Hybrid dual decomposition-machine learning model for forecasting dissolved oxygen in aquaculture

Chong Chen, Chaoyang Liu, Simin Peng*, Yuanliang Wang, Kaiqiang Wang, Zhinan Yin
(*School of Electrical Engineering, Yancheng Institute of Technology, Yancheng 224051, Jiangsu, China*)

Abstract: In aquaculture, dissolved oxygen levels are a primary determinant of water quality; leveraging accurate forecasts for this parameter directly contributes to improved yields and higher-quality harvests. Given that current dissolved oxygen prediction models utilize just one decomposition process, they are unable to sufficiently mitigate residual noise interference or properly handle non-stationary features. To enhance prediction accuracy, our model begins by processing the original dissolved oxygen sequence via an improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN), yielding multiple intrinsic mode functions (IMFs). These IMFs are then clustered according to their fuzzy entropy using the K-means algorithm. The component with the maximum fuzzy entropy value is selected for further refinement through a Sequential Variational Mode Decomposition (SVMD). This dual decomposition approach effectively preserves valid information while enhancing detailed feature extraction. For the forecasting stage, an Extreme Learning Machine (ELM) whose parameters are optimized by the rime optimization algorithm is developed to generate forecasts for each sub-component. Ultimately, the predictions of all components are aggregated through a Back-Propagation (BP) neural network to produce the final dissolved oxygen time series. The proposed model is applied to verify dissolved oxygen in the standard pond of Huludao sea cucumber aquaculture base, Liaoning province, China. Evaluation results show that the proposed model exceeds the performance of existing comparative models. With the RMSE of 0.2443, MAE of 0.1782, MAPE of 0.0352, and R^2 of 0.9835, the model exhibits exceptional predictive accuracy and operational stability, validating that it is well-suited to meet the practical demands of contemporary aquaculture.

Keywords: dissolved oxygen, secondary decomposition, fuzzy entropy, rime, extreme learning machine
DOI: 10.25165/j.ijabe.20261901.10299

Citation: Chen C, Liu C Y, Peng S M, Wang Y L, Wang K Q, Yin Z N. Hybrid dual decomposition-machine learning model for forecasting dissolved oxygen in aquaculture. Int J Agric & Biol Eng, 2026; 19(1): 33–46.

1 Introduction

In aquaculture systems, maintaining appropriate dissolved oxygen concentration is of paramount importance, as it is a fundamental parameter for ensuring high production yields and superior product quality. For benthic organisms such as sea cucumbers, adequate oxygen is crucial during their metabolic activities[1]. When the dissolved oxygen concentration ranges from 3 to 4 mg/L, the condition is classified as mild hypoxia, resulting in a 20%-30% reduction in feeding activity and a decline in growth rate in sea cucumbers. At dissolved oxygen concentrations of 2-3 mg/L, conditions are classified as moderate hypoxia, leading to suppressed immune function and a 10%-15% rise in mortality rates among juvenile sea cucumbers. When dissolved oxygen levels fall below 2 mg/L, severe hypoxia occurs, resulting in mortality rates reaching 50% within 48 h. Similarly, excessively high dissolved oxygen concentrations can lead to eutrophication and adversely affect sea cucumber farming. Therefore, timely monitoring and precise prediction of dissolved oxygen concentrations are economically and practically significant for sea cucumber farming, as they help prevent substantial economic losses for farmers[2].

In recent years, significant advances have been made in dissolved oxygen prediction for aquaculture systems by multiple research groups. Five major challenges hinder precise dissolved oxygen prediction in aquaculture systems. The primary challenge involves multivariable nonlinear coupling, where dissolved oxygen dynamics are strongly interdependent with environmental parameters including temperature, light intensity, and biological activity. For instance, an increase in temperature may reduce dissolved oxygen due to decreased solubility, but it can also increase due to algal photosynthesis. The second challenge involves spatiotemporal heterogeneity, temporally characterized by abrupt meteorological variations and seasonal patterns that demand simultaneous analysis of both short-term fluctuations and long-term trends. Spatially, dissolved oxygen concentrations can vary significantly with pond depth, creating vertical stratification that single-point sensors fail to adequately capture, thus limiting their ability to represent overall water column conditions. The third challenge stems from inherent data quality issues, where dissolved oxygen time series exhibit substantial noise due to sensor drift, environmental interference, and random transmission errors. The fourth challenge involves nonlinear dynamics in dissolved oxygen time series, primarily caused by temporal lag effects commonly observed in photosynthetic and temperature responses. The fifth challenge concerns the prediction of critical threshold breaches, where aquaculture ponds may undergo abrupt dissolved oxygen depletion when concentrations fall below biologically significant levels, typically manifesting as step-function declines below 2 mg/L.

Received date: 2025-10-25 **Accepted date:** 2026-01-06

Biographies: Chong Chen, Professor, research interest: smart agriculture, Email: cc082120@126.com; **Chaoyang Liu**, MS candidate, research interest: aquaculture, Email: lcydw97@163.com; **Yuanliang Wang**, PhD, research interest: machine learning, Email: ylwang_yit@163.com; **Kaiqiang Wang**, MS candidate, research interest: artificial intelligence, Email: 2511238484@qq.com; **Zhinan Yin**, Undergraduate, research interest: smart agriculture, Email: 1691311918@qq.com.

***Corresponding author: Simin Peng**, Professor, research interest: artificial intelligence. Room 703, No.1 Xiwang Middle Avenue, Yancheng, Jiangsu, China, Tel: +86-515-88168194, Email: siminpeng@ycit.edu.cn.

In response to these challenges, existing approaches for dissolved oxygen prediction can be broadly categorized into four groups: mechanism-based models rooted in physical and chemical principles, traditional statistical models relying on data trends, modern machine learning models that extract complex features, and hybrid intelligent models that leverage multimodal data fusion. Regarding mechanism models, Kimberly et al. effectively evaluated the self-purification capacity of the last section of the Culiacán River using the Streeter-Phelps dissolved oxygen model[3]. Although mechanism models have strong explanatory power, they have limitations such as reliance on precise parameters, high computational complexity, and ignoring biological randomness. In terms of statistical models, methods such as time series, multiple regression, and grey prediction can be used, but these methods cannot handle the non-stationary characteristics of dissolved oxygen. In relation to machine learning models, Li et al.[4] introduced advanced machine learning algorithms into the monitoring and prediction analysis of dissolved oxygen levels, providing effective evidence for environmental management in the Yangtze River estuary. Finally, there have been numerous advancements in hybrid intelligent models. For example, Liu et al.[5,6] used the ant colony optimization algorithm to optimize the least squares support vector machine prediction model and proposed the WA-CPSO-LSSVR hybrid prediction model, achieving excellent results in the prediction of dissolved oxygen in crab ponds. Huan et al.[7] proposed a hybrid dissolved oxygen prediction model based on K-means clustering and extreme learning machines. Shi et al.[8] used a regularized extreme learning machine with a clustering mechanism to predict dissolved oxygen in black bass aquaculture ponds. Li et al.[9] enhanced the prediction of dissolved oxygen concentration via a hybrid deep learning method based on sinusoidal geometric mode decomposition, which played an important role in sustainable water management. He et al.[10] proposed a PSG-Crossformer model, which provided more accurate predictions for water quality management. Nowadays, hybrid intelligent models have become the optimal solution for predicting dissolved oxygen in real-world complex water bodies due to their significant advantages in accuracy, robustness, and the ability to handle multi-source heterogeneous data.

Most existing models follow a conventional decomposition-reconstruction-prediction framework, while some merely establish simple optimization-based prediction models. These approaches are generally inadequate for fully extracting and utilizing the information derived from the decomposition process. Dissolved oxygen time series have both nonlinear and non-stationary characteristics. While time series decomposition is effective for revealing the intrinsic characteristics and structure of data, a fundamental constraint of employing a single decomposition technique is its inability to ensure the comprehensive exploitation of information within each time-frequency component[11], which inevitably affects the prediction performance of the model. In addition, although modern intelligent algorithms can achieve good prediction results, they also have the drawbacks of getting trapped in local optima, over-fitting or under-fitting, and the need for manual parameter tuning.

To address these issues, this paper presents a dual-decomposition enhanced extreme learning machine optimized by the RIME algorithm for aquaculture dissolved oxygen prediction. The proposed framework innovatively integrates: primary signal decomposition, modal component clustering, secondary variational decomposition, metaheuristic parameter optimization, neural network forecasting, and nonlinear integration via a neural network, thereby constituting a new hybrid modeling framework. The main contributions of this paper can be summarized as follows:

(1) The analysis of four diagnostic metrics—autocorrelation function, partial autocorrelation function, maximum Lyapunov exponent, and saturation correlation dimension—converges to indicate that the dissolved oxygen time series is characterized by significant nonlinearity, non-stationarity, chaotic dynamics, and inherent predictability.

(2) A novel dissolved oxygen time series preprocessing method is proposed, which first employs fuzzy entropy, sensitive to high-frequency fluctuations, to perform K-means clustering on the IMFs obtained from improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN) decomposition. Then, the clustered high-frequency sequences undergo secondary decomposition using SVMD. This method effectively captures high-frequency components within all IMFs obtained in the first decomposition, and reduces noise, non-stationarity, and complexity in the original dissolved oxygen time series through secondary decomposition.

(3) A hybrid model, named ICEEMDAN-KMEANS-SVMD-RIME-ELM-BP, is proposed. The RIME algorithm automatically optimizes ELM's hidden layer configuration, eliminating manual tuning randomness while determining optimal node numbers. Through ablation, comparison, and generalization experiments, this model significantly enhances prediction accuracy by reducing noise interference and non-stationary characteristics in the dissolved oxygen time series.

This paper is structured into four sections. The construction of the hybrid model, which incorporates Improved CEEMDAN, K-means clustering, successive VMD, the RIME algorithm[12], an Extreme Learning Machine, and a BP neural network, is described in Section 2. Section 3 provides details on the real-world data used, its characteristic analysis, and a comparative assessment of the proposed model's predictive performance. Conclusions and potential future work are outlined in Section 4.

2 Materials and methods

2.1 Data collection

The water quality data of this study is derived from a sea cucumber farm in Muhou Village, Xingcheng City, Liaoning Province, which is equipped with an Internet of Things (IoT) perception and monitoring system. The farm covers an area of approximately 60 hm^2 and is a coastal earthen pond. The water depth is about 1.7-1.8 m in summer and 1.5 m in winter. The stocking density is 75 heads/m^2. Experimental data were acquired using a wireless monitoring system, with its structure detailed in Figure 1. The system is equipped with dissolved oxygen sensors at each water quality monitoring node. The monitoring center transmits the collected data to the server through the control center, allowing users to monitor in real time via mobile phones or computers. The water quality regulation nodes are equipped with aeration pumps and other devices, enabling precise water quality control. The dataset used in this study was collected from the monitoring device (ID: 2066000001, “Nanhe Bridge No. 1”). The data were collected hourly from August 13 to August 27, 2024, spanning 15 consecutive days and accumulating 360 complete datasets. To ensure the fairness of the model evaluation, a fixed split strategy was adopted: for each component, the last 72 samples (accounting for 20% of the total data) were extracted from the end of the time series to form an independent test set, with the remaining data serving as the training set. The test set was kept completely unseen throughout the entire model development process and was only used for the final performance assessment.

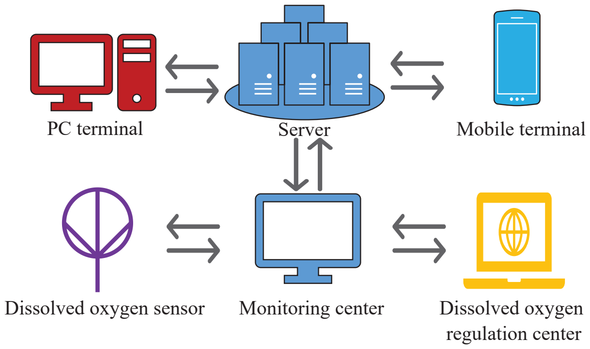


Figure 1　Structure of remote monitoring system

2.2 Improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN)

ICEEMDAN[13], an enhanced signal decomposition algorithm, improves upon CEEMDAN's capability to suppress mode mixing while additionally addressing pseudo-components and residual noise. This extension broadens its applicability to diverse nonlinear and non-stationary signals, achieving superior decomposition accuracy and quality. The decomposition steps are as follows:

Step 1: Add white noise[14,15] to the original signal $Y_0(k)$ to generate multiple perturbed sequences:

$$Y_i(k) = Y_0(k) + W_i(k), (i = 1,2,\ldots,E) \tag{1}$$

where $Y_i(k)$ denotes the i-th perturbed sequence, $W_i(k)$ represents the added white noise, and E denotes the total number of white noise instances.

Step 2: Perform EMD decomposition on each perturbation sequence $Y_i(k)$, and obtain multiple intrinsic mode functions:

$$IMF_{ij}, (j = 1,2,\ldots,n) \tag{2}$$

where $IMF_{ij}(k)$ denotes the j-th intrinsic mode function of the i-th perturbation sequence, and n represents the number of modes.

Step 3: For each mode j, compute the ensemble average of all $IMF_{ij}(k)$:

$$C_j(k) = \frac{1}{E}\sum_{i=1}^{E} IMF_{ij}(k) \tag{3}$$

Step 4: Estimate and remove the local mean of $C_j(k)$ to obtain the corrected mode function:

$$IMF_j^*(k) = C_j(k) - M_j(k) \tag{4}$$

where, $M_j(k)$ denotes the local mean.

2.3 Successive variational mode decomposition (SVMD)

SVMD, an improved variational signal decomposition method[16], was developed to overcome mode mixing and parameter sensitivity issues inherent in traditional variational mode decomposition (VMD) when processing complex signals. Letting $s(t)$ be an original signal, the SVMD signal decomposition steps are as follows:

Step 1: Initialize parameters by setting the maximum number of modes $K_{\max}$, convergence threshold ϵ, initial residual $\gamma_0(t) = s(t)$, and mode index $k = 1$.

Step 2: Perform successive mode extraction by constructing a single-mode variational problem for each $k = 1,2,\ldots,K_{\max}$, where the objective function is minimized to extract the k-th mode $u_k(t)$:

$$\min_{u_k w_k}\left\{\left\|\partial_t\left[\left(\delta(t)+\frac{j}{\pi t}\right)*u_k(t)\right]e^{-\mathrm{j}w_k t}\right\|_2^2 + \lambda\|\,u_k(t)\,\|_1\right\} \tag{5}$$

where w_k denotes the center frequency of mode $u_k(t)$, λ represents the sparsity regularization coefficient, and in the addition operation, the first term represents the mode's bandwidth penalty, while the second term enforces sparsity constraints. The constraint condition is given by $u_k(t) + r_k(t) = r_{k-1}(t)$, where $r_k(t)$ denotes the residual after the k-th decomposition.

Step 3: Perform alternating optimization to iteratively update $u_k(t)$ and w_k using the following equations respectively:

$$u_k^{(n+1)}(t) = \arg\min_{u_k}\left\{\left\|\partial_t\left[\left(\delta(t)+\frac{j}{\pi t}\right)\cdot u_k(t)\right]e^{-\mathrm{j}w_k^{(n)}t}\right\|_2^2 + \lambda\|u_k(t)\|_1\right\} \tag{6}$$

$$w_k^{(n+1)} = \frac{\int_0^{\infty} w\left|\hat{u}_k^{(n+1)}(w)\right|^2 dw}{\int_0^{\infty}\left|\hat{u}_k^{(n+1)}(w)\right|^2 dw}\,(7) \tag{7}$$

where, $\hat{u}_k(w)$ denotes the Fourier transform of $u_k(t)$.

Step 4: Update the residual signal:

$$r_k(t) = r_{k-1}(t) - u_k(t) \tag{8}$$

Step 5: Begin convergence checking. If $\|r_k(t)\|_2 < \epsilon$ or $k = K_{\max}$, terminate the decomposition; otherwise, increment k by 1 and return to Step 2.

2.4 Rime optimization algorithm

The core concept of the RIME optimization algorithm is to simulate the motion trajectories of soft and hard rime particles, then optimize algorithmic performance based on their dynamic behaviors. During the simulation of rime ice growth, the growth and movement behavior of soft rime offers significant advantages in exploring the solution space of optimization problems, enhancing the robustness and adaptability when solving complex problems. Unlike conventional optimization algorithms that often stagnate in local optima, the RIME algorithm demonstrates superior global search capability, enabling rapid convergence to the true global optimum. The flowchart of the RIME optimization algorithm is shown in Figure 2.

The core mechanisms of the RIME optimization algorithm consist of the following four aspects:

1) Initialization of the rime ice crystal population

Step 1: Initialize the rime ice crystal population R. The population R consists of n rime agents P_i, where each agent P_i is composed of d rime particles x_{ij} (see in Equation 9). The population R is represented by the rime particles x_{ij}, whose initial positions are randomly generated within the search space (seen in Equation 10). Subsequently, reset the RIME algorithm's parameters, including the search space range, convergence criteria, and search intensity, to ensure smooth execution of the search process. Step 2: Calculate the fitness value for each initial position to evaluate its quality within the solution space.

$$R = \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_i \end{bmatrix};\quad P_i = \left[x_{i1}, x_{i2}, \ldots, x_{ij}\right] \tag{9}$$

$$R = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} \\ x_{21} & x_{22} & \cdots & x_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} \end{bmatrix} \tag{10}$$

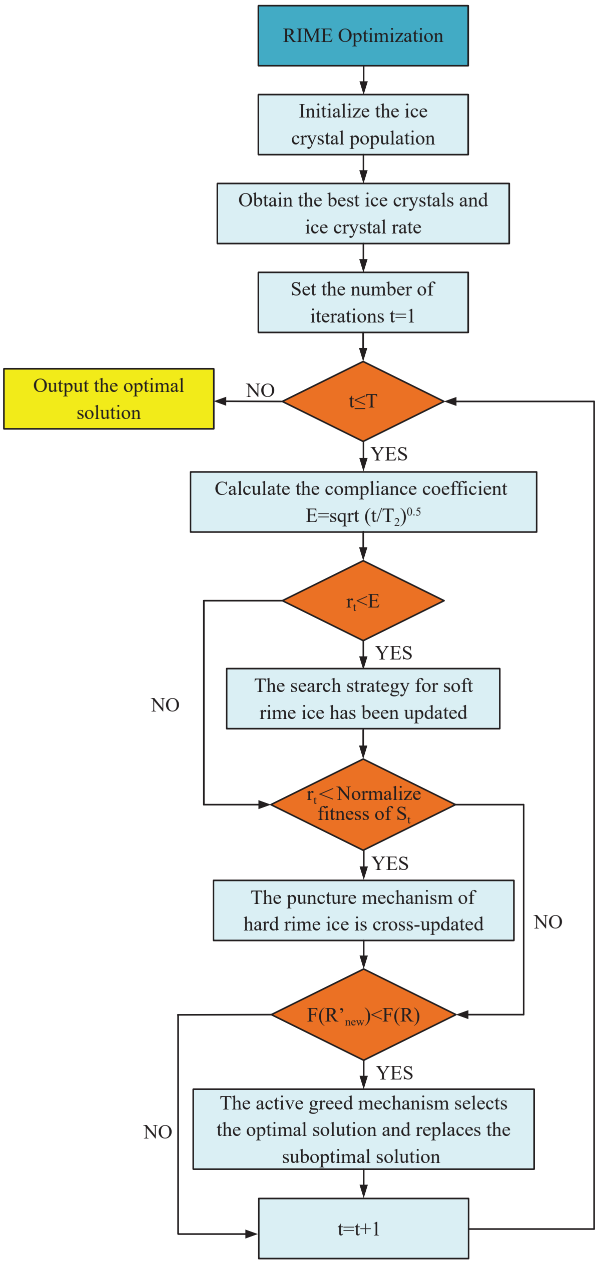


Figure 2 Flowchart of the Rime optimization algorithm

2) Soft rime search strategy

In this phase, rime particle positions in the solution space are calculated by simulating the rime formation process using Equation (11). During this stage, current particle positions are evaluated, subsequent positions are predicted, and the movement directions and step sizes of rime particles in the search space are determined, accounting for environmental conditions and interparticle interactions.

$$R_{ij}^{\text{new}} = R_{\text{best},j} + r_1 \cdot \cos\theta \cdot \beta\left(h\left(Ub_{ij} - Lb_{ij}\right) + Lb_{ij}\right),\ \ r_2 < E \tag{11}$$

where, R_{ij}^{new} represents the updated position of the rime particle; i and j denote the j-th particle in the i-th rime agent; $R_{\text{best},j}$ represents the j-th particle of the best-performing rime agent in population R; r_1 is a random number within the range (−1, 1), which jointly with $\cos\theta$ determines the particle's movement direction; β denotes the environmental factor as defined in Equation (12); Ub_{ij} and Lb_{ij} represent the upper and lower bounds of the escape space, respectively; $h\in(0, 1)$ denotes the adhesion degree coefficient; E denotes the adhesion coefficient, which increases iteratively along with $\cos\theta$ according to Equation (13); r_2 is a random number in the interval (0, 1), which together with the adhesion coefficient E determines whether the particle's position is updated.

$$\beta = 1 - \frac{\left[\frac{w \cdot t}{T_2}\right]}{w} \tag{12}$$

where, the default value of w is 5, β models the step function, and $[\cdot]$ denotes the rounding operator.

$$\theta = \frac{\pi t}{10T_2} \tag{13}$$

where, θ is an angle adjustment factor for calculating particle movement direction, T_2 denotes the maximum iteration number, and t represents the current iteration index.

E is calculated in Equation (14).

$$E = \sqrt{\frac{t}{T_2}} \tag{14}$$

3) Hard rime penetration mechanism

The purpose of this phase is to accelerate algorithm convergence and avoid local optima entrapment by dynamically updating inter-agent interaction rules to enable efficient rime particle exchange. The particle replacement rule is formally defined in Equation (15) and Equation (16).

$$F^{\text{normr}} = \frac{f(S_i) - \min(f)}{\max(f) - \min(f)} \tag{15}$$

$$R_{ij}^{\text{new}} = R_{\text{best},j},\ \ r_3 < F^{\text{normr}}(S_i) \tag{16}$$

where, $F^{\text{normr}}(S_i)$ represents the selection probability of the i-th rime agent, $f(S_i)$ is the fitness value of the i-th individual, and r_3 denotes a random number uniformly distributed in the range (−1, 1).

4) Positive greedy selection mechanism

First update the rime ice crystals positions, then evaluate and compare the fitness values before and after the update. If the updated fitness value is better than the previous one, replace it with the new value to ensure the algorithm consistently progresses toward the optimal solution.

2.5 Extreme learning machine (ELM)

The ELM algorithm, proposed by Huang et al.[17] in 2006, is a classical predictive algorithm based on single-hidden-layer feedforward neural networks (SLFNs) that offers strong nonlinear processing capabilities and significantly faster training speeds. The ELM architecture consists of three layers: an input layer that receives external data, a hidden layer that extracts features and performs nonlinear transformations, and an output layer that generates predictions or classifications based on the hidden layer's output. Unlike traditional neural network training algorithms, ELM differs in two key aspects: first, it does not require iterative adjustment of hidden-layer neuron thresholds, and second, it maintains fixed connection weights between the input and hidden layers throughout training. The only similarity is that both methods require setting the number of hidden-layer neurons during training.

In the ELM algorithm, the selection of activation functions critically influences multiple performance metrics and plays a pivotal role in model effectiveness. Considering four major performance indicators—training speed, model expressiveness, generalization capability, and model stability—the Sigmoid function has become the most widely used activation function[18]. The mathematical expression of the Sigmoid function is given in Equation (17). Its graph exhibits an "S"-shaped curve with an output range of (0,1). As the input x approaches negative infinity, the function value asymptotically approaches 0; at x=0, the function value equals 0.5; and as x approaches positive infinity, the function value converges to 1. The Sigmoid function is differentiable everywhere, and its derivative $S'(x)=S(x)(1-S(x))$ is a key property that significantly contributes to the computational efficiency of the ELM algorithm.

$$S(x)=\frac{1}{1+e^{-x}} \tag{17}$$

The training process of ELM using the Sigmoid activation function is as follows:

1) Random initialization

The process aligns with conventional ELM algorithms: During the initial training phase, random connection weights w_{ij} between the input layer (node i) and hidden layer (node j), along with hidden layer neuron thresholds b_j, are generated. These random parameters typically follow either a uniform or normal distribution.

2) Hidden layer output calculation

For a given input sample $x_i(i=1,2,\ldots,N)$, where N represents the number of samples, the element h_{ij} of the hidden layer output matrix H (at the i-th row and j-th column) is calculated using the Sigmoid activation function as defined in Equation (18).

$$h_{ij}=\frac{1}{1+e^{-\left(\sum_{k=1}^{n} w_{kj}x_{ik}+b_j\right)}} \tag{18}$$

where, w_{kj} represents the connection weight from the k-th input layer node to the j-th hidden layer node, b_j denotes the bias of the j-th hidden neuron, and n indicates the number of input layer nodes. The nonlinear characteristics of the Sigmoid function enable the hidden layer to effectively transform input features, mapping them into a new feature space.

3) Output layer weights calculation

After obtaining the hidden layer output matrix H, the output layer weights β are determined by solving the linear system $H\beta=T$ using the least squares method, where T represents the target output matrix. In practice, the Moore-Penrose pseudoinverse H^{-1} of H is typically employed to compute β.

2.6 Prediction framework of dissolved oxygen

A predictive framework for dissolved oxygen concentration, founded on the proposed model, is developed in this section and depicted schematically in Figure 3.

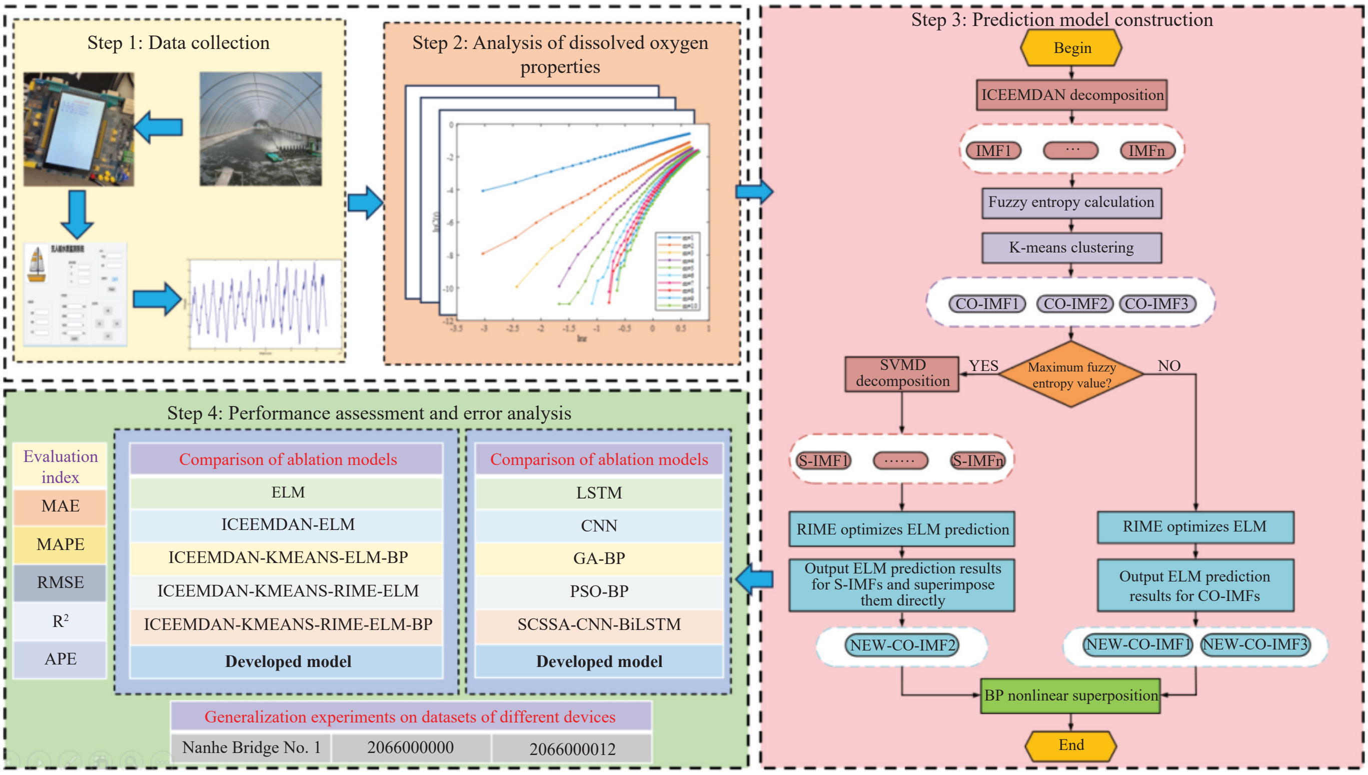


Figure 3 Framework of the proposed dissolved oxygen prediction methodology

Step 1: Data collection. Dissolved oxygen measurements were collected and transmitted to the host computer, yielding the raw time-series dataset.

Step 2: Analysis of dissolved oxygen properties. The acquired raw data were analyzed for key statistical properties to validate predictability.

Step 3: Prediction model construction. The specific building steps are as follows:

1) Read the dataset and set the sampling frequency, then generate the corresponding time-domain axis.

2) Decomposition of the original dissolved oxygen data was performed using the ICEEMDAN technique to obtain multiple intrinsic mode functions (IMFs).

3) Calculate the fuzzy entropy values for all IMFs, cluster IMFs into three groups via K-means, and then reconstruct three components by summing IMFs within each cluster.

4) The component with the maximum fuzzy entropy value after reconstruction is further decomposed into a finite set of S-IMFs using SVMD, thereby enhancing its predictability.

5) Construct a RIME-ELM prediction model using the S-IMFs derived from SVMD decomposition and the other two reconstructed components for forecasting, with data standardization applied to ensure model stability.

6) Output the ELM prediction results of the reconstructed components, and perform a linear superposition of the SVMD-decomposed prediction results for the final output.

7) Utilize a BP neural network to perform nonlinear superposition on all output prediction signals while conducting data re-standardization, and output the final prediction results.

Step 4: Performance assessment and error analysis. The performance of the proposed model was evaluated through error analysis of its predictions and comparisons with multiple benchmark models. Its superiority was validated through comprehensive ablation and comparison experiments, and its stability and generalization capability were further verified via generalization experiments.

3 Results and discussion

All experimental work in this study was performed on a computational server configured with an Intel Core i7-14700HX processor, NVIDIA GeForce RTX 4060 Laptop GPU, and 16 GB DDR4 RAM, utilizing MATLAB R2024b (MathWorks, Inc.) as the computational platform. With the objectives of evaluating the predictability of dissolved oxygen time series, assessing the operational feasibility of the proposed model, and demonstrating its superior performance, three experimental studies were conducted: (1) characteristic analysis of dissolved oxygen time series, (2) model operational process validation, and (3) ablation experiments, comparative performance experiments, and generalization experiments.

3.1 Characteristic analysis of dissolved oxygen time series

In order to empirically verify the predictability of the dissolved oxygen time series, an analysis was conducted based on four key autocorrelation metrics: the autocorrelation function, partial autocorrelation function, maximum Lyapunov exponent, and saturation correlation dimension[19]. The dissolved oxygen time series is shown in Figure 4.

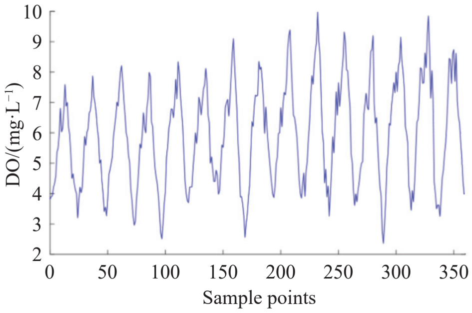
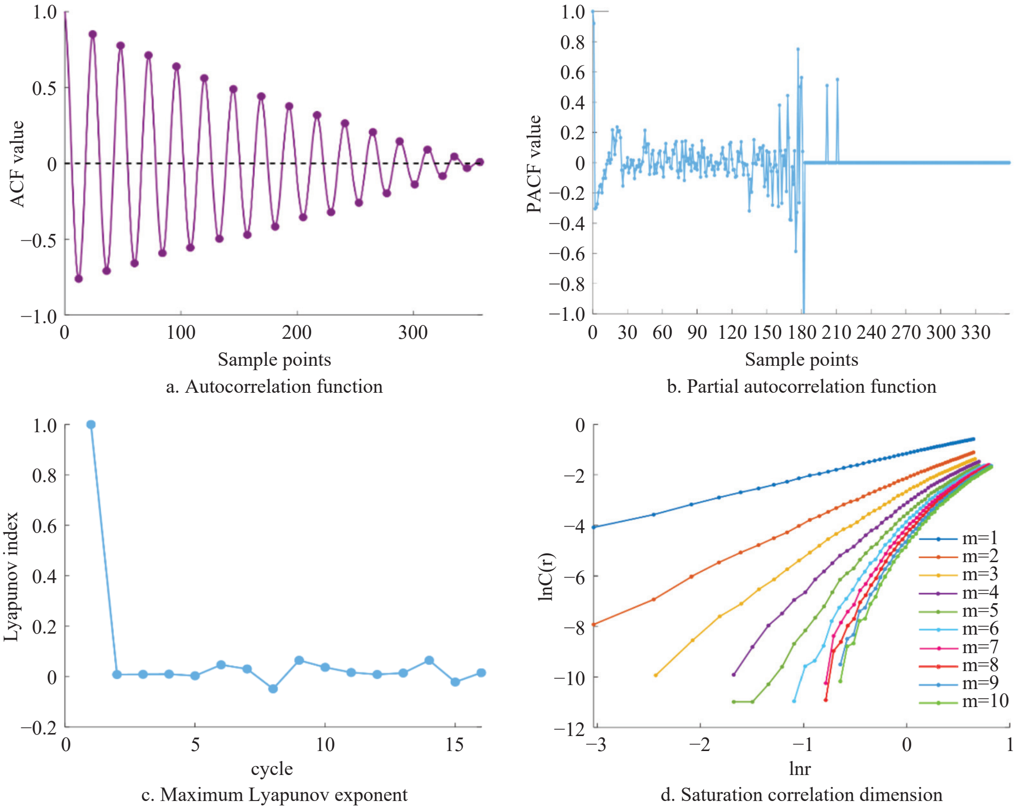


Figure 4 Dissolved oxygen time series plot

1) Nonlinearity: To determine whether a time series is linear, the autocorrelation function is typically employed[20]. This function describes the dependency relationship between a sequence at one time point and another. As shown in Figure 5a, it can be observed how strongly the initial moment correlates with any other moment in the sequence. With the maximum time lag set to 360, the function values turn negative starting from the 7th lag, indicating that the dissolved oxygen series exhibits pronounced nonlinear characteristics.



a. Autocorrelation function b. Partial autocorrelation function

c. Maximum Lyapunov exponent d. Saturation correlation dimension

Figure 5 Dissolved oxygen time series characteristics

2) Nonstationarity: To determine whether a time series is stationary, the partial autocorrelation function is typically employed[21]. As shown in Figure 5b, with the maximum time lag set to 359, if the origin is taken at time *t*, a single prominent peak appears at time *t*+1. The function values begin to converge after time *t*+212 within the confidence interval, indicating that the dissolved oxygen series exhibits nonstationary characteristics.

3) Chaotic characteristics: The maximum Lyapunov exponent is employed to determine whether a time series exhibits chaotic behavior[22]. A positive value indicates the presence of chaotic characteristics in the sequence. Using the Wolf method, this curve was computed over 15 cycles. As evident from Figure 5c, the

maximum Lyapunov exponents of the dissolved oxygen series are predominantly positive except for a few isolated values, demonstrating that the dissolved oxygen series possesses certain chaotic properties.

4) Stochastic characteristics: The saturation correlation dimension[23] serves as an indicator for distinguishing between stochastic and chaotic dynamics in a time series. A saturated correlation dimension with increasing embedding dimension signifies chaos, while unsaturation suggests stochasticity. Figure 5d confirms chaotic characteristics in the hourly dissolved oxygen series, as its correlation dimension clearly saturates when the embedding dimension ranges from 1 to 10.

Through the examination of the four aforementioned metrics, the dissolved oxygen time series demonstrates significant nonlinearity, nonstationarity, and chaotic characteristics. Consequently, this confirms that the time series possesses predictability, thereby enabling reliable forecasting of dissolved oxygen dynamics.

3.2 Operational process validation of the proposed model

3.2.1 ICEEMDAN primary decomposition

The dissolved oxygen time series was decomposed using the ICEEMDAN algorithm with the following parameter settings: a noise standard deviation of 10, 20 noise additions, a maximum of 100 iterations, Mode 2 signal-to-noise ratio processing, and linearly decreasing noise intensity with IMF order. The decomposition produced eight IMFs and residual components (Figure 6). As shown in the Figure, the dissolved oxygen series exhibits distinct nonlinear and nonstationary characteristics. The eight IMFs and residual components exhibit a clear hierarchical frequency structure (from high to low frequency), effectively capturing the multiscale fluctuation patterns of the original dissolved oxygen signal. This decomposition successfully separates the time series into distinct components representing different frequency bands and temporal characteristics.

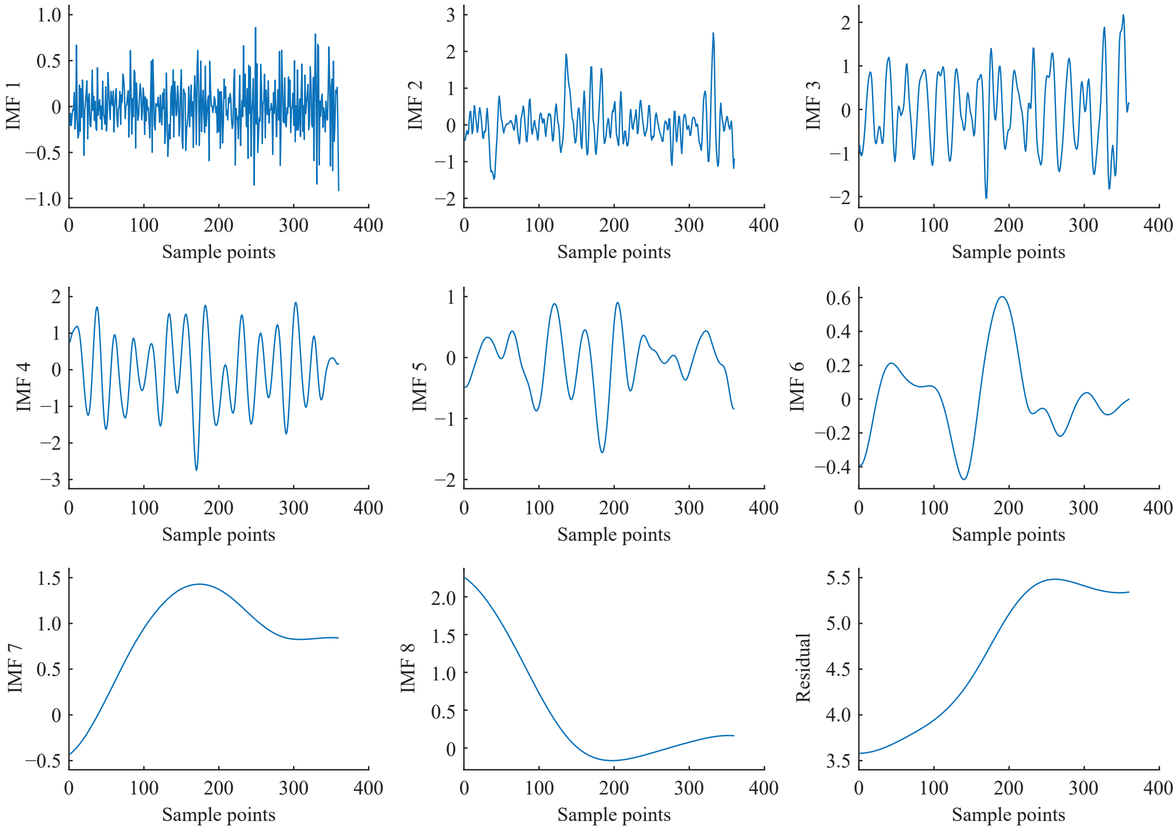


Figure 6　Primary decomposition results of ICEEMDAN

3.2.2 Fuzzy entropy calculation and K-means clustering

Fuzzy entropy is derived from the integration of fuzzy theory and entropy measurement, which evaluates the complexity and uncertainty of dissolved oxygen time series by measuring the frequency of similar patterns. The ICEEMDAN decomposition produces multiple IMF components with distinct complexity levels. The fuzzy entropy value increases with IMF component complexity, and decreases with signal simplicity. In this study, the fuzzy entropy parameters were set as follows: embedding dimension m=2, similarity threshold r=0.2, using the ‘fx’ entropy variant. The number of clusters in the K-means algorithm[7] was set to 3. The final computed fuzzy entropy values and corresponding clustering results are presented in Figure 7.

The reconstructed signals based on K-means clustering are shown in Figure 8. CO-IMF2 clearly exhibits higher frequency characteristics, confirming its classification as a high-frequency component.

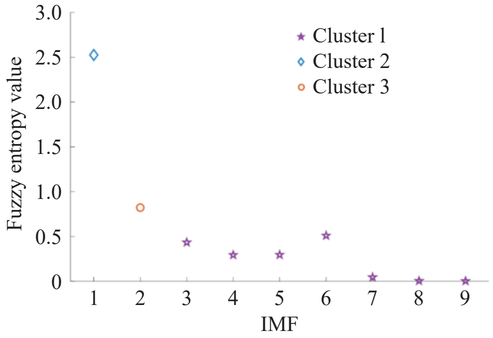


Figure 7　Clustering results of fuzzy entropy values

3.2.3 SVMD secondary decomposition

From the fuzzy entropy values of ICEEMDAN-decomposed IMFs in Figure 7, it is evident that IMF1 has a significantly higher fuzzy entropy value than other components, which typically contains the most complex signal features. Notably, during K-means clustering reconstruction, IMF1 autonomously forms a distinct CO-IMF2 component. This demonstrates that IMF1 (i.e., CO-IMF2) exhibits relatively high complexity, warranting further investigation of its key features. Therefore, this study employs SVMD to further decompose IMF1 (i.e., CO-IMF2).

The SVMD parameters were configured with a maximum number of modes of 2000, a noise tolerance set to 0, a convergence tolerance of 10^{-7}, and a stop criterion type of 4. With these parameter settings, the decomposition produces 19 components. The detailed results are presented in Figure 9.

3.2.4 RIME-ELM prediction

In this study, CO-IMF1, CO-IMF3, and all other SVMD-decomposed components are predicted using ELM. The predicted SVMD components are then linearly superimposed to reconstruct the enhanced NEWCO-IMF2 signal. The ELM prediction curve results are shown in Figure 10.

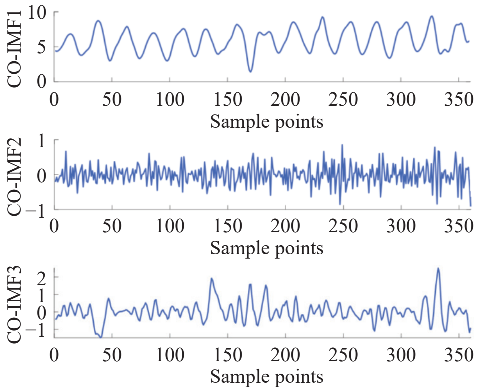


Figure 8 K-means clustering reconstruction results

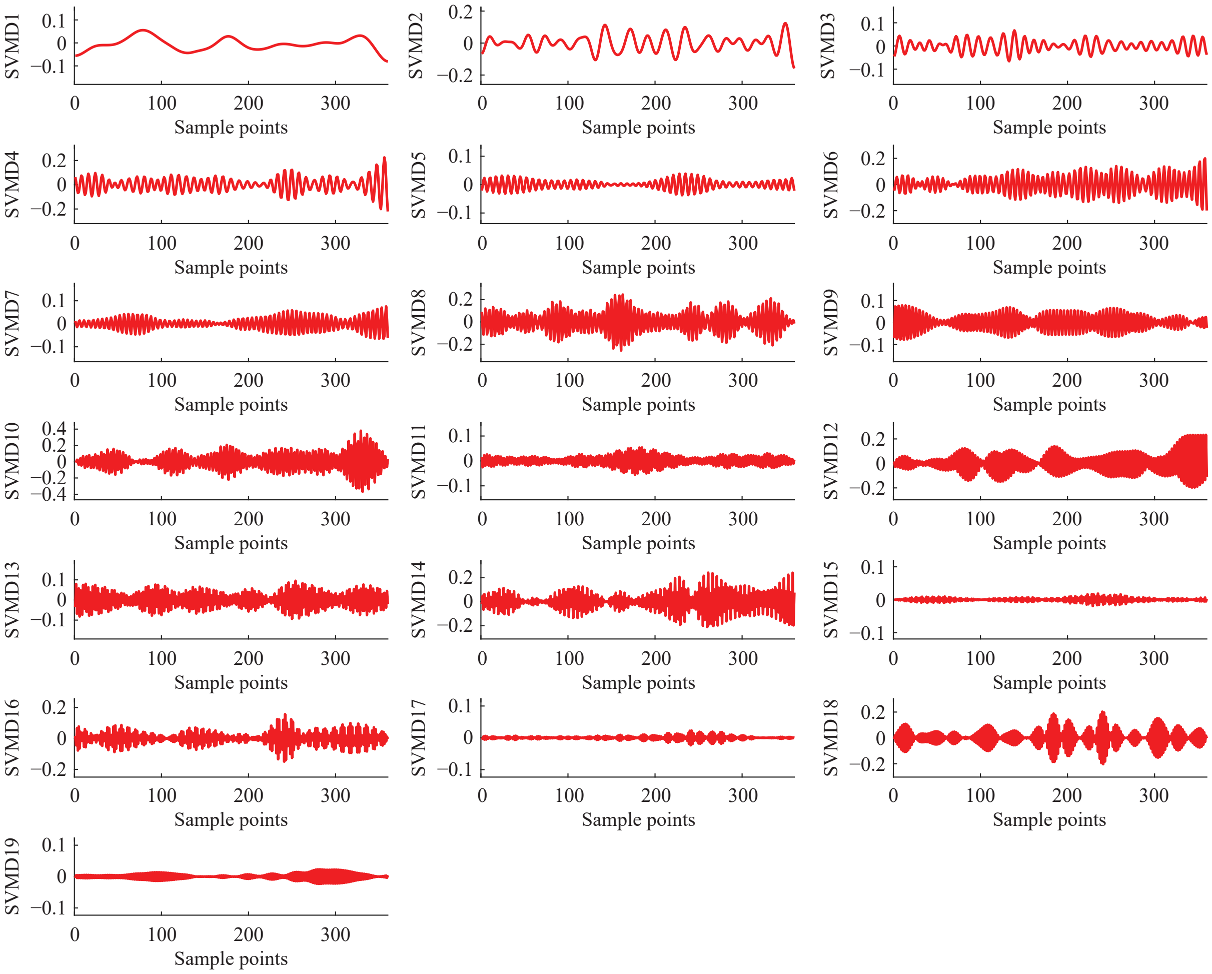


Figure 9 Secondary decomposition results of the maximum fuzzy entropy component CO-IMF2

A key advantage of the Extreme Learning Machine (ELM) is its ability to achieve high generalization performance with substantially reduced training time. This efficiency is attained by randomly setting the hidden layer parameters and determining the output weights through a direct analytical solution. Therefore, this study employed ELM for multivariate time series component prediction, combining sliding window construction and optimization algorithms for parameter tuning. The window length (k_{im}) was set to 6 with a prediction step size (z_{im}) of 1. The dataset was split into 80% for training and 20% for testing. In this framework, the Sigmoid function was utilized as the activation function, while the optimal number of hidden nodes in the ELM was automatically identified by the RIME optimization algorithm. During the optimization process, five-fold cross-validation is adopted to evaluate the performance of candidate parameters on the training set to avoid using test set information. The mean square error of cross-validation is used as the fitness function of the RIME algorithm. The data were normalized using min-max scaling. By optimizing the hidden layer structure and dynamic data partitioning, this approach enhances the model's ability to capture temporal features.

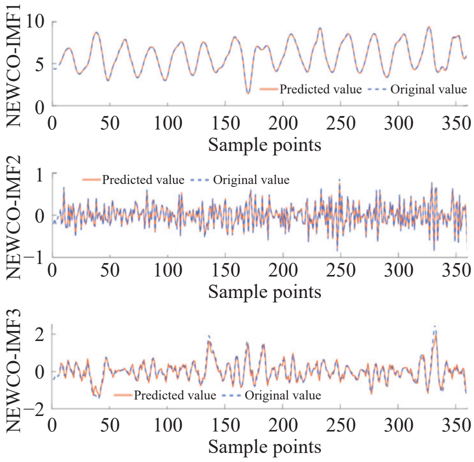


Figure 10 ELM prediction curve results

3.2.5 BP nonlinear superposition

BP neural network is a multi-layer feedforward network based on the supervised learning mechanism. It realizes the gradient optimization of network parameters through the error backpropagation algorithm and is a classic method for solving nonlinear modeling problems. The core lies in applying the chain rule to backpropagate output errors through the network, dynamically adjusting the weights and biases of each layer, so that the model output approaches the target value.

The integration of the individual ELM predictions was accomplished using a Back-Propagation (BP) neural network, substituting the conventional linear superposition method[24]. This nonlinear integration strategy is more suited to the nonlinear and non-stationary nature of dissolved oxygen data, leading to enhanced performance. The network architecture is defined by two hidden layers containing four and three neurons, respectively. The hidden layers employ the tansig (hyperbolic tangent sigmoid) activation function[25] for nonlinear mapping, whereas the output layer uses a purelin (linear) function[26] for regression tasks. Network training leverages the Levenberg-Marquardt (LM) algorithm[27], chosen for its rapid convergence on medium-sized datasets (samples < 1000), albeit with greater memory usage, due to its approximation of second-order derivatives. Key training parameters include a learning rate of 0.05 to maintain stability and avoid oscillations, a maximum of 150 epochs (empirically sufficient for convergence with the LM algorithm), and a mean squared error goal of 1×10^{-3}.

A quantitative evaluation of the final predictions, presented in Figure 11, demonstrates high model accuracy, as evidenced by the following metrics: an MAPE of 0.0352, an RMSE of 0.2443, an MAE of 0.1782, and an R^2 of 0.9835. Each indicator assesses a different aspect of performance. For instance, MAPE, which is preferred for its interpretability, quantifies the average relative error (smaller values are better); MAE indicates the average absolute prediction error; RMSE (smaller values are preferred due to sensitivity to large errors); and R^2 (values closer to 1 denote stronger explanatory power over data variance). By integrating the prediction results with error metrics, it can be concluded that the proposed model demonstrates excellent predictive performance, with outputs closely aligning with actual values.

3.2.6 Model time performance evaluation

Considering that the model needs to be applied in practice in the future, in addition to the error index, the time index of the model is also required. The time spent on each stage of the model is listed in Table 1.

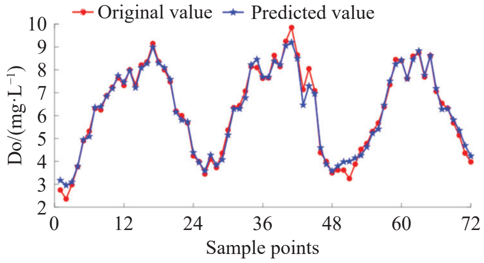


Figure 11 Final prediction curve results for the proposed model

Table 1 Detailed breakdown of model training and inference time

Stage	Steps	Time consumed	Remarks
Training phase	ICEEMDAN decomposition	0.11 s	-
	Fuzzy entropy clustering	0.02 s	-
	SVMD decomposition	0.58 s	-
	RIME optimization for ELM nodes	402.74 s	Total
	Per component optimization	~19.2 s	Average for 21 components
	BP neural network ensemble training	4.83 s	-
	Total training phase	408.27 s	Approx. 6.8 minutes
Inference phase	Single-component ELM inference	0.05-0.47 ms	Time range
	Single-sample BP network inference	2.94 ms	-

From the experimental data, the end-to-end pipeline of the proposed hybrid model exhibits distinct phase-wise characteristics. During the training phase, the total time consumption is approximately 6.8 min, with the vast majority of time concentrated in the RIME algorithm-based optimization process for the 21 ELM components (402.74 s, accounting for 98.6% of the total). This step is a computationally intensive offline process, typically performed on servers or high-performance workstations. In contrast, the time required for signal decomposition (ICEEMDAN, SVMD), feature clustering (fuzzy entropy), and final ensemble network (BP) training is minimal (totaling 5.54 s). During the inference phase, the model demonstrates efficient prediction capability: single-component ELM inference ranges from 0.05 to 0.47 ms, and BP ensemble network inference takes 2.94 ms. Based on this, the estimated total end-to-end inference time for a complete sample is under 15 ms. Given the aforementioned time performance, the deployment of this model on edge devices in farm settings is theoretically feasible.

3.3 Prediction performance

3.3.1 Ablation experiments

The proposed ICEEMDAN-KMEANS-SVMD-RIME-ELM-BP model was evaluated against five progressively complicated variants: ELM, ICEEMDAN-ELM, ICEEMDAN-KMEANS-RIME-ELM, ICEEMDAN-KMEANS-ELM-BP, and ICEEMDAN-KMEANS-RIME-ELM-BP in comprehensive ablation tests. The ablation study evaluates progressively enhanced models starting from the ELM baseline. K-means clusters ICEEMDAN components into three groups. RIME algorithm optimizes the parameters of ELM. SVMD performs secondary decomposition on the component with the maximum fuzzy entropy value. For ELM without backpropagation, final outputs are linearly combined; the incorporation of backpropagation facilitates nonlinear output transformation. In the proposed model, the components from SVMD's secondary decomposition are independently predicted by ELM and then linearly summed.

As seen in Figure 12, the prediction result curve of the proposed model is closer to the true value curve than the forecasting curves of the other five models, the error curve is more stable, and the proposed model has higher fitting precision than the other five. Therefore, this is the prediction model with better prediction data.

Figure 13 compares the predictive performances of various models—including the proposed model, ELM and its variants—using violin plots and APE (Absolute Percentage Error) results statistics plots.

The violin plot analysis reveals that the proposed model's distribution remains significantly closer to the true-value distribution than other ablation models, none of which successfully capture the right-skewed unimodal characteristic of the actual data.

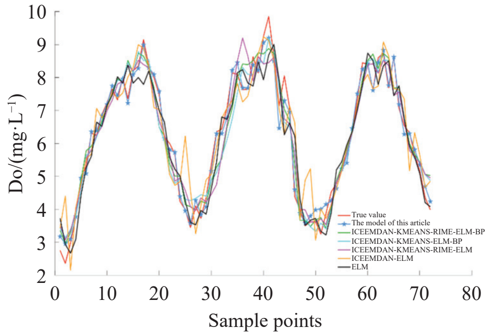


Figure 12 Prediction curve results for ablation models

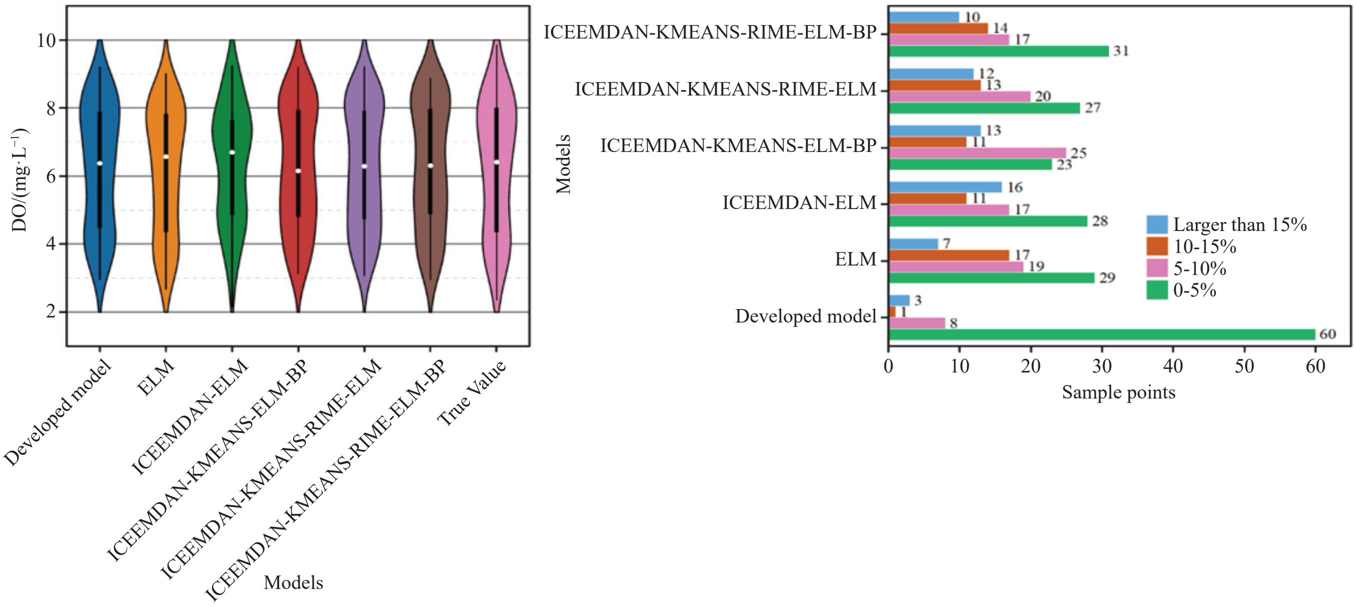


Figure 13 Violin plot and APE results statistics of six models

Based on the standard APE interpretation tiers where ≤5% is excellent, 5-10% is fair, 10-15% is marginal, and >15% is poor, the model's performance distribution in Figure 13 demonstrates robust accuracy. Specifically, 31 predictions (43.1% of 72 total samples) achieved excellent accuracy (APE ≤5%), while 17 (23.6% of 72 total samples), 14 (19.4% of 72 total samples), and 10 (13.9% of 72 total samples) predictions correspond to the fair, marginal, and poor categories, respectively. In ELM, there are 29 points lower than 5%, 19 data points between 5% and 10%, 17 points between 10% and 15%, and only seven points larger than 15. For ICEEMDAN-KMEANS-RIME-ELM, ICEEMDAN-KMEANS-ELM-BP, and ICEEMDAN-ELM, there are 27, 23, and 28 data points within the satisfactory error bound. In the proposed model, there are three points higher than 15%, only one point between 10% and 15%, eight points between 5% and 10%, and most of the points are below 5%. Our proposed model demonstrates overwhelming superiority over the second-best performer (ICEEMDAN-KMEANS-RIME-ELM-BP), which attains only 31 samples (43.1% of 72 total samples) in the high-precision 0%-5% range. Although this model leads other ablation variants, it remains significantly inferior to our approach. For large errors (>15%), our proposed model maintains the best performance with merely three outliers (4.2% of 72 total samples), compared to ELM's seven cases (9.7% of 72 total samples) as the runner-up. The proposed model consequently demonstrates dual competitive advantages in ablation studies: it simultaneously achieves exceptional prediction accuracy and maintains ultra-low risk exposure, attaining optimal performance at both precision boundaries. These experimental results provide conclusive evidence of our proposed model's superior predictive performance in ablation testing.

In order to evaluate the forecasting model more directly, four error metrics are calculated. The error metrics for each prediction model are given in Figure 14 and Table 2, which show that the proposed model has better precision in dissolved oxygen content prediction and exhibits consistently better performance than the other models in terms of various criteria. The proposed model reduces prediction errors by 29% in MAPE, 38% in MAE, and 37% in RMSE compared to the optimal ELM-based model (i.e., ICEEMDAN-KMEANS-RIME-ELM-BP). Although the performance of ICEEMDAN-KMEANS-RIME-ELM-BP is superior to other ELM variants, there is still a gap compared with the proposed model. Therefore, the experimental results demonstrate that applying SVMD-based secondary decomposition to the high-complexity component further separates the modes and significantly improves prediction accuracy.

3.3.2 Comparison experiments

For performance evaluation, the proposed model was compared against five benchmark models in a comparative study, including

Long Short-Term Memory (LSTM)[28], Convolutional Neural Network (CNN)[29], Genetic Algorithm-optimized Backpropagation (GA-BP)[30], Particle Swarm Optimization-Backpropagation (PSO-BP)[31], and Sine Cosine Sparrow Search Algorithm-Optimized Convolutional Neural Network-Bidirectional Long Short-Term Memory (SCSSA-CNN-BiLSTM). Both LSTM and CNN represent classical deep learning architectures capable of processing raw time-series data without feature engineering. GA-BP and PSO-BP are representative optimized prediction models. The SCSSA-CNN-BiLSTM represents a sophisticated hybrid model for time-series forecasting, integrating three core components: (1) sine cosine and cauchy mutation operators[32], (2) sparrow search algorithm[33], and (3) a CNN-BiLSTM neural architecture. This combined framework exhibits enhanced global search capability to escape local optima, making it particularly suitable as a benchmark for comparative evaluation with the proposed model.

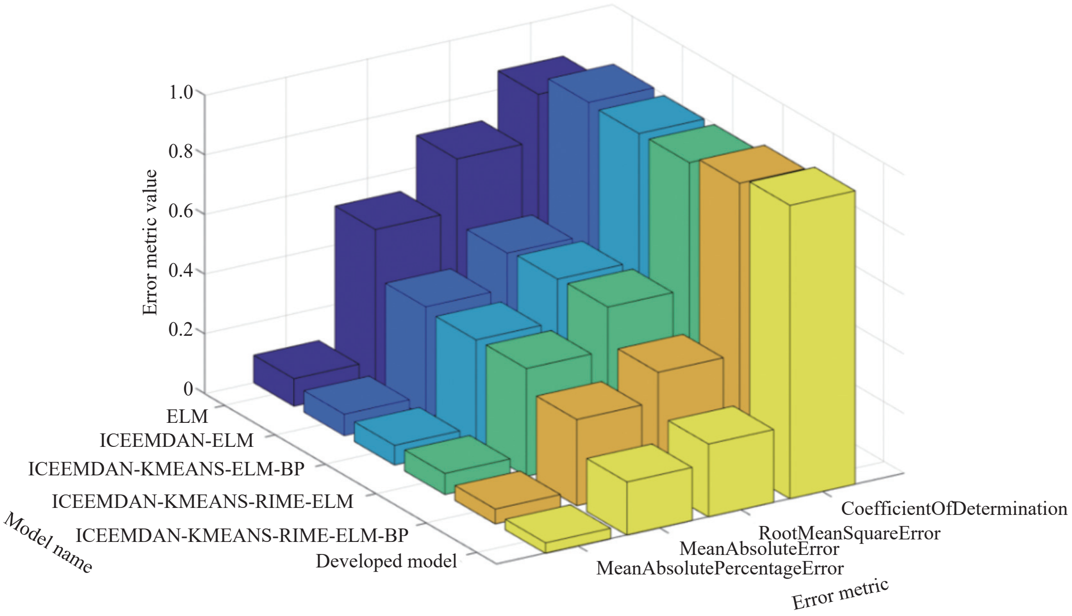


Figure 14 3D metric comparison plot of ablation models by different critical indices

Table 2 Error evaluation indicators for ablation models

Model	MAPE	MAE	RMSE	R^2
The proposed model	0.0352	0.1782	0.2443	0.9835
ELM	0.0919	0.5310	0.7067	0.8623
ICEEMDAN-ELM	0.0715	0.3707	0.4898	0.9339
ICEEMDAN-KMEANS-ELM-BP	0.0658	0.3585	0.5013	0.9282
ICEEMDAN-KMEANS-RIME-ELM	0.0703	0.3612	0.5062	0.9294
ICEEMDAN-KMEANS-RIME-ELM-BP	0.0497	0.2888	0.3861	0.9589

As illustrated in Figure 15, the proposed model yields a prediction curve that most accurately tracks the actual dissolved oxygen values, attaining the highest goodness-of-fit. At the other extreme, the GA-BP model displays the lowest fit to the data. An intermediate level of performance is observed for the SCSSA-CNN-BiLSTM model, which surpasses conventional models yet does not match the accuracy of the proposed hybrid approach, underscoring the advantage of sophisticated hybrid architectures.

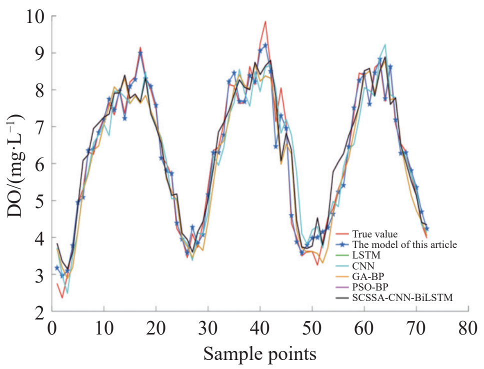


Figure 15 Prediction curve results for different models

This paper still presents the violin plot and APE statistical plot of the different comparison models including LSTM, CNN, GA-BP, PSO-BP, and SCSSA-CNN-BiLSTM to analyze the predictive performance of the model, as shown in Figure 16.

The violin plot reveals a striking similarity between the true-value distribution and the proposed model's output distribution, with near-perfect overlap and matching density peak heights. The proposed model's median of 6.42 most closely approximates the true median value of 6.44. Meanwhile, the median of SCSSA-CNN-BiLSTM is slightly lower than the median of the true value. However, comparative analysis of the violin plots reveals marked disparities in distributional correspondence and median alignment: conventional models (LSTM, CNN, GA-BP, PSO-BP) consistently underperform relative to both the proposed approach and SCSSA-CNN-BiLSTM.

Of the 72 predictions made by the SCSSA-CNN-BiLSTM model (Figure 16), 49 points (68.1% of 72 total samples) had an APE below 5%, 16 points (22.2% of 72 total samples) were between 5%-10%, five points (6.9% of 72 total samples) were between 10-15%, and only two points (2.8% of 72 total samples) exceeded 15%. For PSO-BP and GA-BP, there are 14 and nine data points larger than 15%, and 21 and 26 points lower than 5%. For CNN and LSTM, there are 14 and 11 data points larger than 15%, and 27 and 22 points within the satisfactory error bound. In our proposed model, there are only three points larger than 15%, one data point between 10% and 15%, eight points between 5% and 10%, and most of the points are below 5%. Our proposed model demonstrates overwhelming superiority over the second-best performer (SCSSA-CNN-BiLSTM model), which attains only 49 samples (68.1% of 72 total samples) in the high-precision 0%-5% range. Although this model leads other models, it remains significantly inferior to our approach. Clearly, the proposed model has high prediction accuracy.

A quantitative comparison of the six models was conducted using four error metrics, as summarized in Figure 17 and Table 3. The results indicate that the proposed model significantly outperforms the five benchmark models, achieving the lowest error values (MAPE=0.0352, MAE=0.1782, RMSE=0.2443). It outperforms LSTM, CNN, GA-BP, PSO-BP, and SCSSA-CNN-BiLSTM by 60.45%, 57.44%, 60.32%, 60.54%, and 12.22%, respectively in terms of MAPE index. It is also improved by 65.24%, 63.86%, 66.51%, 66.83%, and 31.04%, respectively in terms of MAE index. Similarly, it is improved by 58.94%, 59.48%, 62.70%, 61.92%, and 37.05%, respectively in terms of RMSE index. In terms of R^2 index, the proposed model has a value of 0.9835, indicating it has strong predictive power. It is obvious that the proposed model outperforms the five counterparts in terms of all precision indicators for dissolved oxygen prediction. Compared with the LSTM, CNN, GA-BP, and PSO-BP, the SCSSA-CNN-BiLSTM hybrid model demonstrates superior performance in dissolved oxygen prediction, achieving higher R^2 (0.9556) and lower MAPE, MAE, and RMSE values (0.0401, 0.2584, and 0.3881). This enhancement primarily stems from its unique hybrid architecture and optimized parameter selection, both of which significantly improve prediction accuracy. The proposed model demonstrates a marked improvement in performance over the established SCSSA-CNN-BiLSTM baseline.

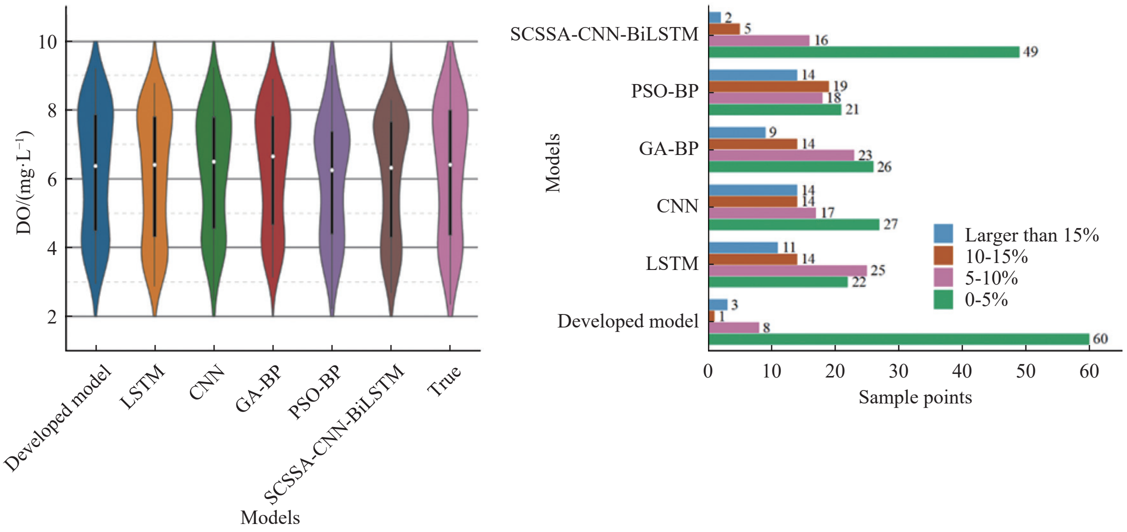


Figure 16 Violin plot and APE results statistics of six models

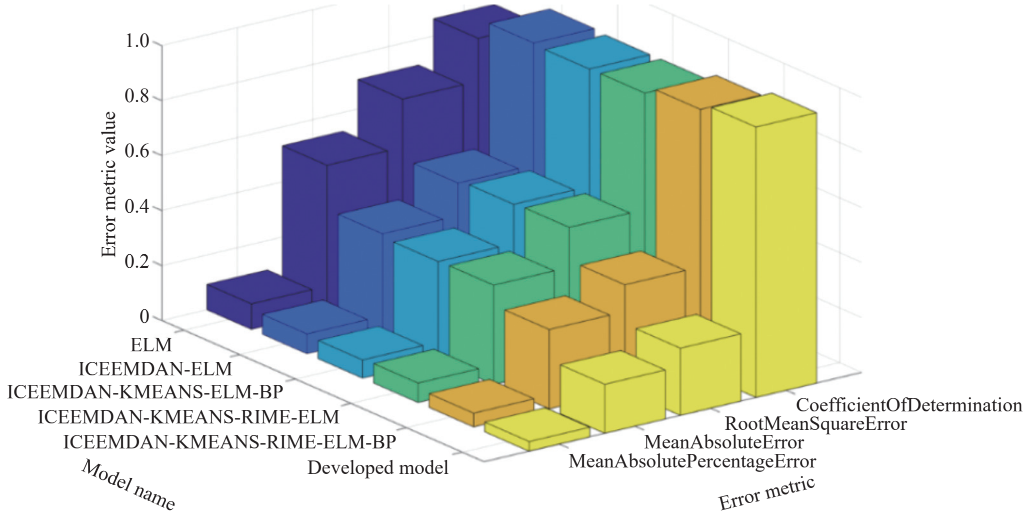


Figure 17 3D metric comparison plot of six models by different critical indices

Table 3 Error evaluation indicators of six models

Model	MAPE	MAE	RMSE	R^2
The proposed model	0.0352	0.1782	0.2443	0.9835
LSTM	0.0890	0.5126	0.5950	0.8854
CNN	0.0827	0.4931	0.6029	0.8967
GA-BP	0.0887	0.5321	0.6549	0.8781
PSO-BP	0.0892	0.5373	0.6416	0.8830
SCSSA-CNN-BiLSTM	0.0401	0.2584	0.3881	0.9556

As evidenced by the 3D metric comparison, the proposed model attains optimal values across all key indicators—lowest errors (MAPE, MAE, RMSE) and highest correlation coefficient—significantly exceeding the performance of competing models. These results collectively attest to its superior predictive capability for dissolved oxygen.

A comprehensive evaluation indicates that the proposed model—a synthesis of ICEEMDAN, K-means clustering, successive VMD, the RIME algorithm, ELM, and a BP network—yields more accurate predictions for dissolved oxygen levels in aquaculture than LSTM, CNN, GA-BP, PSO-BP, and SCSSA-CNN-BiLSTM models.

3.3.3 Generalization experiments

To verify that the excellent predictive performance of the model proposed in this paper on the “Nanheqiao No.1” dataset is not accidental, a generalization ability test was conducted on the model in this paper. The selected data samples were from different monitoring devices belonging to the Sea Cucumber Farm in Mu Hou Village, Xingcheng City, Huludao, Liaoning Province. The names of the monitoring devices are “2066000000” and “2066000012”, respectively.

The monitoring device “2066000000” collected data at an interval of 1 h. The data collection period was from August 3, 2024 to August 17, 2024, totaling 15 d and 360 sets of data. The time series diagram of the dataset is shown in Figure 18.

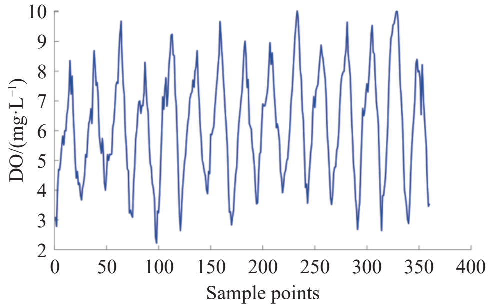


Figure 18 Dissolved oxygen time series plot

The prediction results obtained by using the exact same model operation platform and model structure as that of “Nanheqiao No.1” for this dataset are shown in Figure 19. The model demonstrates strong predictive performance, with an RMSE of 0.4002, an MAE of 0.2977, an MAPE of 0.0561, and an R^2 of 0.9655.

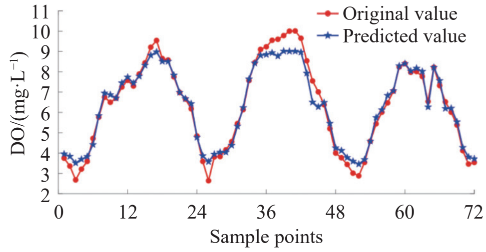


Figure 19 Final prediction curve results for the proposed model

The monitoring device “2066000012” also collects data at an interval of 1 h. The data collection period is from October 1, 2024 to October 15, 2024, totaling 360 sets of data over 15 d. The time series diagram of the dataset is shown in Figure 20.

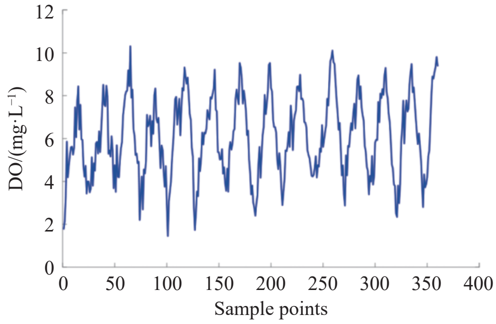


Figure 20 Dissolved oxygen time series plot

The prediction results obtained by using the exact same model operation platform and model structure as that of “Nanheqiao No.1” for this dataset are shown in Figure 21. The model demonstrates strong predictive performance, with an RMSE of 0.3751, an MAE of 0.2782, an MAPE of 0.0479, and an R^2 of 0.9678.

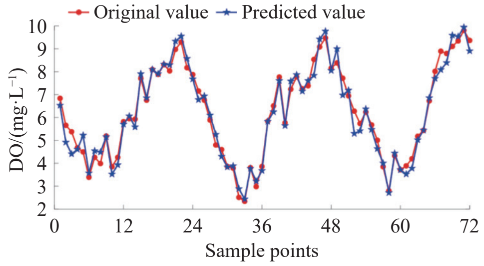


Figure 21 Final prediction curve results for the proposed model

Based on the error metrics of “Nanhe Bridge No.1”, “2066000000”, and “2066000012”, the average R^2 of the model in this paper on the three datasets reaches 0.9723, and the average MAPE reaches only 4.64%, indicating that the model stably explains over 97% of the data variance, with a mean prediction error well below the 5% industrial benchmark for excellence. Particularly noteworthy is the model’s high consistency across datasets: R^2 remains consistently above 0.9655, while MAPE varies within the excellent range of 3.52% to 5.61%, showing no performance breakdown. This cross-dataset stability confirms the model’s strong generalization capability, rather than overfitting to specific data. In terms of error distribution, the reasonable ratio between RMSE and MAE indicates the absence of disruptive extreme outliers in the prediction errors. Overall, the model has reached a top-tier level suitable for industrial applications and fully meets the requirements for deployment in high-precision prediction scenarios.

4 Conclusions

To enhance the accuracy of nonlinear and non-stationary dissolved oxygen prediction, this paper proposes a novel hybrid model named ICEEMDAN-KMEANS-SVMD-RIME-ELM-BP, based on an analysis of dissolved oxygen dynamics. Subsequently the proposed method was implemented on the dissolved oxygen time series data from the Huludao sea cucumber aquaculture base. The primary findings of this research are as follows:

(1) Fuzzy entropy values were computed for the IMFs resulting from the initial ICEEMDAN decomposition. K-means clustering was then performed based on these values, effectively extracting the component with maximum fuzzy entropy. Subsequently, SVMD was applied for secondary decomposition, which significantly enhanced the feature extraction capability for all components and reduced the non-stationarity of the dissolved oxygen time series.

(2) The secondary decomposition process substantially reduced the complexity inherent in the dissolved oxygen time series.

(3) By employing the RIME algorithm to automatically define the optimal hidden layer structure, the ELM’s architecture is optimized, and the subjective randomness of manual parameter setting is eliminated. Compared with LSTM, CNN, GA-BP, PSO-BP, and SCSSA-CNN-BiLSTM prediction models, the proposed hybrid model has better forecasting effect and higher forecasting accuracy, among which MAPE is 0.0352, MAE is 0.1782, RMSE is 0.2443, and R^2 is 0.9835.

(4) Generalization experiments have verified that the proposed model possesses excellent stability and generalization capabilities.

However, the research has limitations that require further investigation. The reliability of a model under abnormal and extreme conditions is crucial for its practical application. Therefore, priority will be given to the robustness validation of the model. This involves collecting long-term on-site data covering different seasons, various pond types, and abnormal events (e.g., typhoons, cold spells, or equipment failures). Additionally, methods such as data augmentation or mechanistic simulation can be explored to generate extreme-case data for testing and optimizing the model's performance at its operational limits. Since season and climate strongly influence dissolved oxygen, these factors can be used to identify historical periods matching forecasted weather patterns, thereby refining the model input. Water temperature, illumination, atmospheric pressure, and pH are also highly correlated with dissolved oxygen. However, as they cannot be adopted as inputs for this model, forecasting accuracy may be affected. Future studies could thoroughly investigate the correlation of these factors with the optimal aquaculture environment. Further research directions may include incorporating additional influencing factors and integrating more advanced artificial intelligence algorithms into the hybrid framework to enhance predictive accuracy. Based on the prediction results, managers can monitor dissolved oxygen levels in real time and develop targeted emergency response plans for sudden dissolved oxygen fluctuations. A three-dimensional study of aquaculture ponds may also be considered. Achieving this requires strategically deploying dissolved oxygen sensors at various locations and depths across large aquaculture ponds. This will enable comprehensive measurement and prediction of three-dimensional water quality parameters, providing managers with more actionable data and operational convenience.

Acknowledgements

This research was financially supported by the National Natural Science Foundation of China (Grant No. 52477226 and 52277213).

[References]

[1] Yu Z H, Zhou Y, Yang H S, Hu C Q. Bottom culture of the sea cucumber *Apostichopus japonicus* Selenka (Echinodermata: Holothuroidea) in a fish farm, southern China. Aquaculture Research, 2014; 45(9): 1434–1441.

[2] Hu J Y, Wang J J, Zhang X S, Fu Z T. Research status and development trends of information technologies in aquacultures. Transactions of the CSAM, 2015; 46(7): 251–263.

[3] Mendivil-García K, Amabilis-Sosa L E, Salinas-Juárez M G, Pat-Espadas A, Rodríguez-Mata A E, Figueroa-Pérez M G, et al. Climate change impact assessment on a tropical river resilience using the Streeter-Phelps dissolved oxygen model. Frontiers in Environmental Science, 2022; 10: 3389.

[4] Li X Y, Wang H, Wang Y Q, Zhang L J, Wu Y. Machine learning-based dissolved oxygen prediction modeling and evaluation in the Yangtze River estuary. Environmental Science, 2024; 45(12): 7123–7133.

[5] Liu S Y, Xu L Q, Li D L, Zeng L H. Dissolved oxygen prediction model of *eriocheir sinensis* culture based on least squares support vector regression optimized by ant colony algorithm. Transactions of the CSAE, 2012; 28(23): 167–175.

[6] Liu S, Xu L Q, Jiang Y, Li D L, Chen Y Y, Li Z B. A hybrid WA–CPSO-LSSVR model for dissolved oxygen content prediction in crab culture. Engineering Applications of Artificial Intelligence, 2014; 29(3): 114–124.

[7] Huan J, Liu X Q. Dissolved oxygen prediction in water based on K-means clustering and ELM neural network for aquaculture. Transactions of the CSAE, 2016; 32(17): 174–181.

[8] Shi P, Kuang L, Yuan L M, Wang Q, Li G H, Yuan Y M, et al. Dissolved oxygen prediction using regularized extreme learning machine with clustering mechanism in a black bass aquaculture pond. Aquacultural Engineering, 2024; 105: 102408.

[9] Li W H, Dong Z T, Chen T, Wang F H, Huang F L. Enhanced prediction of dissolved oxygen concentration using a hybrid deep learning approach with sinusoidal geometric mode decomposition. Water, Air, & Soil Pollution, 2024; 235: 447. DOI: 10.1007/S11270-024-07242-X

[10] He M, Cui M, Zheng Q Y, Xu L Q, Liu S Y. PSG-Crossformer: a hybrid model for long-term dissolved oxygen prediction in aquaculture. Aquaculture International, 2025; 33: 301.

[11] Jiao J G, Ma Q Q, Liu F L, Zhao L Q, Huang S J. A hybrid prediction model of dissolved oxygen concentration based on secondary decomposition and bidirectional gate recurrent unit. Environmental Geochemistry and Health, 2024; 46: 127.

[12] Su H, Zhao D, Asghar H A, Liu L, Zhang X Q, Majdi M, et al. RIME: A physics-based optimization. Neurocomputing, 2023; 532: 183–214.

[13] Colominas M A, Schlotthauer G, Torres M E. Improved complete ensemble EMD: A suitable tool for biomedical signal processing. Biomedical Signal Processing and Control, 2014; 14: 19–29.

[14] Kong Y L, Meng Y, Li W, Yue A Z, Yuan Y. Satellite image time series decomposition based on EEMD. Remote Sensing, 2015; 7(11): 15583–15604.

[15] Wu Z H, Huang N E. Ensemble empirical mode decomposition: A noise-assisted data analysis method. Advances in Adaptive Data Analysis, 2009; 1(1): 1–41.

[16] Dragomiretskiy K, Zosso D. Variational mode decomposition. IEEE Trans. Signal Processing, 2013; 62(3): 531–544.

[17] Huang G B, Chen L, Siew C K. Universal approximation using incremental constructive feedforward networks with random hidden nodes. IEEE Transactions on Neural Networks, 2006; 17(4): 879–892.

[18] Ravikumar R R S, Apparao G. Extreme learning machine with Sigmoid activation function on large data. International Journal of Recent Technology and Engineering, 2019; 8(2): 3523–3526.

[19] Zhou Y T, Zhao X Y, He F, Shi C J. Atmospheric temperature and humidity prediction based on Gaussian process mixture model. Transactions of the CSAE, 2018; 34(5): 219–226.

[20] Hassani H, Yeganegi M R, Mehreyan S Z, Sayyareh A. On the sample autocorrelation function's absolute summability. Fluctuation and Noise Letters, 2022; 21(1). DOI: 10.1142/S0219477522500043

[21] Hassani H, Marvian L, Yarmohammadi M, Yeganegi M R. Unraveling time series dynamics: evaluating partial autocorrelation function distribution and its implications. Mathematical and Computational Applications, 2024; 29(4): 58–58.

[22] Belaire-Franch J. Estimating the maximum Lyapunov exponent with denoised data to test for chaos in the German stock market. Computational Economics, 2025; 66: 3517–3543.

[23] Wei L, Cao H, Wang L D. Optimal choices of reconstruction parameters united improved C-C algorithm and improved G-P algorithm. Journal of Physics: Conference Series, 2023; 2508(1). DOI: 10.1088/1742-6596/2508/1/012057

[24] Adnan R M, Parmar K S, Heddam S, Shahid S, Kisi O. Suspended sediment modeling using a heuristic regression method hybridized with kmeans clustering. Sustainability, 2021; 13(9): 4648–4648.

[25] Ali K B, Morteza F A, Mohamed S, Haj A M E, Natarajan P. Modeling carbon dioxide emission of countries in southeast of Asia by applying artificial neural network. International Journal of Low-Carbon Technologies, 2022; 17: 321–326.

[26] Tümer A E, Edebali S. Modeling of trivalent chromium sorption onto commercial resins by artificial neural network. Applied Artificial Intelligence, 2019; 33(4): 349–360.

[27] Yaldi G, Nur I M, Apwiddhal. Calibrating trip distribution neural network models with different scenarios of transfer functions used in hidden and output layers. International Journal on Advanced Science, Engineering and Information Technology, 2020; 10(6): 2410–2418.

[28] Hu C Y, Zhong Y, Lu Y Q, Luo X T, Wang S R. A prediction model for time series of dissolved gas content in transformer oil based on LSTM. Journal of Physics: Conference Series, 2020; 1659(1): 012030.

[29] Jiang W T, Ling L W, Zhang D B, Lin R B, Zeng L L. A time series forecasting model selection framework using CNN and data augmentation for small sample data. Neural Processing Letters, 2023; 55(5): 5783–5810.

[30] Ma J K, Xu W P. GA-BP-based nonlinear time series forecasting: method and applications. Academic Journal of Computing & Information Science, 2024; 7(8): 15–20.

[31] Gao D, Li K, Cai Y C, Wen T. Landslide displacement prediction based on time series and PSO-BP model in three georges reservoir, China. Journal of Earth Science, 2024; 35(3): 1079–1082.

[32] Yan J L, Hu G, Shu B. MGCHMO: A dynamic differential human memory optimization with Cauchy and Gauss mutation for solving engineering problems. Advances in Engineering Software, 2024; 198: 103793.

[33] Yang X Y, Zhou G F, Ren Z J, Qiao Y X, Yi J C. High-precision air conditioning load forecasting model based on improved sparrow search algorithm. Journal of Building Engineering, 2024; 92: 109809.