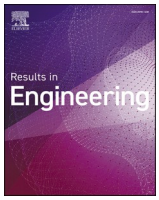
Results in Engineering 22 (2024) 102072

Contents lists available at ScienceDirect

Results in Engineering

journal homepage: www.sciencedirect.com/journal/results-in-engineering



Enhancing riverine load prediction of anthropogenic pollutants: Harnessing the potential of feed-forward backpropagation (FFBP) artificial neural network (ANN) models

Khairunnisa Khairudin [a,f], Ahmad Zia Ul-Saufie [b], Syahrul Fithry Senin [c], Zaki Zainudin [d], Ammar Mohd Rashid [e], Noor Fitrah Abu Bakar [a], Muhammad Zakwan Anas Abd Wahid [f], Syahida Farhan Azha [g], Firdaus Abd-Wahab [h], Lei Wang [i], Farisha Nerina Sahar [f], Mohamed Syazwan Osman [f,j,*]

[a] School of Chemical Engineering, College of Engineering, Universiti Teknologi MARA, 40450, Shah Alam, Selangor, Malaysia
[b] School of Mathematical Sciences, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, 40450, Shah Alam, Selangor, Malaysia
[c] Centre for Civil Engineering Studies, College of Engineering, Universiti Teknologi MARA, Cawangan Pulau Pinang, 13500, Permatang Pauh, Pulau Pinang, Malaysia
[d] Independent Water Quality and Modelling Specialist, Klang, Selangor, Malaysia
[e] AMR Environmental Sdn. Bhd, No. 166-1 Jalan S2 B22, Pusat Dagangan Seremban 2, 70300, Seremban, Negeri Sembilan, Malaysia
[f] EMZI-UiTM Nanoparticles Colloids & Interface Industrial Research Laboratory (NANO-CORE), Chemical Engineering Studies, College of Engineering, Universiti Teknologi MARA, Cawangan Pulau Pinang, 13500, Permatang Pauh, Pulau Pinang, Malaysia
[g] Institute of Energy Infrastructure (IEI) and Department of Civil Engineering, College of Engineering, Universiti Tenaga Nasional (UNITEN), 43000, Kajang, Selangor Darul Ehsan, Malaysia
[h] Faculty of Engineering, International Islamic University Malaysia, P.O. Box 10, Kuala Lumpur, 50728, Malaysia
[i] School of Chemistry and Chemical Engineering, Harbin Institute of Technology, Harbin, 150001, Heilongjiang, PR China
[j] Center of Excellence Geopolymer & Green Technology (CEGeoGTech), Universiti Malaysia Perlis (UniMAP), 01000, Kangar, Perlis, Malaysia

ARTICLE INFO

Keywords:
Riverine load
Artificial neural network (ANN)
Feed-forward backpropagation algorithm
Radial basis neural network
Multiple linear regression

ABSTRACT

Assessing riverine pollutant loads is a more realistic method for analysing point and non-point anthropogenic pollution sources throughout a watershed. This study compares numerous mathematical modelling strategies for estimating riverine loads based on the chosen water quality parameters: Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Suspended Solids (SS), and Ammoniacal Nitrogen (NH_3–N). A riverine load model was developed by employing various input variables including river flow and pollutant concentration values collected at several monitoring sites. Among the mathematical modelling methods employed are artificial neural networks with feed-forward backpropagation algorithms and radial basis functions. The classical multiple linear regression (MLR) statistical model was used for the comparison. Four widely used statistical performance assessment metrics were adopted to evaluate the performance of the various developed models: the root mean square error (RMSE), mean absolute error (MAE), mean relative error (MRE), and coefficient of determination (R^2). The considerable number of errors (with RMSE, MAE, and MRE) discovered in estimating riverine loads using the multiple linear regression (MLR) statistical model can be attributed to the nonlinear relationship between the independent variables (Q and C_x) and dependent variables (W). The feed-forward neural network model with a backpropagation algorithm and Bayesian regularisation training algorithm outperformed the radial basis neural network. This finding implies that, in addition to suspended sediment loads, riverine loads may be predicted using an artificial neural network using pollutant concentration (C_x) and river discharge (Q) as input variables. Other geographical and temporal fluctuation characteristics that may impact river water quality, on the other hand, may be incorporated as input variables to enhance riverine load prediction. Finally, riverine load analyses were successfully conducted to reduce the riverine load.

* Corresponding author. EMZI-UiTM Nanoparticles Colloids & Interface Industrial Research Laboratory (NANO-CORE), Chemical Engineering Studies, College of Engineering, Universiti Teknologi MARA, Cawangan Pulau Pinang, 13500, Permatang Pauh, Pulau Pinang, Malaysia.
E-mail address: syazwan.osman@uitm.edu.my (M.S. Osman).

https://doi.org/10.1016/j.rineng.2024.102072
Received 1 August 2023; Received in revised form 16 March 2024; Accepted 27 March 2024
Available online 4 April 2024

1. Introduction

Rivers, the lifeblood of environmental ecosystems, and the cornerstone of human civilisation are indispensable not only for their ecological importance, but also for their vital roles in personal hygiene, drinking water, agricultural irrigation, industrial processes, and as pivotal sources of hydroelectric power [1–3]. The quality of these vital waterways, encompassing a complex interplay of physical, chemical, and biological characteristics, transcends metrics that reflect the health of ecosystems and their capacity to support human life [4,5]. However, this crucial resource is under siege from anthropogenic activities; urbanization and industrialization have led to the deterioration of river water quality through the discharge of pollutants from urban drains, agricultural runoff laden with pesticides and fertilizers, and industrial effluents [6–8]. This escalating crisis highlights the urgent need for innovative engineering solutions, which are being addressed by exploring new methodologies to assess, monitor, and enhance river water quality. In doing so, it aims to redefine our interactions with these natural resources, merging environmental stewardship with engineering prowess to safeguard the irreplaceable lifelines of our planet for future generations.

Malaysian river systems face a multitude of pollution sources, primarily sewage disposal, effluent discharge from small- and medium-sized industries that lack sufficient effluent treatment facilities, extensive earthwork operations, and land clearing activities [9]. Notably, in 1999, alarming 42% of Malaysia's river basins exhibited pollution attributable to suspended solids (SS), a consequence of poorly planned and unregulated land clearing activities. An additional 30% of these river basins were contaminated by organic matter originating from industrial effluents, whereas 28% were contaminated with ammoniacal nitrogen (NH_3–N) resulting from sewage disposal and animal husbandry practices [10,11]. The adverse consequences of this pervasive degradation of river water quality extend far beyond the environmental domain, encompassing significant ramifications for human health, ecological stability, and national economic well-being [12–14]. Thus, there is an imperative for a comprehensive and continuous assessment and monitoring, serving as the foundational step required to facilitate the formulation and implementation of more efficacious strategies aimed at safeguarding and preserving this invaluable water resource.

Historically, the collection of water quality data hinged upon the labourious practice of on-site monitoring, a method characterised by its resource-intensive nature, necessitating not only a cadre of trained professionals but also the allocation of consistent resources [15] However, as contemporary research endeavours delve into the intricate complexities of water quality dynamics, a pivotal shift in methodology has become apparent. Although deterministic [16,17] and stochastic models [18,19] for water quality prediction have emerged as noteworthy endeavours, they present formidable challenges. Their implementation requires assimilation of extensive datasets and intricate model structures to support robust analysis [20]. Within this evolving landscape, paradigmatic transformation towards statistical modelling has found its genesis. These novel statistical models, underpinned by the assumption of linear relationships and normal distribution properties between the response and predictor variables, represent a significant departure from traditional methods [16]. However, as water quality emerges as an increasingly multifaceted concern, characterised by intricate nonlinear relationships between numerous influencing factors, these conventional linear data analysis methods gradually reveal their limitations [21].

In response to this burgeoning complexity, the scientific community has begun to pivot towards Artificial Neural Networks (ANNs) as a promising alternative approach for water quality parameter prediction. ANNs, which have witnessed significant advancements since their inception, stand at the forefront of data-driven modelling methodologies. Their capacity to model intricate nonlinear interactions between multiple variables in complex systems renders them invaluable tools [11,22–25]. In recent decades, ANNs have been demonstrated to predict a diverse array of water quality parameters across various aquatic environments, including salinity, total dissolved solids (TDS), electrical conductivity, turbidity, chemical oxygen demand (COD), biochemical oxygen demand (BOD), dissolved oxygen (DO), ammoniacal nitrogen (NH_3–N), and water quality indices (WQI) [20,21,24,26,27]. This shift towards ANNs represents not just a departure from traditional methods but a strategic response to the evolving demands of modern water quality assessment, catering to the multifaceted nature of the challenges faced in safeguarding this critical natural resource.

Artificial Neural Networks (ANNs) are remarkable innovations in river water quality prediction that offer several advantages. First, ANNs excel at capturing intricate nonlinear relationships among a multitude of influencing factors, a feature that is unattainable by traditional linear models [21,28]. This nonlinear modelling capability is particularly advantageous in the water quality domain, where the interactions between variables are often complex and dynamic. Additionally, ANNs exhibit adaptability and resilience in handling missing or noisy data, thereby enhancing their robustness in real-world applications [27]. Their inherent ability to self-learn and adapt to evolving conditions renders them suitable for accommodating the temporal and spatial fluctuations characteristic of river ecosystems [20]. The versatility of ANNs has been harnessed across diverse research domains, particularly in water quality prediction, where their application has yielded significant advancements [22,25,29]. This burgeoning body of research underscores the pivotal role of ANNs as a transformative tool in the contemporary landscape of water quality assessment, offering unparalleled predictive capabilities and opening new horizons for understanding and managing vital aquatic ecosystems [30].

The predominant focus on pollutant concentration prediction within artificial neural network (ANN) applications for water quality research has obscured a critical research gap concerning the assessment of pollutant loads, particularly in the context of both point and nonpoint pollution sources within the entire watershed. This gap is particularly relevant, given the increasing global concern regarding water quality and environmental conservation. Therefore, this study introduces a pioneering approach that harnesses the advanced capabilities of a Feed-Forward Backpropagation Neural Network (FFBP NN) for riverine pollutant load modelling, which has not been employed in previous studies. By shifting from a traditional concentration-centric paradigm to one that emphasises pollutant load assessment, this innovation offers several advantages. First, the FFBP NN-based approach provides a more comprehensive and holistic understanding of watershed pollution by capturing the intricate interplay between key pollutants including biochemical oxygen demand (BOD), chemical oxygen demand (COD), ammoniacal nitrogen (NH_3–N), and suspended solids (SS) across all watersheds. This departure from concentration-centric modelling allows for a more accurate evaluation of the contributions of point and nonpoint pollution sources, thereby enhancing the overall accuracy of environmental assessments [31]. Furthermore, the method is equipped with policymakers and environmentalists with a robust and scalable tool for data-driven interventions, thereby establishing new standards for environmental monitoring and river conservation. In an era marked by escalating environmental degradation and the urgent need for sustainable resource management, this innovative FFBP NN-based approach represents a significant leap forward in addressing this critical research gap and advancing the field of riverine pollutant load modelling. The principal aim of this study was to develop a comprehensive riverine load model encompassing the following key water quality parameters based on an Artificial Neural Network model: biochemical oxygen demand (BOD), chemical oxygen demand (COD), ammoniacal nitrogen (NH_3–N), and suspended solids (SS). To achieve this goal, we leveraged a feed-forward backpropagation neural network (FFBP NN), which is a powerful computational tool known for its capacity to effectively capture complex nonlinear relationships within environmental datasets. The inherent ability of the FFBP NN to adapt and optimise model

parameters, coupled with its capacity to handle intricate interdependencies, makes it a robust choice for modelling riverine loading water quality dynamics. To rigorously assess and validate the predictive performance of the model, we conducted comparative analyses with alternative predictive analytical methods, specifically, the radial basis neural network (RBNN) and multiple linear regression (MLR). This research not only advances the field of environmental modelling but also underscores the advantages of utilising FFBP NN for enhancing the accuracy and predictive capabilities of riverine loading water quality assessments.

2. Methodology

2.1. Study area description

The Muda River was chosen as the study area to develop an ANN model for predicting riverine loads. It is the longest river in the state of Kedah, with a total length of 180 km and a catchment area of 4210 km^2 [32] The river basin land is primarily used for forestry and cultivation such as paddy, palm oil, and rubber [33]. This is supported by the ArcGIS mapping of the Muda River, as shown in Fig. 1. In the 1000 m

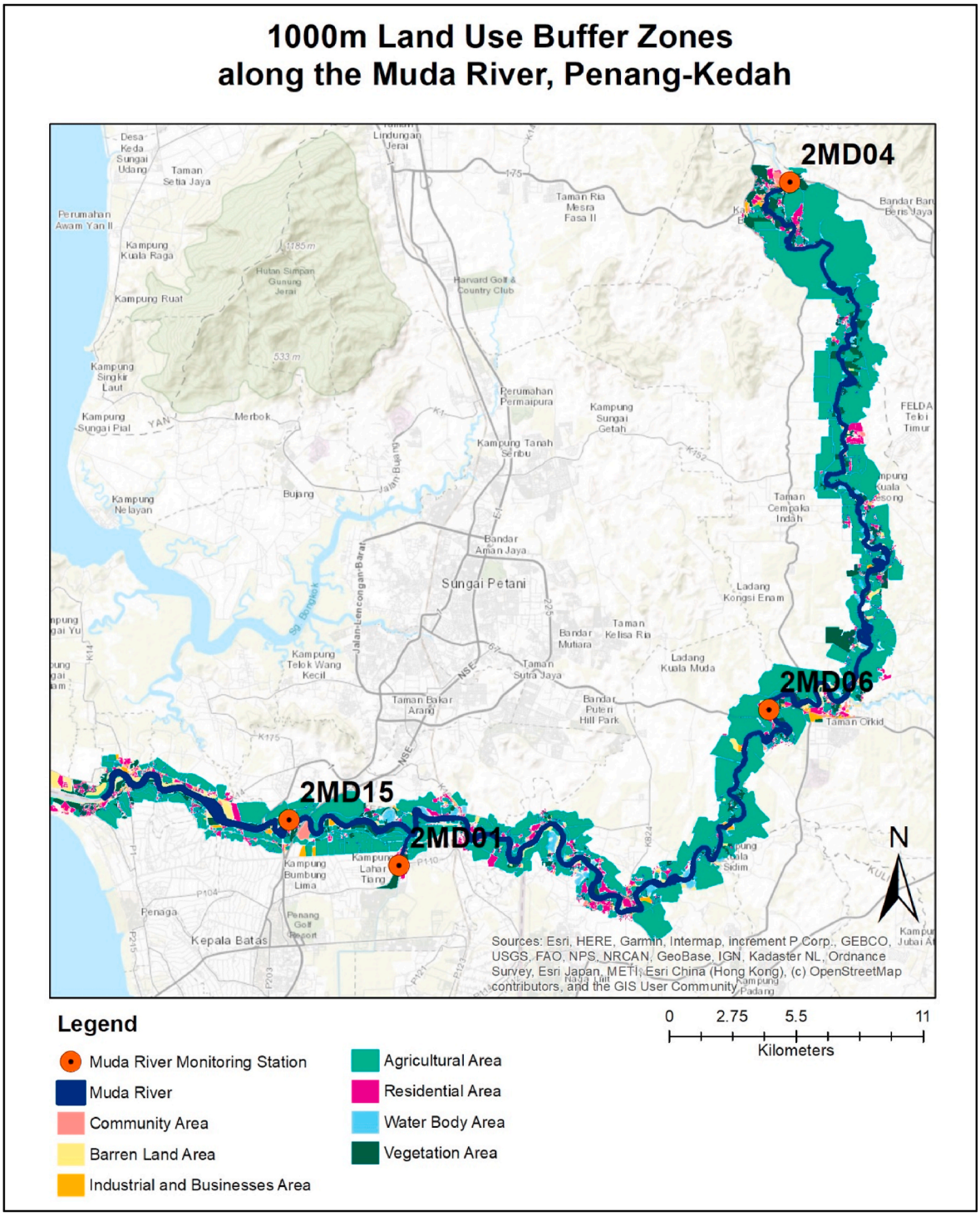


Fig. 1. ArcGIS mapping of monitoring stations in the Muda river basin.

buffer zone scale, the majority of land along the Muda River is agricultural. The land was also covered by vegetation (forests), residential areas, water bodies, community areas, barren land, and industrial areas. In this study, a 1000 m buffer zone was used because Huang et al. discovered that buffer zone scales greater than 1000 m of land cover and land use have a negative correlation with river water quality [34]. This means that land use and cover 1000 m away have a significant impact on river water quality. Non-point sources of pollutants from agricultural areas, vegetation areas, residential areas, water bodies, community areas, and barren land make the Muda River a suitable study area for predicting riverine loads.

The Muda River plays a vital role in distributing water sources for domestic use and irrigation, such as paddy field irrigation in catchment areas [33]. However, the water quality of the Muda River is deteriorating owing to unsustainable human activities, mainly logging, agriculture, and agro-based industries such as rubber and palm oil processing factories [32]. According to the annual Environmental Quality Report (EQR) of the Department of Environment (DOE), the Muda River is classified as a clean (C) river (categorised as Class II), as shown in Table 1 [35]. Although the Muda River is still classified as clean because its WQI values are in the range of 81–100, the WQI values showed a significant decrease from 2013 to 2014, a continuous decrease from 2015 to 2017, and a slight increase in 2018, as shown in Table S1. Water quality and river discharge data used in this study were obtained from the Department of Environment (DOE), Malaysia, and the Department of Irrigation and Drainage (DID), Malaysia. The empirical data were obtained from 2013 to 2018. Fig. 1 shows four (4) monitoring stations that cover most of the Muda River Basin area, and the stations are listed in Table S2.

2.2. Preliminary assessment of water quality data

This study used monthly water quality data from 2013 to 2018 to develop an accurate riverine load prediction model. The parameters used were the BOD (mg/L), COD (mg/L), SS (mg/L), and NH_3–N (mg/L). In developing the riverine load model, 70% of the water quality data (2013–2016) were used for training, and 30% (2017–2018) were used for testing and validation. Table 1 shows the descriptive statistics of the water quality parameters of the Muda River from 2013 to 2018 at different monitoring stations.

To characterise the water quality conditions of the Muda River, the obtained data were compared with the National Water Quality Standards (NWQS). According to the NWQS, the concentrations of BOD, COD, SS, and NH_3–N should not exceed the Class IIA/IIB concentration standards of 3, 25, 50, and 0.3 mg/L, respectively [35]. Breaches in water quality standards depend largely on river discharge. Water quality variations under low-, medium-, and high-flow river discharge conditions were studied using concentration-flow boxplots. Furthermore, the means of BOD, COD, SS, and NH_3–N were plotted using ArcGIS based on individual monitoring stations and years.

2.3. Statistical analysis

Based on the descriptive statistics of the Muda River water quality data shown in Table 1, the kurtosis and skewness were outside the range of −2 to +2; therefore, the water quality data demonstrated a non-normal distribution [36]. Hence, in this study, the water quality data were analysed using non-parametric tests. The Kruskal-Wallis ANOVA test was performed using the statistical package in the OriginPro software. The Bartlett test was conducted to assess the suitability of the water quality data before the Kruskal–Wallis ANOVA test was performed using the R-Studio software. Bartlett's test showed that the water quality data were suitable for the Kruskal Wallis ANOVA test, with a p-value >0.05, indicating that the variance was equal. Because there were four monitoring stations, the Kruskal Wallis ANOVA test was performed with $\alpha = 0.05$, and a degree of freedom (df) equal to 3. The critical chi-squared (χ^2) is shown in Fig. 2.

2.4. Geographical information system (GIS) models

The Muda River land-use data were derived from map adaptation and digitisation using Google Earth Pro (version 7.1.8) and the 2020 Cadastral USGS Projection Map. A Digital Elevation Model (DEM) was obtained from USGS Earth Explorer (SRTM Second-Arc Global Bil TIFF). 90 m) and captured on December 14, 2019 was used to recreate the Muda River polygon. The map was constructed with a single buffer zone that was 1000 m along the Muda River using both the buffer function in

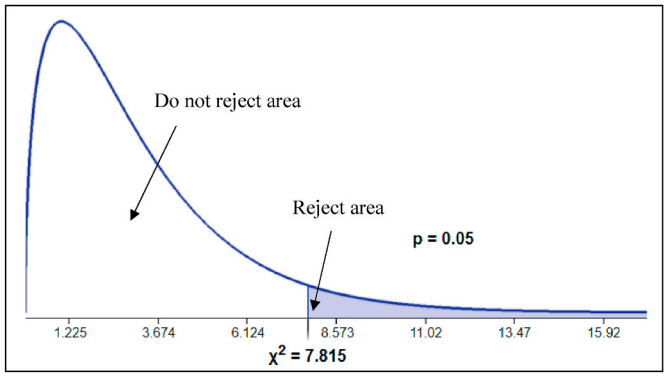


Fig. 2. Chi-squared distribution for Kruskal Wallis ANOVA test.

Table 1
Descriptive statistics for Muda River water quality parameters from 2013 to 2018.

Variables	Unit	Stations	No. of samples	Min. values	Max. values	Median values	Mean values	Kurtosis	Skewness
BOD	mg/L	2MD04	140	2.00	12.19	6.29	6.00	0.32	0.42
		2MD06		2.00	11.58	6.26	7.00	−0.76	−0.14
		2MD01		3.00	19.00	6.85	6.00	6.77	2.20
		2MD15		3.00	11.39	6.56	6.76	−0.58	0.19
COD	mg/L	2MD04	172	5.64	28.00	16.29	16.38	−0.77	0.08
		2MD06		5.09	31.00	15.93	16.00	0.06	0.26
		2MD01		5.55	44.00	18.23	16.89	1.45	0.97
		2MD15		9.99	29.52	19.88	19.19	−0.27	−0.01
SS	mg/L	2MD04	132	12.00	599.00	140.30	156.93	3.84	1.82
		2MD06		14.24	471.00	122.00	166.44	0.21	1.08
		2MD01		14.69	336.00	91.17	116.67	0.24	1.07
		2MD15		11.00	477.00	82.00	112.88	4.03	1.92
NH_3–N	mg/L	2MD04	168	0.02	1.8300	0.45	0.54	−0.28	0.74
		2MD06		0.0100	1.8300	0.1500	0.5014	−0.42	0.81
		2MD01		0.0100	7.6500	0.9000	0.9100	25.47	4.51
		2MD15		0.0100	1.9400	0.6250	0.5752	−0.02	0.69

ArcGIS 10.4.3 and ArcMap.

2.4.1. Inverse Distance Weighting (IDW)

Inverse Distance Weighting (IDW) is a spatial process used in GIS. The IDW is an interpolation model in which statistical values are used to run the model. The algorithm used in the measurement consisted of several setup points, locations (spatial), and probability points, which were later used to test the function of and experiment with the model when a null location was found on a spatial map. The weight is typically determined using the inverse square distance and its functions as shown in Eq. (1), where z_p is the elevation at each point and d^p is the distance at each point [37].

$$z_p = \frac{\sum_{i=1}^{n}\left(\frac{z_i}{d_i^p}\right)}{\sum_{i=1}^{n}\left(\frac{1}{d_i^p}\right)} \quad (1)$$

Interpolation using IDW makes the explicit assumption that items that are close together are more similar than are those that are further apart. IDW forecasts a value for any unmeasured location based on the measured values surrounding the predicted location. The measured values closer to the predicted location had a greater impact on the anticipated value than those farther away. Subsequently, the interpolation analysis results can be visualised in either a stretched or classified layout. However, in this study, the visualisation focused on the classified version of symbology because it was the easiest way to understand the map [38].

2.5. Riverine load equation

The riverine load was determined using a simple mass balance equation by dividing the Muda River system into four segments based on monitoring stations. The assumption behind this method is that the pollutant concentration (C) is uniform with respect to the different Muda River segments. Thus, the concentration of pollutants is related to riverine load (W) and river discharge (Q), as shown in Eq. (2) [39,40]:

$$C = (1/Q) \times W; \quad W = C \times Q \quad (2)$$

In this study, the total riverine load of the Muda River Basin was represented by the summation of the riverine loads (Eq. (3)) from four monitoring stations (Table S2).

$$\sum W = W_{2MD04} + W_{2MD06} + W_{2MD01} + W_{2MD15} \quad (3)$$

Based on the ArcGIS mapping of the monitoring stations in the Muda River Basin shown in Fig. 1, 2MD04 station was near agricultural, vegetation, residential, community, and industrial areas 1000 m away. Meanwhile, 2MD06 station was near agricultural, residential, and community areas 1000 m away. Moreover, 2MD01 station was near agricultural and vegetation areas 1000 m away. Finally, 2MD15 station was near agricultural areas, residential areas, and community areas 1000 m away.

2.6. Development of ANN models for riverine loading pollutant models

The artificial neural network model was constructed using MATLAB R2015a to determine the riverine load. The input and target data were imported from EXCEL Microsoft 365 and saved horizontally in MATLAB workspace. The input data are denoted as 'x' and the target data are denoted as 't'. The network type, training function, adaption learning function, fitting network, division of data, plot function, network training, and testing of the feedforward backpropagation neural network (FFBP NN) were written as algorithms in the MATLAB code in the editor tab. The model was then run to optimise the number of neurones. For the radial basis neural network (RB NN), the training function, spread constant, and optimised hidden layer nodes from the FFBP NN were written as algorithms in the MATLAB code in the Editor tab. Subsequently, the FFBP NN and RB NN models were run ten times. The results were averaged, saved in a MATLAB file, and analysed.

2.7. Feed-forward backpropagation neural network

In 1986, Rumelhart and McClelland introduced the concept of backpropagation neural networks within the framework of artificial neural network (ANN) algorithms, employing nonlinear neuron processing functions. This theoretical architectural design comprises a multilayer feedforward network trained using an error backpropagation algorithm, which is notable for its inherent simplicity, adaptability, and practicality. Such attributes render it particularly well suited for investigating nonlinear phenomena such as surface subsidence resulting from coal mining activities [41]. This artificial intelligence model has also found widespread application in various civil engineering domains [42]. Lee et al. synergistically combined this artificial neural network model with a geographic information system to evaluate and forecast land subsidence patterns within an abandoned coal mining site in South Korea by leveraging existing land subsidence data. Their empirical investigation substantiated the capability of the ANN to accurately predict evolving subsidence trends, which closely aligned with observed real-world conditions [43].

In this study, a feedforward backpropagation neural network (FFBP NN) was employed to develop a riverine load model for BOD, COD, SS, and NH_3–N using five input variables and one output variable, as shown in Fig. 3. The FFBP NN model architecture for riverine load prediction was developed on the basis of the broad conceptual ANN network outlined by Basant et al. [44]. Fig. 3 shows the FFBP NN model architecture for riverine load development, which consists of an input layer, hidden layer, and output layer. The former indicates the river discharge (Q) and the concentration of a pollutant at different monitoring stations (C_x). This is connected to the hidden layer by a specific weight, denoted by w_{ij}, and a tansig transfer function. The hidden layer was connected to the output layer (targeted riverine load) with a specific weight, denoted as w_{jk}, and tansig transfer function. The feedforward neural network was equipped with a backpropagation training function. A feed-forward neural network with a backpropagation algorithm was used to predict the riverine load because it has been extensively used to predict river water quality [45–47]. Fig. 4 illustrates the mechanism of the feedforward backpropagation neural network. The theory behind the feedforward backpropagation algorithm is that the computed outputs from the function of the inputs, weight, and bias are introduced. The error obtained from the subtraction of the output and target was backpropagated, and the weight and bias were adjusted until a specified error tolerance or epoch number was achieved [27,48].

Furthermore, the Bayesian regularisation training function was used with the backpropagation training algorithm to obtain a small error, despite the longer time required to converge than when using Levenberg Marquardt. Subsequently, to accelerate the feed-forward backpropagation neural network convergence with the Bayesian regularisation training function, each layer was connected using the tansig transfer function [49]. In this study, the epoch number and the number of neurones were selected by trial and error. This approach is essential because too few neurones would result in underfitting, whereas too many neurones would result in overfitting [48]. To obtain the optimum number of hidden layer nodes, the numbers of neurones tested were 1, 2, 3, 4, and 5 for the NH_3–N and SS models and 1, 2, 3, 4, 5, and 6 for the COD and BOD models, as suggested by Heaton [50]. As a rule of thumb, the number of neurones in the hidden layer should not exceed twice the number of inputs.

2.8. Radial basis neural network

The radial basis neural network was chosen for comparison with the feed-forward backpropagation because it consists of three layers (input,

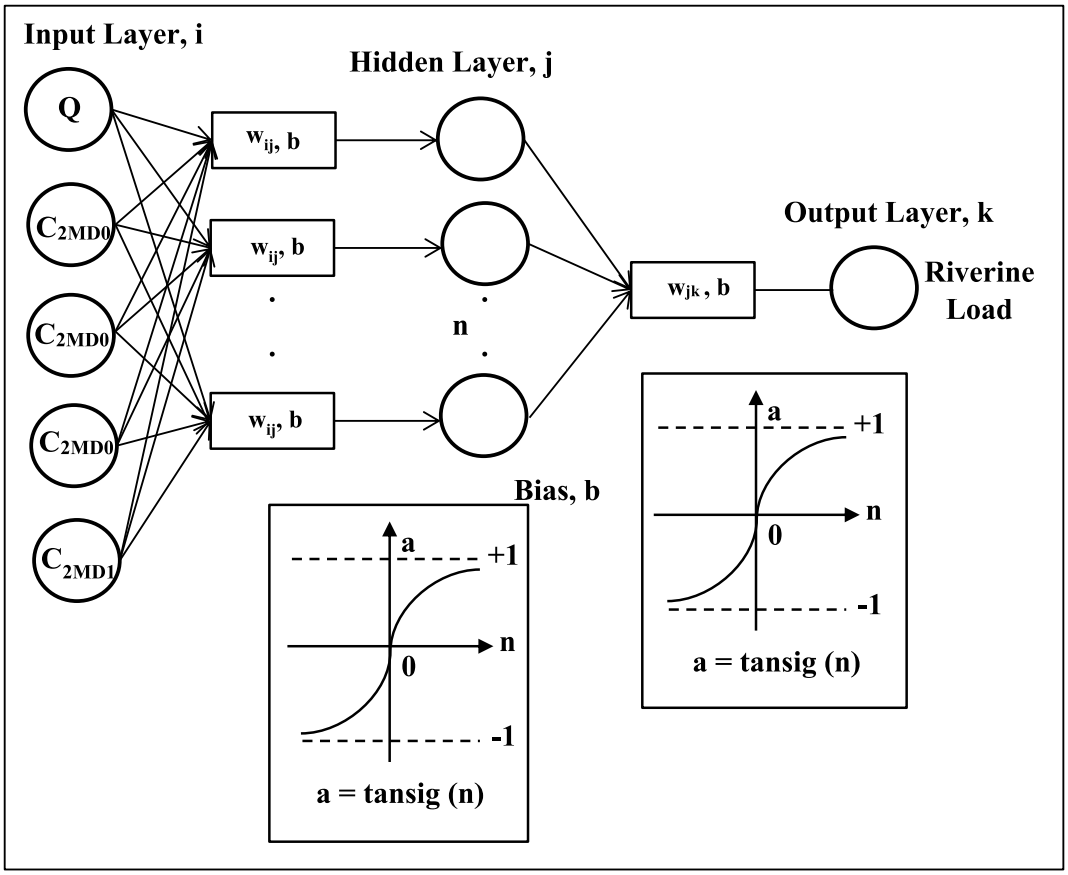


Fig. 3. FFBP NN model architecture for riverine load development.

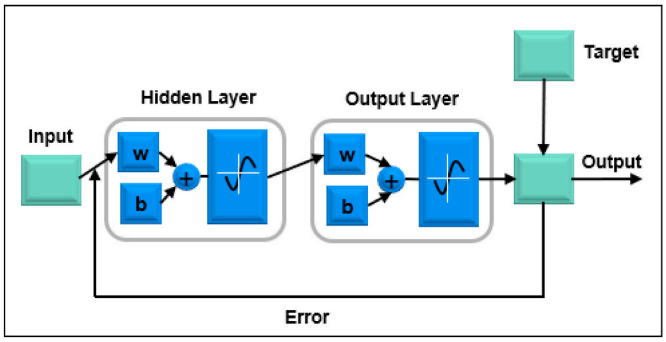


Fig. 4. Mechanism of feed-forward backpropagation neural networks.

hidden, and output). Using an RB NN, the input layer is connected to the hidden layer using a nonlinear Gaussian function. The hidden layer is then connected to the output layer using the purelin transfer function, as shown in Fig. 5 [51]. The RB NN model architecture for riverine load prediction was developed based on the broad conceptual ANN developed by Basant et al. [26]. The nonlinear Gaussian function equation is given by Eq. (4), where μ is the centre of the Gaussian function (mean value of x) and d is the distance (radius) from the centre, providing a measure of the spread of the Gaussian curve [11]. The spread constant d is used in this study as the maximum spread constant for the radial basis function [52]. The hidden layer nodes used for the RB NN are similar to the optimum hidden layer nodes obtained from the trial-and-error approach using the FFBP NN.

$$\varphi(x,\mu) = e^{\frac{x-\mu^2}{2d^2}} \tag{4}$$

2.9. Multiple linear regression

In this study, multiple linear regression (MLR) was used as a reference model for the developed nonlinear neural network models: FFBP NN and RB NN. The linear regression equations used to link Q, C_{2MD06}, C_{2MD04}, C_{2MD01}, and C_{2MD15} are shown in Eq. (5).

$$\text{Riverine Load} = \alpha_0 + \alpha_1(Q) + \alpha_2(C_{2MD06}) + \alpha_3(C_{2MD04}) + \alpha_4(C_{2MD01}) + \alpha_5(C_{2MD15}) \tag{5}$$

where,

α_0 is the intercept;

$\alpha_1, \alpha_2, \alpha_3, \alpha_4$, and α_5 are the coefficients of each input;

Q is the river discharge.where C_X is the concentration of pollutants at different monitoring stations.

2.10. Performance determination parameters

The performance of the developed data-driven models was determined using four statistical equations to quantify the error between the observed (O) and predicted (P) riverine load values, and to quantify the correlation between the observed and predicted riverine load values. The statistical equations used for error quantification were root mean square error (RMSE), mean absolute error (MAE), and mean relative error (MRE). The correlation between predicted and observed values was quantified using the correlation of determination (R^2). Eq. (6) to Eq. (9) provided the error and correlation metrics [8].

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i\text{-}P_i)^2} \tag{6}$$

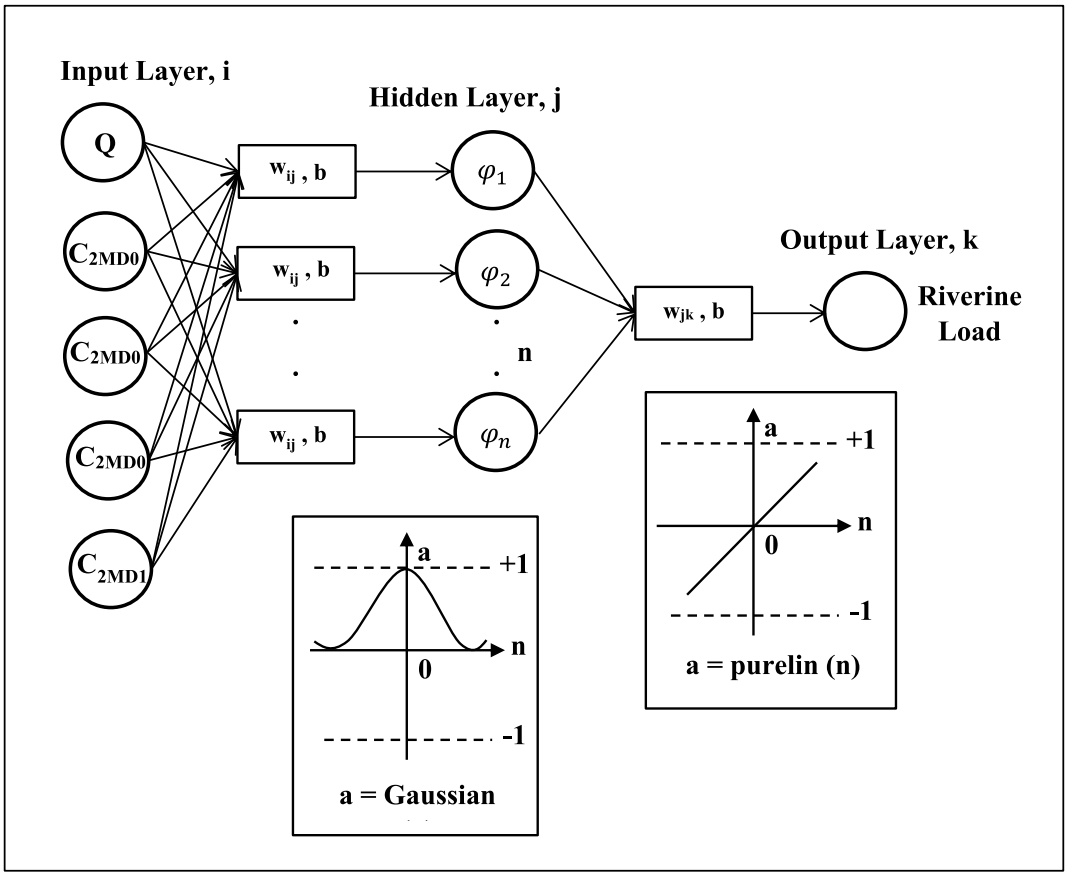


Fig. 5. Rb NN model architecture for riverine load development.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left[O_i\text{-}P_i\right] \quad (7)$$

$$MRE = \frac{1}{n}\sum_{i=1}^{n}\frac{\left[O_i\text{-}P_i\right]}{O_i} \quad (8)$$

$$R^2 = \frac{\sum_{i=1}^{n}\left(O_i\text{-}O_{avg}\right)\left(P_i\text{-}P_{avg}\right)}{\sum_{i=1}^{n}\left(O_i\text{-}O_{avg}\right)^2\sum_{i=1}^{n}\left(P_i\text{-}P_{avg}\right)^2} \quad (9)$$

3. Results and discussion

3.1. Muda River receiving water quality characterization

The Muda River water quality data for the riverine load study indicated violations of Class II of NWQS. Concentration-flow boxplots (Figs. 6–9) were used to document the changes in the concentrations of BOD, COD, SS, and NH_3–N with respect to river discharge. The observed BOD in the Muda River ranged from 3 to 11.39 mg/L. Overall, 94% of the 35 samples had BOD values exceeding Class II of NWQS: 3.0 mg/L. This clearly shows that the Muda River was depleted of dissolved oxygen, signifying low water quality. As shown in Fig. 6, the mean value (7.2 mg/L) at low flow was higher than that at mid-flow (5.2 mg/L) and high-flow (3.7 mg/L).

As shown in Fig. 7, the concentration observed for the chemical oxygen demand (COD) ranged from 9.99 to 29.52 mg/L. Of the 43 samples, 16.3% had COD values exceeding the NWQS for Class II (25 mg/L). The mean value (20.37 mg/L) at low flow was slightly higher than that (20.15 mg/L) at mid-flow, whereas the mean value (11 mg/L) at high flow was the lowest.

The observed suspended solid (SS) concentrations ranged from 11 to

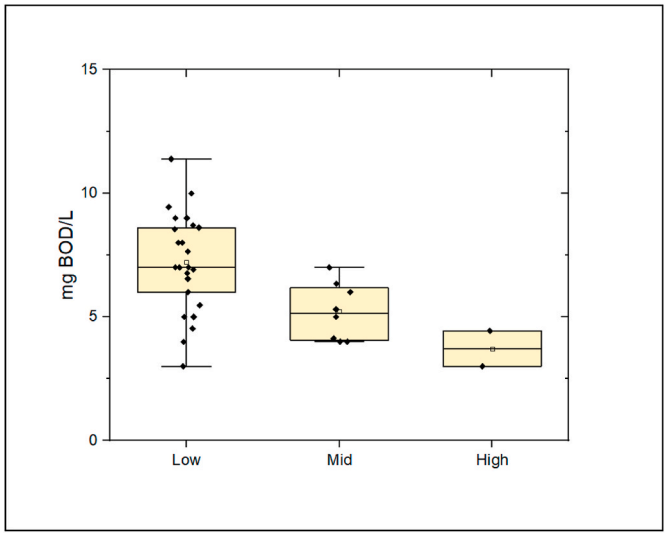


Fig. 6. Comparison of BOD under low-, medium-, and high-flow conditions.

477 mg/L. Of the 33 samples, 70% had SS values exceeding the NWQS for Class II (50 mg/L). This suggests that the Muda River has faced enormous surface runoff and soil erosion from nearby forests and cultivated land. The maximum (477 mg/L), median (100.83 mg/L), and mean (128.74 mg/L) suspended solid values at low flow were higher than the corresponding mid-flow (271, 82, and 95.42 mg/L) and high flow (63, 50.26, and 50.26 mg/L) values, respectively, as shown in Fig. 8.

The ammoniacal nitrogen (NH3–N) concentrations ranged from

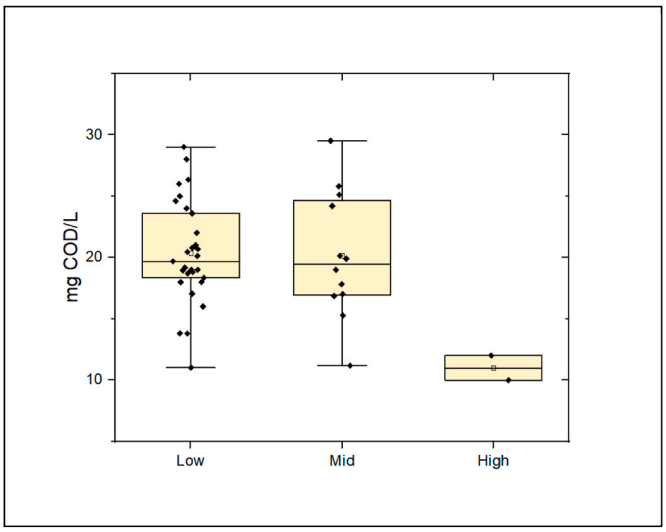


Fig. 7. Comparison of COD under low-, medium-, and high-flow conditions.

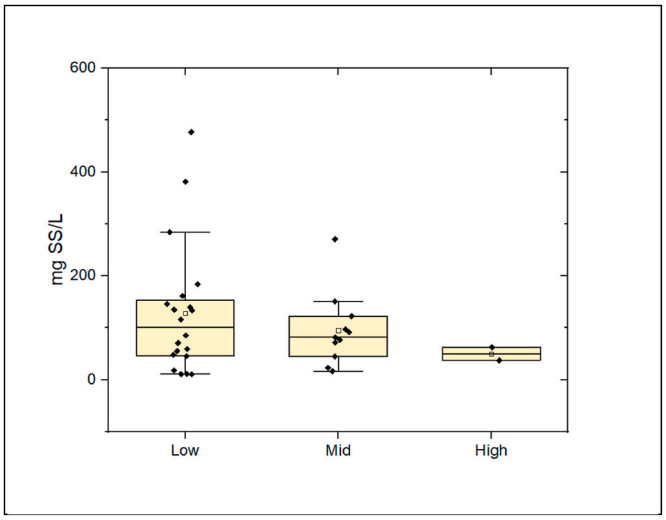


Fig. 8. Comparison of SS under low-, medium-, and high-flow conditions.

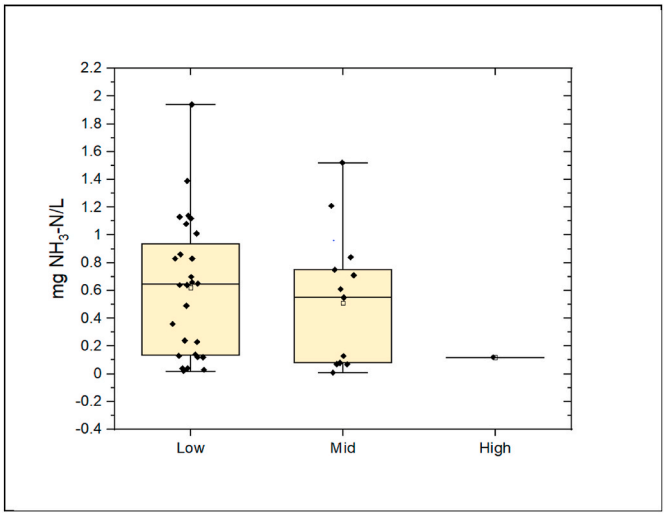


Fig. 9. Comparison of NH_3–N under low-, medium-, and high-flow conditions.

0.01 mg/L to 1.94 mg/L. Of the 42 samples, 62% had NH_3–N values exceeding Class II of the NWQS: 0.3 mg/L. River quality deterioration might be due to nutrient runoff from paddy fields, palm oil plantations, and rubber plantations along the Muda River. As shown in Fig. 9, the mean value (0.62 mg/L) at low flow was higher than the mean values obtained in mid-flow(0.51 mg/L) and high flow (0.12 mg/L) conditions.

As mentioned above, the water quality constituents of the Muda River were found to violate Class II of the NWQS, because the highest pollutant concentration occurred when the streamflow was low. The water quality standard violations occurred between 2013 and 2018. The concentration-flow boxplot (Figs. 6–9) indicates that low streamflow is a critical condition that determines whether violations occur. These findings was generally consistent with previous work [32,33].

Table 2 shows the coloured indicators used on the Muda River GIS map based on the NWQS classification. Figs. 10–13 illustrate GIS maps representing the water quality parameters (BOD, COD, SS, and NH_3–N) of the Muda River at various monitoring stations (2MD04-2MD06-2MD01-2MD15) between 2013 and 2018. Spatial distribution maps (Figs. 10–13) were developed using the monitoring station latitudes and longitudes before the integration of the average water quality data from the Muda River at each monitoring station. Based on the coloured indication of the water quality parameters shown in Table 2 and Figs. 10, 12 and 13, BOD, SS, and NH_3–N were categorised as classes III and IV, indicating that the water supply required extensive treatment and was unsuitable for the long-term conservation of the natural environment and agricultural irrigation. However, according to Fig. 11, the COD value for the Muda River remained Class II because it did not exceed Class II of the NWQS for COD values, which was 25 mg/L. According to the Kruskal Wallis ANOVA test results shown in Table 3, the water quality parameters obtained from the monitoring stations did not differ significantly each year, based on the ($\chi^2 < 7.815$ and $p > 0.05$) at $\alpha = 0.05$ and $df = 3$. This result strongly corresponds to the land-use activities of the Muda River, as shown in Fig. 1 and previous work [53], with agriculture being the most significant source of pollutants along the Muda River 1000 m away.

The Biochemical Oxygen Demand (BOD) is a crucial metric for assessing the extent of organic pollution in aquatic ecosystems. Elevated BOD values signify a heightened presence of organic constituents, often stemming from non-point sources, such as agricultural runoff. As depicted in Fig. 10, the average BOD levels in 2013 ranged from 5.088 to 6.932 mg/L. Specifically, stations 2MD01, 2MD04, 2MD06, and 2MD15 recorded the average BOD values of 6.9325, 5.0875, 6.6025, and 5.775 mg/L, respectively. Statistical analysis using the Kruskal–Wallis ANOVA test revealed no significant variance in the populations among the monitoring stations ($p > 0.05$, $p = 0.5580$, $df = 3$, $\chi 2 = 2.0700$), as presented in Table 4. Similarly, from 2014 to 2018, a lack of substantial disparity in BOD values was observed across different monitoring stations along the Muda River ($p > 0.05$ and $\chi 2 < 7.815$). Notably, in 2014 and 2016, the downstream station (2MD01) of the Muda River exhibited notably high average BOD values of 9.1911 mg/L and 8.7425 mg/L, respectively, surpassing the NWQS Class II BOD limit of 3 mg/L by approximately threefold.

The Chemical Oxygen Demand (COD) is a critical parameter for assessing water body contamination, encompassing both organic and

Table 2
Color indicators for BOD, COD, SS, and NH_3–N values in the GIS map.

Parameter	Value	Unit	Class	Color
BOD	$3 < x \leq 6$	mg/L	III	Yellow – Orange
	$6 < x < 12$	mg/L	IV	Orange – Red
COD	$10 < x \leq 20$	mg/L	II	Dark green – Light green
	$20 < x < 25$	mg/L	II	Orange – Red
SS	$25 < x \leq 50$	mg/L	II	Yellow
	$50 < x \leq 150$	mg/L	III	Orange
	$150 < x \leq 300$	mg/L	IV	Red
NH_3–N	$0.3 < x \leq 0.9$	mg/L	III	Yellow - Red

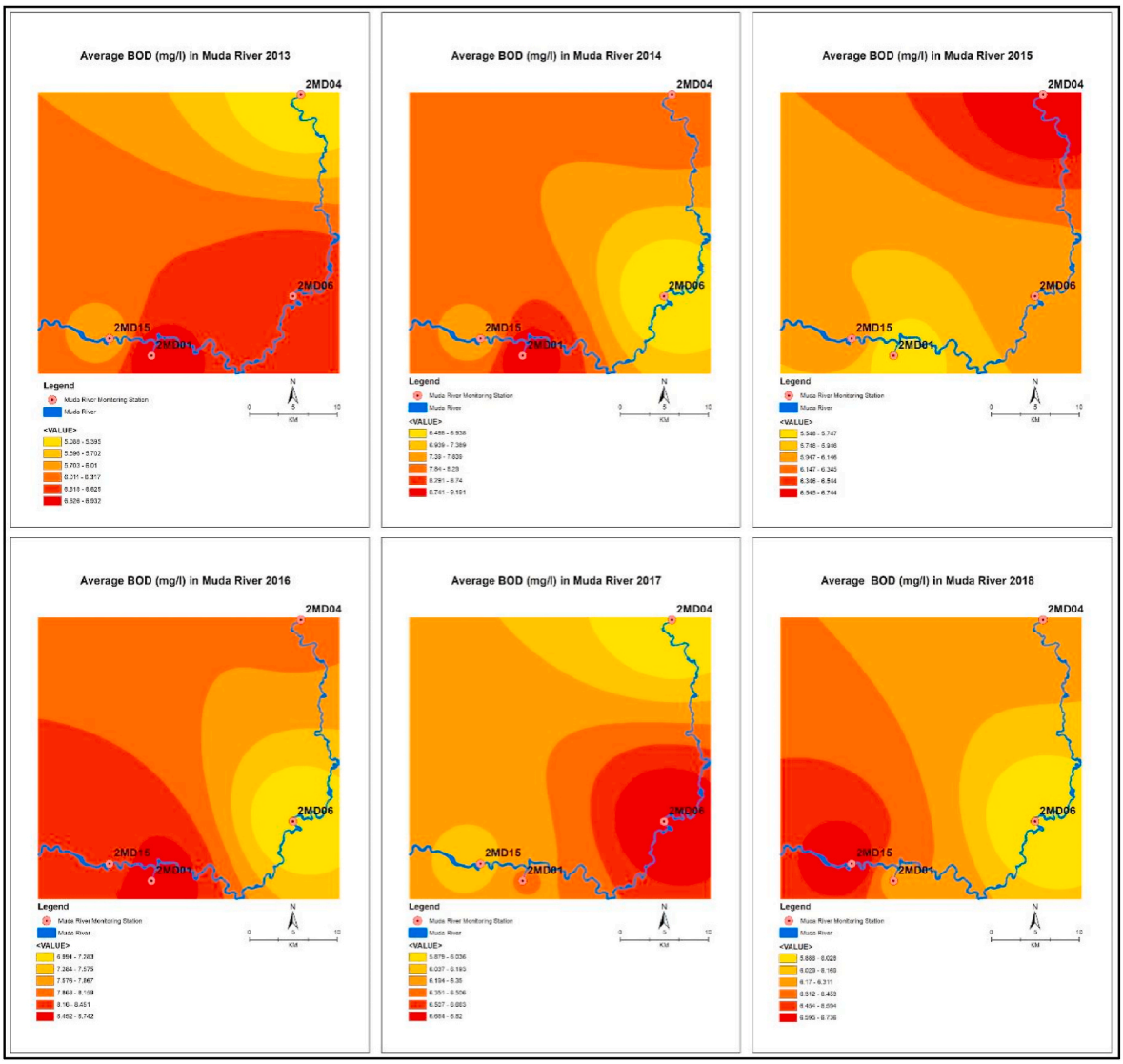


Fig. 10. Average BOD (mg/L) in Muda river at various monitoring stations from 2013 to 2018.

inorganic pollutants. As depicted in Fig. 11, the downstream station (2MD01) of the Muda River recorded the highest average COD values in 2014 and 2016, at 21.4 mg/L and 25.51 mg/L, respectively. In 2014, the average COD value remained within the compliance limit of the NWQS Class II COD limit of 25 mg/L, whereas in 2016 it exceeded this limit by approximately 2.04%. Intriguingly, the same monitoring station reported the highest average COD concentrations in 2014 and 2016. Fig. 11 illustrates that the average COD values in 2013 exhibited a range from 12.64 mg/L to 17.04 mg/L, with stations 2MD01, 2MD04, 2MD06, and 2MD15 recording average COD values of 17.0388 mg/L, 12.6425 mg/L, 15.9313 mg/L, and 13.5888 mg/L, respectively. Statistical analysis using the Kruskal–Wallis ANOVA test revealed no significant variation in populations among the monitoring stations ($p > 0.05$, $p = 0.6664$, $df = 3$, $\chi 2 = 1.5693$). Similarly, from 2014 to 2018, no significant disparities in average COD values were observed among the various monitoring stations along the Muda River ($p > 0.05$ and $\chi 2 < 7.815$).

Suspended Solids (SS) represent a pivotal parameter for evaluating the water quality of the Muda River, particularly given its susceptibility to annual flooding during the rainy seasons of April–May and September–November [54]. As shown in Fig. 12, the average SS values exhibited an annual increase between 2013 and 2017, with ranges of 34.67 mg/L – 84.67 mg/L, 62.68 mg/L – 122.7 mg/L, 69.63 mg/L – 151.1 mg/L, 40.12 mg/L – 162.6 mg/L, and 147 mg/L 254 mg/L, respectively. However, a significant reduction in the average SS value from 147 mg/L to 254 mg/L in 2017 to 90.95 mg/L – 166.7 mg/L in 2018 was observed. This decline was attributed to a comprehensive flood mitigation project undertaken in 2017 by the Department of Irrigation and Drainage of the Ministry of Environment and Water in Malaysia [33]. The project included measures such as river deepening, widening, improved river mouth management, construction of a new barrage, drainage system enhancements, flood control gate installation, riverbank erosion control, and relocation and adjustment of public amenities [33]. Statistical analysis, as detailed in Table 4, demonstrated negligible variations in average SS values across monitoring stations ($p > 0.05$ and $\chi 2 < 7.815$).

Agricultural regions were predominantly concentrated in proximity to the Muda River, as depicted in Fig. 1, and were anticipated to exert a notable influence on average NH_3–N concentrations. The analysis presented in Table 2 and Fig. 13 reveals that the mean NH_3–N concentration consistently surpassed the threshold set by NWQS Class II (0.3 mg/L), as indicated by the gradient of colours from yellow to red. Notably, the NH_3–N classification remained Class III across the monitoring points along the Muda River from 2013 to 2018. This observation is robustly supported by statistical outcomes obtained through the Kruskal–Wallis ANOVA test, as detailed in Table 3, yielding p-values of ($p = 0.8469$, $p =$

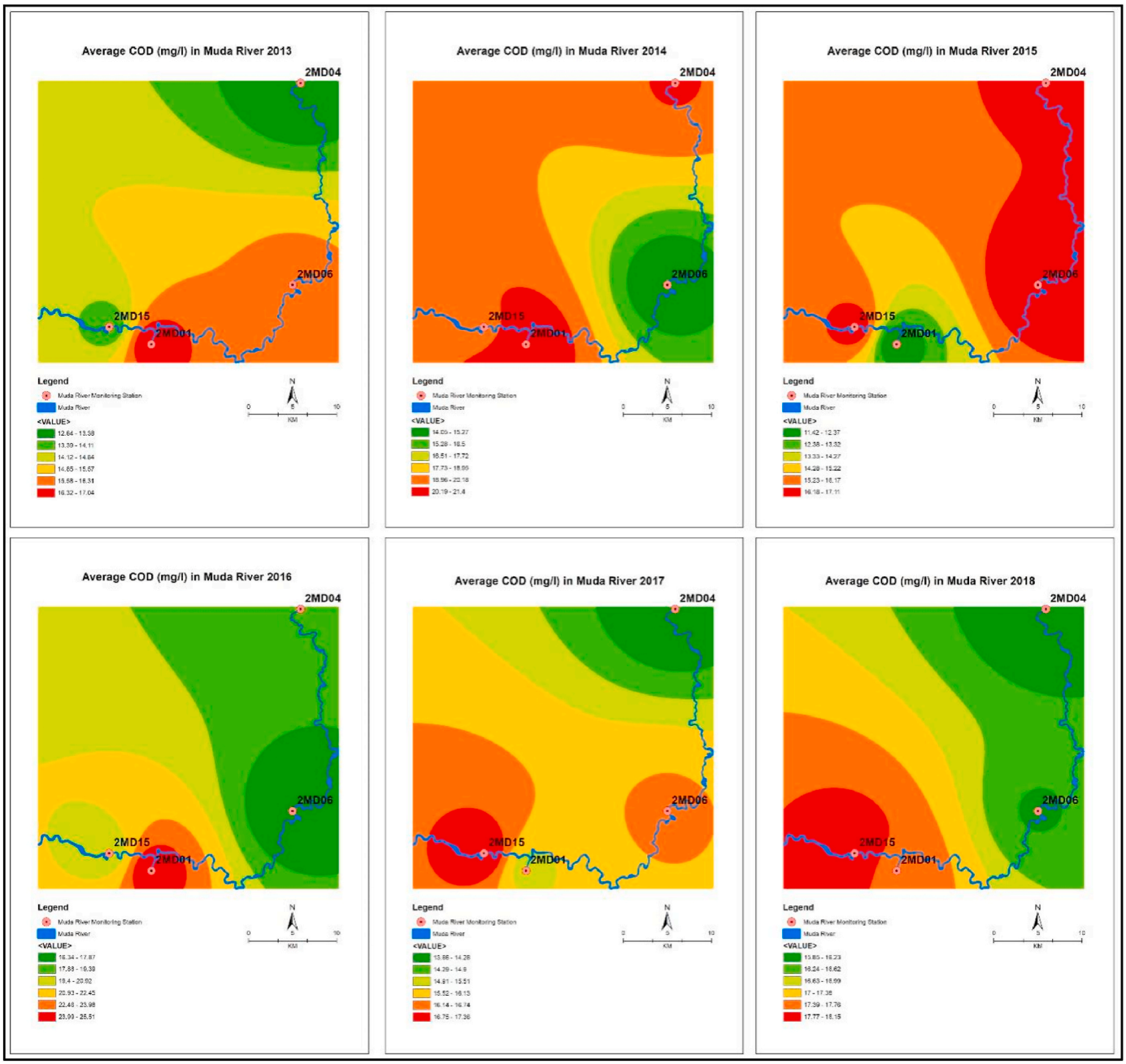


Fig. 11. Average COD (mg/L) in Muda river at various monitoring stations from 2013 to 2018.

0.8065, p = 0.4196, p = 0.8886, p = 0.4897, and p = 0.7204) for the years 2013, 2014, 2015, 2016, 2017, and 2018, respectively. This statistical analysis unequivocally underscores the absence of statistically significant disparities in average NH_3–N concentrations among distinct monitoring stations encompassing upstream (2MD04), midstream (2MD06 and 2MD01), and downstream (2MD15) locations. Shamsuddin et al. recently reported similar findings [55].

3.2. Optimal hidden layer neuron number determination

The optimal number of hidden layer nodes for the Feed-Forward Backpropagation Neural Network (FFBP NN) model was determined using a meticulous trial-and-error process. It is well known that an insufficient number of neurones in the hidden layer can result in underfitting, whereas an excessive number can lead to overfitting [56]. The performance of each neural network model was evaluated using key metrics: the Mean Absolute Error (MAE), Mean Square Error (MSE), Sum Square Error (SSE), and Coefficient of Correlation (R). These evaluations were performed for NH_3–N, COD, BOD, and SS, as shown in Figs. 14–17. The optimal configuration for the NH_3–N and SS models involved four hidden layer nodes, which exhibited the lowest error and the highest correlation with the target, as shown in Figs. 14 and 17. Conversely, the optimum hidden layer configurations for the COD and BOD models were four and five, respectively, as illustrated in Figs. 15 and 16.

In a distinct approach, Nhantumbo et al. (2018) employed the Lippman Rule of Thumb to determine the optimal number of hidden nodes as four, effectively handling three input variables, one output variable, and major ions [57]. Similarly, Heydari et al. achieved an optimal configuration of five hidden nodes when dealing with three independent input variables and one output variable, dissolved oxygen (DO) [58]. Although these studies employed feedforward back-propagation neural networks, they utilised a different activation

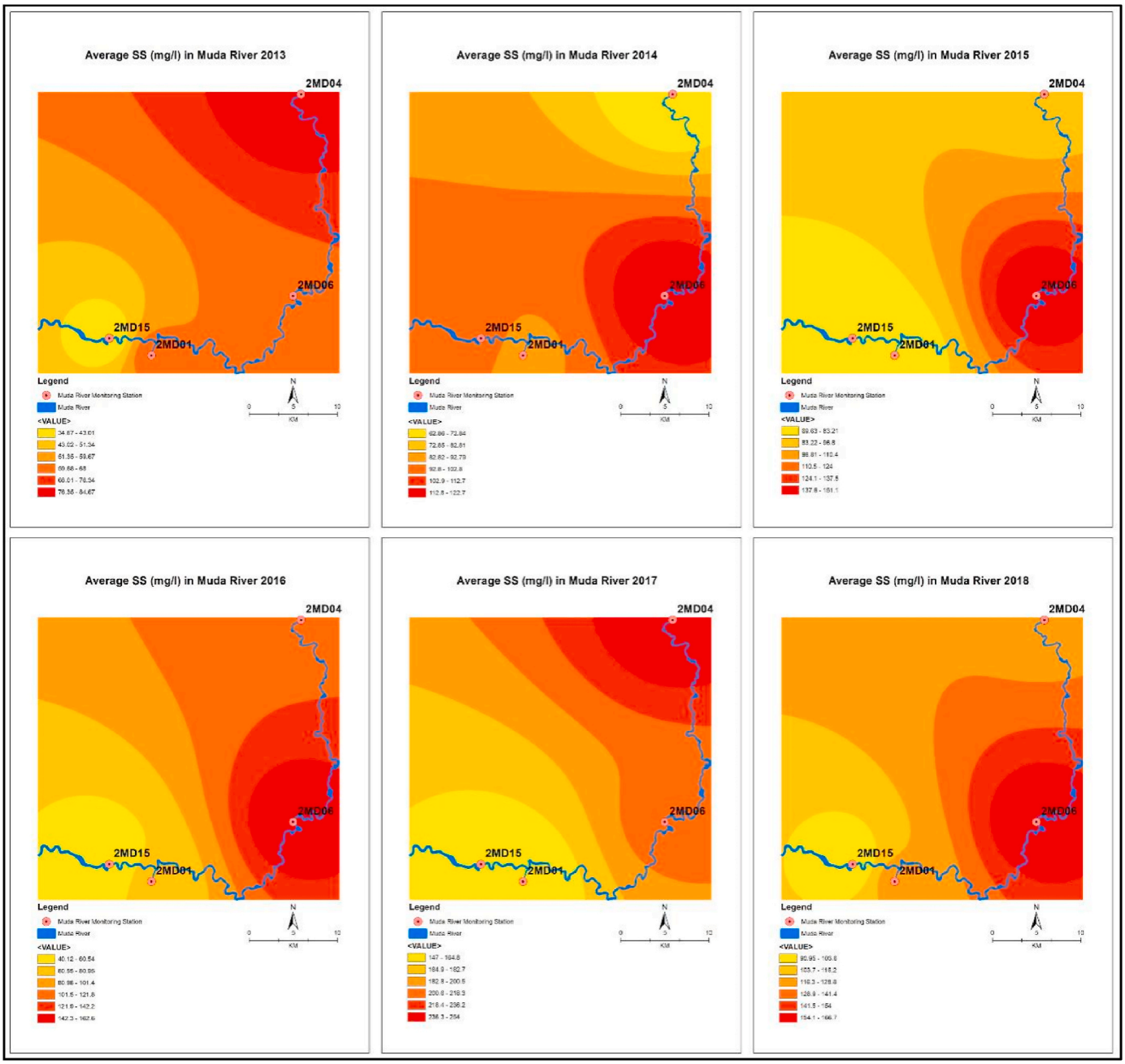


Fig. 12. Average SS (mg/L) in Muda river at various monitoring stations from 2013 to 2018.

function: logsig. The performance metrics displayed in Figs. 14, 15 and 17 indicate a linear decrease in MAE, MSE, and SSE with increasing number of hidden nodes until four nodes were reached, beyond which an increase was observed at five or more hidden nodes. Conversely, the Coefficient of Correlation (R) demonstrated a linear increase with the increase in the number of hidden nodes up to four, after which a decrease was observed beyond five nodes. Accordingly, the optimal number of hidden nodes for NH_3–N, SS, and COD was four, reflecting the lowest error and the highest correlation with the target variable. In the case of the BOD model, the optimal configuration involved five hidden nodes aligned with patterns of minimised error and maximal correlation with the target variable, as shown in Fig. 16. To ensure a consistent comparison, the optimal hidden layer node configurations (4,4,4, and 5) determined through the trial-and-error approach with the FFBP NN were subsequently employed in the development of Radial Basis Neural Network (RB NN) models for NH_3–N, SS, COD, and BOD.

3.3. Model calibration and validation

Riverine load models for Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Suspended Solids (SS), and Ammoniacal Nitrogen (NH3–N) were subjected to rigorous calibration and validation processes involving partitioning of pollutant concentration and river discharge data into distinct subsets. Specifically, 70% of the data was allocated to the training phase, whereas 15% was reserved for the validation and testing phases. The training dataset encompassed data spanning from 2013 to 2016, with 2017 data serving validation purposes and 2018 data designated for testing. To assess the performance of the model, the coefficient of determination (R^2) values were meticulously computed for each iteration. This iterative process was repeated ten times to ensure robust calibration and validation results, which is a necessary step given the random selection of weights and biases in each run [59,37]. Examination of Tables 4–7 reveals the

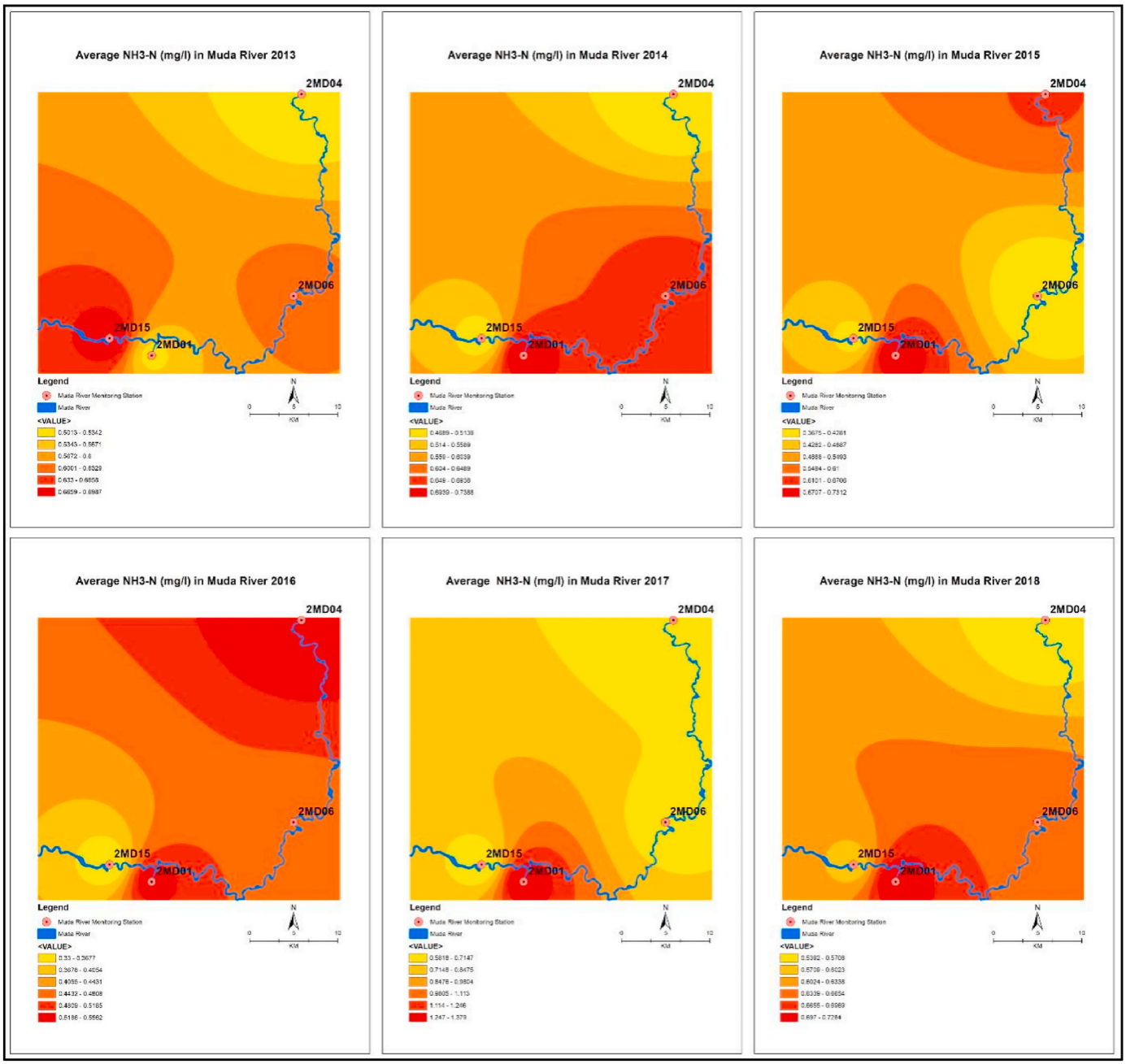


Fig. 13. Average NH_3–N (mg/L) in Muda river at various monitoring stations from 2013 to 2018.

effectiveness of this calibration and validation procedure, as evidenced by the closely aligned average R^2 values for the training, validation, and testing datasets, underscoring the model's robust performance.

3.4. Comparative analysis of riverine load prediction

The primary objective of this investigation was to assess the efficacy of data-driven techniques in predicting riverine loads, with specific emphasis on their inherent simplicity and reduced data requirements. The predictive model chosen for this study was an artificial neural network (ANN), a well-established approach renowned for its capability to model the intricate nonlinear relationships prevalent in water quality applications. Additionally, a multiple linear regression (MLR) model was employed as a reference point to gauge the predictive accuracy of the developed ANN model. A comprehensive comparison of these models is presented in Table 8. As delineated in Table 8, the feed-forward neural network (FFBP NN) model exhibited superior predictive performance across all evaluation criteria when forecasting riverine loads of Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Suspended Solids (SS), and Ammoniacal Nitrogen (NH_3–N). Compared to multiple linear regression, the FFBP NN model yielded notably enhanced predictions. Multiple linear regression yielded considerably larger Root Mean Square Error (RMSE) values when applied to riverine load predictions of BOD, COD, SS, and NH3–N. The FFBP NN model successfully reduced the RMSE by 89.5%, 80.8%, 87.9%, and 79.0% for the BOD, COD, SS, and NH3–N, respectively. Conversely, the radial basis neural network (RB NN) displayed the highest RMSE values among the evaluated models for riverine load prediction parameters, indicating its suboptimal performance in this context. This disparity in performance suggests that the RB NN model exhibits limited sensitivity in capturing the intricate relationships between the input and output variables, particularly when employing

Table 3
Kruskal-Wallis ANOVA of river water quality parameters (BOD, COD, SS, and NH_3–N) at Various Monitoring Stations (2MD04-2MD06-2MD01-2MD15) from 2013 to 2018.

Parameter	2013			2014			2015			2016			2017			2018		
	Chi-Squared	df	p-value	Chi-Squared	df	p-value	Chi-Squared	df	p-value	Chi-Squared	df	p-value	Chi-Squared	df	p-value	Chi-Squared	df	p-value
BOD	2.0700	3	0.5580	1.4918	3	0.6842	1.4749	3	0.6881	2.8820	3	0.4102	1.0902	3	0.7795	1.0988	3	0.7774
COD	1.5693	3	0.6664	5.4048	3	0.1442	4.8922	3	0.1799	6.1711	3	0.1036	2.1202	3	0.5478	1.9888	3	0.5747
SS	6.6409	3	0.0843	2.3874	3	0.4960	5.6864	3	0.1279	5.8143	3	0.1210	4.6419	3	0.1999	3.5748	3	0.3112
NH_3–N	0.8106	3	0.8469	0.9782	3	0.8065	2.8240	3	0.4196	0.6342	3	0.8886	2.4213	3	0.4897	1.3367	3	0.7204

fewer hidden nodes, mirroring the configuration determined for the FFBP NN model [60].

Furthermore, as depicted in Fig. 18(a and b,c,d), a noteworthy concurrence is evident between the observed and predicted riverine load values for Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Suspended Solids (SS), and Ammoniacal Nitrogen (NH_3–N) when employing a feedforward backpropagation neural network (FFBP NN). This alignment signifies a close adherence to the variation patterns exhibited by the observed and predicted riverine load values across these critical water quality parameters. This observation aligns with previous findings by Banejad and Olyaie and He et al., who similarly reported a robust correspondence between observed and predicted variables when utilising the FFBP NN [48,61]. Basant et al. reported minimal disparities between observed and predicted Dissolved Oxygen (DO) values using a feed-forward neural network model characterised by a backpropagation network type, employing distinct transfer functions for the hidden and output layers, specifically Tansig and Purelin [44].

Additionally, it is noteworthy that a substantial portion of the predicted values converged with the observed values when employing Multiple Linear Regression (MLR). However, a limited number of predicted values aligned with the observed values when employing the Radial Basis Neural Network (RB NN), as shown in Fig. 18 (a-d). The suboptimal performance of MLR in prediction can be attributed to the intricate nonlinear relationships inherent in the pollutant concentration, river discharge, and riverine load dynamics. Conversely, the subpar performance of the RB NN can be attributed to the insufficiency of hidden nodes, which hinders the capacity of the model to deliver precise riverine load predictions. Notably, the calculated Coefficients of Determination (R^2) for the riverine loads of BOD, COD, SS, and NH3–N were 0.9998, 0.9990, 0.9996, and 0.9987, respectively, when the feedforward neural network model was employed, whereas the corresponding values for the Radial Basis Neural Network model were 0.8946, 0.7613, 0.8580, and 0.7973, respectively (Table 8). This discrepancy underscores the superior performance of the former compared with the latter.

In summary, these outcomes underscore the capacity of the feedforward neural network, which incorporates a backpropagation algorithm and Bayesian regularisation training methodology, to discern intricate pollutant load patterns across various monitoring stations. This capability facilitates precise prediction of riverine loads of BOD, COD, SS, and NH_3–N within the Muda River. These findings align with the assertions of Senthilkumar et al., who asserted that an artificial neural network model employing a backpropagation algorithm provides superior insights compared to alternative modelling approaches, such as decision tree models, radial basis functions, and fuzzy logic [62].

3.5. Load reduction allocation

Fig. 19 (a-d) provide a comprehensive representation of the riverine loads pertaining to Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Suspended Solids (SS), and Ammoniacal Nitrogen (NH_3–N) within the confines of the Muda River. To determine the actual riverine load, we measured the actual concentrations of BOD, COD, SS, and NH_3–N in the Muda River. The target riverine load was determined using concentration values adhering to Class IIA/IIB standards of the National Water Quality Standards (NWQS). As shown in Fig. 19 (a-d), the riverine load pattern exhibited a direct correlation with the volume of river water discharge. Augmented river water discharge signifies an increased capacity of the river to assimilate pollutants emanating from both point and nonpoint sources. However, it is disconcerting to note that riverine loads of BOD, SS, and NH3–N consistently transgress Class II standards of the NWQS. Furthermore, these parameter violations were exacerbated under conditions of elevated river discharge, as shown in Fig. 19 (a-d).

Typically, when the BOD surpasses the Class II criteria of the NWQS,

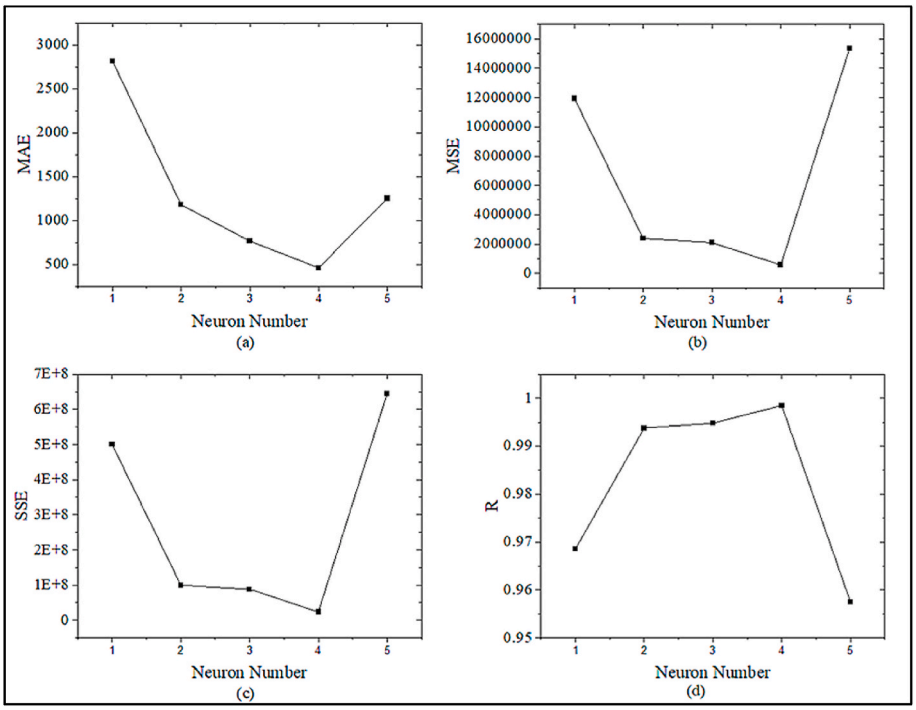


Fig. 14. NH_3–N model performance at various hidden nodes.

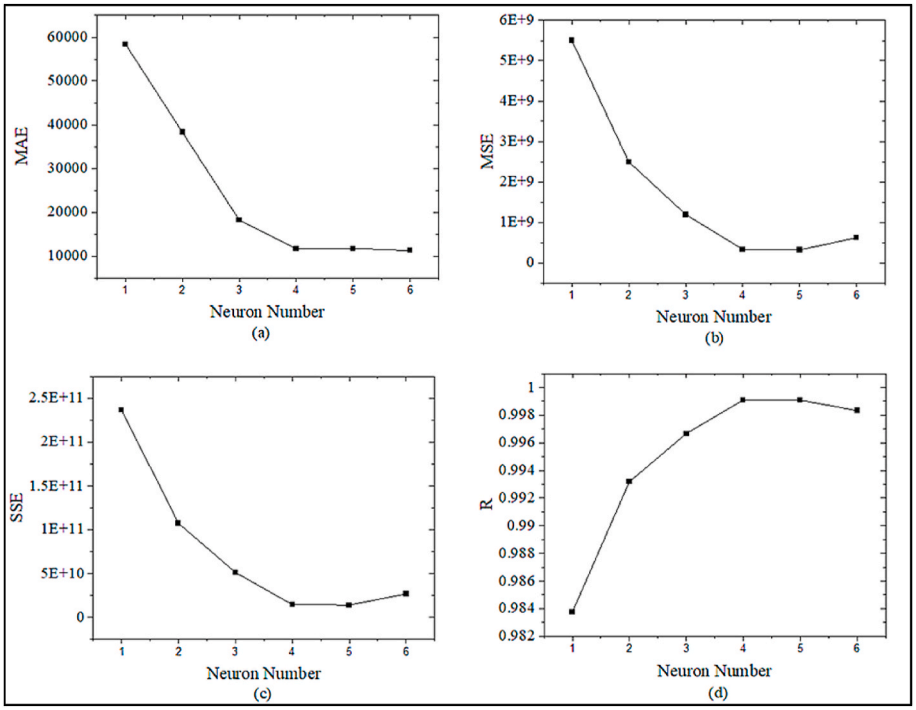


Fig. 15. Cod model performance at various hidden nodes.

it is anticipated that the COD will exceed the stipulated standards by a factor of two to three. However, in this instance, the COD load did not breach Class II standards of the NWQS. This deviation may be attributed to ineffective oxidation of nitrogen during COD determination, suggesting an abundance of nitrogen and nitrate in the Muda River. The principal source of nitrogen predominantly emanates from fertilised agricultural lands, encompassing approximately 55% of the Muda River Basin [33,53]. Additionally, the sampling stations (2MD06, 2MD04, 2MD01, and 2MD15) were situated close to palm oil plantations, palm oil processing facilities, and paddy fields, as substantiated by the GIS data. Consequently, substantial pollution contributions are anticipated from non-point sources, notably agricultural runoff, which further augments ammonium and nitrate levels through the use of organic fertilisers [55].

As shown in Fig. 19 (c), it is noteworthy that 30% of the SS load exceeded 5000 tons per day, with a peak SS load reaching 19,022.39

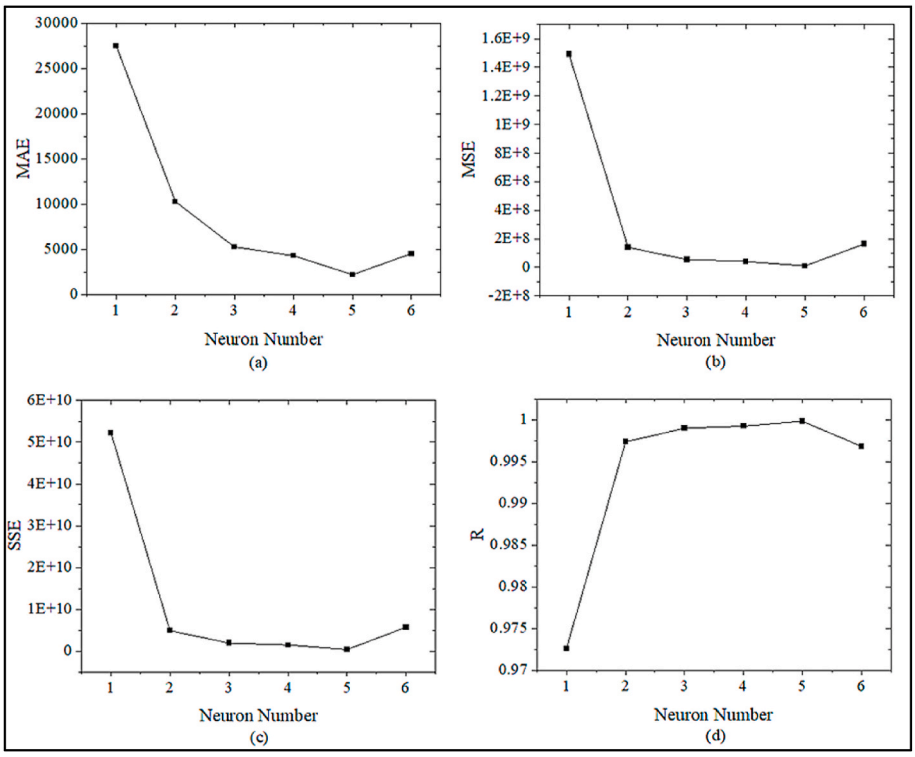


Fig. 16. Bod model performance at various hidden nodes.

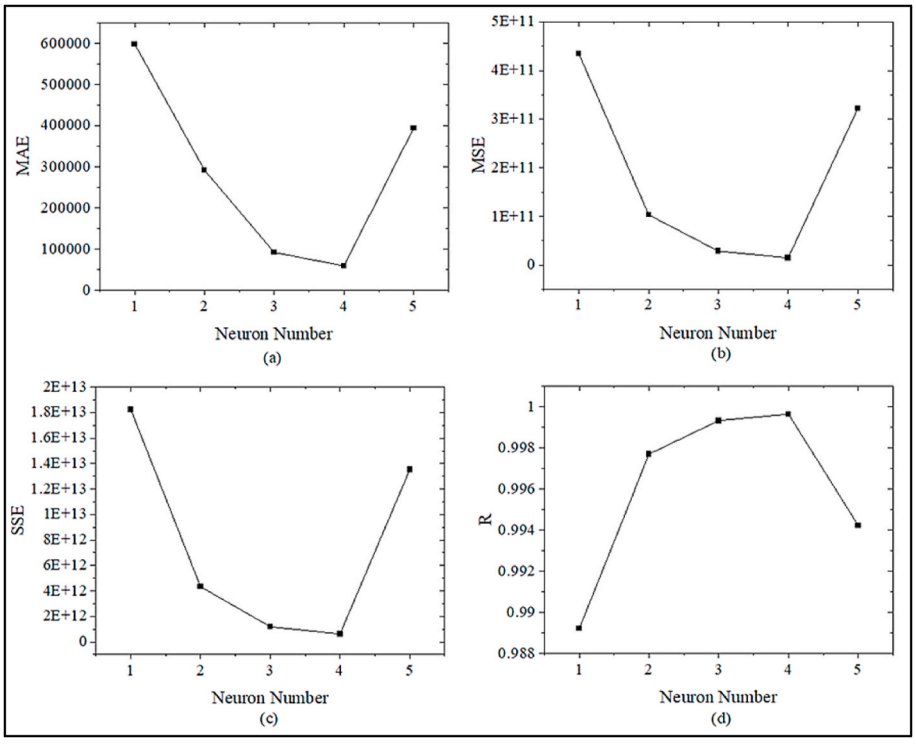


Fig. 17. SS model performance at various hidden nodes.

tons per day at a river discharge of 355.01 m3/s. This observation underscores the substantial daily influx of Suspended Solids into the Muda River. Factors contributing to this heightened sediment load include logging activities, deforestation associated with residential and commercial development, and population expansion within the Muda River Basin. The principal consequence of elevated SS loads in the river is the recurrent inundation during the biannual rainy season.

The core objective of this study was to achieve a reduction in riverine loads to ensure compliance with the Class II standards of the NWQS, stipulating concentration limits of 3, 25, 50, and 0.3 mg NH_3–N/L. It was estimated that an average reduction of 52% in BOD load would be required to achieve compliance. Simultaneously, the COD load required

Table 4
Data calibration and validation of BOD model.

Run Number	Training	Validation	Testing
1	0.99970	0.99998	0.99762
2	0.99074	0.97659	0.98814
3	0.99218	0.99307	0.99159
4	0.99932	0.99997	0.99240
5	0.99801	0.99952	0.98949
6	0.99875	0.99996	0.99503
7	0.99747	0.99980	0.99121
8	0.99896	0.99977	0.99295
9	0.99611	0.99992	0.99438
10	0.99864	0.99999	0.99537
Average	0.99699	0.99686	0.99282

Table 5
Data calibration and validation of COD model.

Run Number	Training	Validation	Testing
1	0.99901	0.99975	0.99749
2	0.99992	0.98912	0.99975
3	0.99799	0.99991	0.99765
4	0.99518	0.99266	0.97985
5	0.99522	0.99789	0.99248
6	0.99669	0.99747	0.98915
7	0.99918	0.99940	0.99697
8	0.99877	0.99845	0.97413
9	0.99274	0.99938	0.99829
10	0.99928	0.99989	0.99940
Average	0.99740	0.99739	0.99252

Table 6
Data calibration and validation of SS model.

Run Number	Training	Validation	Testing
1	0.99958	0.99960	0.99993
2	0.99501	0.99993	0.99423
3	0.99304	0.99097	0.99737
4	0.99827	0.99999	1.00000
5	0.99996	0.96581	0.94749
6	0.99928	0.99476	0.99872
7	0.98013	0.99676	0.90887
8	0.98555	0.99746	0.99664
9	0.99523	0.99990	0.99473
10	0.99936	0.99898	0.99779
Average	0.99454	0.99442	0.98358

Table 7
Data calibration and validation of NH_3–N model.

Run Number	NH_3–N		
	Training	Validation	Testing
1	0.99874	0.99468	0.99807
2	0.99781	0.99214	0.99135
3	0.99862	0.99734	0.99809
4	0.97584	0.99227	0.99879
5	0.99089	0.99848	0.99651
6	0.96769	0.98096	0.99859
7	0.99707	0.99360	0.99832
8	0.99847	0.99884	0.99751
9	0.97817	0.99914	0.99831
10	0.96074	0.99811	0.99689
Average	0.98640	0.99456	0.99724

a nominal reduction of approximately 2% to align with the prescribed 25 mg COD/L standard. Moreover, substantial reductions averaging 60% for SS load and NH_3–N load were essential to attain Class II standards of the NWQS, signifying limits of 50 mg SS/L and 0.3 mg NH_3–N/L, respectively.

Table 8
Comparison criteria for the riverine load prediction models.

Parameter (tonne/d)	Method	RMSE	MAE	MRE	R^2
BOD	MLR	30.5326	18.8764	0.2384	0.9805
	FFBP NN	3.2019	2.2172	0.0471	0.9998
	RB NN	70.9584	52.6285	0.8616	0.8946
COD	MLR	94.7057	69.3731	0.3222	0.9737
	FFBP NN	18.1964	11.6976	0.0757	0.9990
	RB NN	285.3626	222.9565	1.3177	0.7613
SS	MLR	1007.6990	749.3521	0.5453	0.9779
	FFBP NN	121.1806	58.9920	0.0801	0.9996
	RB NN	2556.5280	1876.3930	1.8293	0.8580
NH_3–N	MLR	3.5832	2.5615	0.3786	0.9680
	FFBP NN	0.7511	0.4599	0.0991	0.9986
	RB NN	9.0205	6.9272	1.3055	0.7973

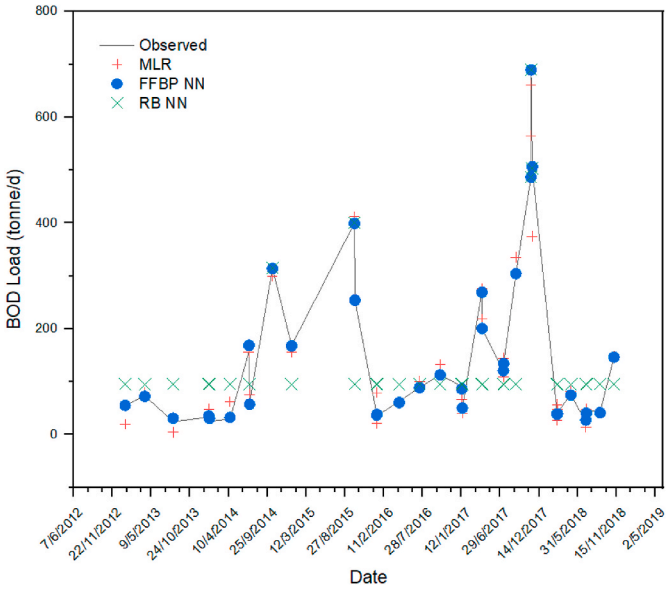


Fig. 18a. Comparison of BOD load prediction results for the FFBP NN, RB NN, and MLR.

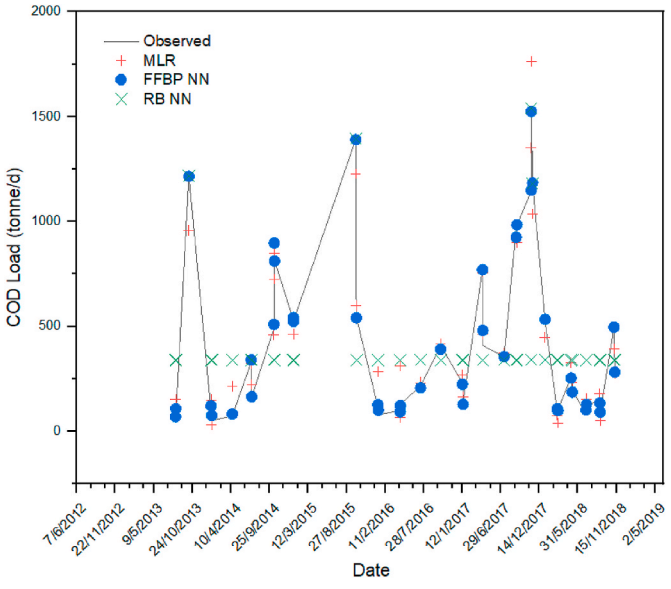


Fig. 18b. Comparison of COD load prediction results for FFBP NN, RB NN, and MLR.

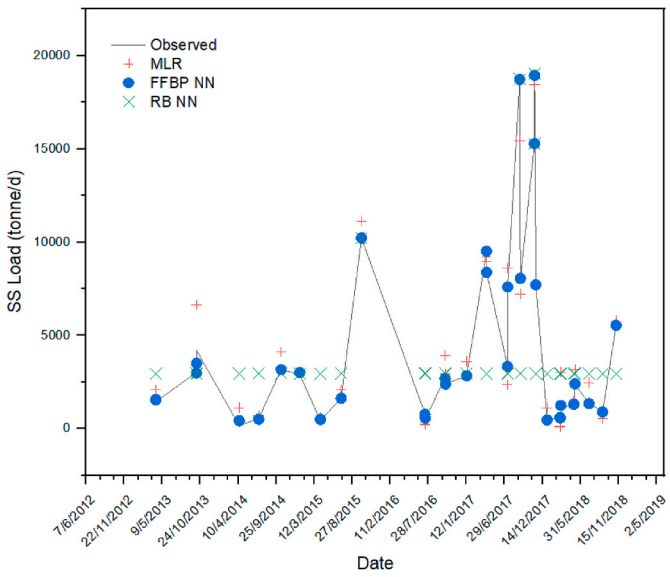


Fig. 18c. Comparison of SS load prediction results for FFBP NN, RB NN, and MLR.

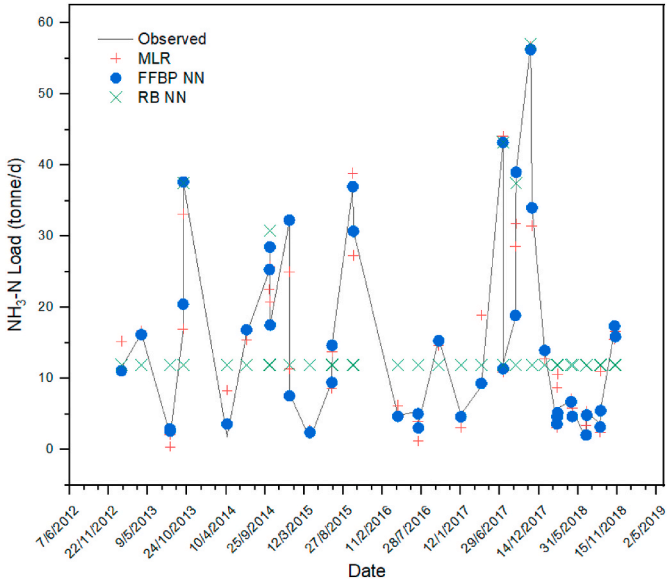


Fig. 18d. Comparison of NH_3–N load prediction results for FFBP NN, RB NN, and MLR.

4. Conclusions

This study presents a comprehensive comparative analysis of diverse mathematical modelling techniques aimed at forecasting riverine loads, focusing on crucial water quality parameters, including Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Suspended Solids (SS), and Ammoniacal Nitrogen (NH_3–N). The riverine load model under examination was established by amalgamating various input variables, notably, river discharge and pollutant concentration data procured from diverse monitoring stations. Among the mathematical modelling methodologies under scrutiny are Artificial Neural Networks (ANNs) that employ feedforward backpropagation algorithms and radial basis functions. To facilitate rigorous comparisons, a conventional Multiple Linear Regression (MLR) statistical model was incorporated. A robust assessment of the model performance was carried out using four standard statistical evaluation metrics: the Root Mean

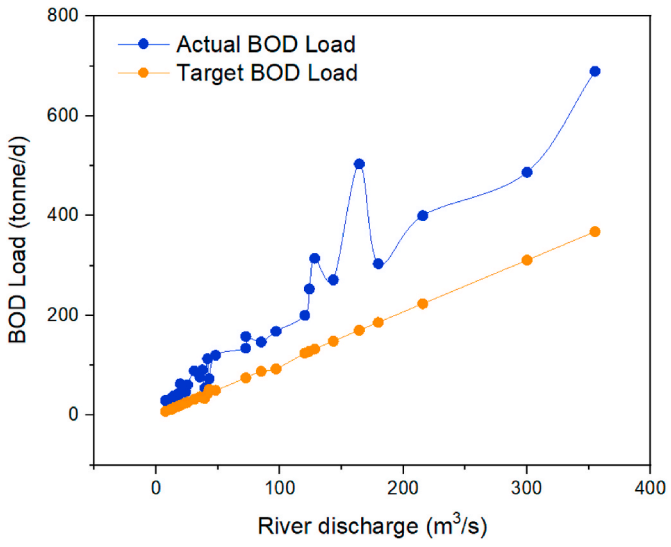


Fig. 19a. Bod load of Muda river with respect to the river discharge from 2013 to 2018.

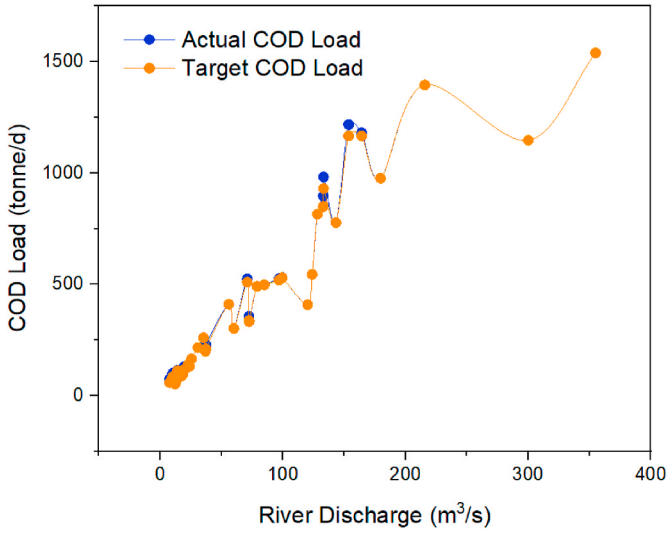


Fig. 19b. Cod load of Muda river with respect to the river discharge from 2013 to 2018.

Square Error (RMSE), Mean Absolute Error (MAE), Mean Relative Error (MRE), and Coefficient of Determination (R^2).

Notably, the elevated error rates (manifested in RMSE, MAE, and MRE) observed in the prediction of riverine loads, attributed to the Multiple Linear Regression (MLR) statistical model, can be attributed to the intricate nonlinear associations inherent in the independent variables (Q and C_x) concerning the dependent variables (W). Furthermore, it is evident that the feedforward neural network model, featuring a backpropagation algorithm and Bayesian regularisation training methodology, outperforms the Radial Basis Neural Network, particularly when operating with a reduced number of hidden nodes. This observation underscores the fact that the Radial Basis Function requires a greater number of hidden nodes to introduce the required nonlinearity into the network. In addition, it is worth noting that the Feedforward Backpropagation Neural Network effectively assimilates the underlying process during calibration. This pivotal finding suggests the feasibility of predicting riverine loads beyond suspended sediment loads using an Artificial Neural Network, wherein the pollutant concentration (C_x) and river discharge (Q) function as key input variables. Such an approach

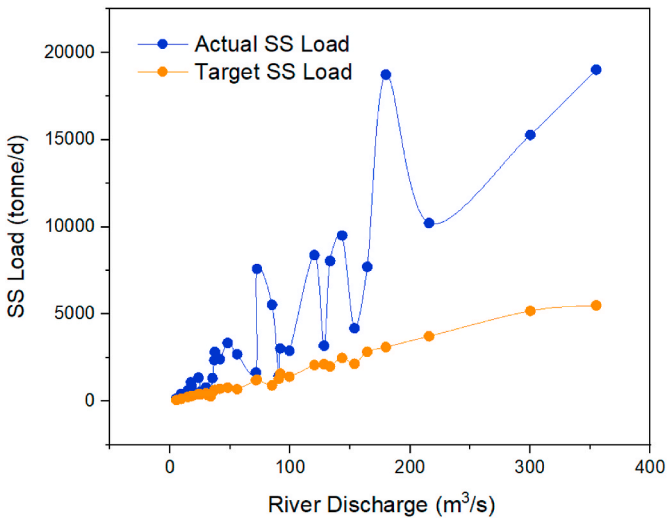


Fig. 19c. SS load of Muda river with respect to the river discharge from 2013 to 2018.

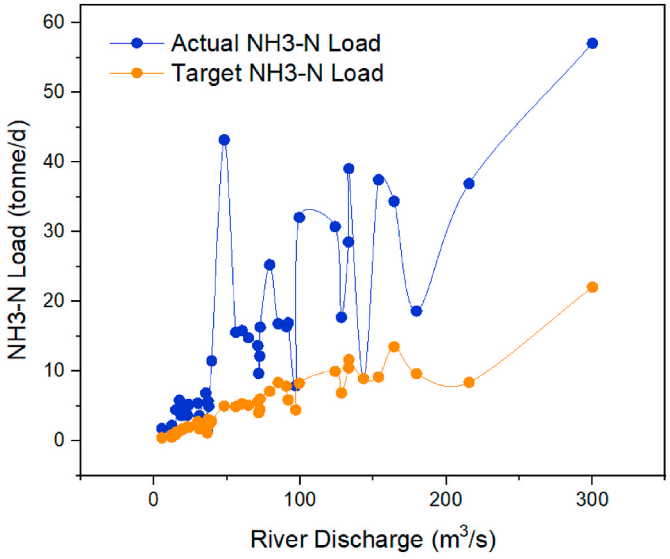


Fig. 19d. NH_3–N load of Muda river with respect to the river discharge from 2013 to 2018.

would obviate the need for extensive data pertaining to variables such as phytoplankton concentration, refractory particulate organic nitrogen, labile particulate organic nitrogen, and kinetic coefficients, which are typically mandated using comprehensive water quality deterministic models for riverine load prediction. However, to enhance riverine load prediction, integration of other spatial and temporal variation parameters that may influence river water quality as additional input variables is recommended. Ultimately, this study underscores the significance of riverine load studies in facilitating informed decisions regarding load-reduction allocation.

Funding

This study did not receive any specific grants from funding agencies in the public, commercial, or not-for-profit sector.

Statement

During the preparation of this work the author(s) used ChatGPT 4.0 in order to improve the structure and clarity of sentences in the manuscript. After using this tool/service, the author(s) reviewed and edited the content as needed and took (s) full responsibility for the content of the publication.

CRediT authorship contribution statement

Khairunnisa Khairudin: Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation. **Ahmad Zia Ul-Saufie: BIN MOHAMAD JAPERI**, Methodology, Conceptualization. **Syahrul Fithry Senin:** Writing – review & editing, Methodology. **Zaki Zainudin:** Writing – review & editing, Conceptualization. **Ammar Mohd Rashid:** Writing – review & editing. **Noor Fitrah Abu Bakar:** Writing – review & editing, Validation. **Muhammad Zakwan Anas Abd Wahid:** Investigation, Formal analysis. **Syahida Farhan Azha:** Writing – review & editing, Visualization. **Firdaus Abd-Wahab:** Writing – review & editing, Validation. **Lei Wang:** Writing – review & editing, Validation. **Farisha Nerina Sahar:** Writing – review & editing, Validation. **Mohamed Syazwan Osman:** Writing – review & editing, Supervision, Resources, Project administration, Conceptualization.

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.

Acknowledgments

We would like to thank the Department of Environment (DOE), Ministry of Environment and Water, Malaysia, for the concentration data and the Department of Irrigation and Drainage (DID), Ministry of Environment and Water, Malaysia, for providing river discharge data.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rineng.2024.102072.

References

[1] M. Song, W. Tao, Y. Shang, X. Zhao, Spatiotemporal characteristics and influencing factors of China's urban water resource utilization efficiency from the perspective of sustainable development, J. Clean. Prod. 338 (Mar) (2022), https://doi.org/10.1016/j.jclepro.2022.130649.

[2] T. Yao, T. Bolch, D. Chen, J. Gao, W. Immerzeel, S. Piao, F. Su, L. Thompson, Y. Wada, L. Wang, T. Wang, The imbalance of the Asian water tower, Nat. Rev. Earth Environ. 3 (10) (2022) 618–632, https://doi.org/10.1038/S43017-022-00299-4. Oct.

[3] G. D'amore, A. Di Vaio, D. Balsalobre-Lorente, F. Boccia, Artificial intelligence in the water–energy–food model: a holistic approach towards sustainable development goals, Sustainability 14 (2) (2022), https://doi.org/10.3390/SU14020867. Jan.

[4] Z. Wang, J. Guo, H. Ling, F. Han, Z. Kong, W. Wang, Function zoning based on spatial and temporal changes in quantity and quality of ecosystem services under enhanced management of water resources in arid basins, Ecol. Indicat. 137 (2022) 108725.

[5] F. García-Ávila, M. Guanoquiza-Suárez, J. Guzmán-Galarza, R. Cabello-Torres, L. Valdiviezo-Gonzales, Rainwater harvesting and storage systems for domestic supply: an overview of research for water scarcity management in rural areas, Results in Engineering 18 (2023) 101153, https://doi.org/10.1016/J.RINENG.2023.101153. Jun.

[6] G. Fu, Y. Jin, S. Sun, Z. Yuan, D. Butler, The role of deep learning in urban water management: a critical review, Water Res. 223 (2022), https://doi.org/10.1016/j.watres.2022.118973. Sep.

[7] A. Rabak, K. Uppuluri, F.F. Franco, N. Kumar, V.P. Georgiev, C. Gauchotte-Lindsay, C. Smith, R.A. Hogg, L. Manjakkal, Sensor system for precision agriculture smart

watering can, Results in Engineering 19 (2023) 101297, https://doi.org/10.1016/J.RINENG.2023.101297. Sep.
[8] M.A. Islam, F. Hossen, M.A. Rahman, K.F. Sultana, M.N. Hasan, M.A. Haque, J. E. Sosa-Hernández, M.A. Oyervides-Muñoz, R. Parra-Saldívar, T. Ahmed, M. T. Islam, An opinion on Wastewater-Based Epidemiological Monitoring (WBEM) with Clinical Diagnostic Test (CDT) for detecting high-prevalence areas of community COVID-19 infections, Curr Opin Environ Sci Health 31 (2023), https://doi.org/10.1016/j.coesh.2022.100396. Feb.
[9] N.A.S. Feisal, N.H. Kamaludin, M.F.A. Sani, D.K.A. Ahmad, M.A. Ahmad, N.F. A. Razak, T.N.B.T. Ibrahim, Anthropogenic disturbance of aquatic biodiversity and water quality of an urban river in Penang, Malaysia, Water Sci. Eng. 16 (3) (2023) 234–242. https://doi.org/10.1016/j.wse.2023.01.003.
[10] M.A. Zali, A. Retnam, H. Juahir, S.M. Zain, M.F. Kasim, B. Abdullah, S.B. Saadudin, Sensitivity analysis for water quality index (WQI) prediction for Kinta River, Malaysia, World Appl. Sci. J. 14 (2011) 60–65.
[11] M. Hameed, S.S. Sharqi, Z.M. Yaseen, H.A. Afan, A. Hussain, A. Elshafie, Application of artificial intelligence (AI) techniques in water quality index prediction: a case study in tropical region, Malaysia, Neural Comput. Appl. 28 (2017) 893–905.
[12] M. Zhu, J. Wang, X. Yang, Y. Zhang, L. Zhang, H. Ren, B. Wu, L. Ye, A review of the application of machine learning in water quality evaluation, Eco-Environment and Health 1 (2) (2022) 107–116, https://doi.org/10.1016/j.eehl.2022.06.001. Jun.
[13] H.A. Mohiyaden, L.M. Sidek, S.M. Noh, M.K. Selamat, M. Marufuzzaman, Water quality index analysis in Kim-Kim River: effects of Covid-19 movement control order, in: IOP Conference Series: Earth and Environmental Science, IOP Publishing, 2023 012002.
[14] W. Zhi, D. Feng, W.P. Tsai, G. Sterle, A. Harpold, C. Shen, L. Li, From hydrometeorology to river water quality: can a deep learning model predict dissolved oxygen at the continental scale? Environ. Sci. Technol. 55 (4) (2021) 2357–2368, https://doi.org/10.1021/ACS.EST.0C06783. Feb.
[15] H. Juahir, S.M. Zain, M.K. Yusoff, T.T. Hanidza, A.M. Armi, M.E. Toriman, M. Mokhtar, Spatial water quality assessment of Langat River Basin (Malaysia) using environmetric techniques, Environ. Monit. Assess. 173 (2011) 625–641.
[16] M. Poursaeid, A.H. Poursaeid, S. Shabanlou, A comparative study of artificial intelligence models and A statistical method for groundwater level prediction, Water Resour. Manag. 36 (5) (2022) 1499–1519, https://doi.org/10.1007/S11269-022-03070-Y. Mar.
[17] H. Zare Abyaneh, Evaluation of multivariate linear regression and artificial neural networks in prediction of water quality parameters, J Environ Health Sci Eng 12 (2014) 1–8.
[18] J. Berbel, A. Expósito, A decision model for stochastic optimization of seasonal irrigation-water allocation, Agric. Water Manag. 262 (Mar) (2022), https://doi.org/10.1016/j.agwat.2021.107419.
[19] A. Kikon, P.C. Deka, Artificial intelligence application in drought assessment, monitoring and forecasting: a review, Stoch. Environ. Res. Risk Assess. 36 (5) (May 2022) 1197–1214, https://doi.org/10.1007/S00477-021-02129-3.
[20] V. Gholami, M.R. Khaleghi, S. Pirasteh, M.J. Booij, Comparison of self-organizing map, artificial neural network, and Co-active neuro-fuzzy inference system methods in simulating groundwater quality: geospatial artificial intelligence, Water Resour. Manag. 36 (2) (2022) 451–469, https://doi.org/10.1007/S11269-021-02969-2. Jan.
[21] A. Najah, A. El-Shafie, O.A. Karim, A.H. El-Shafie, Application of artificial neural networks for water quality prediction, Neural Comput. Appl. 22 (2013) 187–201.
[22] H. Kamyab, T. Khademi, S. Chelliapan, M. SaberiKamarposhti, S. Rezania, M. Yusuf, M. Farajnezhad, M. Abbas, B.H. Jeon, Y. Ahn, The latest innovative avenues for the utilization of artificial Intelligence and big data analytics in water resource management, Results in Engineering 20 (2023) 101566, https://doi.org/10.1016/J.RINENG.2023.101566. Dec.
[23] E. Dogan, B. Sengorur, R. Koklu, Modeling biological oxygen demand of the Melen River in Turkey using an artificial neural network technique, J. Environ. Manag. 90 (2) (2009) 1229–1235.
[24] M. Lowe, R. Qin, X. Mao, A review on machine learning, artificial intelligence, and smart technology in water treatment and monitoring, Water (Switzerland) 14 (9) (2022), https://doi.org/10.3390/W14091384. May.
[25] N.N.M. Rizal, G. Hayder, K.A. Yusof, Water quality predictive analytics using an artificial neural network with a graphical user interface, Water (Switzerland) 14 (8) (2022), https://doi.org/10.3390/W14081221. Apr.
[26] K.P. Singh, A. Basant, A. Malik, G. Jain, Artificial neural network modeling of the river water quality—a case study, Ecol. Model. 220 (6) (2009) 888–895.
[27] J. Park, W.H. Lee, K.T. Kim, C.Y. Park, S. Lee, T.Y. Heo, Interpretation of ensemble learning to predict water quality using explainable artificial intelligence, Sci. Total Environ. 832 (Aug) (2022), https://doi.org/10.1016/j.scitotenv.2022.155070.
[28] K.B. Newhart, J.E. Goldman-Torres, D.E. Freedman, K.B. Wisdom, A.S. Hering, T. Y. Cath, Prediction of peracetic acid disinfection performance for secondary municipal wastewater treatment using artificial neural networks, ACS ES&T Water 1 (2) (2020) 328–338.
[29] B. Beiranvand, T. Rajaee, Application of artificial intelligence-based single and hybrid models in predicting seepage and pore water pressure of dams: a state-of-the-art review, Adv. Eng. Software 173 (2022), https://doi.org/10.1016/j.advengsoft.2022.103268. Nov.
[30] S. Lin, J. Kim, C. Hua, S. Kang, M.H. Park, Comparing artificial and deep neural network models for prediction of coagulant amount and settled water turbidity: lessons learned from big data in water treatment operations, J. Water Proc. Eng. 54 (Aug) (2023), https://doi.org/10.1016/j.jwpe.2023.103949.
[31] A. Ibrahim, A. Ismail, H. Juahir, A.B. Iliyasu, B.T. Wailare, M. Mukhtar, H. Aminu, Water quality modelling using principal component analysis and artificial neural network, Mar. Pollut. Bull. 187 (2023) 114493.
[32] A.A. Ghani, R. Ali, N.A. Zakaria, Z.A. Hasan, C.K. Chang, M.S.S. Ahamad, A temporal change study of the Muda River system over 22 years, Int. J. River Basin Manag. 8 (1) (2010) 25–37.
[33] M.N.A. Zakaria, A.N. Ahmed, M.A. Malek, A.H. Birima, M.M.H. Khan, M. Sherif, A. Elshafie, Exploring machine learning algorithms for accurate water level forecasting in Muda river, Malaysia, Heliyon 9 (7) (2023) 12345.
[34] W. Huang, J. Mao, D. Zhu, C. Lin, Impacts of land use and land cover on water quality at multiple buffer-zone scales in a Lakeside City, Water (Basel) 12 (1) (2020) 47.
[35] D. of Environment, "National Water Quality Standards for Malaysia.".
[36] X. Sun, H. Zhang, M. Zhong, Z. Wang, X. Liang, T. Huang, H. Huang, Analyses on the temporal and spatial characteristics of water quality in a seagoing river using multivariate statistical techniques: a case study in the Duliujian River, China, Int. J. Environ. Res. Publ. Health 16 (6) (2019) 1020.
[37] V.K. Gautam, C.B. Pande, K.N. Moharir, A.M. Varade, N.L. Rane, J.C. Egbueri, F. Alshehri, Prediction of sodium hazard of irrigation purpose using artificial neural network modelling, Sustainability 15 (9) (2023) 7593.
[38] P.P. Sarangi, L. Sabat, B. Naik, Prediction and parametric evaluation of water quality on basin of River Brahmani by ANN and GIS, Mater. Today Proc. (2023Available online 19 June 2023In Press, Corrected Proof). https://doi.org/10.1016/j.matpr.2023.06.041.
[39] M.G. Uddin, S. Nash, A. Rahman, A.I. Olbert, Assessing optimization techniques for improving water quality model, J. Clean. Prod. 385 (2023) 135671.
[40] S.C. Chapra, Surface Water-Quality Modeling, Waveland press, 2008.
[41] X. Zhang, W. Li, X. Zhang, G. Cai, K. Meng, Z. Shen, Application of grey feed forward back propagation-neural network model based on wavelet denoising to predict the residual settlement of goafs, PLoS One 18 (5) (2023) e0281471.
[42] M. Najafzadeh, G. Oliveto, More reliable predictions of clear-water scour depth at pile groups by robust artificial intelligence techniques while preserving physical consistency, Soft Comput. 25 (2021) 5723–5746.
[43] S. Lee, I. Park, J.-K. Choi, Spatial prediction of ground subsidence susceptibility using an artificial neural network, Environ. Manag. 49 (2012) 347–358.
[44] N. Basant, S. Gupta, A. Malik, K.P. Singh, Linear and nonlinear modeling for simultaneous prediction of dissolved oxygen and biochemical oxygen demand of the surface water—a case study, Chemometr. Intell. Lab. Syst. 104 (2) (2010) 172–180.
[45] B.J. Amiri, K. Nakane, Comparative prediction of stream water total nitrogen from land cover using artificial neural network and multiple linear regression, Pol. J. Environ. Stud. 18 (2) (2009).
[46] W. Yang, J. Nan, D. Sun, An online water quality monitoring and management system developed for the Liming River basin in Daqing, China, J. Environ. Manag. 88 (2) (2008) 318–325.
[47] J.-P. Suen, J.W. Eheart, Evaluation of neural networks for modeling nitrate concentrations in rivers, J. Water Resour. Plann. Manag. 129 (6) (2003) 505–510.
[48] B. He, T. Oki, F. Sun, D. Komori, S. Kanae, Y. Wang, H. Kim, D. Yamazaki, Estimating monthly total nitrogen concentration in streams by using artificial neural network, J. Environ. Manag. 92 (1) (2011) 172–177.
[49] T.P. Vogl, J.K. Mangis, A.K. Rigler, W.T. Zink, D.L. Alkon, Accelerating the convergence of the back-propagation method, Biol. Cybern. 59 (1988) 257–263.
[50] J. Heaton, Introduction to Neural Networks with Java, Heaton Research, Inc, 2008.
[51] P.L. Georgescu, S. Moldovanu, C. Iticescu, M. Calmuc, V. Calmuc, C. Topa, L. Moraru, Assessing and forecasting water quality in the Danube River by using neural network approaches, Sci. Total Environ. 879 (2023) 162998.
[52] H. Imanian, H. Shirkhani, A. Mohammadian, J. Hiedra Cobo, P. Payeur, Spatial interpolation of soil temperature and water content in the land-water interface using artificial intelligence, Water (Basel) 15 (3) (2023) 473.
[53] M.N.A. Zakaria, M.A. Malek, M. Zolkepli, A.N. Ahmed, Application of artificial intelligence algorithms for hourly river level forecast: a case study of Muda River, Malaysia, Alex. Eng. J. 60 (4) (2021) 4015–4028.
[54] M.K.N. Shamsuddin, W.N.A. Sulaiman, M.F. Bin Ramli, F.M. Kusin, Geochemical characteristic and water quality index of groundwater and surface water at Lower River Muda Basin, Malaysia, Arabian J. Geosci. 12 (2019) 1–27.
[55] M.K.N. Shamsuddin, W.N.A. Sulaiman, M.F. Ramli, F. Mohd Kusin, K. Samuding, Assessments of seasonal groundwater recharge and discharge using environmental stable isotopes at Lower Muda River Basin, Malaysia, Appl. Water Sci. 8 (2018) 1–12.
[56] E. Sharaf El Din, Y. Zhang, A. Suliman, Mapping concentrations of surface water quality parameters using a novel remote sensing and artificial intelligence framework, Int. J. Rem. Sens. 38 (4) (2017) 1023–1042.
[57] C. Nhantumbo, F. Carvalho, C. Uvo, R. Larsson, M. Larson, Applicability of a processes-based model and artificial neural networks to estimate the concentration of major ions in rivers, J. Geochem. Explor. 193 (2018) 32–40.
[58] M. Heydari, E. Olyaie, H. Mohebzadeh, Ö. Kisi, Development of a neural network technique for prediction of water quality parameters in the Delaware River, Pennsylvania, Middle East J. Sci. Res. 13 (10) (2013) 1367–1376.
[59] A. Ibrahim, A. Ismail, H. Juahir, A.B. Iliyasu, B.T. Wailare, M. Mukhtar, H. Aminu, Water quality modelling using principal component analysis and artificial neural network, Mar. Pollut. Bull. 187 (2023) 114493.

[60] K.A. Tareke, A.G. Awoke, Hydrological drought forecasting and monitoring system development using artificial neural network (ANN) in Ethiopia, Heliyon 9 (2) (2023) 12345.
[61] H. Banejad, E. Olyaie, Application of an artificial neural network model to rivers water quality indexes prediction–a case study, Journal of American science 7 (1) (2011) 60–65.
[62] A.R. Senthil Kumar, C.S.P. Ojha, M.K. Goyal, R.D. Singh, P.K. Swamee, Modeling of suspended sediment concentration at Kasol in India using ANN, fuzzy logic, and decision tree algorithms, J. Hydrol. Eng. 17 (3) (2012) 394–404.