

## Research Article

# Aquaculture Prediction Model Based on Improved Water Quality Parameter Data Prediction Algorithm under the Background of Big Data

Yuan Jiang <sup>1</sup> and Fei Yan <sup>2</sup>

<sup>1</sup>School of Artificial Intelligence, Zhuhai City Polytechnic, Zhuhai 519090, China

<sup>2</sup>School of Electronic Information Engineering, Nantong Vocational University, Nantong 226007, China

Correspondence should be addressed to Yuan Jiang; [jiangyuan@mjc-edu.cn](mailto:jiangyuan@mjc-edu.cn)

Received 21 September 2022; Revised 28 October 2022; Accepted 5 November 2022; Published 25 November 2022

Academic Editor: Theodore E. Simos

Copyright © 2022 Yuan Jiang and Fei Yan. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Computer science and technology under the background of big data are closely related to the development of modern agriculture. The application of information processing technology in aquaculture will promote the scientific development of aquaculture. The aquaculture water quality directly affects the effect of aquaculture. Therefore, on the basis of the dynamic monitoring model of water quality, the relevant factors affecting water quality were analyzed, and a prediction model of aquaculture water quality was constructed. Considering the complex relationship between dissolved oxygen and water quality, combined with principal component analysis, a PCA-BP (principal component analysis back propagation) water quality prediction model was proposed. The parameters of PCA-BP water quality prediction model were optimized by genetic algorithm, the threshold and weight of BP neural network were determined, and an improved PCA-BP water quality prediction model was constructed. The experimental results show that the relative error of the GPCA-BP water quality prediction model for the prediction of dissolved oxygen content is less than 0.76% in water quality prediction experiments in different times and regions, and it has the best prediction accuracy. At the same time, GPCA-BP water quality prediction model also has excellent performance in convergence accuracy, prediction accuracy, and MAE error performance test. The research content has important reference value for the application of information technology in modern aquaculture.

## 1. Introduction

Traditional aquaculture faces problems such as backward technology, many aquatic diseases, and low aquatic product output, which is not conducive to the development of modern aquaculture. In modern freshwater aquaculture, water quality is the key to aquaculture. In aquaculture, the water quality needs to consider many factors. Once there is a problem with the water quality, a large number of fish will die. Therefore, in order to further analyze the complex relationship between water quality factors, the water quality monitoring model was applied to the aquaculture environment. After analyzing a large number of references and aquaculture materials, the aquaculture quality model is based on dissolved oxygen monitoring, and a dissolved oxygen machine

learning model is constructed through big data technology and computer technology. This technology combines the Internet of Things with a water quality monitoring system, monitors the hydrological data environment through hardware, and realizes the prediction and processing of hydrological data through machine learning algorithms. This includes the prediction and analysis of the concentration of dissolved oxygen, oxygen consumption of aquatic organisms and other parameters, and the regulation of aquaculture according to the preset concentration threshold. With the help of big data technology, the precise management and control of aquaculture can be realized, the occurrence of aquaculture diseases can be effectively prevented and controlled, and the quality of aquaculture can be improved in an all-round way. The research content innovatively

combines big data technology and Internet of Things technology to achieve dynamic management and control of aquaculture, providing an important reference for the development of modern agriculture.

## 2. Related Work

With the innovation and development of computer science, big data mining technology is widely used in the fields of data classification, mining, electricity, and modern social prosperity. Experts at home and abroad have done a lot of research on big data mining algorithms. An efficient scheme for massive data storage space using filters is proposed, which uses fuzzy operations to adjust hash data from one filter to another to reduce storage requirements. The experimental test results show that the amount of data processed by this scheme is 1.9 times that of the standard [1]. Chen et al. found that big data systems in reinforcement learning have problems with randomness and reliability, a dynamic coherence quality metric based on an axiomatic framework is proposed, and the proposed metric is applied to three empirical studies of wavelet-based big data systems. In data analysis, experimental performance results show that the adopted scheme has excellent performance in terms of efficiency and robustness [2]. A combined dynamic DNA image encryption algorithm is proposed. Three-neuron fractional discretization is used as a pseudorandom chaotic sequence generator. Experimental test results show that the algorithm has better performance than the algorithm reported in [3]. Deblais et al. discovered that technological advances have enabled the generation and storage of large sets of information, using big data to examine foodborne pathogens in poultry, and experimental results show that genomic approaches reveal the importance of gut microbiota in health and disease [4]. It is a challenge to discover that industrial big data technologies use limited energy to transmit data. Therefore, the  $1 + \alpha$  approximation offline algorithm and the maximum  $\{2, \beta\}$  competitive ratio online algorithm are proposed. The performance test results show that this scheme proves that there is no online algorithm with constant competition ratio [5]. Ni et al. found that using a cyberphysical system to process a large amount of heterogeneous data has a good effect, and a joint network structure is proposed. The experimental performance test shows that the method has good performance [6]. Guan and Zhao studied the trajectories and behavior patterns of shrimp fishing boats in a fishing trajectory system and used big data technology to design a shrimp farm distribution management system based on a back propagation algorithm. Experimental test results show that the solution can achieve effective monitoring of trajectory and shrimp distribution [7]. Kumar found that information data security affects people's daily work. He designed a malware monitoring scheme using big data technology and machine science technology. The simulation test results show that the accuracy rate of the proposed model reaches 99.8% [8]. Gu et al. use big data technology to analyze enterprise organizational resources and improve enterprise performance. The experimental results show that the data analysis ability of the enterprise directly affects the supplier and individual performance development [9]. The origin of the Internet of Things technology and the current situation at home and abroad are

analyzed; combined with the management concept of the Internet of Things, a variety of sensors are built in, including identification and communication technologies. The performance simulation test shows that the program can achieve effective management of employees [10].

Computer information processing technology has important application value in agricultural production, and big data algorithms have outstanding effects in aquaculture. Experts at home and abroad have conducted a lot of research on this. Shi et al. found that water quality directly affects fish survival during long-distance transportation. In the past research, the simulation was mainly aimed at the transportation time, which could not play an effective role. According to the research on aquatic water quality, it is found that parameters such as nitrite and dissolved oxygen affect the survival of fish. Therefore, through data collection, a water quality model is constructed through multiple layers of neural networks. After testing, the proposed model can effectively monitor the water quality status with low error and meet the requirements of aquatic product transportation [11]. Mathisen et al. conducted research on the development of modern aquaculture, and traditional aquaculture faces many challenges. Therefore, an aquaculture case system was developed and designed, based on the analysis of traditional cases, to provide data reference for modern aquaculture and to verify the feasibility of the scheme through the Siamese neural network. Tests show that the proposed scheme is reliable in aquaculture [12]. Li et al. found that dissolved oxygen is an important parameter for freshwater aquaculture, and the content of dissolved oxygen will affect the development of aquaculture. The traditional water quality testing scheme has the characteristics of multivariate, and the prediction effect is low. Therefore, a model is constructed by combining convolution and long- and short-term networks, and the model time series is processed. The test results show that the proposed scheme has better dissolved oxygen and oxygen prediction performance compared with the traditional scheme and at the same time has lower error performance, which meets the requirements of aquaculture [13]. Zhang et al. designed a monitoring system for aquaculture through research on the Internet of Things technology. The equipment includes sensors such as dissolved oxygen monitoring and water temperature monitoring, and they developed a communication platform based on a wireless communication network monitoring and processing. After testing, this scheme can effectively realize the transmission of data and monitor the water environment [14].

Research on related technologies at home and abroad shows that big data mining technology and artificial intelligence algorithm technology are widely used in the field of data analysis and prediction. The application of neural network technology in modern aquaculture has a positive impact on the development of aquaculture.

## 3. Construction of Aquaculture Water Quality Prediction Model Based on Improved PCA-BP

*3.1. Research on the Prediction Model of Dissolved Oxygen in Water Quality.* In modern aquaculture, the dissolved oxygen parameters of fish pond water will determine the quality effect of aquaculture. Therefore, aquaculture needs to master

the content of dissolved oxygen in the pond to ensure the healthy survival of aquatic organisms in the fish pond [15]. The research on the dissolved oxygen level in fish ponds found that many factors in the fish ponds have an impact on the parameters of dissolved oxygen index [16]. Therefore, combining modern aquaculture technology with dissolved oxygen monitoring methods can achieve accurate prediction of dissolved oxygen in fish ponds.

Considering that fish pond water quality prediction is a complex nonlinear mapping problem, it is difficult to achieve accurate prediction results by using common prediction methods, principal component analysis (PCA) method combined with BP (back propagation) neural network. The network will be used to deal with complex nonlinear mapping problems and achieve effective prediction of aquaculture water quality. The idea of principal component analysis is to use the idea of covariance to deal with complex data sample problems and obtain important weights of data and main data [17]. The number of sample variables is defined as the number of samples, the  $m$  total number of samples is  $n$ , the first sample  $j$  variable is  $x_j(j = 1, 2, \dots, m)$ , and the first  $j$  variable index of the sample  $i$  is  $x_{ij}(i=1,2,\dots,j=1,2,\dots,m)$ ; then, the initial matrix expression of the sample is obtained as

$$X = \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ \dots & \dots & \dots \\ x_{n1} & x_{n2} & x_{n3} \end{pmatrix}. \quad (1)$$

The PCA method is used to solve complex data problems, which requires linear transformation of the matrix  $X$  to retain the original sample data information and generate new variable data. Described using parameters  $y_1, y_2, \dots, y_p$  ( $p < m$ ), the new variable data lacks correlation expression and is described using a linear combination of the original variable data, as shown in the following:

$$\begin{cases} y_1 = u_{11}x_1 + u_{12}x_2 + \dots + u_{1m}x_m, \\ y_2 = u_{21}x_1 + u_{22}x_2 + \dots + u_{2m}x_m, \\ \vdots \\ y_p = u_{p1}x_1 + u_{p2}x_2 + \dots + u_{pm}x_m. \end{cases} \quad (2)$$

In Equation (2),  $u$  denotes the variables of the matrix. After obtaining the linear combination description of the original matrix, the matrix expression can be represented by a matrix  $U$  on a matrix  $Y$  as shown in the following:

$$Y = UX. \quad (3)$$

In Equation (3), each row of the matrix  $U$  is a unit row vector, and the matrix  $Y$  is shown as follows:

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix}. \quad (4)$$

Matrix  $U$  is shown in the following:

$$U = \begin{pmatrix} u_{11} & u_{12} & \dots & u_{1m} \\ u_{21} & u_{22} & \dots & u_{2m} \\ \dots & \dots & \dots & \dots \\ u_{p1} & u_{p2} & \dots & u_{pm} \end{pmatrix}. \quad (5)$$

In Equation (3)-Equation (5),  $y_i$  and  $y_j$  are coefficients 0 and  $y \neq j$ . Each row of the matrix  $U$  will be sorted from large to small in the form of linear combination variance, indicating that the original data variables are subsorted.  $y_n$  When the variance of the principal component accounts for the variance of the principal component, the new variable can replace the  $x_n$  total variance of the original variable. it is more than 85%, the dimension of the data can be reduced, thus making data processing simpler and retaining the original data information.

A BP neural network is a prefeedback system that learns to respond to mapping associations between data input and output. The structure of BP network is mainly composed of input, output, and hidden parts. Figure 1 is a schematic diagram of the basic BP neural network structure [17].

In order to illustrate the principle of the three-layer neural network structure, the input layer, hidden layer, and output layer of the BP neural network are defined as  $N, Q$ , and  $M$ , respectively, and the total number of data samples is  $P$ ; then, the  $i$ th input value of the  $P$ th sample is  $x_{ki}$  and the  $i$ th node of the input layer. The  $k$  threshold to the  $i$ th node of the hidden layer is  $v_{ki}$ , and the  $i$  weight of the  $k$ th node of the hidden layer to the  $j$ th node of the output layer is  $\omega_{jk}$ ; then, the formula of the hidden layer node is BP neural network:

$$z_{pk} = f(n_{pk}) = f\left(\sum_{i=0}^n v_{ki}x_{ki} + \theta_k\right), \quad k = 0, 1, \dots, q. \quad (6)$$

In (6),  $n_{pk}$  is the hidden layer neural node, in which  $\theta_k$  represents the hidden layer node parameters. Then, the nodes of the output layer are computed and represented as shown in the following:

$$y_{pk} = f(n_{pj}) = f\left(\sum_{k=0}^q \omega_{jk}z_{pk} + \theta_j\right), \quad k = 0, 1, \dots, m. \quad (7)$$

In Equation (7),  $n_{pj}$  is the output layer neural node, in

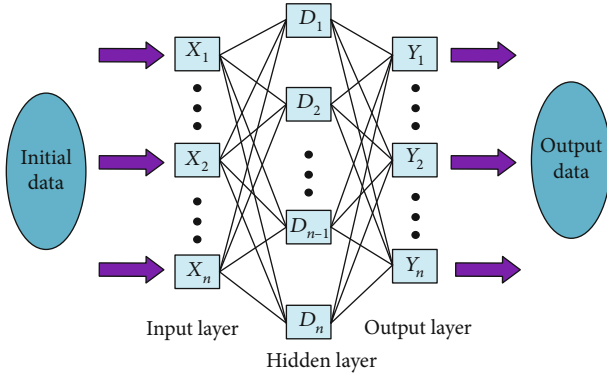


FIGURE 1: The basic structure of the BP neural network.

which  $\theta_j$  represents the output layer node parameters. The global error function of the network is obtained using the activation function, as shown in the following:

$$E = \sum_{p=1}^P E_p = \frac{1}{2} \sum_{p=1}^P \sum_{j=1}^m (t_{pj} - y_{pj})^2. \quad (8)$$

In Equation (8), the  $t_{pj}$  ideal output is  $y_{pj}$  denoted, and the actual output is denoted. The output is adjusted according to the hidden layer node weights, and the learning rate is  $\eta$  taken in the range of 0.1 to 0.3, as shown in the following:

$$\Delta \omega_{jk} = \eta \sum_{p=1}^P \left( -\frac{\partial E_p}{\partial y_{jk}} \cdot \frac{\partial n_{pj}}{\partial \omega_{jk}} \right). \quad (9)$$

In (9), the  $\partial$  variable parameters are represented, and the adjustment formula of the weight of the hidden layer neurons of each neuron weight is finally obtained, as shown in the following:

$$\Delta v_{ki} = \eta \sum_{p=1}^P \left( \sum_{j=1}^m \sigma_{pj} \omega_{jk} \right) z_{pk} (1 - z_{pk}) x_{pi}. \quad (10)$$

In Equation (10), the  $\sigma_{pj}$  output node error signal is denoted. The principle flow of BP neural network is shown in Figure 2.

Dissolved oxygen content in aquaculture is an important indicator of aquaculture, but water quality factors are complex. Therefore, the water quality monitoring model is constructed through BP, the nonlinear factors of water quality are dealt with, the effective monitoring of dissolved oxygen in aquatic products is realized, the data model is established, and the BP principle and data processing process are analyzed.

**3.2. Construction of Aquaculture Water Quality Prediction Model Based on Improved GPCA-BP.** Considering the complex relationship between the factors of aquaculture water quality, BP neural network is used to deal with complex nonlinear problems, which has powerful computing power, but the processing and computing effect of high-

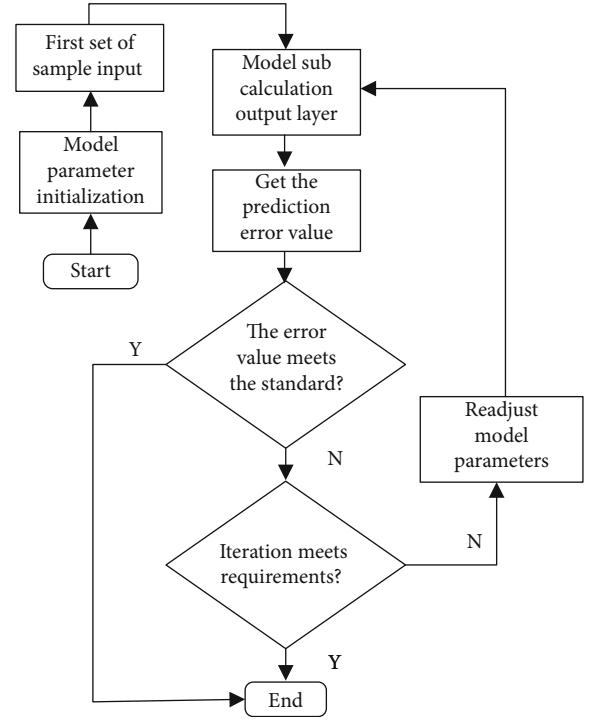


FIGURE 2: The principle flow of BP neural network.

dimensional output variables is not ideal. Therefore, the idea of combining PCA and BP algorithm is adopted to reduce the dimension of the data and establish a new PCA-BP model. The principle of the model is shown in Figure 3.

In the aquaculture water quality prediction model, various factors such as dissolved oxygen content in water quality, pH (hydrogen ion concentration) value of water quality environment, water temperature, air humidity, ammonium carbon concentration, light, and wind are comprehensively considered. It will affect the dissolved oxygen content [18]. Improve the accuracy of water quality testing. The factors affecting dissolved oxygen need to be fully analyzed. Define the original data of fish pond water quality as the  $x'$  minimum value and the  $x_{\min}$  maximum value of the original data of the fish pond  $x_{\max}$ . Use the data compression formula to compress the data  $[0, 1]$  range to better participate in model testing. The compression processing expression is as shown as follows:

$$x = \frac{x' - x_{\min}}{x_{\max} - x_{\min}}. \quad (11)$$

The data of each factor in the pond is compressed, and the standardized data table is used to process the data. The standardized data expression is shown in the following:

$$r_{ij} = \frac{1}{n} \sum_{k=1}^n \frac{(x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sigma_i \sigma_j}. \quad (12)$$

In ((12)), the input value  $\bar{x}_i$  of the  $\bar{x}_j$  normalization process is represented, the output value  $\sigma_i$  of the normalization process is represented, the input error signal is represented,

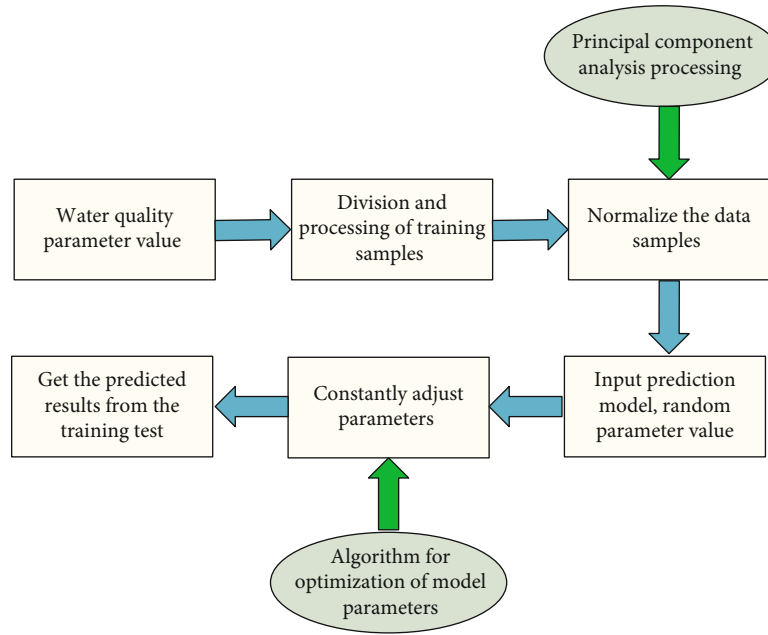


FIGURE 3: Dissolved oxygen prediction flow chart.

and the  $\sigma_j$  output error signal is represented. There are  $m$  eigenvalues defined as  $\lambda_1, \lambda_2, \dots, \lambda_m$ . According to the characteristic equation, we can find  $|\lambda I - R| = 0$  that the range  $\lambda_i \geq 0, i = 1, 2, \dots, m$  of eigenvalues is represented by the eigenvectors  $e_2, e_3, \dots, e_m$ , and then, the eigenvalues are calculated and expressed as shown in the following:

$$e_m = \begin{bmatrix} e_{1m} \\ e_{2m} \\ \dots \\ e_{mm} \end{bmatrix}. \quad (13)$$

The contribution to the data is calculated as shown in the following:

$$\tau_i = \frac{\lambda_1}{\sum_{k=1}^m \lambda_k} \quad (i = 1, 2, \dots, m). \quad (14)$$

The cumulative contribution after processing the data is shown in the following:

$$\delta \tau_i = \frac{\sum_{k=1}^i \lambda_k}{\sum_{k=1}^m \lambda_k} \quad (i = 1, 2, \dots, m). \quad (15)$$

Then, according to formula (11)-formula (15), the principal component loading matrix is calculated, and the dissolved oxygen influence factor expression as shown in formula (16) is obtained.

$$n(z_k, x_i) = \sqrt{\lambda_k u_{ki}} \quad (i, k = 1, 2, \dots, p). \quad (16)$$

In the above analysis, principal component analysis is used to reduce the dimensionality of the pond data to avoid the influence of the mutual interference between complex factors on the prediction effect of the model. However, the weight parameters and threshold parameters will directly affect the performance of the model, so consider using genetic algorithm (GA) to optimize the weights and thresholds of the BP neural network.

The GA realizes the optimal calculation of data by simulating the biological evolution process. In the optimization of BP neural network, the parameters that need BP optimization are arranged and combined into the chromosome factors of the genetic algorithm. A variety of factors constitute a chromosomal population, and the best genetic genes are selected through population fitness rules. Due to the difficulty in the selectivity of the crossover mutation process of the algorithm, the probability adaptation of the crossover mutation process needs to be adjusted to ensure that the inherited base is the optimal gene. The crossover rate adaptation is adjusted as shown in the following:

$$R_c = \begin{cases} R_{c \max} & (f' < f_{\text{avg}}), \\ \frac{R_{c \max} - R_{c \min}}{1 + \exp [2A(f' - f_{\text{avg}}) / (f_{\max} - f_{\text{avg}})]} + R_{c \min} & (f' \geq f_{\text{avg}}). \end{cases} \quad (17)$$

In (17),  $R_{c \max}$  represents the maximum value of the crossover rate,  $R_{c \min}$  represents the minimum value of the crossover rate,  $A$  represents the transformation constant and takes a value of 9.9034,  $f$  represents the change of individual adaptation, and  $f'$  represents the maximum value in

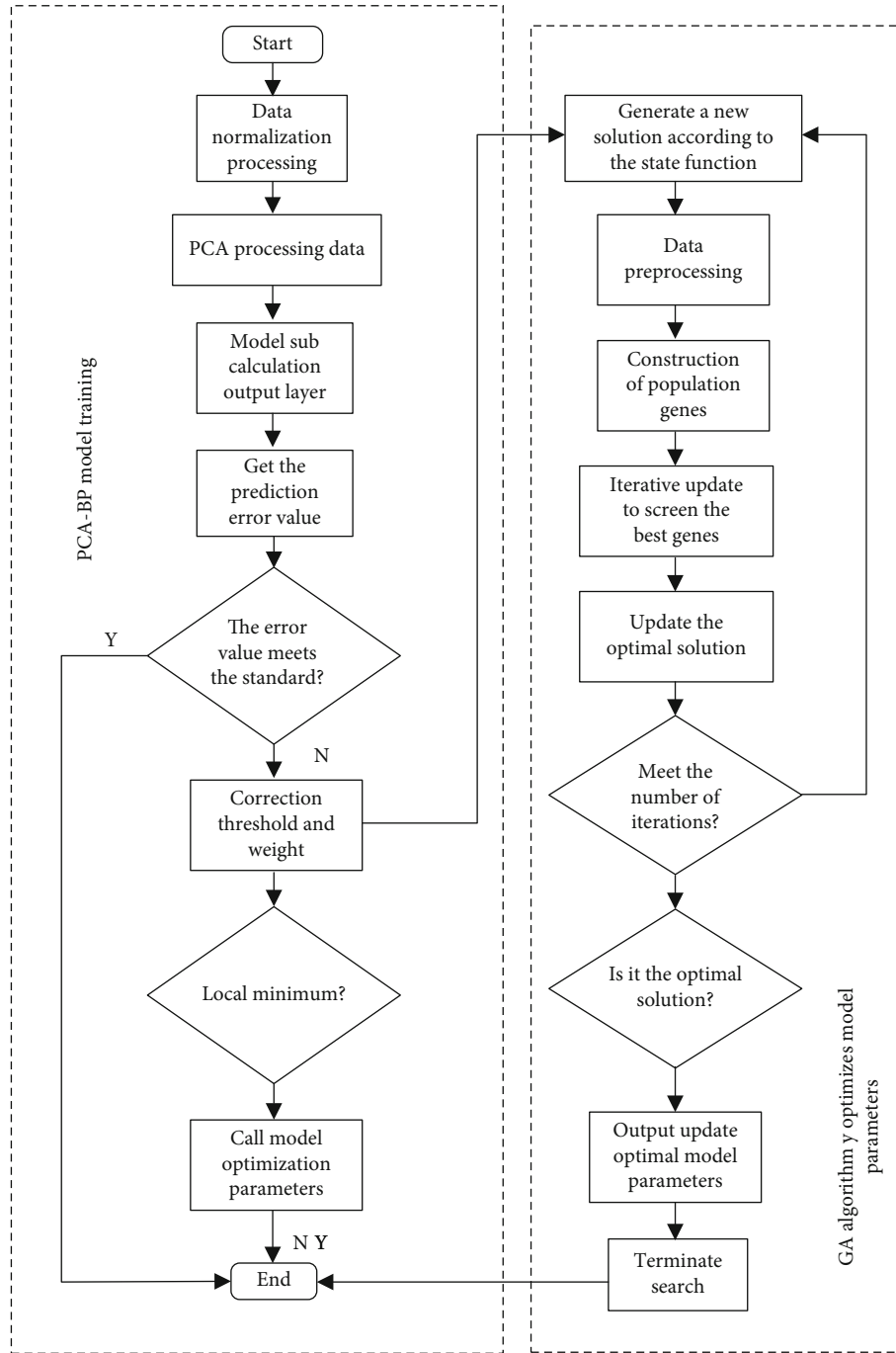


FIGURE 4: GPCA-BP water quality prediction model optimized and improved by GA.

the adaptation. Adjust as shown in Equation (18). The following is the variance rate adaptation:

$$R_m = \begin{cases} R_{m \max} & (f < f_{\text{avg}}), \\ \frac{R_{m \max} - R_{m \min}}{1 + \exp \left[ 2A(f - f_{\text{avg}}) / (f_{\max} - f_{\text{avg}}) \right]} + R_{m \min} & (f \geq f_{\text{avg}}). \end{cases} \quad (18)$$

In ((18)), the maximum value  $R_{m \max}$  of the change rate is shown, and the minimum value  $R_{m \min}$  of the change rate is shown. The parameters of the BP neural network are optimized by the GA, and the parameters are embodied in the form of individual codes, including the input layer weight parameters, the number of hidden threshold layers, the output layer, and the parameter network of the BP neural network. The individual chromosomes of the genetic algorithm are adjusted by the adaptation formulas of Equations (17) and (18). Among them, it is necessary to

TABLE 1: Initial monitoring data of aquaculture water quality.

Data collection time	Water temperature (°C)	Air temperature (°C)	Illuminance (w.m <sup>2</sup> )	Humidity (RH%)	pH	Wind value	Ammonia nitrogen concentration (mg.l)
1 day	Twenty-one	Twenty-two	1251.65	65.45	7.05	2	3.58
2 days	17	20	1345.54	75.45	6.97	2	3.78
3 days	19	Twenty-one	1054.54	68.54	6.45	2	3.89
4 days	Twenty-one	25	1454.56	61.54	6.79	2	3.85
5 days	17	Twenty-one	954.14	73.54	7.06	3	4.15
6 days	18	Twenty-two	1254.52	76.45	6.87	2	4.15
7 days	17	Twenty-three	1354.45	72.45	6.78	2	3.45
8 days	18	Twenty-four	1385.54	64.56	6.65	4	3.56
9 days	19	Twenty-one	1256.54	67.54	6.67	1	3.54
10 days	19	20	1156.5 2	78.54	6.87	2	3.45
11 days	20	Twenty-one	1005.54	76.45	6.79	2	3.75
12 days	Twenty-one	Twenty-two	1234.54	71.54	6.97	3	3.56

judge whether the chromosome settings meet the requirements and whether the decoded thresholds and weights meet the requirements. If the requirements are not met, the fitness needs to be calculated in a loop until the best iterative result is obtained. The final optimal chromosome decoding obtained by the GA is the optimal initial threshold and weight of the BP neural network. As shown in Figure 4, the test principle of the improved GPCA-BP water quality prediction model after the GA is optimized.

The traditional BP model has a good treatment effect on the complex factor relationship of water quality, but it cannot meet the requirements of high-dimensional output variable processing, so the BP model has fallen into local convergence and cannot achieve effective monitoring of dissolved oxygen in water quality. Therefore, PCA is used to reduce the dimension of high-dimensional data of water quality. At the same time, the parameter selection of the BP model affects the overall training effect of the model, so the genetic algorithm is used to optimize the parameters of the BP model to improve the overall applicability of the model.

#### 4. Test of Aquaculture Water Quality Prediction Model

The performance of the improved PCA-BP algorithm is tested, and the three-layer BP neural network structure is used for performance testing, and the data is simulated and analyzed on the MATLAB R2012b platform. The system platform is Win10, the hardware is 3.5 GHz processor, and the running memory is 8 G. Considering the randomness of the traditional BP neural network in the threshold and weight selection, the initial parameter values of the BP neural network are obtained through the GA. The population size value is set to 40, the number of hidden layer nodes of the BP neural network is set to 22, the maximum number of GA iterations is set to 100, and the crossover rate is set to 0.6. The 12-day fish pond data of a breeding base in May was collected as the experimental sample data. The data includes water temperature, humidity, pH, nitrogen, and oxygen concentration collected by the system, and the data

is a total of 3000 tests. The data were normalized and pre-processed to obtain experimental feature data. The initial data of the monitoring part of aquaculture water quality are shown in Table 1.

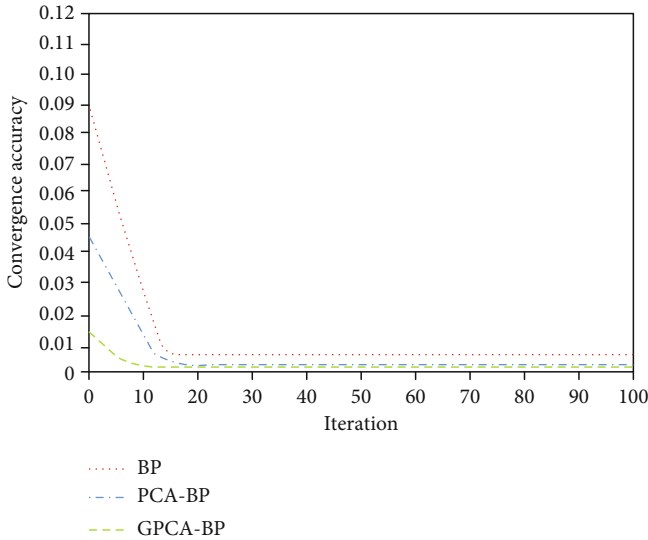
Considering the influence of the weight and threshold parameters of the PCA-BP water quality prediction model on the model performance, in order to improve the data monitoring performance of the model, the GA was used to obtain the optimal parameter neural network of BP after 100 iterations of testing, as shown in Table 2. The thresholds and weights of the BP neural network are obtained from the output of the BP neural network optimized by the GA.

The optimal initial parameters of the improved PCA-BP water quality prediction model are obtained through the GA, and the obtained parameters are the training parameters of the BP model, and the number of iterations of the BP model is set to 100 times.

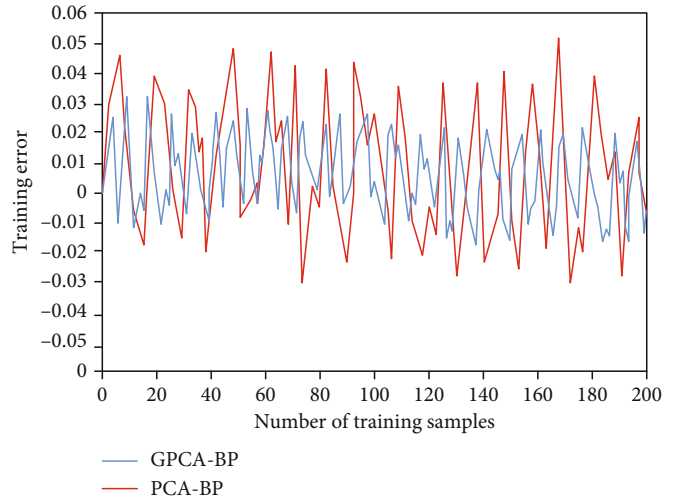
The convergence accuracy of multiple models is compared. As can be seen from Figure 5(a), compared with the traditional BP and PCA-BP models, the proposed GPCA-BP water quality prediction model has the best convergence performance, and the number of training iterations is 10. When the convergence result is reached, the convergence accuracy value is 0.012. The worst performer is the traditional BP model, which tends to converge after 18 iterations, with an accuracy value of 0.009. Figure 5(b) shows the prediction performance results of various water quality prediction models. Since traditional prediction models cannot adapt to high-latitude sample data, only PCA-BP and GPCA-BP models are compared. It can be seen from the figure that the proposed GPCA-BP water quality prediction model has more accurate prediction performance, and the overall fluctuation range is -0.02~0.03, while the traditional PCA-BP has a larger overall fluctuation, with an interval of -0.04~0.05. It can be seen that the proposed GPCA-BP water quality prediction model has good performance. Select the time period from 9:00 to 10:00 in May, and conduct 20 dissolved oxygen concentration tests on the data of aquaculture fish ponds. Before the experiment, the GPCA-BP water quality prediction model was improved, the dimension of

TABLE 2: BP neural network threshold and weight results obtained by GA optimizing BP neural network output.

Network layer type	Tier 1	Tier 2	Tier 3	Tier 4	Tier 5	Tier 6	Tier 7
Weights between input and hidden layers	-1.48648	0.01476	1.325956	0.93158	0.02403	-0.02426	-0.89025
Hidden layer node threshold	-0.62552	1.274165	1.74146	-1.38956	-1.26545	/	/
Weights between hidden and output layers	0.282906	-2.7065	-1.67065	2.675564	2.695482	/	/
Output layer node threshold	7.70154	/	/	/	/	/	/



(a) Comparison results of convergence accuracy



(b) Error comparison results

FIGURE 5: Comparison results of model convergence and accuracy prediction errors.

the data was reduced, and the initial threshold and weight were obtained through GA optimization. The predicted results of dissolved oxygen in fish ponds at a certain time are shown in Figure 6.

The red in the figure is the real measurement value, and a total of 20 groups of real data samples were measured. It can be seen from the data in the figure that the dissolved oxygen content of fish ponds in different time periods is different. Considering the influence of temperature, light, ammonia nitrogen concentration, and pH value on dissolved oxygen content, these factors were ignored, and the experimental average was taken to ensure the accuracy of the experiment. There is a big difference between the traditional BP neural network prediction model and the measured concentration curve. Sample 8 has the largest relative error, with a relative error of 1.26%. At this time, the actual value of dissolved oxygen in the fish pond was 7.76 mg.l, and the measured value was 8.06. The overall prediction performance is poor. In the PCA-BP water quality prediction model, the prediction curve is closer to the true value curve, but there are still large fluctuations. The best prediction effect is the GPCA-BP water quality prediction model, which is basically consistent with the actual predicted value curve, and the overall fluctuation is small. In sample 2, sample 4, sample 6, and sample 8, the measured dissolved oxygen concentrations were 6.95 mg.l, 7.42 mg.l, 7.16 mg.l, and 7.68 mg.l, respectively, with a rela-

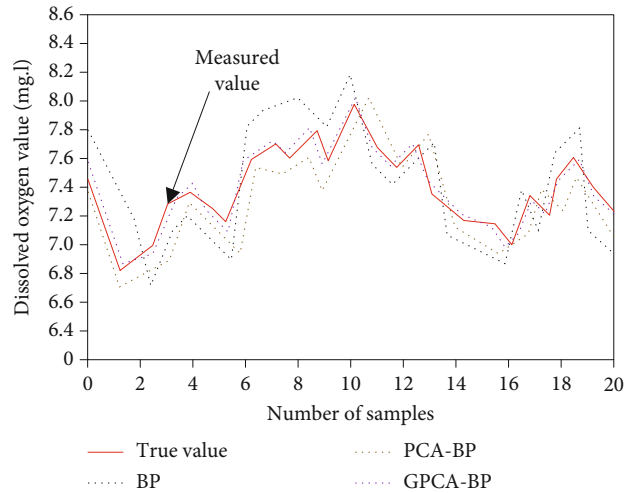
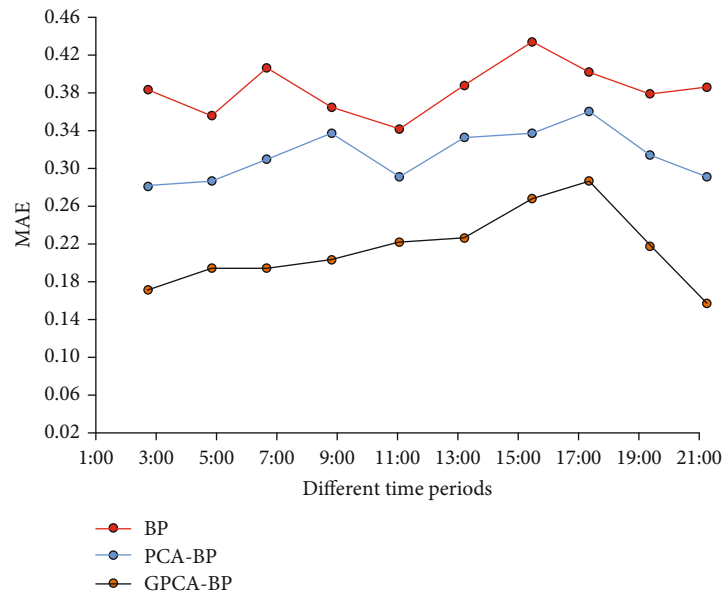


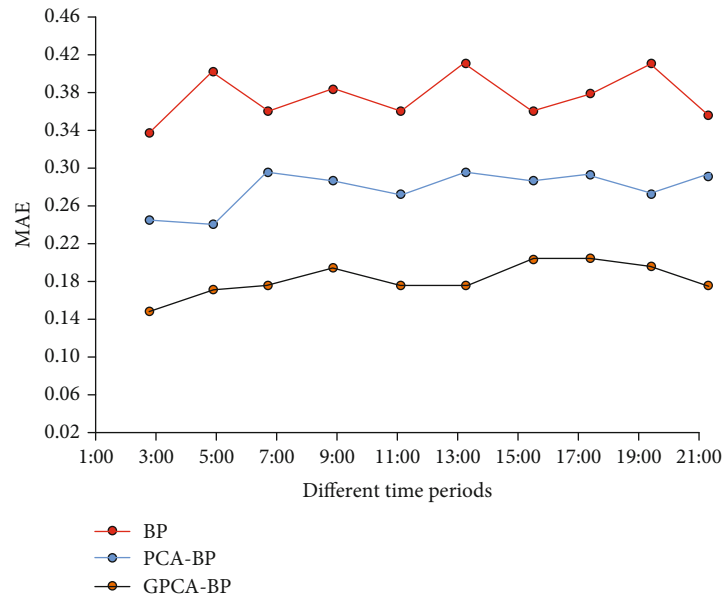
FIGURE 6: Prediction results of dissolved oxygen in fish ponds in a certain period of time.

tive error of less than 0.8%. Finally, it is shown that the GPCA-BP forecasting model has good forecasting performance in actual forecasting, considering the short test range time, less sample data, and other effects. Therefore, the mean square error (MSE) performance of different time periods and different fish pond waters was tested, and the test results are shown in Figure 7.



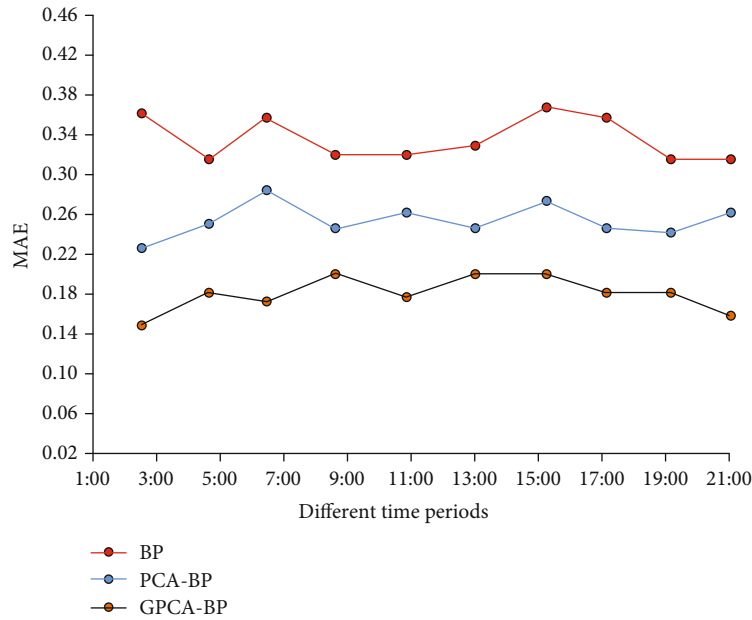


(a) Test location 1



(b) Test location 2

FIGURE 7: Continued.



(c) Test location 3

FIGURE 7: Mean square error results in different periods and regions.

Three representative areas of the pond were selected for model error performance testing. Figure 7(a) shows the results of water quality testing in the fish rearing area. The maximum error value is obtained at 15:00. Considering that the test area is a feeding area and is easily affected by complex factors, the results of the three model tests all fluctuate greatly at this moment, mainly due to the influence of air temperature and water temperature. But the MAE error performance of the proposed GPCA-BP water quality prediction model is the best. At 15:00, the MAE value is less than 0.196. Figure 7(b) is the water intake area of the fish pond. This area has the least influencing factors. Due to the high dissolved oxygen content in the water intake area, the interference of sunlight and temperature on the test results can be reduced. Therefore, the prediction accuracy of PCA-BP and GPCA-BP water quality prediction models has been improved, but the traditional BP water quality prediction model cannot handle complex high temperature data, the overall error is large, and the MAE value is greater than 0.34. Figure 7(c) is the central area of the fish pond, and the influencing factors are relatively balanced, which can reflect the predicted actual effect of most fish ponds. The traditional BP neural network has the largest error value at 15:00 MAE with a value of 0.379, and the prediction error performance is the worst. The PCA-BP prediction model also reached the maximum MAE error value at 7:00 with a value of 0.286, and the GPCA-BP water quality prediction model reached the maximum MAE error at 13:00 with a value of 0.192. It can be seen that the proposed GPCA-BP water quality prediction model has excellent prediction accuracy performance in different time periods and different regions. Finally, fish ponds were selected for 20 water

quality tests at different times and in different regions. The results of each water quality prediction model are shown in Table 3.

It can be seen from the table data that the GPCA-BP water quality prediction model has the best prediction performance for dissolved oxygen content, and the relative error between the actual value and the predicted value can be reduced to less than 0.76%, while the BP model and the GPCA-BP model have the largest errors, respectively, which were 1.96% and 1.16%. To sum up, compared with the PCA-BP model, the algorithm performance of the GPCA-BP model is stronger and the error is lower. In the actual aquatic product prediction, the accuracy rate is high, the detected real value is close to the actual value, and the water quality prediction effect is the best, which meets the requirements of modern aquaculture.

The PCA-BP model optimized by principal components and the GPCA-BP model optimized by principal components and genetic algorithms are compared. In the simple model performance test, the hybrid GPCA-BP model outperforms the first two models in both iterative performance and convergence performance. In water quality detection, the hybrid GPCA-BP model has better prediction accuracy and lower error in different time periods and water quality environments. Aquaculture has very high requirements on water quality, especially differences in dissolved oxygen concentration, pH value, water temperature, etc., which will directly affect the survival of fish. Overall, the GPCA-BP model is more in line with the requirements of aquaculture development. Compared with the traditional water quality detection scheme, the proposed scheme combines machine learning and data

TABLE 3: Results of water quality prediction model for 20 water quality experiments conducted in different periods and regions.

Data sample	BP model			PCA-BP model			GPCA-BP model		
	Actual value	Estimate	Relative error (%)	Actual value	Estimate	Relative error (%)	Actual value	Estimate	Relative error (%)
1	8.04	7.98	-1.89	8.02	7.58	-1.12	7.99	7.98	-0.19
2	8.04	9.07	1.76	8.03	8.46	1.11	8.01	8.07	0.66
3	8.01	9.06	1.76	8.01	7.66	-1.16	7.98	8.06	0.66
4	8.03	9.06	1.56	7.76	8.16	0.96	8.01	8.06	0.56
5	7.98	6.83	-1.25	7.91	7.53	-0.95	7.99	7.83	-0.25
6	7.96	6.90	-1.12	7.99	7.60	-0.72	7.96	7.90	-0.12
7	7.98	6.91	-1.56	7.99	7.51	-0.96	8.08	7.91	-0.56
8	8.01	9.89	1.45	8.02	7.49	-0.85	8.03	7.89	0.45
9	7.98	9.09	1.75	7.96	8.69	1.05	8.08	8.09	0.75
10	7.99	6.08	-1.96	7.94	7.08	-0.96	8.03	8.05	-0.66
11	8.01	9.10	1.46	8.05	8.36	0.89	8.03	8.10	0.46
12	8.04	9.06	1.35	8.01	8.96	1.03	7.97	8.06	0.35
13	7.96	9.05	1.65	7.99	8.85	1.15	8.01	8.05	0.65
14	7.99	6.90	-1.26	7.94	7.40	-0.86	7.99	7.90	-0.26
15	8.07	6.01	-1.46	8.04	7.61	-0.96	8.01	8.01	-0.46
16	8.04	9.12	1.56	8.01	8.82	1.06	8.04	8.12	0.56
17	8.04	6.97	-1.75	8.01	7.57	-0.75	8.03	7.97	-0.65
18	8.06	9.11	1.86	8.01	8.61	0.76	8.01	8.11	0.76
19	8.04	9.07	1.65	8.02	8.77	0.85	8.02	8.07	0.65
20	8.03	9.13	1.75	8.03	7.63	-1.05	8.01	8.13	0.65

mining technology and improves the traditional BP model, the overall monitoring accuracy is better, and the accuracy is better.

## 5. Discussion and Conclusion

Modern aquaculture is inseparable from the support of advanced intelligent technology. The application of information technology in aquaculture is of great significance to the development of modern fisheries. Dissolved oxygen content is an important criterion for measuring the aquaculture water quality. The relevant factors affecting water quality were analyzed, and a BP neural network water quality prediction model was constructed. Considering that the water quality monitoring data is high-dimensional data, the principal component analysis method is used to process the high-dimensional data, and the PCA-BP water quality prediction model is constructed. The weight and threshold parameters of PCA-BP water quality prediction model were optimized by genetic algorithm, and an improved GPCA-BP water quality prediction model was constructed. The experimental test results show that in the model performance test, the GPCA-BP prediction model has a shorter convergence time than the BP and PCA-BP prediction models and tends to converge after 10 iterations, with a convergence accuracy of 0.012, while in the traditional BP, the prediction model tends to converge after 18 training iterations with a convergence accuracy value of 0.009, the worst performance. In the predicted results of dissolved oxygen in fish ponds in a cer-

tain period of time, the GPCA-BP prediction model is in sample 2, sample 4, sample 6, and sample 8, the predicted value of dissolved oxygen concentration is 6.95 mg.l, 7.42 mg.l, and 7.16 mg.l, and the relative errors of 7.68 mg.l were less than 0.8%, respectively, with the best prediction accuracy. According to the mean square error results of different periods and regions, the MAE value of the GPCA-BP water quality prediction model is less than 0.196. Compared with the BP and PCA-BP models, the prediction accuracy is the best. It can be seen that, compared with the traditional aquatic product water quality monitoring scheme, the proposed scheme is more intelligent and the monitoring is more accurate. At the same time, a more advanced combined learning scheme can be used to deal with more complex water quality variables, which cannot be satisfied by traditional schemes. However, there are also problems in the research. The water quality environment still needs to fully consider the relationship between pH, temperature, and other data and needs to be further improved in the later work.

## Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

## Conflicts of Interest

The authors declare no conflict of interest in this article.

## Acknowledgments

In 2020, “Intelligent Aquaculture System Based on Artificial Intelligence Technology” was funded by the Special Project of Key Areas of Colleges and Universities in Guangdong Province (Rural Revitalization) (Project No. 2020ZDZX1077). In 2022, “Research on Key Technologies of Adaptive Intelligent Oxygenation Control System” was funded by the Special Project of Key Areas of Colleges and Universities in Guangdong Province (Rural Revitalization) (Project No. 2022ZDZX4105). In 2021, “Development and Application of Zhuhai Baijiao Sea Bass Intelligent Breeding System Under the Background of Rural Revitalization” was funded by the Social Development Science and Technology Plan Project in Zhuhai (Project No. ZH22036201210141PWC). In 2021, Zhuhai City Polytechnic supported the Natural Sciences Key Scientific Research Project “Research on DO Intelligent Monitoring and Prediction System Based on Deep Learning Algorithm in Complex Breeding Environment” (Project No. KY2021Z02Z).

## References

- [1] A. Singer, S. Garg, K. Kaur, S. Batra, N. Kumar, and K. K. R. Choo, “Fuzzy-folded bloom filter-as-a-service for big data storage in the cloud,” *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2338–2348, 2019.
- [2] Y. T. Chen, E. W. Sun, and Y. B. Lin, “Coherent quality management for big data systems: a dynamic approach for stochastic time consistency,” *Annals of Operations Research*, vol. 277, no. 1, pp. 3–32, 2019.
- [3] L.-p. Chen, H. Yin, L.-g. Yuan, A. M. Lopes, J. A. T. Machado, and R.-c. Wu, “A novel color image encryption algorithm based on a fractional-order discrete chaotic neural network and DNA sequence operations,” *Frontiers of Information Technology & Electronic Engineering*, vol. 21, no. 6, pp. 866–879, 2020.
- [4] L. Deblais, D. Kathayat, Y. A. Helmy, G. Closs Jr., and G. Rajashekara, “Translating big data: better understanding of host-pathogen interactions to control bacterial foodborne pathogens in poultry,” *Animal Health Research Reviews*, vol. 21, no. 1, pp. 15–35, 2020.
- [5] H. W. Hui, C. C. Zhou, S. G. Xu, and F. H. Lin, “A novel secure data transmission scheme in industrial Internet of things,” *China communications*, vol. 17, no. 1, pp. 73–88, 2020.
- [6] P. Ni, Y. Li, G. Li, and V. Chang, “A hybrid Siamese neural network for natural language inference in cyber-physical systems,” *ACM Transactions on Internet Technology*, vol. 21, no. 2, pp. 1–25, 2021.
- [7] H. Guan and X. Zhao, “Study on the prediction system of shrimp field distribution in the East China Sea based on big data analysis of fishing trajectories,” *Journal of Ocean University of China*, vol. 20, no. 1, pp. 228–234, 2021.
- [8] M. Kumar, “Scalable malware detection system using big data and distributed machine learning approach,” *Soft Computing*, vol. 26, no. 8, pp. 3987–4003, 2022.
- [9] V. C. Gu, B. Zhou, Q. Cao, and J. Adams, “Exploring the relationship between supplier development, big data analytics capability, and firm performance,” *Annals of Operations Research*, vol. 302, no. 1, pp. 151–172, 2021.
- [10] X. Lv and M. Li, “Application and research of the intelligent management system based on Internet of things technology in the era of big data,” *Mobile Information Systems*, vol. 2021, no. 16, Article ID 6515792, 6 pages, 2021.
- [11] P. Shi, Y. M. Yuan, H. Y. Zhang, and L. Kuang, “Monitoring system of aquaculture water quality based on wireless sensor network,” *Journal of Anhui Agricultural Sciences*, vol. 49, no. 5, pp. 207–210, 2021.
- [12] B. M. Mathisen, K. Bach, and A. Aamodt, “Using extended Siamese networks to provide decision support in aquaculture operations,” *Applied Intelligence*, vol. 51, no. 11, pp. 8107–8118, 2021.
- [13] J. Zhang, “Exploration on coal mining-induced rockburst prediction using Internet of things and deep neural network,” *Journal of Supercomputing*, vol. 78, no. 12, pp. 13988–14008, 2022.
- [14] S. Zhang, Y. Guo, S. Li et al., “Investigation on environment monitoring system for a combination of hydroponics and aquaculture in greenhouse,” *Information Processing in Agriculture*, vol. 9, no. 1, pp. 123–134, 2022.
- [15] Y. Zhang, W. H. Chu, and M. Ahmad, “A seismic energy-based prediction model for soil liquefaction using neural networks,” *Environmental Geosciences*, vol. 81, no. 4, pp. 1–12, 2022.
- [16] S. Moon, H. Kim, Y. H. You, C. H. Kim, and I. Hwang, “Deep neural network for beam and blockage prediction in 3GPP-based indoor hotspot environments,” *Wireless Personal Communications*, vol. 124, no. 4, pp. 3287–3306, 2022.
- [17] E. Elbasani, S. N. Njimbouom, T. J. Oh, E. H. Kim, H. Lee, and J. D. Kim, “GCRNN: graph convolutional recurrent neural network for compound-protein interaction prediction,” *BMC Bioinformatics*, vol. 22, Supplement 5, pp. 1–14, 2022.
- [18] Z. J. Li, C. Y. Zhao, B. C. Wen, and Z. C. Lu, “Heat source rate identification and thermal error predictions of ball screw feed drive system for CNC machine tools,” *Journal of Northeastern University (Natural Science)*, vol. 40, no. 9, pp. 1305–1309, 2019.