Hindawi
Advances in Civil Engineering
Volume 2022, Article ID 6955271, 14 pages
https://doi.org/10.1155/2022/6955271

Research Article

Hybrid Technique to Improve the River Water Level Forecasting Using Artificial Neural Network-Based Marine Predators Algorithm

Sarah J. Mohammed,[1] Salah L. Zubaidi ,[1] Nadhir Al-Ansari ,[2] Hussein Mohammed Ridha ,[3] and Nabeel Saleem Saad Al-Bdairi [1]

[1]*Department of Civil Engineering, Wasit University, Wasit 52001, Iraq*
[2]*Department of Civil Environmental and Natural Resources Engineering, Lulea University of Technology, Lulea 971 87, Sweden*
[3]*Department of Electrical and Electronics Engineering, Faculty of Engineering, Universiti Putra Malaysia, Serdang 43400, Malaysia*

Correspondence should be addressed to Nadhir Al-Ansari; nadhir.alansari@ltu.se

Received 1 July 2022; Revised 22 August 2022; Accepted 8 September 2022; Published 10 October 2022

Academic Editor: Sadik Ozgur Degertekin

Water level (WL) forecasting has become a difficult undertaking due to spatiotemporal fluctuations in climatic factors and complex physical processes. This paper proposes a novel hybrid machine learning model based on an artificial neural network (ANN) and the Marine Predators algorithm (MPA) for modeling monthly water levels of the Tigris River in Al-Kut, Iraq. Data preprocessing techniques are employed to enhance data quality and determine the optimal input model. Historical data for water level and climatic factors data are utilized from 2011 to 2020 to build and assess the model. MPA-ANN algorithm's performance is compared with recent constriction coefficient-based particle swarm optimization and chaotic gravitational search algorithm (CPSOCGSA-ANN) and slime mold algorithm (SMA-ANN) to reduce uncertainty and raise the prediction range. The finding demonstrated that singular spectrum analysis is a highly effective method to denoise time series. MPA-ANN outperformed CPSOCGSA-ANN and SMA-ANN algorithms based on different statistical criteria. The suggested novel methodology offers good results with scatter index (SI) = 0.0009 and coefficient of determination ($R^2 = 0.98$).

1. Introduction

Providing municipal water is essential for achieving a sustainable environment in modern cities, especially under the effect of global warming and socioeconomic factors. For the reasons stated, freshwater scarcity is a typical challenge for policymakers [1]. According to the World Economic Forum, water scarcity is one of the most important international problems due to limited freshwater availability (approximately 0.014 percent of Earth's total amount of water). Climate change, water pollution, and ineffective management of freshwater resources are also vital contributors to water scarcity [2]. There are various methods for measuring water depth in rivers and lakes. These methods can simultaneously record depth at several hundred points along the bottom, such as sonar and echo sounder [3, 4]. Thus, water level prediction accuracy is a critical requirement for scientific decision-making and planning. Water level prediction aims to develop a reliable prediction model and uncover the changing laws of rivers and lakes [5, 6].

Iraq is an Arab Country in the Middle East that is one of the most sensitive to climate change due to its location in an arid to semiarid region. This area is generally experiencing a water shortage, which is expected to worsen due to various factors such as climate change, the oil industry, urbanization, and rapid population growth [7]. The Euphrates and Tigris rivers are Iraq's principal freshwater supplies. Between 2009 and 2014, these rivers had substantial water shortages, which are predicted to worsen due to climate change and the policy of water source countries (i.e., Turkey and Iran) that were constructing and running new dams on the route Tigris and Euphrates Rivers [8].

In the past decade, various data-driven models have been developed and implemented for water level prediction in watersheds, such as adaptive neurofuzzy inference systems (ANFIS) [9, 10], support vector regression (SVR) [11], artificial neural networks (ANN) [11–13], random forest [14]. From a theoretical standpoint, one single forecasting method is inappropriate for modeling complicated and unknown latent patterns in world time series. As a result, how to accurately estimate time series remains a difficult subject on which many scholars have focused. Significant efforts have been made to solve these difficulties and achieve a high level of accuracy, resulting in remarkable advancements in forecasting methods. Hybridization has emerged as a promising technique for overcoming the obvious drawbacks of standalone methods while also improving predicting accuracy [15] and the hybrid techniques' performance better than single models, such as

(1) Azad et al. [16] employ a hybrid technique that includes the Box–Jenkins autoregressive seasonal autoregressive integrated moving average (SARIMA) and ANN to forecast reservoir water level (RWL) in the Red Hills Reservoir, Tamil Nadu, India. A new SARIMA–ANN model was created to solve this limitation of the SARIMA model and compared with ANN and SARIMA models. The results revealed that the SARIMA–ANN model outperformed the other models.

(2) Mohammadi et al. [10] enhanced the capacity of the support vector regression (SVR) model by utilizing the grey wolf algorithm (GWO) to predict the lake water level. In addition, three data pretreatments methods were used to pick the optimal scenario of predictors: random forest, relief algorithm, and principal component analysis. The results reveal that the random forest method provides the optimum model input scenario with four lags. The hybrid model SVR-GWO simulates water levels better than the standalone SVR based on different statistical criteria.

(3) Lineros et al. [17] used a multiobjective genetic algorithm (MOGA) framework to integrate the artificial neural network (ANN) technique for river water level prediction. The design process is a semiautomated method for splitting data into datasets and finding a near-optimal model with a suitable topology and inputs. The results show that the suggested framework can create low-complex models with great performance on unseen data, with an RMSE of $2.5 * 10^{-3}$, which compares favorably to results obtained using an alternative design.

(4) Nguyen et al. [18] developed hybrid models (GA-XGBoost and DE-XGBoost) for water level prediction that combine two evolutionary models, a genetic algorithm (GA) and a differential evolution (DE) algorithm. The GA-AND DE-XGBoost models' results were compared with classification and registration tree (CART) and random forest (RF) models. The results indicated that two hybrid models, GA-XGBoost and DE-XGBoost, outperformed RF and CART in multistep-ahead water level prediction. This research shows that hybrid XGBoost models for hourly water level prediction may be superior to many existing models despite some shortcomings in terms of long-step-ahead prediction and model complexity.

(5) Ehteram et al. [19] utilized a sunflower optimization (SO) algorithm to improve the ability of the adaptive neurofuzzy inference system (ANFIS) and multilayer perceptron (MLP) models to predict lake water levels. The performance of the hybrid models was also validated using the firefly algorithm (FFA) and practical swarm optimization (PSO). The optimization algorithms were used to find the best tuning hyperparameters for the ANFIS and MLP models. The ANFIS-SO model was discovered to have a lower level of uncertainty based on the percentage of more responses in the confidence band and the model's smaller bandwidth.

(6) Tao et al. [20] applied the improved grasshopper optimization algorithm (IGOA) to combine relevance vector machine (RVM) and artificial neural network (ANN) models to predict catchment water levels in Malaysia. To validate the performance of the IGOA, the classical GOA and particle swarm optimization (PSO) algorithm were used. When different statistical criteria were considered, the IGOA enhanced the models' performance, and the RVM-IGOA was better.

(7) Ghorbani et al. [21] evaluated the prediction performance of a combined model integrating the multilayer perceptron (MLP) with the firefly algorithm (FFA) in Lake Egirdir, Turkey. Monthly data from 1961 to 2016 were utilized for training and testing the proposed hybrid MLP-FFA model to develop and investigate its veracity. The average mutual information technique was used to predict four lagged combinations of historical data. The results indicate that the MLP-FFA model is superior to the MLP model alone.

ANN is considered to be a robust method for predicting water levels because of its capacity to handle vast volumes of nonlinear data and its noise-handling capabilities [22]. Multiple researchers in the field of water level prediction have recommended using the ANN model such as [20, 23, 24].

Additionally, various metaheuristic optimization algorithms could be used to solve multiple issues in various application sectors. The capacity of optimization algorithms to pick the ideal values of system parameters under different conditions and their time-saving capabilities are their key advantages. The constriction coefficient-based particle swarm optimization and chaotic gravitational search algorithm (CPSOCGSA) was created by Rather and Bala [25], it is one of the updated algorithms, and this algorithm is being used for the first time in this study. In addition, The slime mold algorithm (SMA) was proposed recently by Li et al. [26]. SMA is employed in several optimization issues, such as water demand prediction [2, 27], and optimal power flow problems [28].

Another essential issue to consider is that most studies in this field have not applied all the steps of the data preprocessing techniques, where the data preprocessing is vital to improving the quality of time series and determining the best independent variables; for example, singular spectrum analysis (SSA) [29] and wavelet transform (WT) [30] were used to denoise raw time series. Principal component analysis (PCA) [10], mutual information (MI) [21], and tolerance technique [2] were employed to choose the most appropriate independent variables.

Furthermore, several studies in water level forecasting [21, 31–34] are advised to apply the climatic factors for forecasting water levels to improve forecasting accuracy so that climate variables will be included as input in this research.

This paper aims to critically examine a novel hybrid methodology that includes a combination of data preprocessing techniques, a metaheuristic algorithm, and a machine learning model to predict the Tigris River water level, upstream of the Al-Kut barrage in Al-Kut City.

The principal objective of this project was to

(1) Investigate 12 climate variables over ten years to determine the impact of climate change on water levels
(2) Improve the quality of raw data and determine the best scenario of predictors
(3) Integrate the ANN model by MPA algorithm by selecting the best ANN hyperparameters
(4) Evaluate the updated MPA-ANN algorithm by comparing it with recent SMA-ANN and CPSOCGSA-ANN algorithms to increase the prediction range and decrease the uncertainty
(5) Provide further insight into the application and interpretation of the recent prediction model to the stakeholders

To the best of the authors' knowledge, this study is the first to explore a novel methodology for predicting water levels utilizing updated algorithms (MPA, CPSOCGSA, and SMA).

This paper is divided into four sections. The study area and data set are described in Section 2. The proposed approach for estimating water level is described in Section 3. Results and discussion of the data are presented in Section 4. The conclusions are stated in Section 5.

2. Study Area and Data Collection

Iraq depends on Tigris and Euphrates Rivers as the primary source of freshwater for agricultural, residential, commercial, and industrial. Tigris River is one of the major rivers in the Middle East, with a length of 1718 Kilometers, which is shared by Turkey, Syria, and Iraq, and approximately 85% of the overall river basin lies in Iraq. The Tigris basin's climate ranges from semihumid in the headwaters to semiarid around its confluence with the Euphrates River in the south [35].

Al-Kut is the capital of the Wasit Governorate, Iraq, which lies between two latitudes ($32^o\ 21'$ and $32^o\ 34'$) north and two longitudes ($45^\circ\ 54'$ and $45^o\ 45'$) east, with an average height of around 20 meters. The governorate's area is 17,153 square kilometers. In comparison, the city of Kut's built-up area is about 40 square kilometers (4000 hectares) [36]. Al-Kut barrage was established on the Tigris River. Its function is to raise the water level upstream to provide the Dujaila and al-Gharraf branches with water responsible for the prosperity of several cities that lie beside these branches (see Figure 1).

Precipitation varies from less than 100 mm/year in the middle and southern deserts to over 1,000 mm/year in northern Iraq's mountainous region. In the winter, the average minimum temperature varies from near freezing in the northern and northeastern foothills to 4–5°C in the alluvial plains of southern Iraq. In the summer, the average minimum temperature is around 27–31°C, with maximum temperatures between 41 and 45°C (FAO, 2003). The combination of a lack of rain and high heat has turned much of Iraq into a desert [37].

Historical monthly water level data (upstream data of the Kut barrage) was provided by the directorate of water resources in the Wasit Governorate. Generally, data from metrological stations in Iraq were lost due to abnormal conditions (i.e., terrorism). Accordingly, to be with Ahmad et al. [38] and Capt et al. [39], monthly weather variables were collected from the National Oceanic and Atmospheric Administration (NASA) [40]. The data were collected for the period 01vJanuary 2011 to 31vDecember 2020. It includes Water level (WL) (m), maximum temperature ($T_{\max}$) (°c), minimum temperature ($T_{\min}$) (°c), mean temperature (Tmean) (°c), rainfall (rain) (mm/day), dew forest (DF) (°c), relative humidity (RH) (percent), specific humidity (SH) (percent), surface pressure (P) (kpa), wind speed (WS) (m/s), maximum wind speed (WSmax) (m/s), minimum wind speed (WSmin) (m/s), and range wind speed (WSrange) (m/s). Figure 2 shows the monthly time series of water level and the box plot in sections a and *b*, respectively.

3. Methodology

The suggested methodology consists of four parts: data preprocessing, artificial neural networks, MPA algorithm, and model evaluation. The structure of the proposed methodology for predicting monthly data of water level based on meteorological parameters is shown in Figure 3.

3.1. Data Preprocessing. The three strategies used in this study as data preprocessing techniques are normalization, cleaning, and determination of the best model input:

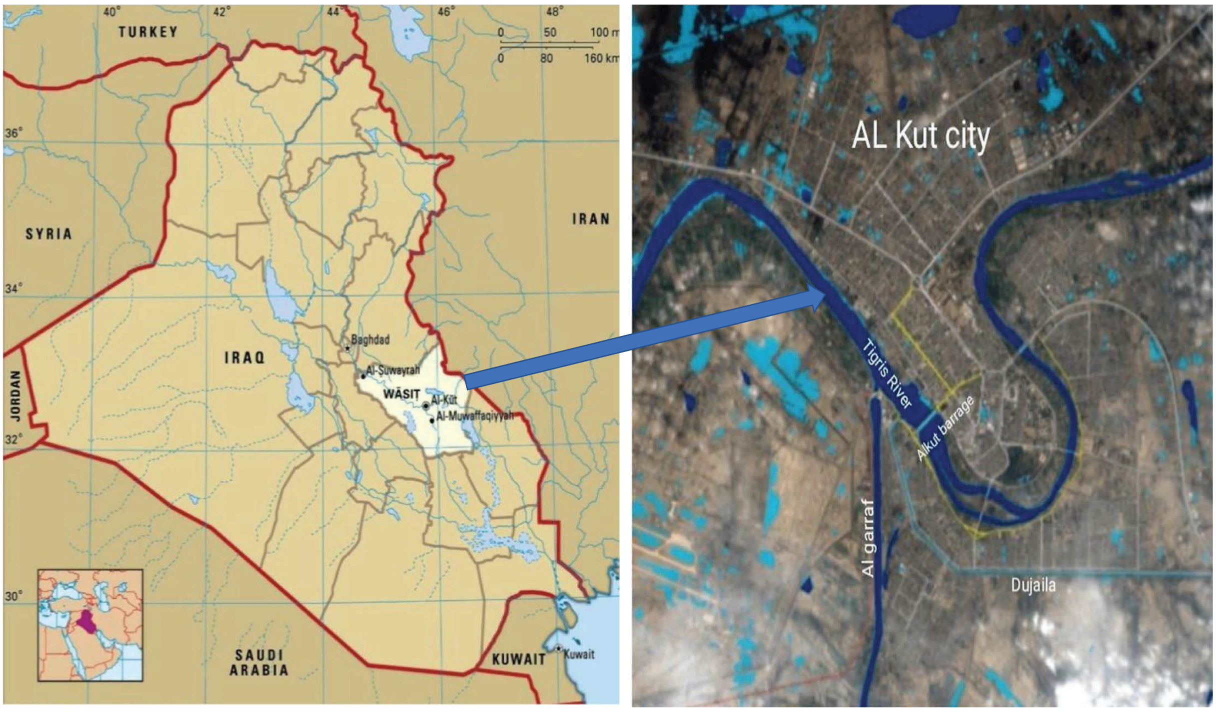

Figure 1: Location of case study, AL-Kut City, Tigris River, Iraq.

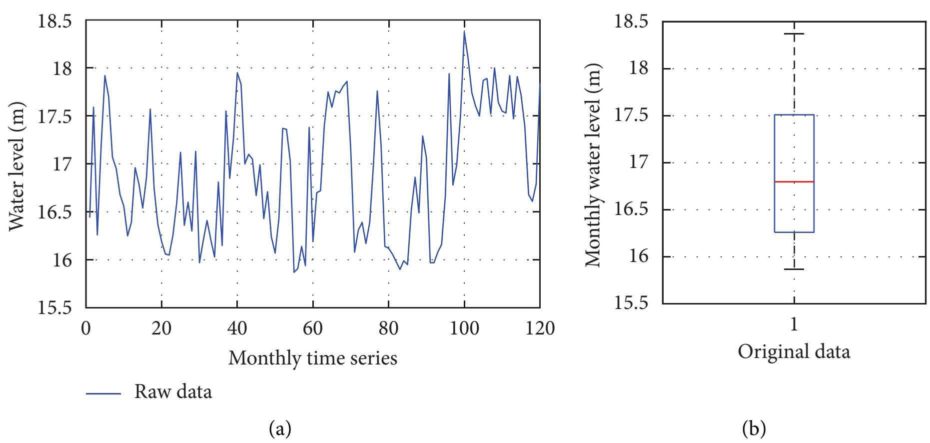

Figure 2: (a) Monthly time series and (b) box plot for water level.

3.1.1. Normalization. The goal of this method is to smooth out the answer space and reduce outliers effects in order to make the time series normally or nearly normally distributed [41, 42]. The natural logarithm approach was employed to normalize the time series because it makes data more static and reduces collinearity across predictors [43].

3.1.2. Cleaning. Noise and outliers cause a negative impact on the performance of the suggested model. As a result, data cleaning is necessary to detect and treat noisy data [44]. Accordingly, the box and whisker approach was utilized via the SPSS version (24) statistics package to clean data from outliers. This step has a significant favorable impact on the precision of the suggested prediction model. Singular spectrum analysis (SSA) was also employed to denoise raw data.

SSA is a strong tool for analyzing time series in order to identify relevant predictive properties. It applies to both linear and nonlinear time series and tiny sample sizes. It does not rely on any statistical assumptions based on the series' stationarity and linearity or the residuals' normality [45]. It detects and removes noise from time series in order to improve the regression coefficient and reduce the scale of error. All time series have many noise components, and the SSA is one of the pretreatment signal techniques for denoising the raw time series by decomposing them into separate components [46]. This strategy has proven to be effective in a variety of fields, including drought forecasting [47], industry [46], economics [48], streamflow prediction

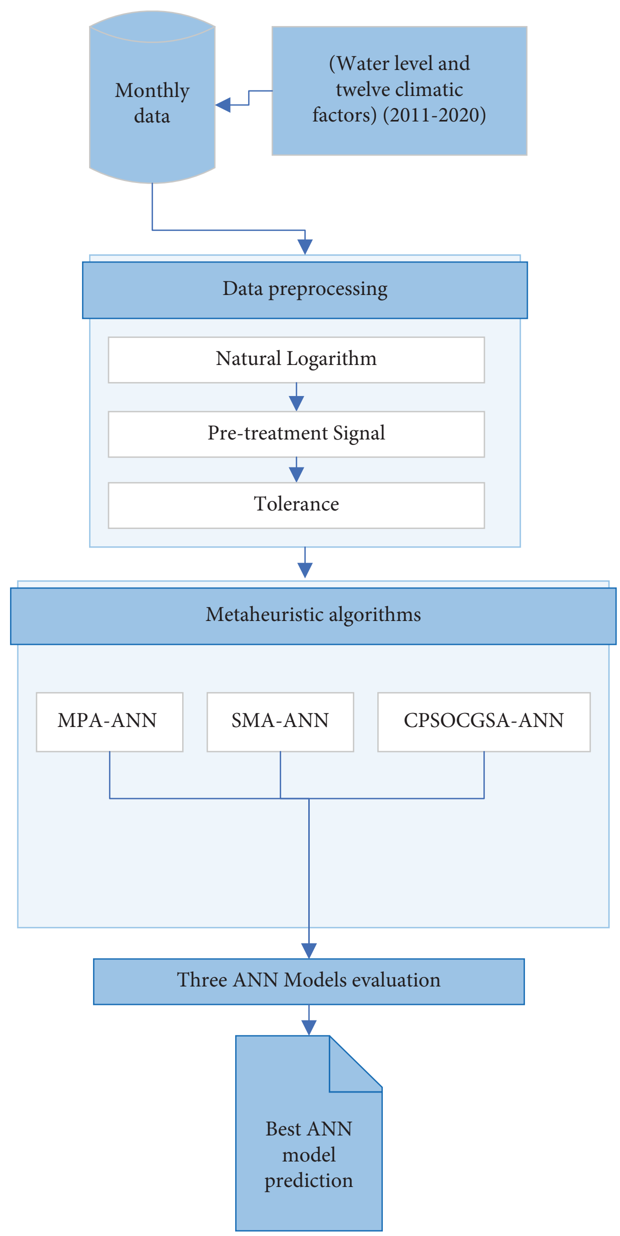

FIGURE 3: The suggested methodology for predicting water level.

[49], and water demand prediction [41, 42]. More information on SSA can be found in [50].

3.2. Selection of Best Model Input. Determining the explanatory variables that influence water level as model input data is crucial in creating any successful prediction model, not only for the ANN prediction model [51]. Cross-correlation and tolerance techniques were used to pick the best model input scenario. Pallant [52] V advised utilizing the tolerance approach to select independent variables with a tolerance coefficient equal to and more than 0.2.

3.3. Marine Predators Algorithm (MPA). MPA is one of the newest nature-inspired optimization algorithms proposed by Faramarzi et al. [53]. The concept of the MPA algorithm is taken from the forging movement of marine predators such as sharks and sunfish (Figure S1).

3.3.1. Step 1: Prey's Population Initialization. The MPA begins with setting an initial random solution set X_0 as defined by (1). This random set is made within the search space:

$$X_0 = X_{lb} + r * \left(X_{ub} - X_{lb}\right), \tag{1}$$

where X_{lb} an d X_{ub} are the lower and upper bond of variables, respectively, and r is a random vector in a range of zero to one.

3.3.2. Step 2: Creation of the Predator Matrix. In the MPA, both the predators and the prey are considered search agents because they look for their own food. The top predator, which is naturally more talented than other search agents, in the search agents is termed as Elite. Mathematically, the Elite matrix is updated based on information on the prey's locations. The formulation of the Elite and prey matrix is shown as follows:

$$\begin{gathered} Elite = \begin{bmatrix} X_{11}^1 & X_{12}^1 & . & X_{1d}^1 \\ X_{21}^1 & X_{22}^1 & . & X_{2d}^1 \\ . & . & . & . \\ . & . & . & . \\ X_{n1}^1 & X_{n2}^1 & . & X_{nd}^1 \end{bmatrix}, \\ X = \begin{bmatrix} X_{11} & X_{12} & . & X_{1d} \\ X_{21} & X_{22} & . & X_{2d} \\ . & . & . & . \\ . & . & . & . \\ X_{n1} & X_{n2} & . & X_{nd} \end{bmatrix}. \end{gathered} \tag{2}$$

3.3.3. Step 3: MPA Optimization Process. After constructing the Elite and prey matrices, the positions of the predators and preys are updated according to three phases. These phases are related to the velocity ratio between the predator and the prey. The three phases are the high-velocity ratio, unit velocity ratio, and low-velocity ratio.

3.3.4. Phase 1: High-Velocity Ratio. In this phase, the movement of the predator is faster than the prey. The step size of prey movement is updated as in the following equation:

$$\begin{aligned} S_i &= R_B \otimes \left(\text{Elite}_i - R_B \otimes X_i\right), i = 1, 2, \ldots, n, \\ X_i &= X_i + P.R \otimes S_i. \end{aligned} \tag{3}$$

where R is a random vector where the values of its elements are in a range of zero to one value. P is a constant number. R_B is a random vector refereeing to Brownian motion.

The $\otimes$ symbol refers to the element-wise multiplication process.

This phase occurs in one-third of the total number of iterations (i.e., $1/3t_{\max}$).

3.3.5. Phase 2: Unit Velocity Ratio. In this phase, the process of searching for prey/food is simulated. The prey movement is represented by Levy flight while the predator is represented by Brownian motion. This phase occurs in the second third of the total iterations (i.e., $1/3t_{\max} < t < 2/3t_{\max}$. The following equations are applied to the first half of the population.

$$\begin{aligned} S_i &= R_L \otimes \left(\text{Elite}_i - R_L \otimes X_i\right), i = 1, 2, \ldots, n, \\ X_i &= X_i + P.R \otimes S_i, \end{aligned} \tag{4}$$

where R_{L} represents numbers following Levey distribution. The second half of the population is subjected to the following equations:

$$\begin{aligned} S_i &= R_B \otimes \left(R_B \otimes \text{Elite}_i - X_i\right), i = 1, 2, \ldots, n, \\ X_i &= X_i + P.CF \otimes S_i, CF = \left(1 - \frac{t}{t_{\max}}\right)^{2\left(t/t_{\max}\right)}. \end{aligned} \tag{5}$$

CF represents a parameter that controls the movement step size of the predator.

3.3.6. Phase 3: Low-Velocity Ratio. This phase is considered the last part of the optimization, and it simulates predator movements when it is faster than the prey. It occurs in the last third of the total iterations (i.e., $2/3t_{\max}$):

$$\begin{aligned} S_i &= R_L \otimes \left(R_L \otimes \text{Elite}_i - X_i\right), i = 1, 2, \ldots, n, \\ X_i &= X_i + P.CF \otimes S_i, CF = \left(1 - \frac{t}{t_{\max}}\right)^{2\left(t/t_{\max}\right)}. \end{aligned} \tag{6}$$

3.3.7. Step 4: Eddy Formation and FADs. The environment parameters can also be taken in the simulation, such as the eddy formation and the fish aggregating device. The FAD's effect is

$$X_i = \begin{cases} X_i + CF\left[X_{\min} + R \otimes \left(X_{\max} - X_{\min}\right)\right] \otimes \overline{U}, & \text{if } r < \text{FAD}, \\ X_i + \left[\text{FADS}(1 - r) + r\right]\left(X_{r1} - X_{r2}\right), & \text{if } r > \text{FAD}, \end{cases} \tag{7}$$

where r is a random value in a range of zero to 1. R1 and r2 refer to the random indices from prematrix. FADS refers to the FAD's probability. The $\overline{U}$ is a binary vector.

3.3.8. Step 5: Marine Memory. The marine predators have good memory for successful locations for foraging. In the MPa, this was simulated by enabling the MPA to save the solution fitness values in each iteration and comparing it to other fitness values in successive iterations.

3.4. Artificial Neural Network (ANN). An artificial neural network defined is a set of simple processing units acting as a parallel distributed processor. Neurons are the units in charge of storing experimental knowledge for subsequent disposal. The ANNs, like the brain, replicate the biological nervous system; they learn by examples and have knowledge stored in the connection weights between neurons [54]. When analytically demonstrating the relationship between the dependent and independent variables for any physical occurrence is difficult, the ANN tool could be particularly useful. ANN can use previous data to make a reasonably accurate prediction of the modeled parameters. It may be used to simulate any physical occurrence [55], making it suitable for a wide range of hydrological applications, including water demand forecasting [29, 44], streamflow forecasting [56, 57], drought prediction [58, 59], and water quality predictions [60, 61].

This research applied the multilayer perceptron (MLP) network (a feed-forward, backpropagation network) to simulate water level. It has been used to anticipate water resources and highly effective hydrology applications. Learning method Levenberg–Marquardt (LM) is used to train data. The MLP comprises four layers; the first is the input layer, which contains the model inputs, after which there are two hidden layers and one output layer, which includes the water level [2]. Different researchers have successfully used ANN with two hidden layers in different contexts, and the findings showed that these models accurately captured the nonlinear pattern between predictors and targets [23, 41, 62] (Figure S2). The trial-and-error procedure does not always provide the best solution [41]. As a result, the learning rate (Lr) and the number of neurons in the first (N1) and second (N2) hidden layers are determined using metaheuristic algorithms. The hybrid techniques improved the performance of the ANN model and were time-saving.

3.5. Model Performance Assessment. The forecast model's performance was evaluated in various statistical tests because no global performance test is suitable for specific usage [63]. Therefore, mean absolute error (MAE, (9)), root mean squared error (RMSE, (10)), coefficient of determination (R^2, Equation (11)), mean bias error (MBE, Equation (13)), mean absolute relative error (MARE, Equation (12)), nash Sutcliffe coefficient (NSC, Equation (14)), and scatter index (SI, Equation (13)) are used to assess the model. The Taylor diagram and residual analysis plot tests were also applied for more confirmation.

$$\text{MBE} = \frac{1}{N}\sum_{i=1}^{N}\left(O_i - F_i\right), \tag{8}$$

$$\text{MAE} = \frac{\sum_{i=1}^{N}\left|O_i - F_i\right|}{N}, \tag{9}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}\left(O_i - F_i\right)^2}{N}}, \tag{10}$$

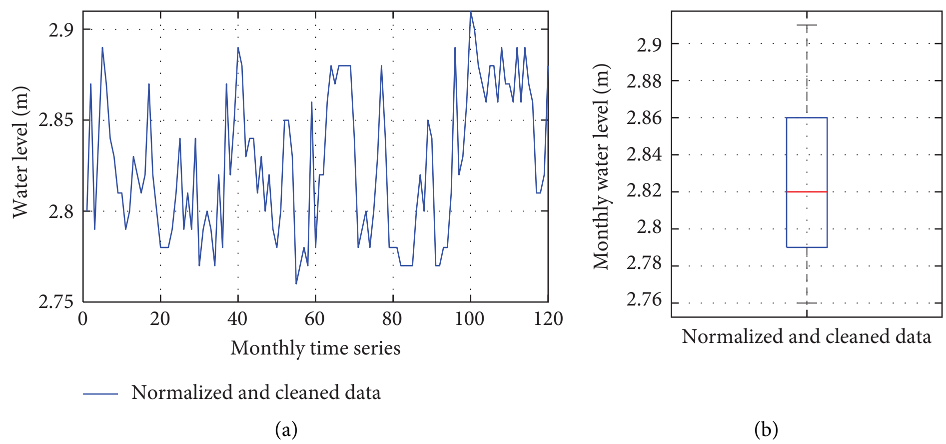

FIGURE 4: (a) Monthly time series and (b) box plot of normalized and cleaned water level data.

$$R^2 = \left[\frac{\sum_{i=1}^{N}(O_i - \overline{O}_i)(F_i - \overline{F}_i)}{\sqrt{\sum (O_i - \overline{O}_i)^2 \sum (F_i - \overline{F}_i)^2}}\right]^2, \tag{11}$$

$$\mathrm{MARE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|O_i - F_i|}{O_i}, \tag{12}$$

$$SI = \frac{\mathrm{RMSE}}{\overline{O}} \times 100, \tag{13}$$

$$\mathrm{NSC} = 1 - \frac{\sum_{i=1}^{N}(O_i - F_i)^2}{\sum_{i=1}^{N}(O_i - \overline{O}_i)^2}, \tag{14}$$

where O_i and F_i were the observed and estimated data. The length of the time series is N. The mean value of the observed data was $\overline{Oi}$. The average value of the predicted data was $\overline{Fi}$. The model's performance is good when the value of $R^2 \geq 0.85$ [64]. The best model with values approaches zero for MBE, MAE, and RMSE criteria [65, 66]. NSE can range between -∞ and 1, NSE = 1 denotes a perfect match between the predictions and the observations, NSE = 0 implies that the predictions are as accurate as the mean of the observations, and NSE <0 indicates that the predictions are less accurate than the mean of the observations [67, 68]. When the value of SI is less than 10%, the model accuracy is excellent, it is good if the value is between 10% and 20%, fair when it is between 20% and 30%, and poor when it is more than 30% [69, 70].

4. Results and Discussion

4.1. Data Preprocessing Analysis. According to Tabachnick and Fidell [71], all the time series were transformed to reduce the impact of outliers and make the distribution of time series normal or close to normal. Then, the remaining outliers after the transformation (if found) were rescaled. Figure 4 demonstrates the time series and box plot for normalized and clean water level time series.

After that, the SSA technique was employed to denoise the time series. The normalized and clean time series (top row), the trend signal (second component), the seasonal signal (third component), the stochastic signal (fourth component), and the noise signal (fourth component), respectively. Figure 5 shows the normalized and cleaned water level data and its deconstructed components, such as trend, seasonality, stochasticity, and noise.

In the final stage of data preprocessing, a tolerance method was employed to find the optimal scenario of independent factors (i.e., climate factors) that could accurately simulate water level data and avoid multicollinearity. After performing different scenarios, the best scenario was selected that offered tolerance coefficients of more than 0.2 for all nominated predictors, as shown in Table 1. The table reveals that $\mathrm{WL}_{t\text{-}1}$, DF, and WS were chosen as the best model input.

Table 2 displays the correlation coefficients between dependent and independent factors in both the raw and preprocessed data stages. The table demonstrates that data preprocessing techniques improved the data quality significantly, such as raising the correlation coefficient between water level and lag of water level time series (from 0.643 to 0.993) and dew forest time series (from 0.378 to 0.457). The correlation coefficient values confirmed the link between water level and climatic variables.

According to Tabachnick, B. G., and Fidell [71], the connection between the size of the sample (N) and the number of independent variables should follow the following equation:

$$N \geq 50 + 8\mathrm{m}, \tag{15}$$

where the number of predictor variables is given by m, and N presents the size of simple, in this study, $N = 119$, which is more than the 74 needed.

4.2. Application of the Hybrid MPA-ANN Algorithm. The data was divided into three sets: training (70%), testing (15%), and validation (15%) [29, 44]. The recent hybrid algorithms MPA, CPSOCGSA, and SMA, were conducted using the MATLAB toolbox to find optimum hyperparameters of the ANN model. Five population sizes, 10, 20,

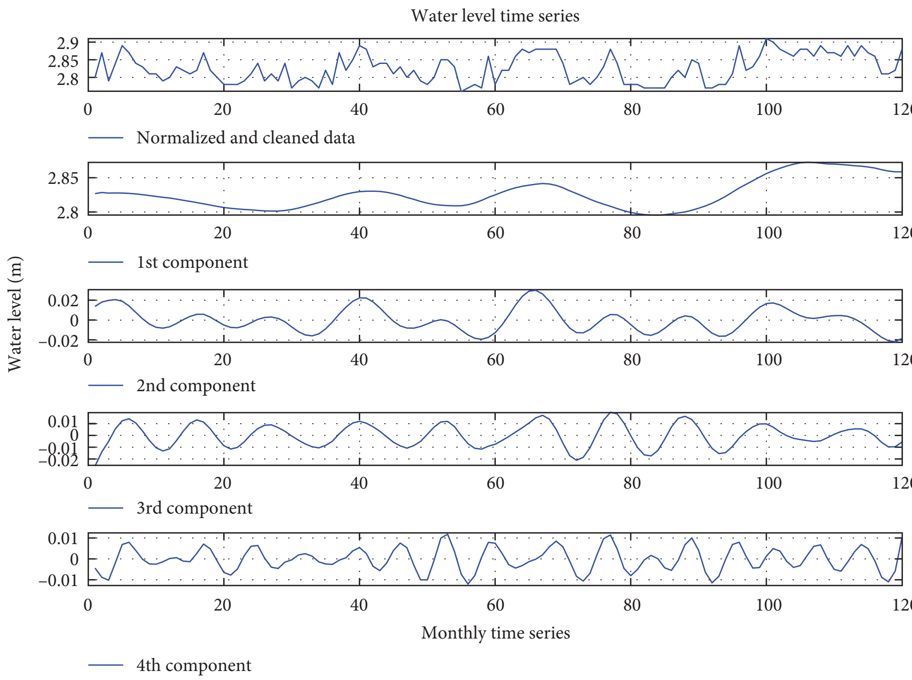

Figure 5: Normalized and clean time series and the 1st four components obtained by SSA.

Table 1: Collinearity statistics to the nominated predictors.

Climatic factors	Tolerance
Water _1	0.721
Dew	0.498
ws	0.420

Table 2: The correlation coefficients between independent and dependent factors.

Data	WL_{t-1}	DF	WS
Raw	0.643**	0.378**	−0.026
Preprocessed	0.993**	0.457**	−0.555**

**Correlation is significant at the 0.01 level (2-tailed).

30, 40, and 50, were used in all algorithms to evaluate the best number of hidden neurons and the best learning rate coefficient of the ANN model. It is worth noting that these population sizes refer to the size of the swarm, not the sample size given earlier. Five swarm sizes (10, 20, 30, 40, and 50) were used to conduct the combined model for each algorithm. Each swarm size was repeated five times to diminish the uncertainty and increase the predicting range (e.g., see Figure 6 for MPA-ANN algorithm). For each swarm size, the better one that offers the lowest error (e.g., the application second for the 10 swarm size of MPA-ANN is the best) was chosen and incorporated with the best implementation for the rest swarm sizes.

Figure 7(a) displays that the (30–5) swarm size provides the best solution for the MPA-ANN algorithm (MSE = 0.0007427, after 193 iterations), while in Figure 7(b), the (40–2) swarm size gives the best solution for the CPSOCGSA -ANN algorithm (MSE = 0.0006644, after 74 iterations). Figure 7(c) presents the swarm size (50–3) and offers the best solution for the SMA-ANN algorithm (MSE = 0.0006644, after 42 iterations). The ANN models' hyperparameters for the best swarm for each metaheuristic algorithm were tabulated in Table 3.

4.3. Application of the ANN Model. To be with Tao et al. [20] and Ghorbani et al. [21] techniques, three ANN models were configured based on the hyperparameters in Table 3. Each ANN model was performed multiple times to determine the best network that offers an accurate solution. To evaluate the performance of models, five statistical tests were used. The results of the performance criteria indices (RMSE, MBE, MAE, NSC, R^2, and MARE) of the CPSOCGSA-ANN, MPA-ANN, and SMA-ANN models in the validation stage are displayed in Table 4. [64]. Across these indices in Table 4 and according to the limitations in Section 3.4, it is clear to see that the MPA-ANN technique outperforms other techniques, confirmed by the higher R^2 values and the lower RMSE, MAE, MARE, and MBE values, as well as the value of NSC >0 for the validation stage.

The Taylor diagram was also used to assess the performance of various hybrid models in the validation stage. Figure 8 depicts the obtained results. The observed WL is represented by the red character (*A*) on the Taylor diagram's *X*-axis. A model is thought to be better if it is close to the observed point. Taylor diagram graphically compares three statistics (standard deviation (SD), correlation coefficient (R), and root mean square error difference (RMSD)) [20, 21]. It thus provides a reliable assessment of the relative performance of different models. According to the Taylor diagram, the performance of the MPA-ANN model (point B) was found to be the closest to the observed point compared

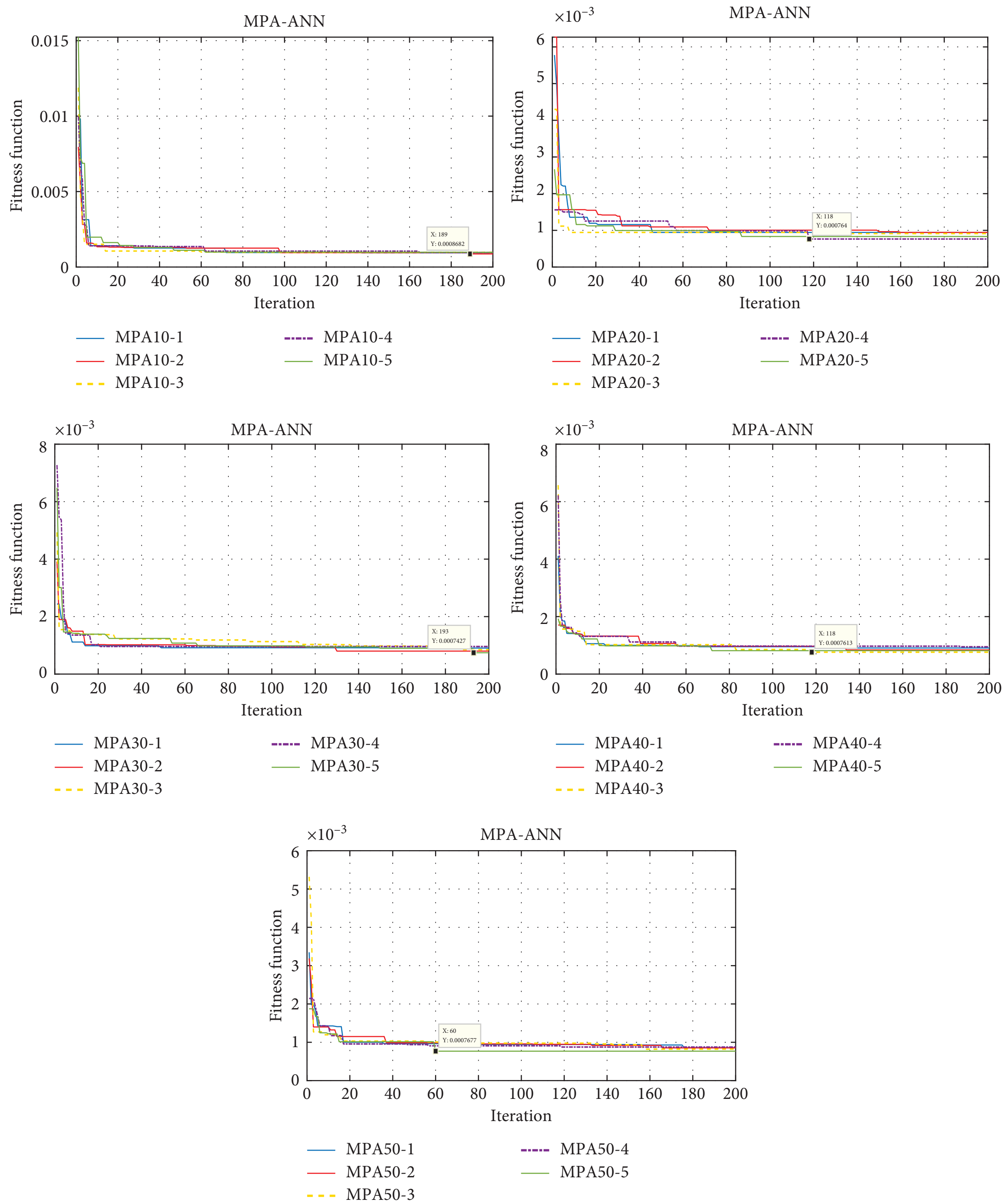

Figure 6: Performance of MPA-ANN algorithm with five trials for each swarm size.

with the CPSOCGSA-ANN model (point C) and SMA-ANN model (point D).

Also, Figure 9 depicts a residual comparison of the three models in the validation phase. As shown in the figure, the hybrid MPA-ANN model has a minimum error (0.0012–0.0037) meter, while the error for CPSOCGSA-ANN (0.0016–0.0127) and error is (-0.00009–0.0173) for SMA-ANN model. All the above three techniques confirm that the MPA-ANN model is the best.

Moreover, Figure 10 depicts the observed and forecasted water level time series from all models in the validation stage. The observed water level data are in blue, and the predicted water level data by MPA-ANN, CPSOCGSA-ANN, and SMA-ANN are in red, black, and pink, respectively. The

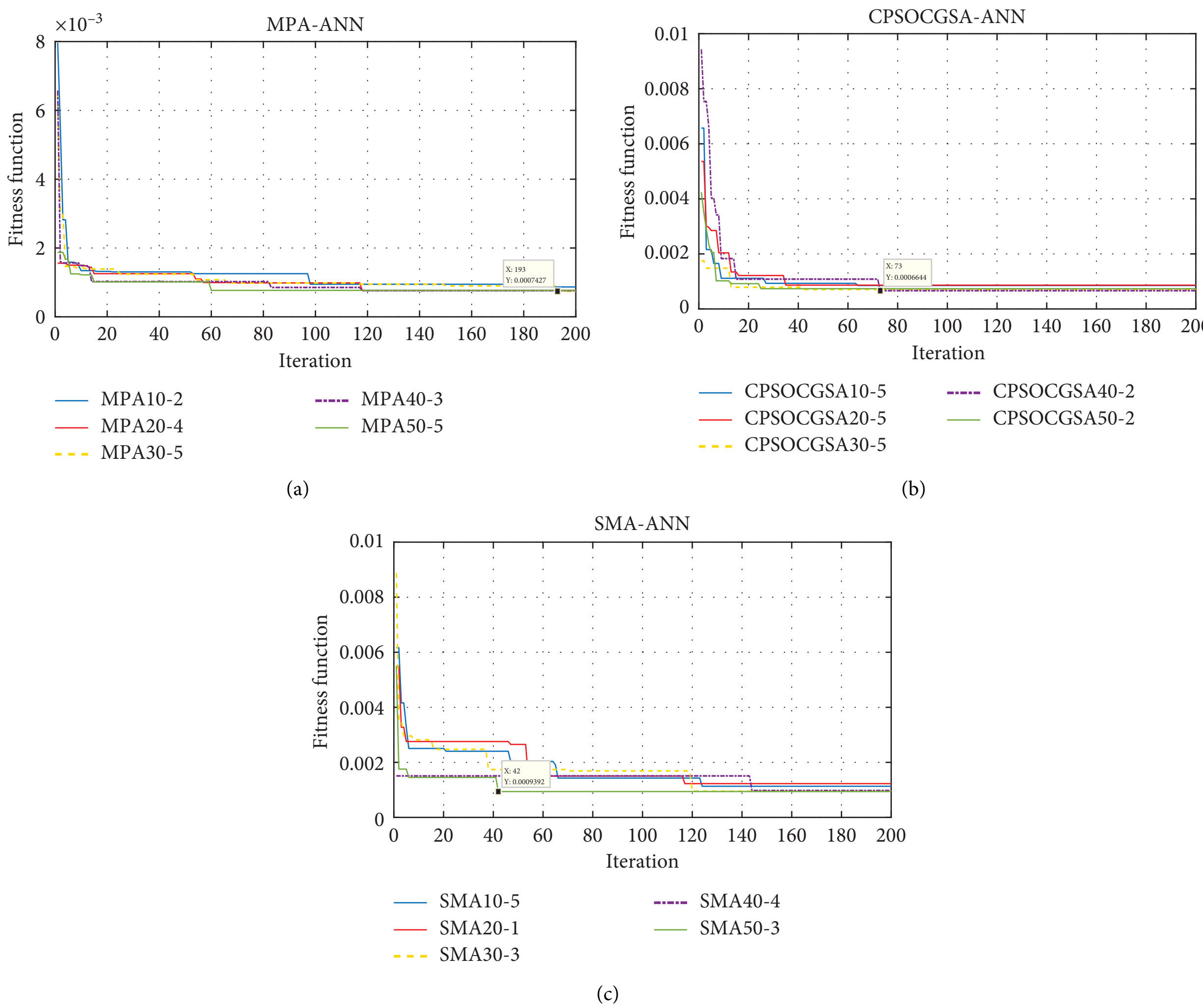

Figure 7: The best swarm for MPA, CPSOCGSA, and SMA algorithms.

Table 3: ANN hyperparameters of all algorithms.

Model	Lr	N1	N2
MPA-ANN	0.5758	1	2
CPSOCGSA-ANN	0.3350	2	4
SMA-ANN	0.1202	2	10

Lr = learning rate, and N1 and $N2$ = number of nodes in the first and second hidden layers, respectively.

Table 4: Performance assessment for validation data stage.

Model	MAE (m)	RMSE (m)	R^2	MBE (m)	MARE	NSC
MPA-ANN	0.0025	0.0026	0.98	0.0025	0.00086	0.66
CPSOCGSA-ANN	0.0078	0.0085	0.94	0.0078	0.0027	−2.76
SMA-ANN	0.0088	0.0098	0.76	0.0088	0.0031	−3.97

figure shows an excellent fit between observed and forecasted water level time series, demonstrating the capacity of the MPA-ANN technique to simulate observed time series. This technique can precisely capture the pattern of water level more than CPSOCGSA-ANN and SMA-ANN models, a result which is in line with the scale of error for each series, as detailed previously.

For extra validation for the MPA-ANN model, a scatter index was employed. The value SI = 0.0009 means that the MPA-ANN has an excellent performance according to the limitations in Section 3.4. Furthermore, the Kolmogorov–Smirnov and Shapiro–Wilk tests agree that the residual data are normally distributed depending on the significant values (p value) being more than 0.05 [72, 73].

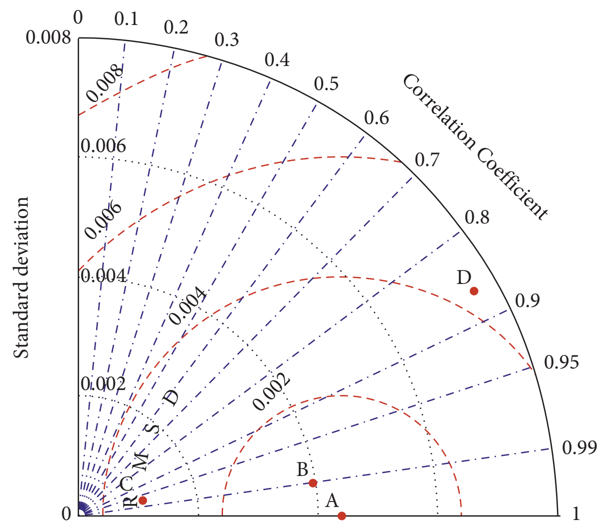

Figure 8: Taylor diagram of MPA-ANN, CPSOCGSA-ANN, and SMA-ANN predicted models.

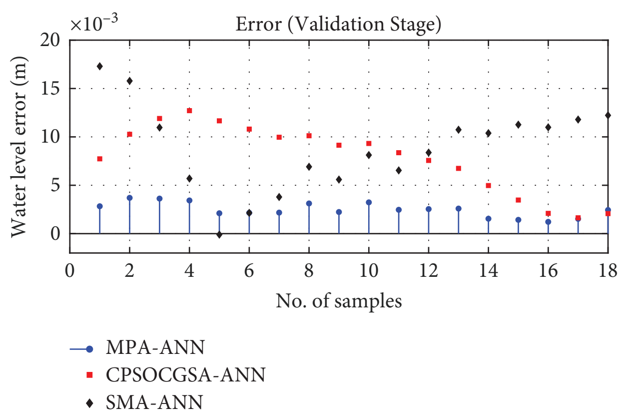

Figure 9: SMA-ANN, MPA-ANN, and CPSOCGSA-ANN error distributions.

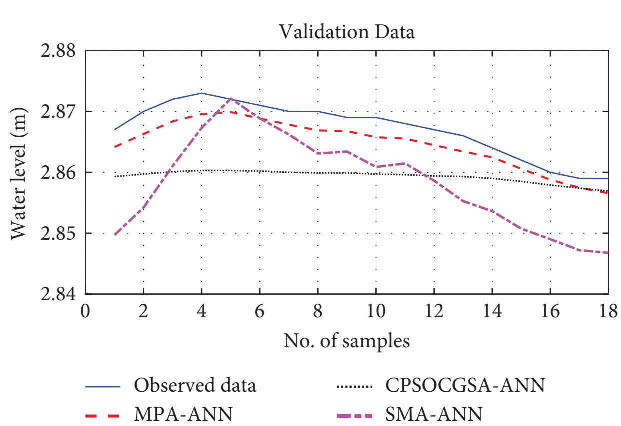

Figure 10: The comparison between actual and predicted data for MPA-ANN, CPSOCGSA-ANN, and SMA-ANN for the validation stage.

The study's statistical testing findings show the following: (1) Data preprocessing techniques play an essential role in improving data quality, particularly the SSA method. (2) The tolerance method successfully selected the optimal independent factor scenario. (3) MPA-ANN is an efficient and reliable technique for monthly water level forecasting. (4) Applying each swarm five times to each algorithm increased the selection accuracy of the best ANN model parameters. (5) The results confirm the relationship between climate variables and water level. (6) Using different scenarios and techniques increase the reliability and range of prediction, decreases the uncertainty, and gives a scientific view to the decision-maker.

Wolpert and Macready [74] declared that based on the No Free Lunch (NFL) theorem, there is no specific method that can bring the optimum result compared with other methods for all the issues of optimization. According to NFL, Faramarzi et al. [53] evolved the combined MPA technique for warranting the global solution, considering different techniques and strategies throughout the process of optimization. Various strategies of foraging have significantly inspired MPA in the biological interaction between predators and prey. Accordingly, the Brownian and LF distributions were designed not only to have a systematic explorer-exploiter tendency professionally but also to significantly improve the search ability in each performance. These allowed the MPA technique to precisely detect the global optima of the issues of optimization considered in this study.

5. Conclusion

This study used a novel methodology, including a combination of preprocessing techniques and the hybrid updated MPA-ANN algorithm to estimate the monthly water level. The historical data for Tigris River water levels and climate variables from 2011 to 2020 in Al-Kut City was considered to build and assess the methodology. The updated MPA-ANN algorithm has been compared with the recent two hybrid algorithms, CPSOCGSA-ANN and SMA-ANN. The findings indicate that data pretreatment techniques are powerful in improving data quality. However, the MPA-ANN algorithm outperforms both CPSOCGSA-ANN and SMA-ANN algorithms based on several statistical criteria (e.g., $R^2 = 0.98$, $\text{MARE} = 0.00086$, and $\text{SI} = 0.0009$), indicating that the performance of the MPA-ANN algorithm is good to excellent. The present study draws the following conclusions: a further investigation into data preprocessing techniques is strongly recommended. More broadly, research is also needed to investigate the MPA-ANN performance in the other fields of hydrology or even in the water level prediction in different regions.

Data Availability

Tigris water level data was provided by Water Resources Department/Wasit Province. Additionally, climatic factors data were collected from the National Oceanic and Atmospheric Administration (NASA) (https://www.ncdc.noaa.gov/cdo-web/datatools/findstation).

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors are grateful to the Water Resources Department/Wasit Province for providing the Tigris water level data for this research. The APC was funded by the Lulea University of Technology.

Supplementary Materials

This section has the flowchart of the proposed MPA-ANN algorithm and the structure of the ANN model. (*Supplementary Materials*)

References

[1] S. L. Farhan, M. G. Abdelmonem, and Z. A. Nasar, "The urban transformation of traditional city centres: holy karbala as a case study," *International Journal of Architectural Research: ArchNet-IJAR*, vol. 12, no. 3, p. 53, 2018.

[2] S. L. Zubaidi, I. H. Abdulkareem, K. S. Hashim et al., "Hybridised Artificial Neural Network Model with Slime Mould Algorithm: A Novel Methodology for Prediction of Urban Stochastic Water Demand," *Water*, vol. 12, no. 10, pp. 1–18, 2020.

[3] R. de Zeeuw, E. van Ettinger, S. W. Frye et al., "A machine learning approach for estimation of shallow water depths from optical satellite images and sonar measurements," *Journal of Hydroinformatics*, vol. 15, no. 4, pp. 1408–1424, 2013.

[4] A. Grządziel, "The Importance of Under-keel Sound Velocity Sensor in Measuring Water Depth with Multibeam Echosounder," *Energies*, vol. 14, no. 5267, 2021.

[5] Z. Liu, L. Cheng, K. Lin, and H. Cai, "A Hybrid Bayesian Vine Model for Water Level Prediction Environmental Modelling & Software," *A hybrid bayesian vine model for water level prediction*, vol. 142, 2021.

[6] C. Zhong, T. Guo, Z. Jiang, X. Liu, and X. Chu, "A hybrid model for water level forecasting: a case study of wuhan station," in *Proceedings of the 2017 4th International Conference on Transportation Information and safety(ICTIS)*, banff, canada, August 2017.

[7] L. A. Al-Maliki, S. L. Farhan, I. A. Jasim, S. K. Al-Mamoori, N. Al-Ansari, and S. L. Fegade, "Perceptions about water pollution among university students: a case study from Iraq," *Cogent Engineering*, vol. 8, no. 1, 2021.

[8] S. Ethaib, S. L. Zubaidi, N. Al-Ansari, and S. L. Fegade, "Evaluation water scarcity based on GIS estimation and climate-change effects: a case study of Thi-Qar Governorate," *Iraq Cogent Engineering*, vol. 9, no. 1, 2022.

[9] M. Çimen and O. Kisi, "Comparison of two different data-driven techniques in modeling lake level fluctuations in Turkey," *Journal of Hydrology*, vol. 378, no. 3-4, pp. 253–262, 2009.

[10] B. Mohammadi, Y. Guan, P. Aghelpour, S. Emamgholizadeh, R. Pillco Zolá, and D. Zhang, "Simulation of titicaca lake water level fluctuations using hybrid machine learning technique integrated with grey wolf optimizer algorithm," *Water*, vol. 12, no. 11, p. 3015, 2020.

[11] E. S. K. Tiu, Y. F. Huang, J. L. Ng, N. AlDahoul, A. N. Ahmed, and A. Elshafie, "An Evaluation of Various Data Pre-processing Techniques with Machine Learning Models for Water Level Prediction," *Natural Hazards*, vol. 110, no. 7, 2021.

[12] K. Park, Y. Jung, Y. Seong, and S. Lee, "Development of Deep Learning Models to Improve the Accuracy of Water Levels Time Series Prediction through Multivariate Hydrological," *Data Water*, vol. 14, 2022.

[13] H. Yasuda, A. A. Fenta, M. L. Berihun, K. Inosako, T. Kawai, and A. S. Belay, "Water level change of Lake Tana, source of the Blue Nile: prediction using teleconnections with sea surface temperatures," *Journal of Great Lakes Research*, vol. 48, no. 2, pp. 468–477, 2022.

[14] Y. Zhang, X. Dai, R. Wan, G. Yang, and B. Li, "Comparison of random forests and other statistical methods for the prediction of lake water level: a case study of the Poyang Lake in China," *Hydrology Research*, vol. 47, pp. 69–83, 2016.

[15] Z. Hajirahimi and M. Khashei, "Hybridization of Hybrid Structures for Time Series Forecasting: A Review," *Artificial Intelligence Review*, vol. 86, 2022.

[16] A. S. Azad, R. Sokkalingam, H. Daud et al., "Water level prediction through hybrid SARIMA and ANN models based on time series analysis: red Hills reservoir case study," *Sustainability*, vol. 14, no. 3, p. 1843, 2022.

[17] M. L. Lineros, A. M. Luna, P. M. Ferreira, and A. E. Ruano, "Optimized Design of Neural Networks for a River Water Level Prediction," *System Sensors*, vol. 21, 2021.

[18] D. H. Nguyen, X. Hien Le, J.-Y. Heo, and D.-H. Bae, "Development of an extreme gradient boosting model integrated with evolutionary algorithms for hourly water level prediction," *IEEE Access*, vol. 9, pp. 125853–125867, 2021.

[19] M. Ehteram, A. Ferdowsi, M. Faramarzpour et al., "Hybridization of artificial intelligence models with nature inspired optimization algorithms for lake water level prediction and uncertainty analysis," *Alexandria Engineering Journal*, vol. 60, no. 2, pp. 2193–2208, 2021.

[20] H. Tao, N. K. Al-Bedyry, K. M. Khedher, S. Shahid, and Z. M. Yaseen, "River water level prediction in coastal catchment using hybridized relevance vector machine model with improved grasshopper optimization," *Journal of Hydrology*, vol. 598, Article ID 126477, 2021.

[21] M. A. Ghorbani, R. C. Deo, V. Karimi, Z. M. Yaseen, and O. Terzi, "Implementation of a hybrid MLP-FFA model for water level prediction of Lake Egirdir, Turkey," *Stochastic Environmental Research and Risk Assessment*, vol. 32, no. 6, pp. 1683–1697, 2017.

[22] B. Deng, S. H. Lai, C. Jiang, P. Kumar, A. El-Shafie, and 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*, vol. 14, no. 4, pp. 1987–2001, 2021.

[23] F. Farzad and A. H. El-Shafie, "Performance enhancement of rainfall pattern – water level prediction model utilizing self-organizing-map clustering method," *Water Resources Management*, vol. 31, pp. 945–959, 2016.

[24] M. A. Ghorbani, R. C. Deo, V. Karimi, M. H. Kashani, and S. Ghorbani, "Design and implementation of a hybrid MLP-GSA model with multi-layer perceptron-gravitational search algorithm for monthly lake water level forecasting," *Stochastic Environmental Research and Risk Assessment*, vol. 33, no. 1, pp. 125–147, 2018.

[25] S. A. Rather and P. S. Bala, *Applied Soft Computing and Communication Networks*, Springer, New York, NY, USA, 2020.

[26] S. Li, H. Chen, M. Wang, A. A. Heidari, and S. Mirjalili, "Slime mould algorithm: A new method for stochastic optimization," *Future Generation Computer Systems*, vol. 111, pp. 300–323, 2020.

[27] K. Yu, L. Liu, and Z. Chen, "An Improved Slime Mould Algorithm for Demand Estimation of Urban Water Resources," *Mathematics*, vol. 9, 2021.

[28] S. Khunkitti, A. Siritaratiwat, and S. Premrudeepreechacharn, "Multi-Objective Optimal Power Flow Problems Based on Slime Mould Algorithm," *Sustainability*, vol. 13, no. 13, 2021.

[29] S. L. Zubaidi, S. Ortega-Martorell, H. Al-Bugharbee et al., "Urban Water Demand Prediction for a City that Suffers from Climate Change and Population Growth: Gauteng Province Case Study," *Water*, vol. 12, 2020.

[30] E. C. Loh, S. B. Ismail, A. Mustafa, and S. M. Shaharudin, "Application of discrete wavelet transformation with artificial neural network in kelantan water level prediction," *Journal of Advanced Research in Dynamical & Control Systems*, vol. 11, no. 11, pp. 82–88, 2019.

[31] R. Adnan, F. A. Ruslan, A. m Samad, and Z. M. Zain, "Artificial neural network modelling and flood water level prediction using extended kalman filter," in *Proceedings of the International Conference on Control System, Computing and Engineering*, Penang, Malaysia, November 2012.

[32] P. Coulibaly, "Reservoir Computing approach to Great Lakes water level forecasting," *Journal of Hydrology*, vol. 381, no. 1-2, pp. 76–88, 2010.

[33] M. El-Diasty and S. Al-Harbi, "Development of wavelet network model for accurate water levels prediction with meteorological effects," *Applied Ocean Research*, vol. 53, pp. 228–235, 2015.

[34] P. Páliz Larrea, X. Zapata-Ríos, and L. Campozano Parra, "Application of Neural Network Models and ANFIS for Water Level Forecasting of the Salve Faccha Dam in the Andean Zone in Northern Ecuador," *Water*, vol. 13, no. 15, 2021.

[35] H. Tao, A. O. Al-Sulttani, A. M. Salih Ameen et al., "Training and Testing Data Division Influence on Hybrid Machine Learning Model Process: Application of River Flow Forecasting," *Complexity*, vol. 2020, Article ID 8844367, 22 pages, 2020.

[36] S. F Balket and N. M Asmael, "Study the characteristics of public bus routes in Al kut city," *Journal of Engineering and Sustainable Development*, vol. 25, pp. 3–94, 2021.

[37] S. A. Salman, S. Shahid, T. Ismail, E. Chung, and A. M. Al-Abadi, "Long-term Trends in Daily Temperature Extremes in Iraq," *Atmospheric Research*, vol. 198, pp. 97–107, 2017.

[38] H. Q. Ahmad, S. A. Kamaruddin, S. B. Harun, N. Al-Ansari, S. Shahid, and R. M. Jasim, "Assessment of spatiotemporal variability of meteorological droughts in northern Iraq using satellite rainfall data," *KSCE Journal of Civil Engineering*, vol. 25, no. 11, pp. 4481–4493, 2021.

[39] T. Capt, A. Mirchi, S. Kumar, and W. S. Walker, "Urban Water Demand: Statistical Optimization Approach to Modeling Daily Demand," *Journal Of Water Resources Planning And Management*, vol. 147, no. 2, 2021.

[40] NOAA, "National Oceanic and Atmospheric Administration," *Data Tools: Find a Station,* National Oceanic and Atmospheric Administration, 2021.

[41] S. L. Zubaidi, P. Kot, R. M. Alkhaddar, M. Abdellatif, and H. Al-Bugharbee, "Short-Term water demand prediction in residential complexes: case study in columbia city, USA," in *Proceedings of the 2018 11th International Conference on Developments in eSystems Engineering (DeSE)*, pp. 31–35, Cambridge, UK, September 2018.

[42] S. L. Zubaidi, S. Ortega-Martorell, P. Kot et al., "A Method for Predicting Long-Term Municipal Water Demands Under Climate Change," *Water Resources Management*, vol. 34, pp. 1265–1279, 2020.

[43] S. Behboudian, M. Tabesh, M. Falahnezhad, and F. A. Ghavanini, "A long-term prediction of domestic water demand using preprocessing in artificial neural network," *Journal of Water Supply: Research & Technology - Aqua*, vol. 63, no. 1, pp. 31–42, 2014.

[44] S. L. Zubaidi, K. Hashim, S. Ethaib, N. S. S. Al-Bdairi, H. Al-Bugharbee, and S. K. Gharghan, "A novel methodology to predict monthly municipal water demand based on weather variables scenario," *Journal of King Saud University - Engineering Sciences*, vol. 34, no. 3, pp. 163–169, 2022.

[45] H. Hassani, S. Heravi, and A. Zhigljavsky, "Forecasting European industrial production with singular spectrum analysis," *International Journal of Forecasting*, vol. 25, no. 1, pp. 103–118, 2009.

[46] H. Al-Bugharbee and I. Trendafilova, "A fault diagnosis methodology for rolling element bearings based on advanced signal pretreatment and autoregressive modelling," *Journal of Sound and Vibration*, vol. 369, pp. 246–265, 2016.

[47] Q. B. Pham, T.-C. Yang, C.-M. Kuo, H.-W. Tseng, and P.-S. Yu, "Coupling Singular Spectrum Analysis with Least Square Support Vector Machine to Improve Accuracy of SPI Drought Forecasting," *Water Resources Management*, vol. 35, pp. 847–868, 2021.

[48] H. Hassani, A. Webster, E. S. Silva, and S. Heravi, "Forecasting U.S. Tourist arrivals using optimal singular spectrum analysis," *Tourism Management*, vol. 46, pp. 322–335, 2015.

[49] H. Apaydin, M. Taghi Sattari, K. Falsafian, and R. Prasad, "Artificial intelligence modelling integrated with Singular Spectral analysis and Seasonal-Trend decomposition using Loess approaches for streamflow predictions," *Journal of Hydrology*, vol. 600, Article ID 126506, 2021.

[50] J. B. Elsner, "Analysis of time series structure: SSA and related techniques," *Journal of the American Statistical Association*, vol. 97, no. 460, pp. 1207-1208, 2002.

[51] H. R. Maier and G. C. Dandy, "Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications," *Environmental Modelling & Software*, vol. 15, pp. 101–24, 2000.

[52] J Pallant, *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using Spss for Windows*, Open University Press/ McGraw-Hill, London, UK, 2016.

[53] A. Faramarzi, M. Heidarinejad, S. Mirjalili, and A. H. Gandomi, "Marine predators algorithm: a nature-inspired metaheuristic," *Expert Systems With Applications*, vol. 152, 2020.

[54] M. Khan, N. Muhammad, and A. El-Shafie, "Wavelet-ANN versus ANN-based model for hydrometeorological drought forecasting," *Water*, vol. 10, 2018.

[55] M. Y. A. Khan, F. Hasan, S. Panwar, and G. J. Chakrapani, "Neural network model for discharge and water-level prediction for Ramganga River catchment of Ganga Basin, India," *Hydrological Sciences Journal*, vol. 61, no. 11, pp. 2084–2095, 2016.

[56] M. B. Gunathilake, C. Karunanayake, A. S. Gunathilake et al., "Hydrological Models and Artificial Neural Networks (ANNs) to Simulate Streamflow in a Tropical Catchment of Sri Lanka," *Applied Computational Intelligence and So P Computing*, vol. 9, 2021.

[57] X. Huang, Y. Li, Z. Tian et al., "Evaluation of Short-Term Streamflow Prediction Methods in Urban River Basins," *Physics And Chemistry Of the Earth*, vol. 123, Article ID 103027, 2021.
[58] F. Ahmadi, S. Mehdizadeh, and B. Mohammadi, "Development of Bio-Inspired- and Wavelet-Based Hybrid Models for Reconnaissance Drought Index Modeling," *Water Resources Management*, vol. 35, pp. 4127–4147, 2021.
[59] M. M. H. Khan, N. S. Muhammad, and A. El-Shafie, "Wavelet based hybrid ANN-ARIMA models for meteorological drought forecasting," *Journal of Hydrology*, vol. 590, Article ID 125380, 2020.
[60] L. J. Stamenković, "Application of ANN and SVM for prediction nutrients in rivers," *Journal of Environmental Science and Health*, vol. 56, pp. 867–873, 2021.
[61] H. Tahraoui, A.-E. Belhadj, A.-e Hamitouche, M. Bouhedda, and A. Amrane, "Predicting the Concentration of Sulfate (SO42-) in Drinking Water Using Artificial Neural Networks: A Case Study: Médéa-Algeria," *Desalination And Water Treatment*, vol. 217, pp. 181–194, 2021.
[62] S. Zubaidi, H. Al-Bugharbee, S. Ortega-Martorell et al., "A Novel Methodology for Prediction Urban Water Demand by Wavelet Denoising and Adaptive Neuro-Fuzzy Inference System Approach," *Water*, vol. 12, 2020.
[63] Y. Seo, S. Kwon, and Y. Choi, "Short-Term water demand forecasting model combining variational mode decomposition and extreme learning machine," *Hydrology*, vol. 5, no. 4, p. 54, 2018.
[64] C. W. Dawson, R. J. Abrahart, and L. M. See, "HydroTest: A web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts," *Environmental Modelling & Software*, vol. 22, pp. 1034–1052, 2007.
[65] K. Ahmed, S. Shahid, X. Wang, N. Nawaz, and K. Najeebullah, "Evaluation of Gridded Precipitation Datasets over Arid Regions of Pakistan," *Water*, vol. 11, 2019.
[66] E. Eze, S. Halse, and T. Ajmal, "Developing a Novel Water Quality Prediction Model for a South African Aquaculture Farm," *Water*, vol. 13, no. 13, 2021.
[67] M. Pan, H. Zhou, J. Cao et al., "Water Level Prediction Model Based on GRU and CNN," *IEEE Access*, vol. 8, pp. 60090–100, 2020.
[68] T. Ren, X. Liu, J. Niu, X. Lei, and Z. Zhang, "Real-time water level prediction of cascaded channels based on multilayer perception and recurrent neural network," *Journal of Hydrology*, vol. 585, Article ID 124783, 2020.
[69] A. B. Heinemann, P. A. J. Van Oort, D. S. Fernandes, and A. D. H. N. Maia, "Sensitivity of APSIM/ORYZA model due to estimation errors in solar radiation," *Bragantia*, vol. 71, no. 4, pp. 572–582, 2012.
[70] M.-F. Li, X.-P. Tang, W. Wu, and H.-B. Liu, "General models for estimating daily global solar radiation for different solar radiation zones in mainland China," *Energy Conversion and Management*, vol. 70, pp. 139–148, 2013.
[71] B. G. Tabachnick and Fidell, *Using Multivariate Statistics*, pearson education inc, New York, NY, USA, 2013.
[72] N. Sirajudin, "suratno j and pamuti 2021 developing creativity through stem education," *Journal Of Physics: Conference Series*, vol. 1806, Article ID 012211.
[73] M. Valentini, G. B. dos Santos, and B. Muller Vieira, "Multiple linear regression analysis (MLR) applied for modeling a new WQI equation for monitoring the water quality of Mirim Lagoon, in the state of Rio Grande do Sul—Brazil," *SN Applied Sciences*, Springer, New York, NY, USA, 2021.
[74] D. H. Wolpert and W. G. Macready, "No Free Lunch Theorems for Optimization," *IEEE Transactions on Evolutionary Computation*, vol. 1, no. 1, 1997.