Computers and Electronics in Agriculture 198 (2022) 106971

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

A hybrid model for the prediction of dissolved oxygen in seabass farming

Jianjun Guo [a,c,d,1], Jiaqi Dong [a,b,c,d,1], Bing Zhou [a,c,d], Xuehua Zhao [b], Shuangyin Liu [a,c,d], Qianyu Han [a,c,d], Huilin Wu [e], Longqin Xu [a,c,d,*], Shahbaz Gul Hassan [a,c,d,*]

[a] Guangzhou Key Laboratory of Agricultural Products Quality & Safety Traceability Information Technology, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China

[b] School of Digital Media, Shenzhen Institute of Information Technology, Shenzhen 518172, China

[c] College of Information Science and Technology, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China

[d] Academy of Intelligent Agricultural Engineering Innovations, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China

[e] National S&T Innovation Center for Modern Agricultural Industry (Guangzhou) (Short for GNASIC), China

ARTICLE INFO

Keywords:
Seabass farming
Dissolved oxygen
Pathfinder algorithm
Principal component analysis
Gated recurrent unit
Parameter optimization

ABSTRACT

Perch is a relatively valuable aquatic product with high economic value. Dissolved oxygen follows a complex, dynamic and non-linear system. To solve the problems of low prediction accuracy and poor generalization ability of traditional dissolved oxygen prediction methods, a dissolved oxygen hybrid prediction model for perch culture water quality based on principal component analysis and pathfinder optimization algorithm is proposed in this paper. Firstly, the key influencing factors affecting the dissolved oxygen of bass were extracted by PCA to eliminate redundant variables and reduce the data dimension and complexity. Then the PFA optimization algorithm is used to automatically optimize the key parameters of GRU neural network to obtain the optimal parameter combination. Finally, a combined prediction model based on PCA-PFA-GRU is constructed to predict the dissolved oxygen in perch culture water quality. The MSE, MAE, RMSE and R^2 are 0.010, 0.060, 0.100 and 0.983, respectively. The simulation results show that the proposed PCA-PFA-GRU model has a small fluctuation of prediction error and high prediction accuracy. In conclusion, the proposed model has good prediction accuracy and generalization and has achieved excellent prediction effect in short-term prediction to avoid huge losses, reduce growth risks and promote the development of fishery modernization.

1. Introduction

China is the world's largest aquaculture country, the total aquatic products ranking first in the world for 29 consecutive years in which fishery occupies a dominant position (Hu et al., 2021; Li et al., 2018). Perch is a kind of precious aquatic product, which has high economic value. However, it requires high water quality and is sensitive to dissolved oxygen content (Zhang et al., 2021; Fang et al., 2016). In the intensive culture pond, the amount of dissolved oxygen (DO) seriously affects the development and health of bass. Low DO may lead to the decrease of feeding, growth stagnation, physical decline and even death of bass. Dissolved oxygen content is too saturated, easy to cause bubble disease harm (Abdel-Tawwab et al., 2019; Zhang et al., 2019). Changes in dissolved oxygen content in intensive aquaculture ponds are susceptible to the influence of other water quality parameters with complex coupling relationships (Shi et al., 2019), and the dissolved oxygen prediction models constructed by existing prediction methods (Kisi et al., 2020) have certain limitations, constrained generalization ability, easy to fall into local extremes, and low prediction accuracy, which are difficult to meet the needs of intensive bass aquaculture. Therefore, establishing an accurate model for predicting future trend of dissolved oxygen is not only an effective means to prevent the deterioration of water quality and the outbreak of diseases, but also an important way to avoid huge losses, reduce growth risks and promote the development of fisheries modernization (Rahman et al., 2020).

Dissolved oxygen (DO), as one of the critical elements of the aquatic environment, is affected by complex, dynamic, and non-linear factors (Yin et al., 2021; Sun et al., 2021). At the moment, scholars at home and abroad have conducted many experiments on the complex characteristics of dissolved oxygen and developed several effective prediction models. Wu presented a BP neural network model based on particle swarm optimization to forecast dissolved oxygen in water (Wu et al.,

* Corresponding authors at: College of Information Science and Technology, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China.
E-mail addresses: xlqlw@126.com (L. Xu), mhasan387@zhku.edu.cn (S.G. Hassan).

[1] Contributed equally to this work.

https://doi.org/10.1016/j.compag.2022.106971
Received 19 January 2022; Received in revised form 3 April 2022; Accepted 11 April 2022
Available online 1 June 2022
0168-1699/

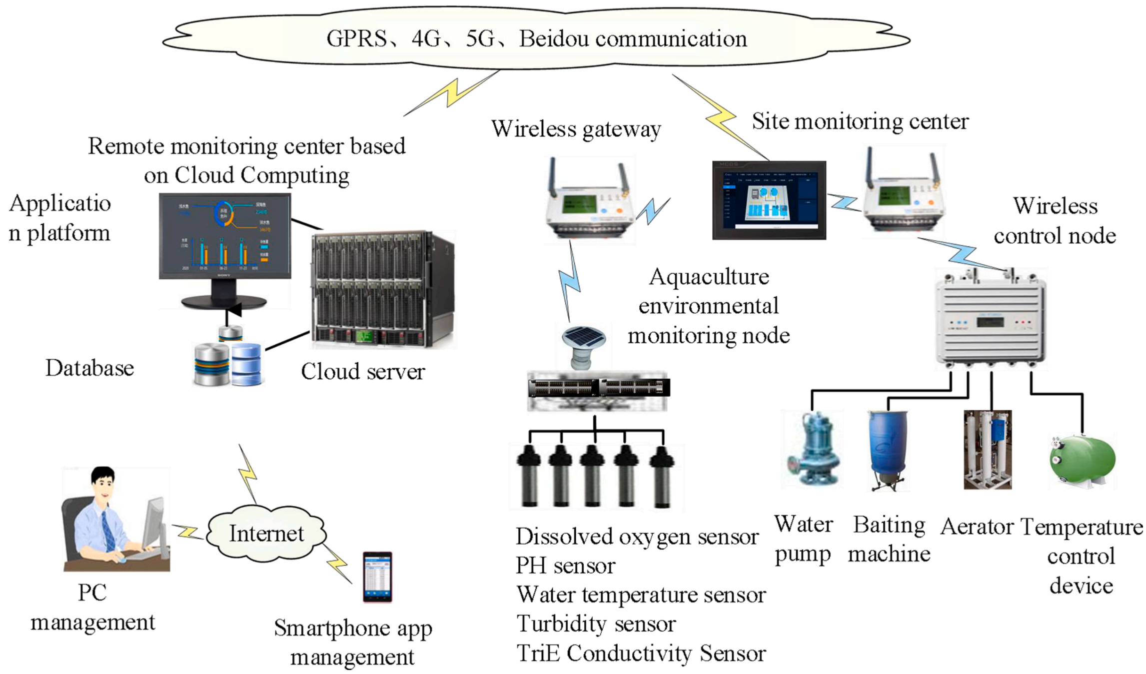

Fig. 1. Topology structure diagram of a platform.

2018). Cao employed a combination of grey correlation, empirical mode decomposition (EEMD), and regularized extreme learning machine (RELM) to make predictions and obtained favorable results in order to comprehend the time-dependent trend in dissolved oxygen in aquaculture (Cao et al., 2019). Juan constructed a dissolved oxygen prediction model based on K-means clustering and an ELM neural network using a small data sample. Liu utilized a combined model of ant colony optimization and least squares support vector regression to predict and get accurate results for the trend of dissolved oxygen content in crab culture (Liu et al., 2013). The artificial neural network and mixed wavelet prediction methods were utilized to forecast dissolved oxygen during the monitoring and assessment of river water quality (Ravansalar et al., 2016).

Additionally, an inverse understanding convolutional neural network prediction model is presented to handle the problem of dissolved oxygen prediction (Ta and Wei, 2018). In contrast, these approaches have produced accurate predictions. There are still some concerns with the actual prediction process, such as model parameter unpredictability, prediction result instability, and applicability. It is not easy to manage the dissolved oxygen concentration well enough to match the real bass production.

Deep learning has been extensively used in various fields in recent years. The LSTM (Long and Short Term Memory Network) model is particularly adaptable to time series prediction among the several deep learning models (Fang et al., 2021; Zhang et al., 2021; Chaudhary et al., 2018). Huan developed a dissolved oxygen prediction model for conventional ponds using a mix of gradient boosting decision trees (GBDT) and long short-term memory networks (LSTM). This model has a high prediction accuracy and generalization (Huan et al., 2020). However, model parameters must be modified depending on subjective experience, which is unpredictable. Li proposed a hybrid model for predicting the mass concentration of dissolved oxygen in aquaculture using sparse autoencoders (SAE) and long short-term memory networks (LSTM) (Li et al., 2018). This approach is more accurate and stable than conventional procedures. However, multiple experiments are required to discover the model's best parameters, which is time-consuming. Li developed GRU algorithm based on a prediction model for dissolved oxygen (DO) in fisheries ponds (Li et al., 2021). As a version of LSTM, this approach performs similarly to LDM. It is more appropriate for practical use due to its simple structure and reduced time cost. However, this work does not examine the redundancy of other elements. Long short-term memory networks have shown outstanding performance in various disciplines (Shi et al., 2021; Banik et al., 2021; Huang et al., 2020). As a version of LSTM, GRU may enhance prediction accuracy and efficiency, making it more appropriate for practical use. However, since there is a lack of research on the prediction of dissolved oxygen in sea bass aquaculture water, this study is significant in terms of research.

According to the aforementioned existing studies, there is a lack of accuracy and reliability of dissolved oxygen prediction in aquaculture ponds, raising the issue of redundant information, complex calculation, and poor prediction performance of traditional dissolved oxygen prediction models. For this purpose, a combination prediction model based on Principal Component Analysis (PCA), Pathfinder Algorithm (PFA), and Gated Current Unit (GRU) is proposed. Additionally, Principal component analysis filtered out the critical parameters influencing dissolved oxygen in sea bass culture to minimize the model's complexity. Then, using the pathfinder optimization algorithm, the gating loop unit's critical parameters were optimized to reach the best solution. Finally, a combination PCA-PFA-GRU prediction model was developed to estimate the dissolved oxygen content of sea bass aquaculture water. This model was used for an empirical analysis of water quality in Guangzhou City, Guangdong Province, and the experimental results showed that the combined prediction model proposed in this research outperforms existing models in the field of dissolved oxygen prediction, which may provide a novel approach to the challenging issue of water quality dissolved oxygen prediction in bass culture.

2. Data acquisition

2.1. Experiment area

The experimental data were collected from an experimental aquaculture base (23.11° N, 113.27° E) in Nansha District, Guangzhou City, Guangdong Province, with around 200 m and a maximum water depth of 1.5 m. The pond has a deep-water pump, a dead fish collector, a current sensor, an oxygen sensor, a UV filter, a mechanical filter, and an oxygen cylinder, among other features. And, as a complete test foundation for intelligent fisheries breeding, it is based on the Internet of

Table 1
Raw data of perch culture water quality and environment.

Time	Dissolved oxygen/ (mg·L-1)	Water temperature/°C	Conductivity/(S·m-1)	Turbidity/NTU	pH value
2020-12-07 18:57:09	10.6	17.6	1.2	4.8	6.6
2020-12-07 18:47:09	10.4	18	1.2	6.5	6.7
2020-12-07 18:37:09	10.4	17.9	1.2	0.3	6.6
……	……	……	……	……	……
2021-01-18 22:34:42	11.2	15.9	821.5	5.8	7.4
2021-01-18 22:24:42	11.2	15.9	802.7	1.6	7.5
2021-01-18 22:14:42	11.2	15.9	813.8	4.1	7.5

Things intensive sea bass aquaculture water quality real-time remote online monitoring system.

2.2. Data collection

A cloud service platform of intelligent agricultural IoT for water quality parameters of bass culture in the fishing ground is developed by Zhongkai University of Agriculture and Engineering. The data is collected by sensors every 10 min in the pond and stored on a server that can be viewed in real-time at the terminal using an IoT architecture shown in Fig. 1. The collected data include dissolved oxygen, water temperature, conductivity, turbidity and pH. Some environmental parameters of sea bass aquaculture water are shown in Table 1.

3. Models and principles

3.1. Principal component Analysis (PCA)

Principal Component Analysis (PCA) is based on matrix factorization. N feature vectors are mapped to a few new features that reflect the comprehensive information (Lee and Jemain, 2021). The main difference between principal component analysis and feature selection is that the matrix does not select the features with the most information in the original matrix but reconstructs the existing features to create new features not present in the original matrix. The following are the specific steps.

Step1: Let dataset as:

$$X = \begin{matrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{matrix} \tag{1}$$

The original data is normalized to eliminate dimensional effects.

Step2: Computing covariance matrix.

$$R_{ij} = \frac{\sum_{k=1}^{n}(X_{ij}-X_i)(X_{kj}-X_j)}{\sqrt{\sum_{k=1}^{n}(X_{kj}-X_j)^2(X_{kj}-X_j)^2}} \tag{2}$$

Determine the number of principal components according to the contribution rate of cumulative variance, usually $\geq$ 85%. The cumulative variance contribution rate size indicates that the first N principal components contain the original data information. Among them, the contribution rate of the i$^{\text{th}}$ principal component F_i is $\frac{\lambda_i}{\sum_{i=1}^{n}\lambda_i}$, The formula for calculating the cumulative contribution rate of the first N principal components is:

$$\sum_{i=1}^{n} F_i = \frac{\sum_{i=1}^{n}\lambda_i}{\sum_{i=1}^{m}\lambda_i} \tag{3}$$

Step 4: Perform dimensionality reduction processing on the original data.

3.2. Pathfinder optimization

Yapici and Cetinkaya proposed a new meta-heuristic algorithm, the Pathfinder Algorithm (PFA), in 2019 (Yapici and Cetinkaya, 2019). The concept is derived from the collective movement of animal groups. It imitates the group's leadership in locating the best prey in the region. PFA is easy to build, simple to comprehend, capable of converging to the global optimum solution and demonstrates significant advantage in addressing real-world problems with unknown search space and a high degree of difficulty. The mathematical representation is as follows: Assuming that each member has a location in two-dimensional, three-dimensional, or D-dimensional space, a member will be selected as the leader if it is in the most promising region at any time. Suppose individual position vectors represent all possible solutions to a problem. In that case, group members may step in two-, three-, or D-dimensional space. To locate prey or foraging regions and follow the pathfinder, its model is as follows:

$$x_i^{k+1} = x_i^k + R_1\cdot(x_j^k - x_i^k) + R_2\cdot(x_p^k - x_i^k) + \varepsilon, i \geq 2 \tag{4}$$

In the formula, k is the current iteration number, x_i^k is the current position of the follower, x_i^{k+1} is the updated position, x_i^k is the neighbor of x_j^k, x_p^k represents the Kth time position vector of the iterative pathfinder ; R_1 and R_2 are random vectors, where $R_1 = \alpha\cdot r_1$, $R_2 = \beta\cdot r_2$, α is the interaction coefficient between the member and the neighbor, β the attraction of the pathfinder to the follower The coefficients, α and β are randomly selected in the range of (Hu et al., 2021; Li et al., 2018); r_1 and r_2 are random variables uniformly generated in the range of [0, 1] ; ε is the randomness of the movement of the follower, $\varepsilon = (1 - \frac{k}{k_{max}})\cdot u_1\cdot D_{ij}$, $D_{ij} = ||x_i - x_j||$, where u_1 is a random vector in the range of [−1, 1] , which affects the individual's moving step length, D_{ij} is the current follower and other members the distance.

In PFA, no member is required to specify their pace. PFA has a leader, and members follow the leader based on their neighbor's location and the leader's movement while simultaneously updating the search space. The members' interaction decides their subsequent place. Members and pathfinders are on the same level and operate independently. The distance between neighboring members is lowered during continuous iteration. Random movement may be carried out according to the parameters specified by the user. A random search is used to locate the global optimum.

3.3. Gated Recurrent Unit (GRU)

Gated Recurrent Unit (GRU) is a variant of Long Short-term Memory (LSTM), which not only preserves the ability of LSTM to solve the problem of gradient disappearance and gradient explosion, but also in the central processing unit Time convergence and parameter updating and generalization are superior than LSTM (Sak et al., 2014; Chung et al., 2014). The internal unit structure of GRU is somewhat different from LSTM. LSTM has three gate structures, but GRU has two gate structures: update gate and reset gate (Dey and Salem, 2017). The update gate is used to manage the retention of the state information at the previous instant. The reset gate controls whether the current state information is mixed with the prior state information. The GRU unit

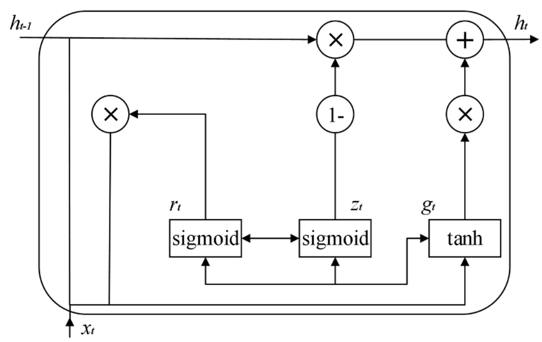

Fig. 2. The structure of the GRU cell.

structure is presented in Fig. 2:

In the figure, x_t and h_t are the input and output of the hidden layer, r_t and z_t are the reset gate and the update gate, g_t is the candidate hidden state at time t, h_{t-1} is the output of the hidden layer at the previous time.

GRU can calculate the input as:

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \tag{5}$$

$$z_t = \sigma(U_z h_{t-1} + W_z x_t + b_z) \tag{6}$$

$$g_t = tanh(U_c(r_t \cdot h_{t-1}) + W_c x_t + b_c) \tag{7}$$

$$h_t = (1 - z_t) \cdot g_t + z_t \cdot h_{t-1} \tag{8}$$

where σ is the sigmoid activation function, *tanh* is the activation function of the candidate hidden state, W_r, U_r are the reset gate weights, W_z, U_z are the update gate weights, W_c, U_c are the weights when the cell state g_t is formed, b_r, b_z, b_c are its bias vector.

3.4. PCA-PFA-GRU neural network prediction model

This article proposes a combined prediction model based on PCA, PFA, and GRU neural networks for forecasting the future trend of dissolved oxygen in sea bass breeding ponds. PCA reduces the data dimension while maintaining the integrity of adequate information and increasing calculation speed. The GRU neural network has a strong time-series capability. Simultaneously, the Pathfinder optimization technique is used to optimize the parameters of the GRU neural network in order to increase the model's prediction accuracy, guarantee that the model has a global optimum, and avoid doing a large number of repeated tests. The specific processes of the procedure are shown in Fig. 3. The following are the particular steps:

Step1: Real-time online collection of time series data on water quality parameters for sea bass aquaculture using an intelligent agricultural IoT cloud service platform, repairing and normalizing the original data before dividing the training and test sets.

Step2: The principal component analysis approach filters for significant influencing variables of dissolved oxygen in sea bass water quality. Furthermore, remove duplicate and unnecessary data and minimize computing complexity. If the cumulative variance contribution rate is more than 85%, dimensionality reduction is conducted; otherwise, the number of primary components is reselected.

Step3: Optimize the GRU neural network model using PFA: Clarify the GRU neural network's startup settings and the hyperparameters to optimize. Initialize the PFA's search range to the preset range, establish the population, compute fitness, and assign the explorer and follower roles. Compare and investigate the fitness value and position of the follower and follower, determine their optimal positions, calculate a new round of fitness values, and update the global optimal. When the search process reaches the preset maximum range or the fitness value no longer changes significantly; Stop updating and retrieve the GRU model's optimum hyperparameter combination at this point.

Step4: Input the optimized hyperparameters obtained in Step 3 into the GRU model and apply the optimized PCA-PFA-GRU combined prediction model to the sea bass breeding water quality dissolved oxygen prediction, analyzing the evaluation indicators to realize the future of sea bass breeding water quality dissolved precise oxygen prediction.

4. Simulation experiment verification

4.1. Experiment setup and data collection

The computer is equipped with an Intel I5-5200U processor with a processor speed of 1.26 GHz and the operating system Windows 10. (64-bit). Anaconda3 is used as the test software platform, while Python 3.7 is used as the programming language (64-bit). Keras is used to develop a hybrid model based on the PCA-PFA-GRU to predict dissolved oxygen in sea bass farming, because keras has advantage of good scalability, modularity, and free combination of training model layers.

The sample data for this article is from December 7, 2020 to January 18, 2021, and includes five environmental factors, totaling 6051 dissolved oxygen parameter data for sea bass breeding water quality, of

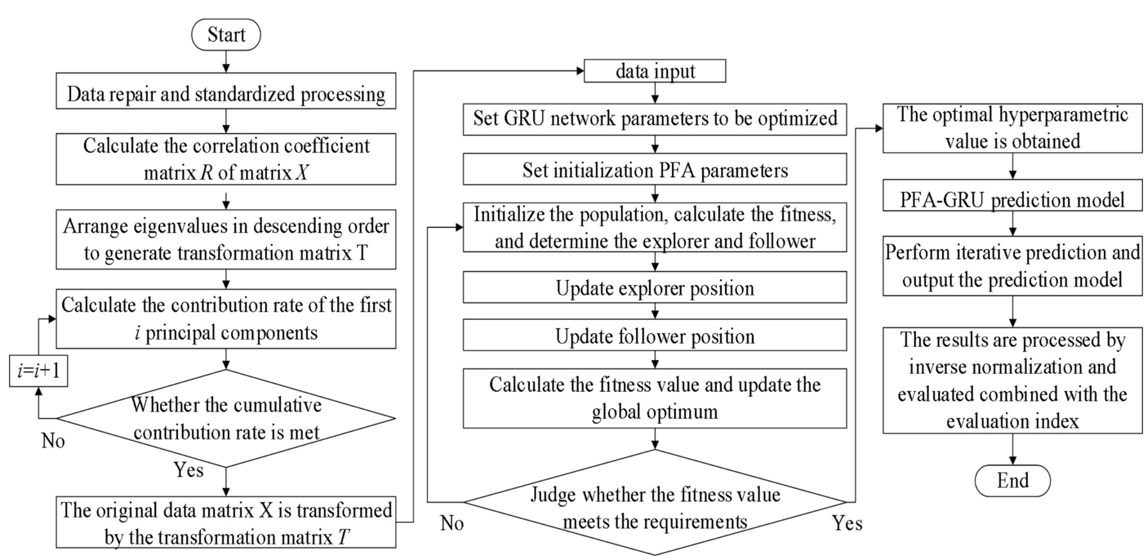

Fig. 3. Flow chart of PFA-GRU algorithm.

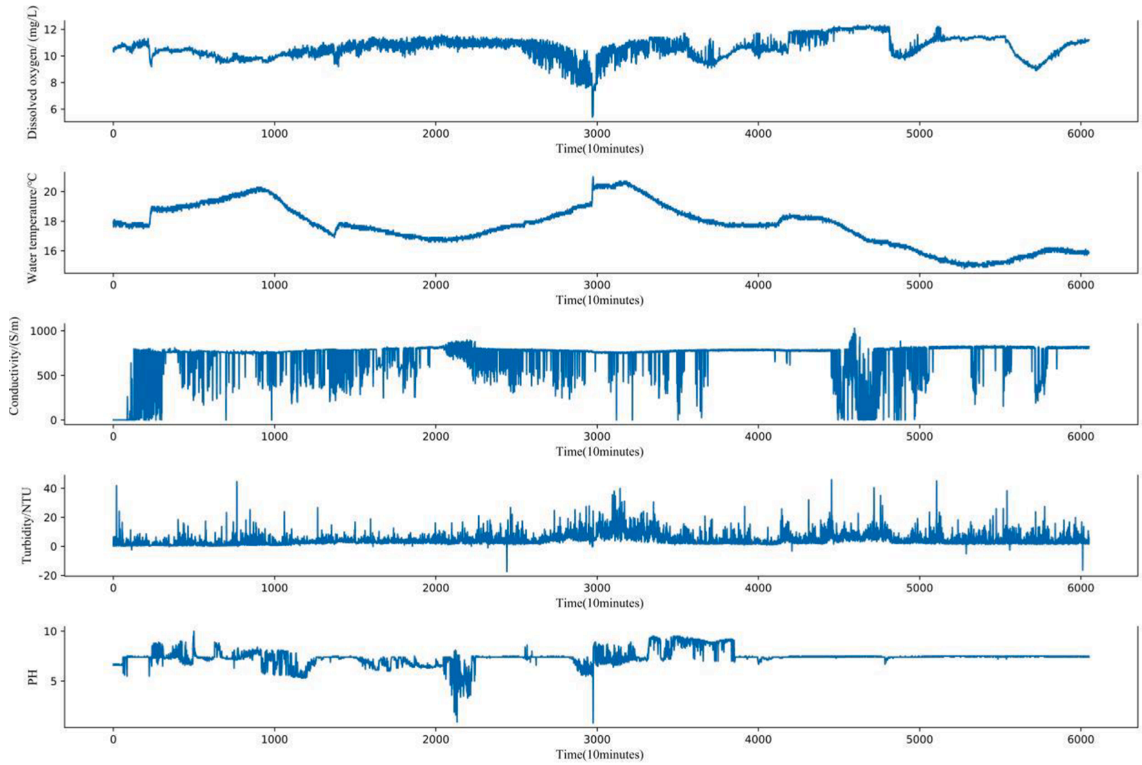

Fig. 4. Raw data of water quality parameters of perch culture.

Table 2
Eigenvalue and principal component contribution rate.

Component	Initial eigenvalues			Extraction sums of squared loading		
	characteristic value	Variance %	Cumulative %	characteristic value	Variance %	Cumulative %
1	0.0548233	47.3	47.3	0.0548233	47.3	47.3
2	0.03917745	33.8	81.1	0.03917745	33.8	81.1
3	0.01106025	9.5	90.6	0.01106025	9.5	90.6
4	0.00659623	5.7	96.3			
5	0.00419794	3.7	100			

which 1,000 are utilized for testing. The 5051 data points are utilized for training purposes. The water quality environmental parameters used in sea bass farming are non-linear and non-stationary. Fig. 4 illustrates the original data change curve.

4.2. Date processing

Due to the effect of sensor aging, line failure, network delay, and difficult weather, data loss and irregularity are likely to occur during the online collection of water quality environmental data for bass breeding. Repair pond water quality environmental data using linear interpolation and mean smoothing. In order to eliminate the adverse effects caused by singular sample data and accelerate the speed of solving the optimal solution, the original data is normalized and mapped to the range of [0, 1]. The procedure of normalizing is as follows:

$$x_n = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (9)$$

where x_{max} is the maximum value of the variable x in the sample, x_{min} is the minimum value. In the same way, demoralize the predicted result to get the actual result value.

Table 3
Principal component load matrix.

Parameters	Main ingredient 1	Main ingredient 2	Main ingredient 3
Dissolved oxygen	−0.230	0.972	−0.015
Water temperature	0.084	0.005	−0.996
Conductivity	−0.944	−0.216	−0.079
Turbidity	0.022	0.082	0.022
pH value	0.021	−0.039	−0.013

4.3. Dimensionality reduction

Numerous variables influence dissolved oxygen, and the mechanism of action is complicated. If a certain element is ignored directly, prediction errors may occur. Simultaneously, principal component analysis (PCA) analyzes the signal to increase efficiency and minimize processing complexity. The critical influencing variables of dissolved oxygen in aquaculture water quality were screened. The eigenvalues and contribution rate of the major component were determined as shown in Table 2 and the main component loading matrix as shown in Table 3.

As shown in Table 2, the cumulative contribution rate of the first three impact factors has reached 90.6 percent, ensuring that the majority of the original data is retained and according to the premise of principal component analysis. As a result, the first three indications are

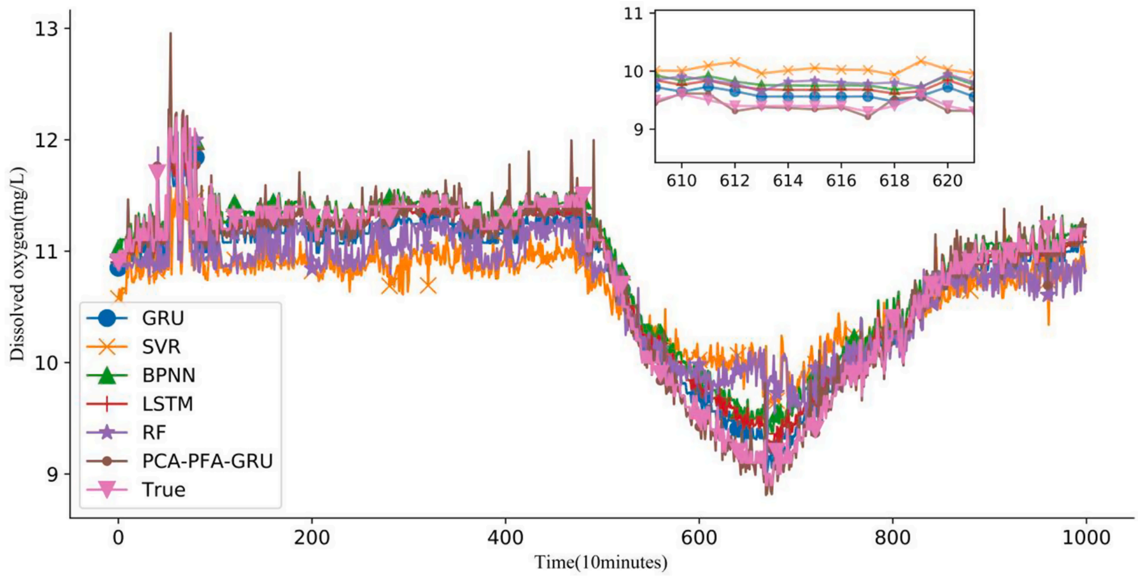

Fig. 5. Comparison of the first mock exam and PCA-PFA-GRU error curve.

chosen to substitute for the original variables.

After extracting the components, the larger the load associated with each major component, the bigger the quantity of original information reflected by the principal component. As shown in Table 3, electrical conductivity has the greatest effect on the first principal component, dissolved oxygen has the greatest effect on the second principal component, and water temperature affects the third major component. As a result, this study uses dissolved oxygen, water temperature, and electrical conductivity as the primary component indicators, mostly compatible with the major influencing elements of dissolved oxygen chosen by aquaculture experts based on their experience. These variables serve as input for the PFA-GRU prediction model, ensuring the prediction's correctness. Simultaneously, variable dimensions are minimized and computation speed is raised.

4.4. Simulation results and analysis

After reducing the dimensionality of the data through the preceding experiments, the method proposed in Section 3.4 is used to establish a prediction model of dissolved oxygen in sea bass breeding water quality in order to predict the dissolved oxygen content of sea bass breeding water quality in the next ten minutes. The initial parameters of the Pathfinder algorithm are as follows: population size is 100, maximum number of iterations is 750, interaction coefficient between neighbors is 1, and Pathfinder's attraction coefficient to followers is 2. The ideal learning rate for the GRU neural network was determined to be 0.001, the optimal number of iterations to be 100, the optimal sample batch to be 128, and the optimal number of neurons to be 99. The prediction effect is at its peak at the moment.

The MSE (Mean Square Error), MAE (Mean absolute error), RMSE (Root Mean Squared Error), and R2 (coefficient of determination) are all

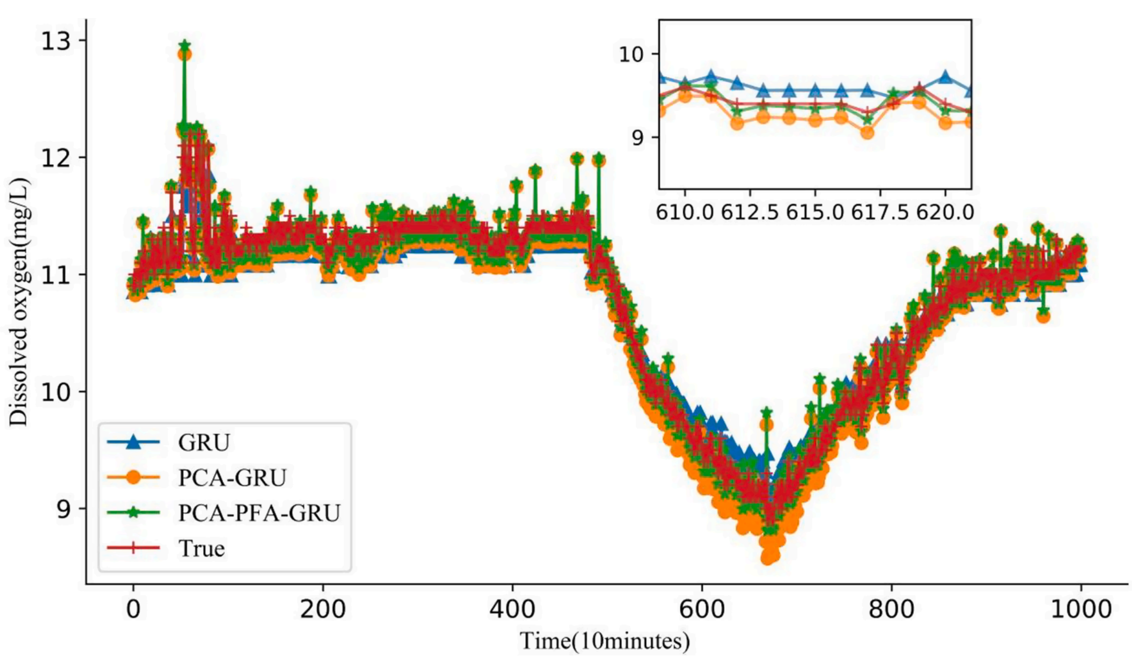

Fig. 6. Comparison of GRU, PCA-GRU and PCA-PFA-GRU error curve.

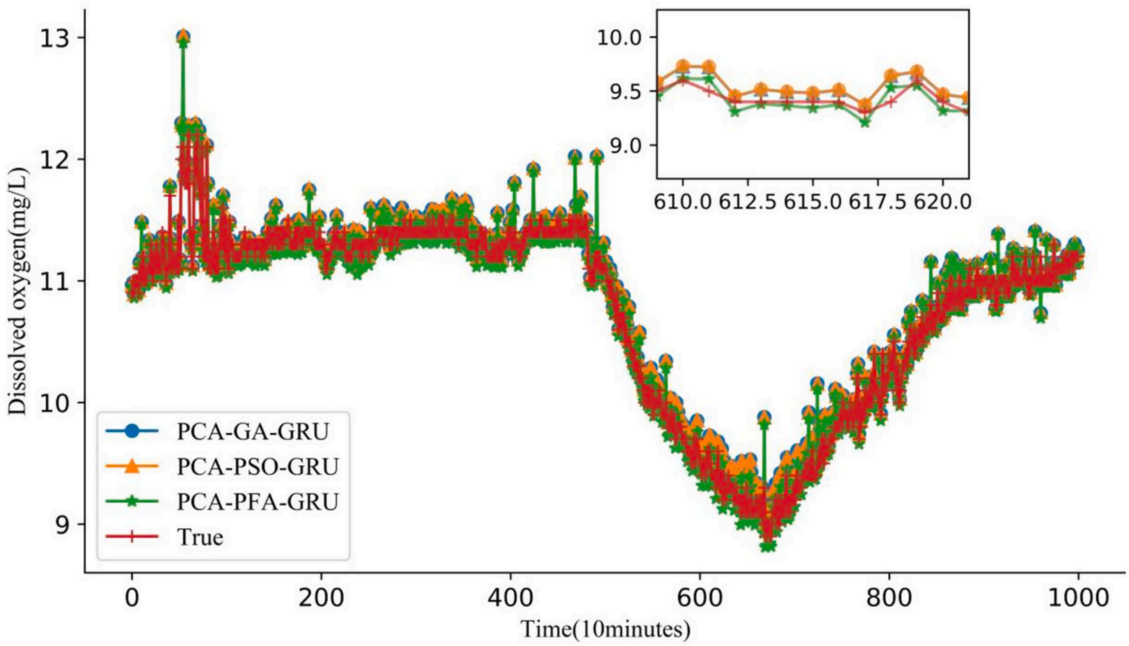

Fig. 7. Comparison diagram of optimized GRU neural network after dimensionality reduction based on PCA.

used in this article (Afrin and Yodo, 2021). The following formula is used to calculate the evaluation index for the forecast result:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}[y_i - g(x_i)]^2 \tag{10}$$

$$MAE = \frac{1}{n}|y_i - g(x_i)| \tag{11}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[y_i - g(x_i)]^2} \tag{12}$$

$$R^2 = 1 - \frac{\sum[y_i - g(x_i)]^2}{\sum[y_i - \bar{y}]^2} \tag{13}$$

In the formula, y_i is the true value of dissolved oxygen content in the water quality of sea bass culture, $g(x_i)$ is the predicted value of dissolved oxygen content in the water quality of sea bass culture, $\bar{y}$ is the average value of the real results, and n is the number of samples. Among these, a bigger RMSE, MAE, or MSE value indicates a wider variance between the model's and the actual value; a larger R2number indicates a stronger correlation between the model's and the actual value a better prediction effect.

To demonstrate this model's superiority, we compare it to other models, including a single model GRU, LSTM, SVR, RF, and BPNN, as well as a PCA-based dimensionality reduction model, PCA-GRU, PCA-GA-GRU, PCA-PSO-GRU, and PCA-PFA-GRU, which all generate predictions on the same data set. As shown in Fig. 5, the prediction results of many prediction models are consistent with the actual value trends. In contrast, the prediction results of a single model have significant performance variances. The PCA-PFA-GRU prediction curve proposed in this research is the closest to the real curve, followed by the GRU model; the random forest (RF) and support vector machine (SVR) models show considerable fluctuation, with the poorest impact of fitting the actual curve. As shown in Fig. 6, the fitting curve of the GRU neural network after PCA is more similar to the real curve than the fitting curve of the GRU single model. The data after PCA dimension reduction eliminates the interference dimension with less information. Increase the predictive model's accuracy. However, the PCA-PFA-GRU combined prediction model provided in this study remains the best at the moment. When the GRU hyperparameters are adjusted using PFA, a more optimal hyperparameter combination based on PCA-PFA may be achieved. The GRU combined prediction model exhibits superior generalization ability to predict dissolved oxygen in sea bass culture water quality.

Table 4
Experimental results for different prediction models.

Model	Error type			
	MSE	MAE	RMSE	R^2
SVR	0.191	0.386	0.438	−0.251
BPNN	0.044	0.156	0.210	0.895
RF	0.120	0.282	0.346	0.556
LSTM	0.029	0.125	0.169	0.932
GRU	0.029	0.138	0.170	0.933
PCA-GRU	0.019	0.121	0.139	0.969
PCA–GA-GRU	0.012	0.067	0.110	0.978
PCA–PSO-GRU	0.013	0.070	0.113	0.977
PCA–PFA-GRU	0.010	0.060	0.100	0.983

Furthermore, The PCA-GA-GRU and PCA-PSO-GRU models were compared to demonstrate this model superiority further. Fig. 7 illustrates a comparison chart of the model prediction effect. The findings demonstrate that PFA outperforms PSO and GA in terms of optimization. Twenty randomly selected samples of predicted data are enlarged and shown for comparison to demonstrate the predictive accuracy of this model more simply and intuitively.

According to the statistical findings of the model performance assessment in Table 4, the evaluation indicators based on the PCA-PFA-GRU combined prediction model proposed in this work outperform other models in predicting the dissolved oxygen content of sea bass culture water quality. It has a mean square error (MSE), a mean absolute error (MAE), a root mean square error (RMSE), and R^2 of 0.010, 0.060, 0.100, and 0.983, respectively. For a single model, GRU has a root mean square error of 0.170, which is 61.19 %, 19.05 %, and 50.87 % less than SVR, BPNN, RF, and LSTM, respectively. However, the predictions of the GRU and LSTM models are almost identical to the actual curve. Nevertheless, GRU outperforms LSTM in time convergence and central processing unit parameter optimization. It is more efficient at predicting dissolved oxygen in perch farming water quality parameters, indicating that GRU can better mine and fit perch farming water quality

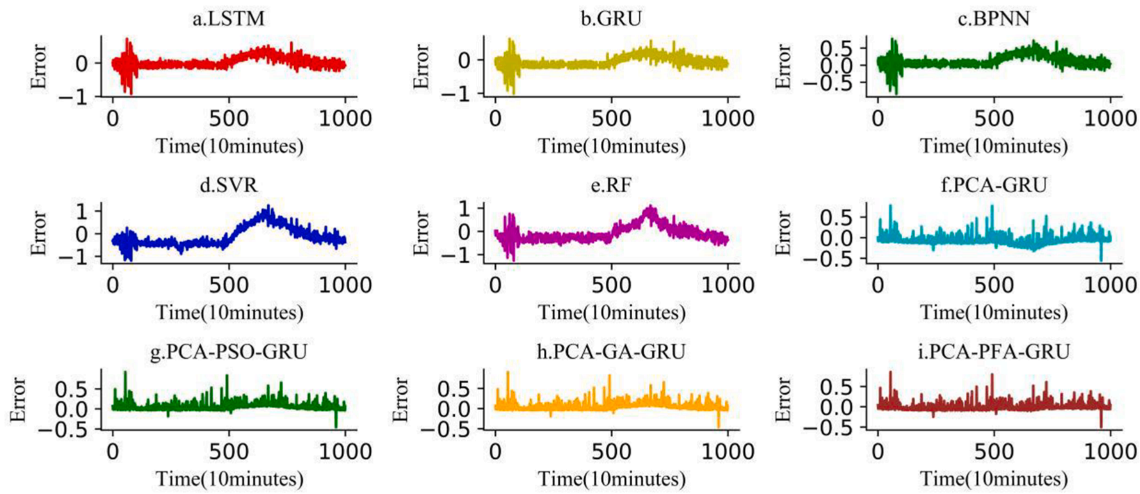

Fig. 8. Prediction error analysis of different models.

parameters. When PCA-GRU and GRU models are compared, the MSE, MAE, and RMSE of PCA-GRU are decreased by 34.48 %, 12.32 %, and 18.24 %, respectively, while R2increases by 3.86 %. This is because the original data contains some useless or redundant data, which complicates our computations and affects the model's efficiency. PCA may be used to identify the critical influencing factors that impact dissolved oxygen concentration, reduce the dimensionality of sea bass breeding environmental data, and increase prediction accuracy to a certain degree. Compared to the unoptimized model, it is obvious that the optimized GRU model performs substantially better, has a minimal error between the predicted and real values, and has an optimum prediction effect. Among them, PCA-PFA-GRU performs better than PCA-For GRU, PCA-GA-GRU, and PCA-PSO-GRU in terms of mean square error reduction of 47.36 %, 16.67 %, and 23.08 %, respectively, and root mean square error reduction of 28.06 %, 9.09 %, and 11.50 %, indicating that PFA can achieve GRU ultra-precision. The ideal parameter combination results in superior prediction accuracy and generalization performance for this method. Simultaneously, Fig. 8 illustrates the prediction errors associated with various models. The figure shows that the PCA-PFA-GRU model suggested in this study has the least error curve variations, typically hanging around 0.

The above results demonstrate that the PCA-PFA-GRU combined model proposed in this paper has a higher prediction accuracy. The proposed model can accurately capture the change information of dissolved oxygen content, meet actual needs and provide critical decision-making support for predicting and early warning of sea bass culture water quality.

5. Conclusion

The Internet of Things is used in this article to remotely access the real-time acquisition of parameters for sea bass water quality influencing elements. A combination prediction model based on PCA-PFA-GRU is presented and validated for non-linearity and non-stationarity of dissolved oxygen. Further conclusions are as follows:

(1) Using the principal component analysis algorithm reduces the dimension of the data, reduces training time, and eliminates irrelevance and redundancy between variables while preserving as much data as possible.
(2) PFA is simple to implement, simple to understand, capable of converge to the global optimal solution, and has a strong advantage in solving challenging and realistic problems with unknown search space.
(3) The GRU time series model is comparable to the LSTM in terms of prediction accuracy, but it has fewer GRU parameters and a quicker convergence rate, which may eliminate the lag associated with conventional models and enhance prediction accuracy;

The PCA-PFA-GRU-based combined prediction model proposed in this paper has a high prediction accuracy and a high capacity for generalization. It addresses the shortcomings of traditional methods regarding prediction accuracy and robustness. It enables powerful decision-making for water quality monitoring and regulation. According to this, it plays a significant role in regulating the growth circumstances of sea bass.

CRediT authorship contribution statement

Jianjun Guo: Methodology, Writing – original draft. **Jiaqi Dong:** Methodology, Writing – original draft. **Bing Zhou:** Data curation, Investigation. **Xuehua Zhao:** Supervision, Validation. **Shuangyin Liu:** Supervision, Validation. **Qianyu Han:** Resources, Visualization, Software. **Huilin Wu:** Resources, Visualization, Software. **Longqin Xu:** Writing – review & editing. **Shahbaz Gul Hassan:** Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grant 61871475, 52175240, 51775358, E0506, in part by the special project of laboratory construction of Guangzhou Innovation Platform Construction Plan under Grant 201905010006, Guangzhou key research and development project under Grant 202103000033, 201903010043, National Key Technologies R & D Program of China under Grant 2016YFD0500510,Guangdong Science and Technology Project under Grant 2020A1414050060, 2020B0202080002, 2017B010126001, Innovation Team Project of Universities in Guangdong Province under Grant 2021KCXTD019, Rural Science and technology correspondent project of Zengcheng District, Guangzhou city under Grant 2021B42121631.

References

Abdel-Tawwab, M., Monier, M.N., Hoseinifar, S.H., Faggo, C., 2019. Fish response to hypoxia stress: growth, physiological, and immunological biomarkers. Fish Physiol. Biochem. 45 (3), 997–1013.

Afrin, T., Yodo, N., 2021. A Long Short-Term Memory-based correlated traffic data prediction framework. In: Knowl.-Based Syst., p. 107755.

Banik, S., Sharma, N., Mangla, M., Mohanty, S.N., Shitharth, S., 2021. LSTM based decision support system for swing trading in stock market. Knowl.-Based Syst. 107994.

Cao, W., Huan, J., Liu, C., Qin, Y., Wu, F., 2019. A combined model of dissolved oxygen prediction in the pond based on multiple-factor analysis and multi-scale feature extraction. Aquacultural Eng. 84, 50–59.

Chaudhary, V., Deshbhratar, A., Kumar, V., Paul, D., 2018. Time series based LSTM model to predict air pollutant's concentration for prominent cities in India. UDM.

Chung, J., Gulcehre, C., Cho, K.H., Bengio, Y., 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555.

Dey, R., Salem, F.M., 2017. Gate-variants of gated recurrent unit (GRU) neural networks. In: 2017 IEEE 60th international idwest symposium on circuits and systems (MWSCAS). IEEE, pp. 1597–1600.

Fang, Z., Crimier, N., Scanu, L., Midelet, A., Alyafi, A., Delinchant, B., 2021. Multi-zone indoor temperature prediction with LSTM-based sequence to sequence model☆. Energy Build. 245, 111053.

Fang, J., Qin, J., Ren, M., Zhang, F., Jing, Y., Li, H., Zhai, W., Wang, Y., Tian, C., 2016. Experiment on large-scale seedling cultivation and healthy pond culture of Seven Star bass. Aquaculture 37 (11), 9–11 in Chinese with English abstract.

Hu, F., Zhong, H., Wu, C., Wang, S., Guo, Z., Tao, M., Zhang, C., Gong, D., Gao, X., Tang, C., Wei, Z., Wen, M., Liu, S., 2021. Development of fisheries in China. Reprod. Breeding 1 (1), 64–79.

Huan, J., Li, H., Li, M., Chen, B.o., 2020. Prediction of dissolved oxygen in aquaculture based on gradient boosting decision tree and long short-term memory network: A study of Chang Zhou fishery demonstration base, China. Comput. Electron. Agric. 175, 105530.

Huang, P., Wen, C., Fu, L., Peng, Q., Tang, Y., 2020. A deep learning approach for multi-attribute data: A study of train delay prediction in railway systems. Inform. Sci. 516, 234–253.

Kisi, O., Alizamir, M., Docheshmeh Gorgij, AliReza, 2020. Dissolved oxygen prediction using a new ensemble method. Environ. Sci. Pollut. Res. 27 (9), 9589–9603.

Lee, L.C., Jemain, A.A., 2021. On overview of PCA application strategy in processing high dimensionality forensic data. Microchem. J. 169, 106608.

Li, C., Li, Z., Wu, J., Zhu, L., Yue, J., 2018. A hybrid model for dissolved oxygen prediction in aquaculture based on multi-scale features. Inform. Process. Agric. 5 (1), 11–20.

Li, Z., Peng, F., Niu, B., Li, G., Wu, J., Miao, Z., 2018. Water quality prediction model combining sparse auto-encoder and LSTM network. IFAC-PapersOnLine 51 (17), 831–836.

Li, W., Wu, H., Zhu, N., Jiang, Y., Tan, J., Guo, Y., 2021. Prediction of dissolved oxygen in a fishery pond based on gated recurrent unit (GRU). Inform. Process. Agric. 8 (1), 185–193.

Liu, S., Xu, L., Li, D., Li, Q., Jiang, Y., Tai, H., Zeng, L., 2013. Prediction of dissolved oxygen content in river crab culture based on least squares support vector regression optimized by improved particle swarm optimization. Comput. Electron. Agric. 95, 82–91.

Rahman, A., Dabrowski, J., McCulloch, J., 2020. Dissolved oxygen prediction in prawn ponds from a group of one step predictors. Inform. Process. Agric. 7 (2), 307–317.

Ravansalar, M., Rajaee, T., Ergil, M., 2016. Prediction of dissolved oxygen in River Calder by noise elimination time series using wavelet transform. J. Exp. Theor. Artif. Intell. 28 (4), 689–706.

Sak, H., Senior, A., Beaufays, F., 2014. Long short-term memory recurrent neural network architectures for large scale acoustic modeling.

Shi, P., Li, G., Yuan, Y., Huang, G., Kuang, L., 2019. Prediction of dissolved oxygen content in aquaculture using Clustering-based Softplus Extreme Learning Machine. Comput. Electron. Agric. 157, 329–338.

Shi, X., Li, Y., Yang, Y., Sun, B., Qi, F., 2021. Multi-models and dual-sampling periods quality prediction with time-dimensional K-means and state transition-LSTM network. Inf. Sci. 580, 917–933.

Sun, L., Wu, Y., Li, D., Wang, B., Sun, X., Luo, B., 2021. Prediction model for the number of crucian carp hypoxia based on the fusion of fish behavior and water environment factors. Comput. Electron. Agric. 189, 106386.

Ta, X., Wei, Y., 2018. Research on a dissolved oxygen prediction method for recirculating aquaculture systems based on a convolution neural network. Comput. Electron. Agric. 145, 302–310.

Wu, J., Li, Z., Zhu, L., Li, G., Niu, B., Peng, F., 2018. Optimized BP neural network for dissolved oxygen prediction. IFAC-PapersOnLine 51 (17), 596–601.

Yapici, H., Cetinkaya, N., 2019. A new meta-heuristic optimizer: pathfinder algorithm. Appl. Soft Comput. 78, 545–568.

Yin, L., Fu, L., Wu, H., Xia, Q., Jiang, Y., Tan, J., Guo, Y.a., 2021. Modeling dissolved oxygen in a crab pond. Ecol. Model. 440, 109385.

Zhang, Z., Fang, X., Sun, L., 2021. Comparison for ecological economic performance of Chinese sea perch (Lateolabrax Maculatus) under different aquaculture systems. Aquac. Fisheries.

Zhang, Y., Fitch, P., Vilas, M.P., Thorburn, P.J., 2019. Applying multi-layer artificial neural network and mutual information to the prediction of trends in dissolved oxygen. Front. Environ. Sci. 7, 46.

Zhang, L., Liu, P., Zhao, L., Wang, G., Zhang, W., Liu, J., 2021. Air quality predictions with a semi-supervised bidirectional LSTM neural network. Atmos. Pollut. Res. 12 (1), 328–339.