Results in Engineering 23 (2024) 102541

Contents lists available at ScienceDirect

Results in Engineering

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

Metaheuristic algorithms applied in ANN salinity modelling

Zahraa S. Khudhair [a], Salah L. Zubaidi [a,b,*], Anmar Dulaimi [c,d], Hussein Al-Bugharbee [e], Yousif Raad Muhsen [a,f], Ramadhansyah Putra Jaya [g,**], Hussein Mohammed Ridha [h], Syed Fawad Raza [i], Saleem Ethaib [j]

[a] Department of Civil Engineering, Wasit University, Wasit, 52001, Iraq
[b] College of Engineering, University of Warith Al-Anbiyaa, Karbala, 56001, Iraq
[c] College of Engineering, University of Kerbala, Karbala, 56001, Iraq
[d] School of Civil Engineering and Built Environment, Liverpool John Moores University, Liverpool, L3 2ET, United Kingdom
[e] Department of Mechanical Engineering, Wasit University, Wasit, 52001, Iraq
[f] Technical Engineering College, Al-Ayen University, Thi-Qar, 64001, Iraq
[g] Faculty of Civil Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah, Kuantan, 26300, Pahang, Malaysia
[h] Department of Computer Engineering, University of Al-Mustansiriyah, 10001 Baghdad, Iraq
[i] College of Earth and Environmental Sciences, University of the Punjab, Lahore, Pakistan
[j] Department of Civil Engineering, College of Engineering, University of Thi-Qar, 64001, Al-Nassiriya, Iraq

ARTICLE INFO

Keywords:
Artificial neural network
Metaheuristic algorithm
Data pre-processing
Surface water quality
Euphrates river

ABSTRACT

Salinity is a classic problem in planning the quality of freshwater resources management. Recent studies related to hybrid machine learning models have shown it's capability to simulate salinity dynamics. However, previous studies of metaheuristic algorithms have not dealt with comparing single- and hybrid-based algorithms in much detail. The present study aimed to develop univariate salinity by applying an artificial neural network model (ANN) integrated with (hybrid-based) coefficient-based particle swarm optimisation and chaotic gravitational search algorithm (CPSOCGSA). The methodology was developed and tested using electrical conductivity (EC) and total dissolved solids (TDS) data collected from the Euphrates River in Babylon Province, Iraq, from 2010 to 2019. The CPSOCGSA performance was evaluated by various single-based ones, including multi-verse optimiser (MVO), marine predator's optimisation algorithm (MPA), particle swarm optimiser (PSO), and the slim mould algorithm (SMA). The principal finding here confirms that hybrid-based outperformed four single-based algorithms based on different criteria. The outcomes for TDS were 0.004, 0.0248, and 0.98 for CPSOCGSA-ANN technique concern scatter index (SI), root-mean-squared error (RMSE), and correlation coefficient (R^2), respectively. For EC, the results were 0.96 for R^2, 0.0386 for RMSE, and 0.006 for SI. Due to its predictive accuracy, the proposed CPSOCGSA-ANN approach is suggested as a potential strategy for predicting monthly salinity data. Considering agriculture's vital role in Babylon Province's economy, this study may help inform future freshwater quality management decisions.

1. Introduction

Water quality (WQ) refers to water's chemical, physical, and biological features and suitability for specific purposes [1]. Various studies have revealed the potential WQ degradation caused by increased activities that demand large amounts of water due to the continued growth of the world's population. In addition, the drop-in river flow reduced the pollutants' dilution and increased their concentration in various world rivers. Thus, several of these actions pollute the environment and breach the sustainability limits around water resource employment [2–4]. Furthermore, many river systems, which are used for drinking, irrigation, and industrial, worldwide have recently suffered from water pollution due to increased salinity levels [5,6].

Since 2003, Iraq's rivers (the Euphrates and Tigris) have significantly

* Corresponding author. Department of Civil Engineering, Wasit University, Wasit, 52001, Iraq.
** Corresponding author.
E-mail addresses: zahra.asadik2@gmail.com (Z.S. Khudhair), salahlafta@uowasit.edu.iq (S.L. Zubaidi), anmar.f@uokerbala.edu.iq (A. Dulaimi), hrazzaq@uowasit.edu.iq (H. Al-Bugharbee), Yousif@uowasit.edu.iq (Y.R. Muhsen), ramadhansyah@umpsa.edu.my (R. Putra Jaya), hussain_mhammad@uomustansiriyah.edu.iq (H.M. Ridha), fawad.raza@gmail.com (S.F. Raza), dr.saleem@utq.edu.iq (S. Ethaib).

https://doi.org/10.1016/j.rineng.2024.102541
Received 19 May 2024; Received in revised form 27 June 2024; Accepted 8 July 2024
Available online 10 July 2024

reduced discharge and water levels due to terrorist attacks on numerous barrages and dams. In addition, between 2009 and 2014, several dams were built on the paths of rivers (i.e., Syria, Iran, and Turkey), which had a negative effect on the water control of the rivers in recent years. Consequently, WQ has degraded, creating concerns about excessive salinity levels [2,7]. Also, salinity levels in the Euphrates River in Iraq have risen [8]. The problem at hand is that high concentrations of total dissolved solids and electrical conductivity (TDS and EC) lead to low WQ indices, which are essential criteria in defining water salinity for municipal and agricultural water [9]. As a result, trusty prediction techniques are a critical need for policymakers, as they enable improved management and conservation of water quality. In response, an open approach would be to develop predictive models that rely on machine learning (ML) algorithms built on comprehensive datasets, including all the relevant parameters, allowing for effective WQ management.

Conventional modelling approaches are inadequately practical in WQ issues because they only deal with linear relationships [10]. Meanwhile, employing ML creates a flexible mathematical expression capable of detecting non-linear and complicated correlations between predictor and target factors [11,12]. ML approaches have been applied in the prediction of WQ parameters, including adaptive neuro-fuzzy inference system [13], support vector regressions [14], random forest [5,15], and artificial neural network (ANN) [16]. Several studies on WQ forecasting, including Sha et al. [17], Wang et al. [18], Choi et al. [19], and Monteiro and Costa [20], have used historical WQ data as predictors in their forecast models because of their simplicity and limited data needs. As a basic and functional model, ANN is suited to handling non-linear, uncertain problems and can capture functional correlations between WQ parameters [21,22]. Therefore, its application in hydrological modelling has been extensive [23]. Unfortunately, single models do not yield accurate results because of the intricacy of the data structure and the use of the trial-and-error methodology in choosing the hyperparameters [24]. Hence, the tendency toward using hybrid ANN models has been increasing. These techniques play an essential role in the simulations of WQ parameters. Additionally, they can be combined with metaheuristic algorithms (MHAs) to create efficient and flexible models [21]. In these hybrid systems, one of the techniques is usually considered the major, with the others serving as pre- or post-processing procedures [25]. Multiple researchers applied hybrid models in hydrological prediction and improved outperformed them on the same single technique, such as Zhou et al. [26] and Raheli et al. [27].

In the same context, the pre-processing data methods are another crucial factor to consider. They may effectively overcome the WQ issue and choose the appropriate independent scenario, as proven by Sha et al. [17]. As a result, different pre-treatment signal techniques have been applied to reduce noise in WQ data, including the singular spectrum analysis (SSA) [28] and ensemble empirical mode decomposition [29]. Another essential part of pre-processing data is choosing the optimal set of model input factors, e.g. a univariate procedure using mutual information (MI) [30]. A non-linear statistical dependency approach, i.e., MI, is appropriate for choosing model input factors for ANN models [31].

Moreover, different MHAs are available for usage in diverse application settings. The optimisation algorithms aim to find the best system parameter values under many scenarios [32]. Among these techniques is particle swarm optimisation (PSO), which has been applied to handle multiple optimisation issues since it can deal with complex problems, has a fast convergence rate, and has good generalisation capabilities for various situations [33]. Thus, several fields of study have benefited from its implementation, such as forecasting floods [34], water quality [35], electric vehicles [36], and industrial design [37]. In addition, the slim mould algorithm (SMA) is one of the newest nature-inspired algorithms developed by Li et al. [38]. It has been applied to a wide range of optimisation problems, including those arising in engineering design [39] and solar photovoltaic systems [40]. Also, the multi-verse optimiser (MVO) developed by Mirjalili et al. [41], which has been efficiently used in several fields, such as stream flow field [42], and hydrology [43]. Moreover, a marine predator's algorithm (MPA) was proposed by Faramarzi et al. [44] and has multiple uses, such as water level [45].

Furthermore, Khudhair et al. [46] reviewed a combined technique to predict WQ and indicated that space for development concerning the WQ parameter prediction exists. Thus far, few combination strategies (data pre-processing approaches, ML models, and metaheuristic algorithms) have been applied to predict WQ parameters. Additionally, the use of SSA as a pre-treatment signal method has also been suggested. Support using various methods to choose the optimum predictors to enhance the model's performance has also been provided. Along with creating new (single-based, i.e., SMA) metaheuristic algorithms, alternative strategies combine several algorithms' best features to create a superior algorithm (hybrid-based, i.e., CPSOCGSA).

Hajirahimi and Khashei [47] mentioned that by combining two or more hybrid classes rather than just joining the conventional individual predicting approaches, a new concept called hybridisation of hybrid models has been put out in the literature to achieve highly accurate results. The hybridisation of parameter optimisation-based with preprocessing-based hybrid models (HOPH) is one of the effectively implemented methods that has several current gaps that need to be filled.

Accordingly, this paper aims to build a new methodology to forecast water salinity precisely utilising previous WQ data lags (TDS and EC). To achieve this aim, the following objectives will be carried out: (1) Apply pre-processing data approaches to increase quality data by SSA technique and choose the best predictor (lags) by MI method. (2) Integrate the ANN method with the CPSOCGSA algorithm for hyper-parameters tuning and structure configuration to forecast better water salinity. (3) Evaluate the performance of the recent CPSOCGSA-ANN method by comparing it with SMA-ANN, MPA-ANN, PSO-ANN, and MVO-ANN techniques to raise the forecast range and reduce uncertainty. (4) Apply the novel hybrid technique of HOPH to simulate monthly water salinity considering multiple time lags. (5) Test various new metaheuristic techniques to expand the range of possible outcomes from monthly water salinity simulations and reduce the associated uncertainty (i.e., one hybrid-base algorithm and four single-based algorithms).

In order to accomplish all of the aforementioned objectives, this research contributes to the body of knowledge: (1) This study examines a new HOPH model including SSA, MI, and ANN integrated with CPSOCGSA technique to forecast monthly salinity data. (2) Applying and comparing a hybrid-based algorithm (CPSOCGSA-ANN) with four single-based algorithms (SMA, MPA, MVO, and PSO). (3) The province of Babylon relies economically on agriculture but is already experiencing water salinity stress, and this is the first time that the salinity of the Al-Euphrates River has been predicted at Babylon Governorate using data with multiple time lags.

2. Case study and data description

Iraq is a country in southwest Asia that covers 438,320 square kilometres (km^2). Discharge rates in the Euphrates and Tigris rivers, Iraq's principal sources of surface water, have already dropped to less than a third of their typical capacity due to climate change. As a result, investigating river water quality is important because decreased river water leads to increased salinity. The catchment area in this research is the Al-Musayyab District, north Babylon Governorate, between longitudes (44°20′43“E and 44°29′32“E), latitudes (32°31′50“N and 33°7′36“N), and it has an area of land of 1008 km^2. Babylon Governorate, which covers 5119 km^2. Generally, the climate in the Babylon district is dry; the temperatures exceed 40 °C in summer, and agriculture is an essential economic resource [48]. Monthly time series of TDS (milligram/litter, mg/l) and EC (micromhos/centimetre, μ mhos/cm) parameters were used in the study. The data were gathered over ten years (2010–2019) from the AL-Musayyab point at the Euphrates River

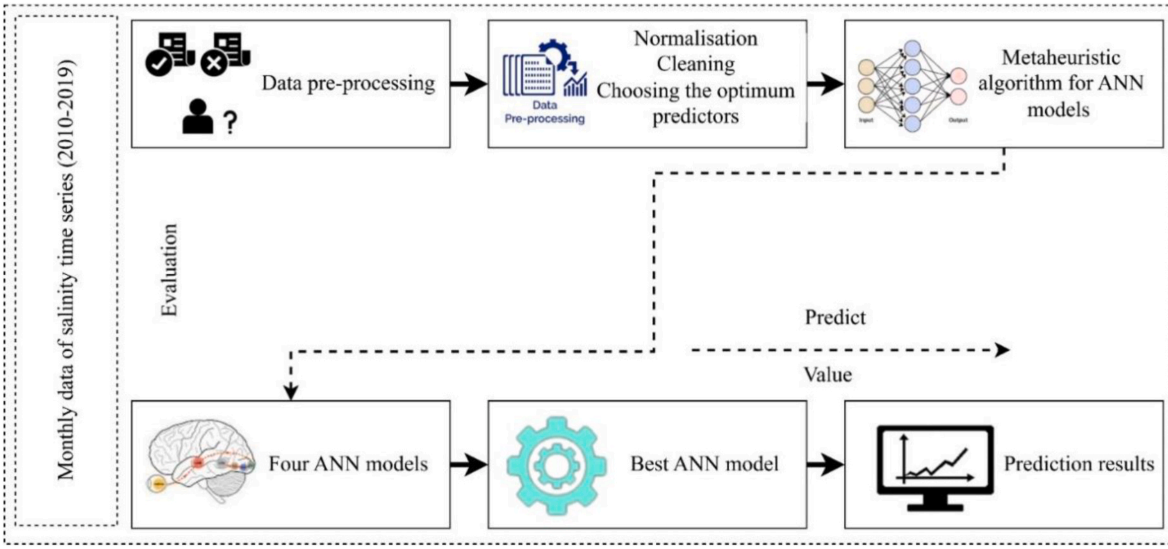

Fig. 1. An illustration of the proposed method to foretell monthly salinity data.

by the Ministry of the Environment.

3. Methodology

Four categories describe the approach proposed for monthly salinity forecasting: (1) data pre-processing, (2) CPSOCGSA algorithm, (3) ANN model, and (4) model evaluation metrics (see Fig. 1).

3.1. Data pre-processing

Data should be appropriately formatted and pre-processed before being utilised in an ANN model. These techniques guarantee that each input in the learning period receives equal attention [49]. It can be classified into three phases: normalisation, cleaning, and choosing the optimum predictors. Normalisation is performed to ensure that the time series follows a normal distribution or a distribution that is extremely near to it. The natural logarithm method has been used to reduce multicollinearity among independent parameters, as in Zubaidi et al. [25]. Cleaning data is compulsory to detect and treat unacceptable values because these values may have an adverse effect on data analysis and the performance of the proposed technique [25]. Data is subsequently denoised using the SSA approach.

SSA successfully analyses the normalised and clean data into several principal components (PCs). The 1st PC has the biggest variation value, and the latter PC has the smallest percentage; hence, each PC explains a piece of the variation of the original data. It can be utilised with nonlinear and linear time series data and a decent sample size. SSA may be used to denoise data by choosing the PCs with the highest proportions of variation and ignoring the lowest proportions of variance, which often account for the time series' structureless noise. In addition, SSA optimises the coefficient of regression and reduces the scale of error by detecting and removing noise in the data [50]. Extra details regarding SSA are available in Golyandina and Zhigljavsky [51].

The final step of the data pre-processing technique, which is considered one of the most critical phases in developing a suitable prediction technique, is the choice of the best scenario's input. In this study, the best explanatory factors are identified using the MI approach. Using this technique, we can find out how well the goal and delayed data correlate statistically. This feature enables the selection of the most highly correlated components with the highest MI [52].

3.2. Overview of the hybridised Constriction Coefficient-Based Particle Swarm Optimisation and chaotic gravitational search algorithm (CPSOCGSA)

This algorithm integrates constriction-based PSO with chaotic GSA to address the intensification, randomisation, and local minimum difficulties that plague traditional GSA and PSO. The component of the current combined technique is investigated in this section.

A. Constriction Coefficient-Based Particle Swarm Optimisation (CPSO)

The PSO method was developed after observing how schools of fish or flocks of birds locate sources of food. There are limitations to this approach, such as how to properly take into account particle motions that occur outside of the solution space. For example, the time of convergence through the optimisation procedure can be addressed by developing constriction coefficients to enhance the PSO exploitation stage [53].

B. Chaotic Gravitational Search Algorithm CGSA

GSA is one of the optimisation approaches based on physical phenomena. It is influenced mainly via Newton's theory of gravity and motion. This method begins the optimisation procedure by modelling the seeking agents as masses. It is necessary to introduce the constant G to ensure that the solution space is sufficiently constrained to support a viable region. Rather and Bala [54] and Rather and Bala [55] used the chaotic normalisation process to describe how G will behave over time.

C. Combination of CCPSO and CGSA

Combining the two methods (CPSO and CGSA) can help mitigate each method's drawbacks and reinforce both positives. More details about the CPSOCGSA algorithm can be found in Rather and Bala [55].

3.3. Artificial neural network (ANN) technique

This technique is a ML strategy similar to a human brain simulation. It can handle large datasets, link inputs and outputs, and can quickly learn a pattern and predict a model's output in a dimensional space [32, 56]. This research uses a multilayer perceptron, feed-forward network (MLFFNN), and Levenberg-Marquart (LM) method is used to train algorithms because it is flexible computation with high demands [57]. By

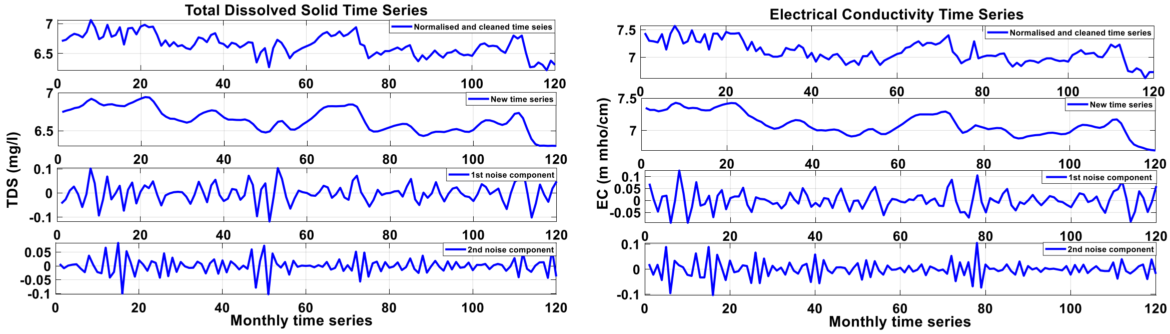

Fig. 2. Normalised and clean time series and three components after pre-processing data for TDS and EC time series.

comparing MLFFNNs with one and two hidden layers, Thomas et al. [58] sought to determine whether the latter had superior generalisation performance. Researchers discovered that two-hidden-layer networks achieved exceptionally well in nine out of ten situations, even while the exact degree of enhancement varies from case to case. Additionally, ANNs with two hidden layers are effective in several studies in representing the nonlinear connection between observed and predicted data [32,59]. Thus, the proposed ANN has four layers: (i) an input layer where the data are first introduced to the network; (ii) two hidden layers where the data are processed; and (iii) an output layer, which serves as the target and is activated using a linear activation function. ANN is beneficial for many different kinds of hydrology, such as water and wastewater management [60], riverine load forecast [61], and prediction of water quality index [62].

The time series were segmented into three categories: seventy percent of the data was used for training, 15 percent for testing, and 15 percent for validation, respectively, as earlier done by Tahraoui et al. [63]. The trial-and-error strategy does not always produce the optimum result. Several metaheuristic techniques were utilised with ANN to determine the ideal neuron number in the first and second hidden layers (N1 and N2), respectively, and the ideal learning rate value (Lr) to create the best independent/dependent mapping and minimise over- and underestimation [25].

3.4. Model performance indicators

Due to the lack of universal performance metrics suited for specific usage, the effectiveness of the proposed methodologies was validated using a wide range of statistical criteria. In this research, five criteria were used: root mean square error (RMSE) (Equation (1)), mean absolute error (MAE) (Equation (2)), mean absolute relative error (MARE) (Equation (3)), coefficient of determination (R^2) (Equation (4)), and scatter index (SI) (Equation (5)). In addition, the statistical graphs were used to inspect the precision of the predicted model, and several tests were considered to inspect the stationarity and normality of the residual data.

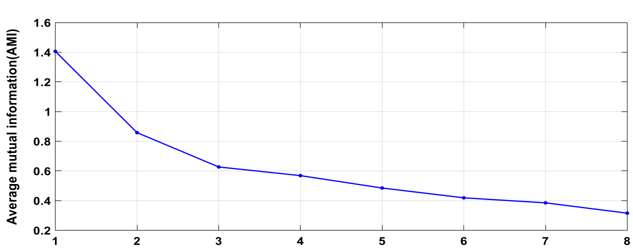

Fig. 3. Average mutual information (AMI) function of the electrical conductivity data.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(O_i - F_i)^2}{N}} \tag{1}$$

$$\mathrm{MAE} = \frac{\sum_{i=1}^{N}|O_i - F_i|}{N} \tag{2}$$

$$\mathrm{MARE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|O_i - F_i|}{O_i} \tag{3}$$

$$R^2 = \left[\frac{\sum_{i=1}^{N}(O_i - \overline{O}_i)(F_i - \overline{F}_i)}{\sqrt{\sum(O_i - \overline{O}_i)^2\sum(F_i - \overline{F}_i)^2}}\right]^2 \tag{4}$$

$$\mathrm{SI} = \frac{\mathrm{RMSE}}{\overline{O}} \times 100 \tag{5}$$

Where F_i represent simulated WQ parameters, O_i is the measured WQ variables, $\overline{O}_i$ represents the mean of measured WQ variables, $\overline{F}_i$ is the mean of simulated WQ variables, and N is the data length. Model performance is considered to be good when $R^2 > 0.85$ [64]. The best model in which the MAE, MARE, and RMSE metrics are all close to zero [29]. Besides, when SI is less than 10 %, the model's accuracy is excellent, between 10 and 20 % is good, between (20–30)% is suitable, and above 30 % is poor [65]. Additionally, graphical plots (i.e., Taylor diagram and box plot) are utilised to estimate the efficiency of the suggested strategy.

4. Results

The results from the implementation of our module are summarised in this section, along with a detailed analysis of those results. Our module's implementation led to significant results demonstrating its usefulness and potential impact. The three subsections are described below.

4.1. Improvement model input

According to Tabachnick and Fidell [66], the TDS and EC data were normalised by applying the natural logarithm to minimise the effects of outliers and make the distribution of the time series close to the normal distribution. Then, the remaining outliers (if found) were rescaled. After that, the SSA technique was used to denoise the time series. Fig. 2 shows the decomposition time series for TDS and EC parameters: the top row (normalised and cleaned data), the second row (the modified time series), and the third and fourth rows (two noise components).

The quality of the raw data was improved via data pre-processing techniques and raised the correlation coefficient (R) for Lag1 between target and model input factors for various lags of monthly WQ data (TDS and EC) (such as the R of raw TDS data of Lag1 improved (from 0.788 to 0.969). The R values for the first three lags of denoise TDS are 0.969,

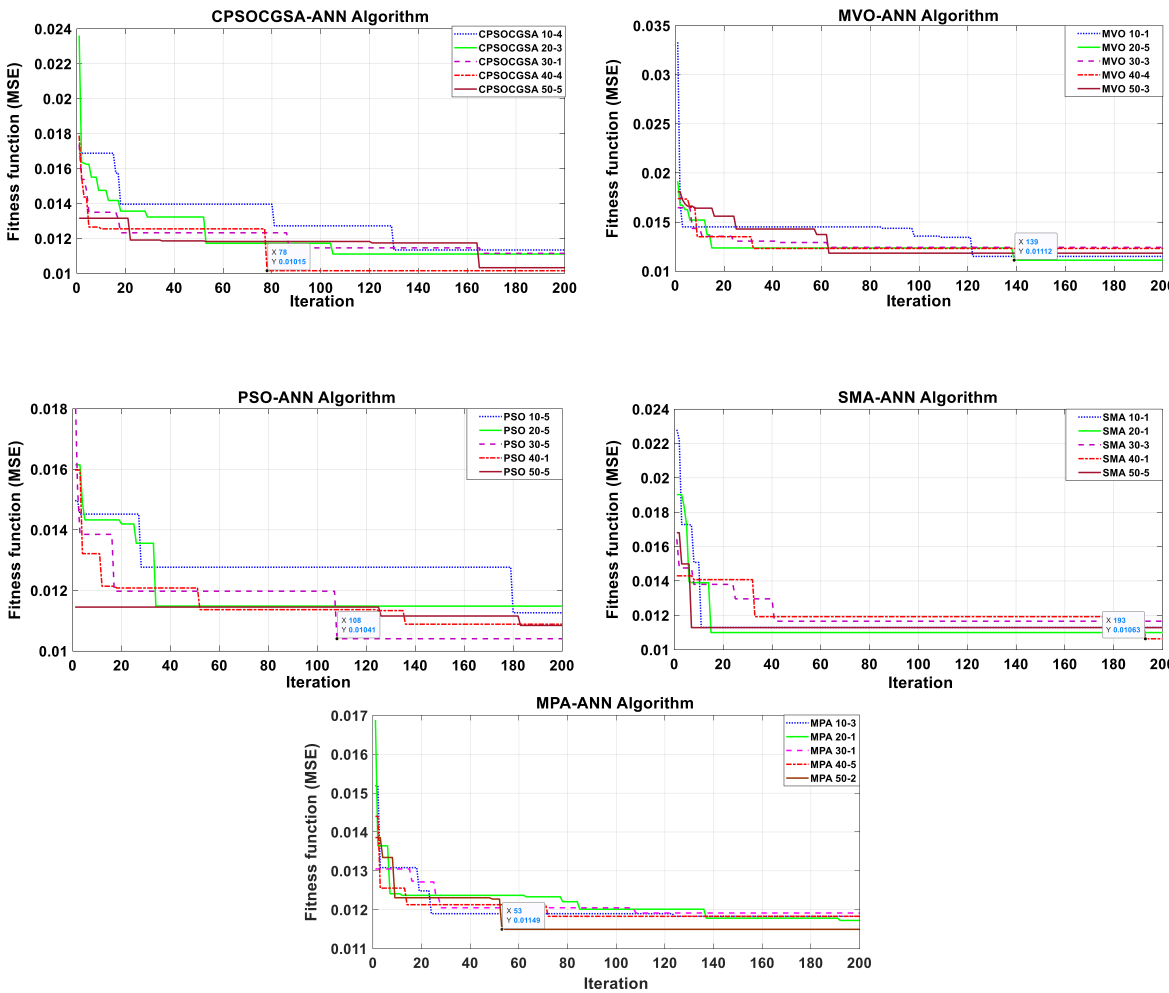

Fig. 4. The best swarm for CPSOCGSA, MVO, PSO, and SMA techniques in modelling TDS parameter.

0.888, and 0.783. Moreover, for EC, the R values for the first three lags are 0.976, 0.916, and 0.839. This study's outcomes of improving the quality of raw data by preprocessing technique agree with previous research that conducted by Pham et al. [67] and Apaydin et al. [68].

The MI approach was also employed to determine the optimum model input scenarios for the TDS and EC forecasting techniques. The time lag is nominated as an initial minimum of average mutual information (AMI), as described in the literature [69]. Therefore, three monthly water salinity data lags were utilised to simulate future water salinity based on the AMI figure, as demonstrated in Fig. 3 for the EC parameter.

Also, choosing the optimal sample size for evolving a suitable model according to Tabachnick and Fidell [66] suggested utilising a sample size that is based on the predictors' number, as revealed in Equation (6):

$$N \geq 50+8\ m \tag{6}$$

Where N represents the sample size, and m is the number of predictor factors, in this paper, N = 117, which is more than the wanted (i.e., 74).

4.2. Application hybrid Algorithms-ANN models

CPSOCGSA-ANN, MVO-ANN, SMA, MPA, and PSO are the five hybrid methods used to determine the optimal ANN hyperparameters. This work investigated swarm sizes ranging from 10 to 50 by hybridising several metaheuristic algorithms with the ANN model. To get the best possible fitness function (MSE) (for example, Fig. S1 for simulating TDS, CPSOCGSA-ANN technique), each algorithm's swarm was performed five times.

As can be seen in Figs. 4 and 5, the optimal swarm for the TDS and EC models was selected and compared to other swarms for the same algorithm. Fig. 4 shows that the CPSOCGSA-ANN, MVO-ANN, PSO-ANN, SMA-ANN, and MPA-ANN algorithms each have optimal swarm sizes of 40-4, 20-5, 30-5, 40-1, and 50-2, respectively, for the TDS model. That means, for example, the best solution for CPSOCGSA-ANN algorithm is swarm 40, the fourth trial.

Fig. 5 also shows that the CPSOCGSA-ANN, MVO-ANN, PSO-ANN, SMA-ANN, and MPA-ANN methods all have optimal swarm sizes of 40-2, 20-1, 50-4, 10-5, and 20-3, respectively, for each EC model. Swarm40, the second trial, is the optimal solution for the CPSOCGSA-ANN algorithm, for instance.

The ideal hyperparameters of ANN obtained from the four hybrid models based on the best swarm for TDS and EC models are listed in Table 1.

4.3. Performance evaluation

Five ANN models were constructed using the hyperparameter values listed in Table 1. Every ANN technique was executed many times to locate the optimal network that delivers precise results. In addition, multiple statistical standards were calculated to inspect and compare the performance of the configured techniques. Table 2 displays the R^2, MAE, RMSE, and MARE of all techniques. Based on Dawson et al. [64], the findings of approaches demonstrated a good level of simulation of both TDS and EC with R^2 of more than 0.85, which means good outcomes

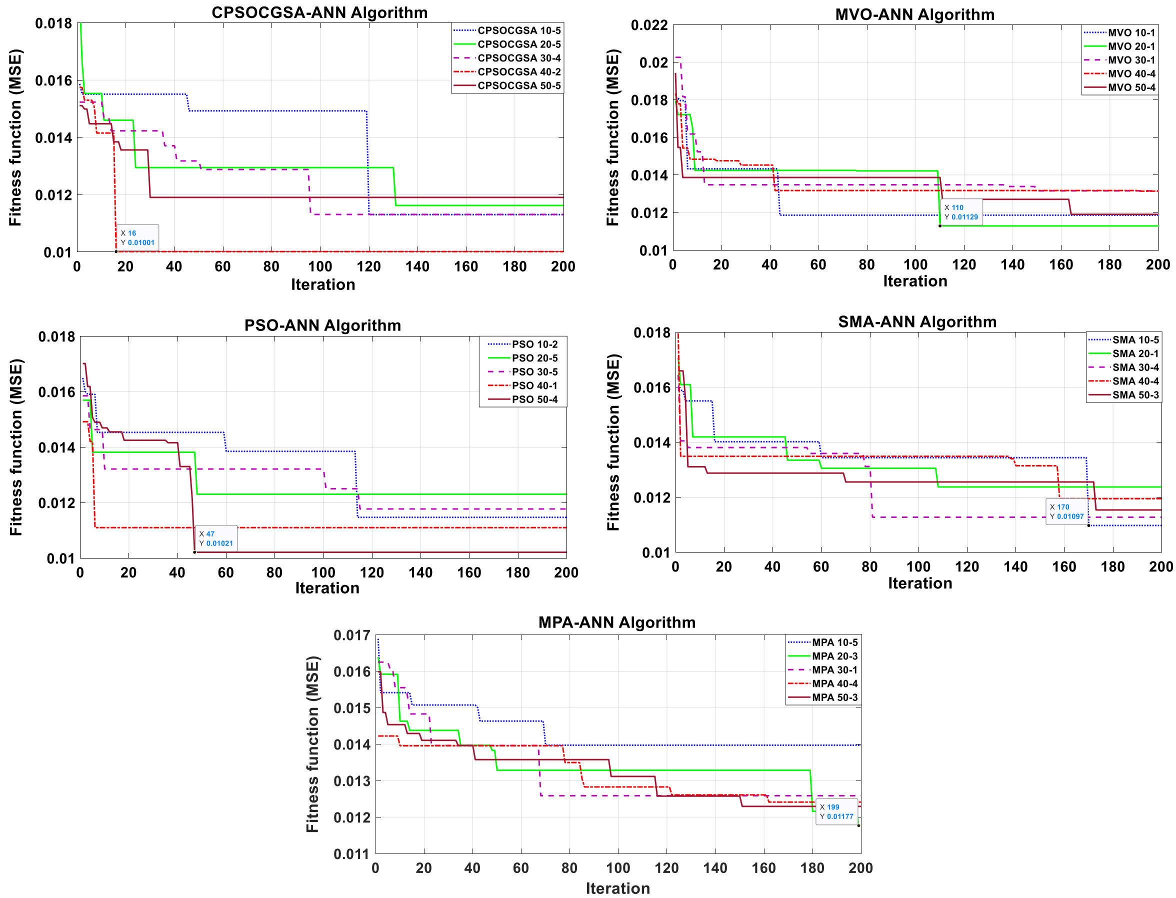

Fig. 5. The best swarm for CPSOCGSA, MVO, PSO, and SMA techniques in modelling EC parameters.

Table 1
Results of hyperparameters for all meta-heuristic techniques.

WQ Parameter	ANN Hyperparameter	CPSOCGSA-ANN	SMA-ANN	MVO-ANN	PSO-ANN	MPA-ANN
TDS	Lr	0.3289	0.2982	0.5143	0.3496	0.5785
	N1	8	3	5	3	17
	N2	2	16	5	11	1
EC	Lr	0.1783	0.0615	0.7618	0.4090	0.4414
	N1	6	12	3	6	1
	N2	7	3	14	4	15

Lr: learning rate, N1 and N2: number of neurons hidden for the 1st and 2nd layers, respectively.

Table 2
Performance assessment of suggested models for validation data phase.

WQ Parameter	Hybrid Models	R^2	MAE (mg/l)	RMSE (mg/l)	MARE
TDS	CPSOCGSA-ANN	0.98	0.0189	0.0248	0.0029
	PSO-ANN	0.81	0.0495	0.0739	0.0078
	MVO-ANN	0.89	0.0644	0.0855	0.0101
	SMA-ANN	0.86	0.0716	0.0969	0.0112
	MPA-ANN	0.96	0.0471	0.0587	0.0074
Parameter	Models	R^2	MAE (µ mhos/cm)	RMSE (µ mhos/cm)	MARE
EC	CPSOCGSA-ANN	0.96	0.0302	0.0386	0.0044
	PSO-ANN	0.94	0.0684	0.0905	0.0100
	MVO-ANN	0.94	0.0689	0.0867	0.0101
	SMA-ANN	0.93	0.0646	0.0863	0.0095
	MPA-ANN	0.94	0.0652	0.0879	0.0096

except for the PSO-ANN technique in modelling TDS data with R^2 less than 0.85. For the rest of the indices, including MAE, RMSE, and MARE, the CPSOCGSA-aANN algorithm offers the lowest values for TDS and EC models. However, it is worth noting that the new proposed model (CPSOCGSA-ANN) exhibited the best forecast performance among all other suggested models. This may be because the GSA optimisation of the PSO algorithm assisted PSO in identifying the best hyperparameters of the ANN model.

In addition, Fig. 6 (Taylor diagram) shows the performance of all hybrid techniques for TDS and EC at the validation phase. This figure summarises the degree of agreement between the patterns of observed and forecast data, accounting for correlation coefficient (R), standard deviation (SD), and root-mean-squared error (RMSD). On the X-axis of the Taylor diagram, the measured WQ (Reference) indicates that if a model is near the observed node, it is regarded to be superior. As a result, it is possible to compare the relative performances of several techniques. According to the diagram, the CPSOCGSA-ANN model produced best performed with high R, and low SD, RMS when compared to the PSO-ANN, MVO-ANN, MPA-ANN, and SMA-ANN models at TDS and EC

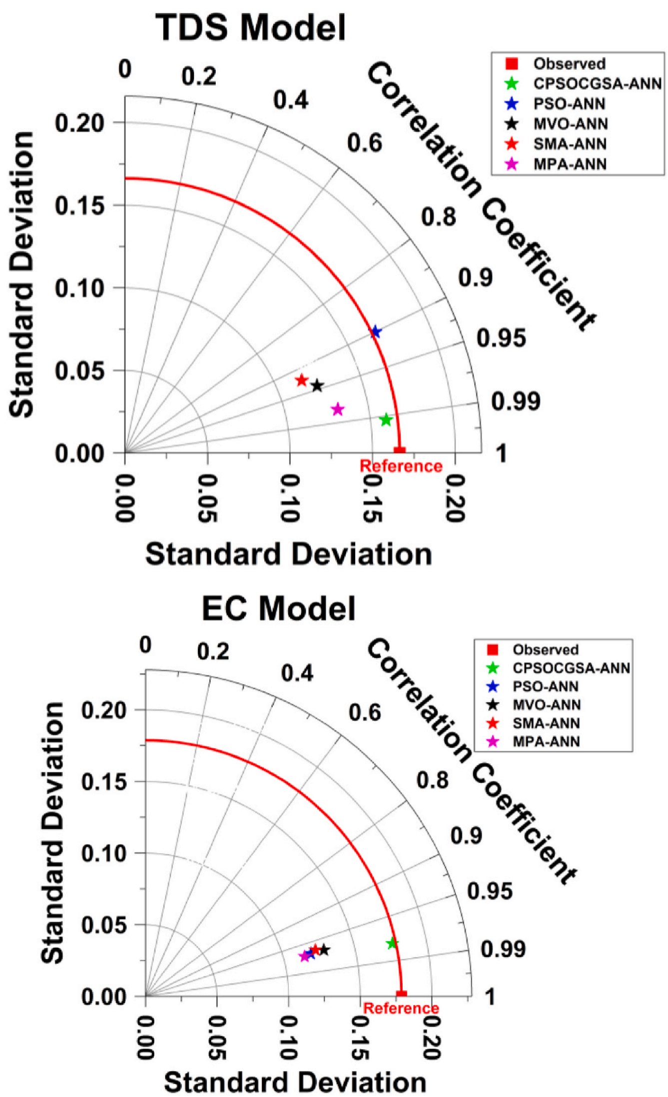

Fig. 6. Taylor diagram to compare the performance of the suggested hybrid models.

forecasting. For EC model, it can be seen that the performance of MPA-ANN, PSO-ANN, SMA-ANN, and MVO-ANN are nearly equal compared with TDS model, which may relate to the patterns of data and their nonlinear relationship.

Also, the Box–Whisker plot was used to assess the models' performance. Fig. 7 displays the box plots of the measured and predicted data for TDS and EC. It can be observed when modelling TDS that all models have median and upper borders similar to measured values but different in the lower borders and extremes. However, the CPSOCGSA-ANN technique is close to observing, revealing the highest forecast model, and the PSO-ANN technique is the lowest forecast model. Besides that, the medians for all suggested models were close to the observed EC values for modelling EC. While the extreme values for the CPSOCGSA-ANN model and upper border were closest to the observed EC values, followed by the MPA-ANN, MVO-ANN, PSO-ANN, and SMA-ANN models. It means that the CPSOCGSA-ANN model is superior to other strategies.

5. Discussion

Providing an accurate water salinity methodology is still one of the important research topics that has attracted many researchers. In Iraq, where the correlation between salinity levl and the water quality index is strong, the Ministry of Water Resourse's decision-makers need an accurate measure for water quality to intervene in a suitable form. Therefore, this study presents a new methodology of five scenarios for a univariate prediction for each of TDS and ED. These parameters are clear indicators of the salinity level. The methodology has taken into consideration the necessity of pretreatment of time series. This was done through the reduction of structureless noise via the application of SSA, where the components of the lowest variance were neglected. In addition, the correlation among the lags of the univariate series (i.e. of TDS and ED) has been increased. For instance, the correlation coefficient (R) for the first three lags of denoise TDS are 0.969, 0.888, and 0.783, while for EC, they are 0.976, 0.916, and 0.839. After achieving the necessary improvement for the time series through the pretreatment stage, another enhancement for the prediction process was conducted through a combination of the conventional ANN with various metaheuristic techniques (single-based and hybrid-based). The purpose behind this combination is to optimise the ANN hyperparameter, which no doubt reflects positively on the hybridised model performance. Among the five model scenarios, the CPSCOGSA-ANN method (hybrid-based) was shown to be the most accurate in predicting the time series with minimum error compared to other model scenarios. This significant performance comes from the combined capabilities of GSA's exploration and the PSO's exploitation. In the CPSO, the tuning of coefficients will control the movement of its particles, while the CGSA provides more diversity to avoid the local optima. Supporting earlier studies [70,71], the present study's results show that hybrid-based metaheuristic algorithms outperform single-based algorithms. This outperforming was proven clearly in several statistical metrics. As for the SI, the hybridised model shows 0.004 for the TDS prediction and 0.006 for the ED prediction. In addition, the R2, MSE, MARE, and MAE have supported this superiority. These results provide credence to the theory put forward in the literature [72,73], which states that hybrid-based algorithms can avoid local minima while achieving greater precision, stability, and reliability in solving real-world problems.

Further studies could be considered to examine the performance of different hybrid-based metaheuristic algorithms integrated with other machine learning techniques, such as random forest and support vector regression. These new hybrid models can be combined with various data

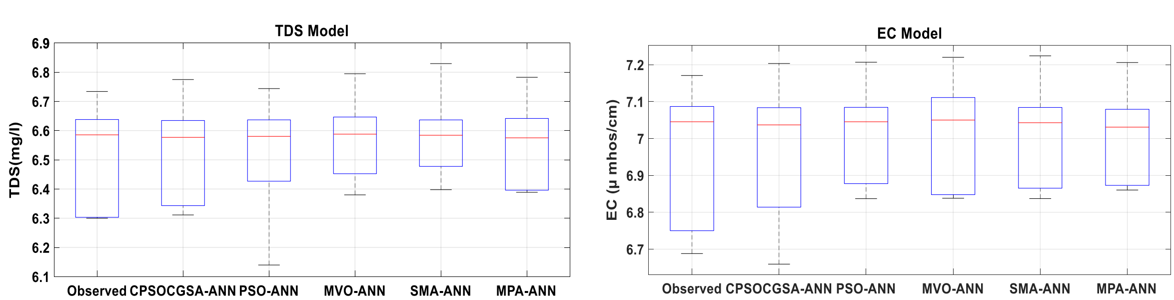

Fig. 7. Box plots of predictions were used during the validation stage to assess the models for TDS and EC.

preprocessing techniques to create different HOPH models.

6. Practical implications

The reliability of water quality models is essential for effective water resource management. TDS and EC are necessary standards used to manage and control the Euphrates River. This means that changes in water quality can be predicted before they occur; Thus, poor water quality can be avoided. This is particularly the case in Babil Governorate, where water is called blue gold due to its use for irrigation, domestic use, as well as industrial use. Thus, Effective forecasts help preserve the environment and protect public health by reducing the negative effects of high levels of TDS and EC. Efficient water quality management can be enhanced through proper prediction of TDS and EC. Thus, knowing the time and place where the water quality is more likely to decline, the government can direct their efforts more efficiently. Accurate predictions assist sustainable management by showing the long-run tendencies and possible developments in the future. This vision enables the stakeholders to design measures that can foster water use for agricultural, industrial, and domestic purposes without jeopardising the river's health.

7. Conclusion

The importance of properly portraying salinity behaviour in water quality research has sparked a growing interest in the modelling requirements. The current study investigates the capacity of hybridisation of ANN with metaheuristic techniques for TDS and EC time series prediction in Iraq. A novel method for estimating univariate salinity time series was proposed in this study, which contains data preprocessing procedures and an ANN method integrated with a CPSOCGSA (hybrid-based) algorithm. The performance of the CPSOCGSA algorithm was compared with four single-based algorithms (PSO, MVO, SMA, and MPA).

Because the MHAs applied in the optimisation process follow various strategies, the hybridisation process results in different values for the hyperparameters, which in turn yield different model scenarios. In general, all the predicting techniques performed well, and there are a few potential reasons for that. One probable reason for this edge could be that the data pretreatment method improved the data quality, leading to more accurate predictions. Another possible reason is that each algorithm's swarm was run five times to find the optimal solution, which resulted in a broader range of predictions and less uncertainty.

Overall, the results for TDS and EC models showed that the pretreatment processing method had improved the time series quality through data denoising using the SSA technique, where the structureless noise components were neglected. In addition, the outperformance of CPSOCGSA-ANN compared to other techniques. In terms of several statistical metrics, CPSOCGSA-ANN (hybrid-based) performance was found to be better than other single-based models (i.e. SMA-ANN, PSO-ANN, MPA-ANN, and MVO-ANN), where the R2 was the highest (0.98 for TDS and 0.96 for EC). In addition, the MAE, RMSE, and MARE were the minimum, as shown in Table 2.

The province of Babylon relies economically on agriculture but is already experiencing water salinity stress, and this is the first time that the salinity of the Al-Euphrates River has been predicted at Babylon Governorate using data with multiple time lags. This study may help inform future freshwater quality management decisions. For future research direction, the present study forms a stepping stone towards more research, including exploring more hybrid forecast methodologies over various time scales. Accordingly, these outcomes support the hypothesis that a hybrid-based metaheuristic algorithm performs better than a single-based one. Using multi-criteria decision-making in order to select the best algorithms for this data from multiple perspectives.

CRediT authorship contribution statement

Zahraa S. Khudhair: Writing – original draft. **Salah L. Zubaidi:** Writing – review & editing. **Anmar Dulaimi:** Methodology. **Hussein Al-Bugharbee:** Investigation. **Yousif Raad Muhsen:** Formal analysis. **Ramadhansyah Putra Jaya:** Supervision. **Hussein Mohammed Ridha:** Software. **Syed Fawad Raza:** Conceptualization. **Saleem Ethaib:** Project administration.

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 researchers are thankful to the Ministry of Environment for providing the data.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rineng.2024.102541.

References

[1] Y. Wang, Y. Yuan, Y. Pan, Z. Fan, Modeling daily and monthly water quality indicators in a canal using a hybrid wavelet-based support vector regression structure, Water 12 (2020) 1476, https://doi.org/10.3390/w12051476.
[2] S. Ethaib, S.L. Zubaidi, N. Al-Ansari, S.L. Fegade, Evaluation water scarcity based on GIS estimation and climate-change effects: a case study of Thi-Qar Governorate, Iraq, Cogent Engineering 9 (2022) 2075301, https://doi.org/10.1080/23311916.2022.2075301.
[3] M.T.H. Van Vliet, W.H.P. Franssen, J.R. Yearsley, F. Ludwig, I. Haddeland, D. P. Lettenmaier, P. Kabat, Global river discharge and water temperature under climate change, Global Environ. Change 23 (2013) 450–464, https://doi.org/10.1016/j.gloenvcha.2012.11.002.
[4] S. Bi, L. Wang, Y. Li, Z. Zhang, Z. Wang, X. Ding, J. Zhou, K. Hatfield, A comprehensive method for water environment assessment considering Trends of water quality, Adv. Civ. Eng. 2021 (2021) 1–8, https://doi.org/10.1155/2021/5548113.
[5] M. Ali Khan, M. Izhar Shah, M. Faisal Javed, M. Ijaz Khan, S. Rasheed, M.A. El-Shorbagy, E. Roshdy El-Zahar, M.Y. Malik, Application of random forest for modelling of surface water salinity, Ain Shams Eng. J. 13 (2022) 101635, https://doi.org/10.1016/j.asej.2021.11.004.
[6] J.M. Hunter, H.R. Maier, M.S. Gibbs, E.R. Foale, N.A. Grosvenor, N.P. Harders, T. C. Kikuchi-Miller, Framework for developing hybrid process-driven, artificial neural network and regression models for salinity prediction in river systems, Hydrol. Earth Syst. Sci. 22 (2018) 2987–3006, https://doi.org/10.5194/hess-22-2987-2018.
[7] A.H. Kamel, M.A. Sulaiman, A.S. Mustaffa, Study of the effects of water level Depression in Euphrates River on the water quality, Journal of Civil Engineering and Architecture 7 (2013) 238–247.
[8] K.A. Rahi, T. Halihan, Changes in the salinity of the Euphrates River system in Iraq, Reg. Environ. Change 10 (2009) 27–35, https://doi.org/10.1007/s10113-009-0083-y.
[9] M. Al-Mukhtar, F. Al-Yaseen, Modeling water quality parameters using data-driven models, a case study abu-ziriq marsh in south of Iraq, Hydrology 6 (2019) 24, https://doi.org/10.3390/hydrology6010024.
[10] J. Ren, Q. Zhu, C. Wang, Y. Huang, Edge computing for water quality monitoring systems, Mobile Inf. Syst. 2022 (2022) 1–7, https://doi.org/10.1155/2022/5056606.
[11] K. Han, Y. Wang, A review of artificial neural network techniques for environmental issues prediction, J. Therm. Anal. Calorim. 145 (2021) 2191–2207, https://doi.org/10.1007/s10973-021-10748-9.
[12] A. Najah Ahmed, F. Binti Othman, H. Abdulmohsin Afan, R. Khaleel Ibrahim, C. Ming Fai, M. Shabbir Hossain, M. Ehteram, A. Elshafie, Machine learning methods for better water quality prediction, J. Hydrol. 578 (2019) 124084, https://doi.org/10.1016/j.jhydrol.2019.124084.
[13] A.A.M. Ahmed, S.M.A. Shah, Application of adaptive neuro-fuzzy inference system (ANFIS) to estimate the biochemical oxygen demand (BOD) of Surma River,

Journal of King Saud University - Engineering Sciences 29 (2017) 237–243, https://doi.org/10.1016/j.jksues.2015.02.001.
[14] R. Dehghani, H. Torabi Poudeh, Z. Izadi, Dissolved oxygen concentration predictions for running waters with using hybrid machine learning techniques, Modeling Earth Systems and Environment 8 (2021) 2599–2613, https://doi.org/10.1007/s40808-021-01253-x.
[15] A. Alqahtani, M.I. Shah, A. Aldrees, M.F. Javed, Comparative assessment of individual and ensemble machine learning models for efficient analysis of river water quality, Sustainability 14 (2022), https://doi.org/10.3390/su14031183.
[16] M. Azrour, J. Mabrouki, G. Fattah, A. Guezzaz, F. Aziz, Machine learning algorithms for efficient water quality prediction, Modeling Earth Systems and Environment 8 (2022) 2793–2801, https://doi.org/10.1007/s40808-021-01253-x.
[17] J. Sha, X. Li, M. Zhang, Z.-L. Wang, Comparison of forecasting models for real-time monitoring of water quality parameters based on hybrid deep learning neural networks, Water 13 (2021) 1547, https://doi.org/10.3390/w13111547.
[18] Wang, J.-d.; Li, P.-y.; Zhang, Y.-m.; Qi, W.-g. River water turbidity forecasting based on phase space reconstruction and support vector regression. In Proceedings of 2010 International Conference on Intelligent Computation Technology and Automation; pp. 217-220.
[19] H. Choi, S.-I. Suh, S.-H. Kim, E.J. Han, S.J. Ki, Assessing the performance of deep learning algorithms for short-term surface water quality prediction, Sustainability 13 (2021) 10690, https://doi.org/10.3390/su131910690.
[20] M. Monteiro, M. Costa, A time series model comparison for monitoring and forecasting water quality variables, Hydrology 5 (2018), https://doi.org/10.3390/hydrology5030037.
[21] Y. Chen, L. Song, Y. Liu, L. Yang, D. Li, A review of the artificial neural network models for water quality prediction, Appl. Sci. 10 (2020) 5776, https://doi.org/10.3390/app10175776.
[22] H.M. Mustafa, A. Mustapha, G. Hayder, A. Salisu, Applications of IoT and artificial intelligence in water quality monitoring and prediction: a review, in: Proceedings of 2021 6th International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 20-22 January 2021, pp. 968–975.
[23] O.T. Faloye, A.E. Ajayi, Y. Ajiboye, M.O. Alatise, B.S. Ewulo, S.S. Adeosun, T. Babalola, R. Horn, Unsaturated hydraulic conductivity prediction using artificial intelligence and multiple linear regression models in biochar amended sandy clay loam soil, J. Soil Sci. Plant Nutr. 22 (2022) 1589–1603, https://doi.org/10.1007/s42729-021-00756-x.
[24] H. Apaydin, M. Taghi Sattari, K. Falsafian, R. Prasad, Artificial intelligence modelling integrated with Singular Spectral analysis and Seasonal-Trend decomposition using Loess approaches for streamflow predictions, J. Hydrol. 600 (2021) 126506, https://doi.org/10.1016/j.jhydrol.2021.126506.
[25] S.L. Zubaidi, N.S.S. Al-Bdairi, S. Ortega-Martorell, H.M. Ridha, N. Al-Ansari, H. Al-Bugharbee, K. Hashim, S.K. Gharghan, Assessing the benefits of nature-inspired algorithms for the parameterization of ANN in the prediction of water demand, J. Water Resour. Plann. Manag. 149 (2023) 1–10, https://doi.org/10.1061/(asce)wr.1943-5452.0001602.
[26] J. Zhou, Y. Wang, F. Xiao, Y. Wang, L. Sun, Water quality prediction method based on IGRA and LSTM, Water 10 (2018) 1148, https://doi.org/10.3390/w10091148.
[27] B. Raheli, M.T. Aalami, A. El-Shafie, M.A. Ghorbani, R.C. Deo, Uncertainty assessment of the multilayer perceptron (MLP) neural network model with implementation of the novel hybrid MLP-FFA method for prediction of biochemical oxygen demand and dissolved oxygen: a case study of Langat River, Environ. Earth Sci. 76 (2017) 503, https://doi.org/10.1007/s12665-017-6842-z.
[28] Y. Kang, J. Song, Z. Lin, L. Huang, X. Zhai, H. Feng, Water quality prediction based on SSA-MIC-SMBO-ESN, Comput. Intell. Neurosci. 2022 (2022) 1264385, https://doi.org/10.1155/2022/1264385.
[29] E. Eze, S. Halse, T. Ajmal, Developing a novel water quality prediction model for a South African aquaculture farm, Water 13 (2021) 1782, https://doi.org/10.3390/w13131782.
[30] H.-G. Han, J.-F. Qiao, Q.-L. Chen, Model predictive control of dissolved oxygen concentration based on a self-organising RBF neural network, Control Eng. Pract. 20 (2012) 465–476, https://doi.org/10.1016/j.conengprac.2012.01.001.
[31] H.R. Maier, A. Jain, G.C. Dandy, K.P. Sudheer, 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. Software 25 (2010) 891–909, https://doi.org/10.1016/j.envsoft.2010.02.003.
[32] M.I. Shah, M.F. Javed, A. Alqahtani, A. Aldrees, Environmental assessment based surface water quality prediction using hyper-parameter optimised machine learning models based on consistent big data, Process Saf. Environ. Protect. 151 (2021) 324–340, https://doi.org/10.1016/j.psep.2021.05.026.
[33] A. Azad, H. Karami, S. Farzin, S.-F. Mousavi, O. Kisi, Modeling river water quality parameters using modified adaptive neuro fuzzy inference system, Water Sci. Eng. 12 (2019) 45–54, https://doi.org/10.1016/j.wse.2018.11.001.
[34] S. Anupam, P. Pani, Flood forecasting using a hybrid extreme learning machine-particle swarm optimisation algorithm (ELM-PSO) model, Modeling Earth Systems and Environment 6 (2019) 341–347, https://doi.org/10.1007/s40808-019-00682-z.
[35] B. Aghel, A. Rezaei, M. Mohadesi, Modeling and prediction of water quality parameters using a hybrid particle swarm optimisation–neural fuzzy approach, Int. J. Environ. Sci. Technol. 16 (2018) 4823–4832, https://doi.org/10.1007/s13762-018-1896-3.
[36] Y.-H. Lin, M.-T. Lee, Y.-H. Hung, A thermal management control using particle swarm optimisation for hybrid electric energy system of electric vehicles, Results in Engineering 21 (2024), https://doi.org/10.1016/j.rineng.2023.101717.
[37] Y. Tian, Y. Chang, Application of the particle swarm optimisation algorithm-back propagation neural network algorithm introducing new parameter terms in the application field of industrial design, Results in Engineering 21 (2024), https://doi.org/10.1016/j.rineng.2023.101728.
[38] S. Li, H. Chen, M. Wang, A.A. Heidari, S. Mirjalili, Slime mould algorithm: a new method for stochastic optimisation, Future Generat. Comput. Syst. 111 (2020) 300–323, https://doi.org/10.1016/j.future.2020.03.055.
[39] E.H. Houssein, M.A. Mahdy, M.J. Blondin, D. Shebl, W.M. Mohamed, Hybrid slime mould algorithm with adaptive guided differential evolution algorithm for combinatorial and global optimisation problems, Expert Syst. Appl. 174 (2021) 114689, https://doi.org/10.1016/j.eswa.2021.114689.
[40] C. Kumar, T.D. Raj, M. Premkumar, T.D. Raj, A new stochastic slime mould optimisation algorithm for the estimation of solar photovoltaic cell parameters, Optik 223 (2020) 165277, https://doi.org/10.1016/j.ijleo.2020.165277.
[41] S. Mirjalili, S.M. Mirjalili, A. Hatamlou, Multi-Verse Optimizer: a nature-inspired algorithm for global optimisation, Neural Comput. Appl. 27 (2015) 495–513, https://doi.org/10.1007/s00521-015-1870-7.
[42] T. Peng, J. Zhou, C. Zhang, W. Fu, Streamflow forecasting using empirical wavelet transform and artificial neural networks, Water 9 (2017) 406, https://doi.org/10.3390/w9060406.
[43] Y. Tikhamarine, A. Malik, K. Pandey, S.S. Sammen, D. Souag-Gamane, S. Heddam, O. Kisi, Monthly evapotranspiration estimation using optimal climatic parameters: efficacy of hybrid support vector regression integrated with whale optimisation algorithm, Environ. Monit. Assess. 192 (2020) 696, https://doi.org/10.1007/s10661-020-08659-7.
[44] A. Faramarzi, M. Heidarinejad, S. Mirjalili, A.H. Gandomi, Marine predators algorithm: a nature-inspired metaheuristic, Expert Syst. Appl. 152 (2020), https://doi.org/10.1016/j.eswa.2020.113377.
[45] S.J. Mohammed, S.L. Zubaidi, N. Al-Ansari, H.M. Ridha, N.S.S. Al-Bdairi, S. O. Degertekin, Hybrid technique to improve the river water level forecasting using artificial neural network-based marine predators algorithm, Adv. Civ. Eng. 2022 (2022) 1–14, https://doi.org/10.1155/2022/6955271.
[46] Z.S. Khudhair, S.L. Zubaidi, S. Ortega-Martorell, N. Al-Ansari, S. Ethaib, K. Hashim, A review of hybrid soft computing and data pre-processing techniques to forecast freshwater quality's parameters: current trends and future directions, Environments 9 (2022) 85, https://doi.org/10.3390/environ-ments9070085, 10.3390/environments9070085.
[47] Z. Hajirahimi, M. Khashei, Hybridization of hybrid structures for time series forecasting: a review, Artif. Intell. Rev. 56 (2023) 1201–1261, https://doi.org/10.1007/s10462-022-10199-0.
[48] A.H. Hashim, O.Z. Jasim, M.M. Salih, The establishing of geospatial database for agricultural lands of islamic WAQF in Iraq: case study Babil province, IOP Conf. Ser. Earth Environ. Sci. 961 (2022), https://doi.org/10.1088/1755-1315/961/1/012025.
[49] H.R. Maier, G. Dandy, Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications, Environ. Model. Software 15 (2000) 101–124, doi:PII: S1364-8152(99)00007- 9.
[50] H. Al-Bugharbee, I. Trendafilova, A fault diagnosis methodology for rolling element bearings based on advanced signal pretreatment and autoregressive modelling, J. Sound Vib. 369 (2016) 246–265, https://doi.org/10.1016/j.jsv.2015.12.052.
[51] N. Golyandina, A. Zhigljavsky, Singular Spectrum Analysis for Time Series; New York, NY, USA, 2013.
[52] J.B. Kinney, G.S. Atwal, Equitability, mutual information, and the maximal information coefficient, Proc. Natl. Acad. Sci. U.S.A. 111 (2014) 3354–3359, https://doi.org/10.1073/pnas.1309933111.
[53] M. Clerc, J.J.I.t.o.E.C. Kennedy, The Particle Swarm-Explosion, Stability, and Convergence in a Multidimensional Complex Space, vol. 6, 2002, pp. 58–73.
[54] Rather, S.A.; Bala, P.S. Hybridization of constriction coefficient-based particle swarm optimisation and chaotic gravitational search algorithm for solving engineering design problems. In Proceedings of International Conference on Advanced Communication and Networking; pp. 95-115.
[55] S.A. Rather, P.S. Bala, Hybridization of constriction coefficient-based particle swarm optimization and chaotic gravitational search algorithm for solving engineering design problems, Applied Soft Computing and Communication Networks (2020) 95–115, https://doi.org/10.1007/978-981-15-3852-0_7pp.
[56] U. Bello, N.A. Amran, M.S. Hazwan Ruslan, E.H. Yáñez, U. Suparmaniam, H. Adamu, S.I. Abba, U.I. Tafida, A.A. Mahmoud, Enhancing oxidative stability of biodiesel using fruit peel waste extracts blend: comparison of predictive modelling via RSM and ANN techniques, Results in Engineering 21 (2024), https://doi.org/10.1016/j.rineng.2024.101853.
[57] M. Bayatvarkeshi, K. Mohammadi, O. Kisi, R. Fasihi, A new wavelet conjunction approach for estimation of relative humidity: wavelet principal component analysis combined with ANN, Neural Comput. Appl. 32 (2018) 4989–5000, https://doi.org/10.1007/s00521-018-3916-0.
[58] A.J. Thomas, M. Petridis, S.D. Walters, S.M. Gheytassi, R.E. Morgan, Two hidden layers are usually better than one, Engineering Applications of Neural Networks, Springer 744 (2017) 279–290.
[59] B. Deng, S.H. Lai, C. Jiang, P. Kumar, A. El-Shafie, R.J. Chin, Advanced water level prediction for a large-scale river–lake system using hybrid soft computing approach: a case study in Dongting Lake, China, Earth Science Informatics 14 (2021) 1987–2001, https://doi.org/10.1007/s12145-021-00665-8.
[60] F. Rastegaripour, M.S. Saboni, S. Shojaei, A. Tavassoli, Simultaneous management of water and wastewater using ant and artificial neural network (ANN) algorithms, Int. J. Environ. Sci. Technol. 16 (2018) 5835–5856, https://doi.org/10.1007/s13762-018-1943-0.
[61] K. Khairudin, A.Z. Ul-Saufie, S.F. Senin, Z. Zainudin, A.M. Rashid, N.F. Abu Bakar, M.Z. Anas Abd Wahid, S.F. Azha, F. Abd-Wahab, L. Wang, et al., Enhancing

riverine load prediction of anthropogenic pollutants: harnessing the potential of feed-forward backpropagation (FFBP) artificial neural network (ANN) models, Results in Engineering 22 (2024), https://doi.org/10.1016/j.rineng.2024.102072.
[62] M. Yilma, Z. Kiflie, A. Windsperger, N. Gessese, Environment. Application of artificial neural network in water quality index prediction: a case study in Little Akaki River, Addis Ababa, Ethiopia, Modeling Earth Systems and Environment 4 (2018) 175–187, https://doi.org/10.1007/s40808-018-0437-x.
[63] H. Tahraoui, A.-E. Belhadj, A.-e. Hamitouche, M. Bouhedda, A. Amrane, Predicting the concentration of sulfate (SO42-) in drinking water using artificial neural networks: a case study: Médéa-Algeria, Desalination Water Treat. 217 (2021) 181–194, https://doi.org/10.5004/dwt.2021.26813.
[64] C.W. Dawson, R.J. Abrahart, L.M. See, HydroTest: a web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts, Environ. Model. Software 22 (2007) 1034–1052, https://doi.org/10.1016/j.envsoft.2006.06.008.
[65] M.-F. Li, X.-P. Tang, W. Wu, H.-B. Liu, General models for estimating daily global solar radiation for different solar radiation zones in mainland China, Energy Convers. Manag. 70 (2013) 139–148, https://doi.org/10.1016/j.enconman.2013.03.004.
[66] B.G. Tabachnick, L.S. Fidell, Using Multivariate Statistics; Boston, MA,USA, 2013.
[67] Q.B. Pham, T.-C. Yang, C.-M. Kuo, H.-W. Tseng, P.-S. Yu, Coupling singular spectrum analysis with least square support vector machine to improve accuracy of SPI drought forecasting, Water Resour. Manag. 35 (2021) 847–868, https://doi.org/10.1007/s11269-020-02746-7.
[68] H. Apaydin, M.T. Sattari, K. Falsafian, R. Prasad, Artificial intelligence modelling integrated with Singular Spectral analysis and Seasonal-Trend decomposition using Loess approaches for streamflow predictions, J. Hydrol. 600 (2021) 1–16, https://doi.org/10.1016/j.jhydrol.2021.126506.
[69] C. Aldrich, L. Auret, Unsupervised Process Monitoring and Fault Diagnosis with Machine Learning Methods, vol. 16, Springer, Berlin, Germany, 2013.
[70] H.E. Khairan, S.L. Zubaidi, Y.R. Muhsen, N. Al-Ansari, Parameter optimisation-based hybrid reference evapotranspiration prediction models: a systematic review of current implementations and future research directions, Atmosphere 14 (2022), https://doi.org/10.3390/atmos14010077.
[71] B. Abdul Kareem, L. Zubaidi, S, N. Al-Ansari, Y. Raad Muhsen, Review of recent trends in the hybridisation of preprocessing-based and parameter optimisation-based hybrid models to forecast univariate streamflow, Comput. Model. Eng. Sci. 138 (2024) 1–41, https://doi.org/10.32604/cmes.2023.027954.
[72] E.-S.M. El-kenawy, B. Zerouali, N. Bailek, K. Bouchouich, M.A. Hassan, J. Almorox, A. Kuriqi, M. Eid, A. Ibrahim, Improved weighted ensemble learning for predicting the daily reference evapotranspiration under the semi-arid climate conditions, Environ. Sci. Pollut. Control Ser. 29 (2022) 81279–81299, https://doi.org/10.1007/s11356-022-21410-8.
[73] M. Zeinolabedini Rezaabad, S. Ghazanfari, M. Salajegheh, ANFIS modeling with ICA, BBO, TLBO, and IWO optimization algorithms and sensitivity analysis for predicting daily reference evapotranspiration, J. Hydrol. Eng. 25 (2020), https://doi.org/10.1061/(asce)he.1943-5584.0001963.