Predicting the Degree of Dissolved Oxygen Using Three Types of Multi-Layer Perceptron-Based Artificial Neural Networks

Fen Yang 1, Hossein Moayedi 2,3,* and Amir Mosavi 4,*

1 School of Economics and Management, Beijing University of Technology, Beijing 100124, China; yangf666@126.com
2 Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam
3 Faculty of Civil Engineering, Duy Tan University, Da Nang 550000, Vietnam
4 John von Neumann Faculty of Informatics, Obuda University, 1034 Budapest, Hungary
* Correspondence: hosseinmoayedi@duytan.edu.vn (H.M.); amir.mosavi@mailbox.tu-dresden.de (A.M.)

Abstract: Predicting the level of dissolved oxygen (DO) is an important issue ensuring the sustainability of the inhabitants of a river. A prediction model can predict the DO level using a historical dataset with regard to water temperature, pH, and specific conductance for a given river. The model can be built using sophisticated computational procedures such as multi-layer perceptron-based artificial neural networks. Different types of networks can be constructed for this purpose. In this study, the authors constructed three networks, namely, multi-verse optimizer (MVO), black hole algorithm (BHA), and shuffled complex evolution (SCE). The networks were trained using the datasets collected from the Klamath River Station, Oregon, USA, for the period 2015–2018. We found that the trained networks could predict the DO level of 2019. We also found that both BHA- and SCE-based networks could predict the level of DO using a relatively simple configuration compared to that of MVO. From the viewpoints of absolute errors and Pearson’s correlation coefficient, MVO- and SCE-based networks performed better than BHA-based networks. In synopsis, the authors recommend MVO- and MLP-based artificial neural networks for predicting the DO level of a river.

Keywords: water quality; dissolved oxygen; neural network; machine learning; artificial intelligence; deep learning; big data; data science; hydrological model; water treatment

1. Introduction

1.1. Problem Statement and Background

As is known, acquiring an appropriate forecast for water quality parameters such as dissolved oxygen (DO) is an important task due to their effects on aquatic health maintenance and reservoir management [1]. Constraints such as the influence of various environmental factors on the DO concentration [2] have driven many scholars to replace conventional models with sophisticated artificial intelligent techniques [3–6].

From a general point of view, the objective of many recent developments lies in facilitating complex analysis [7–9]. These efforts have resulted in finding promising approaches for various analyses, such as water treatment [10,11], energy [12,13], construction and waste reduction [14–16], optics-related simulations [17,18], environmental modeling [19–21], and remote sensing [22]. In this sense, different intelligent methods have been devised by experts to give a reliable simulation of engineering problems [23–25].

Going beyond typical soft computing, metaheuristic algorithms have been designed for optimization purposes [26–29]. Most of these techniques rely on the herding behavior of animals [30–33] (e.g., Harris hawks [34,35] and grey wolf [36,37]). When incorporated with conventional soft computing, metaheuristic algorithms aim to optimize the hyperparameters of a base model to achieve optimum configuration [38–41]. Some examples of these algorithms can be found in research by Hu et al. [42], Shen et al. [43], Wang and Chen [44], and Li et al. [45], dedicated to medical diagnosis.
1.2. Similar Works

Through applying support vector regression (SVR), Liu et al. [46] showed the efficiency of the maximal information coefficient technique used for feature selection in the estimation of DO concentration. The results of the optimized dataset were much more reliable (28.65% in terms of root mean square error, RMSE) than the original input configuration. Csábrági et al. [47] showed the appropriate efficiency of three conventional notions of artificial neural networks (ANNs), namely, multi-layer perceptron (MLP), radial basis function (RBF), and general regression neural network (GRNN), for this purpose. Similar efforts can be found in [48,49]. Heddam [50] introduced a new ANN-based model, namely, evolving fuzzy neural network, as a capable approach for DO simulation in a river ecosystem. The suitability of fuzzy-based models has been investigated in many studies [51]. Adaptive neuro-fuzzy inference system (ANFIS) is another potent data mining technique that has been discussed in many studies [52–54]. More attempts regarding the employment of machine learning tools can be found in [55–58].

Ouma et al. [59] compared the performance of a feed-forward ANN with multiple linear regression (MLR) in simulating DO in Nyando River, Kenya. It was shown that the correlation of the ANN is considerably greater than the MLR (i.e., 0.8546 vs. 0.6199). Zhang et al. [60] combined a recurrent neural network (RNN) with kernal principal component analysis to predict hourly DO concentration. Their suggested model was found to be more accurate than regular data mining techniques, including feed-forward ANN, SVR, and GRNN, by around 8%, 17%, and 12%. Additionally, the largest accuracy (the coefficient of determination \( R^2 = 0.908 \)) was obtained for DO in the upcoming hour. Ali et al. [61] combined a so-called denoising method, namely, “complete ensemble empirical mode decomposition with adaptive noise”, with two popular machine learning models, namely, random forest (RF) and extreme gradient boosting, to analyze various water quality parameters. It was shown that the RF-based ensemble is a more accurate approach for the simulation of DO, temperature, and specific conductance. They also proved the viability of the proposed approaches by comparing them with some benchmark tools. Likewise, Ahmed [62] showed the superiority of RF over MLR for DO modeling. He also revealed that water temperature and pH play the most significant roles in this process. Ay and Kiş [63] conducted a comparison among MLP, RBF, ANFIS (sub-clustering), and ANFIS (grid partitioning). The respective \( R^2 \) values of 0.98, 0.96, 0.95, and 0.86 for one station (Number: 02156500) revealed that the outcomes of MLP are better correlated with the observed DOs.

Synthesizing conventional approaches with auxiliary techniques has led to novel hybrid tools for various hydrological parameters [64–66]. Ravansalar et al. [67] showed that linking the ANN with discrete wavelet transform results in an improvement of accuracy (i.e., Nash–Sutcliffe coefficient) from 0.740 to 0.998. A similar improvement was achieved for the SVR applied to estimate biochemical oxygen demand in Karun River, Western Iran. Antanasijević et al. [68] presented a combination of Ward neural networks and a local similarity index for predicting DO in the Danube River. They noted the better performance of the proposed model compared to the multi-site DO evaluative approaches presented in the literature.

1.3. Novelty and Objective

Metaheuristic search methods such as teaching–learning based optimization [69] have provided suitable approaches for intricate problems. Ahmed and Shah [52] suggested three optimized versions of ANFIS using differential evolution, genetic algorithm (GA), and ant colony optimization for predicting water quality parameters, including electrical conductivity, sodium absorption ratio, and total hardness. In similar research, Mahmoudi et al. [70] coupled SVR with the shuffled frog leaping algorithm (SFLA) for the same objective. Zhu et al. [71] compared the efficiency of the fruit fly optimization algorithm (FOA) with the GA and particle swarm optimization (PSO) for optimizing a least-squares SVR for forecasting the trend of DO. Referring to the obtained mean absolute percentage
errors of 0.35%, 1.3%, 2.03%, and 1.33%, the proposed model (i.e., FOA-LSSVR) surpassed
the benchmark techniques. In this work, three stochastic search techniques of multi-verse
optimizer (MVO), black hole algorithm (BHA), and shuffled complex evolution (SCE) are
used to optimize an MLP neural network for predicting DO using recent data collected from
the Klamath River Station. According to Sullivan et al. [72], the reach of interest is classified
as very poor water quality based on the Oregon Water Quality Index. Additionally, the
reach of Keno dam (downstream of the river) is labeled as “water quality limited” for
ammonia and dissolved oxygen year-round, as well as pH and chlorophyll a in summer. It
clearly highlights the importance of water quality assessments in this area. To the best of
the authors’ knowledge, up to now, few metaheuristic algorithms have been used for training
the ANN in the field of DO modeling (e.g., firefly algorithm [73] and PSO [74]). Therefore,
the models suggested in this study are deemed as innovative hybrids for this purpose.

2. Methodology

The steps of this research are shown in Figure 1. After providing the appropriate
dataset, the MLP is submitted to MVO, BHA, and SCE algorithms to adjust its parameters
through metaheuristic schemes. During an iterative process, the MLP is optimized to
present the best possible prediction of the DO.

![Figure 1. The steps taken in this research for predicting DO.](image)

2.1. The MVO

As is implied by its name, the MVO is obtained from multi-verse theory in physics [75].
According to this theory, there is more than one big bang event, each of which has initiated
a separate universe. The algorithm was introduced by Mirjalili et al. [76]. The main
components of the MVO are wormholes, black holes, and white holes. The concepts of
black and white holes run the exploration phase, while the wormhole concept is dedicated
to the exploitation procedure. The pseudo code of the MVO is presented as Algorithm 1.
Algorithm 1. Pseudo code of the MVO [77]

Initialize the parameters (population size, iterations, wormhole existence probability (WEP), travelling distance rate (TDR))

While maximum iteration not reached
    Compute the fitness for each universe
    AU = Sort the population
    BI = Normalize the fitness values
    for i = 2:N
        Black hole = i
        for j = 1:size(i)
            Generate r1 as a random value
            if r1 < BI(U_i)
                white hole = Roulette Wheel Selection(-B)
                U(black hole, j) = AU(white hole, j)
            end if
            Generate r2 as a random value
            if r2 < WEP
                Generate r3 and r4 as a random value
                if r3 < 0.5
                    Update the position of the universe (Equation (4))
                else
                    Update the position of the universe (Equation (4))
                end if
            end if
        end for
    end for
end while
Response = Best solution.

In the MVO, the so-called parameter of “rate of inflation” (ROI) is defined for each universe. The objects are transferred from the universes with larger ROIs to those with lower values to improve the average ROI of the whole cosmos. During an iteration, the organization of the universes is carried out with respect to their ROIs, and after a roulette wheel selection (RWS), one of them is deemed the white hole. In this relation, a set of universes can be defined as:

\[
U = \begin{bmatrix}
    x_1^1 & x_1^2 & \ldots & x_1^g \\
    x_2^1 & x_2^2 & \ldots & x_2^g \\
    \vdots & \vdots & \ddots & \vdots \\
    x_k^1 & x_k^2 & \ldots & x_k^g
\end{bmatrix}
\]  

where \(g\) symbolizes the number of objects and \(k\) stands for the number of universes. The \(jth\) objective in the \(ith\) solution is generated according to the below equation:

\[
x_j^i = \text{lb}_j + \text{rand}()((\text{ub}_j - \text{lb}_j) + 1) \forall i \in (1, 2, \ldots, k) \forall j \in (1, 2, \ldots, g)
\]  

where \(\text{ub}_j\) and \(\text{lb}_j\) denote upper and lower bounds, and the function \(\text{rand}()\) produces a discrete randomly distributed number.

In each repetition, there are two options for the \(x_j^i\): (i) it is selected from earlier solutions using RWS (e.g., \(x_j^i \in (x_1^j, x_2^j, \ldots, x_{i-1}^j)\)) and (ii) it does not change. It can be written as follows:

\[
x_j^i \begin{cases} 
    x_i^m & \text{rand}_1 < \text{Norm}(U_i) \\
    x_j^i & \text{rand}_1 \geq \text{Norm}(U_i)
\end{cases}
\]  

In the above equation, \(U_i\) stands for the \(ith\) universe, \(\text{Norm}(U_i)\) gives the corresponding normalized ROI, and \(\text{rand}_1\) is a random value in \([0, 1]\).
Equation (4) expresses the measures considered to deliver the variations of the whole universe. In this sense, the wormholes are supposed to enhance the ROI.

\[
x_j^i = \begin{cases} 
  x_j + TDR \times ((u_{bjj} - l_{bjj}) \times r_4 + l_{bjj}) & \text{if } r_3 < 0.5 \text{ if } r_2 < WEP \\
  x_j - TDR \times ((u_{bjj} - l_{bjj}) \times r_4 + l_{bjj}) & \text{if } r_3 \geq 0.5 \text{ if } r_2 \geq WEP 
\end{cases}
\]

where \( x_j \) signifies the \( j \)th best-fitted universe obtained so far, and \( r_2, r_3, \) and \( r_4 \) are random values in \([0, 1]\). Moreover, the two parameters of WEP and TDR stand for the wormhole existence probability and the traveling distance rate, respectively. Given \( \text{Iter} \) as the running iteration, and \( \text{Iter}_{\text{max}} \) as the maximum number of \( \text{Iter} \)s, these parameters can be calculated as follows:

\[
WEP = a + \text{Iter} \times \left( \frac{b - a}{\text{Iter}_{\text{max}}} \right)
\]

\[
TDR = 1 - \frac{\text{Iter}^{1/q}}{\text{Iter}_{\text{max}}^{1/q}}
\]

where \( q \) is the accuracy of exploitation, and \( a \) and \( b \) are constant pre-defined values \([78,79]\).

2.2. The BHA

Inspired by black hole incidents in space, Hatamlou \([80]\) proposed the BHA in 2013. Emerging after the collapse of massive stars, a black hole is distinguished by a huge gravitational power. The stars move toward this mass, and it explains the pivotal strategy of the BHA for achieving an optimum response. A randomly generated constellation of stars represents the initial population. Based on the fitness of these stars, the most powerful one is deemed as the black hole that will absorb the surrounding ones. The pseudo code of the BHA is presented as Algorithm 2.

**Algorithm 2. Pseudo code of the BHA \([81]\)**

1. Initialize the stars \( x_i \)
2. Initialize the parameters (fitness function and iterations (\text{Iter}))
3. Select the best-fitted star \((x_b)\) as the black hole (BH)
4. While maximum \text{Iter} not reached
   1. For each \( x_i \)
      1. Compute the fitness
      2. if fitness \((x_i)\) > fitness \((x_b)\)
         1. \( x_b = x_i \)
      end if
   end for
5. Update fitness and compute Equation (8)
6. if \( \sqrt{(x_b - x_i)^2} < R_{BH} \)
   1. Replace \( x_i \) with a new star
end if
7. end while
8. \( \text{Response} = \text{Best solution} \)

In this procedure, the positions change according to the relationship below:

\[
x_i^{\text{Iter} + 1} = x_i^{\text{Iter}} + \text{rand} \times (x_{BH} - x_i^{\text{Iter}}) \quad i = 1, 2, \ldots, Z
\]

where rand is a random number in \([0, 1]\), \( x_{BH} \) is the black hole’s position, \( Z \) is the total number of stars, and \( \text{Iter} \) symbolizes the iteration number.
Once the fitness of a star surpasses that of the black hole, they exchange their positions. In this regard, Equation (8) calculates the radius of the event horizon for the black hole.

\[
R = \frac{F_{BH}}{\sum_{i=1}^{Z} F_i}
\]

where \( F_i \) is the fitness of the \( i^{th} \) star, and \( F_{BH} \) is the value for the black hole [82].

### 2.3. The SCE

Originally proposed by Duan et al. [82], the SCE has been efficiently used for dealing with optimization problems with high dimensions. The SCE can be defined as a hybrid of complex shuffling and competitive evolution concepts with the strengths of the controlled random search strategy. This algorithm (i.e., the SCE) benefits from a deterministic strategy to guide the search. Additionally, utilizing random elements has resulted in a flexible and robust algorithm. The pseudo code of the SCE is presented as Algorithm 3.

**Algorithm 3. Pseudo code of the SCE [83]**

<table>
<thead>
<tr>
<th>Initialize the population (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample the population ( {x_1, x_2, \ldots, x_s} )</td>
</tr>
<tr>
<td>Compute the fitness (f) for each member</td>
</tr>
<tr>
<td>Sort based on the obtained f values</td>
</tr>
<tr>
<td>Create ( D^0 = {x_i, f_i \mid i = 1, 2, \ldots, s} )</td>
</tr>
<tr>
<td>Create complexes W.R.T Equation (9)</td>
</tr>
<tr>
<td>While maximum ( It ) not reached</td>
</tr>
<tr>
<td>for ( q = 1:i )</td>
</tr>
<tr>
<td>Evolve ( C^q ) using Competitive Complex Evolution</td>
</tr>
<tr>
<td>end for</td>
</tr>
<tr>
<td>( D^i = D^{i+1} )</td>
</tr>
<tr>
<td>end while</td>
</tr>
<tr>
<td>Response = Best solution.</td>
</tr>
</tbody>
</table>

The SCE is implemented in seven steps. Assuming \( N_C \) as the number of complexes and \( N_P \) as the number of points existing in one complex, the sample size of the algorithm is generated as \( S = N_C \times N_P \). In this sense, \( N_C \geq 1 \) and \( N_P \geq 1 + \) the number of design variables. Next, the samples \( x_1, x_2, \ldots, x_s \) are created in the viable space (i.e., within the bounds). The fitness values are also calculated using sampling distribution. In the third step, these samples are arranged with reference to their fitness. An array-like \( D = \{x_i, f_i \mid \text{where } i = 1, 2, \ldots, s\} \) can be considered for storing them. This array is then divided into \( N_C \) complexes \( \{C^1, C^2, \ldots, C^{N_C}\} \), each of which contains \( N_P \) samples (Equation (9))

\[
C^q = \{x_j^q, f_j^q \mid x_j^q = x_q + N_C(j - 1), f_j^q = f_q + N_C(j - 1), j = 1, 2, \ldots, N_P\}
\]

In the fifth step, each complex is evolved by the competitive complex evolution algorithm. Later, in a process named the shuffling of the complexes, all complexes are replaced in array \( D \). This array is then sorted based on the fitness values. Lastly, the algorithm checks for stopping criteria that terminate the process [84].

### 2.4. Accuracy Criteria

The quality of the results is lastly evaluated using Pearson’s correlation coefficient \( (R_P) \) along with mean absolute error (MAE) and RMSE. They analyze the agreement and difference between the observed and predicted values of a target parameter. In the present...
work, given $DO_{\text{predicted}}$ and $DO_{\text{observed}}$ as the predicted and observed DOs, the $R_p$, $MAE$, and $RMSE$ are expressed by the following equations:

$$R_p = \frac{\sum_{i=1}^{K} (DO_{\text{predicted}} - \bar{DO}_{\text{predicted}})(DO_{\text{observed}} - \bar{DO}_{\text{observed}})}{\sqrt{\sum_{i=1}^{K} (DO_{\text{predicted}} - \bar{DO}_{\text{predicted}})^2 \sqrt{\sum_{i=1}^{K} (DO_{\text{observed}} - \bar{DO}_{\text{observed}})^2}}}$$  \hspace{1cm} (10)

$$MAE = \frac{1}{K} \sum_{i=1}^{K} |DO_{\text{observed}} - DO_{\text{predicted}}|$$  \hspace{1cm} (11)

$$RMSE = \sqrt{\frac{1}{K} \sum_{i=1}^{K} [(DO_{\text{observed}} - DO_{\text{predicted}})^2]}$$  \hspace{1cm} (12)

where $K$ signifies the number of compared pairs.

3. Data

As a matter of fact, intelligent models should first learn the pattern of the intended parameter in order to predict it. This learning process is carried out by analyzing the dependence of the target parameter on some independent factors. In this work, DO is the target parameter for water temperature (WT), pH, and specific conductance (SC). This study uses the data belonging to a US Geological Survey (USGS) station, namely, the Klamath River (station number: 11509370). As Figure 2 illustrates, this station is located in Klamath County, Oregon State.

Figure 2. Location of the studied USGS station.
Pattern recognition is fulfilled by means of data obtained between 1 October 2014 and 30 September 2018. After training the models, the DO for the subsequent year (i.e., from 1 October 2018 to 30 September 2019) is predicted. Since the models do not know this data, the accuracy of this process will reflect their capability to predict DO in unseen conditions. Hereafter, these two groups are categorized as training data and testing data, respectively. Figure 3 depicts DO vs. WT, PH, and SC for the (a, c, and e) training and (b, d, and f) testing data. Based on the available data for the mentioned periods, the training and testing groups contain 1430 and 352 records, respectively. Moreover, the statistical description of these datasets is presented in Table 1.

![Figure 3](image-url)
Table 1. Statistical indicators of DO and independent factors.

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Train Data</th>
<th></th>
<th>Test Data</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WT (°C)</td>
<td>pH</td>
<td>SC (µS/cm)</td>
<td>DO (mg/L)</td>
</tr>
<tr>
<td>Average</td>
<td>12.20</td>
<td>8.05</td>
<td>160.46</td>
<td>7.74</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>7.38</td>
<td>0.64</td>
<td>39.81</td>
<td>2.78</td>
</tr>
<tr>
<td>Sample Variance</td>
<td>54.43</td>
<td>0.41</td>
<td>1584.75</td>
<td>7.72</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.07</td>
<td>0.48</td>
<td>1.54</td>
<td>−0.54</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.40</td>
<td>6.90</td>
<td>105.00</td>
<td>0.30</td>
</tr>
<tr>
<td>Maximum</td>
<td>25.70</td>
<td>9.60</td>
<td>332.00</td>
<td>14.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>WT (°C)</td>
<td>pH</td>
<td>SC (µS/cm)</td>
<td>DO (mg/L)</td>
</tr>
<tr>
<td>Average</td>
<td>11.54</td>
<td>7.97</td>
<td>174.20</td>
<td>7.59</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>7.43</td>
<td>0.65</td>
<td>45.80</td>
<td>2.69</td>
</tr>
<tr>
<td>Sample Variance</td>
<td>55.23</td>
<td>0.42</td>
<td>2097.97</td>
<td>7.24</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.11</td>
<td>0.96</td>
<td>1.49</td>
<td>−0.57</td>
</tr>
<tr>
<td>Minimum</td>
<td>1.40</td>
<td>7.20</td>
<td>116.00</td>
<td>0.60</td>
</tr>
<tr>
<td>Maximum</td>
<td>23.10</td>
<td>9.60</td>
<td>387.00</td>
<td>13.20</td>
</tr>
</tbody>
</table>

Moreover, the effect of the inputs on DO is investigated using a tree-based ensemble method. To do this, a bagged ensemble of 200 regression trees is implemented, and the outcome is reported as a value of importance. Figure 4 shows the results. As can be seen, SC is smaller than WT and the effect of WT smaller than pH. In other words, pH has the greatest impact on DO concentration in this case.

Figure 4. Importance assessment of the inputs.

4. Results and Discussion

4.1. Optimization and Weight Adjustment

As explained, the proposed hybrid models are designed in the way that MVO, BHA, and SCE algorithms are responsible for adjusting the weights and biases of the MLP. To do this, a raw MLP structure (that wants to predict DO from WT, pH, and SC) should be given as the optimization problem to the mentioned algorithms. In this work, a three-layered MLP with three neurons in the input layer (each for one input), six neurons in the middle layer, and one neuron (for DO) in the last layer was considered, noting that the value of 6 was determined after a trial and error process; this is schematically shown in Figure 5.
The overall formulation of a neuron can be expressed as follows:

\[ R_N = f(I_N \times W + b) \]  

where \( f(x) \) is the activation function used by the neurons in a layer; additionally, \( R_N \) and \( I_N \) denote the response and input of the neuron \( N \), respectively. With a large number of these equations, each algorithm first suggests a stochastic response for the \( W \) and \( b \) values. In the next iterations, the algorithms improve this response in order to build a more accurate MLP. The formulation of a trained MLP network is presented at the end of the study to better illustrate this concept.

The created hybrids are implemented with different population sizes (\( N_{\text{Pop}} \)) of the trainer algorithm to achieve the best results. Figure 6 shows the values of the objective function obtained for the \( N_{\text{Pop}} \)s of 10, 25, 50, 75, 100, 200, 300, 400, and 500. In the case of this study, the objective function is reported by the RMSE criterion. Figure 6 shows that unlike the SCE, which gives more quality training with small \( N_{\text{Pop}} \)s, the MVO performs better with the three largest \( N_{\text{Pop}} \)s. The BHA, however, did not show any specific behavior. Overall, the MVO, BHA, and SCE with the \( N_{\text{Pop}} \)s of 300, 50, and 10, respectively, could adjust the MLP parameters with the lowest error.

Figure 5. The used MLP structure.

Figure 6. The quality of training for different configurations of the MVO, BHA, and SCE.
As stated, metaheuristic algorithms minimize the errors in an iterative process. Figure 7 shows the convergence curve plotted for the selected configurations of the MLP-MVO, BHA-MVO, and SCE-MVO. To this end, the training RMSE is calculated for a total of 1000 iterations. According to Figure 7, the optimum values of the objective function are 1.314816444, 1.442582978, and 1.33041779 for the MLP-MVO, BHA-MVO, and SCE-MVO, respectively. These configurations are applied in the next section to predict DO. Their results are then evaluated for accuracy assessment.

Figure 7. Error reduction carried out by the selected configurations of the MVO, BHA, and SCE.

4.2. Assessment of the Results

Figure 8 shows a comparison between the observed DO rates and those predicted by the MLP-MVO, MLP-BHA, and MLP-SCE for the whole five years. At a glance, all three models could properly capture DO behavior. It indicates that the algorithms have designated appropriate weights for each input parameter (WT, PH, and SC). The results of the training and testing datasets are presented in detail in the following figure.

Focusing on the training results, an acceptable level of accuracy is reflected by the RMSEs of 1.3148, 1.4426, and 1.3304 for the MLP-MVO, MLP-BHA, and MLP-SCE. In this sense, the values of the MAE (0.9687, 1.0931, and 0.9929) confirmed this statement and showed that all three models had understood the DO pattern with good accuracy. By comparison, it can be deduced that both error values of the MLP-MVO are lower than the two other models. Based on the same reason, the MLP-SCE outperformed the MLP-BHA.

Figure 9 depicts the errors obtained for the training data. This value is calculated as the difference between the predicted and observed DOs. The errors of the MLP-MVO, MLP-BHA, and MLP-SCE have the following ranges: \([-4.6396, 4.7003]\), \([-4.4964, 4.9537]\), and \([-4.5585, 4.5653]\).
Figure 8. The predicted and observed DO rates from 1 October 2014 to 30 September 2019.

Figure 9. Training errors for the (a) MLP-MVO, (b) MLP-BHA, and (c) MLP-SCE.
As stated previously, the quality of the testing results shows how successful a trained model can be in confronting new conditions. The data of the fifth year were considered as these conditions in this study. Figure 10 depicts the histogram of the testing errors. In these charts, $\mu$ stands for the mean error, and $\sigma$ represents the standard error. In this phase, the RMSEs of 1.3187, 1.4647, and 1.3085, along with the MEAs of 1.0161, 1.1997, and 1.0122, imply the power of the used models for dealing with strange data. It means that the weights (and biases) determined in the previous section have successfully mapped the relationship between DO and WT, PH, and SC for the second phase. From the comparison point of view, unlike the training phase, the SCE-based hybrid outperformed the MLP-MVO. The MLP-BHA, however, presented the poorest prediction of DO, again.

![Histograms of testing errors](image)

**Figure 10.** The histogram of the testing errors for the (a) MLP-MVO, (b) MLP-BHA, and (c) MLP-SCE.

The third accuracy indicator (i.e., the $R_P$) shows the agreement between the predicted and observed DO rates. This index can range within $[-1, +1]$, where $-1 (+1)$ indicates a totally negative (positive) correlation, and 0 means no correlation. Figure 11 shows a scatterplot for each model containing both training and testing results. As can be seen, all outputs are positively aggregated around the best-fit line (i.e., the black line). For the training results, the $R_P$s of 0.8808, 0.8545, and 0.8778 indicate the higher consistency of the MLP-MVO results, while the values of 0.8741, 0.8453, and 0.8775 demonstrate the superiority of the MLP-SCE in the testing phase.
4.3. Cross-Validation

In order to further verify the potential of the implemented models, the trained models (i.e., the MVO, BHA, and SCE, with the N_{Pop}s of 300, 50, and 10, respectively) are applied to another station called Fanno Creek (station number: 14206950). Similar to Klamath River, it is located in the Oregon State but in the northern part (longitude: 122°45′13″ W and latitude: 45°24′13″ N). The models predicted the DO for the water year 2019 using the daily records of WT, SC, and pH. Notably, the measured DO ranged from 5.40 to 13.20 mg/L in this site.

The results are shown in Table 2. As can be seen, the performance of all three models is satisfying, and the predicted values are well-correlated with the observed DO rates. It professes that the models can be used for new prediction cases. However, the combination of inputs (i.e., taking WT, pH, and SC into account) should be respected for all study cases.

**Table 2.** The results of cross-validation.

<table>
<thead>
<tr>
<th>Model</th>
<th>RMSE</th>
<th>MAE</th>
<th>RP</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLP-MVO</td>
<td>0.7090</td>
<td>0.5767</td>
<td>0.9413</td>
</tr>
<tr>
<td>MLP-BHA</td>
<td>0.6980</td>
<td>0.5686</td>
<td>0.9430</td>
</tr>
<tr>
<td>MLP-SCE</td>
<td>0.6989</td>
<td>0.5686</td>
<td>0.9431</td>
</tr>
</tbody>
</table>
4.4. Further Discussion

Due to the water quality situation in the Klamath River, this study was dedicated to suggesting novel predictive tools for analyzing the DO concentration in this reach. Notably, this river is important for irrigation, flows, and hydropower generation [72]. It was discussed that DO should be evaluated in response to different key parameters. Another difficulty is that the DO does not follow a certain pattern over time. Hence, it is necessary to tackle these complexities by selecting appropriate models.

Many scholars have stated their tendency regarding the use of hybrid methods for similar issues (e.g., sediment concentration [85] and salinity [86] predictions). The reason for the development of such models can be the use of an optimization technique in the position of a trainer algorithm. In the case of DO modeling in this study, each of the MVO, BHA, and SCE algorithms drove an MLP neural network. In other words, their specific solutions were employed to analyze the relationship between DO and influential parameters through a neural framework.

The high level of accuracy achieved shows that DO can be promisingly predicted. However, there are some suggestions that should be applied for even more efficient solutions. The first part is related to methodology. Optimization using metaheuristic techniques needs to be carefully monitored to select appropriate parameters. For example, the number of iterations and the complexity of the population are better optimized vs. time, where time is as important as accuracy. However, the focus of this study was mostly on the accuracy of prediction (Figure 6). Another idea in this regard can be to test different metaheuristic techniques to find the quickest [87]. Another potential idea would be conducting comparisons with benchmark machine learning solutions such as the ANFIS and SVR as well as their hybridized versions.

The dataset is also a key item. The factors that can be regarded are optimizing the number of input parameters, selecting the appropriate time period, and the pre-processing and purification of misleading samples, among others. In this work, a valid 5-year dataset with three inputs was used, and the results showed that it could provide sufficient samples to be analyzed by the algorithms. Since the models could predict the DO of the testing period without prior knowledge of it, they can be used for further unseen events. Additionally, these results were achieved with the effect of only three influential parameters. Such simplicity is effective for avoiding complicated simulations and reducing the cost of computations.

4.5. SCE-Based Formula

In this section, the formulation of the MLP-SCE model is presented due to its superiority in the prediction phase as well as a simpler configuration compared to the MVO and BHA. Moreover, considering time efficiency, the SCE could optimize the ANN in a meaningfully shorter time. The elapsed times were approximately 5980, 675, and 531 s for the selected configurations of the MLP-MVO, MLP-BHA, and MLP-SCE, respectively (under a core i7 (at 1.8 GHz) operating system with 16 gigs of RAM).

This formula can serve as a predictive equation that estimates DO for given values of WT, pH, and SC. This model predicts DO based on Equation (14), in which Input stands for three inputs, and $LW$ and $b1$ are their corresponding weights and bias vectors, respectively. $LW$ and $b2$ symbolize the same values but for the output layer (Figure 5). These numbers are optimally tuned by the SCE for the MLP so that the lowest training error is achieved (Figure 7). Additionally, $f(x)$ (i.e., the activation function) is presented in Equation (20). It is worth noting that due to the neural network mechanism, this formula must be fed with normalized data.

$$ DO = [LW] \times (f([(IW] \times [Input]) + [b1])) + [b2] $$ (14)

$$ LW = [\begin{array}{c}
-0.0317 \\
0.4963 \\
0.6755 \\
0.4120 \\
0.5135 \\
0.0709 \\
\end{array}] $$ (15)
\[ IW = \begin{bmatrix} -1.9543 & 0.0595 & -1.6276 \\ -0.6330 & 2.2098 & 1.0898 \\ -1.0687 & -1.5362 & 1.7233 \\ -1.5501 & -0.3963 & 1.9778 \\ -2.1836 & -0.9834 & 0.8583 \\ -0.7249 & 1.8623 & -1.5742 \end{bmatrix} \quad (16) \]

\[ Input = \begin{bmatrix} WT \\ pH \\ SC \end{bmatrix} \quad (17) \]

\[ b_1 = \begin{bmatrix} 2.5440 \\ 1.5264 \\ 0.5088 \\ -0.5088 \\ -1.5264 \\ -2.5440 \end{bmatrix} \quad (18) \]

\[ b_2 = [-0.5142] \quad (19) \]

\[ f(x) = \frac{2}{1 + e^{-2x}} - 1 \quad (20) \]

5. Conclusions

This research points out the suitability of metaheuristic strategies for analyzing the relationship between DO and three influential factors (WT, PH, and SC) through the principles of a multi-layer perceptron network. The algorithms used were multi-verse optimizer, black hole algorithm, and shuffled complex evolution, which showed high applicability for optimization objectives. A finding of this study was that while the MVO needs \( N_{\text{Pop}} = 300 \) to give proper training to the MLP, two other algorithms can do this with smaller populations (\( N_{\text{Pop}} \)'s of 50 and 10). According to the findings of the training phase, the MVO can achieve a more profound understanding of the mentioned relationship. The RMSE of this model was 1.3148, which was found to be smaller than MLP-BHA (1.4426) and MLP-SCE (1.3304). However, different results were observed in the testing phase. The SCE-based model came up with the largest accuracy (the RPs were 0.8741, 0.8453, and 0.8775). All in all, the authors believe that the tested models can serve as promising ways for predicting DO. However, assessing other metaheuristic techniques and other hybridization strategies is recommended for future studies.

**Author Contributions:** F.Y., writing—review and editing, visualization, supervision; H.M., methodology, software validation, writing—original draft preparation; A.M., writing—review and editing, visualization, supervision, project administration. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.
Abbreviations

DO Dissolved oxygen
SVR Support vector regression
ANN Artificial neural networks
MLP Multi-layer perceptron
RBF Radial basis function
GRNN General regression neural network
ANFIS Adaptive neuro-fuzzy inference system
GA Genetic algorithm
SFLA Shuffled frog leaping algorithm
FOA Fruit fly optimization
PSO Particle swarm optimization
MVO Multi-verse optimizer
BHA Black hole algorithm
SCE Shuffled complex evolution
MLR Multiple linear regression
RNN Recurrent neural network
ROI Rate of inflation
RWS Roulette wheel selection
R\textsubscript{P} Pearson’s correlation coefficient
MAE Mean absolute error
RMSE Root mean square error
WT Water temperature
SC Specific conductance

References


9. Zhao, C.; Li. Equilibrium Selection under the Bayes-Based Strategy Updating Rules. *Symmetry* 2020, 12, 739. [CrossRef]


40. Shan, W.; Qiao, Z.; Heidari, A.A.; Chen, H.; Turabieh, H.; Teng, Y. Double adaptive weights for stabilization of moth flame optimizer: Balance analysis, engineering cases, and medical diagnosis. *Knowl.-Based Syst.* 2021, 214, 106728. [CrossRef]


49. Heddam, S. Simultaneous modelling and forecasting of hourly dissolved oxygen concentration (DO) using radial basis function neural network (RBFNN) based approach: A case study from the Klamath River, Oregon, USA. *Model. Earth Syst. Environ.* 2016, 2, 1–18. [CrossRef]


60. Zhang, Y.-F.; Fitch, P.; Thorburn, P.J. Predicting the Trend of Dissolved Oxygen Based on the kPCA-RNN Model. *Water* 2020, 12, 585. [CrossRef]


