

## NARX-based water quality index model of Air Busuk River using chemical parameter measurements

Muhammad Ierfan Hasnan<sup>1</sup>, Azhar Jaffar<sup>2</sup>, Norashikin M. Thamrin<sup>3</sup>, Mohamad Farid Misnan<sup>4</sup>, Ahmad Ihsan Mohd Yassin<sup>5</sup>, Megat Syahirul Amin Megat Ali<sup>6</sup>

<sup>1,3</sup>School of Engineering, Universiti Teknologi MARA, Shah Alam, Malaysia

<sup>2</sup>Department of Electrical Engineering, Politeknik Ungku Omar, Ipoh, Malaysia

<sup>4</sup>School of Engineering, Universiti Teknologi MARA Cawangan Johor, Kampus Pasir Gudang, Malaysia

<sup>5,6</sup>Microwave Research Institute, Universiti Teknologi MARA, Shah Alam, Malaysia

### Article Info

#### Article history:

Received May 19, 2021

Revised Jul 23, 2021

Accepted Jul 28, 2021

#### Keywords:

Ammoniacal nitrogen

Artificial neural network

Biological oxygen demand

Dissolved oxygen

Water quality index

### ABSTRACT

Water quality plays a major role in issues related to public health and marine life. Hence, monitoring river for contaminations is vital for ensuring safe and sustainable water resources. Conventional method for assessing water quality index is costly as it requires considerable amount of time and laboratory resources. Therefore, this study proposes a water quality index model based on artificial neural network. A six-year data for Air Busuk River is obtained from the Department of Environment. Dissolved oxygen, biological oxygen demand, and ammoniacal nitrogen has shown high correlation with water quality index. The water quality index model is then developed based on these parameters, employing the non-linear autoregressive with exogeneous input structure. Generally, the model which is based on three chemical parameters has shown satisfactory performance with overall regression of 0.8767 and passed the correlation function tests. The model offers a potentially efficient method for assessing water quality with cost-saving benefits for government agencies and monitoring authorities.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



### Corresponding Author:

Megat Syahirul Amin Megat Ali

Microwave Research Institute

Universiti Teknologi MARA

40450 Shah Alam, Malaysia

Email: megatsyahirul@uitm.edu.my

## 1. INTRODUCTION

Water quality has a significant impact in issues related to public health and marine life. Freshwater bodies are under constant threat of pollution from the expanding population, urban development, as well as the increasing agricultural and industrial activities [1]. Therefore, assessment of water quality is essential for ensuring safe and sustainable water resources. Generally, the overall quality of water bodies such as rivers and lakes are described by a parameter known as water quality index (WQI) [2]. In Malaysia, the Department of Environment (DOE) assesses WQI of rivers based on the national water quality standards (NWQS). The parameter is derived through a combination of physico-chemical parameters. The chemical parameters include dissolved oxygen (DO), biological oxygen demand (BOD), chemical oxygen demand (COD), ammoniacal nitrogen (NH<sub>3</sub>-N), and potential of hydrogen (pH). Meanwhile, suspended solids (SS) are the only physical parameter used for calculating WQI [3]. These parameters are to an extent, influenced by changes to water temperature. Increasing temperature will reduce the absorption rate of oxygen and other gases [4]. These affect photosynthesis by aquatic plants and thus, resulting in lower DO [5].

Generally, DO is a chemical parameter that assesses the level of contamination based on the lack of oxygen due to consumption by organic substances in water. The parameter is positively correlated with WQI, thereby indicating that a high level of DO would be desirable [6]. Meanwhile, BOD and COD each indicate the amount of oxygen consumed by microorganisms when performing aerobic decomposition of organic matters [7], and oxidation of particles in water [8]. The reduced amount of oxygen indicates increased contamination of organic matter and particles in the water. Both parameters directly influence the level of DO in water. Therefore, lack of DO in water leads to anaerobic condition in water that would adversely affect aquatic life [9]. Conversely, pH indicates the level of free hydrogen ions in the water. Increased acidity has been observed when there is a reduction in photosynthetic activities, assimilation of carbon dioxide and bicarbonate ions [10]. Meanwhile,  $\text{NH}_3\text{-N}$  is essentially a measure of ammonia, a toxic pollutant from organic waste, fertilizers, and other industrial effluents. Increased acidity has also shown to increase toxicity of ammonia, leading to potential death of aquatic life [11]. These further reduces DO and water quality [12].

SS refers to the concentration of small particles that are suspended in water. These suspended particles absorb more heat from solar radiation than water particles. Therefore, a high concentration of SS will result in increase in water temperature and reduced DO [13]. Conversely, concentration of much smaller particles is measured in terms of dissolved solids (DS). These comprise of a variety of inorganic salts and dissolved organic matter [14]. Elevated levels of DS may not pose health hazard but increases the hardness of water [15]. Both SS and DS are smaller elements to a parameter known as total solids (TS) [16]. Conversely, turbidity is defined as the measure of relative clarity of water. The parameter is positively correlated with SS in which it assesses the intensity of light that can penetrate through the surface of water body [17]. Increased turbidity affects the growth of aquatic plants due to reduced amount of light required for photosynthesis. Coupled with increasing temperature and acidity, these will adversely impact the level of DO in water [18]. While not directly used for determining WQI in Malaysian rivers, DS, TS, and turbidity are additional parameters regularly monitored by the Department of Environment [3].

Various efforts have been made to integrate water quality monitoring with artificial intelligence (AI) due to its capability in handling non-linear data, robustness, reliability, cost-effectiveness, efficiency, as well as effectiveness in problem-solving, and decision-making [19]. Realization of these intelligent systems have seen an exponential rise over the past two decades due to challenges from the lack of funding, data deficiency, time-series pattern that are unique to geological locations, as well as increasing contaminants, and variables [14]. Utilization of AI for water quality monitoring in Malaysian rivers has also been observed. To date, numerous AI methods have been tested. These include artificial neural network [20], fuzzy and hybrid models [21], as well as nature inspired algorithms [22]. Despite the recent progresses in AI paradigms, artificial neural network (ANN) still retains its effectiveness for modelling non-linear time-series data. These are realizable through the non-linear autoregressive with exogeneous inputs (NARX) model structure. Such approach has been successfully implemented to capture the dynamics of  $\text{NH}_3\text{-N}$  on river water pollution. Comparative analysis has also shown the effectiveness of NARX over the static models [23].

The standard approach for determining WQI is expensive since it requires considerable amount of time and laboratory resources to extract the physico-chemical information. A reduction of input variables to the model will not only optimize the resources, but also enable efficient computation by the model [24]. Based on the extensive review, two major gaps in the literature have been identified. First, the parameters acquired by the Department of Environment comprise of DO, BOD, COD,  $\text{NH}_3\text{-N}$ , pH, temperature, SS, DS, TS, and turbidity. However, the strength of relationship between each of these parameters to WQI based on NWQS has not been determined. Second, a dynamic model that captures the relationship between the significant water parameters and WQI has yet to be developed. Therefore, this study sets out to achieve the following objectives: 1) to analyze the correlation between the parameters acquired by the Department of Environment and WQI and 2) to develop a NARX-based WQI model based on the significant water quality variables. The study is conducted on Air Busuk River, a branch that feeds water into the larger Klang River Basin [25]. This of significant importance as any fluctuation in the WQI of the river will adversely affect the major river branch. Therefore, the model would serve as an early warning system to alert the authorities of pending contamination.

## 2. RESEARCH METHOD

### 2.1. Data acquisition and correlation analysis

Water parameters for Air Busuk River from 2012 to 2018 has been obtained from the Department of Environment. The parameters include WQI, DO, BOD, COD,  $\text{NH}_3\text{-N}$ , pH, temperature, SS, DS, TS, and turbidity. Throughout the six-years period, measurements are performed using calibrated equipments and standard operating procedure at station 1K35. Sampling is performed approximately once in every two weeks. To ensure optimum computational performance, an analysis is required to reduce these parameters to

the most significant few. The strength of relationship between each of the water parameters with WQI is performed using correlation test. This can be mathematically expressed by (1), where  $r_{xy}$  is the correlation coefficient of the relationship between variables  $x$  and  $y$ .  $x_i$  and  $\bar{x}$  are each the individual values and mean of variable  $x$ , while  $y_i$  and  $\bar{y}$  are each the individual values and mean of variable  $y$ .

$$r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \tag{1}$$

The strength of relationship between the water parameters and WQI is then determined based on the categories in Table 1 [26]. Parameters with coefficients of less than 0.3 will be discarded as it is uncorrelated with WQI. Subsequently, cross-correlation will be performed among the remaining water parameters. Should strong correlation be obtained for any two of the water parameters, only one will be used in the development of NARX-based WQI model. Highly correlated variables will only increase computational complexity, but not significantly improve model performance [27].

Table 1. Strength of relationship based on correlation coefficient

Correlation Coefficient	Strength of the Relationship
$r_{xy} \geq 0.7$	Strong
$0.5 \leq r_{xy} < 0.7$	Moderate
$0.3 \leq r_{xy} < 0.5$	Weak
$r_{xy} < 0.3$	None

**2.2. ANN and NARX model**

Generally, ANN mimics the function of biological neural network in the brain. The network is therefore capable of learning from given examples and generalize solution for a particular problem. Thus far, the method has been widely used for pattern classification and function approximation. ANN is essentially a black box modelling technique that performs non-linear mapping of  $m$ -dimensional input space onto  $n$ -dimensional output space, with no prior knowledge of the input-output relationships [28].

NARX is an effective approach for modelling time-series phenomenon and has been extensively used in a variety of applications. The structure which utilizes information from past memories to capture the dynamic behavior of non-linear systems. When developing the NARX model using ANN, two architectures will be used. The open-loop series-parallel structure is used for training, while the closed-loop parallel form is used for testing the model [29]. The architectures are illustrated in Figure 1.

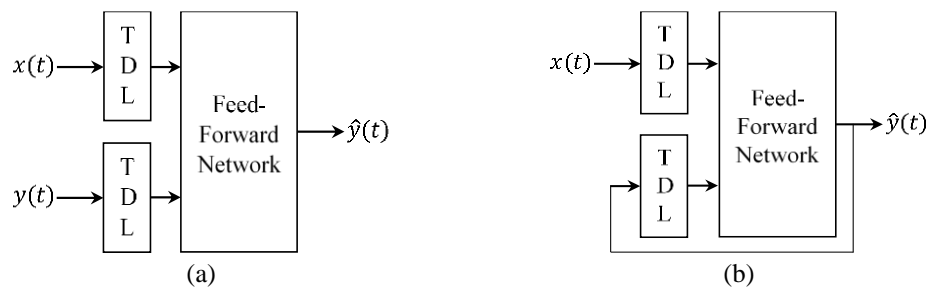


Figure 1. These figure are, (a) series-parallel, (b) parallel NARX architectures with tapped delay lines (TDL) [29]

The input-output relationship for series-parallel and parallel structures can each be expressed by (2) and (3), where  $F(\cdot)$  is the estimated model, and  $\hat{y}(t)$  is the predicted output at time  $t$ . The desired past output values are denoted by  $y(t - 1), y(t - 2), \dots, y(t - n_y)$ , and the inputs are  $x(t - 1), x(t - 2), \dots, x(t - n_x)$ . Conversely, for parallel architecture,  $\hat{y}(t - 1), \hat{y}(t - 2), \dots, \hat{y}(t - n_y)$  represents the lagged predicted output values from the NARX model.  $n_x$  and  $n_y$  each denotes the number of delays for input and output [30].

$$\hat{y}(t) = F[y(t - 1), y(t - 2), \dots, y(t - n_y), x(t), x(t - 1), x(t - 2), \dots, x(t - n_x)] \tag{2}$$

$$\hat{y}(t) = F[\hat{y}(t - 1), \hat{y}(t - 2), \dots, \hat{y}(t - n_y), x(t), x(t - 1), x(t - 2), \dots, x(t - n_x)] \tag{3}$$

In the series-parallel structure, the value of times series is predicted from present and past inputs, as well as the desired past outputs. The structure is used when training the model, making use of desired past values of the time-series. This feed-forward architecture allows for the standard multilayered perceptron (MLP) network to be implemented. After the training phase is completed, the NARX model is converted to parallel architecture. Prediction of time-series from this parallel structure relies on present and past inputs, as well as past predicted outputs [31]. The mapping function  $F(\cdot)$  is initially unknown. This can be approximated by training the MLP network within the internal architecture. Figure 2 shows the structure of an MLP network which comprises of an input layer, two hidden layers, and an output layer. Studies have shown that in certain applications, single hidden layer would suffice for approximating a function. However, more hidden layers will be required for modelling complex problems [32].

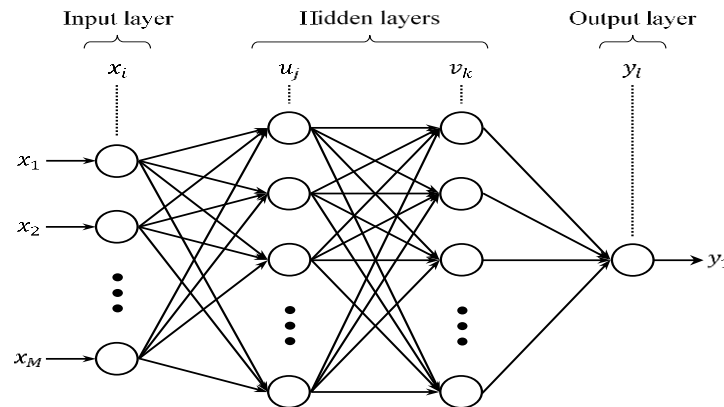


Figure 2. Structure of two-hidden layer MLP network

Initially, the vector of input variables,  $x_i$  is converted to a vector of variables for the first hidden layer,  $u_j$ , via activation function,  $\Gamma_1$ . This is shown by (4), where  $w_{ij}$  are weights connecting  $i$ th input node to  $j$ th node in the first hidden layer.  $\theta_j$  represents the bias, and  $M$  denotes the number of input nodes [33].

$$u_j = \Gamma_1 \left( \sum_{i=1}^M w_{ij} x_i + \theta_j \right) \quad (4)$$

Subsequently, the vector of variables from the first hidden layer,  $u_j$  is transformed to a vector of variables for the second hidden layer,  $v_k$ , through activation function,  $\Gamma_1$ . This is expressed by (5), where  $w_{jk}$  are weights connecting  $j$ th node in the first hidden layer to  $k$ th node in the second hidden layer.  $\theta_k$  denotes the bias, and  $N$  is the number of nodes in the first hidden layer.

$$v_k = \Gamma_1 \left( \sum_{j=1}^N w_{jk} u_j + \theta_k \right) \quad (5)$$

Similar transformation procedure is performed between the vector of variables from the second hidden layer,  $v_k$  to the output layer,  $y_l$ , via activation function,  $\Gamma_2$ . This is shown by (6), where  $w_{kl}$  are weights connecting  $k$ th node in the second hidden layer to  $l$ th output node, and  $\theta_l$  represents the bias, and  $O$  denotes the number of nodes in the second hidden layer [33].

$$y_l = \Gamma_2 \left( \sum_{k=1}^O w_{kl} v_k + \theta_l \right) \quad (6)$$

Hyperbolic tangent is used as the activation function for the hidden layers. Meanwhile, the output layer utilizes the pure linear function.  $\Gamma_1$  and  $\Gamma_2$  are each be expressed by (7) and (8).

$$\Gamma_1(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (7)$$

$$\Gamma_2(x) = x \tag{8}$$

During training, the network learns to generalize a solution based on a given set of input and output. This is realized through backpropagation weight update procedure that considers the error between the computed and actual output. The weights and biases within the network are iteratively adjusted until the error is sufficiently minimized. To avoid the network from over-fitting, early stopping criterion is also implemented. An unseen validation data is used to periodically assess the network during training. Increasing validation error indicates that the network is beginning to over-fit and training is immediately stopped [34]. Subsequently, the testing data is used to assess the model performance with final weights and biases from the training phase. The time-series data is initially normalized using min-max scaling. 70% of the data is used for training, 15% for validation, and the remaining 15% for testing [35]. To preserve the dynamic characteristics of time-series information, data is divided in block format. Levenberg-Marquardt algorithm [36] is implemented for network training. Various lag values are tested for the best network performance. The model is assessed via regression, time-series response, and error histogram. Autocorrelation function (ACF) test [37] and cross-correlation function (CCF) [38] test is performed to verify that the residuals and predicted output is not correlated with each other. Development of the WQI model is performed in MATLAB.

### 3. RESULTS AND DISCUSSION

#### 3.1. Correlation analysis

The strength of relationship between each water quality parameter and WQI is determined through correlation analysis. Table 2 shows the correlation coefficients for DO, BOD, COD, NH<sub>3</sub>-N, pH, temperature, SS, DS, TS, and turbidity based on a six-year period data. Based on the results, pH, temperature, SS, DS, TS, and turbidity have shown no correlation with WQI. These indicate that the parameters do not have a direct influence on water quality. Conversely, BOD and COD has shown strong relationship with WQI, followed by DO with moderate correlation. Meanwhile, NH<sub>3</sub>-N has also demonstrated some degree of correlation, albeit a weak one. Generally, the correlation pattern between WQI with BOD, COD, DO, and NH<sub>3</sub>-N shows conformity with the literature. WQI is positively correlated with DO, but negatively correlated with BOD, COD, and NH<sub>3</sub>-N. To ensure that the NARX model is computationally efficient, a cross-correlation analysis is performed among DO, BOD, COD, and NH<sub>3</sub>-N. Highly correlated features will not only increase computational complexity but reduces model performance [39]. Table 3 summarizes the correlation coefficients obtained between the selected parameters.

Table 2. Correlation coefficient between water quality parameters and WQI

Parameter	Correlation Coefficient, $r_{xy}$
DO (% sat)	0.5720
DO (mg/l)	0.5803
BOD (mg/l)	-0.7567
COD (mg/l)	-0.7448
NH <sub>3</sub> -N (mg/l)	-0.4035
pH	0.1833
Temperature (°C)	-0.0116
SS (mg/l)	-0.0040
DS (mg/l)	-0.1799
TS (mg/l)	-0.1048
Turbidity (NTU)	-0.0523

Table 3. Cross-correlation among DO, BOD, COD, and NH<sub>3</sub>-N

Parameter	DO (% sat)	DO (mg/l)	BOD (mg/l)	COD (mg/l)	NH <sub>3</sub> -N (mg/l)
DO (% sat)	1				
DO (mg/l)	0.9902	1			
BOD (mg/l)	-0.0826	-0.0862	1		
COD (mg/l)	-0.1570	-0.1621	0.8504	1	
NH <sub>3</sub> -N (mg/l)	-0.0879	-0.0892	0.2937	0.3310	1

The standard DO measurement has shown high correlation with the same parameter measured as percentage saturation. Strong relationship between BOD and COD has also been observed. Meanwhile, NH<sub>3</sub>-N did not show high degree of correlation with other selected parameters. To increase computational efficiency and improve model performance, only DO, BOD, and NH<sub>3</sub>-N are selected. The selection of these chemical parameters is valid as each are correlated with WQI.

### 3.2. NARX-based WQI model

The study adopts a two-hidden layer MLP network with seven TDL for lagged input and past output values during the training phase. Each hidden layer comprises of twenty hidden nodes. This optimum structure has been determined through exhaustive experiment for every randomtwister setting. Lagged input and past output of 14 weeks has been adopted to develop the model. These were based on extensive trial-run experiments that produces the most optimum results. After the network is successfully trained, the NARX model is converted from open-loop series-parallel architecture to the closed-loop parallel form. Generally, the model has attained satisfactory performance with overall regression of 0.8767. Figure 3 shows the regression obtained for training, validation, and testing.

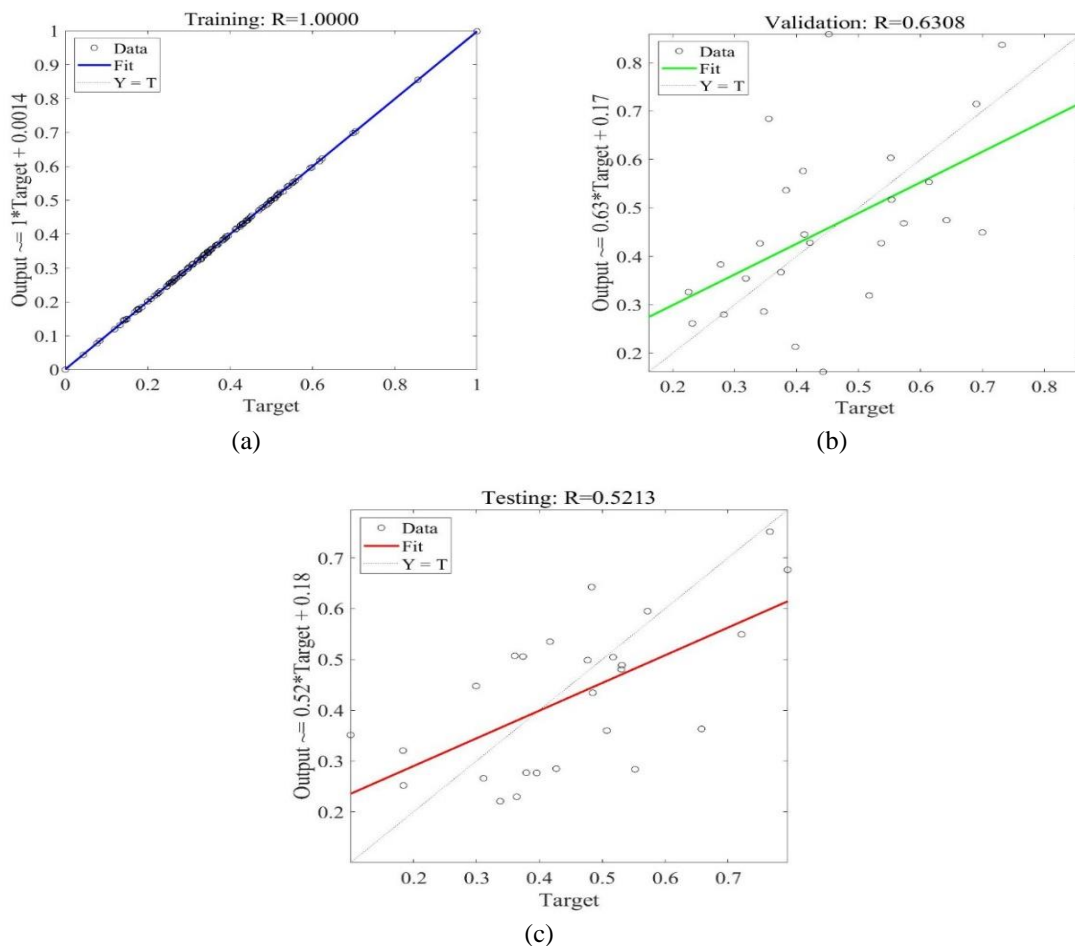


Figure 3. Regression for; (a) training, (b) validation, (c) testing

Excellent performance has been observed for training, attaining regression of 1. However, the results are less than satisfactory for validation and testing, each yielding regression of 0.7. These may be attributed by several factors. The frequency of sampling for Air Busuk River is approximately once in every two weeks. Therefore, the model was not able to perfectly capture the dynamic information from current and past inputs of DO, BOD,  $NH_3-N$ , resulting in deteriorating performance for validation, and testing. Moreover, there may be additional information that has yet to be addressed by the model but is not present in the data provided by the Department of Environment. Data such as rainfall volumes which directly relates to current and water level will also influence the measured parameters.

Meanwhile, Figure 4 revealed the model fit and error obtained for training, validation, and testing. These are further supported by the error histogram in Figure 5, illustrating the bias and distribution of error during the different phases of model development. Excellent model fit has been achieved for the training phase. However as predicted from the regression results, less than satisfactory error has been attained for validation and testing. Further investigation revealed that all error related to network training are close to

zero. However, higher errors are much prominent for validation and training. These support the results from preceding time-series response. The model is unbiased as majority of the error are distributed near zero.

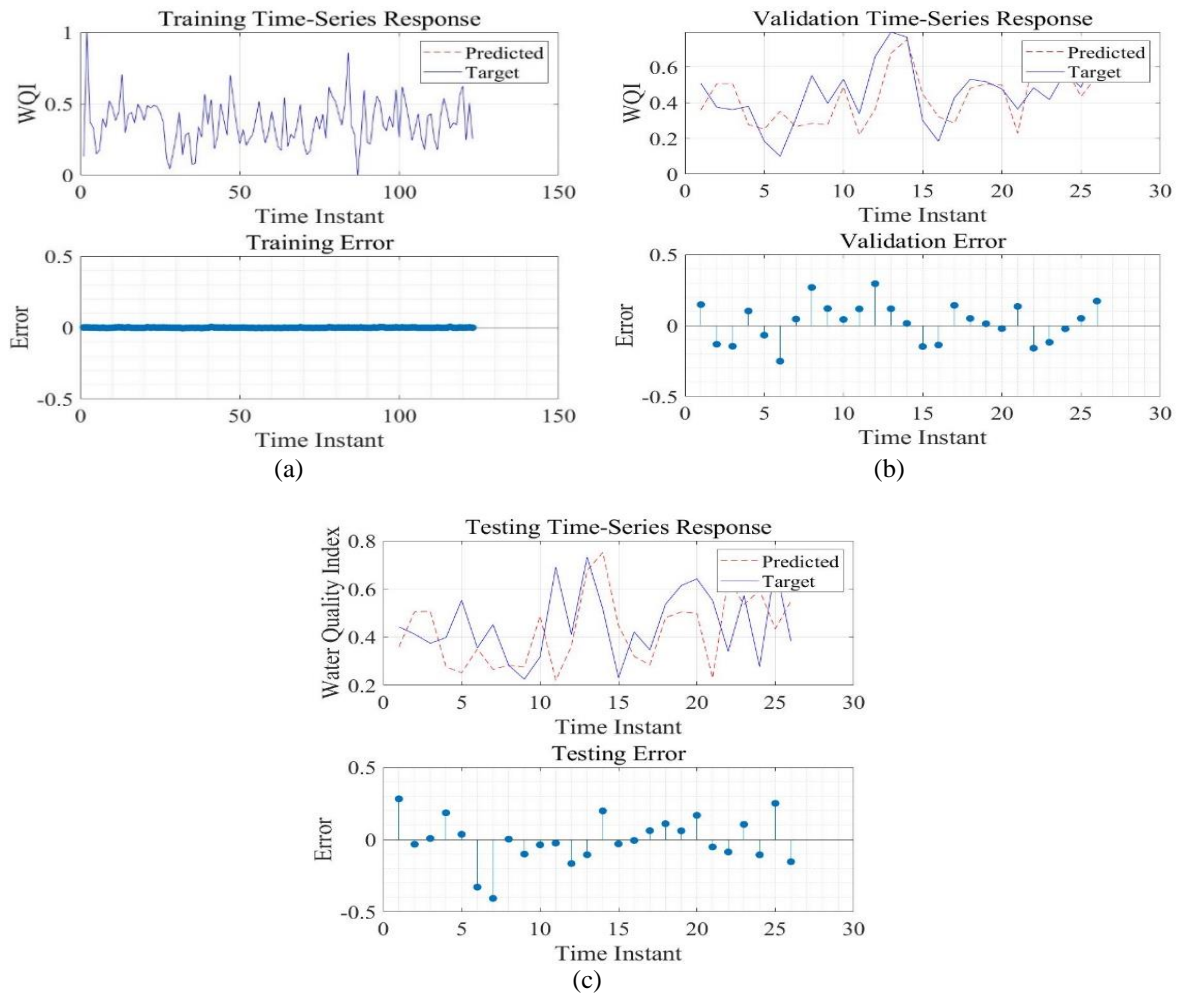


Figure 4. Time-series response for; (a) training, (b) validation, (c) testing

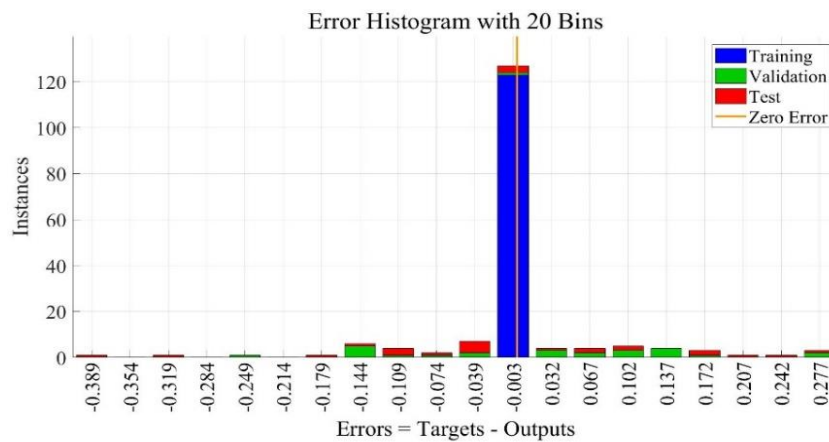


Figure 5. Error histogram

For the model to be accepted, the residuals and predicted output should not be correlated between each other. Results for ACF and CCF tests during training are shown in Figure 6. For ACF test, the results are acceptable as correlations for majority of other lags are within the 95% confidence limit except for lag 0, at 1. The results at lag 1 and lag 2 are still acceptable as the correlation still falls close to defined boundary. Conversely for CCF test, majority of correlation is also within the 95% confidence limit. Similarly, correlations at lag 0, lag 2, and lag 3 are also acceptable as the results is near to the defined boundary.

Figure 7 demonstrates the results for ACF and CCF test for validation data. For ACF test, correlations for all lag falls within the 95% confidence limit except for lag 0, at 1. The CCF test has also shown desirable results with majority of correlations within the defined boundary. Correlation at lag -1 is also acceptable as the result is close to the 95% confidence limit. Conversely, ACF and CCF tests for testing is shown by Figure 8. The ACF test demonstrated similar pattern of results to that of model validation. Correlations for all lags are within the 95% confidence limit except for lag 0, at 1. CCF test has also shown desirable results where correlations for majority of the lags fall within the 95% confidence limit. However, moderate correlation is shown at lag 0, which can still be accepted.

Based on the results presented, the model has sufficient capabilities to predict WQI of Air Busuk River based on DO, BOD, and  $\text{NH}_3\text{-N}$ . The study has therefore, established a new and efficient method for assessing water quality with reduced number of parameters and consequently, saving cost for measurement of unrelated parameters. These cost saving measures could instead be redirected for increasing frequency of sampling and other factors that have yet to be considered by the model. Capability to predict WQI of Air Busuk River will have a significant impact as it directly affects the quality of Klang River Basin.

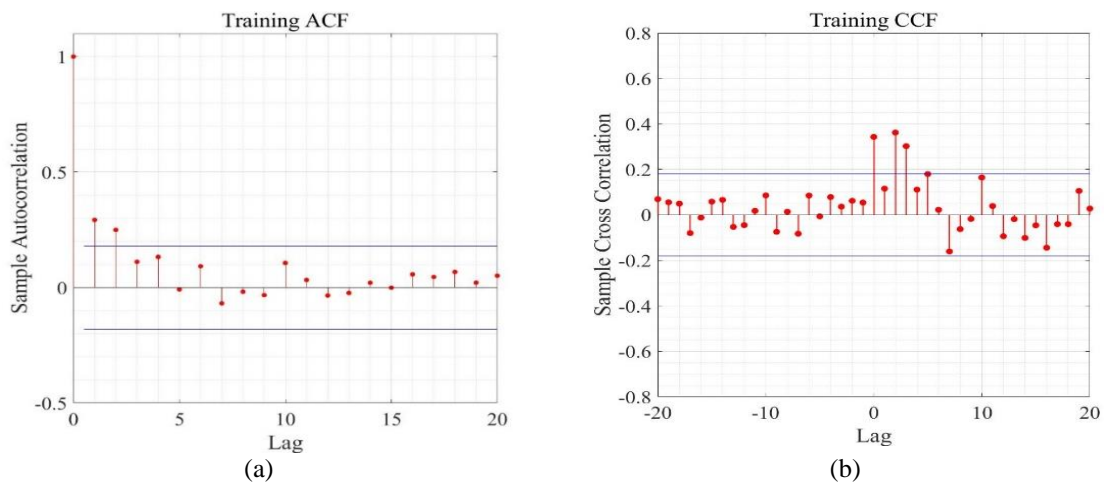


Figure 6. These figure are; (a) ACF, (b) CCF tests for training

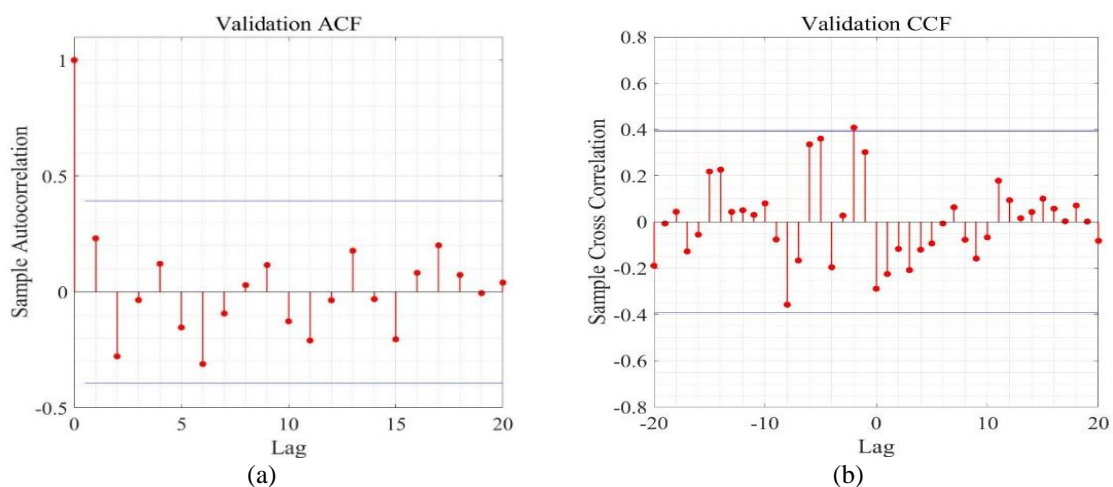


Figure 7. These figure are; (a) ACF, (b) CCF tests for validation



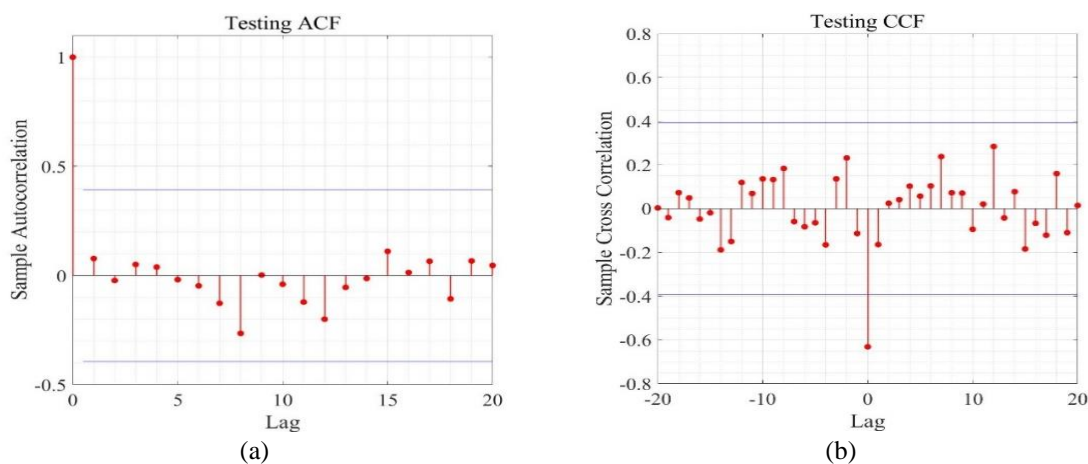


Figure 8. These figure are; (a) ACF, (b) CCF tests for testing

#### 4. CONCLUSION

The study has set out to analyze the correlation between DO, BOD, COD, NH<sub>3</sub>-N, pH, temperature, SS, DS, TS, and turbidity from Air Busuk River with WQI, and develop a NARX-based WQI model based on significant water quality parameters. Based on the analysis, DO, BOD, and NH<sub>3</sub>-N has shown correlation and is used to develop the WQI model. The NARX-based model is then developed using two-hidden layer MLP network with TDL of seven. The model which is based on three chemical parameters has shown satisfactory performance with overall regression of 0.8767 and passed the correlation function tests. Based on the results, the model has shown to be capable of predicting fluctuations in WQI by considering from the past fourteen weeks based on only three chemical parameters. Hence, these would serve as an early indicator to alert the authorities on the pending pollution that will reach Klang River with reduced cost of operations. Despite the successful development of model using artificial neural network and NARX structure, results indicate that there is still opportunity for improvement. The model was developed based on data that has been sampled at a frequency of two weeks. Therefore, predictions can only be performed using data from the past fourteen weeks. A more responsive model, however, should rely on data sampled at higher frequency. Hence, the Department of Environment should use the findings as a motivation for increasing the rate of water quality measurement. These would allow future models to capture the unseen dynamics of chemical parameters which would lead more accurate predictions. Additional parameters related to external elements such as rainfall volume can also be considered in the WQI model. The proposed considerations, however, will be subject to future investigations. Thus far, this is the only study that proposed a model based on the NARX structure. Therefore, new developments related to improved sampling frequency, prediction performance, and adoption of other modelling methods will be benchmarked with this study.

#### ACKNOWLEDGEMENTS

Authors would like to thank the Department of Environment for providing data on Air Busuk River. This study is funded by the Institute of Research Management and Innovation, Universiti Teknologi MARA through the LESTARI SDG Triangle@UiTM Research Grant (600-RMC/LESTARI SDG-T 5/3 (131/2019)).

#### REFERENCES

- [1] O. T. Dede, I. T. Telci, and M. M. Aral, "The use of water quality index models for the evaluation of water surface quality: A case study for Kirmir Basin, Ankara, Turkey," *Water Quality, Exposure and Health*, vol. 5, pp. 41-56, February 2013, doi: 10.1007/s12403-013-0085-3.
- [2] Y. F. Huang, S. Y. Ang, K. M. Lee, and T. S. Lee, "Quality of water resources in Malaysia," in *Research Practices in Water Quality*, Ed. London: IntechOpen, vol. 3, pp. 65-94, 2015, doi: 10.5772/58969.
- [3] "National Water Quality Standards for Malaysia," *Department of Environment, Putrajaya*, 2019.
- [4] L. Zlatanovic, A. Moerman, J. P. van der Hoek, J. Vreeburg, and M. Blokker, "Development and validation of a drinking water temperature model in domestic drinking water supply systems," *Urban Water Journal*, vol. 14, no. 10, pp. 1031-1037, June 2017, doi: 10.1080/1573062X.2017.1325501.
- [5] W. Bi, *et al.*, "Evolution characteristics of surface water quality due to climate change and LUCC under scenario simulations: A case study in the Luanhe River Basin," *International Journal of Environmental Research and Public Health*, vol. 15, no. 8, pp. 1-18, August 2018, doi: 10.3390/ijerph15081724.

- [6] M. Arief, "Development of dissolved oxygen concentration extraction model using Landsat data case study: Ringgung coastal waters," *International Journal of Remote Sensing and Earth Sciences*, vol. 12, pp. 1-12, 2015, doi: 10.30536/ijreses.2015.v12.a2667.
- [7] N. Xiao, R. Wu, J. J. Huang, and P. R. Selvaganapathy, "Development of a xurographically fabricated miniaturized low-cost, high-performance microbial fuel cell and its application for sensing biological oxygen demand," *Sensors and Actuators B: Chemical*, vol. 304, pp. 128-131, February 2020, doi: 10.1016/j.snb.2019.127432Get.
- [8] J. Li, G. Luo, L. He, J. Xu, and J. Lyu, "Analytical approaches for determining chemical oxygen demand in water bodies: A review," *Critical Reviews in Analytical Chemistry*, vol. 48, no. 1, pp. 47-65, 2018, doi: 10.1080/10408347.2017.1370670.
- [9] A. Jaffar, N. M. Thamrin, M. S. A. M. Ali, M. F. Misnan, and A. I. M. Yassin, "The influence of physico-chemical parameters to determine water quality: A review," *Journal of Electrical and Electronic Systems Research*, vol. 17, pp. 108-113, 2020.
- [10] V. S. Kale, "Consequence of temperature, pH, turbidity and dissolved oxygen water quality parameters," *International Advanced Research Journal in Science, Engineering and Technology*, vol. 3, no. 8, pp. 186-190, 2016.
- [11] A. R. Purwono, M. Hibbaan, and M. A. Budihardjo, "Ammonia-nitrogen (NH<sub>3</sub>-N) and ammonium-nitrogen (NH<sub>4</sub><sup>+</sup>-N) equilibrium on the process of removing nitrogen by using tubular plastic media," *Journal of Materials and Environmental Sciences*, vol. 8, pp. 4915-4922, 2017.
- [12] L. Huff, C. Delos, K. Gallagher, and J. Beaman, "Aquatic Life Ambient Water Quality Criteria for Ammonia-Freshwater 2013," *US Environmental Protection Agency*, Washington D.C., 2013.
- [13] G. S. Bilotta and R. E. Brazier, "Understanding the influence of suspended solids on water quality and aquatic biota," *Water Research*, vol. 42, no. 12, pp. 2849-2861, June 2008, doi: 10.1016/j.watres.2008.03.018Get.
- [14] Tiyasha, A. M. Tung, and Z. M. Yaseen, "A survey on river water quality modelling using artificial intelligence models: 2000-2020," *Journal of Hydrology*, vol. 585, p. 124670, June 2020, doi: 10.1016/j.jhydrol.2020.124670Get.
- [15] F. Kozisek, "Regulations for calcium, magnesium or hardness in drinking water in the European Union member states," *Regulatory Toxicology and Pharmacology*, vol. 112, p. 104589, 2020, doi: 10.1016/j.yrtph.2020.104589Get.
- [16] R. Noori, R. Berndtsson, M. Hosseinzadeh, J. F. Adamowski, and M. R. Abyaneh, "A critical review on the application of the National Sanitation Foundation Water Quality Index," *Environmental Pollution*, vol. 244, pp. 575-587, January 2019, doi: 10.1016/j.envpol.2018.10.076Get.
- [17] T. Nasrabadi, H. Ruegner, Z. Z. Sirdari, M. Schwientek, and P. Grathwohl, "Using total suspended solids (TSS) and turbidity as proxies for evaluation of metal transport in river water," *Applied Geochemistry*, vol. 68, pp. 1-9, March 2016, doi: 10.1016/j.apgeochem.2016.03.003.
- [18] T. Leziart, Pierre-Marie D. de la Rochere, R. Cheswick, P. Jarvis, and A. Nocker, "Effect of turbidity on water disinfection by chlorination with the emphasis on humic acids and chalk," *Environmental Technology*, vol. 40, no. 13, pp. 1734-1743, March 2019, doi: 10.1080/09593330.2019.1585480.
- [19] K. W. Chau, "A review on integration of artificial intelligence into water quality modelling," *Marine Pollution Bulletin*, vol. 52, no. 7, pp. 726-733, July 2006, doi: 10.1016/j.marpolbul.2006.04.003.
- [20] M. Hameed, S. S. Sharqi, Z. M. Yaseen, H. A. Afan, A. Hussain, and A. Elshafie, "Application of artificial intelligence (AI) techniques in water quality index prediction: A case study in tropical region, Malaysia," *Neural Computing and Applications*, vol. 28, pp. S893-S905, 2017, doi: 10.1007/s00521-016-2404-7.
- [21] A. A. Najah, A. El-Shafie, O. A. Karim, and O. Jaafar, "Water quality prediction model utilizing integrated wavelet-ANFIS model with cross-validation," *Neural Computing and Applications*, vol. 21, pp. 833-841, 2012, doi: 10.1007/s00521-010-0486-1.
- [22] C. J. Perez, M. A. Vega-Rodriguez, K. Reder, and M. Flörke, "A multi-objective artificial bee colony-based optimization approach to design water quality monitoring networks in river basins," *Journal of Cleaner Production*, vol. 166, pp. 579-589, November 2017, doi: 10.1016/j.jclepro.2017.08.060.
- [23] F. J. Chang, Yu-Hsuan Tsai, Pin-An Chen, A. Coynel, and G. Vachaud, "Modeling water quality in an urban river using hydrological factors," *Journal of Environmental Management*, vol. 151, pp. 87-96, March 2015, doi: 10.1016/j.jenvman.2014.12.014.
- [24] S. de Franciscis, S. Johnson, and J. J. Torres, "Enhancing neural-network performance via assortativity," *Physical Review E*, vol. 83, no. 3, p. 036114, 2011, doi: 10.1103/PhysRevE.83.036114.
- [25] "Environmental Quality Report 2018," *Department of Environment*, Putrajaya, Malaysia, 2018.
- [26] D. S. Moore and S. Kirkland, "The Basic Practice of Statistics," New York: Macmillan Learning, 2018.
- [27] A. Ahmad, N. Javid, N. Alrajeh, Z. A. Khan, U. Qasim, and A. Khan, "A modified feature selection and artificial neural network-based day-ahead load forecasting model for a smart grid," *Applied Sciences*, vol. 5, pp. 1756-1772, 2015, doi: 10.3390/app5041756.
- [28] A. H. Jahidin, M. N. Taib, N. Md Tahir, and M. S. A. M. Ali, "IQ classification via brainwave features: Review on artificial intelligence techniques," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 5, no. 1, pp. 84-91, February 2015.
- [29] Z. Boussaada, O. Curea, A. Remaci, H. Camblong, and N. M. Bellaaj, "A nonlinear autoregressive exogenous (NARX) neural network model for the prediction of the daily direct solar radiation," *Energies*, vol. 11, no.3, p. 620, 2018, doi: 10.3390/en11030620.

- [30] I. Mohd Yassin, M. N. Taib, and R. Adnan, "Extended analysis of BPSO structure selection of nonlinear autoregressive model with exogenous inputs (NARX) of direct current motor," *Songklanakarin Journal of Science and Technology*, vol. 36, no. 6, pp. 683-699, 2014.
- [31] Y. Amellas, O. El Bakkali, A. Djebil, and A. Echchelh, "Short-term wind speed prediction based on MLP and NARX network models," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 18, no. 1, pp. 150-157, April 2020, doi: 10.11591/ijeecs.v18.i1.pp150-157.
- [32] J. M. P. Menezes Jr. and G. A. Barreto, "Long-term time series prediction with the NARX network: An empirical evaluation," *Neurocomputing*, vol. 71, no. 16-18, pp. 3335-3343, October 2008, doi: 10.1016/j.neucom.2008.01.030Get.
- [33] F. S. Ab Gani, M. K. Nordin, A. I. M. Yassin, I. P. Ibrahim, and M. S. A. M. Ali, "Electrocardiogram profiling of myocardial infarction history using MLP and HMLP networks," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 17, no. 1, pp. 183-190, January 2020, doi: 10.11591/ijeecs.v17.i1.pp183-190.
- [34] M. S. A. M. Ali, A. Jahidin, M. Taib, N. M. Tahir, and I. Yassin, "Classification of Kolb's learning styles using EEG sub-band spectral centroid frequencies and artificial neural network," *Asian Journal of Scientific Research*, vol. 9, no. 5, pp. 234-241, 2016.
- [35] A. H. Jahidin, *et al.*, "Classification of intelligence quotient using EEG sub-band power ratio and ANN during mental task," in *2013 IEEE Conference on Systems, Process & Control (ICSPC)*, 2013, pp. 204-208, doi: 10.1109/SPC.2013.6735132.
- [36] M. S. Sulaiman and Z. Saad, "Classification of healthy and white root disease infected rubber trees based on relative permittivity and capacitance input properties using LM and SCG artificial neural network," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 19, no.1, pp. 222-228, July 2020, doi: 10.11591/ijeecs.v19.i1.pp222-228.
- [37] M. S. Salman, A. Eleyan, and B. Al-Sheikh, "Discrete wavelet transform-based RI adaptive algorithm for system identification," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 3, pp. 2383-2391, June 2020, doi: 10.11591/ijece.v10i3.pp2383-2391
- [38] T. Igasaki, S. Shimai, and M. Kobayashi, "Measuring cardiorespiratory information in sitting position using multiple piezoelectric sensors," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 6, no. 1, pp. 132-138, April 2017, doi: 10.11591/ijeecs.v6.i1.pp132-138.
- [39] L. Tolosi and T. Lengauer, "Classification with correlated features: Unreliability of feature ranking and solutions," *Bioinformatics*, vol. 27, no. 14, pp. 1986-1994, July 2011, doi: 10.1093/bioinformatics/btr300.