ANN and a learning algorithm of fuzzy logic systems (Azad et al. 2019).

Previous study also showed great results when using ANFIS to predict water quality in their research. For example, Abba et al. (2017) have used ANN and ANFIS techniques in order to predict the concentration of dissolved oxygen (DO) in Yamuna River. However, it has shown that ANFIS outperformed ANN in performances. The authors have achieved satisfying results for ANFIS with the value of 0.94 and 0.7 for correlation coefficient (R) and root mean square error (RMSE), respectively, in the calibration phase and the value of 0.81 and 1.38 for R and RMSE, respectively, in the validation phase. Even though ANFIS is better than ANN, the authors concluded that it still can be applied in modelling the DO concentration in the river (Abba et al. 2017). Other previous research that conducted by Abba et al. (2019), Sonmez et al. (2018) and Ranković et al. (2012) also predicted river water quality only by using historical data and have achieved good results in their research (Abba et al. 2019; Sonmez et al. 2018; Ranković et al. 2012). In the current research, ANFIS models have been established for the prediction of six different water quality parameters in Langat River, Malaysia. In the next section, study area and the methods are explained. In Sect. 3, the results of the modellings are presented while the discussion is deliberated in Sect. 4. The conclusion of the study is detailed out in Sect. 5.

2 Study Area and Methods

Langat River, Malaysia is the study area chosen as shown in Fig. 1. The historical data (from 1981 to 2019) of the river water quality parameters have been retrieved from the water quality station (Station No. 2917601) at Department of Irrigation and Drainage (DID), Malaysia. Magnesium, pH, total solid (TS), conductivity, colour, ammonia, nitrate, turbidity, dissolved solid (DS), chloride, solids, alkalinity, fluoride, calcium, hardness, biochemical oxygen demand (BOD_{5day}), chemicals, potassium, sodium, manganese, silica, iron, phosphate, total suspended solid (TSS) and sulphate were the parameters of water quality applied as inputs for the modelling. Moreover, only 161 data for all 25 parameters of the water quality were available due to the missing value. A basic statistics of the raw data of the parameters of the water quality in Langat River are shown in Table 1.

The ANFIS models have been developed using Neuro-Fuzzy Designer app in MATLAB 2020b and all of the historical data have been cleaned and pre-processed to avoid errors while running the modelling. The missing values in the historical data have been replaced with a constant value which was a zero value. Furthermore, the historical data have been normalized in the range of [0,1] and have been divided into 3 sets of data namely, training data (70%), testing data (15%) and checking data (15%) sets. While developing the ANFIS models, subtraction clustering method was used due to large input data in the modelling. Backpropagation method and 100 epochs were chosen for the optimization method and the number of training epochs, respectively.

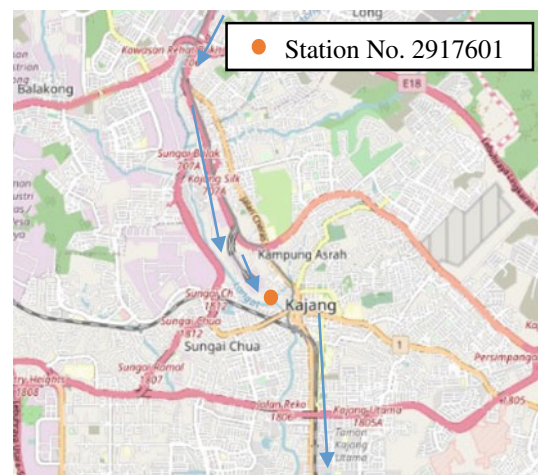


Fig. 1 The study area

Table 1 Basic statistic of the raw data of the parameters of water quality in Langat River

Parameters (unit)	Average	Std. Dev
Alkalinity (mg/L × 100)	3801.43	1661.62
Ammonia (mg/L × 100)	193.18	168.51
BOD5day (mg/L × 100)	7.47	13.37
Calcium (mg/L × 10)	177.68	147.42
Chemical (mg/L × 100)	3207.64	2542.24
Chloride (mg/L × 10)	87.80	49.59
Colour (Hazen)	66.30	89.18
Conductivity (µs/cm)	154.85	68.94
Dissolved solid (mg/L × 100)	1053.99	3818.74
Fluoride (mg/L × 100)	27.19	10.34
Hardness	49.79	54.58
Iron (mg/L × 10)	57.56	86.57
Magnesium (mg/L × 10)	15.95	53.75
Manganese (mg/L × 100)	41.2	150.21
Nitrate (mg/L × 100)	822.98	769.82
pH (pH × 10)	68.91	41.55
Phosphate (mg/L × 100)	51.08	245.42
Potassium (mg/L × 10)	199.54	888.25
Silica (mg/L × 100)	1809.54	1103.56
Sodium (mg/L × 100)	1328.65	5091.28
Solids (mg/L)	264.21	298.22
Sulphate (mg/L × 10)	136.93	93.87
Total solid (mg/L)	360.16	269.97
Total suspended solid (mg/L)	348.27	274.88
Turbidity (Fullers × 10)	1520.68	2336.71

3 Results

In the current research, six models of ANFIS have been established to predict six different water quality parameters, viz. nitrate, phosphate, BOD, TS, DS and TSS. Table 2 shows the results of the ANFIS models according to the respective targets.

Based on Table 2, it has shown that ANFIS Model 5 has achieved the highest value of R^2 with the value of 0.9712. It

also obtained low values of RMSE which were 0.0028, 0.0144 and 0.0924 for training, testing and checking data sets, respectively. Compared to ANFIS Model 3 that has been used to predict DS, it has achieved the lowest value of R^2 (0.4662). ANFIS Model 1, 2, 4 and 6 also obtained satisfying outcome by gaining good value of R^2 which were 0.9501, 0.8903, 0.8342 and 0.7588, respectively. However, all of the ANFIS models have obtained low values of RMSE for all data sets. Figure 2 shows the regression plot of ANFIS Model 5 that has been used to predict nitrate.

Table 2 The outcomes of the ANFIS models according to the respective targets

ANFIS model No.	Target	R^2	RMSE		
			Training	Testing	Checking
1	BOD	0.9501	0.0037	0.0103	0.07868
2	TSS	0.8903	0.0045	0.0491	0.1236
3	DS	0.4662	0.0013	0.0110	0.2422
4	TS	0.8342	0.0024	0.1318	0.1119
5	Nitrate	0.9712	0.0028	0.0144	0.0924
6	Phosphate	0.7588	0.0020	0.0899	0.1169

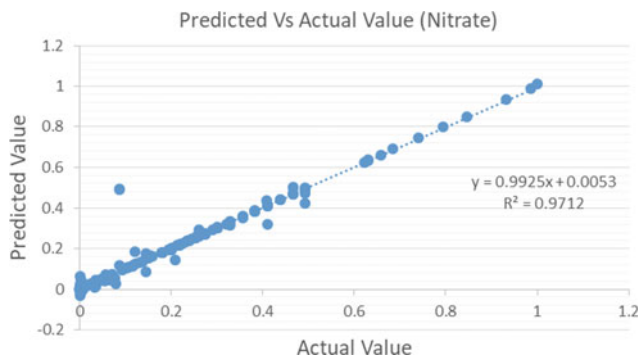


Fig. 2 The regression plot for ANFIS Model 5

4 Discussion

Based on the results in Sect. 3, it has shown that most of ANFIS models have achieved wonderful outcomes. Most of the models also obtained line of best fit in regression plots and achieved high accuracy in prediction. Past researches also have proven that using ANFIS to predict river water quality do obtained good outcome. Ranković et al. (2012) have used ANFIS method to predict dissolved oxygen (DO) in Gruža reservoir in Serbia. Eight parameters of water quality as inputs for the model have been used by the authors. The authors have achieved the respective values for the mean square error (MSE) and the mean absolute error (MAE) for the comparisons of measured and ANFIS predicted values of DO which were 0.6670 and 1.23 for test set while 1.0373 and 2.1831 for training and test set (Ranković et al. 2012). Meanwhile, Sonmez et al. (2018) have applied ANFIS for prediction of cadmium (Cd) concentrations in Fliyos River, Turkey. The performance parameters that have been used were R^2 , mean absolute deviation (MAD), mean absolute percentage error (MAPE), MSE and Nash–Sutcliffe efficiency. The results obtained shown that relatively higher correlation value, $R^2 = 0.91$ was found between modelled and observed Cd concentrations. Thus, it indicated that ANFIS model gave good estimations with high degree robustness and accuracy for the concentrations of cadmium (Sonmez et al. 2018). Abba et al. (2019) on the other hand have used ANFIS, auto regressive integrated moving average (ARIMA) and ANN model in modelling WQI at Yamuna River, India and Kinta River, Malaysia. The result has shown that ANFIS-III with 5 input parameters, triangular membership function and 2 membership function (5, trimf, 2) was the best model in predicting WQI for calibration and verification phase for both Kinta and Agra stations. Thus, ANFIS has proven to be the optimal performance in predicting the water quality index for both regions (Abba et al. 2019). Therefore, ANFIS models have been proven as an excellent approach to predict water quality parameters in

river since it able to provide high accuracy prediction with low errors while developing the modelling.

5 Conclusion

Six ANFIS models have been developed successfully for prediction of six different parameters of water quality in Langat River. Five of the ANFIS models have achieved high value of R^2 and low value for RMSE in all three data sets. ANFIS Model 3 is the only model that has obtained poor results and achieved low accuracy to predict the river water quality. For future study, the models can be upgraded to an advance predictive modellings to predict the water quality parameters.

Acknowledgements The work is supported by research grant provided by Fundamental Research Grant Scheme (FRGS) with the project code 20190105FRGS from the Ministry of Higher Education.

References

- S.I. Abba, S.J. Hadi, J. Abdullahi, River water modelling prediction using multi-linear regression, artificial neural network, and adaptive neuro-fuzzy inference system techniques. *Procedia Computer Science* **120**, 75–82 (2017)
- S.I. Abba, A.S. Maihula, M.B. Jibril, A.M. Sunusi, M.A. Ahmad, M.A. Saleh, Application of data driven algorithms for the forecasting of non-linear parameter. *International Journal of Recent Engineering Science* **6**(2) (2019)
- S.I. Abba, S.J. Hadi, S.S. Sammen, S.Q. Salih, R.A. Abdulkadir, Q.B. Pham, Z.M. Yaseen, Evolutionary computational intelligence algorithm coupled with self-tuning predictive model for water quality index determination. *J. Hydrol.* **587**, 124974 (2020)
- A.N. Ahmed, F.B. Othman, H.A. Afan, R.K. Ibrahim, C.M. Fai, M.S. Hossain et al., Machine learning methods for better water quality prediction. *Journal of Hydrology* **578**, 124084 (2019)
- F. Al-Badaii, M. Shuhaimi-Othman, M.B. Gasim, Water quality assessment of the Semenyih river, Selangor, Malaysia. *Journal of Chemistry* (2013)
- A. Azad, H. Karami, S. Farzin, S.F. Mousavi, O. Kisi, Modeling river water quality parameters using modified adaptive neuro fuzzy inference system. *Water Science and Engineering* **12**(1), 45–54 (2019)
- E.K. Lafdani, A.M. Nia, A. Ahmadi, Daily suspended sediment load prediction using artificial neural networks and support vector machines. *J. Hydrol.* **478**, 50–62 (2013)
- V. Ranković, J. Radulović, I. Radojević, A. Ostojić, L. Čomić, Prediction of dissolved oxygen in reservoirs using adaptive network-based fuzzy inference system. *J. Hydroinf.* **14**(1), 167–179 (2012)
- N.N.M. Rizal, River water quality prediction using artificial intelligence approach: literature review. *Journal of Energy and Environment* (2020)
- B.H.Z. Sami, B.F.Z. Sami, C.M. Fai, Y. Essam, A.N. Ahmed, A. El-Shafie, Investigating the reliability of machine learning algorithms as a sustainable tool for total suspended solid prediction. *Ain Shams Engineering Journal* (2021)

-
- A.Y. Sonmez, S. Kale, R.C. Ozdemir, A.E. Kadak, An adaptive neuro-fuzzy inference system (ANFIS) to predict of cadmium (Cd) concentrations in the Filyos River, Turkey. *Turk. J. Fish. Aquat. Sci.* **18**(12), 1333–1343 (2018)
- S. Tyagi, B. Sharma, P. Singh, R. Dobhal, Water quality assessment in terms of water quality index. *American Journal of Water Resources* **1**(3), 34–38 (2013)