Ecological Indicators 156 (2023) 111190

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Enhancing river health monitoring: Developing a reliable predictive model and mitigation plan

Syahida Farhan Azha[a,*], Lariyah Mohd Sidek[a,*], Zainal Ahmad[b], Jie Zhang[c], Hidayah Basri[a], Mohd Hafiz Zawawi[a], Nurshahira Mohammad Noh[a], Ali Najah Ahmed[a]

[a] Institute of Energy Infrastructure (IEI) and Department of Civil Engineering, College of Engineering, Universiti Tenaga Nasional (UNITEN), 43000 Kajang, Selangor Darul Ehsan, Malaysia

[b] School of Chemical Engineering, Engineering Campus, Universiti Sains Malaysia (USM), 14300 Nibong Tebal, Penang, Malaysia

[c] School of Engineering, Merz Court, Newcastle University, Newcastle upon Tyne NE1 7RU, United Kingdom

ARTICLE INFO

Keywords:
Water quality index
Feedforward artificial neural network
Mitigation plans
Batu Pahat River
SDG
Water Energy Security

ABSTRACT

The escalating environmental harm inflicted upon rivers is an unavoidable outcome resulting from climate fluctuations and anthropogenic activities, leading to a catastrophic impact on water quality and thousands of individuals succumb to waterborne diseases. Consequently, the water quality monitoring stations have been established worldwide. Regrettably, the real-time evaluation of Water Quality Index (WQI) is hindered by the intricate nature of off-site water quality parameters. Thus, there is a pressing need to create a precise and robust water quality prediction model. The dynamic and non-linear characteristics of water quality parameters pose significant challenges for conventional machine learning algorithms like multi-linear regression, as they struggle to capture these complexities. In this particular investigation, machine learning model called Feedforward Artificial Neural Networks (FANNs) was employed to develop WQI prediction model of Batu Pahat River, Malaysia exclusively utilizing on-site parameters. The proposed method involves a consideration of whether to include or exclude parameters such as BOD and COD, which are not measured in real time and can be costly to monitor as model inputs. Validation accuracy values of 99.53%, 97.99%, and 91.03% were achieved in three different scenarios: the first scenario utilized the full input, the second scenario excluded BOD, and the third scenario excluded both BOD and COD. It was suggested that the model has better predictive power between input variables and output variables. Factor contributed to river pollution has been identified and mitigation plan for Batu Pahat river pollution has been proposed. This could provide an effective alternative to compute the pollution, better manage water resources and mitigate negative impacts of climate change of river ecosystems.

1. Introduction

The mandate of the United Nations has specified the ambitious 17 Sustainable Development Goals (SDGs) through the development pathway by balancing the economic growth with social inclusion and earth's environmental sustainability, end poverty, ensure societies is in conditions of peaceful and prosperity, as well as to realize the human rights (DSDG, 2023). The goals set out to tackle worldwide issues such as poverty, climate change, inequality, education, energy, water, peace, and justice, with the aim of achieving them by 2030. The progress towards these goals is measured using a comprehensive framework of indicators. The sixth Sustainable Development Goal (SDG) aims to achieve clean water and sanitation, with a particular focus on sub-section 6.3. This target aims to enhance water quality by reducing the amount of toxic waste, eliminating dumping and reducing the release of hazardous chemicals and materials. It also aims to reduce the amount of untreated wastewater and substantially increase recycling and the safe reuse of water. Further into sub-section 6.6, everybody has to take part in protecting and restoring water-related ecosystems, including mountain, forest, wetlands, aquifers, lakes and river (DSDG, 2023). Therefore, this noble vision requires support from all human being for a better future and generation.

Nonetheless, due to over-exploitation and deterioration of natural sources especially river water, the environmental degradation issue has become a major cause for concern, particularly due to the rising levels of pollutants which have put it in jeopardy. Human activities, forest

* Corresponding authors.
E-mail addresses: syahida.farhan@uniten.edu.my (S.F. Azha), lariyah@uniten.edu.my (L.M. Sidek).

https://doi.org/10.1016/j.ecolind.2023.111190
Received 8 August 2023; Received in revised form 26 October 2023; Accepted 29 October 2023
Available online 6 November 2023

exploration and agriculture activities have causes and gives negative effect in water quality that can harm the health and public safety, which then damaging the ecosystem (Fitri et al., 2020; Hazritauding and Adnan, 2022; Moradi et al., 2022). River pollution can stem from both point sources, such as discharges from industrial and sewage treatment plants, as well as non-point sources, including surface water runoff from agricultural land use, housing areas, commercial developments, and industries (Khullar and Singh, 2021). Efficient management of watersheds is crucial to ensure effective control and management of rivers. Rivers play a crucial role in the ecosystem and provide numerous benefits to both human and wildlife. The key importance of rivers includes as a primary source of fresh water (drinking, irrigation), biodiversity to support a diverse array of aquatic life, provide transportation of goods and people since thousands of years, opportunities for a range of recreational activities (fishing, boating), helps in erosion and flood control by carrying sediment downstream and prevent flooding, respectively and last but not least for energy generation such as hydroelectric power as a renewable source of energy (Fitri et al., 2020; Ho et al., 2019; Sidek et al., 2022). Thus, due to these beneficial roles of river towards mankind and ecosystem, the river water quality has to be monitored accordingly and precisely.

The Water Quality Index (WQI) is a numerical metric used to evaluate and describe the overall quality of water in a given water body such as river, lake, ocean, etc (Nasir et al., 2022; Venkata Vara Prasad et al., 2020). The WQI is calculated by taking into account on several physical, chemical, and biological parameters/factors (turbidity, pH, total nitrogen, COD, BOD) that are known to impact the suitability of water for various uses, such as consumption/drinking, household chores, irrigation, recreation, and aquatic life. It is typically calculated by assigning a weight or score to different water quality parameters and then combining these score into a single value. The WQI provides a comprehensive, easy-to-understand and practical assessment of water quality and can be used by water management agencies, public health organizations, and other stakeholders to track changes in water quality over time and to make informed decisions about water use and management (Khouri and Bashar Al-Moufti, 2022; Mohammed et al., 2022). In addition, by controlling and monitoring the quality of water, it could help in protecting human health, save the environment and close-control and micromanage the pollution. As a result, the general public will be able to take extra care and attention to the state of their local water bodies, and indirectly contribute to ensuring the local water bodies are maintained and protected (Marselina et al., 2022; Nong et al., 2023).

However, evaluating multiple water quality parameters can be time-consuming and costly, as it requires collecting water samples and conducting laboratory analyses, and on the basis of large calculation volume, the process required lots of effort and time (Liu et al., 2019). Moreover, the accuracy and reliability of the results depend on the quality of the sampling and analytical methods used as well as the equipment and expertise of the personnel involved. In recent years, Artificial Intelligence (AI) provides promising prospects to enhance the identification and forecasting of environmental issues, including water quality management, through automation. Several AI algorithms have been evaluated to analyze water quality data obtained over a prolonged duration and develop a dependable method for predicting water quality as accurately and efficiently as possible. Machine Learning (ML) is a robust tool for assessing water quality, offering a more efficient and accurate alternative to traditional methods of calculating Water Quality Index (WQI) (Ahmed et al., 2019; Zaini et al., 2022). Recent advancements in ML techniques have been utilized to tackle the aforementioned challenges in water quality prediction. These methods are capable of efficiently addressing highly nonlinear issues without requiring prior knowledge of the physical processes involved in the system being studied. One of the frequently employed ML models is the Artificial Neural Networks (ANNs) technique, which can precisely approximate the nonlinear relationship between inputs and outputs through a network architecture consisting of multiple layers of interconnected nodes, and trained on large amounts of historical data. Empirical studies have demonstrated the ANN method's effective in enhancing water quality prediction accuracy, establishing it as a crucial alternative (Chen et al., 2020; Ezemagu et al., 2021; Khoi et al., 2022).

Feedforward Artificial Neural Networks (FANNs) are a type of machine learning models inspired by the structure and function of the human nervous system and are intended to replicate its behavior (Chen et al., 2020). FANN is a specific type of ANN that uses a forward-only connection scheme, allowing for quick and accurate predictions, making it suitable for water quality monitoring and air quality management, lake and reservoir modelling, and hydrologic forecasting, due to their ability to learn from historical data, flexibility, and adaptability. The FANN with a multi-input single-output (MISO) structure is a type of neural network that can take multiple inputs and produce a single output. In this architecture, the input layer receives data from multiple sources, and each input is processed independently through a set of hidden layers before producing a single output in the output layer (Djarum et al., 2022). The MISO structure is often used in applications where there are multiple input variables that affect a single output variable, such as in water quality monitoring, where various environmental parameters influence the quality of water. This architecture can handle large amounts of data and identify complex patterns in input variables to produce accurate predictions, automating the prediction process and reducing the need for human intervention. Unlike other types of ANNs, FANN does not allow for feedback loops, which can cause instability and slow down the training process. While other types of ANNs may be better suited for other problems, FANNs are more computationally efficient for quick and accurate predictions.

In recent studies, the ANN models have been developed to predict various water quality parameters, such as DO and BOD in order to emphasize and enhance the applicability and reliability of water quality prediction/modelling (Chen et al., 2020; Maier et al., 2010). Abba et al. utilized monthly data from 1999 to 2005 to forecast the dissolved oxygen (DO) downstream of the Yamuna River in Agra city, India (Abba et al., 2018). In a study comparing multilinear regression (MLR), adaptive neuro fuzzy inference system (ANFIS), and artificial neural network (ANN) models in predicting dissolved oxygen (DO), the input variables used were DO, pH, BOD, and water temperature. Results showed that ANN was the most accurate predictor of DO, achieving up to 94% accuracy compared to ANFIS and MLR, which had an average accuracy of 81%. Another study by Kanda and Kosgei (2016) on the Nzoia River in the Lake Victoria basin in Kenya used multilayer perception (MLP), a type of feedforward backpropagation ANN, to predict DO using monthly data from 2003 to 2013 and input variables including pH, turbidity, temperature, and electrical conductivity(Kanda and Kosgei, 2016). The exclusion of pH resulted in acceptable DO prediction, even though the number of input neurons ranged from 1 to 4 and the number of neurons in the hidden layer varied from 22 to 28. Sarkar and Pandey predicted DO in the River Yamuna, India, using the feedforward backpropagation network architecture and discharge, temperature, pH, BOD, and COD as input variables. They concluded that the ANN model performed best with optimum input variables, and values above or below the optimum would lead to overfitting and inaccurate prediction (Sarkar and Pandey, 2015). In previous studies, various machine learning approaches have been utilized to predict water quality parameters. Khan and See (2016) developed an ANN-based WQ model that included dissolved oxygen, chlorophyll, conductivity, and turbidity (Khan and See, 2016). Abyaneh (2014) utilized both ANN and regression techniques to forecast COD (Abyaneh, 2014). Meanwhile Sakizadeh (2015) applied ANN in conjunction with Bayesian regularization (Sakizadeh, 2015), Mohammadpour et al. (2015) focusing on constructed wetlands (Mohammadpour et al., 2015), (Khoi et al., 2022) delved into surface water quality using four ANN-based algorithms (multilayer perceptron, radial basis function, deep feed-forward neural network, and convolutional neural network), and additionally (Kadam et al., 2019; Singh et al., 2021) ventured into the prediction of groundwater quality for drinking to estimate the water quality index (WQI). These studies collectively contribute to our understanding of water quality assessment through diverse methodologies. Therefore, there are numerous other studies that employed different ANN architecture models and water quality parameters to simulate and predict WQI due to the scarcity of water quality monitoring data and the intricate and nonlinear nature of interactions between water quality parameters.

Despite the fact that numerous studies have been conducted on the Batu Pahat River, encompassing various scopes of research, these remains a dearth of investigations focusing on the prediction on WQI utilizing diverse artificial intelligence approaches. Previous assessment on Batu Pahat river studied on textural characteristics and sedimentation at eroded and deposited coastline (Wan Mohtar et al., 2017), sediment load and sediment properties (Mokhtar et al., 2022; Tjahjanto et al., 2008), Tidal effect on Suspended Sediment dispersion (Ismail, 2007), a normal water quality assessment and water quality index (Salim and Kasmin, 2022; Sidek et al., 2022), flood water level modelling (Adnan et al., 2012), flood estimation studies (Mohammed et al., 2021), flood prone area detection using GIS and water balance model (Rahman and Yusof, 2015), application of total maximum daily load (TMDL) (Adnan et al., 2022), assessment of water quality parameters due to high or low flow of river (Hazritauding and Adnan, 2022), prediction of future temperature and rainfall characteristics (Latiff et al., 2021), wave hindcasting (Bateni et al., 2009) and few more can be searched in literatures. Therefore, the research on the predicting WQI values using ML algorithm in which to improve the prediction accuracy is very relevant and novel topic specifically for the assessment of Batu Pahat river.

The development of a predictive WQI model using a feedforward neural network marks a pivotal stride in fortifying the ability to address the challenges posed by deteriorating river water quality. The study's primary goal was to precisely forecasting the WQI model to emerge as a potent tool for pinpointing areas that demand immediate pollution control interventions. These interventions are pivotal not only for safeguarding the environment but also for protecting human health and the economy. The mitigation plan for polluted rivers encompasses a comprehensive strategy that entails the identification of pollution sources, the establishment of reduction targets, and the implementation of targeted pollution control measures. Crucially, the effectiveness of such mitigation plan hinges on the accuracy of WQI predictions. Regular monitoring and ongoing assessment of progress in curtailing pollution levels, combined with heightened public awareness and collaborative efforts among stakeholders, amplify the chances of success in implementing the mitigation plan. In sum, mitigating the repercussions of polluted river water necessitates a holistic approach. Accurate WQI prediction, as demonstrated in this study, stands as a linchpin in this holistic endeavor, enhancing capability to monitor, understand, and proactively address river health concerns.

2. The study area and materials

2.1. The Batu Pahat River Basin

Located in the State of Johor, Batu Pahat serves as the administrative capital of the Batu Pahat District, situated southeast of Muar, southwest of Kluang, northwest of Pontian, and south of Segamat. The population of almost half of million people. Batu Pahat River basin is located within the longitudes 102° 47′E and 103° 16′E, and within the latitudes 1° 46′N and 2° 25′N. The basin area is 2,049 km^{2,} and the length of the main river is 125.16 km. Batu Pahat River basin is composed of four (4) districts namely Batu Pahat District, Kluang District, Muar District & Segamat District. The basin also consists of twenty-three (23) sub-districts. The river basin in which Batu Pahat is the major town extends from the northern to the southern part of the district before finally discharging into the Straits of Malacca. Several other small towns, including Sri Medan, Yong Peng, Senggarang, Parit Raja, Air Hitam, Tongkang Pecah, Parit Yaani, and Chaah, are also situated within this river basin. The Simpang river system consists of two main tributaries, namely Simpang Kiri River and Simpang Kanan River. The Simpang Kiri River has an elongated basin with an area coverage of 815 km^2, and its upper reaches are known as Lenik River. On the other hand, Simpang Kanan River has a sub-basin area of 645 km^2 and is drained by two main tributaries, Bekok River and Sembrong River.

The Batu Pahat River basin has been dealing with serious ecological degradation due to incessant development, industrialization, the presence of a tide and the encroachment of development into the floodplain. The catastrophic events of flooding in the Batu Pahat River basin were most severe in December 2006 and January 2007. During this period, Batu Pahat River and its related tributaries, such as Simpang Kanan River, Simpang Kiri River, Bekok and Sembrong River were the most affected. In total, 53,000 people were displaced and disrupted the transportation on major roads.

The Batu Pahat River (marked in Fig. 1 as SBP) basin is divided into 25 sub-basins, with their own distinct characteristics in terms of size, land use activities, topography, and main rivers and tributaries. SBP 01 being the largest covering 841.486 ha (37.55% of the whole basin). SBP 02 and SBP 03 are smaller, covering approximately 29.86% and 15.06% of the basin respectively, while SBP 04 to SBP 09 cover less than 10% of the basin each. Each sub-basin has a main river and several tributaries. SBP 02 and SBP 03 each have an existing dam, the Bekok Dam and the Sembrong Dam respectively. The sub-basins have varying land use activities, with SBP 01 being mostly covered by forest and agricultural land use, while SBP 05, has a host of transportation land use activities while SBP 22 and SBP 25 still has a lot of vacant land. SBP 06 to SBP 12 are relatively small sub-basins separated by main canals in Batu Pahat, with agricultural activities being the major land use activity in all sub-basins. The topography also varies, with SBP 01 being quite hilly due to its upstream location where the highest peak reaches almost 700 m, while the other sub-basins are relatively flat with the highest peaks ranging from 3.1 m to 410 m. Agricultural activities are also one of the major land use activities in all sub-basins especially at SBP 06 till SBP 21. At SBP 23, even agricultural activities are also one of the major land use activities (32.40%) there also mostly vacant land (43.87%) at this sub-basin. Fig. 1 presents the map of Batu Pahat river basin and the signage of sub-basins were marked as SBP 01 to SBP 25.

Several site visits to the Batu Pahat river basin were conducted to gain a better understanding of the surrounding and did certain analyses in order to verify at the site especially in terms of location of sand mining, and flood mark. Pictures 1–11 depict various scenes from the Batu Pahat River basin, including the river mouth and sand mining activities. As depicted by a Fig. 2, Picture No. 1 (small image in figure) shows the Batu Pahat River at Batu Pahat Town, while Picture No. 2 captures the intersection of the Bekok and Sembrong Rivers, where two bay tidal control gates are located. Picture No. 4 and No. 5 are the locations of Pekan Ayer Hitam and Pekan Parit Raja, respectively. Picture No. 6 captures the Bekok River at Yong Peng Town. Picture No. 7 is the confluence point of the Simpang Kiri and Simpang Kanan Rivers, which is the start of the Batu Pahat River and where it begins its downstream flow. Picture No. 8 is the river mouth of the Batu Pahat River, while Picture No. 9 shows one of the sand mining activities near Ayer Hitam River. Picture No. 10 is the location of the Sembrong River and Picture No. 11 is the location of the Simpang Kanan River, near Jalan Tongkang Pecah. Fig. 2 shows the Batu Pahat River basin layout of the pertinent pictures taken during the site visit via drone/ground for clearer view of the site.

2.2. The sampling point/data collection

A sample of the water was taken and data was collected by the Department of Environment (DOE) of Malaysia in accordance with the

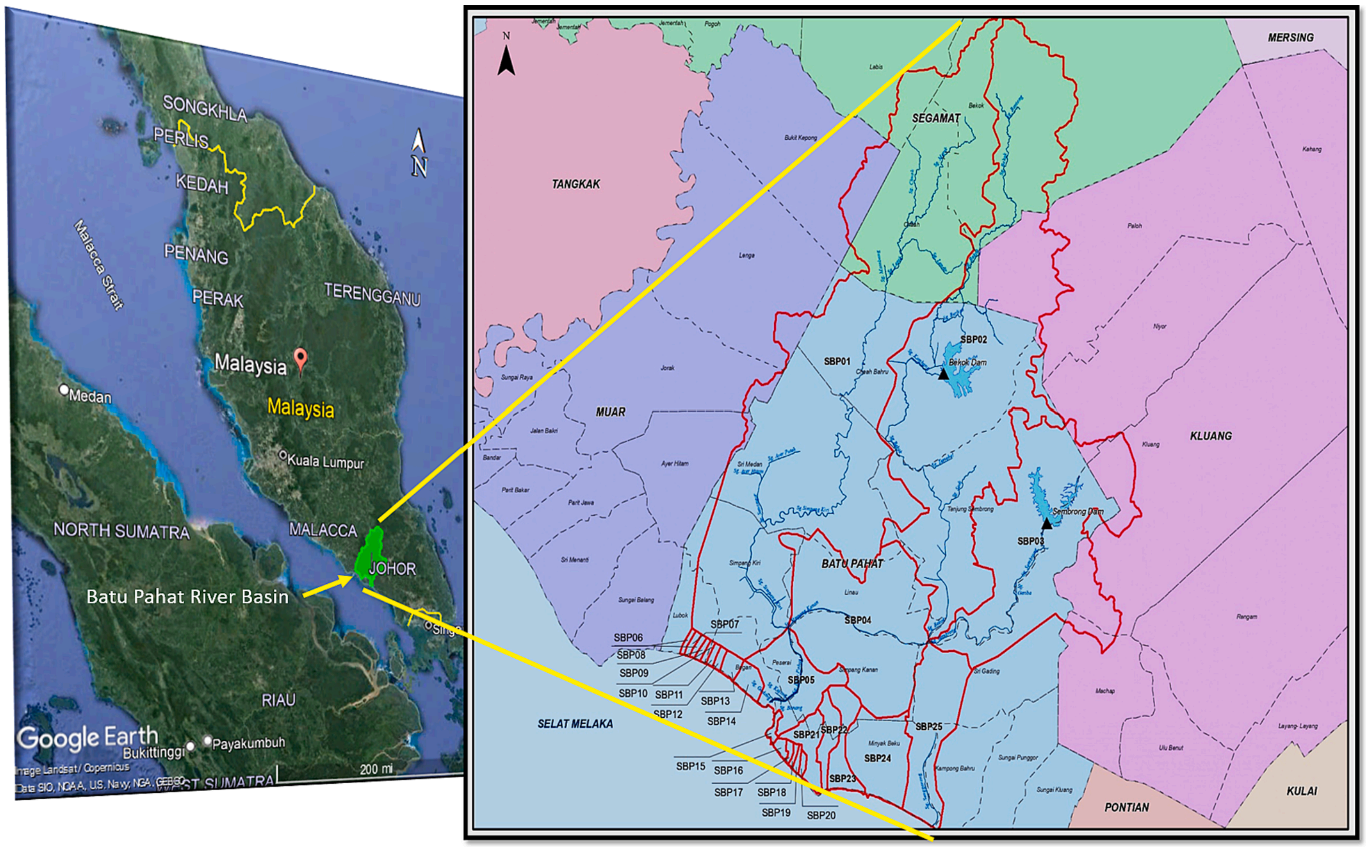


Fig. 1. Map of Batu Pahat River Basin from SBP 01 to SBP 25.

prescribed procedures. The collected data were from 25 sub-basin monitoring station for about 25 water quality variables encompassing dissolved oxygen (DO), biochemical oxygen demand (BOD), chemical oxygen demand (COD), ammoniacal nitrogen (NH_3-N), suspended solids (SS), dissolved solids (DS), pH, temperature, calcium (Ca), iron (Fe), chloride (Cl^-), phosphate (PO_4^-), potassium (K), conductivity, salinity, magnesium (Mg), sodium (Na), turbidity, total solids (TS), nitrate (NO_3^-), arsenic (As), chromium (Cr), zinc (Zn), *Escherichia coli* (*E. coli*) and coliform. DOE carries out in situ measurements of several water quality parameters such as DO (mg/l), turbidity (NTU), conductivity (uS/cm), salinity (ppt), pH, and temperature. However, laboratory analysis is required to determine the remaining chemical and biological parameters. Due to the lengthy analysis process, BOD and COD variables are not used in this study for real-time WQI prediction. BOD, which stands for BOD5 or BOD7, requires at least 5–7 days to obtain results, while COD analysis takes several hours to complete. Therefore, there have 6 out of 25 water quality parameters of Batu Pahat river were applied for WQI calculation (DO, BOD, COD, NH_3-N, pH, and temperature) which were taken since 2008 to 2018 (bimonthly from January 2008 to July 2018) and consisted of 1181 water quality samples. The flow chart of overall research included WQI forecasting and river mitigation plan were finalized such in Fig. 3.

2.3. WQI calculation

In accordance with the National Water Quality Standards (NWQS), the Water Quality Index (WQI) measures water pollution and suitability. The DOE of Malaysia introduced the WQI monitoring approach since 1978 (Ahmad et al., 2016). In this approach, six variables were taken into consideration, including biological oxygen demand (BOD), chemical oxygen demand (COD), dissolved oxygen (DO), ammoniacal nitrogen (NH3-N), suspended solid (SS), and pH value. The resulting WQI values provide a measure of water contamination that can be used by policymakers and environmentalists to evaluate water quality. After laboratory analysis results were recorded, the WQI equation presented in Equation 1 was used to calculate the water quality index for a given sample. There are six typical physicochemical water quality parameters used;

$$WQI = (0.22*SIDO) + (0.19*SIBOD) + (0.16*SICOD) + (0.15*SIAN) + (0.16*SISS) + (0.12*SIpH) \quad (1)$$

The equation in Table 1 are used to measure all the sub-indices specified in Equation 1. The sub-indices for different parameters have different ranges (Salim and Kasmin, 2022). An index value ranging from zero to one hundred is produced through the assessment and valuation of these parameters. The resulting WQI value allows for the categorization of water quality into five classes: excellent (Class I), good (Class II), moderate (Class III), poor (Class IV), and very poor (Class V), which are presented in Table 2. Based on the index value, the river water quality is then classified into three main categories: clean, slightly polluted, and polluted, as outlined in Table 3.

3. Feedforward artificial neural network (FANN) model development

An iterative process is typically used to choose the appropriate model architecture or topology in order to achieve the best model structure. This process involves selecting a network with a specific structure, including the number of hidden nodes and transfer function, calibrating the chosen model, evaluating its performance, and then repeating the calibration and evaluation steps for different network configurations (López et al., 2022). The network configuration's performance is considered optimal when the results indicate the lowest MSE and high regression values for both training and testing data. The model is then

Fig. 2. Batu Pahat River Basin Site Visit Layout Map (1. Batu Pahat River in Batu Pahat Town, 2. Intersection of Bekok River and Sembrong River, 3. Upstream Parit Karjo, 4. Pekan Ayer Hitam, 5. Pekan Parit Raja, 6. Bekok River at Yong Peng Town, 7. Confluence of Simpang Kiri River and Simpang Kanan River (Starting of Batu Pahat River), 8. River mouth of Batu Pahat River, 9. Sand Mining at Ayer Hitam River, 10. Sembrong River, 11. Simpang Kanan River at Tongkang Pecah Road).

validated using an independent or unseen dataset to confirm the network's performance (Ahmed et al., 2019; Ghaedi and Vafaei, 2017).

A feedforward artificial neural network with a multi-input single-output (MISO) structure was chosen for the prediction model. The network takes multiple input variables to predict a single output variable, based on the results of the input feature selection (Bolboacă and Haller, 2023). The output of the FANN network is an estimation of WQI. The MISO structure was selected because it can significantly reduce the size and complexity of the neural network, thereby reducing the risk of overfitting. This structure also helps the model to identify the most important input variables, leading to a simpler and more interpretable model while maintaining high predictive accuracy. This approach is consistent with the current understanding of best practices in machine learning.

3.1. Model configuration

To achieve optimal performance of the feedforward artificial neural network (FANN), a number of factors were studied and divided into different elements. These elements included the determination of model architecture and other parameters such as the number of hidden nodes. The aim of the FANN-based model was to predict the water quality index (WQI) using input data that were grouped into three categories. One of the key features of the FANN architecture is the number of neurons in the hidden layer, which plays a critical role in modelling complex data. Determining the optimal number of neurons is important to prevent underfitting or overfitting of the data. In this study, the number of neurons in the hidden layers was determined through trial and error to meet the precision criteria (Juahir et al., 2004; López et al., 2022). Other important elements of the FANN model development included the selection of transfer functions, training algorithms, network training and validation, and network performance evaluation.

The input data was split into training, testing, and validation subsets to achieve optimal performance. The training subset was utilized to estimate the unknown connection weights while the testing subset was employed to identify the best network structure without overfitting the data. Meanwhile, the validation subset, or unseen data, was used to assess the model's generalization ability which consisted of 1181 water quality samples collected from various sub-basin states in Malaysia between 2008 and 2018. The complete data set was partitioned into training, testing, and validation sets using the 'divideint' function in Matlab, with a ratio of 70:15:15 (70 %: 827 samples for training, 15 % (177 samples) for testing and 15 % (177 samples), respectively. This allowed for repeated network training while maintaining the validation data as unseen data.

A trial and error approach was adopted in this study to select the number of hidden nodes in the ANN by incrementally varying their number. According to Hecht-Nielsen (1987), the number of hidden nodes, M, should range between I and 2I + 1, where I represents the number of input nodes (Hecht-Nielsen, 1987). M should not be less than the maximum of I/3 and the number of output nodes (Kavzoglu and Mather, 2003). For this study, the number of neurons in the hidden layer was varied between 6 and 25, and 15 neurons were found to be the optimal number. The selection of the learning rate value does not follow a specific rule, but (Sahoo et al., 2005) showed that unstable learning occurs for rates greater than 0.035. Therefore, a learning rate of 0.05 was chosen and kept constant during training. The training process stops when the iteration number reaches the maximum number of epochs, the

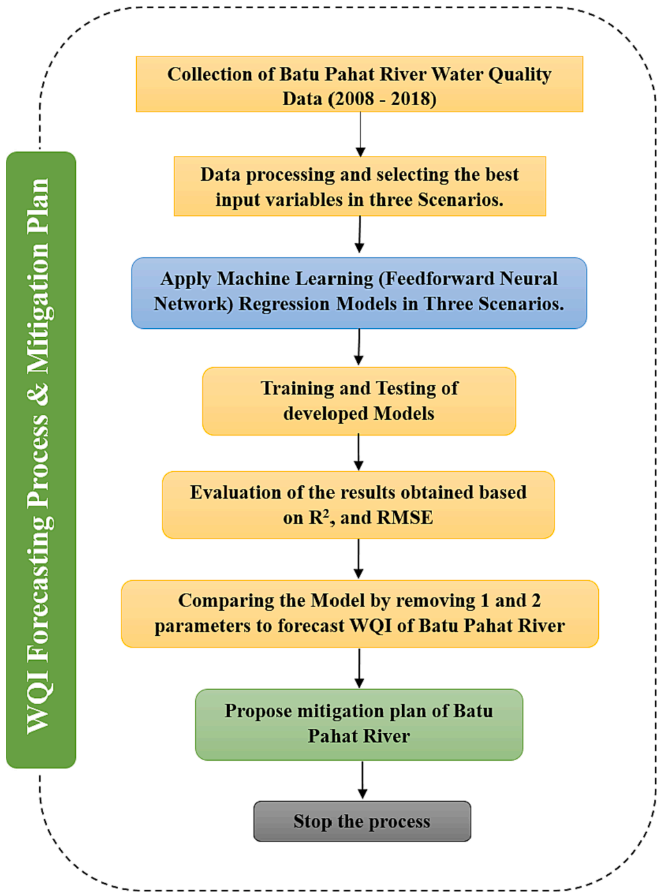


Fig. 3. Flow Chart of WQI forecasting and river mitigation plan methods.

Table 1
The equations for the estimation of various sub-index values ("DOE," 2020).

Parameter	Value	Sub-index equation
DO (in % saturation)	for $x \leq 8$; for $x \geq 92$ for $8 < x < 92$	$SIDO = 0$, $SIDO = 100$, $SIDO = -0.395 + 0.030x^2 - 0.00020x^3$,
BOD	for $x \leq 5$ for $x > 5$	$SIBOD = 100.4–4.23x$, $SIBOD = 108*e\ (-0.055x) - 0.1x$,
COD	for $x \leq 20$ for $x > 20$	$SICOD = -1.33x + 99.1$; $SICOD = 103*e\ (-0.0157x) - 0.04x$,
NH_3-N	for $x \leq 0.3$ for $0.3 < x < 4$ for $x \geq 4$	$SIAN = 100.5 - 105x$; $SIAN = 94*e\ (-0.573x) - 5\|x - 2\|$, and $SIAN = 0$,
SS	for $x \leq 100$; for $100 < x < 1000$ for $x \geq 1000$	$SISS = 97.5*e\ (-0.00676x) + 0.05x$, $SISS = 71*e\ (-0.0061x) - 0.015x$, $SISS = 0$
pH	for $x < 5.5$ for $5.5 \leq x < 7$; for $7 \leq x < 8.75$ for $x \geq 8.75$	$SIpH = 17.2–17.2x + 5.02x^2$, $SIpH = -242 + 95.5x - 6.67x^2$, $SIpH = -181 + 82.4x - 6.05x^2$, $SIpH = 536–77.0x + 2.76x^2$,

target error goal MSE, or the minimum performance gradient of 10-5, which was set according to (Juahir et al., 2004). During training, the connection weights were continually updated until they reached the defined iteration number or acceptable error (Kavzoglu and Mather, 2003).

To determine the optimal neural network for each output, several other parameters were varied during FANN network development. These parameters were listed in Table 4, and the following steps were taken:

The development of the FANN network involved varying some parameters to identify the optimal network for each output. The parameters were listed in Table 4, and the following steps were taken:

1. The fixed and varied FANN network parameters in Table 4, except for the training algorithm, were used. The Levenberg-Marquardt back-propagation (*trainlm*) algorithm was used to determine the best number of hidden neurons with the best transfer function.
2. The fixed FANN network parameters in Table 4 were used with the newly identified best transfer function for each output and case study. The training algorithm was varied as specified in Table 4, along with the number of hidden neurons, to discover the optimal network for each output.
3. The best network, with the determined number of hidden neurons, training algorithm, and transfer function, was selected as the final network setting for MISO, with various potential inputs, as shown in the next section.

The Levenberg-Marquardt algorithm is a highly regarded optimization technique used for training artificial neural networks which marries the steepest descent and Gauss-Newton algorithms (Sapna et al., 2012). This confluence of methods results in faster convergence times compared to alternate algorithms such as gradient descent. The Levenberg-Marquardt algorithm is especially effective for nonlinear model optimization, as it can accommodate non-convex optimization problems and dynamically adjust the step size to prevent divergence and oscillations. Additionally, it boasts robustness when working with noisy data - a common occurrence in real-world applications. Utilizing this algorithm can enhance the ability of the proposed model to optimize the neural network's weights and biases, ultimately reducing the difference between predicted and actual outputs. Overall, the Levenberg-Marquardt method is a suitable optimization algorithm for training models due to its ability to handle nonlinearities, adjust step size, and resistance to noisy data.

The neural network's hidden layer neurons utilize a logarithmic sigmoid activation function (Ranjan et al., 2023), whereas the output layer neurons utilize a linear activation function (Gao and Zhang, 2023). The number of hidden neurons is selected using cross-validation, whereby the number of hidden nodes is varied up to 15 to calculate corresponding mean squared errors (MSE) and R^2 values for both training and testing data sets. These values are then plotted against the number of hidden nodes, and the network that yields the lowest MSE value on both training and testing data sets is identified as the optimal network topology. Furthermore, the performance of the developed models is evaluated using MSE on unseen validation data as the criterion.

During the investigation of predictive modelling for estimating the WQI of Batu Pahat River, six variables (BOD, COD, DO, SS, pH, and NH_3-N) were used as inputs. Before feeding the data into the model, the variables were normalized using the Z-transformation, which involves subtracting the mean from all values and dividing them by the standard deviation. This transformation ensures that the original data distribution is preserved and that the modeling is not affected by outliers, as explained in (Kotu and Deshpande, 2014).

Table 2
DOE Water Quality Index Classification ("DOE," 2020).

Parameter	Unit	Class				
		I	II	III	IV	V
Ammoniacal Nitrogen	Mg/L	<0.1	0.1–0.3	0.3–0.9	0.9–2.7	>2.7
Biological Oxygen Demand	Mg/L	<1	1–3	3–6	6–12	>12
Chemical Oxygen Demand	Mg/L	<10	10–25	25–50	50–100	>100
Dissolved Oxygen	Mg/L	>7	5–7	3–5	1–3	< 1
pH	–	>7.0	6–7	5.0–6.0	< 5.0	>5.0
Total Suspended Solid	Mg/L	<25	25–50	50–150	150–300	>300
Water Quality Index (WQI)		>92.7	76.5–92.7	51.9–76.5	31.0–51.9	<31.0

Table 3
DOE Water Quality Classification based on Water Quality Index ("DOE," 2020).

Sub Index & Water Quality Index	Index Range		
	Clean	Slightly Polluted	Polluted
Biological Oxygen Demand	91–100	80–90	0–79
Ammoniacal Nitrogen	92–100	71–91	0–70
Suspended Solid (SS)	76–100	70–75	0–69
Water Quality Index (WQI)	81–100	60–80	0–59

Table 4
FANN network parameters.

Fix parameters	
Learning rate	0.05
Epochs	1000
Target error goal	10^{-5}
Minimum performance gradient	10^{-5}
Varied parameters	
Number of hidden neuron	6 to 25 (15) optimum based on training and validation data
Transfer function (first layer)	Log-sigmoid (*logsig*)
Transfer function (second layer)	Linear (*purelin)*
Training algorithm	Levenberg Marquardt backpropagation (*trainlm*)

3.2. Output and error analysis

To comprehensively assess and scrutinize the errors in the network's output, this study employed two distinct statistical methods for analysis. These methods encompass the mean square error (MSE), and coefficient of determination (R2). These metrics were used to evaluate the disparity between the network's output and the target output (Abba et al., 2020). Mean Square Error (MSE) is a widely used metric to measure the average squared difference between the values (or predictions) and the actual values in a dataset (Nong et al., 2023). It is a measure of the average squared deviation or error between the predicted or estimated values and the true values. MSE is particularly useful in regression problems, to estimate a continuous numerical value. The formula for calculating MSE is in the Eq. (2):

$$MSE = (1/n)\sum (yi - \widehat{y}i)^2 \tag{2}$$

Where: n is the total number of data points. yi represents the actual or observed value for the i-th data point, ŷi represents the predicted or estimated value for the i-th data point and the Σ symbol denotes the sum over all data points.

The MSE provides a measure of how well a predictive model is performing. Lower MSE values indicate that the predictions are closer to the actual values, implying better model accuracy. Conversely, higher MSE values suggest that the model's predictions are farther away from the actual values, indicating poorer model performance (Nong et al., 2023).

3.3. Water Quality Index (WQI) prediction model

For WQI prediction, there was one output variable involved in the modelling procedure and six input variables (Table 5). As for this study, FANN full is applied where BOD and COD are part of the input that has been excluded from the initial inputs.

Initially, all six water quality parameters were used as input variables in the modelling to evaluate model performance. However, to achieve higher effectiveness and accuracy in predicting the WQI based on a smaller number of parameters, the focus was shifted to reducing the number of input parameters. This was done by developing scenarios with only five or four input parameters, instead of the original six stipulated in the Department of Irrigation and Drainage (DID) manual. By using less input variables in this feedforward neural network model, forecasting the river water quality's water quality would be cheaper and less time consuming. It would also result in a smaller amount of time consuming laboratory testing. It is also anticipated that these scenario analysis analyses will help to identify the relationship between water quality parameters and WQI classes in the future (Ho et al., 2019).

There have few series of steps taken to reduce the input variables in the calculation and prediction of WQI (Table 6). The data are firstly being analysed to get the correlation matrix to identify parameters with the highest and lowest correlation with WQI. As for the next step, based on the obtained result from correlation coefficient analysis, if their contribution to WQI is low, they can be removed from the calculation. The impact of removing selected parameters must be evaluated before deciding whether to exclude them. If their removal does not significantly impact the accuracy of WQI, they can be excluded. As for last step, the accuracy of the new WQI calculation have to be validated by comparing it to the original equation and assessing its accuracy in predicting water quality.

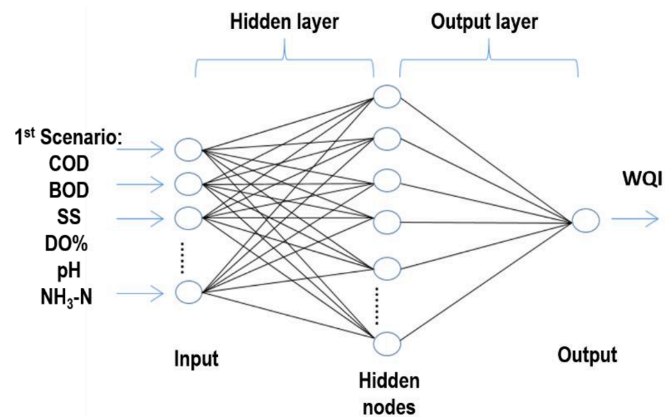
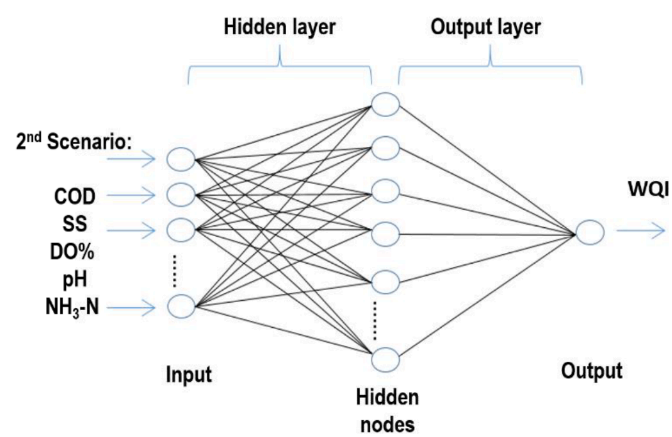
3.4. First modelling scenario (all input included)

The model is set up with all six parameters (BOD, COD, DO %, NH_3-N, SS and pH) as input variables in order to determine the model performance. In the first scenario, six parameters retained in the calculation and used as input (Table6A). The second scenario has 5 inputs by excluding the BOD (Table 6B), and in the third scenario, the number of neurons was decreased to four parameters by excluding both BOD and COD (Table 6C).

Table 5
Input and output dataset for WQI Prediction.

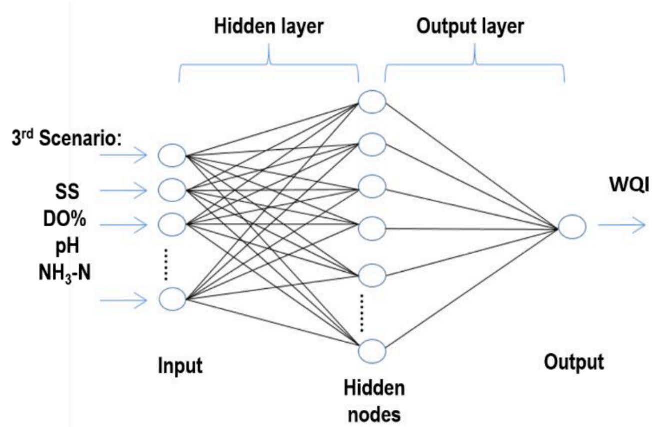
	FANN_full
Input	Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Dissolved Oxygen (DO), ammoniacal nitrogen (NH_3-N), suspended solid (SS) and potential for Hydrogen (pH)
Output	WQI

Table 6
FANN structure for (a) all inputs, (b) excluding BOD and (c) excluding BOD and COD and WQI as output.

Modelling Scenario	FANN Structure
A **1st Scenario (all inputs)** FANN_full, [(WQI) ̂(t)] = Fn [BOD(t), COD(t), DO % Sat(t), SS(t), pH(t), NH_3-N(t)]	
B 2^{nd} scenario (excluding BOD as inputs) FANN_full, [(WQI) ̂(t)] = Fn [COD(t), DO % Sat(t), SS(t), pH(t), NH_3-N(t)]	

(continued on next page)

Table 6 (*continued*)

Modelling Scenario	FANN Structure
C 3^{rd} Scenario (excluding BOD and COD as inputs) FANN_full, [(WQI)`(t)] = Fn [DO % Sat(t), SS(t), pH(t), NH_3-N(t)]	

4. Result and discussion

The Department of Environment (DOE) in Malaysia has established the Water Quality Index (WQI) Classification to assess the water quality of rivers, which is based on several parameters such as BOD, DO, pH, and TSS. In a 10-year study of Batu Pahat River, the WQI was found to vary significantly, with values ranging from Class I to Class V, depending on the sampling locations and time periods. This suggests that the river is facing various environmental pressures, including point and non-point source pollution from agricultural runoff, industrial discharges, and domestic sewage. The WQI classification could provide valuable insights into the water quality status and potential risks to human health and the environment. However, continuous monitoring and assessments are crucial to identifying emerging water quality issues and implementing effective management strategies to ensure the sustainable use of water resources.

4.1. The WQI classification and environmental analysis

4.1.1. Dissolved oxygen (DO)

The collected data for DO parameter throughout 10 years of study at Batu Pahat River at all sub-basin was tabulated in Fig. 4. The DOE Water Quality Index Classification categorizes the quality of water based on the amount of DO present in it, with Class I being the highest quality and Class V being the lowest quality, based on a range of DO levels in milligrams per liter (mg/L). Class I has DO levels greater than 7 mg/L, Class II has levels between 5 and 7 mg/L, Class III range around 3 and 5 mg/L, Class IV has levels between 1 and 3 mg/L, and Class V has levels below 1 mg/L (Table 2). Referring to the graph, the DO values were scattered at all level of water Quality Index Classification ranging from Class 1 to Class V, where the most were tabulated under Class I, II and III and the least at Class V. This can be concluded that the Batu Pahat river had undergo various condition which depend on several factors that can affect the level of dissolved oxygen in a river. One of the major factor is due to the pollution. Pollutants like sewage, agricultural runoff, and industrial discharge can introduce organic matter into the water, leading

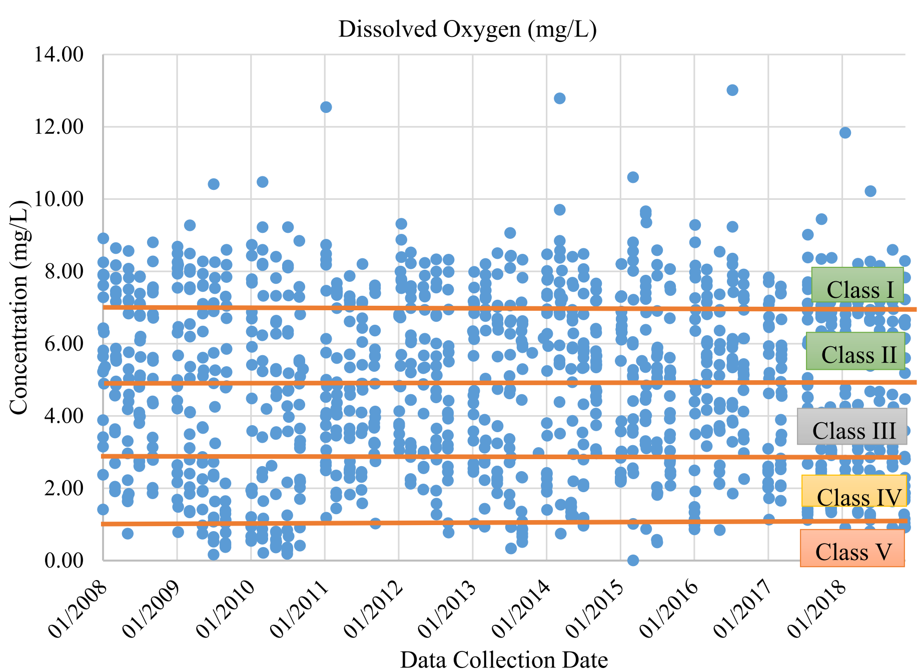


Fig 4. The DOE WQI classification for DO concentration in 10 years study of Batu Pahat river.

to increased bacterial activity and subsequent oxygen depletion (Juahir et al., 2011). Scientifically, aquatic plants and algae produce oxygen during photosynthesis, while bacteria and other organisms consume oxygen during respiration. If there is excessive plant or algae growth, or if there is a large population of oxygen-consuming organisms, or if there is excessive nutrient such as nitrogen and phosphorus, it ultimately depleted the dissolved oxygen levels as these organisms die and decompose (Nong et al., 2023). Furthermore, rainfall also could be one of the major reason that effect the concentration and distribution of dissolved oxygen in river water, where indirectly turn impact a water quality (Jia et al., 2021). When there is heavy rainfall, this has led to the higher velocity and turbulence by altering the flow rate and volume of water in the river (Li et al., 2015). Therefore, water movement and turbulence increases the amount of oxygen that can dissolve in water, so areas with greater water movement typically have higher DO levels. The changes in water temperature due to rainfall might also contribute to the fluctuation of dissolve oxygen values where colder water holds more oxygen, resulting in decreasing of water temperature (Jia et al., 2021; Verma and Singh, 2013). Based on the historical data, there have the catastrophic events of flooding in the Batu Pahat River basin, where the most severe in December 2006 and January 2007. Therefore, the unremitting development, industrialization and climate change alongside the sub-basin of Batu Pahat river may contributed to the uncertainty of DO concentration towards various classification.

4.1.2. Biological Oxygen Demand (BOD)

The Water Quality Index Classification categorizes the quality of water based on one of the parameter which is BOD that indicate the amount of organic matter present in milligrams per liter of water (mg/L). Referring to Table 2, Class I being the highest quality and Class V being the lowest quality. Class I has BOD levels less than 1 mg/L, Class II has levels between 1 and 3 mg/L, Class III ranges from 3 and 6 mg/L, Class IV has levels between 6 and 12 mg/L, and Class V has levels greater than 12 mg/L. According to the Fig. 5, the Batu Pahat river can be categorized under Class IV since the highest value scattered at around 6 to 12 mg/L. Higher BOD levels indicate greater levels of organic matter present in the water, which can lead to lower dissolved oxygen levels and potentially harm aquatic life. In contrast, higher level of DO can help to promote the breakdown of organic matter by aerobic bacteria, which lead to the higher BOD levels. Organic matter can originate from different sources, including sewage (failing septic systems), agricultural runoff that may included dead plants and animals, and industrial discharge. Agricultural activities, which are the major land use activity in all sub-basins of Batu Pahat river, contribute to an increase in nutrient levels, sediment runoff, water use, pesticide, and fertilizer use, all of which can impact the BOD level of nearby water bodies (Bhateria and Jain, 2016; Juahir et al., 2011).

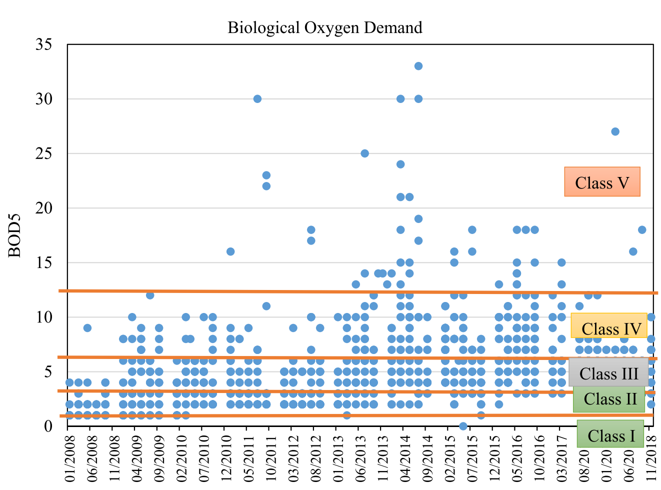


Fig. 5. The DOE WQI classification for BOD concentration in 10 years study of Batu Pahat river.

4.1.3. Chemical Oxygen Demand (COD)

Over the course of 10 years, the water quality index for Batu Pahat River in Malaysia consistently falls under Class II in the COD Water Quality Index Classification, with COD levels ranging between 10 and 25 mg/L. This indicates that the water quality is fair, with a moderate level of organic matter present. However, even though the COD levels are within an acceptable range, it is important to note that any increase in organic matter can lead to a decrease in dissolved oxygen levels and create conditions that are harmful to aquatic life. Referring to the graph in Fig. 6, there have been at certain condition that the COD reading at Batu Pahat river experiencing Class V during year 2021 and few recorded data were classified under Class IV. The high levels of organic matter can contribute to algae growth, unpleasant odors and tastes, and the buildup of toxic substances in the water. Therefore, it is important to continue monitoring COD levels in Batu Pahat River to ensure that the organic matter levels remain within a safe range and to identify any potential sources of pollution that may lead to an increase in COD levels. This can help to maintain the health of the river ecosystem and ensure that it remains a valuable resource for both humans and aquatic organisms, kill the flora and fauna and collapse the ecosystem.

4.1.4. Total suspended solid (TSS)

Total suspended solid is one of a water quality index that measures the amount of solid particles, such as silt and organic matter, that are suspended in water column. The TSS level is measured in miligrams per liter (mg/L) and it can provide an indication of the water quality and the level of contamination in a river. The Water Quality Index Classification for TSS is divided into five classes, with Class I being the highest quality and Class V being the lowest quality. Class I has TSS levels less than 25 mg/L, Class II has levels between 25 and 50 mg/L, Class III has levels between 50 and 150 mg/L, Class IV has levels between 150 and 300 mg/L, and Class V has levels greater than 300 mg/L (Table 2). Over a period of 10 years from DOE data collection at Batu Pahat river, the classification index consistently falls between Class I and Class II, with TSS levels ranging below 50 mg/L, which indicates the river water quality is in a fair to good with moderate level of suspended solid (Fig. 7). The level of TSS in a river can be influenced by various factors, such as natural erosion, urban and agricultural runoff, and industrial discharges. As in Batu Pahat river, agricultural activities are also one of the major land use activities in all sub-basins. Scientifically, higher TSS levels indicate a higher level of pollution and directly contribute to the negative impacts on aquatic life and river ecosystems. The TSS particles able to block sunlight, reducing the amount of light available for photosynthesis and leading to a decline in the growth of aquatic plants. Additionally, TSS can cause sedimentation, smothering the benthic organisms and habitats, and reducing biodiversity. Besides, TSS can also transport pollutants and nutrients such as heavy metals and phosphorus, which can accumulate in the sediment and affect aquatic organisms and public health. Excessive nutrients from TSS can lead to eutrophication, promoting the growth of algae and other aquatic plants, which can cause algal blooms and reduce oxygen levels in the water, leading to the death of aquatic organisms. Therefore, it is essential to monitor TSS levels in rivers and implement measures to reduce the sources of TSS pollution to maintain healthy river ecosystems and protect public health.

4.1.5. pH

The pH level of a river indicates the acidity or basicity of the water, which can have a significant impact on aquatic life and the ecosystem. In Batu Pahat river, the WQI for pH has consistently remained under Class II for the past ten years (based on majority recorded collected data), indicating a slightly acidic to neutral water condition (Fig. 8). However, the fluctuation in pH reading seldomly occurred since the recorded data

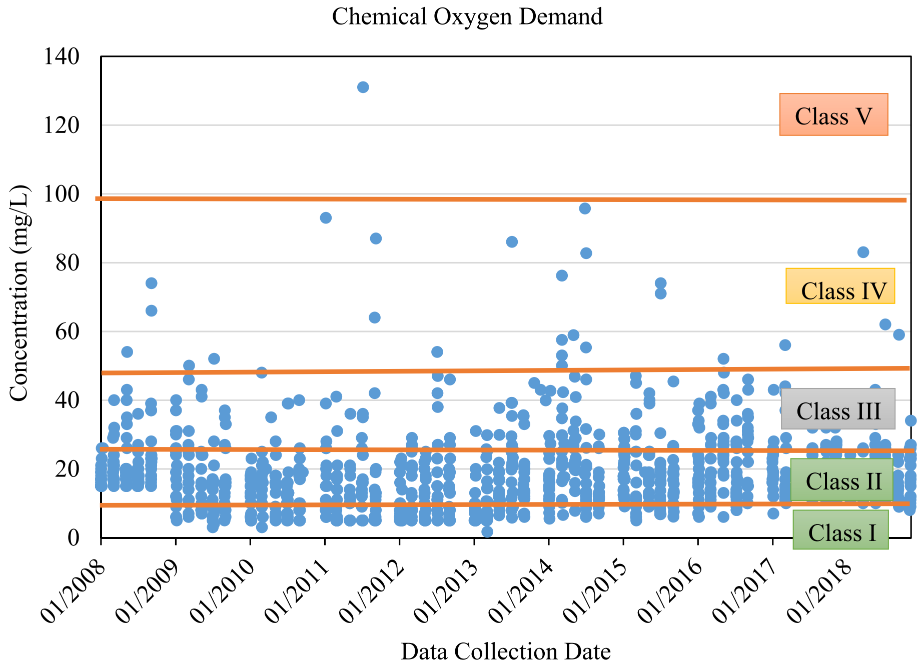


Fig. 6. The DOE WQI classification for COD concentration in 10 years study of Batu Pahat river.

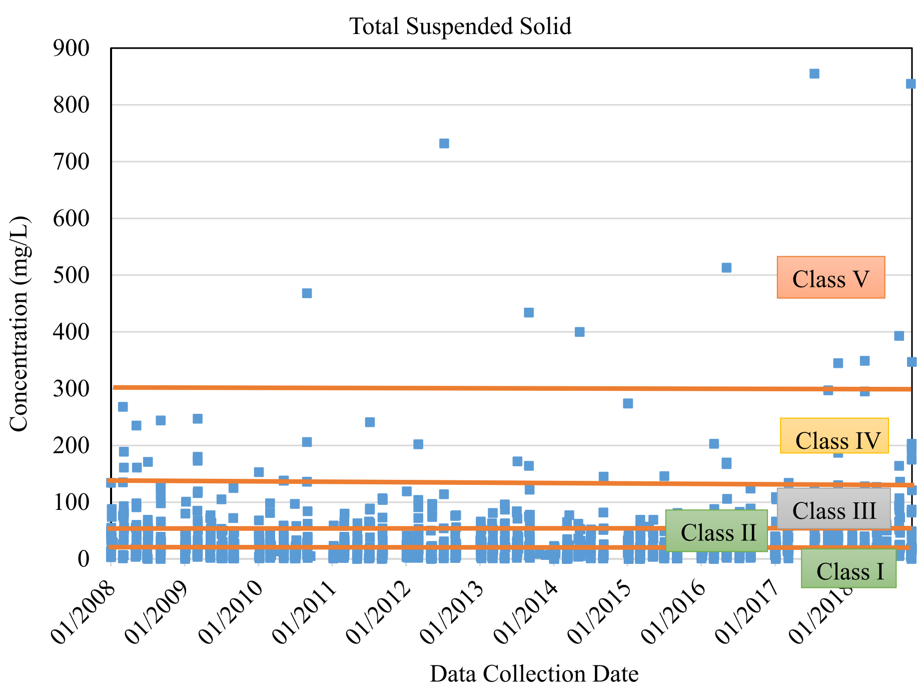


Fig. 7. The DOE WQI classification for TSS concentration in 10 years study of Batu Pahat river.

also varies at pH below 5 and lies under Class V (Fig. 12). Changes in pH can have various effects on river water quality. Biomicrobial productivity is optimal between pH 7 and 8.5, while pH 4 is detrimental to aquatic life (Yona et al., 2023). As pH changes with temperature, dissolved oxygen levels in water are affected, which can lead to biochemical reactions such as photosynthesis being impaired or killed (Yona et al., 2023). Some fish species are sensitive to even minor pH changes, so acidic water can damage their gills and scales, causing them to perish. Changes in pH can also affect the solubility and toxicity of certain chemicals in the water, making them more or less harmful to aquatic organisms (Verma and Singh, 2013). Additionally, pH can affect the biological processes that occur in rivers, such as photosynthesis, respiration, and nutrient uptake. Certain microorganisms that play an essential role in these processes are also sensitive to changes in pH levels, which can lead to imbalances in the ecosystem. Therefore, maintaining a balanced pH level is crucial for protecting the health of river ecosystems and ensuring the sustainability of aquatic life and protect public health. pH is a crucial water quality parameter that is strictly monitored in Malaysia. It is the most accessible parameter to measure as it can be obtained on-site without requiring extensive laboratory analysis (Ho et al., 2019). Additionally, Malaysia's tropical climate with high levels of rainfall throughout the year contributes to

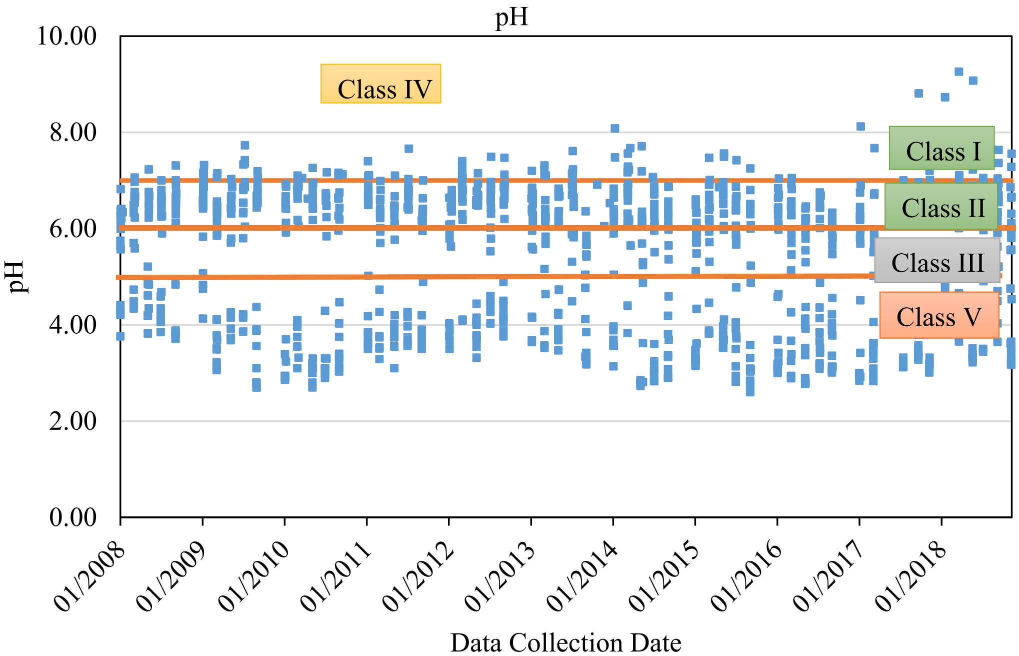


Fig. 8. The DOE WQI classification for pH concentration in 10 years' study of Batu Pahat river.

the rapid dilution and neutralization of pH values in rivers. There have been studies demonstrating some degree of deterioration of water quality in some natural rural lakes due to rainfall (Brias et al., 2018). According to Ahmed et al. (2018), the pH usually drops by 5 % after heavy rains of over 25 mm in one day (Ahmed et al., 2018).

4.1.6. Ammoniacal nitrogen

The WQI for ammoniacal nitrogen in Batu Pahat river has been classified as Class V for over a period of 10-years from the DOE recorded data, with levels exceeding 2.7 mg/L (Fig. 9). It is not a good sign where it can cause severe damage to aquatic life and river ecosystem. However, there still have numbers of recorded data that scattered around Class I, II and III which indicate the fluctuation of river condition, due to weather, urbanization, and surrounding activities. High levels of ammoniacal nitrogen can lead to eutrophication, which promotes the growth of algae and other aquatic plants, reducing oxygen levels in the water and leading to the death of aquatic organisms (Back et al., 2023). It has been reported that agricultural and livestock wastewater, landfill leachate, and municipal and industrial wastewater contain high levels of ammoniacal nitrogen (Tonetti et al., 2016). Moreover, ammoniacal nitrogen can cause toxic effects on aquatic life, affecting their growth, development, and reproduction. In human and animal waste, ammoniacal nitrogen is formed from the breakdown of urea and other nitrogen-containing compounds in urine and feces. In agriculture, ammoniacal nitrogen can come from the application of nitrogen-rich fertilizers and manure to fields, which can leach into nearby waterways. In industries

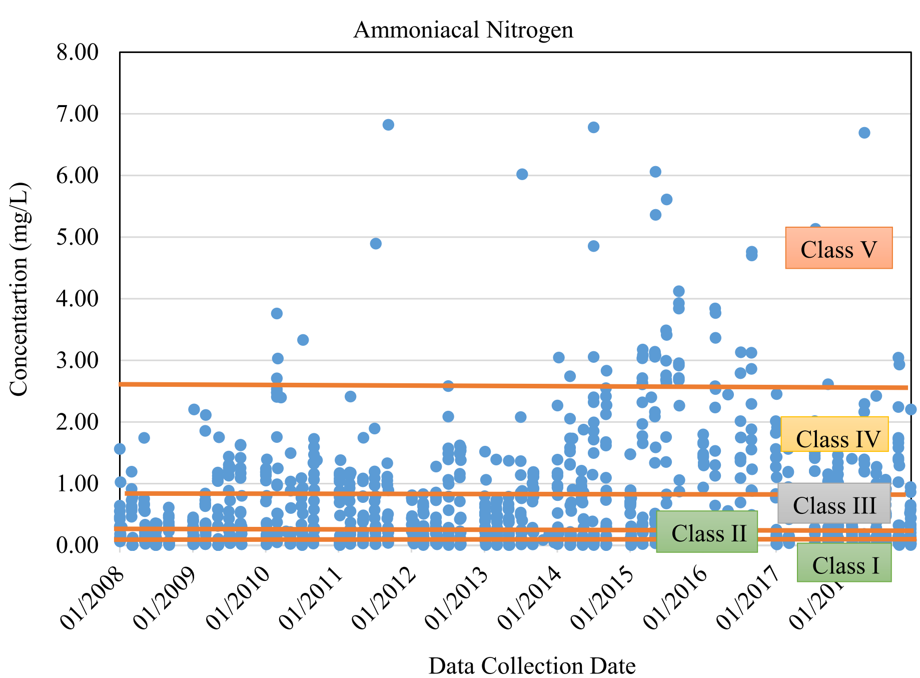


Fig. 9. 10 years' distribution of Ammoniacal Nitrogen Batu Pahat River Basin.

such as food processing and chemical manufacturing, ammoniacal nitrogen can be a byproduct of production processes and wastewater discharges (Adam et al., 2019). Irrespective of where it dwells in a water stream, ammoniacal nitrogen must be removed from it before it can be discharged to the environment.

4.1.7. Water Quality Index (WQI)

Fig. 10 presents the WQI of Batu Pahat River during period of 2008 and 2018. The index was separated based on individual sub-basin in order to identify in details the WQI value at each of the studied location. Referring to the graph, throughout the years, the trend of WQI were scattered at almost similar values of WQI. For example, at SBP01 station, the WQI value were tabulated at approximately 55 to 70 and categorized under Class III. Similar observation can be seen at almost all stations especially at SBP 10, SBP 17, SBP 23, SBP 25, and SBP 26, where it tabulated at almost similar number. As overall, all sub-basins were categorized at Class II and Class III, which indicated that Batu Pahat river was slightly clean and slightly polluted. And most importantly that the river is excluded from Class V throughout the 10 years of sampling. It was a good sign for the whole ecosystem that the river can be managed and been take care of for at least to improve the WQI classification to be maintained at Class II or Class I in the near future.

The WQI is an important tool for assessing water quality and is often used to classify water bodies into different categories based on their suitability for different uses. In this study, the WQI values of Batu Pahat river suggest that the river is currently suitable for recreational activities, such as fishing and swimming, but may require management interventions to improve its water quality for other uses, such as drinking water or irrigation. Furthermore, there parameters are important indicators of water quality and can help identify potential sources of pollution. The slightly polluted classification observed in this study may be attributed to human activities such as agriculture, industrial activities and urbanization, which can be contribute to the degradation of water quality.

4.2. Model correlation and calibration

An output WQI is calculated using the correlation coefficients of the input parameters as shown in Fig. 11 and Table 7, which indicates the degree of linear relationships between two variables. This analysis provides crucial insights into the strength and direction of the relationship between each input parameter and the WQI. The presentation is clear and concise, with each input parameter, including DO (mg/L), BOD (mg/L), COD (mg/L), SS (mg/L), pH, NH_3-N, and their corresponding correlation coefficient values, listed in Table 3. The precision of the correlation coefficient values to two decimal places makes it easy to compare the strength of the relationships between different input parameters and the WQI. The results indicate that some input parameters exhibit a stronger correlation with the WQI than others, with DO (% sat) demonstrating a strong positive correlation with a correlation coefficient of 0.85, while pH exhibits the second-highest correlation with a value of 0.69. Conversely, the negative correlation coefficients for BOD (-0.4), COD (-0.34), TSS (-0.05) suggested that these parameters have a negative impact on the WQI, with NH_3-N showing a moderate negative correlation of −0.59. The remaining input parameters exhibit correlation coefficients ranging from −0.05 to −0.4. Overall, this analysis provides valuable information on the relationship between the input parameters and the WQI, highlighting the significance of considering multiple input parameters to assess water quality and to develop strategies to improve it based on these findings.

4.3. First modelling scenario (all input included)

Fig. 12 depicts the actual (solid line) and predicted (dashed line) water quality index (WQI) values on the training and testing datasets, while Fig. 13 shows the actual and predicted WQI values, as well as the model residuals, on the unseen validation dataset. The figures illustrate the strong predictive performance of the feedforward artificial neural network (FANN) model, as the predicted values closely match the actual values of WQI. Table 8 summarizes the performance of the model across all stages (training, validation, and testing), revealing excellent predictive accuracy when all input variables are included. The mean square error (MSE) values obtained during the training, testing, and validation stages were 0.0009, 0.0024, and 0.0045, respectively, indicating that the model has effectively learned to predict the WQI. Furthermore, the coefficient of determination (R^2) achieved during all stages (0.9991, 0.9976, and 0.9953 for training, testing, and validation, respectively) support the model's ability to capture the complex relationships between input variables and WQI. By including all input variables, the FANN model identifies the most significant predictors of water quality, providing valuable insight for water resource management and policy making. The low MSE values obtained during the training and testing stages demonstrate that the model generalizes well to new, unseen data. However, a higher MSE value during the validation stage would indicate

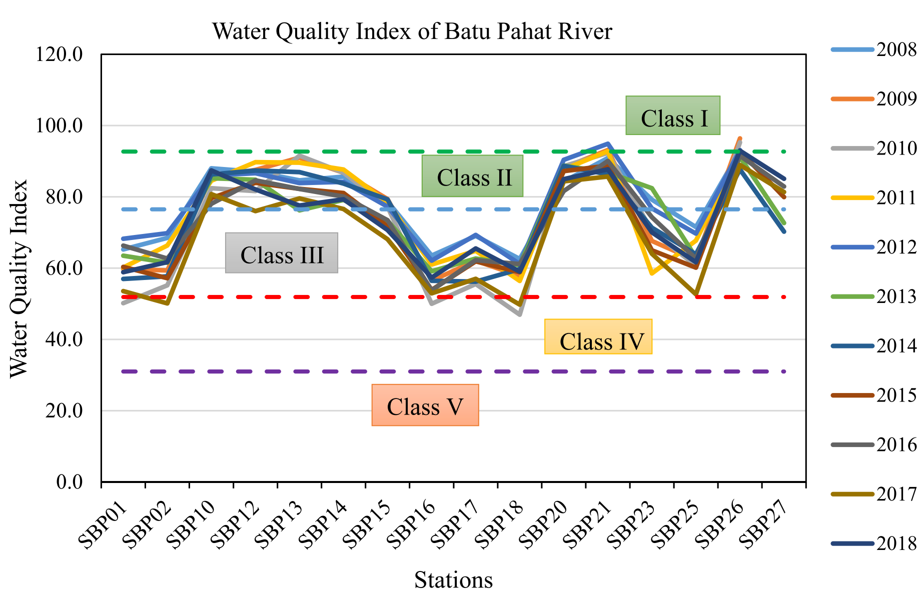


Fig. 10. Water Quality Index for Batu Pahat River since years 2008 and 2018. Frequency of class classification for Class I, II, III, IV and V.

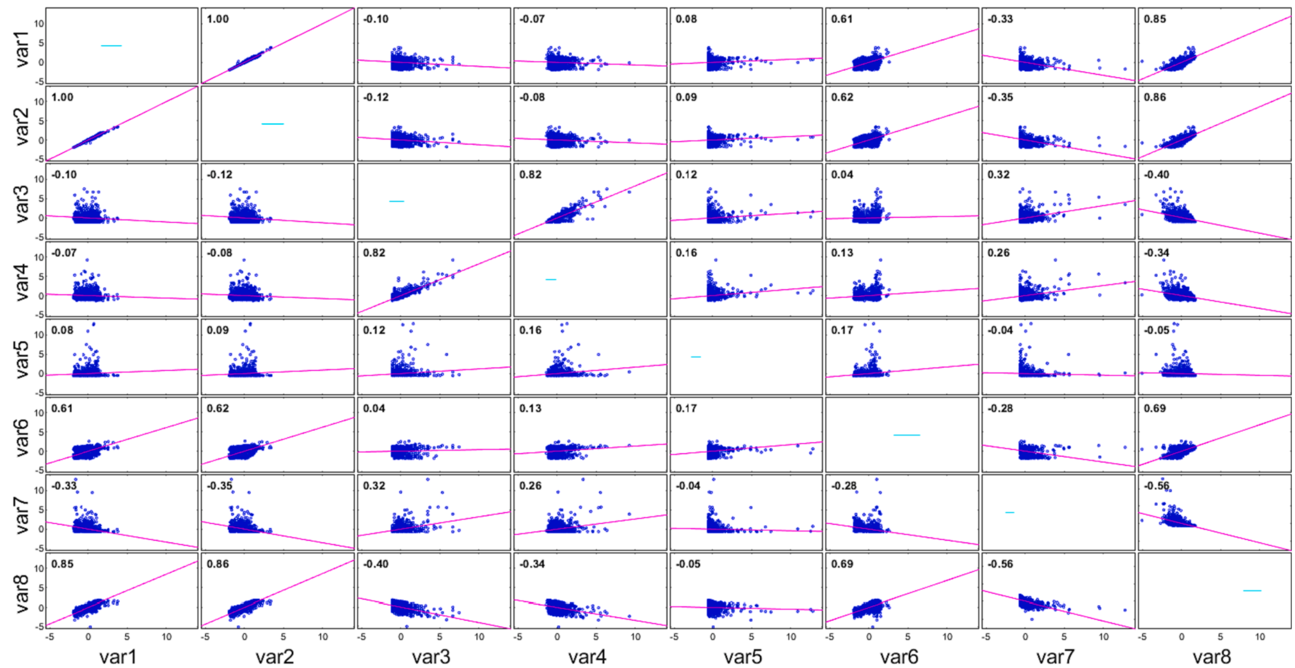


Fig. 11. Correlation coefficient matrix for the inputs and output.

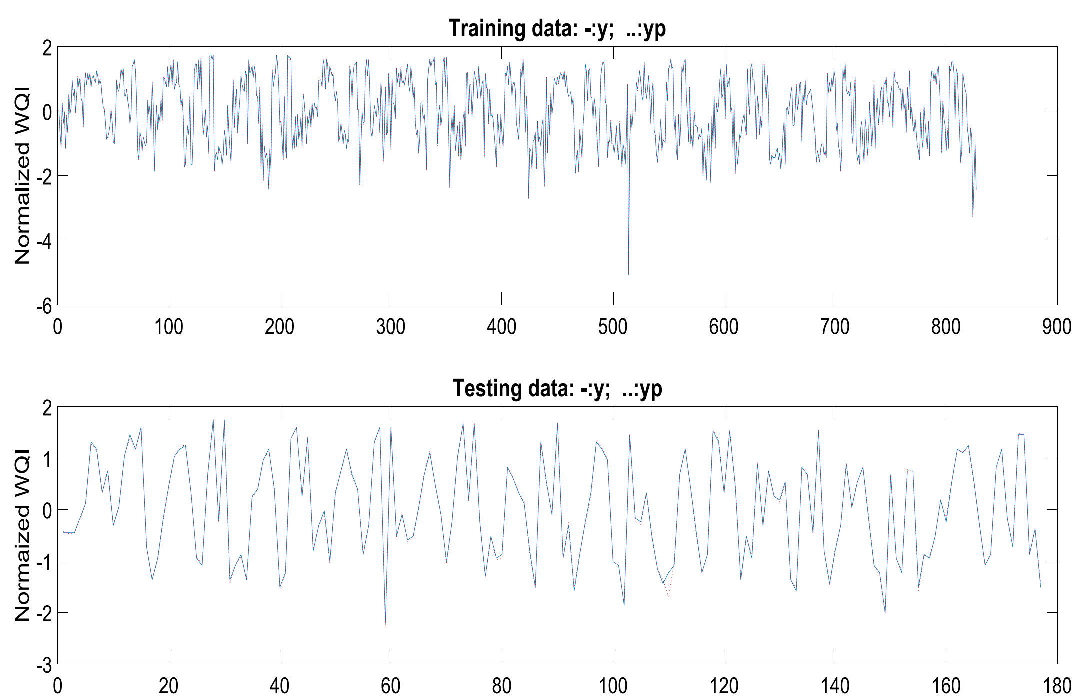


Fig. 12. Actual and predicted values for training and testing data for all inputs.

overfitting to the training data and suggest a need for further model refinement, such as incorporating additional data or tuning model parameters.

4.4. Second modelling scenario (Excluding BOD)

The primary objective of conducting scenario II was to enhance the model performance by minimizing the number of input variables. This was accomplished by excluding the BOD parameter and testing different input combinations using only five water quality parameters. Fig. 14 displays the actual and predicted values for water quality prediction based on the training and testing data for all inputs. The X-axis represents the input data, while the Y-axis represents the actual and predicted values. The graph illustrates that the predicted values for the water quality are relatively close to the actual values for both the training and testing data, which suggests that the model used for water quality prediction is effective in accurately predicting the water quality based on the input data.

In the training data, the predicted values conform to the trend of the actual values, with some variations in between. In the testing data, there is little to no deviation between the actual and predicted values, indicating the model's effectiveness. It is important to note that the accuracy

Table 7
Correlation coefficient for each input to output WQI.

Variables	Attributes		Correlation toward WQI
1	DO (% Sat)	Input	0.85
2	BOD (mg/l)		−0.4
3	COD (mg/l)		−0.34
4	TSS (mg/l)		−0.05
5	pH		0.69
6	NH_3-N (mg/l)		−0.56
7	WQI	Output	1

of the model may depend on the quality and quantity of the input data. Overall, the graph provides valuable insights into the accuracy of the water quality prediction model, indicating that it can effectively predict water quality based on the input data (Fig. 15). The results of the analysis in the second scenario further confirmed that BOD is the least effective parameter in predicting WQI. Furthermore, BOD has been shown to have low correlation to WQI prediction. Moreover, BOD is a challenging parameter to monitor, as its measurement can be obtained through tedious laboratory analysis, which may take up to 6–7 days. Hence, the exclusion of BOD as input data for predictive modelling has little to no effect on the prediction of WQI class.

Table 9 provides an account of the outcomes from training, testing and validation of the WQI using FANN methodology, with the elimination of BOD data as an input. The training, testing and validation mean squared error (MSE) values, being 0.0170, 0.0141, and 0.0193, respectively, are relatively low, demonstrating that the model's predictions are reasonably precise. The R^2, 0.9832, 0.9853, and 0.9799, respectively, indicate that the model's predictions are strongly linked with the actual WQI values. Notably, the exclusion of BOD data as an input factor may suggest that other input parameters used in the model could be adequately predictive of WQI independently. This implication may be valuable in water quality management as it offers an opportunity to develop more cost-effective monitoring programs focusing on a subset of input parameters instead of measuring all parameters. Collectively, the findings signify that the FANN approach is useful in predicting WQI values utilizing a subset of input parameters, and that the omission of BOD data does not appear to significantly impact the model's accuracy.

Table 8
Training, testing performance of WQI prediction.

Performance	*FANN-all*	
	MSE	R^2
Training	0.0009	0.9991
Testing	0.0024	0.9976
Validation	0.0045	0.9953

4.5. Third modelling scenario (Excluding BOD and COD as inputs)

In scenario III, the input parameter was further reduced to four parameters. The performance of FANN with 15 hidden neurons is presented in Figs. 16 and 17. According to Fig. 16 across the training and testing data, the actual values are displayed as solid lines while the predictions are shown as dashed lines, whereas Fig. 17 displays the actual values, as well as model residues, on the unseen validation data as solid lines and dashed lines respectively. The analysis of the data reveals that the performance of the FANN model is excellent as the model predictions are closely aligned with the actual values of QWI. The network exhibits the ability to generalize and adapt to new input data, as depicted in the figures. Also, the results of the statistical analysis presented in Table 10 demonstrate the relative effectiveness of the FANN model in predicting WQI based on the unseen validation data. The results illustrate how well the models perform at each stage of the analysis (training, testing, and validation), as indicated by the statistical analysis results. As illustrated by the provided data, the FANN model performed well in predicting the WQI values, with the testing phase having the lowest MSE (0.0663) and the R^2 (0.9311) when compared to the training phase. This finding indicates that the FANN model has effectively learned the relationship between the input parameters and the WQI values, allowing it to make accurate predictions.

The results obtained in scenario III provide insights into the effectiveness of the FANN model in predicting WQI values using a reduced number of input parameters. This finding could have important implications for water quality management, as it suggests that a more cost-effective monitoring program could be developed that focuses on a subset of input parameters rather than measuring all parameters. It is important to note that the accuracy of the model is dependent on the quality and quantity of the input data. Thus, further research is needed to validate the effectiveness of the FANN model in predicting WQI values under different scenarios.

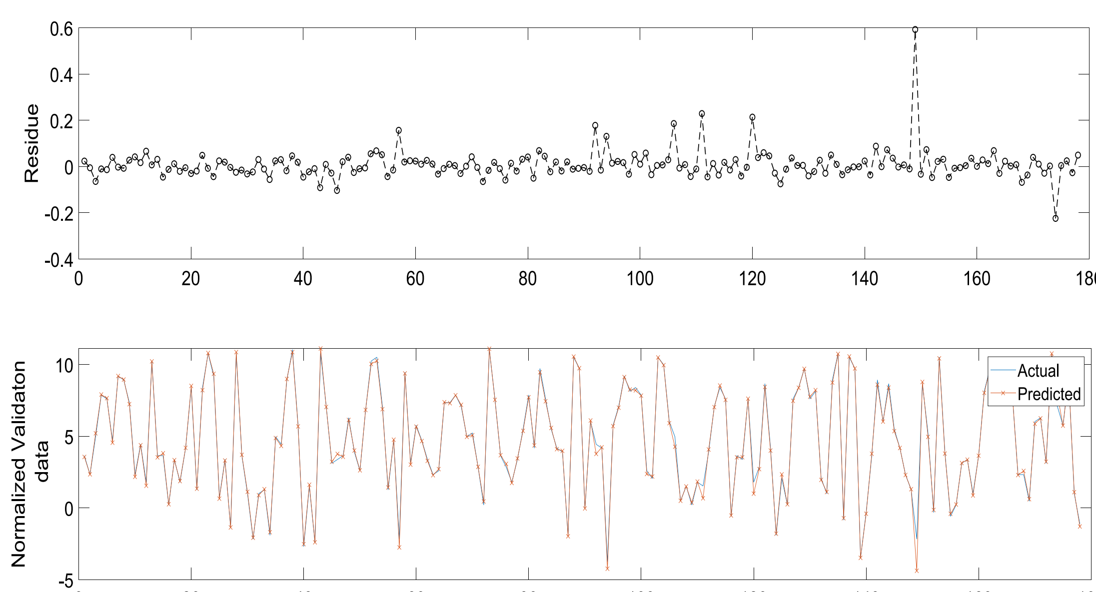


Fig. 13. Actual and predicted values and residues for validation data for all inputs.

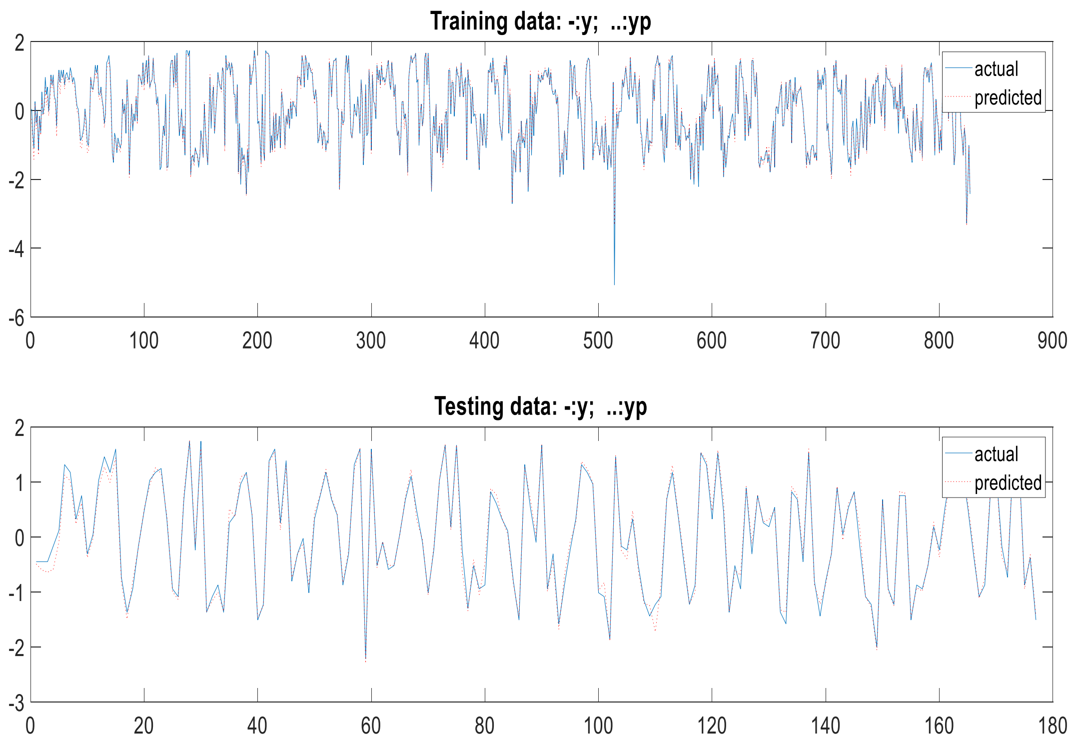


Fig. 14. Scaled actual and predicted values for training and testing data without BOD in the inputs.

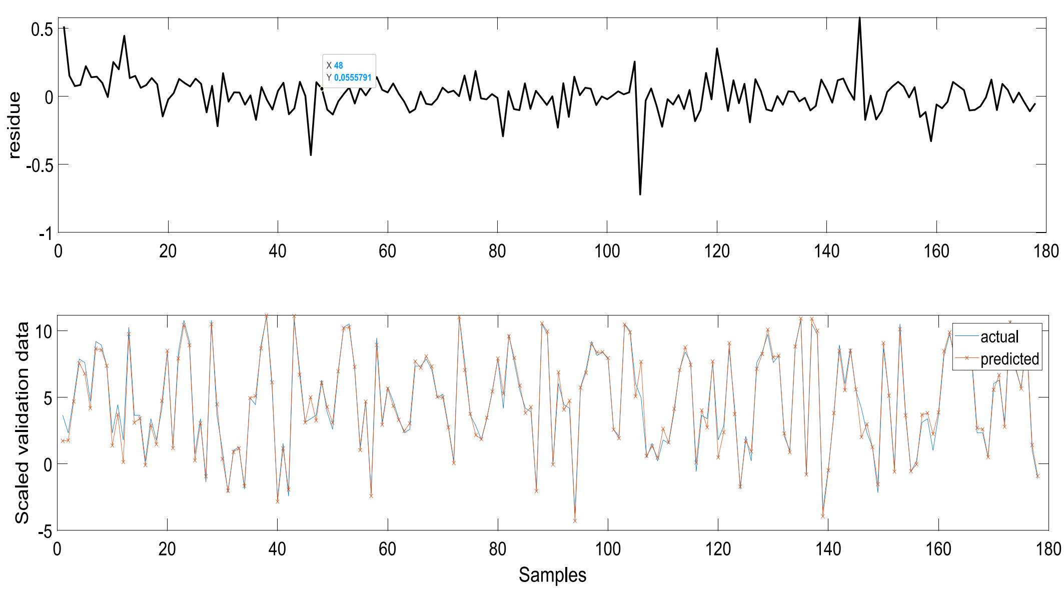


Fig. 15. Actual and predicted values and residues for validation data without BOD in the inputs.

Table 9
Training, testing and validation performance of WQI prediction.

Performance	*FANN- exclude BOD*	
	MSE	R^2
Training	0.0170	0.9832
Testing	0.0141	0.9853
Validation	0.0193	0.9799

As a conclusion, the result given has appeared that removing one parameter from the input variables for WQI calculation only results in a small decrease in accuracy, with a reduction of R^2 from 0.9953 to 0.9799 which is reciprocate to 1.55 % (Table 11). On the other hand, removing two parameters resulted in a more significant decrease in accuracy, with a R^2 of 0.9103 (8.4 %). Removing one parameter has only a small impact on the accuracy of WQI calculation, removing 2 parameters slightly reduce the accuracy. And it is therefore important to carefully evaluate the contributions of each parameter to the overall calculation and to only remove parameters that have a low impact on the accuracy.

Besides, based on the findings of this study, in order to accurately represent the water quality index of a river, which serves as a crucial metric for assessing the river's environmental health, it is recommended to utilize all six parameters as input variables. Nevertheless, the proposed model can still be deemed reliable even in situations where one or

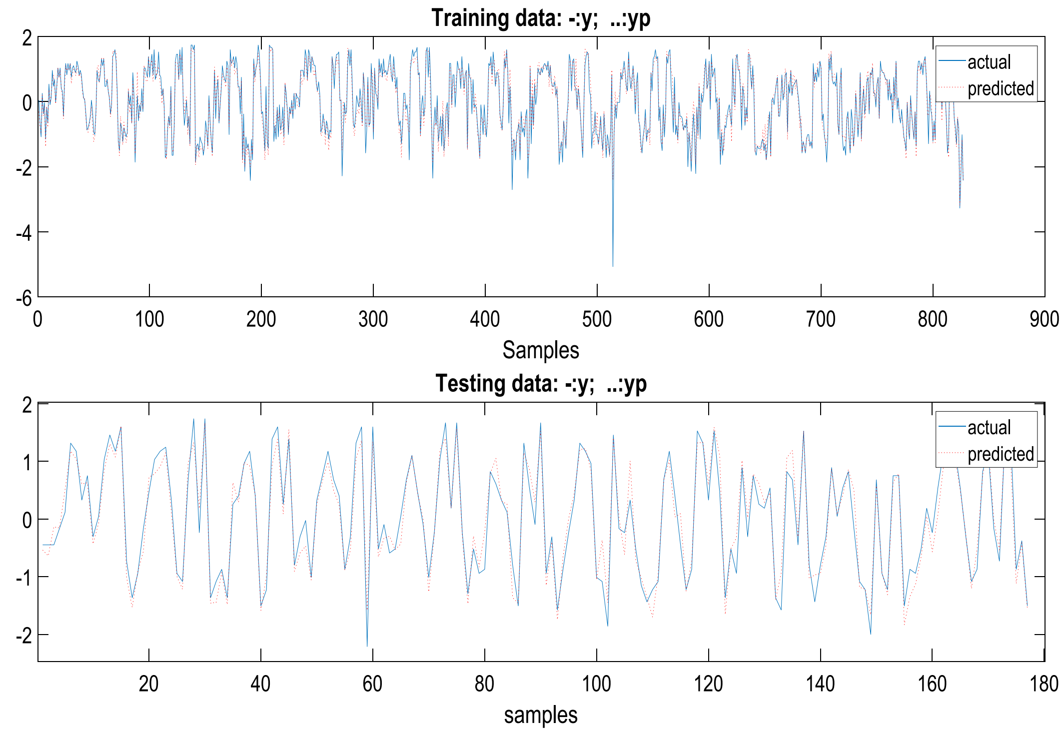


Fig. 16. Actual and predicted values for training and testing data without BOD and COD in the inputs.

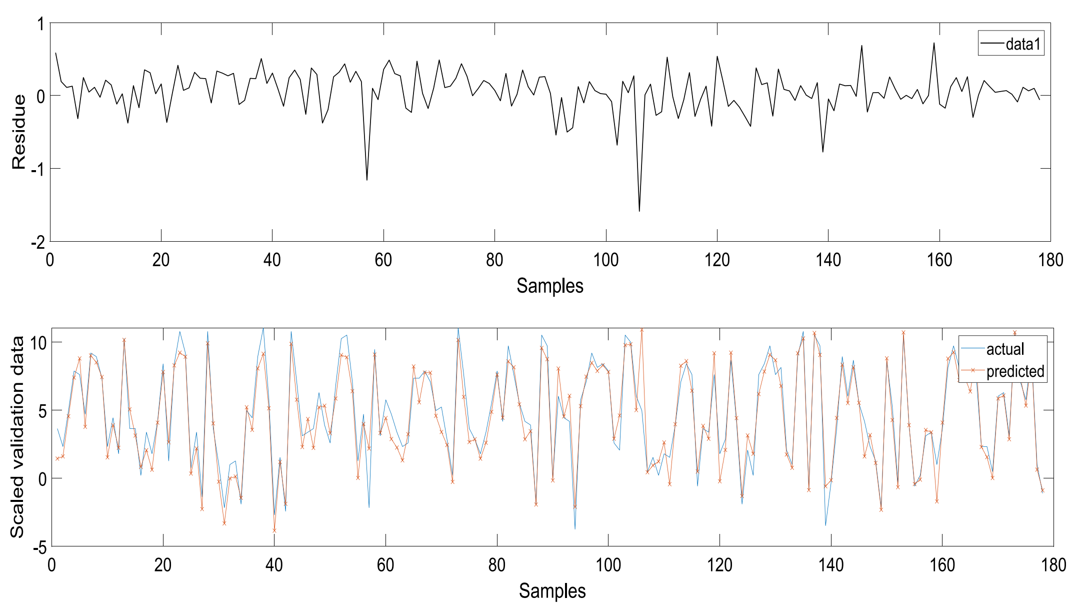


Fig. 17. Actual and predicted values and residues for validation data without BOD and COD in the inputs.

Table 10
Training, testing and validation performance of WQI prediction.

	FANN-(exclude BOD and COD)	
	MSE	R^2
Training	0.0744	0.9266
Testing	0.0633	0.9311
Validation	0.0860	0.9103

two parameters are unavailable. As part of this investigation to ascertain the robustness of the proposed model, two of the most significant parameters, namely COD and BOD, were deliberately omitted from the model development. It is noteworthy that even with the exclusion of these parameters, the developed model consistently exhibited an acceptable level of accuracy, with performance ranging between 91% and 98%. This suggests that the model retains its predictive capabilities and can be considered reliable when dealing with incomplete input data.

Table 11
Accuracy improvement of WQI prediction in removing water quality parameters.

Validation performance	R^2 (scenario 3)	R^2 (scenario 2)	R^2 (scenario 1)	Accuracy improvement by removing 1 parameter (%)	Accuracy improvement by removing 2 parameters (%)
	0.9103	0.9799	0.9953	1.55	8.46

5. Batu pahat river basin mitigation plan

Rivers play a vital role in the water cycle by collecting precipitation runoff from surrounding areas and moving it towards the oceans. When it rains, the water is absorbed by the soil, and some of it flows over the surface into streams and rivers, which eventually lead to the ocean. In certain cases, once the water reaches a river, it begins to flow downstream, and reaching a dam. In this area of studies, there have Bekok Dam and Sembrong Dam that located at SPB 02 and SBP 03, respectively. The main river for SBP 02 covers Bekok River and consists of several tributaries, e.g. Terusan River, Sedi River, Temehel River, Puroh River, Berlian River, etc. While at SBP 03, sub-basin covers Sembrong River and consists of several tributaries, e.g. Biuh River, Geriba River and Merpo River. Dams are built to control the flow of water and create a reservoir of water behind them. Water is stored in the reservoir until it is needed for human use. The water can be released from the reservoir through the dam and into a water treatment plant, where it is cleaned and treated to remove any contaminants. After treatment, the water is distributed through a network of pipes and pumps to homes and businesses for human use. Once the water has been used, it is returned to the environment through the sewer system, where it is eventually treated again before being released back into rivers or the ocean.

The amount and intensity of rainfall are critical factors in soil erosion, especially in areas with steep slopes or sparse vegetation. As they determine the amount of water and energy available to erode the soil, by also carrying pollutants such as pesticides, fertilizers, and sediment into water streams. As for Batu Pahat River Basin the activities that majorly impact the quality of the river are from oil palm plantations, sewage treatment plant (STP) effluents, commercial dischargers, industrial effluents and also from greywater usage in the basin. Oil palm plantations often rely on the use of chemical fertilizers and pesticides to maximize the crop yield. These chemical can leach into the soil and eventually into nearby rivers and waterways through surface runoff or groundwater. Besides, palm oil processing mills generate amounts of waste which contain high level of organic matters and nutrients. The effluent from STP, commercial and industrial discharges also contributed to river pollution in several ways since it contains high level of nutrients (nitrogen and phosphorus), pathogen (bacteria, viruses) as well as pharmaceuticals, personal care products and industrial chemicals. Fig. 18 describe the condition of water cycle in Batu Pahat River Basin.

A combination of strategies that focus on reducing pollution at the source, controlling runoff, intercepting and treating pollutants, preventing future pollution, and preventing erosion and sediment issues can help mitigate river pollution and protect water quality. These strategies can be tailored to specific river systems and local conditions to achieve the most effective results. Therefore, all activities that occurred in the basin will greatly impacts the water quality of river.

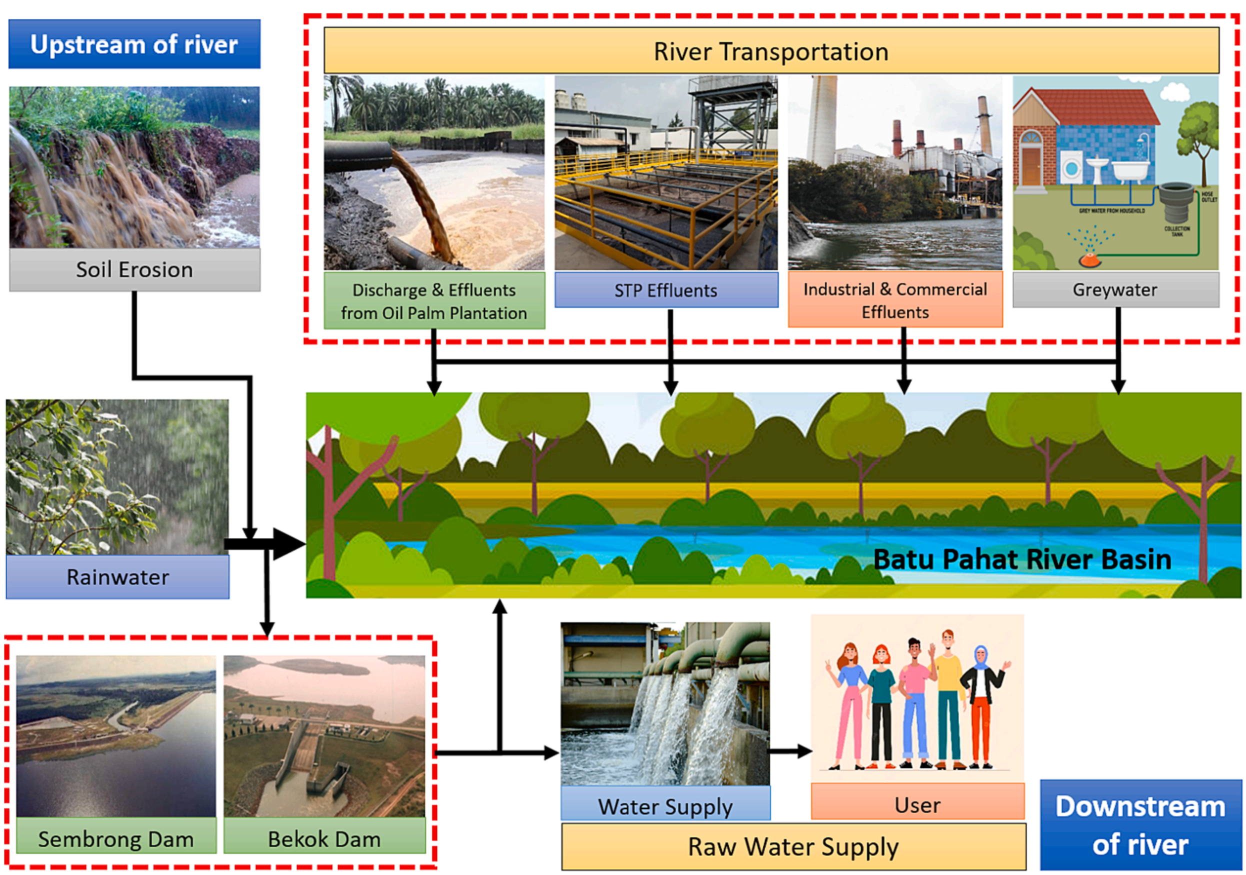


Fig. 18. Current Water Cycle in Batu Pahat River Basin.

5.1. Batu Pahat river basin mitigation plan for oil palm plantation area through good agricultural practices (GAPs)

Good Agricultural Practices (GAPs) are a set of guidelines and practices that aim to promote the sustainable and safe production of crops and livestock. The goal of GAPs is to reduce the risk of contamination and improve the overall quality of agricultural products while minimizing the impact on the environment and human health. The mitigation plan that focusing oil palm plantation area for Batu Pahat river may include the implementation of GAP during planning, development and planting of palm oil plantation as well as for biodiversity conservation, water and waste management around the plantation.

The authorities may consider implementing several mitigation plans, such as monitoring refueling and lubrication operations for pumping equipment to prevent watercourse pollution, and carefully monitoring the output from effluent ponds to prevent oxygen demand issues. It is also important to prevent run-off from entering watercourses when recycled, digested effluent is used. Furthermore, mulching application using empty fruit bunches is recommended as organic fertilizer to recycle organic matter, conserve moisture and prevent erosion. All organic materials such as empty fruit bunches (EFB), mill effluent and decanter solids as well as surplus shell and fibre is encouraged to be recycled to the field. Apart from that, the treated palm oil mill effluent (POME) and solid waste (biomass) can be apply for nutrient recycling and moisture conservation. Therefore, with these proposed mitigation plan, serious action has to be taken in order to minimize the risk and pollution towards river water.

5.2. Batu pahat river basin mitigation plan to reduce pollution load from major wet market effluent

The activities that take place at the market, such as cleaning, washing, and food preparation, generate a significant volume of wastewater on a daily basis. This wastewater is often disposed of improperly and can have detrimental effects on the environment. The pollution caused by this wastewater can take the form of unpleasant odours, contamination of water sources, and the spread of waterborne diseases. The main contributions of wastewater load from wet market are normally organic loads from food sources and slaughtering and minor detergent traces from washing. The absence of particular regulations to regulate the release of certain substances can worsen the problems associated with their discharge. According to the water sampling measurement, the wet market and food court in Pasar Besar Batu Pahat generate weighty impact to the environment especially through wastewater. There have previous studied from (Jais et al., 2020) for Public Market Parit Raja, Batu Pahat. The studied monitored the wastewater characteristics where BOD reading was measured at 89 $\pm$ 3.61 mg/L, COD has exceeding the DOE limit, and recorded at 456 $\pm$ 8.19 mg/L, TSS was 132.3 $\pm$ 1.7 mg/L, sulfate and total chlorine were 32.3 $\pm$ 0.78 mg/L and 32 $\pm$ 0.69 mg/L, respectively, pH at 6.0 $\pm$ 0.1, turbidity verified at 66.0 $\pm$ 8.9 mg/L, and oil and grease was 5.22 $\pm$ 0.07 mg/L. This can be concluded that wastewater from the wet markets was classified as class V and obviously not suitable for any purposes. To reduce pollution loading from wet markets, wastewater treatment plant (WWTP) is proposed to be constructed at major wet market in Pekan Yong Peng sub-basin. The treatment plant will reduce pollutants in the wastewater which is currently not being treated. Harmful pollutants and bacteria, such as BOD, ammoniacal nitrogen, oil, grease, suspended solids and E. Coli bacteria are among the substances filtered before the water is released into the river. With the wastewater treatment plant, the treated water will be improved to class II, making it suitable for recreational purposes to be compliance with the National Water Quality Standard.

5.3. Reducing pollution load from rural sub-basin using constructed wetland

Constructed wetlands are now recognized as a reliable and adaptable wastewater treatment technology and a suitable solution for the treatment of many different types of wastewater (Vymazal, 2011). Constructed wetlands are effective at attenuating small storm flows, which refers to the gradual reduction in the rate and volume of runoff that occurs as it passes through the wetland. This process is achieved through a combination of physical, chemical and biological processes that remove pollutants from the runoff and slow down the flow rate (Vymazal, 2011). The constructed wetland area could also provide some volume of attenuation of small storm flows. However, the effectiveness of constructed wetlands in attenuating storm flows is limited, and it varies depending on factors such as the size of the wetland and the intensity of the storm event. This is because the inundation of wetland plants with high volumes of water for extended periods can cause physical damage and loss of vegetation (Donchez et al., 2017).

In Batu Pahat river basin, there are 61 numbers of villages and husbandries which are located outside of Indah Water Konsortium (Malaysian national wastewater and sanitation company) service area. The population density in rural areas is typically lower than in urban areas, which makes it less practical to have a large-scale centralized sewage treatment system. Instead, a decentralized sewage treatment system that can cover a relatively large area is needed. Decentralize systems typically consist of smaller treatment units, such as septic tanks, that are installed at individual homes or clusters of homes. Given the economic situation of rural areas, systems with low construction, maintenance, and operation costs are required. In order to reduce pollution loading from rural house and husbandry, a constructed wetland has to be constructed at final outlet of each village.

A constructed wetland system can remove pollutants from storm water by using natural processes, like sedimentation, filtration, and biological uptake. These systems typically consist of shallow water bodies that are extensively vegetated with emergent plants (Vymazal, 2010; Vymazal, 2011, 2007). The system is designed to slowly release water after rainfall events by slowly raising the water levels in the wetland. Furthermore, constructed wetlands can be used for the treatment of storm water, for the preservation of wildlife habitat, for passive recreation, and for the improvement of landscape amenity (Vymazal, 2011).

5.4. Reducing sediment load by establishing river bank erosion hazard zone

The effects of anthropogenic activities on sediment concentration of rivers including land use practices, dam construction, footpaths, water conservation practices, construction of roads and other infrastructures. It was revealed that anthropogenic activity is a cause for concern, especially in industrial countries, as it is responsible for considerable variation in sediment concentration trend along relatively small subcatchments. The increase of sediment concentration due to the excessive wash load entering the Sg Chaah and Sg Bekok river reaches are mainly caused by the palm oil replantation, vegetable cultivation and agriculture activities along Sg Chaah and Sg Bekok. The effects of anthropogenic activities on streamflow hydrology and morphology also revealed that sand mining for road and infrastructural construction as well as over withdrawal of particles may enhance sediment concentration in rivers. To reduce the impact of excessive wash load delivered to the river system, a series of actions must be taken, which include preventing illegal cultivation along the river by enforcing guidelines and procedures, enforcing the Erosion and Sediment Control Plan (ESCP) on agricultural activities such as replantation of palm oil and others, prohibiting livestock farming along river reserves, not allowing mining activities in the river, treating mining wastewater to comply with standards before discharging it into the watercourse, and enforcing the

Erosion and Sediment Control Plan (ESCP) on mining and quarry stakeholders to implement primary erosion and sediment control measures.

6. Conclusion

In conclusion, this study demonstrates the efficacy of a feedforward artificial neural network in predicting the water quality index of the Batu Pahat River in real-time, using a reduced number of water quality parameters as model inputs. The study conducted revealed that the prediction accuracy was found to be at its peak of 97.99% when BOD was excluded from the input variables. In addition, an exceptional prediction accuracy of 91.03% was achieved when both COD and BOD were not included as input variables. These outcomes were more favorable than the benchmark value of 90% prediction accuracy, thereby suggesting the feasibility of a reduction in the number of water quality parameters in a given monitoring process. The findings suggest that omitting parameters such as COD and BOD, which cannot be measured in real-time, does not significantly affect the accuracy of the model. The study's results have important implications for improving water resources management by reducing the cost and time required for monitoring while maintaining high accuracy levels. It simplifies the data collection and analysis process, reducing the resource required to monitor and measure a large number of water quality parameters. This can make real-time WQI prediction more accessible and cost-effective for water management authorities and other stakeholders. Besides, the reduced number of parameters as inputs also could help to address the issue of missing or incomplete data, which is often a challenge in water quality monitoring. By relying on a smaller set of input parameters, the model still can provide reliable WQI predictions even when some data is missing or unavailable. Moreover, this can be particularly beneficial for areas with limited resources or where water quality monitoring is not a priority. Besides, emerging algorithms such as Random Tree (RT), Random Forest (RF), M5P, Reduced Error Pruning Tree (REPTree), Random Committee (RC), Bagging, and Instance-Based k-Nearest Neighbors (IBK) have found application in addressing challenges within the domains of hydrology, climatology, and hydraulics.

Furthermore, this study provides valuable insights for the management and enhancement of the water quality in Batu Pahat River. This research has recognized and identified specific sources of pollution and developing targeted mitigation plans to improve water quality. Thus, effective regulation and enforcement of environmental laws are needed to hold industries accountable for proper waste handling and disposal. Additionally, industries should adopt sustainable practices, such as implementing efficient wastewater treatment systems, reducing the use of harmful chemicals, and promoting responsible waste management, to prevent river pollution and ensure a healthy environment for imminent generations.

7. Declarations

Ethics Approval Not applicable.
Consent to Participate All authors give their consent to participate.
Consent to Publish All authors give their consent to publish.

Funding

The work was supported by Tenaga Nasional Berhad (TNB) and UNITEN through the BOLD Refresh Postdoctoral Fellowship and BOLD Refresh Publication Fund under the project code J510050002-IC-6 BOLDREFRESH2025-Centre of Excellent andunder project of National Flood Forecasting and Warning System (NaFFWS) 2B – Project Code (JPS/IP/C/H/01/2022).

CRediT authorship contribution statement

Syahida Farhan Azha: Conceptualization, Methodology, Validation, Formal analysis, Writing – original draft, Writing – review & editing. **Lariyah Mohd Sidek:** Supervision, Methodology. **Zainal Ahmad:** Conceptualization, Methodology, Validation, Formal analysis, Writing – review & editing. **Jie Zhang:** Supervision, Methodology. **Hidayah Basri:** Supervision, Methodology. **Mohd Hafiz Zawawi:** Supervision, Methodology.**Nurshahira Mohammad Noh:** Methodology, Validation. **Ali Najah Ahmed:** Conceptualization, Methodology, Validation, Formal analysis, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The authors extend their sincere appreciation to the Department of Irrigation and Drainage (DID) of Malaysia for their invaluable support and collaboration throughout this project. Gratitude is also expressed to ZHL Engineers Sdn. Bhd. for their expert technical guidance and contributions, which significantly contributed to the success of this study.

References

Abba, S.I., Jasim, S., Abdullahi, J., 2018. River water modelling prediction using multi-linear regression, artificial neural network, and adaptive neuro-fuzzy inference system techniques. Procedia Comput. Sci. 120, 75–82. https://doi.org/10.1016/j.procs.2017.11.212.

Abba, S.I., Pham, Q.B., Saini, G., Linh, N.T.T., Ahmed, A.N., Mohajane, M., Khaledian, M., Abdulkadir, R.A., Bach, Q.V., 2020. Implementation of data intelligence models coupled with ensemble machine learning for prediction of water quality index. Environ. Sci. Pollut. Res. 27, 41524–41539. https://doi.org/10.1007/s11356-020-09689-x.

Abyaneh, H.Z., 2014. Evaluation of multivariate linear regression and artificial neural networks in prediction of water quality parameters. J. Environ. Heal. Sci. Eng. 12, 40. https://doi.org/10.1186/2052-336X-12-40.

Adam, M.R., Othman, M.H.D., Abu Samah, R., Puteh, M.H., Ismail, A.F., Mustafa, A., A. Rahman, M., Jaafar, J., 2019. Current trends and future prospects of ammonia removal in wastewater: A comprehensive review on adsorptive membrane development. Sep. Purif. Technol. 213, 114–132. doi: 10.1016/j.seppur.2018.12.030.

Adnan, M., Roslen, H., Samsuri, S., 2022. The Application of Total Maximum Daily Load (TMDL) Approach in Water Quality Assessment for The Batu Pahat River, in: IOP Conference Series: Earth and Environmental Science. doi: 10.1088/1755-1315/1022/1/012074.

Adnan, R., Ruslan, F.A., Samad, A.M., Zain, Z., 2012. Flood water level modelling and prediction using artificial neural network: Case Study of Sungai Batu Pahat in Johor. IEEE 22–25.

Ahmad, Z., Rahim, N.A., Bahadori, A., Zhang, J., 2016. Improving water quality index prediction in Perak River basin Malaysia through a combination of multiple neural networks: Improving water quality index prediction in Perak River basin Malaysia through a combination of multiple neural networks. Intl. J. River Basin Manag. 1–9. https://doi.org/10.1080/15715124.2016.1256297.

Ahmed, A.N., Binti Othman, F., Abdulmohsin Afan, H., Khaleel Ibrahim, R., Ming Fai, C., Shabbir Hossain, M., Ehteram, M., Elshafie, A., 2019. Machine learning methods for better water quality prediction. J. Hydrol. 578 https://doi.org/10.1016/j.jhydrol.2019.124084.

Ahmed, W., Payyappat, S., Cassidy, M., Besley, C., Power, K., 2018. Novel crAssphage marker genes ascertain sewage pollution in a recreational lake receiving urban stormwater runoff. Water Res. 145, 769–778. https://doi.org/10.1016/j.watres.2018.08.049.

Back, S., Saito, N., Lee, S., 2023. A facile and efficient approach for the removal of high concentrations of ammonia nitrogen in wastewater: Liquid-phase plasma treatment. J. Environ. Chem. Eng. 11, 109075 https://doi.org/10.1016/j.jece.2022.109075.

Bateni, N., Hashim, N.B., Asfar, J., Selaman, O.S., 2009. Comparison of Wave Hindcasting of Linear Wave Theory and Swan Modeling at Batu Pahat River Estuary, in: International Conference on Water Resources (ICWR 2009).

Bhateria, R., Jain, D., 2016. Water quality assessment of lake water: A review. Sustain. Water Resour. Manag. doi: 10.1007/s40899-015-0014-7.

Bolboacă, R., Haller, P., 2023. Performance analysis of long short-term memory predictive neural networks on time series data. Mathematics. https://doi.org/10.3390/math11061432.

Brias, A., Mathias, J.-D., Deffuant, G., 2018. Inter-annual rainfall variability may foster lake regime shifts: An example from Lake Bourget in France. Ecol. Modell. 389, 11–18. https://doi.org/10.1016/j.ecolmodel.2018.10.004.

Chen, Y., Song, L., Liu, Y., Yang, L., Li, D., 2020. A Review of the artificial neural network models for water quality prediction. Appl. Sci. 10.

Djarum, D.H., Ahmad, Z., Zhang, J., 2022. Performance analysis of neural network architecture in developing real-time Malaysian river water quality model. IOP Conf. Ser. Mater. Sci. Eng. 1257, 012022 https://doi.org/10.1088/1757-899x/1257/1/012022.

DOE [WWW Document], 2020. Stand. Kual. Air Sungai Kebangs. dan Indeks Kual. Air Sungai. URL https://www.doe.gov.my/en/regulation/ (accessed 4.17.22).

Donchez, R., Callahan, B.G., Colón, M.G., Evans, E.R., Martell, S.M., Negrón, O., Reynolds, J.W., Waldron, A.R., 2017. Green Infrastructure Guideline. Lehigh Val. Comm. Plan. 1–64.

DSDG, 2023. The 17 Sustainable Development Goals [WWW Document]. accessed 2.24.23. https://sdgs.un.org/.

Ezemagu, I.G., Ejimofor, M.I., Menkiti, M.C., Nwobi-Okoye, C.C., 2021. Modeling and optimization of turbidity removal from produced water using response surface methodology and artificial neural network. South African J. Chem. Eng. 35, 78–88. https://doi.org/10.1016/j.sajce.2020.11.007.

Fitri, A., Abdul Maulud, K.N., Pratiwi, D., Phelia, A., Rossi, F., Zuhairi, N.Z., 2020. Trend of water quality status in Kelantan River Downstream, Peninsular Malaysia. J. Rekayasa Sipil 16, 178. https://doi.org/10.25077/jrs.16.3.178-184.2020.

Gao, F., Zhang, B., 2023. Data-Aware Customization Of Activation Functions Reduces Neural network error [WWW Document]. Cornell Univ.

Ghaedi, A.M., Vafaei, A., 2017. Applications of artificial neural networks for adsorption removal of dyes from aqueous solution: A review. Adv. Colloid Interface Sci. 245, 20–39. https://doi.org/10.1016/j.cis.2017.04.015.

Hazritauding, M.A., Adnan, M.S., 2022. Assessment of Water Quality Parameters due to High and low Flow for Batu Pahat River, Johor. Recent Trends Civ. Eng. Built Environ. 3, 596–605.

Hecht-Nielsen, R., 1987. Kolmogorov's Mapping Neural Network Existence Theorem.

Ho, J.Y., Afan, H.A., El-Shafie, A.H., Koting, S.B., Mohd, N.S., Jaafar, W.Z.B., Lai Sai, H., Malek, M.A., Ahmed, A.N., Mohtar, W.H.M.W., Elshorbagy, A., El-Shafie, A., 2019. Towards a time and cost effective approach to water quality index class prediction. J. Hydrol. 575, 148–165. https://doi.org/10.1016/j.jhydrol.2019.05.016.

Ismail, M.A.N., 2007. Modelling of Tidal Effect on Suspended Sediment Dispersion at Sungai Batu Pahat, Johor.

Jais, N.M., Miswan, M.S., Radin Mohamed, R.M.S., Al-Gheethi, A.A.S., 2020. Treatment technologies of fresh market wastewater, in: Al-Gheethi, A.A.S., Radin Mohamed, R. M.S., Noman, E.A., Mohd Kassim, A.H. (Eds.), Prospects of Fresh Market Wastes Management in Developing Countries. Springer International Publishing, Cham, pp. 59–76. doi: 10.1007/978-3-030-42641-5_4.

Jia, Z., Chang, X., Duan, T., Wang, X., Wei, T., Li, Y., 2021. Water quality responses to rainfall and surrounding land uses in urban lakes. J. Environ. Manage. 298, 113514 https://doi.org/10.1016/j.jenvman.2021.113514.

Juahir, H., Zain, S.M., Toriman, M.E., Mokhtar, M., Man, H.C., 2004. Application of artificial neural network models for predicting water quality index. J. Kejuruter. Awam 16, 42–55.

Juahir, H., Zain, S.M., Kamil, M., Hanidza, T.I.T., Armi, A.S.M., Ekhwan, M., Mazlin, T., 2011. Spatial water quality assessment of Langat River Basin (Malaysia) using environmetric techniques. Environ. Monit. Assess 173, 625–641. https://doi.org/10.1007/s10661-010-1411-x.

Kadam, A.K., Wagh, V.M., Muley, A.A., Umrikar, B.N., Sankhua, R.N., 2019. Prediction of water quality index using artificial neural network and multiple linear regression modelling approach in Shivganga River basin, India. Model. Earth Syst. Environ. 5, 951–962. https://doi.org/10.1007/s40808-019-00581-3.

Kanda, E.K., Kosgei, E.C.K.J.R., 2016. Dissolved oxygen modelling using artificial neural network: A Case o River Nzoia, Lake Victoria Basin, Kenya. J. Water Secur. 2, 1–7.

Kavzoglu, T., Mather, P.M., 2003. The use of backpropagating artificial neural networks in land cover classification classification. Int. J. Remote Sens. 24, 4907–4938. https://doi.org/10.1080/0143116031000114851.

Khan, Y., See, C.S., 2016. Predicting and analyzing water quality using Machine Learning: A comprehensive model. In: 2016 IEEE Long Island Systems, Applications and Technology Conference (LISAT). pp. 1–6. doi: 10.1109/LISAT.2016.7494106.

Khoi, D.N., Quan, N.T., Linh, D.Q., Nhi, P.T.T., Thuy, N.T.D., 2022. Using machine learning models for predicting the water quality index in the La Buong River, Vietnam. Water (Switzerland) 14, 1–12. https://doi.org/10.3390/w14101552.

Khouri, L., Bashar Al-Moufti, M., 2022. Selection of suitable aggregation function for estimation of water quality index for the Orontes River. Ecol. Indic. 142, 109290 https://doi.org/10.1016/j.ecolind.2022.109290.

Khullar, S., Singh, N., 2021. Machine learning techniques in river water quality modelling: A research travelogue. Water Sci. Technol. Water Supply 21. https://doi.org/10.2166/ws.2020.277.

Kotu, V., Deshpande, B., 2014. Predictive Analytics and Data Mining: Concepts and Practice with Rapidminer. Morgan Kaufmann.

Latiff, N.H., Aini, N., Arish, M., 2021. Prediction of Future Temperature and Rainfall Characteristics Using Statistical Downscaling Model (SDSM) for Empangan Sg. Sembrong in Batu Pahat Catchment. Prog. Eng. Appl. Technol. 2, 300–310.

Li, X., Huang, T., Ma, W., Sun, X., Zhang, H., 2015. Effects of rainfall patterns on water quality in a stratified reservoir subject to eutrophication: Implications for management. Sci. Total Environ. 521–522, 27–36. https://doi.org/10.1016/j.scitotenv.2015.03.062.

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, 2058. https://doi.org/10.3390/su11072058.

López, O.A.M., López, A.M., Crossa, J., 2022. Fundamentals of artificial neural networks and deep learning. In: Montesinos López, O.A., Montesinos López, A., Crossa, J. (Eds.), Multivariate Statistical Machine Learning Methods for Genomic Prediction. Springer International Publishing, Cham, pp. 379–425. https://doi.org/10.1007/978-3-030-89010-0_10.

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. Environ. Model. Softw. 25, 891–909. https://doi.org/10.1016/j.envsoft.2010.02.003.

Marselina, M., Wibowo, F., Mushfiroh, A., 2022. Water quality index assessment methods for surface water: A case study of the Citarum River in Indonesia. Heliyon 8. https://doi.org/10.1016/j.heliyon.2022.e09848.

Mohammadpour, R., Shaharuddin, S., Chang, C.K., Zakaria, N.A., Ghani, A.A., Chan, N. W., 2015. Prediction of water quality index in constructed wetlands using support vector machine. Environ. Sci. Pollut. Res. 22, 6208–6219. https://doi.org/10.1007/s11356-014-3806-7.

Mohammed, A.F.M., Adnan, M.S., Muneer, A., Sadeq, S., 2021. Flood estimation studies using hydrologic modelling system (HEC-HMS) for Batu Pahat River Malaysia. In: 2021 3rd Int. Sustain. Resil. Conf. Clim. Chang. 225–230. doi: 10.1109/IEEECONF53624.2021.9667988.

Mohammed, I., Hashim Al-Khalaf, S.K., Alwan, H.H., Samir Naje, A., 2022. Environmental assessment of Karbala water treatment plant using water quality index (WQI). Mater. Today Proc. 60, 1554–1560. https://doi.org/10.1016/j.matpr.2021.12.065.

Mokhtar, M., Tun, U., Onn, H., Daud, M.E., Tun, U., Onn, H., Kaamin, M., Tun, U., Onn, H., Azim, M., Azmi, M., Tun, U., Onn, H., 2022. Sediment properties of eroded coastal area at Batu. Environ. Ecol. Res. 10, 248–259. https://doi.org/10.13189/eer.2022.100214.

Moradi, N., Moezzi, A., Khajavi-Shojaei, S., Khaji, P., 2022. Cadmium immobilization in contaminated soil by nano-biohar and Fe-modified nano-biochar. Iran. J. Soil Water Res. https://doi.org/10.22059/ijswr.2022.337907.669193.

Nasir, N., Kansal, A., Alshaltone, O., Barneih, F., Sameer, M., Shanableh, A., Al-Shamma'a, A., 2022. Water quality classification using machine learning algorithms. J. Water Process. Eng. 48, 102920. doi: 10.1016/j.jwpe.2022.102920.

Nong, X., Lai, C., Chen, L., Shao, D., Zhang, C., Liang, J., 2023. Prediction modelling framework comparative analysis of dissolved oxygen concentration variations using support vector regression coupled with multiple feature engineering and optimization methods: A case study in China. Ecol. Indic. 146, 109845 https://doi.org/10.1016/j.ecolind.2022.109845.

Rahman, M.A., Yusof, S., 2015. Flood prone areas detection through geographical information system (GIS) and water balance model in Batu Pahat. International Conference on Sustainable Environment & Water Research.

Ranjan, P., Khan, P., Kumar, S., Das, S.K., 2023. log-sigmoid activation-based long short-term memory for time series data classification. IEEE Trans. Artif. Intell. 1–12 https://doi.org/10.1109/TAI.2023.3265641.

Sahoo, G.B., Ray, C., Wade, H.F., 2005. Pesticide prediction in ground water in North Carolina domestic wells using artificial neural networks. Ecol. Modell. 183, 29–46. https://doi.org/10.1016/j.ecolmodel.2004.07.021.

Sakizadeh, M., 2015. Artificial intelligence for the prediction of water quality index in groundwater systems. Model. Earth Syst. Environ. 2, 8. https://doi.org/10.1007/s40808-015-0063-9.

Salim, N.F.A., Kasmin, H., 2022. Study on the Water Quality Index (WQI) of Parit Besar River in Batu Pahat. J. Adv. Environ. Solut. Resour. Recover. 2, 8–14.

Sapna, S., Tamilarasi, A., Kumar, M.P., 2012. Backpropagation learning algorithm based on Levenberg Marquardt algorithm. Comput. Sci. Inf. Technol. 393–398.

Sarkar, A., Pandey, P., 2015. River water quality modelling using artificial neural network technique. Aquat. Procedia 4, 1070–1077. https://doi.org/10.1016/j.aqpro.2015.02.135.

Sidek, L.M., Jajarmizadeh, A., Mohiyaden, H.A., Noh, S.M., Puad, A.H.M., Shahrir, A.M. D., 2022. Water Quality Assessment for Batu Pahat River Basin, Malaysia. Lect. Notes Civ. Eng. 179, 9–28. https://doi.org/10.1007/978-981-16-5041-3_2.

Singh, B., Sihag, P., Singh, V.P., Sepahvand, A., Singh, K., 2021. Soft computing technique-based prediction of water quality index. Water Supply 21. https://doi.org/10.2166/ws.2021.157.

Tjahjanto, D.W.I., Rahmat, N., Mohd, M.I., Musa, S., Manan, I.A.B.D., 2008. A Study of Sediment Load : Case Study at Parit Botak Channel, Batu Pahat Johor Malaysia. iI: International Conference on Environment 2008 (ICENV 2008). pp. 1–8.

Tonetti, A.L., De Camargo, C.C., Guimarães, J.R., 2016. Ammonia removal from landfill leachate by struvite formation: An alarming concentration of phosphorus in the treated effluent. Water Sci. Technol. 74, 2970–2977. https://doi.org/10.2166/wst.2016.490.

Venkata Vara Prasad, V.V.P., Y Venkataramana, L., Kumar, P.S., G, P., K, S., A.J, P., 2020. Water quality analysis in a lake using deep learning methodology: prediction and validation. Int. J. Environ. Anal. Chem. doi: 10.1080/03067319.2020.1801665.

Verma, A.K., Singh, T.N., 2013. Prediction of water quality from simple field parameters. Environ. Earth Sci. 69, 821–829. https://doi.org/10.1007/s12665-012-1967-6.

Vymazal, J., 2007. Removal of nutrients in various types of constructed wetlands. Sci. Total Environ. 380, 48–65. https://doi.org/10.1016/j.scitotenv.2006.09.014.

Vymazal, 2010. Constructed wetlands for wastewater treatment. Water. doi: 10.3390/w2030530.

Vymazal, J., 2011. Constructed wetlands for wastewater treatment: five decades of experience. Environ. Sci. Technol. 45, 61–69. doi: 10.1021/es101403q.
Wan Mohtar, W.H.M., Nawang, S.A.B., Abdul Maulud, K.N., Benson, Y.A., Azhary, W.A. H.W.M., 2017. Textural characteristics and sedimentary environment of sediment at eroded and deposited regions in the severely eroded coastline of Batu Pahat, Malaysia. Sci. Total Environ. 598, 525–537. https://doi.org/10.1016/j.scitotenv.2017.04.093.
Yona, C., Makange, M., Moshiro, E., Chengula, A., Misinzo, G., 2023. Water pollution at Lake Natron Ramsar site in Tanzania: A threat to aquatic life. Ecohydrol. Hydrobiol. 23, 98–108. https://doi.org/10.1016/j.ecohyd.2022.11.001.
Zaini, N., Ean, L.W., Ahmed, A.N., Malek, M.A., 2022. A systematic literature review of deep learning neural network for time series air quality forecasting. Environ. Sci. Pollut. Res. 29, 4958–4990. https://doi.org/10.1007/s11356-021-17442-1.