IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 7, 2021, accepted February 25, 2021, date of publication March 4, 2021, date of current version March 17, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3064029

Prediction of Dissolved Oxygen Content in Aquaculture Based on Clustering and Improved ELM

SHOUQI CAO, LIXIN ZHOU, AND ZHENG ZHANG
College of Engineering Science and Technology, Shanghai Ocean University, Shanghai 201306, China
Shanghai Engineering Research Center of Marine Renewable Energy, Shanghai 201306, China

Corresponding author: Zheng Zhang (z-zhang@shou.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2019YFD090080, and in part by the Shanghai Engineering Research Center of Marine Renewable Energy under Grant 19DZ2254800.

ABSTRACT In the aquaculture industry, dissolved oxygen is an important water quality parameter index. The prediction of dissolved oxygen can reduce the operation cost of aquatic product management to a certain extent. In this paper, a hybrid method is proposed to predict the change of dissolved oxygen from the perspective of time series in aquaculture, which based on k-means clustering and improved Softplus extreme learning machine (SELM) with particle swarm optimization (PSO). We use k-means algorithm to divide the dataset into several clusters by calculating the similarity among variables, to find the periodic change rule and trend of variables. Softplus is employed as the activation function of ELM to make the model closer to the biological activation model. Meanwhile, partial least square (PLS) method is utilized to solve the strong collinearity among variables. In addition, we introduce PSO algorithm to optimize the model parameters. The experimental results show that our model can achieve better prediction performance and accuracy of prediction compared with other single models. Compared with the counterpart model, the improved model can tolerate some data loss and uncertain outliers of sensor time series. Our work provides an accurate predictive model framework for researchers to track dissolved oxygen.

INDEX TERMS Dissolved oxygen, prediction, clustering, extreme learning machine, particle swarm optimization.

I. INTRODUCTION

In aquaculture, dissolved oxygen (DO) has become an important parameter to predict water quality [1]. The excessive or insufficient of dissolved oxygen content in water have an impact on the metabolism and other physiological functions of breeding organisms, and even seriously affect the normal growth of organisms [2], [3]. However, dissolved oxygen is easily affected by many factors [4] such as weather, water quality, human activities and has the characteristic of nonlinear, large inertia, strong coupling and time-varying [5]–[7]. It has become the key content of agricultural production to strengthen the research on key technologies of dissolved oxygen prediction [8] to improve the ability of aquaculture disaster reduction and prevention, and ensure the safety production of aquaculture [9].

The associate editor coordinating the review of this manuscript and approving it for publication was Xi Peng.

Prediction of water quality change is a daunting task, involving multi parameter and dynamic delay process and it is difficult to describe its model by mechanism method with simple mathematical formula or transfer function. Also, the performance of the sensor is often reduced by environmental factors, resulting in data loss or uncertain outliers, which challenges the prediction accuracy. We need to utilize the existing data to establish DO prediction model that can grasp the change trend of DO in a timely manner, and tolerate some degree of sensor data quality problems.

Among these existing methods to solve above two problems, Support Vector Machine (SVM) and Artificial Neural Network (ANN) are two widely used water quality prediction methods [10]. To increase prediction accuracy, a new hybrid dissolved oxygen content forecasting model based on the radial basis function neural networks (RBFNN) data fusion method and a least squares support vector machine (LSSVM) was developed by Yu *et al.* [11]. But the SVM is vulnerable to

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

data loss, long training time and poor prediction performance, so it is difficult to select suitable kernel function for various applications. ANN can tolerate a certain degree of data missing [12]. Faruk [10] proposed a combination of seasonal ARIMA model and neural network back propagation model for monthly prediction of water quality parameters. However, due to the complexity of network topology and data, neural network models often have the problems of over fitting, poor stability and time-consuming.

Motivated by above challenges, considering the large area and wide range of water environment monitoring scenarios, this paper uses the existing Internet of things monitoring information, big data analysis and artificial intelligence algorithm to accurately predict the future trend of dissolved oxygen in water, and make reasonable management decisions in time. In this case, we mainly study how to predict dissolved oxygen efficiently and accurately in a certain time domain under the influence of multiple factors. To achieve this goal, a multi-scale environmental data acquisition scheme is adopted to analyze the characteristics of dissolved oxygen time series, and a single factor prediction model of dissolved oxygen time series is constructed. This paper proposes a scheme based on similar day clustering, and combines the Softplus extreme learning machine improved by partial least squares method with adaptive particle swarm optimization algorithm to predict dissolved oxygen in aquaculture.

The remainder of the paper is organized as follows. In section 2, we review some related work. In Section 3, research areas and data sources are described. We demonstrate the accuracy and efficiency of the forecasting results and discuss the differences among the models in Section 4. In Section 5, concludes the paper and mentions the future work.

II. RELATED WORK

Currently, water quality prediction methods can be divided into two main types: one is the traditional prediction method based on classical mathematical theory; the other is the intelligent prediction method based on modern computational intelligence [13].

Traditional methods mainly include Markov method, regression analysis, time series analysis and function model prediction method. The Markov method is essentially a prediction method based on random processes, which describes the state transition trend of a complex system according to the state and time parameters of the system [14]. Yue and Li [15] introduced the concept of level eigenvalues in the fuzzy set theory into the Markov model, so that it can be used for quantitative prediction of water quality. This method does not require a large number of data sets, and can predict its future short-term change trend of water quality based on historical data [16]. Regression analysis is used for statistical analysis of historical data, and the functional relationship between historical water quality data and prediction data is established by regression equation [17]. The advantages of the model are that the prediction model is simple, fast and has good short-term prediction effect. But this method requires more historical data as a sample set, and has higher requirements for the distribution of water quality changes [18], [19]. The time series method [20] analyzes the change trend of the time series, selects the appropriate algorithm to establish the prediction model, and predicts the future change trend of water quality [21]. Arya and Zhang [22] applied the time series analysis method to model and predict the univariate dissolved oxygen and temperature time series for the four water quality assessment stations of the Stilaguamish River. The results show that the univariate water quality time series of the basin has three different structures. The algorithm of this paper is relatively intuitive, and the predicted value can be obtained quickly. Its disadvantage is that the precision is not high and the new trend of water environment parameters with time is not considered. The function model [23] summarizes the interrelationships among various factors in the changes of water quality parameters by mathematical methods, and describe the complex water environment system with a set of appropriate mathematical equations [24], [25]. This method can well reflect the dynamic trend of water quality parameters, and obtaining all parameters will improve the prediction accuracy. However, the model needs more parameters, and the calculation of the model is time-consuming. When the parameters are not available, the prediction accuracy of the model will be reduced.

Scholars have done a lot of investigations on prediction methods. Nevertheless, the variation of water quality parameters is more complicated. Traditional methods can no longer meet the requirement of aquaculture fine management. Intelligent optimization algorithms can better solve the above problems. Therefore, we need to develop predictive algorithms that can apply intelligent computing to aquaculture.

Grey model (GM) [26], artificial neural network (ANN) [27] and support vector machine (SVM) [28]. These methods have made many achievements and have been widely used in the field of water quality parameters prediction. In the work [29], the GM uses a small amount of sample information to establish a prediction model based on differential equations to make fuzzy long-term predictions of the changing trends of water quality parameters. The prediction accuracy of this method is related to the change of historical data. When the change is large, the prediction accuracy of the model will decrease [30]. In the work [31], the artificial neural network simulates the information processing mode of the human brain, which is suitable for solving complex nonlinear regression estimation, and has a good prediction effect on water quality parameters. Gomolka *et al.* [32] solve the problem of water quality control in the river by using the characteristics of classical ANN, and the interpolation of river ecological state parameters which are difficult to measure is realized. This method can perform autonomous learning and knowledge reasoning, has strong fitting ability and strong robustness. However, the prediction effect of the limited sample set is poor, and it is easy to fall into the local minimum. The SVM is based on the structural risk

minimization and VC dimension theory in the work [33]. It has good prediction accuracy for small sample sets and high generalization ability. Liu *et al.* [34] construct an online prediction model of dissolved oxygen based on LSSVR using SVM, which can adaptively select the best sub model as the online prediction model. This method can solve the problems of under learning, over learning, and local minimum in other algorithms, and obtain the global optimal solution. Unfortunately, the two parameters of the SVM have great influence on the prediction accuracy of the model, and there is no authoritative method to select these two parameters.

The above methods can realize the prediction of water quality parameters to a certain extent. But these methods are not effective in analyzing the impact of changing environment on dissolved oxygen. Most of existing methods cannot determine the key characteristics of the water environment, thus resulting in poor generalization [35]. Due to the aquaculture environment is a dynamic, non-linear and complex system [36], the dissolved oxygen in the water is affected by various factors. It is necessary to consider the influence of changes in various factors on the dissolved oxygen in the dynamic environment. Actually, each feature has its advantages and disadvantages on describing dissolved oxygen contents [37]. Comprehensive analysis of dissolved oxygen related characteristics, we can achieve better predictive performance. Extreme learning machine (ELM) is a new neural network training model proposed by Huang *et al.* [38]. Compared with the traditional ANN model, it has the characteristics of simple network structure and high operation efficiency [39]. Huan *et al.* [4] proposed a new statistical variable similarity based on Pearson correlation analysis. The results show that the historical data can be divided into several clusters by the k-means method, and various potential relationships can be found. In fact, under the similar environment, the daily trend of DO also has convergent changes. Therefore, effective classification of test samples can optimize training samples, thus improving the accuracy of the dissolved oxygen prediction model. This method is of practical significance and can be used for reference.

In this research, a new hybrid method is proposed to achieve accurate prediction of dissolved oxygen. Some properties of our work that differ from the previous ones were summarized as follows:

(1) Based on the analysis of the time series characteristics of DO, a similarity clustering method combining Euclidean distance and cosine angle is defined to divide the monitoring data into different clusters accurately and effectively.

(2) Softplus function introduced into ELM network to replace Sigmoid activation function to solve the nonlinear and continuous problems of time series data. Meanwhile, partial least squares (PLS) method is employed to eliminate the redundancy among variables, which is rarely achieved in other work.

(3) PLS-SELM model based on particle swarm optimization is proposed for short-term DO forecasting, which takes advantages of PSO to search global optimal solution for fast deal with the variable relationship within the network. The single prediction models (SVR, BP, LSTM) and corresponding models (PLS-SELM, SELM and ELM) are compared and analyzed. The efficiency and accuracy of the scheme in the actual dissolved oxygen prediction are verified.

III. MATERIALS AND METHODS

A. TEST AREA AND DATA SOURCE

The aquaculture environment data comes from the remote wireless monitoring system for aquaculture developed by Shanghai Ocean University as shown in Figure 1. The monitoring system includes three layers, namely the perception layer, the transmission layer and the application layer. In the data perception layer, multi-scale sensors are used to collect water quality parameters and meteorological parameters. In the application layer, we process the aquaculture parameter data transmitted from the perception layer through the transmission layer.

All data were obtained from the Guangming breeding base located in Chongming Island, Shanghai. The water resources in aquaculture area are extremely rich and the hydration factors such as petroleum and heavy metals in the water meet the requirements of fishery water quality standards. The total area of Guangming crab breeding base 280 acres. The depth of the pond was 1.8m, and the DO sensor, pH sensor, and water temperature sensor are all placed 1.0m underwater. All these data have been transmitted to the wireless sensor monitoring system.

Large-scale environmental data has been recorded from February to March. A total of 5760 data records were collected over 60 days from 1 February to 31 March in 2020 to predict the dissolved oxygen in 07 Apr, 08 Apr, 09 Apr, 10 Apr and 11 Apr 2020, respectively. The data collection interval of the system is 15 minutes, and the system can obtain 96 pieces of data each day. Since the monitoring data has the nature of daily cycle, the daily data set is defined as a standard unit as a sample. Among 60 samples, the ratio of training set to test set is 7:3. After the dissolved oxygen prediction model is trained on the training set, the test set is used to verify its performance. The water quality of aquaculture is mainly affected by dissolved oxygen, pH, temperature, humidity, air pressure, illumination, wind speed, wind direction and other physical and chemical factors. The seven impact factors and DO values of the historical days constitute the input vector of the prediction model, and the result of the model is the dissolved oxygen value of the prediction day, as shown in Figure 2.

B. PROPOSED METHOD

The function flow chart of the prediction model in this paper is illustrated in Figure 3, is a non-linear predictive model. The specific steps involved in the prediction process are as follows.

Step 1: Data preprocessing. In aquaculture, the data collected by IoT sensors are usually affected by

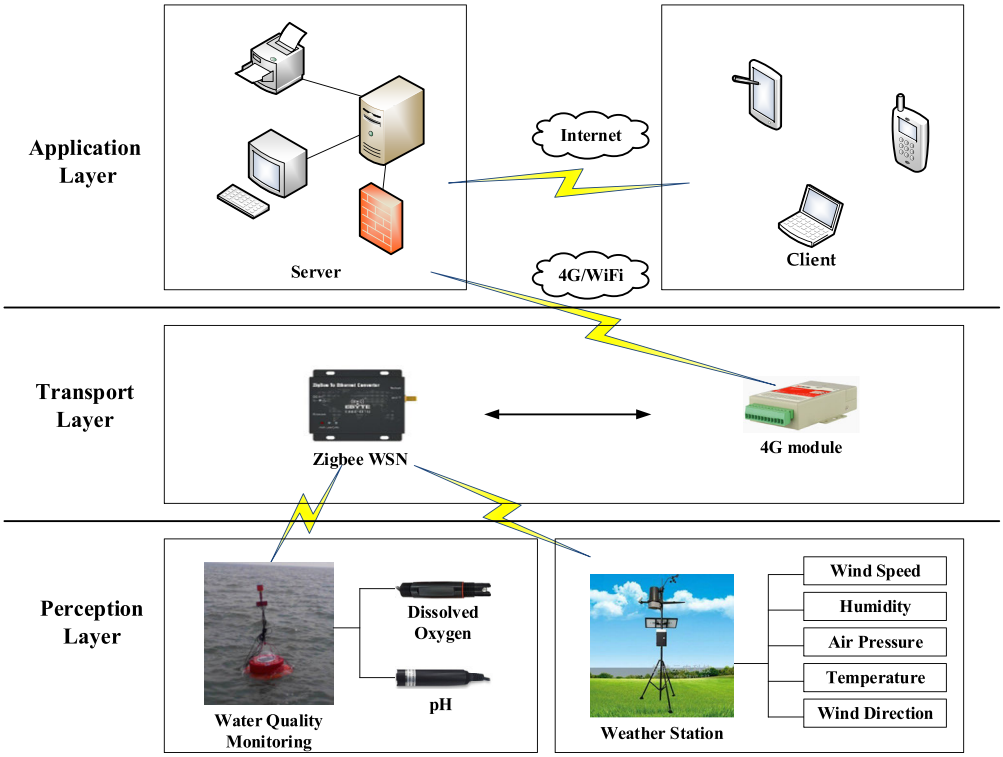


FIGURE 1. The IoT monitoring system for collecting aquaculture data.

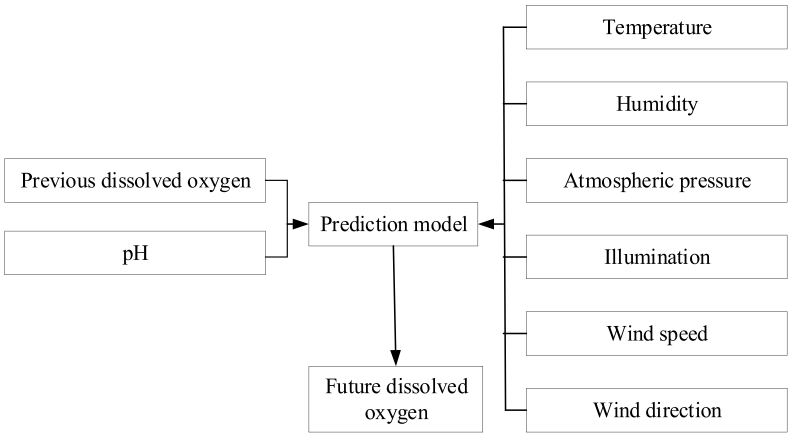


FIGURE 2. Input-output structure of forecasting model.

harsh environment. These factors lead to inaccurate data or inconsistent data formats. Therefore, it is necessary to preprocess the data.

Step 2: Selection of key factors. After data normalization, Pearson correlation coefficient is introduced to determine the weight of environmental factors, and the statistics of similar days are constructed, namely statistical degree. The factors with large statistics will be used as input parameters of the model.

Step 3: K-means clustering. K-means clustering method is applied to cluster historical day data samples according to the similarity, and the appropriate samples are selected. The historical day data samples are divided into several clusters. The samples with the largest similarity are used as the clusters of prediction days to form training samples.

Step 4: Establishment of PLS-SELM Neural Network. Softplus function is utilized as the activation function of the network, and PLS is used to establish the linear relationship between the output matrix and the hidden layer.

Step 5: PSO improves the model parameters. The PSO-PLS-SELM neural network model is established by training samples, and the model is verified by test samples. Finally, the final predicted value is obtained by compensation and denormalization.

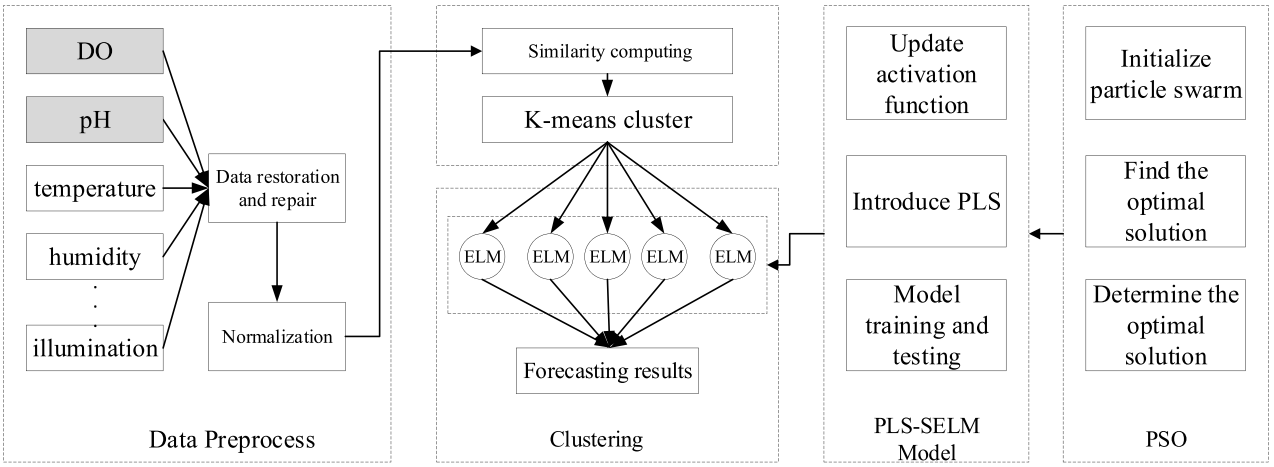


FIGURE 3. The detail process of dissolved oxygen content forecasting.

C. DATA PREPROCESSING

Due to the influence of equipment failure, poor network, bad weather and human factors, the raw data collected by sensors will inevitably be lost or abnormal. This kind of data is called ''dirty data'', which will increase the costs and time of data processing. In addition, the data series of different dimensions will also affect the final predictive performance of the model [40], [41]. Therefore, it is necessary to perform the following processing on the data before conducting model research.

For missing data with short time interval before and after, the linear interpolation method is used to repair it.

$$x_{a+i} = x_a + \frac{i \cdot (x_{a+j} - x_a)}{j} \quad 0 < i < j \tag{1}$$

In formula (1), x_{a+i} is the missing data at time $a + i$, x_a and x_{a+j} are the original data at time a and $a+j$, respectively. There are a large number of missing data or the time interval is long, the data of the same weather type in adjacent days should be used to supplement.

For abnormal data, mean smoothing method is used to solve the step change of water quality parameters in a short time, as shown in formula (2).

$$x_b = \frac{x_{b-1} + x_{b+1}}{2} \quad |x_b - x_{b-1}| > \theta_1 \ \ |x_b - x_{b+1}| > \theta_2 \tag{2}$$

where, x_b is abnormal data, x_{b-1} and x_{b+1} represent its adjacent valid data respectively. Also θ_1 and θ_2 are the error thresholds of adjacent data, respective.

In particular, with the same meteorological conditions, the parameter values of water quality factors on adjacent dates fluctuate less. If the variation range of water quality parameters at the same time on the second day increases or decreases by 20% compared with the previous day, it is considered that the data is abnormal [42]. The mean value method was used to deal with the problem, as shown in formula (3).

$$x_{(d,k)} = \begin{cases} \overline{x_k}' + \theta_3 \\ \overline{x_k}' - \theta_3, \end{cases} \quad \left| x_{(d,k)} - \overline{x_k}' \right| > \theta_3 \tag{3}$$

where, $x_{(d,k)}$ is the water quality sample collected at k time on the day d, θ is the error threshold of the same weather condition at the same time, and $\overline{x_k}'$ is the average value of water quality parameters at the same time in adjacent days with similar weather conditions.

To eliminate different dimensions and quantities of the environmental factors, formula (4) is used for data standardization.

$$\overline{x_k} = \frac{x_k - x_{\min}}{x_{\max} - x_{\min}} \tag{4}$$

where x_k and $\overline{x_k}$ are the original data and the normalized data respectively, $x_{\min}$ and $x_{\max}$ are minimum and maximum values of the original data.

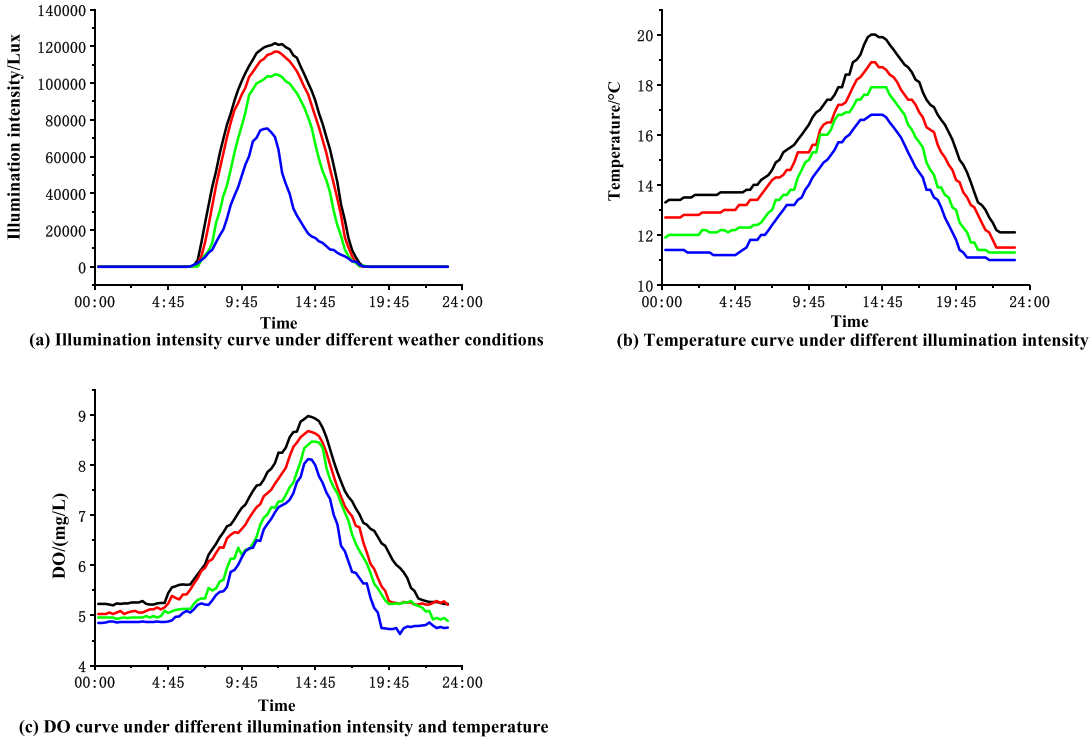
D. DEFINING THE WEIGHTS OF THE ENVIRONMENTAL FACTORS FOR DO

Obviously, weather conditions have the greatest impact on dissolved oxygen. The reason is that different weather conditions result in different solar radiation intensity, leading to differences in plant photosynthesis. To visualize the data, the weather is quantified as illumination intensity. Figure 4(a) is the change curve of different illumination intensity under different weather conditions. Figure4(b) is the change curve of different temperature corresponding to different illumination intensity. Figure 4(c) shows the changes of dissolved oxygen in different weather conditions. Affected by illumination intensity and temperature, it can be seen that dissolved oxygen changes with the same trend as the two changes. And it can be seen that the dissolved oxygen curve changes greatly with different weather conditions. This difference is not only reflected in the changing trend of dissolved oxygen, but also in the output concentration value.

When multiple environmental factors coexist, dissolved oxygen will be affected by many factors such as weather condition, wind speed, temperature, humidity, air pressure, pH value, wind direction and other factors. If all the collected data are taken as input, the model will become very complex, the prediction time will be longer, and the prediction effect

TABLE 1. Correlation coefficients among DO and influencing factors.

DO	Correlation Coefficients						
	pH	Temperature	Humidity	Wind Direction	Wind Speed	Air Pressure	Illumination
σ	0.2280	0.2641	-0.4295	0.1870	0.1958	-0.1992	0.5592



(a) Illumination intensity curve under different weather conditions

(b) Temperature curve under different illumination intensity

(c) DO curve under different illumination intensity and temperature

FIGURE 4. Curves of influencing factors under different weather conditions.

will be poor. In this paper, the distance analysis method of Pearson similarity is used to identify the correlation between variables.

We input 5760 data into the formula for the Pearson correlation coefficient, as shown in formula (5), and calculate the correlation coefficient between dissolved oxygen and environmental factors. The results are shown in Table 1.

$$\sigma_{xy} = \frac{\sum_{i=1}^{m}(x_i - \overline{x_i})(y_i - \overline{y_i})}{\sqrt{\sum_{i=1}^{m}(x_i - \overline{x_i})^2(y_i - \overline{y_i})^2}} \tag{5}$$

where x and y are vectors with a size of $1 * m$, the i-th element of vectors x and y are x_i and y_i respectively, $\overline{x_i}$ and $\overline{y_i}$ are the average values of the elements in vectors x and y, respectively.

Table 1 shows the different correlations between dissolved oxygen content and various environmental factors. Therefore, we determined the weight of these environmental factors on DO.

E. CLUSTERING OF SIMILAR DAYS

To avoid the problems of low accuracy and slow convergence caused by the introduction of bad samples, this paper proposes to use K-means clustering to select similar day with highly similar characteristics to the prediction day, and then establish ELM model for dissolved oxygen prediction.

1) K-MEANS CLUSTERING

Based on a given clustering objective function, the K-means algorithm adopts iterative updating method. Each iteration process is carried out in direction of reducing the objective function [43]. The final clustering result makes the objective function obtain a minimum value and achieve a better classification effect. As proposed by Peng *et al.* [44], the purpose of data clustering is to group a collection of samples into different clusters by simultaneously minimizing intercluster similarity and maximizing intracluster similarity. And the main idea of k-means algorithm is shown in Figure 5. And the iterative process described as follows.

First, k points are randomly selected as the initial clustering center. Second, the distance between each data and each cluster center is calculated, and the data is divided into the nearest cluster center. Third, the location of k centers are determined according to the data class calculated in the previous step. Fourth, if the objective function does not converge, repeat the process from step 2. If there is no change, proceed to step 5. Fifth, the iteration is completed and the algorithm is ended.

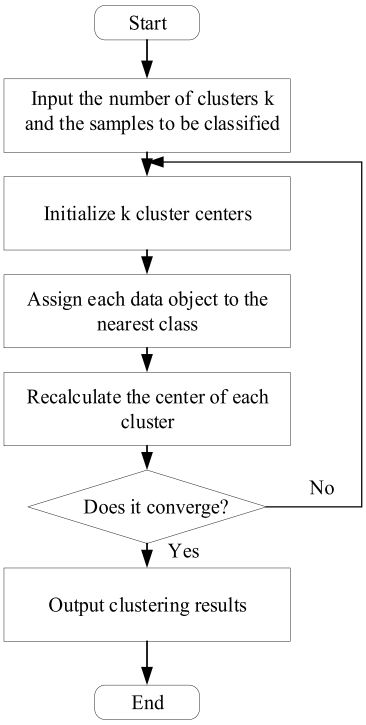


FIGURE 5. Flow chart of K-means algorithm.

In this paper, the sample set D is composed of n data to be classified, namely, $D = \{x_1, \ldots, x_n\}$.The monitoring data of 96 time periods per day constitute a sample unit. Also, each sample has m indicators to represent its attributes, so the data matrix of each sample can be obtained as follows.

$$x_i = \begin{bmatrix} x_{11i} & x_{12i} & \cdots & x_{1mi} \\ x_{21i} & x_{22i} & \cdots & x_{2mi} \\ \vdots & \vdots & & \vdots \\ x_{t1i} & x_{t2i} & \cdots & x_{tmi} \end{bmatrix} \tag{6}$$

where, x_i is the sample on day i, m is the characteristic factor, t is the monitoring time, and $t = 96$.

The following pseudo code is used to describe the process of the K-means algorithm 1:

Algorithm 1 Function of K-Means

Input: sample set $D\{x_1, \ldots, x_n\}$; number of clusters k

1. Randomly select k samples from D as the initial mean vector $\{\mu_1, \ldots, \mu_k\}$
2. Repeat
3. Let $C_i = \emptyset (1 \le i \le k)$
4. **For** $j = 1, 2, \ldots, n$ **do**
5. Calculate the distance between sample x_j and each mean vector $\mu_i (1 \le i \le k) : d_{ji} = \left\| x_j - \mu_i \right\|_2$
6. Determine the cluster label of x_j according to the nearest mean vector: $\lambda_j = \arg\min_{i \in 1,2,\ldots,k} d_{ji}$
7. Divide sample x_j into corresponding cluster: $C_{\lambda_j} = C_{\lambda_j} \cup \{x_j\}$
8. **End for**
9. **For** $i = 1, 2, \ldots, k$ **do**
10. Calculate the new mean vector: $\mu_i' = \frac{1}{|C_i|} \sum_{x \in C_i} x$
11. **If** $\mu_i' \neq \mu_i$ **then**
12. Update the current mean vector μ_i to μ_i'
13. **Else**
14. Keep the current mean vector unchanged
15. **End if**
16. **End for**
17. Until none of the current mean vectors are updated

Output: cluster division $C = \{C_1, \ldots, C_k\}$

2) IMPROVEMENT OF CLUSTER SIMILARITY STATISTICS

In the process of data analysis and data mining, it is often necessary to understand the difference among individuals, and then evaluate the similarities and cluster of individuals. Euclidean distance and angle cosine are commonly used to measure the similarity between two samples in clustering analysis. Suppose that the two individuals to be compared are dissolved oxygen X and any influence factor Y. They all contain t-dimensional features, namely, $X = (x_1, x_2, \ldots, x_t)$, $Y = (y_1, y_2, \ldots, y_t)$.

Euclidean distance measures the absolute distance between points in multidimensional space. Reflect the numerical difference of the sample. The formula is as follows.

$$dist(X, Y) = \sqrt{\sum_{i=1}^{t} (x_i - y_i)^2} \tag{7}$$

Angle cosine uses the cosine value of the angle between two vectors in the vector space as a measure of the difference between two individuals. Compared with the distance measure, cosine similarity pays more attention to the difference of two vectors in shape, rather than distance or length. The formula is as follows.

$$\mathrm{Cos}(X, Y) = \frac{\sum_{i=1}^{t} x_i y_i}{\sqrt{\sum_{i=1}^{t} x_i^2 \sum_{i=1}^{t} y_i^2}} \tag{8}$$

Unfortunately, the complex variation trend of dissolved oxygen under different environmental conditions, it is limited to use only one distance index to calculate the similarity between samples, and the importance of sample characteristics is not considered in both methods.

To comprehensively consider the numerical information and shape information in the historical data, combined with the Euclidean distance and the angle cosine calculation method, this paper uses an improved similarity statistic D_{xy}, to calculate the trend similarity between samples. The calculation formula is as follows.

$$D_{xy} = \alpha d_{xy} + \beta D_{\cos xy} \tag{9}$$

TABLE 2. DBI values of different k.

Cluster number	2	3	4	5	6
DBI	1.2824	0.8906	0.9510	0.8243	0.9014

where,

$$d_{xy} = 1 - \frac{1}{t}\sum_{i=1}^{t}\sqrt{\frac{1}{m}\sum_{j=1}^{m}\sigma_j(x_{ij} - y_{ij})^2} \tag{10}$$

$$D_{\cos xy} = \frac{1}{t}\sum_{i=1}^{t}\frac{\sum_{j=1}^{m}\sigma_j x_{ij} y_{ij}}{\sqrt{\sum_{j=1}^{m}x_{ij}^2}\sum_{j=1}^{m}y_{ij}^2} \tag{11}$$

x_{ij} and y_{ij} are samples of x and y with characteristics of j at the i-th moment, and their values are between [0, 1]. α and β are the weight coefficients between the angle cosine $D_{\cos xy}$ and the Euclidean distance d_{xy}, respectively. These two values will vary with weather conditions. If the weather changes violently, α is close to 1, otherwise β is close to 1. And $\alpha + \beta = 1, \sigma_j$ is the weight of the j-th characteristic factor in the sample.

Judging the validity of clustering results is a difficult and complex problem. In this paper, we set the value of k is determined by our extensive experimental experience. The Davies-Bouldin index proposed by David L. Davis and Donald Bouldin is a clustering measure that describes the similarity of the same class and the difference in different clusters [45]. Small index value can better reflect the clustering effect. Its definition is as follows:

$$DBI = \frac{1}{k}\sum_{i=1}^{k}\max_{j\neq i}\left(\frac{avg(C_i) + avg(C_j)}{d_{cen}(\mu_i, \mu_j)}\right) \tag{12}$$

$$avg(C) = \frac{2}{|C|(|C|-1)}\sum_{1\le i\le j\le |C|} dist(x_i, x_j) \tag{13}$$

$$d_{cen}(C_i, C_j) = dist(\mu_i, \mu_j) \tag{14}$$

where $avg(C)$ represents the average distance between class, $dist(\cdot,\cdot)$ represents the distance between samples and μ stands for the cluster center of class C.

In our study, to establish optimal cluster for model, we select the clustering with the smallest DBI. In Table 2, the number of clusters ranges from 2 to 6, and different numbers of clusters correspond to different DBIs. When the number of clusters is 5, the minimum DBI value is 0.8243. Therefore, it is most appropriate to divide the data set into five categories. And the clustering results are shown in Figure 6.

As can be seen from Figure 6 that the data objects in each cluster are concentrated, and the degree of dispersion among the clusters is relatively high.

F. PLS SOFTPLUS ELM NEURAL NETWORK

1) ELM

The extreme learning machine (ELM) proposed by Halkidi *et al.* [46] is a commonly used learning algorithm, which is a single hidden layer feedforward neural network. Meanwhile, the weights and thresholds of the hidden layer are randomly generated [47]. The network model is usually divided into three layers: input layer, hidden layer and output layer. And the ELM three-layer network structure can approximate any nonlinear function. The topology of the network is shown in Figure 7.

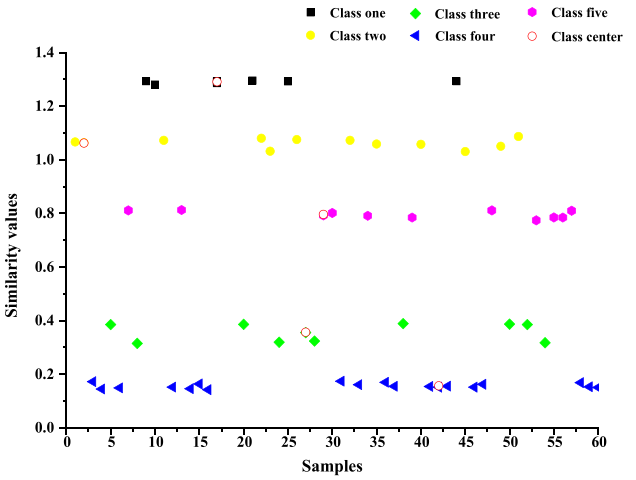


FIGURE 6. The results of K-means clustering.

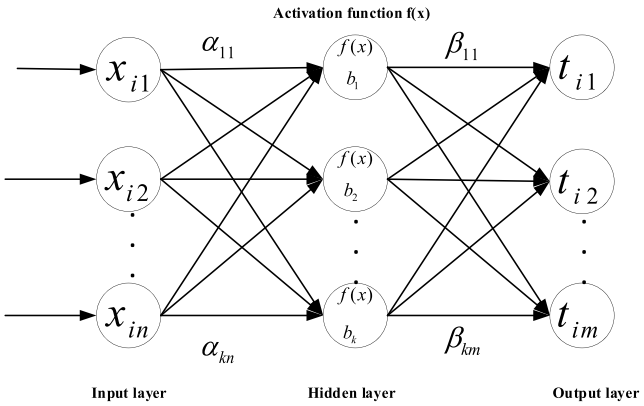


FIGURE 7. ELM neural network topology.

There are N input layer neurons, m output layer neuron, and k hidden layer neurons in the network. The activation function is $f(x)$ between the input layer and the hidden layer [40]. For the total number of N samples, $\{(x_i, t_i)\}_{i=1}^{N}$ is the data sample set, where $x_i = [x_{i1}, x_{i2}, \ldots, x_{in}]^T \in R^n$, $t_i = [t_{i1}, t_{i2}, \ldots, t_{im}]^T \in R^m$. Then the input and output of the network can be expressed as equation (15).

$$\sum_{j}^{k}\beta_j f(\alpha_j, b_j, x_i) = t_i \quad i = 1, 2, \ldots, N \; j = 1, 2, \ldots, k \tag{15}$$

where, $\alpha_j = \left[\alpha_{j1}, \ldots, \alpha_{jn}\right]^T$ is the weight vector between the j-th hidden layer neuron and the input vector x_i, b_j is the biases of the j-th hidden layer neuron. Generally, the Sigmoid function in equation (16) is used as the activation function, and the activation function is used to approximate these N samples with 0 error, the existence of α, β and b makes the

equation (15) true.

$$f(x)\frac{1}{1+e^{-x}} \tag{16}$$

The above formula (15) can be written in matrix form as equation (17)

$$H\beta = T \tag{17}$$

where,

$$H = \begin{pmatrix} f(\alpha_1, b_1, x_1) & \cdots & f(\alpha_k, b_k, x_k) \\ \vdots & \ddots & \vdots \\ f(\alpha_1, b_1, x_N) & \cdots & f(\alpha_k, b_k, x_N) \end{pmatrix}_{N\times k} \tag{18}$$

$$\beta = \left[\beta_1', \beta_2', \ldots, \beta_k'\right]_k' \tag{19}$$

$$T = \left[t_1', t_2^1, \ldots, t_N'\right]_N' \tag{20}$$

$H = \{h_{ij}\}$ is the hidden layer output matrix of the network, $\beta_j = \left[\beta_{j1}, \beta_{j2}, \ldots, \beta_{jm}\right]^T$ When the training x_i is given and the parameters (α_j, b_j) are randomly generated, then we can obtain matrix H.

Then, the output weights β are generated as formula (21).

$$\beta = H^+T \tag{21}$$

where H^+ is the inverse matrix of H, there is a negative correlation between the generalization ability and the norm of the weights β.

The ELM network algorithm consists of 3 steps:

1) Randomly generate the output weight α_j and biases b_j of the hidden unit, $j = 1, 2, \ldots, k$,

2) Calculate the output matrix H of the hidden unit,

3) Calculate the output weight $\beta : \beta = H^+T$.

2) SOFTPLUS ELM

Activation function is a function running on the neurons of neural network, which maps the input of neurons to the output. The activation function plays an important role in the ELM neural network. The earliest idea of ELM is to use the Sigmoid function as the activation function, the output is bounded, and it is easy to serve as the next layer of input. The function is defined as follows:

$$f(x) = \frac{1}{1+e^{-x}} \tag{22}$$

The Sigmoid function is the most commonly used activation function in traditional neural networks and is regarded as the core of neural networks. But the gradient of the function becomes smaller when the input is far from the origin of the coordinates. This will cause the weight to have almost no effect on the loss function, which is not conducive to the optimization of the weight. This problem is called gradient saturation. Actually, the activation function is not unique [48]. However, it is still a daunting task to choose an appropriate activation function [49]. The Rectified Linear Unit Function (ReLU) is currently a popular activation function. The function is defined as follows:

$$f(x) = \max(0, x) \tag{23}$$

Compared with the Sigmoid function, when a positive number is input, there is no gradient saturation problem. In addition, since the ReLU has only a linear relationship, the calculation speed will be faster. However, in the training process, when the function passes through the ReLU unit, the negative gradient will be set to zero, and then it will not be activated by any data. A certain proportion of neurons will die irreversibly, and the parameter gradient cannot be updated, which will lead to the failure of the whole training process. Softplus can be regarded as the smoothness of ReLU [50]. It is defined as follows:

$$f(x) = \ln(1 + e^x) \tag{24}$$

According to the relevant research of neuroscientists, Softplus is similar to the activation frequency function of brain neurons. That is to say, compared to the early activation function Sigmoid, Softplus is closer to the activation model of brain neurons. In order to improve prediction accuracy and save prediction time, this paper uses Softplus as the activation function of ELM, referred to as SELM.

3) PLS-SELM

Partial least squares (PLS) method, by projecting the high-dimensional data space of the independent variable and the dependent variable to the corresponding low-dimensional feature space to obtain the mutually orthogonal feature vectors of the independent variable and the dependent variable [40], and then establish the univariate linear regression relationship of the eigenvectors of independent and dependent variables.

The algorithm flow of PLS-SELM is shown in Figure 8. Obtain the linear relationship between output matrix $H_{n\times L}$ and output $Y_{n\times m}$ of hidden layer nodes in SELM through PLS, as shown below.

$$Y = H\beta_{PLS} + e \tag{25}$$

where, β_{PLS} is the output weight coefficient matrix, e is the noise generated in the model. The external model of PLS, that is, the linear relationship between the output matrix H and the output Y of the hidden layer is as follows.

$$H = TP^T + E = \sum_{i=1}^{a} t_i p_i^T + E \tag{26}$$

$$Y = UQ^T + F = \sum_{i=1}^{a} u_i q_i^T + F \tag{27}$$

The internal model of PLS is:

$$u_i = \beta_i t_i + E_i \tag{28}$$

$$\widehat{\beta}_i = u_i^T t_i / t_i^T t_i \tag{29}$$

The internal model can be expressed as a matrix:

$$U = TB \tag{30}$$

where, $T = [t_1, \ldots, t_a] \in R^{n\times a}$ and $U = [u_1, \ldots, u_a] \in R^{n\times a}$ represent the score matrix of the hidden layer and output layer, respectively. $P = [p_1, \ldots, p_a] \in R^{L\times a}$ is the load matrix of H corresponding to t, and $Q = [q_1, \ldots, q_a] \in$

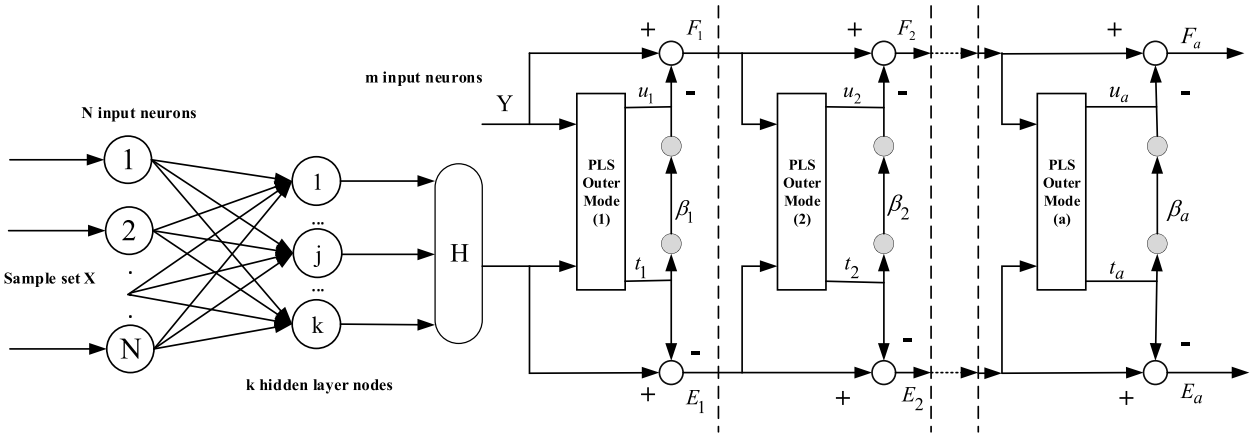


FIGURE 8. The schematic diagram of PSO-PLS-SELM.

$R^{m\times a}$ is the load matrix of Y corresponding to u. a is the number of selected components. $\beta = [\beta_1, \ldots, \beta_a] \in R^{a\times a}$ is the regression coefficient matrix. $E \in R^{n\times L}$ is the residual matrix of H. $F \in R^{n\times L}$ is the residual matrix of Y.

Finally, the output weight $\widehat{\beta_{PLS}}$ can be calculated by

$$\widehat{\beta_{PLS}} = W(P^T W)^{-1} B Q^T \tag{31}$$

where W is the input weights. For each new sample, we divide it into its class and then calculate the prediction result by formula (25).

G. PARTICLE SWARM OPTIMIZATION PLS-SELM NEURAL NETWORK

1) PARTICLE SWARM OPTIMIZATION

Particle Swarm Optimization (PSO) is a swarm intelligence optimization algorithm in the field of computational intelligence. The algorithm was first proposed by Kennedy and Eberhart [51], and was derived from the study of bird predation problems. Particle swarm optimization (PSO) algorithm makes the movement of the whole group evolve from disorder to order in the problem solving space through the sharing of personal information in the group, so as to obtain the optimal solution [52].

Assuming that d and n are the dimensions of the search space and the number of particles respectively, then $Y = (Y_1, \ldots, Y_n)$ is n particles in the d-dimensional space, where $Y_i = (y_{i1}, y_{i2}, \ldots, y_{id})^T$ represents the potential optimal solution of the optimization problem and the particle in the solution space position. And the velocity $V = (V_1, \ldots, V_N)$ of particles represents its distance and direction, where $V_i = (v_{i1}, \ldots, v_{id})^T$. The particle characteristics are represented by three indicators: "position" (x_i^k), "velocity" (v_i^k) and fitness value. The particles move in the search space, and the personal position is updated by tracking the personal best solution *Pbest* and the global best solution *Gbest*. The fitness value is calculated every time a particle updates its position. During each iteration, the particle updates its speed and position through personal extreme values and group extreme values. The update formula is as follows:

$$v_{id}^{k+1} = \omega v_{id}^k + c_1 r_1 (p_{id}^k - y_{id}^k) + c_2 r_2 (p_{gd}^k - y_{gd}^k) \tag{32}$$

$$x_{id}^{k+1} = x_{id}^k + v_{id}^k \tag{33}$$

where k is the number of hidden layer nodes, ω represents iteration weight, c_1 and c_2 stand for the learning factors, r_1 and r_2 are randomly set constant values, and their range is in the interval $[0, 1]$. $P_i = (p_{i1}, \ldots, p_{in})^T$ and $P_g = (p_{g1}, \ldots, p_{gn})^T$ represent personal best position and group best position respectively. $[v_{\min}, v_{\max}]$ is the range of the velocity. Algorithm 2 shows the basic structure of the PSO algorithm.

Algorithm 2 PSO Algorithm for Solving d-Dimensional Optimization Problems

1. Initialize P particles at random positions
2. Evaluate the fitness function of particles
3. *Gbest* = global best solution
4. **For** l = 1 to maximum number of iterations **do**
5. **For** j = 1 to P **do**
6. Update the velocity and position of the j-th particle using formulas (33) and (34)
7. Evaluate the fitness function of j-th particles
8. Update the personal best solution *Pbest* of j-th particles
9. Update the *Gbest*
10. Keep *Gbest* as the best problem solution
11. **End for**
12. **End for**

2) PSO-PLS-SELM

PLS-SELM improves the activation function of the traditional ELM and solves the strong collinearity problem of the hidden layer output matrix. However, since the input weights and hidden layer biases of the network in ELM are initialized randomly, this may make some input weight matrices and hidden layer deviations to 0, resulting in some hidden layer

nodes invalid. In practical applications, a large number of hidden layer nodes are needed to achieve the expected accuracy in ELM. ELM has insufficient generalization ability when dealing with new samples in the training process. Therefore, in order to solve the above problems, this paper proposes PSO to improve the parameters of ELM to improve the prediction accuracy.

The particles in the group consist of input weight matrix and hidden layer biases. The particle length $d = k(n+1)$, where k is the number of hidden layer nodes, and n is the dimension of the input vector, that is, the number of input neurons. Let $\theta^i = [\alpha^i_{11}, \alpha^i_{12}, \ldots, \alpha^i_{1n}, \alpha^i_{21}, \alpha^i_{22}, \ldots, \alpha^i_{2n}, \alpha^i_{k1}, \alpha^i_{k2}, \ldots, \alpha^i_{kn}, b^i_1, b^i_2, \ldots, b^i_k]$ be the i-th particle in the group, where α^i_{ij} and b^i_j are random numbers in the range of $[-Z_{\max}, Z_{\max}]$, $Z_{\max} = 1$. Construct fitness function $f(\theta^i)$ during training. The iteration will not stop until the number of iterations exceeds the maximum or the adaptability value is less than the minimum. The training steps of PSO-PLS-SELM are detailed in Algorithm 3.

Algorithm 3 PSO-PLS-SELM Algorithm

1. Establish PLS-SELM network
2. Set the number k of hidden layer neuron and activation function $f(x)$
3. Initialize PSO particles θ
4. Determine the particle size n, inertia weight ω, the maximum number of iterations T and constant coefficients c_1 and c_2
5. Input training sample
6. **While** ending conditions false **do**
7. **For all** particle i of population **do**
8. Calculate the fitness function $f(\theta^i)$
9. Get the personal extremum P_i
10. **End for**
11. Calculate the global extremum P_g
12. **For all** particle i of the population **do**
13. Adjust the velocity V_i
14. Update the position θ^i
15. **End for**
16. **End while**
17. Separate the global extremum P_g
18. Get input weight α_j and hidden layer biases b_j

IV. RESULTS AND DISCUSSIONS

A. PARAMETER SETTINGS OF ALGORITHM

In this experiment, we use the clustering model proposed in the previous section to divide all data (5760 data sets) into five clusters. Different training sets and test sets correspond to different numbers of nodes that affect the hidden layer. In PLS-SELM, we use trial and error to determine the number of hidden layer nodes in different classes. The number of hidden layer nodes in the five classes are $k_1 = 25$, $k_2 = 30$, $k_3 = 28$, $k_4 = 40$, $k_5 = 29$,respectively, as shown in Table 2.

The selection of PSO parameters is based on a large number of experiments. And the population size of PSO-PLS-SELM is set to 50, acceleration constants $c_1 = 1$ and $c_2 = 2$ respectively, inertia weight = 0.9, iteration number is set to 200 and the fitness accuracy of the normalized samples are equal to 0.005 [53]. These values provide better computational efficiency on the training data set. All experiments are implemented by MATLAB and run on a PC equipped with 2.4 GHz Core (TM) processor, 8.0G memory and Microsoft Windows 10.

B. MODEL PERFORMANCE EVALUATION

In order to further test the prediction performance of the model, the prediction results are evaluated by mean absolute percentage error, root mean square error [54], [55], average percentage error and Nash coefficient [56]. These metrics can reflect the error and fitting degree between the original data and the predicted data from different mathematical perspectives. Due to the larger dimension of each solution in the algorithm, more training time is needed to achieve smaller error. The length of each prediction time is also an index to measure the prediction performance of the model. And these indexes are computed by formula (34) − (37).

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{Y_i - f_i}{Y_i}\right| \tag{34}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Y_i - f_i)^2} \tag{35}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|Y_i - f_i| \tag{36}$$

$$NSC = 1 - \frac{\sum_{i=1}^{N}(Y_i - f_i)^2}{\sum_{i=1}^{N}(Y_i - \overline{Y_i})^2} \tag{37}$$

where N is the number of prediction time points in each data set, Y_t is the original data, f_t is the predicted value and $\overline{Y_t}$ represent the average of original value. The higher NSC, the MAPE, RMSE and MAE values, the more accurate the model.

C. RESULTS OF K-MEANS CLUSTERING

From the DBI results, the 60 days sample set is clustered into five clusters, and then used as the input of PLS-SELM neural network. The optimized network structure is shown in Table 3. The network structure corresponding to different training sets and test sets is also different. With the change of sample number, the hidden layer nodes also change from 25 to 40.

Among them, class 1 has 25 (minimum number) hidden layer nodes, the structure is 7-25-1, and its RMAE value is the smallest. We can find that the weather type of all the data points in class 1 is sunny, the equipment is less affected by the environment, and the periodicity and regularity of the data are good, so the accuracy of the prediction results is also improved accordingly.

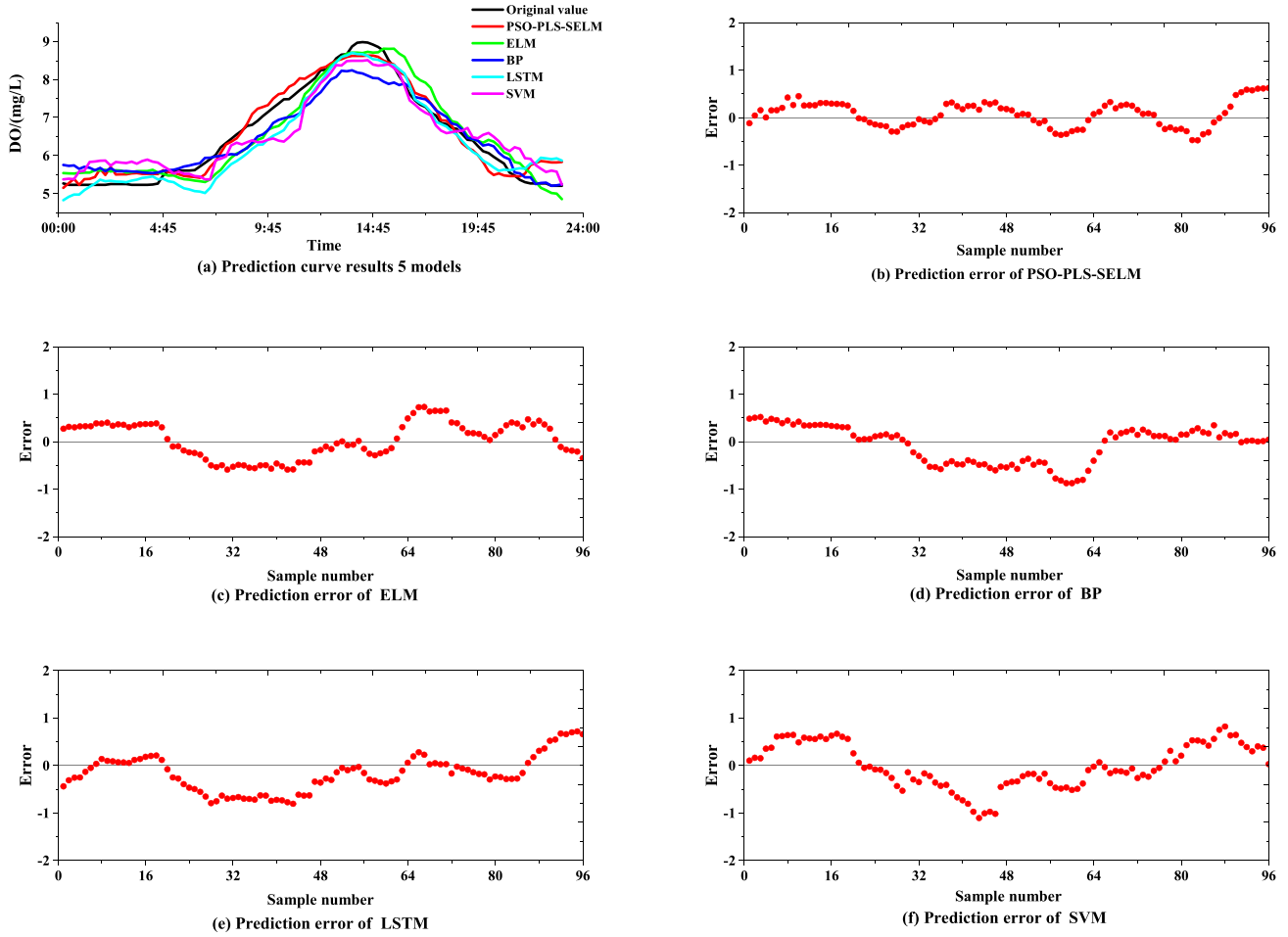


FIGURE 9. The prediction curves of the five models and their prediction error curves.

TABLE 3. Neural network structures in five clusters.

Serial number of clusters	Total data points	Training data points	Testing data points	Structure of neural network	RMSE
1	768	538	230	7-25-1	0.2704
2	1152	806	346	7-30-1	0.2983
3	960	672	288	7-28-1	0.3157
4	1824	1277	547	7-40-1	0.3321
5	1056	739	317	7-29-1	0.3630

D. MODEL COMPARISION AND ANALYSIS

In order to verify the prediction effect of improved PLS-SELM (PSO-PLS-SELM) model based on clustering and PSO, different prediction models are compared in this paper. The comparison models include extreme learning machine (ELM) model, back propagation (BP) neural network model, support vector machine (SVM) model and Long Short-Term Memory (LSTM) model. To testify the performance of the models in this paper, data set is utilized in these models to predict the data of April 8, 2020, that is, the dissolved oxygen content of 24 hours on that day. Figure 9 shows the prediction curves of the five models and their prediction error curves It can be seen from the figure that the prediction result curve of PLS-SELM hybrid model based on clustering and PSO optimization is closer to the original value than the other five models, and has better prediction accuracy. It can be seen from the prediction error chart that the error fluctuation of PSO-PLS-SELM model is smaller than that of the other four models, indicating that the hybrid model improves the prediction accuracy of the model. RMSE, MAE, MAPE and NSC of the five models are calculated for further comparison of the prediction results of the five models. The accuracy of the results is shown in Table 4.

It can be seen from the table that the RMSE, MAPE, MAE and NSC of the improved PLS-SELM model based on clustering and PSO are 0.2704, 0.0371, 0.2284 and 0.9527, respectively. NSC of PSO-PLS-SELM of the five models is the best. However, from the Table 4, we can see that compared with ELM, the running time of the model has been greatly prolonged.

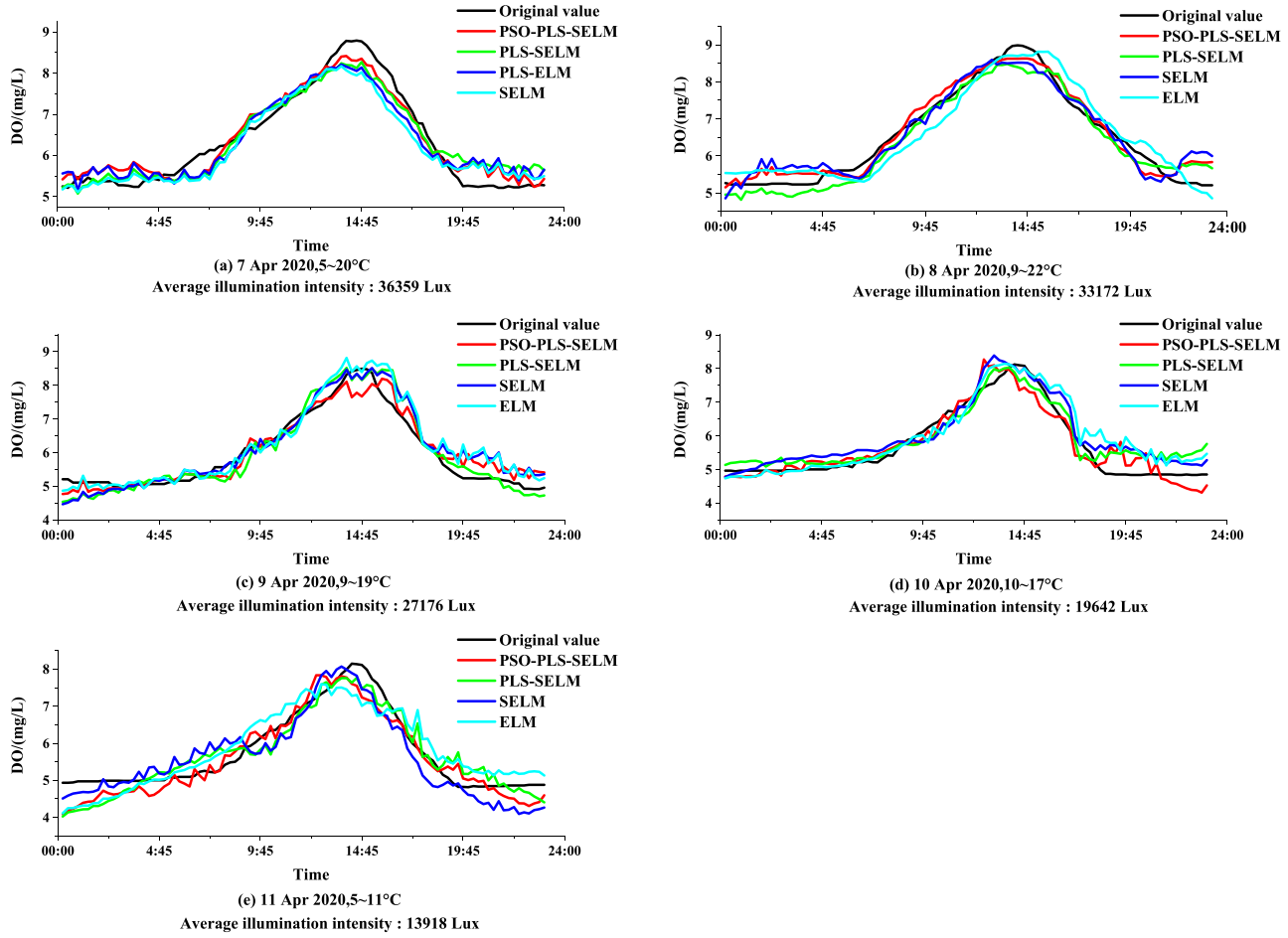


FIGURE 10. The forecasting results of dissolved oxygen content by PSO-PLS-SELM and counterpart models.

TABLE 4. Precision statistics on five models.

Index	PSO-PLS-SELM	ELM	BP	LSTM	SVM
RMSE	0.2704	0.3771	0.3908	0.4164	0.4668
MAPE	0.0371	0.0526	0.0478	0.0533	0.0626
MAE	0.2284	0.3333	0.3228	0.3362	0.3918
NSC	0.9527	0.9081	0.9013	0.8879	0.8592
Time/s	4.760	2.379	4.650	5.686	7.455

Compared with the traditional models, the running time of BP and SVM is longer, and under the same conditions, the prediction results are slightly inferior to the method in this paper. Compared with single LSTM, the hybrid PSO-PLS-SELM model has better performance. The results show that in a multi-factor environment, the similar sample set constructed by this model can eliminate sample differences. Therefore, the prediction model proposed in this paper has high prediction accuracy and strong fitting ability, which can effectively predict the DO content in aquaculture, and can provide help for the prediction and early warning of aquaculture.

In order to further prove the accuracy of the model proposed in this paper, 07 Apr, 08 Apr, 09 Apr, 10 Apr and 11 Apr 2020 are selected as the prediction days of the model, and MAPE, MAE and RMSE are compared and the PSO-PLS-SELM is compared with the three counterpart methods, as shown in Figure 10 (a-e). In these graphs, the X coordinate represents the time of day and the Y coordinate represents the dissolved oxygen content. As can be seen from Figure 10 (a-e), all models can well predict dissolved oxygen, but the prediction results are different. The prediction curve of PSO-PLS-SELM is closer to the original data than that of PLS-SELM, SELM and ELM. Figure 11 shows that PSO-PLS-SELM achieves the best prediction effect by clustering and optimizing model parameters.

Further more, we also give the average prediction accuracy evaluation index as shown in Table 5. Compared with PLS-SELM (0.0500), SELM (0.0567) and ELM (0.0563), the average MAPE of PSO-PLS-SELM is the best (0.0435).

TABLE 5. Average precision results on three counterpart models.

Index	PSO-PLS-SELM	PLS-SELM	SELM	ELM
RMSE	0.3159	0.3534	0.3941	0.4171
MAPE	0.0435	0.0500	0.0567	0.0563
MAE	0.2556	0.2911	0.3289	0.3356
NSC	0.9160	0.8958	0.8712	0.8555

TABLE 6. Comparison of forecasting results on different weather conditions.

Index	PSO-PLS-SELM		PLS-SELM		SELM		ELM	
	08 Apr	11 Apr	08 Apr	11 Apr	08 Apr	11 Apr	08 Apr	11 Apr
MAPE	0.0371	0.0523	0.0428	0.0596	0.0480	0.0677	0.0526	0.0701
RMSE	0.2704	0.3630	0.3075	0.3954	0.3591	0.4343	0.3771	0.4570
MAE	0.2284	0.2929	0.2661	0.3249	0.2890	0.3773	0.3333	0.3989
NSC	0.9527	0.8782	0.9389	0.8555	0.9167	0.8257	0.9081	0.807

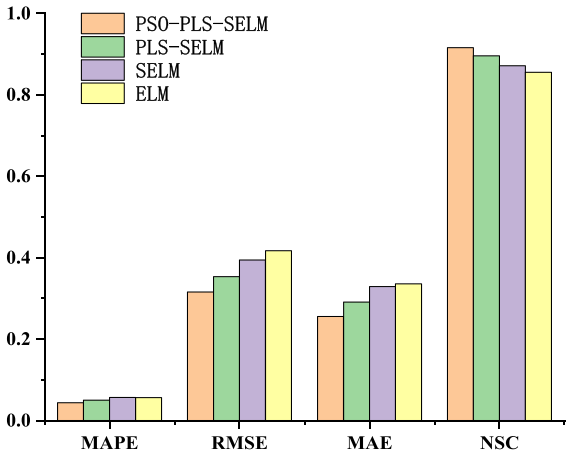


FIGURE 11. Average precision on three counterpart models.

Compared with PLS-ELM, SELM and ELM, the average RMSE of the algorithm is improved by 11.87%, 22.22% and 32.03%, respectively. Similarly, the average MAE of PSO-PLS-SELM is improved 13.89% of PLS-SELM, 28.68% of SELM, and 31.30% of ELM. Obviously, the proposed PSO-PLS-SELM model is superior to the other three models in all accuracy indexes of DO prediction. Simultaneously, the NSC value of this model (0.9160) is better than that of the three models.

For different weather conditions, choose April 8th (The highest and lowest temperatures are 22 °C and 9 °C, respectively. The average illumination intensity is 36359 Lux) and April 11th (The highest and lowest temperatures are 11 °C and 5 °C, respectively. The average illumination intensity is 13918 Lux). The predicted and true values of dissolved oxygen for the four comparison models are shown in Figure 10(b, e). Overall, the results of the PSO-PLS-SELM model are closer to the true values. From Table 6, under two different weather conditions, the RMSE of the proposed PSO-PLS-SELM (0.2704) is the smallest, and the MAPE and MAE also have the same conclusion.

It shows that the model introduced in this paper performs similar daily clustering under the influence of multiple factors, which can adjust input and eliminate sample differences. On April 11, the NSC of PSO-PLS-SELM had the largest value compared with the other three control models. Although, the forecast curve fluctuates relatively greatly in the first few hours. The model we proposed can identify sudden changes in weather on the forecast day. The model increases the frequency of data updating, and quickly obtains the appropriate training samples, thus reduces the prediction error in the later stage, which makes the result closer to the real value.

These data prove the good performance of PSO-PLS-SELM prediction model. The main reason is that the activation function in the traditional ELM is replaced by the Softplus function, making the model closer to the neural activation model. In addition, PLS is applied to the Softplus ELM model, can not only effectively overcome the collinearity problem of ordinary least squares regression, but also emphasize the explanation and prediction of independent variables to dependent variables when selecting feature vectors, and eliminate the influence of useless noise on regression model. The model contains the least number of variables, so the PLS model has better robustness and predictive stability. Meanwhile, PSO is utilized to optimize the output weight and hidden layer bias of the network. These factors greatly improve the accuracy and robustness of the algorithm. In conclusion, compared with the other methods, the accuracy of the proposed PSO-PLS-SELM model is greatly improved. In particular, PSO-PLS-SELM can select the best input samples, find the periodic change pattern and trend change of water quality and meteorological data (aquaculture), combined with the advantages of various single models, so as to provide better prediction results.

V. CONCLUSION

We propose a prediction model for predicting the change of dissolved oxygen content in time series data. The PLS-SELM model is improved based on clustering and PSO. A combined similarity clustering method is adopted to divide the original data into different clusters, which improves the sample selection of the model. Meanwhile, Softplus function is utilized as the activation function of the network, PLS is applied to construct the regression model which can explain

the system and predict the set of dependent variables, and PSO is introduced to effectively improve the generalization performance of short-term dissolved oxygen prediction model. This ability endows the model with better performance than other counterpart models. Experimental research shows that in terms of RMSE, MAPE, MAE and NSC, the performance of the method introduced in this paper is better than the single BP neural networks, SVR, LSTM and the traditional ELM algorithm. Its effectiveness has been verified in practical applications. It can predict the dissolved oxygen content, thereby reducing and avoiding unnecessary losses, such as fish mortality rate, caused by excessive or insufficient dissolved oxygen content. For PSO-PLS-SELM, we mainly consider to improve the prediction scheme of dissolved oxygen. In the future work, we will experiment and discuss in other similar application fields, and promote them. At the same time, we will pay more attention to the comparison of different models, so that the performance of the model can be more in-depth analysis and discussion.

ACKNOWLEDGMENT

The authors thank the anonymous reviewers for their valuable comments.

REFERENCES

[1] X. Zhu, D. Li, D. He, J. Wang, D. Ma, and F. Li, "A remote wireless system for water quality online monitoring in intensive fish culture," *Comput. Electron. Agricult.*, vol. 71, pp. S3–S9, Oct. 2010.

[2] H. C. Yuan, J. H. Huang, and Y. T. Zhao, "Prediction of dissolved oxygen based on PCA-NARX neural network," *J. Shandong Agricult. Univ., Natural Sci. Ed.*, vol. 50, no. 5, pp. 902–907, Apr. 2019.

[3] M. Lipizer, E. Partescano, A. Rabitti, A. Giorgetti, and A. Crise, "Qualified temperature, salinity and dissolved oxygen climatologies in a changing Adriatic sea," *Ocean Sci.*, vol. 10, no. 5, pp. 771–797, Oct. 2014.

[4] J. Huan, W. J. Cao, and X. Q. Liu, "A dissolved oxygen prediction method based on k-means clustering and the ELM neural network: A case study of the ChangDang Lake, China," *Appl. Eng. Agricult.*, vol. 33, no. 4, pp. 461–469, May 2017.

[5] Z. Xiao, L. Peng, Y. Chen, H. Liu, J. Wang, and Y. Nie, "The dissolved oxygen prediction method based on neural network," *Complexity*, vol. 2017, pp. 1–6, Oct. 2017.

[6] W.-B. Chen and W.-C. Liu, "Artificial neural network modeling of dissolved oxygen in reservoir," *Environ. Monitor. Assessment*, vol. 186, no. 2, pp. 1203–1217, Feb. 2014.

[7] S. R. Poulson and A. B. Sullivan, "Assessment of diel chemical and isotopic techniques to investigate biogeochemical cycles in the upper Klamath River, Oregon, USA," *Chem. Geol.*, vol. 269, nos. 1–2, pp. 3–11, Jan. 2010.

[8] L. Q. Xu, Q. C. Li, S. Y. Liu, and D. L. Li, "Prediction of pH value in industrialized aquaculture based on ensemble empirical mode decomposition and improved artificial bee colony algorithm," *Trans. Chin. Soc. Agricult. Eng.*, vol. 32, no. 3, pp. 202–209, Feb. 2016.

[9] J. Huan and X. Q. Liu, "Dissolved oxygen prediction in water based on K-means clustering and ELM neural network for aquaculture," *Trans. Chin. Soc. Agricult. Eng.*, vol. 32, no. 17, pp. 174–181, Sep. 2016.

[10] D. Ö. Faruk, "A hybrid neural network and ARIMA model for water quality time series prediction," *Eng. Appl. Artif. Intell.*, vol. 23, no. 4, pp. 586–594, Jun. 2010.

[11] H. Yu, Y. Chen, S. Hassan, and D. Li, "Dissolved oxygen content prediction in crab culture using a hybrid intelligent method," *Sci. Rep.*, vol. 6, no. 1, p. 27292, Jun. 2016.

[12] S. Amid and T. M. Gundoshmian, "Prediction of output energies for broiler production using linear regression, ANN (MLP, RBF), and ANFIS models," *Environ. Prog. Sustain. Energy*, vol. 36, no. 2, pp. 577–585, Jan. 2017.

[13] J. J. Pan, "Study on the model of dissolved oxygen prediction based on RBF neural network," M.S. thesis, Dept. Comput. Sci. Technol., Shanghai Ocean Univ, Shanghai, China, 2016.

[14] Y. Z. Zhu and N. Chen, "Dynamic forecast of regional groundwater level based on grey Markov chain model," *Chin. J. Geotech. Eng.*, vol. 33, no. S1, pp. 78–82, Aug. 2011.

[15] Y. Yue and T. H. Li, "The application of a fuzzy-set-theory based Markov model in the quantitative prediction of water quality," *J. Basic Sci. Eng.*, vol. 19, no. 2, pp. 231–242, Apr. 2011.

[16] P. S. Xue, M. Q. Feng, and X. P. Xing, "Water quality prediction model based on Markov chain improving gray neural network," *Eng. J. Wuhan Univ.*, vol. 45, no. 3, pp. 319–324, Jun. 2012.

[17] R. Grbic, D. Kurtagic, and D. Sliskovic, "Stream water temperature prediction based on Gaussian process regression," *Expert Syst. Appl.*, vol. 40, no. 18, pp. 7407–7414, Dec. 2013.

[18] B. Luo, Y. Zhao, K. Chen, and X. Zhao, "Partial least squares regression model to predict water quality in urban water distribution systems," *Trans. Tianjin Univ.*, vol. 15, no. 2, pp. 140–144, Sep. 2009.

[19] X. Yang and W. Jin, "GIS-based spatial regression and prediction of water quality in river networks: A case study in Iowa," *J. Environ. Manage.*, vol. 91, no. 10, pp. 1943–1951, Oct. 2010.

[20] F. Evrendilek and N. Karakaya, "Regression model-based predictions of diel, diurnal and nocturnal dissolved oxygen dynamics after wavelet denoising of noisy time series," *Phys. A, Stat. Mech. Appl.*, vol. 404, pp. 8–15, Jun. 2014.

[21] J. Wu, J. Lu, and J. Wang, "Application of chaos and fractal models to water quality time series prediction," *Environ. Model. Softw.*, vol. 24, no. 5, pp. 632–636, May 2009.

[22] F. K. Arya and L. Zhang, "Time series analysis of water quality parameters at Stillaguamish river using order series method," *Stochastic Environ. Res. Risk Assessment*, vol. 29, no. 1, pp. 227–239, Jan. 2015.

[23] J. Paredes-Arquiola, J. Andreu-Álvarez, M. Martín-Monerris, and A. Solera, "Water quantity and quality models applied to the Jucar river Basin, Spain," *Water Resour. Manage.*, vol. 24, no. 11, pp. 2759–2779, Sep. 2010.

[24] S. Missaghi, M. Hondzo, and W. Herb, "Prediction of lake water temperature, dissolved oxygen, and fish habitat under changing climate," *Climatic Change*, vol. 141, no. 4, pp. 747–757, Apr. 2017.

[25] S. N. Chan, W. Thoe, and J. H. W. Lee, "Real-time forecasting of Hong Kong beach water quality by 3D deterministic model," *Water Res.*, vol. 47, no. 4, pp. 1631–1647, Mar. 2013.

[26] Y. Zhang and Q. Q. Gao, "Comprehensive prediction model of water quality based on grey model and fuzzy neural network," *Chin. J. Environ. Eng.*, vol. 9, no. 2, pp. 537–545, Feb. 2015.

[27] M. J. Alizadeh and M. R. Kavianpour, "Development of wavelet-ANN models to predict water quality parameters in Hilo Bay, Pacific Ocean," *Mar. Pollut. Bull.*, vol. 98, nos. 1–2, pp. 171–178, Sep. 2015.

[28] S. Liu, H. Tai, Q. Ding, D. Li, L. Xu, and Y. Wei, "A hybrid approach of support vector regression with genetic algorithm optimization for aquaculture water quality prediction," *Math. Comput. Model.*, vol. 58, nos. 3–4, pp. 458–465, Aug. 2013.

[29] Y. H. Du, K. P. Wei, Y. J. Shi, E. H. Liu, Q. Y. Feng, and G. Y. Dong, "Infrared detection and clustering grey fusion prediction model of water quality turbidity," *Infr. Laser Eng.*, vol. 45, no. 10, pp. 272–278, Oct. 2016.

[30] T. Y. Pai, R. S. Wu, C. H. Chen, L. Chen, C. Y. Lin, H. Y. Lee, L. H. Shih, Y. Z. Jiang, and C. Y. Shen, "Predicting hardness of four groundwater monitoring stations in Kaohsiung city of Taiwan using seven types of GM (1, 1) model," *Adv. Mater. Res.*, vol. 905, pp. 314–317, Apr. 2014.

[31] N. V. Kumar, S. Mathew, and G. Swaminathan, "Analysis of groundwater for potability from Tiruchirappalli city using backpropagation ANN model and GIS," *J. Environ. Protection*, vol. 1, no. 2, pp. 136–142, 2010.

[32] Z. Gomolka, B. Twarog, E. Zeslawska, A. Lewicki, and T. Kwater, "Using artificial neural networks to solve the problem represented by BOD and DO indicators," *Water*, vol. 10, no. 4, pp. 1–26, Jan. 2018.

[33] N. Mahmoudi, H. Orouji, and E. Fallah-Mehdipour, "Integration of shuffled frog leaping algorithm and support vector regression for prediction of water quality parameters," *Water Resour. Manage.*, vol. 30, no. 7, pp. 2195–2211, May 2016.

[34] S. Y. Liu, L. Q. Xu, D. L. Li, and L. H. Zeng, "Online prediction for dissolved oxygen of water quality based on support vector machine with time series similar data," *Trans. Chin. Soc. Agricult. Eng.*, vol. 30, no. 3, pp. 155–162, Feb. 2014.

[35] J. T. Zhou, H. Zhang, D. Jin, and X. Peng, "Dual adversarial transfer for sequence labeling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 2, pp. 434–446, Feb. 2021.
[36] S. Liu, L. Xu, D. Li, Q. Li, Y. Jiang, H. Tai, and L. Zeng, "Prediction of dissolved oxygen content in river crab culture based on least squares support vector regression optimized by improved particle swarm optimization," *Comput. Electron. Agricult.*, vol. 95, pp. 82–91, Jul. 2013.
[37] P. Hu, X. Peng, H. Zhu, J. Lin, L. Zhen, and D. Peng, "Joint versus independent multiview hashing for cross-view retrieval," *IEEE Trans. Cybern.*, early access, Oct. 29, 2020, doi: 10.1109/TCYB.2020.3027614.
[38] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, Dec. 2006.
[39] J. Zhou, Y. Liu, and T. Zhang, "Analytical redundancy design for aeroengine sensor fault diagnostics based on SROS-ELM," *Math. Problems Eng.*, vol. 2016, pp. 1–9, Apr. 2016.
[40] P. Shi, G. Li, Y. Yuan, G. Huang, and L. Kuang, "Prediction of dissolved oxygen content in aquaculture using clustering-based softplus extreme learning machine," *Comput. Electron. Agricult.*, vol. 157, pp. 329–338, Feb. 2019.
[41] S. T. Wu, F. H. Hou, and F. Dai, "Linear approxmiating method in the transacting process of nonlinear standardization of data," *J. Inf. Eng. Univ.*, vol. 8, no. 2, pp. 250–253, Jun. 2007.
[42] L. Q. Xu and S. Y. Liu, "Water quality prediction model based on APSO-WLSSVR," *J. Shandong Univ., Eng. Sci.*, vol. 42, no. 5, pp. 80–86, Oct. 2012.
[43] X. Cao, Y. Liu, J. Wang, C. Liu, and Q. Duan, "Prediction of dissolved oxygen in pond culture water based on K-means clustering and gated recurrent unit neural network," *Aquacultural Eng.*, vol. 91, Sep. 2020, Art. no. 102122.
[44] X. Peng, H. Zhu, J. Feng, C. Shen, H. Zhang, and J. T. Zhou, "Deep clustering with sample-assignment invariance prior," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 11, pp. 4857–4868, Nov. 2020.
[45] M. Sassi, "Towards fuzzy-hard clustering mapping processes," *Comput. Sci.*, vol. 1, pp. 37–63, Apr. 2012.
[46] M. Halkidi, Y. Batistakis, and M. Vazirgiannis, "On clustering validation techniques," *J. Intell. Inf. Syst.*, vol. 17, nos. 2–3, pp. 107–145, 2001.
[47] G.-B. Huang, L. Chen, and C.-K. Siew, "Universal approximation using incremental constructive feedforward networks with random hidden nodes," *IEEE Trans. Neural Netw.*, vol. 17, no. 4, pp. 879–892, Jul. 2006.
[48] K. Hornik, "Approximation capabilities of muitilayer," *Neural Netw.*, vol. 4, pp. 157–251, Oct. 1991.
[49] P. Hu, X. Peng, H. Y. Zhu, and J. Lin, "Cross-modal discriminant adversarial network," *Pattern Recognit.*, vol. 112, pp. 1–14, Oct. 2020.
[50] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, vol. 15, May 2011, pp. 315–323.
[51] R. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in *Proc. 6th Int. Symp. Mach. Hum. Sci.*, vols. 4–6, Oct. 1995, pp. 39–43.
[52] E. M. N. Figueiredo and T. B. Ludermir, "Investigating the use of alternative topologies on performance of the PSO-ELM," *Neurocomputing*, vol. 127, pp. 4–12, Aug. 2014.
[53] Q. Wu and R. Law, "Cauchy mutation based on objective variable of Gaussian particle swarm optimization for parameters selection of SVM," *Expert Syst. Appl.*, vol. 38, no. 6, pp. 6405–6411, Jun. 2011.
[54] Q. Y. Shao, "Research on architecture selection of ELM networks," M.S. thesis, Comput. Appl. Technol., Hebei Univ., Hebei, China, 2013.
[55] S. X. Wang, Y. M. Wang, Y. Liu, and N. Zhang, "Hourly solar radiation forecasting based on EMD and ELM neural network," *Electr. Power Autom. Equip.*, vol. 34, no. 8, pp. 7–12, Aug. 2014.
[56] L. Benyahya, A. St-Hilaire, T. B. M. J. Quarda, B. Bobée, and B. Ahmadi-Nedushan, "Modeling of water temperatures based on stochastic approaches: Case study of the Deschutes River," *J. Environ. Eng. Sci.*, vol. 6, no. 4, pp. 437–448, Sep. 2007.

SHOUQI CAO received the bachelor's degree in mechanical manufacturing technology and equipment and the M.S. degree in mechanical manufacturing and automation from Sichuan University, in 1996 and 1999, respectively, and the Ph.D. degree in control science and engineering from Shanghai University, in 2009. He is currently a Professor and a Ph.D. Supervisor with the College of Engineering Science and Technology, Shanghai Ocean University. His main research interests include marine Internet of Things engineering, fisheries engineering, and automation technology research.

LIXIN ZHOU received the bachelor's degree in information and computing science from Shanghai Ocean University, in 2018, where she is currently pursuing the degree with the College of Engineering Science and Technology. Her main research interest includes ocean engineering and information.

ZHENG ZHANG received the B.S. and M.S. degrees from Zhengzhou University, China, in 2003 and 2007, respectively, and the Ph.D. degree from Shanghai Jiao Tong University, China, in 2015. Since 2015, he has been a Lecturer with the College of Engineering Science and Technology, Shanghai Ocean University. His research interests include wireless sensor networks, ad hoc networks, and embedded systems.