



---

*Research article*

## **Prediction of dissolved oxygen concentration in aquaculture based on attention mechanism and combined neural network**

Wenbo Yang<sup>1</sup>, Wei Liu<sup>1,2,\*</sup> and Qun Gao<sup>1</sup>

<sup>1</sup> School of Information and Electronic Engineering, Shandong Technology and Business University, Yantai 264005, China

<sup>2</sup> Key Laboratory of Sensing Technology and Control in University of Shandong, Yantai 264005, China

\* **Correspondence:** Email:liuwei@sdtbu.edu.cn.

**Abstract:** As an essential water quality parameter in aquaculture ponds, dissolved oxygen (DO) affects the growth and development of aquatic animals and their feeding and absorption. However, DO is easily influenced by external factors. It is not easy to make scientific and accurate predictions of DO concentration trends, especially in long-term predictions. This paper uses a one-dimensional convolutional neural network to extract the features of multidimensional input data. Bidirectional long and short-term memory neural network propagated forward and backward twice and thoroughly mined the before and after attribute relationship of each data of dissolved oxygen sequence. The attention mechanism focuses the model on the time series prediction step to improve long-term prediction accuracy. Finally, we built an integrated prediction model based on convolutional neural network (CNN), bidirectional long and short-term memory neural network (BiLSTM) and attention mechanism (AM), which is called CNN-BiLSTM-AM model. To determine the accuracy of the CNN-BiLSTM-AM model, we conducted short-term (30 minutes, one hour) and long-term (6 hours, 12 hours) experimental validation on real datasets monitored at two aquaculture farms in Yantai City, Shandong Province, China. Meanwhile, the performance was compared and visualized with support vector regression, recurrent neural network, long short-term memory neural network, CNN-LSTM model and CNN-BiLSTM model. The results show that compared with other comparative models, the proposed CNN-BiLSTM-AM model has an excellent performance in mean absolute error, root means square error, mean absolute percentage error and determination coefficient.

**Keywords:** dissolved oxygen prediction; deep learning; multi-step prediction; combinatorial neural network; attention mechanism

---

## 1. Introduction

According to the Food and Agriculture Organization of the United Nations (FAO), more than one billion people worldwide rely on fish to supplement their bodies with animal protein [1]. Over the past 30 years, aquaculture has been the fastest-growing sector in agriculture. It is one of the pillar industries driving China's economy, creating many jobs in rural areas and bringing stable income to farmers [2]. With the development of artificial intelligence and big data technology, how to increase aquaculture production based on modern information technology and improve the information management of fisheries and fishery administration has become a hot research topic.

For aquatic animals, DO is essential in sustaining their lives, and survival and reproduction can only occur under oxygenated conditions. At the same time, too high or too low dissolved oxygen concentrations can be fatal to the health of aquatic products and must be kept within a reasonable range [3]. When DO levels are too high, fish are prone to bubble disease [4]. On the contrary, when DO is below the standard index for a long time, the growth of aquatic organisms will be slowed down, disease resistance will be reduced, and death will result in severe cases. Therefore, aquaculturists would be very convenient if the trend of DO concentration could be accurately predicted in advance. However, accurately predicting DO concentration trends are challenging. Since aquaculture is in an open-air environment some microorganisms in the water will increase the DO content through photosynthesis; meanwhile, fish and phytoplankton will also accelerate the oxygen consumption through respiration [5]. Different depths and water temperatures will also lead to uneven DO distribution within the culture water [6]. Therefore, the DO time series monitored by water quality sensors will show a nonlinear characteristic [7,8]. With the increase in the prediction time, the nonlinear characteristics will gradually decrease the model's accuracy [9, 10].

In order to reduce the influence of nonlinear characteristics on prediction results, researchers have gradually designed various water quality prediction models for various application scenarios, which can be divided into mechanistic and nonmechanistic models according to their working principles [11, 12]. The mechanical model is derived from the system structure based on the physical, chemical, biological and other reaction processes of the water environment system with the help of many hydrological, water quality, meteorological and other monitoring data [13]. Because the mechanistic model requires a large amount of basic information about the water environment, which is usually very complex, it limits the further application in water quality prediction.

With the development and application of computer technology, more and more non-mechanical models are being applied to water quality prediction [14]. The non-mechanical water quality prediction method does not consider the physical and chemical changes of the water body. It builds a corresponding model based on historical data to predict the changing trend. The process is simple, and the effect is good. Mainly including time series, regression, probabilistic statistical, machine learning, and deep learning [15–17]. For example, Shi et al. [18] propose the softplus extreme learning machine (ELM) model based on clustering to accurately predict changes in DO for the nonlinear characteristics of the DO data in aquaculture waters. The model employs partial least squares while using a new Soft-plus activation function to improve the ELM, which solves the nonlinearity problem in time series data streams and avoids the instability of the output weight coefficients. However, this model is not suitable for training a large amount of historical data, so that the deep learning model may be more suitable for actual aquatic data. Li et al. [19] developed three deep learning models, recurrent neural network

(RNN), long short-term memory neural network (LSTM) and gated recurrent unit (GRU), to predict DO in fish ponds. The results showed that the performance of GRU is similar to LSTM, but the time cost and the number of parameters used by GRU are much lower than LSTM, which is more suitable for DO prediction in natural fish ponds. Although the above three deep learning models can obtain excellent prediction results while training a large amount of aquatic data, the accuracy will decrease significantly with the increase in prediction time.

In recent years, many scholars have also decomposed the raw DO data first and then used different machine learning models or deep learning models to predict the characteristics of the decomposition results. For example, Li et al. [20] proposed a hybrid model of integrated empirical mode decomposition based on multiscale features. Ren et al. [21] proposed a prediction model based on variational modal decomposition and deep belief network combined prediction model. Huang et al. [22] proposed a combined prediction model based on fully integrated empirical mode decomposition, adaptive noise and GRU combined with an improved particle swarm algorithm. Compared with various benchmark models, this method can effectively consider complex time-series data and predict DO variation trends reliably. This decomposition-based method can effectively separate and denoise the original data and enhance the neural network input data quality, thus improving the model's prediction accuracy. However, the method based on decomposition before prediction can cause the problem of boundary distortion, and there is a possibility that future data will be used in the model training [23]. Because of the excellent performance of the attention mechanism in artificial intelligence, many scholars have introduced it to DO prediction in aquaculture. For example, Bi et al. [24] added an attention mechanism after the LSTM network for multi-steps prediction of DO. Liu et al. [25] built a combined prediction model combining attention mechanism and RNN network and obtained excellent short-term and long-term prediction.

Based on previous studies, we investigate a combined model combining convolutional neural network, attention mechanism, and bidirectional long-term and short-term memory neural networks for short-term and long-term prediction of DO concentrations in aquaculture. It consists of a one-dimensional convolutional neural network (1D-CNN), a BiLSTM, and the AM, called the CNN-BiLSTM-AM model. The 1D-CNN helps the model extract important feature data from multiple input vectors. The BiLSTM adds forward and backward propagation to the LSTM, which can learn the input data's backward and forward adjacency relationships in the input data to achieve the purpose of fully mining the multiple data features. After this, the BiLSTM network is combined with the AM. The model's attention is focused on the moving step to capture the effect of different time steps on DO concentration prediction and improve the accuracy and stability of the model in long-term prediction.

The remainder of this paper is organized as follows: Section 2 describes the materials and methods. Section 3 details the experiments and analysis of results. Finally, Section 4 gives conclusions and directions for future work.

## 2. Materials and methods

### 2.1. Problem definition

The time series is a set of numerical sequences formed by arranging them in chronological order [26]. Time series are divided into unary time series and multivariate time series, and multivariate time series is the combination of multiple unary time series, which can be regarded as a sampling

process to obtain multiple observed variables from different sources [27]. The multidimensional DO concentration prediction model for aquaculture proposed in this paper is a multivariate time series prediction problem. For a given time series data with  $N$  characteristic variables, multivariate single-step prediction can be defined by the Eq (2.1):

$$\hat{y}_{t+1} = f(x_0^0, x_1^1, \dots, x_t^{N-1}) \quad (2.1)$$

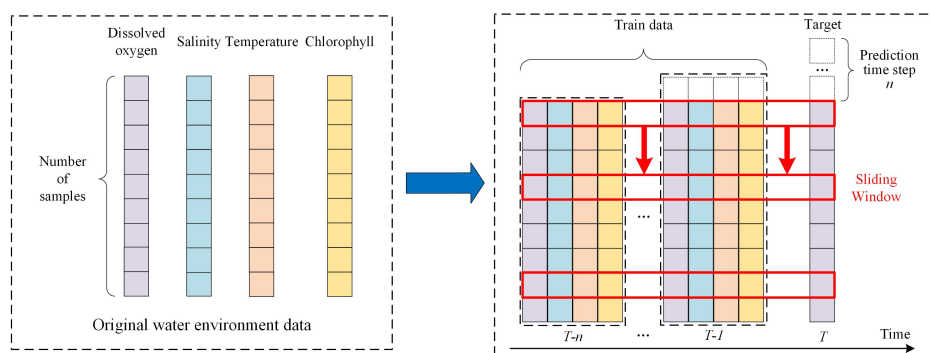
where  $\hat{y}_{t+1}$  represents the model's estimate of DO concentration for one future moment,  $x_i^j = (x_0^j, x_1^j, \dots, x_t^j)^T$  represents the data vector for the  $j \in [0, N)$  eigenvariable at the moment  $i \in \{0, 1, 2, \dots, t\}$ .

The multivariate multi-step prediction is based on the multivariate single-step prediction. It uses  $k \cdot N$  characteristic variables in the last  $k$  moments as inputs to predict the DO concentration for the next  $k$  moments. The Eq (2.2) demonstrates the  $k$  steps prediction.

$$(\hat{y}_{t+1}, \dots, \hat{y}_{t+k}) = f(X_0, X_1, \dots, X_k) \quad (2.2)$$

where  $X_i \in \{X_1, X_2, \dots, X_k\}$  is a multi-step supervised learning dataset constructed  $i$  moments in advance, model  $f(\cdot)$  is usually estimated using a supervised learning strategy and multi-step prediction, using training data and corresponding labels for estimation.

The data used in this paper were obtained from two aquaculture farms in Yantai, Shandong Province, China. The marine farm is equipped with multi-parameter water quality monitoring sensors to collect various water environment data in real-time, including water temperature, salinity, chlorophyll concentration, and DO concentration. All data are collected at a frequency of 10 minutes. This paper uses the first 80% of the data collected in the experiment as the training set samples. The last 20% of the data collected is divided into validation and test set samples. The training set is used to adjust the model's internal parameters, such as the weights and bias vectors of the neural network. The validation data check whether the model is overfitted or underfitting. The test set data is used to evaluate the model's predictive performance.



**Figure 1.** Multi-step predictive supervised learning data construction.

Multivariate multi-step prediction can be defined as a supervised learning problem. The multi-dimensional data are stitched together to form a matrix. Specifically, the inputs to the model are DO concentration, salinity, water temperature and chlorophyll content at the past time. The output of the

model is the DO concentration at the current time.  $T$  denotes the current moment and  $n$  represents the prediction time step. Figure 1 illustrates the construction process of the supervised learning dataset in this paper, using multi-dimensional data from the previous  $(T - n)$  moments to predict the DO concentration at the moment. The data surrounded by red boxes represent a set of constructed training data and target labels, and the sliding window technique is used to repeat the operation to complete the construction of all data.

At the same time, the DO concentration at the current moment can be predicted in several time steps; the one-time step is 10 min, using sliding windows of different sizes. In order to verify the performance of the model in short-term prediction, three-step prediction (30 minutes) and six-steps prediction (1 hour) are selected for experimental analysis. The three-step prediction uses the historical data of the first three moments to predict the DO trend of the current moment 30 minutes in advance. The six-step prediction uses the historical data of the first six moments in advance to predict the DO trend of the current moment. Based on this, this paper chooses 36 steps ahead (6 hours) and 72 steps ahead (12 hours) to verify the performance of the CNN-BiLSTM-AM model in long-term prediction.

## 2.2. CNN

The CNN is an improvement of Lecun on multilayer perceptron (MLP) [28]. Due to its structural features of local area connectivity, weight sharing, and downsampling, the CNN excels in image processing. The application scenarios of CNN are specifically in image classification, face recognition, autonomous driving, and target detection [29–31]. There are three types of convolution operations: 1D, 2D convolution and 3D convolution [32]. The 1D convolution is used to process sequential data, such as in natural language processing; 2D convolution is often used in computer vision and image processing; 3D convolution is often used in medicine and video processing.

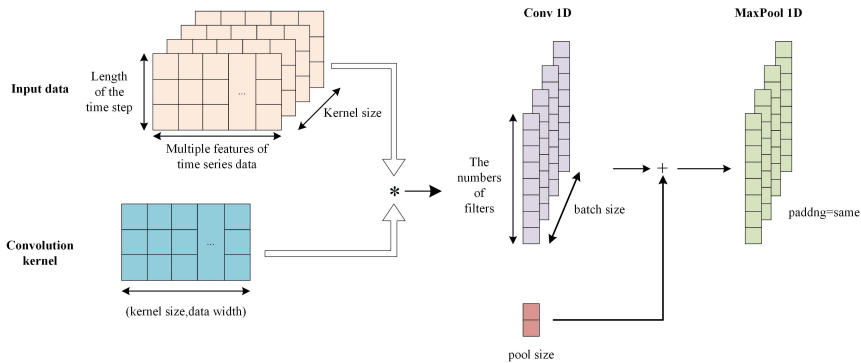
Since this paper focuses on predicting DO concentration sequences in aquaculture and is one-dimensional data, 1D-CNN is used for feature extraction. The 1D-CNN is calculated as shown in Eq (2.3):

$$h_t = \sigma(W * x_t + b) \quad (2.3)$$

where  $W$  denotes the convolution kernel, also called the weight coefficient of the filter in the convolution layer.  $b$  denotes the bias vector.  $x_t$  represents the sample data of the  $t_{th}$  input network.  $*$  denotes the convolution operator.  $\sigma$  represents the activation function.  $h_t$  represents the output result after the convolution operation.

Figure 2 shows the structure of 1D-CNN. In this paper, the 1D-CNN is used to extract the features of DO concentration for the characteristics of time series data. Firstly, the historical data of the aquaculture water environment with multiple features are stitched into a matrix. In order to feed the constructed supervised learning data into the interior of the CNN network, the dimensions of the matrix are therefore converted into  $k$  tensors. At the same time, each tensor has  $m$  rows and  $n$  columns. The  $k$  represents the size of the one-dimensional convolution kernel,  $m$  is a variety of water environment data from the past few days, and  $n$  represents the length of the time step. After the convolution operation, this paper uses maximum pooling to retain the most vital features and eliminate the weak ones. The purpose is to reduce the complexity and avoid overfitting the model. The step of maximum pooling is to place the pooling window on the sequence and use the maximum value within that window as the

output value of the pooling. The window is then slid, and the above steps are repeated until the end of the sequence.



**Figure 2.** The structure of 1D-CNN model.

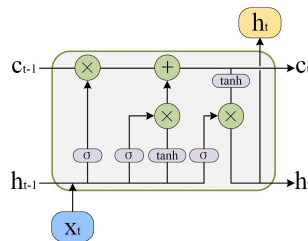
2.3. LSTM

Although the RNN can learn the relationship between the current moment and the earlier moment information in the long time series prediction problem, the longer the time goes, the more difficult it is for RNN to learn this relationship. Researchers call this phenomenon the long-term dependence problem [33]. It is like a person with weak memory who cannot remember the past is the same. The root cause of the long-term dependency problem is that the gradient tends to disappear or explode after the RNN has propagated through many stages. To solve the long-term dependency problem, Hochreiter and Schmidhuber [34] proposed the LSTM in 1997. LSTM is a modification of RNN, and Figure 3 illustrates the basic structural unit of LSTM. The LSTM network consists of several structural units, each containing three gating mechanisms: forgetting, input, and output.

The forgetting gate mainly determines the degree of forgetting of previous information. The forgetting gate decides which information from the past is discarded after receiving the last moment output  $h_{t-1}$  and the current moment input  $x_t$ . The forgetting gate of LSTM is calculated as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{2.4}$$

where  $f_t$  represents the output of the forgetting gate and  $\sigma$  represents the activation function. The output of the sigmoid function in the forgetting gate ranges between (0, 1) so that selective discarding of the data can be achieved.



**Figure 3.** The structure of one LSTM neuron.

The role of the input gate is to select which current information is input to the internal network after the forgetting gate has discarded some of the information [35, 36]. The input gate of LSTM consists of two steps. Firstly, determining which values need to be updated according to Eqs (2.5) and (2.6). Secondly, updating the cell state of the last moment to the cell state of the current moment according to Eq (2.7).

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2.5)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (2.6)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (2.7)$$

As the last part inside the LSTM network, the output gate obtains the model's output value, and the calculation formula is as follows: Eqs (2.8) and (2.9). The initial output value  $o_t$  is first calculated using the output  $h_{t-1}$  of the previous moment and the input  $x_t$  of the current moment, which is used to control the information to be output. The hyperbolic tangent activation function ( $\tanh$ ) then scales the cell state  $c_t$  at the current moment between  $-1$  and  $1$  and multiplies the result by the initial output value  $o_t$  to obtain the final output of the output gate.

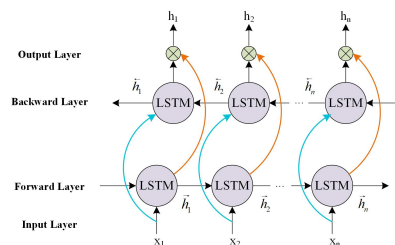
$$o_t = \sigma(W_o [h_{t-1}, x_t] + b_o) \quad (2.8)$$

$$h_t = o_t * \tanh(C_t) \quad (2.9)$$

where  $w_f, w_i, w_c, w_o$  are the weight matrices of oblivion gate, input gate, state update, and output gate.  $b_f, b_i, b_c,$  and  $b_o$  represent the corresponding bias vectors.

#### 2.4. BiLSTM

Although the LSTM network internally solves the long-term dependency problem by adding a threshold mechanism, the LSTM can only handle temporal information in one direction. For the long-time sequence problem, the current state may relate to the previous information and involve the following information. The BiLSTM combines the bi-directional recurrent neural network [37] and LSTM to fully use the sequence's historical and future information to improve the model's prediction performance [38]. Since BiLSTM networks can obtain the before-and-after features of the input sequences, they have been widely used in machine translation and speech recognition.



**Figure 4.** The structure of one LSTM neuron.

The BiLSTM neural network adds a forward layer and backward layer to the LSTM structure, and the main structure is shown in Figure 4. The principle of each propagation layer in BiLSTM is precisely

the same as the structure of the one-way propagation LSTM model. As shown in Eq (2.10), the final output consists of a superposition of LSTMs in both directions.

$$h_t = \vec{h}_t \otimes \overleftarrow{h}_t \quad (2.10)$$

where  $t$  represents the moment in the time series;  $\vec{h}$  and  $\overleftarrow{h}$  represent the output values of the forward and backward propagation layers, respectively.  $h_t$  represents the output results after the forward and backward superposition at moment  $t$ .

## 2.5. AM

When we look through a photo album quickly, we may not see the whole picture but often focus on the most beautiful parts of the picture. The AM references the above idea explained in mathematical language as  $X = [x_1, x_2, x_3, \dots, x_n]$  representing  $n$  inputs of information. In order to save computational resources, we do not want the neural network to process all the inputs but only select some information from the most relevant inputs to the task. In recent years, the AM mechanism has been widely used in image enhancement, text classification, and machine translation [39, 40].

In this paper, we use soft attention, the most common attention method. It means that when information selection is performed, instead of selecting one of the  $n$  input information, the weighted sum of the  $n$  input information is calculated and then input to the neural network for calculation. In this paper, attention is assigned to the prediction step of the model based on the CNN-BiLSTM network.

The calculation steps are as follows: the first step is to calculate the similarity between the output  $h_i$  ( $i = 1, 2, \dots, n$ ) of the BiLSTM network at each time and the output  $h_t$  of the current moment to get the corresponding weight  $s_i$  ( $i = 1, 2, \dots, n$ ). The second step normalizes the weights  $s_i$  using the *soft max* ( $\cdot$ ) function. Obtain the output  $h_t$  at the current moment with the weights  $\alpha_i$  ( $i = 1, 2, \dots, n$ ) of the outputs  $h_i$  at each moment. The third step weights the weight  $\alpha_i$  and the output  $h_i$  of the BiLSTM neural network at each moment to obtain the final output  $c_i$ . The formulas for the attention mechanism are given in Eqs (2.11)–(2.13).

$$s_i = \tanh(W_{h_i} h_t + b_{h_i}) \quad (2.11)$$

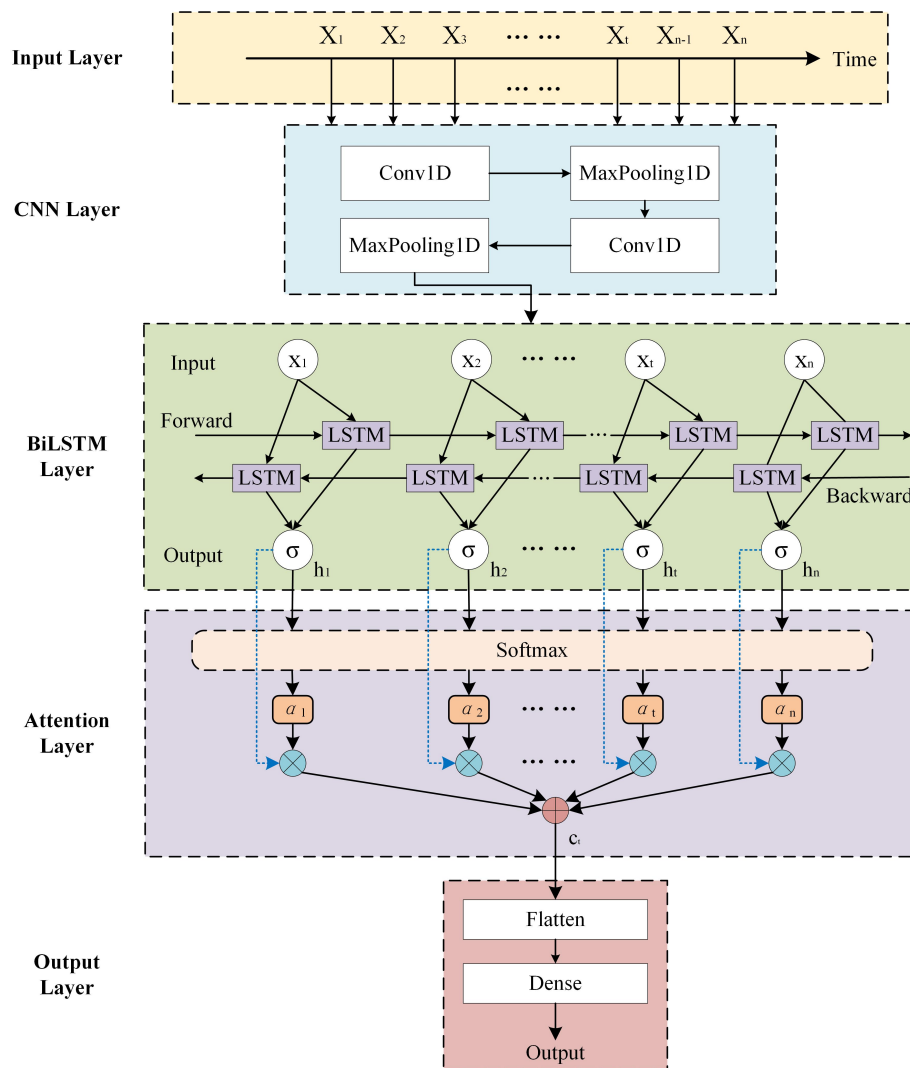
$$\alpha_i = \text{soft max}(s_i) \quad (2.12)$$

$$c_t = \sum_{i=1}^n \alpha_i h_i \quad (2.13)$$

## 2.6. CNN-BiLSTM-AM model

Figure 5 illustrates the CNN-BiLSTM-AM model framework built in this paper. Table 1 collates the details of the main parameters of the CNN-BiLSTM-AM model. In addition to the input and output layers, the model contains a convolutional layer, a BiLSTM layer, and an Attention layer.





**Figure 5.** CNN-BiLSTM-AM model structure framework diagram.

The first step is the data input layer, which uses multi-feature data from aquaculture water environments to construct a supervised learning dataset  $\{(X_i, Y_i) | i = 1, 2, \dots, n\}$ , and constructs the features  $X$  and labels  $Y$  for model training. The second step is the convolution layer, which extracts the features in the sequence using one-dimensional convolution (Conv1D) while adding a pooling (MaxPooling) layer to reduce the complexity of the features to avoid overfitting of the model. Although the convolution and pooling stages of the convolution layer can fully extract the time-series features and enrich the diversity of features, CNN only considers the correlation links between adjacent data in the sequence and does not consider the problem of long-term information dependence of the time series.

In order to remedy this deficiency, the third step of this paper connects the BiLSTM layer to the CNN layer. The forward and reverse LSTM networks in the BiLSTM layer are utilized to fully consider the sequences' past and future information features. Since the BiLSTM network is based on the bi-directional recurrent neural network and the LSTM neural network, the core of the network is still the LSTM, so the feature information of the output of the CNN layer needs to be changed accordingly, as

the tensor form of the data is consistent with the input format [number of samples, prediction time step size, and input feature dimension] required by the LSTM. In the BiLSTM layer, the BiLSTM network firstly traverses the data output from the CNN layer from left to right, and secondly traverses the data from right to left in the reverse direction, and finally stitches the output results from the above two directions together and submits them to the attention layer.

**Table 1.** Parameter setting of the model.

Hyperparameter	Value
Filter size of CNN	[128,64]
Kernel size of CNN	1
Padding	same
Activation function	Relu
Unit number for BiLSTM	128
Optimization function	adam
Learning-rate	0.001
Batch size	64
Epoch number	100

In this paper, we study the multi-dimensional multi-step time series prediction of DO concentration, and different time steps affect the model's accuracy. Therefore, the fourth step of this paper incorporates the AM after the BiLSTM layer, specifically, the soft attention mechanism. The AM is used to capture the importance of different prediction steps in the time series on the impact of the model and improve the model's overall prediction accuracy. The fifth step uses the Flatten Layer to transform the multi-dimensional input data into one dimension. Finally, the fully connected layer is added to obtain the model's output.

### 2.7. Evaluation criteria

In this paper, the mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE) and determination coefficient ( $R^2$ ) are chosen to assess the prediction accuracy of the model. Their calculation methods are shown in Eqs (2.14)–(2.17).

$$MAE = \frac{1}{N} \sum_{k=1}^N |\hat{y}_k - y_k| \quad (2.14)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{k=1}^N (y_k - \hat{y}_k)^2} \quad (2.15)$$

$$MAPE = \frac{1}{N} \sum_{k=1}^N \left| \frac{\hat{y}_k - y_k}{y_k} \right| \times 100\% \quad (2.16)$$

$$R^2 = 1 - \frac{\sum_{k=1}^N (y_k - \hat{y}_k)^2}{\sum_{k=1}^N (y_k - \bar{y})^2} \quad (2.17)$$

where  $N$  is the number of samples in the test set,  $\widehat{y}_k$  is the actual value, and  $y_k$  represents the prediction result. The model has higher prediction accuracy when the MAE and RMSE results are negligible. Conversely, a higher  $R^2$  value represents a better fit of the model on the test set and more accurate prediction results.

### 3. Experiments and analysis of results

#### 3.1. Data preprocessing

Aquaculture enterprises in the daily production and operation process, power outages, poor network signals and water quality sensor failure and other factors can easily lead to the collection of raw water environmental data missing values and abnormal. This paper uses the Lagrangian interpolation method to fill in the missing and abnormal data, calculated as in Eq (3.1). Table 2 collates the statistical information on DO concentrations monitored in aquaculture farms above.

**Table 2.** Statistical characteristics of dissolved oxygen in the aqueous environment.

Cases	Datasets	Numbers	Statistical indicators					
			Mean	Std.	Max.	Min.	Kurtosis	Skewness
Dataset 1	All Samples	20851	8.09	1.20	10.25	5.02	-0.64	-0.62
	Training Set	16681	8.57	0.77	10.25	6.37	-0.37	-0.47
	Validation Set	2085	6.54	0.25	7.09	5.84	-0.53	-0.04
	Testing Set	2085	5.81	0.35	6.44	5.02	-0.88	-0.21
Dataset 2	All Samples	21673	7.28	1.35	10.32	4.12	-0.99	0.46
	Training Set	17339	6.75	0.94	9.08	4.12	-0.69	0.48
	Validation Set	2167	8.99	0.24	9.71	8.49	-0.03	0.97
	Testing Set	2167	9.70	0.22	10.32	9.32	-1.47	0.19

At the same time, the different units of data in different water environments and the possible existence of odd sample data in the original data can negatively affect the model’s training. For example, when gradient descent is performed, the gradient direction tends to deviate from the direction of the minimum value, resulting in a long training time for the model. Therefore, this paper adopts the normalization operation for the preprocessed data, and Eq (3.2) describes its calculation process.

$$L_n(x) = \sum_{k=0}^n y_k \cdot \frac{(x - x_0)(x - x_1) \cdots (x - x_{k-1})(x - x_{k+1}) \cdots (x - x_n)}{(x_k - x_0)(x_k - x_1) \cdots (x_k - x_{k-1})(x_k - x_{k+1}) \cdots (x_k - x_n)} \tag{3.1}$$

$$z' = \frac{z - \min(z)}{\max(z) - \min(z)} \tag{3.2}$$

where  $x_k (k = 0, 1, \dots, n)$  is several independent variables,  $y_k$  is the value of the dependent variable corresponding to the independent variables,  $x$  is the interpolation node, and  $L_n(x)$  is the result after interpolation.  $z$  represents the different water environment parameters in the original data,  $\min(z)$  and

$\max(z)$  represent the minimum and maximum values under the same feature, respectively,  $z'$  representing the normalized data results.

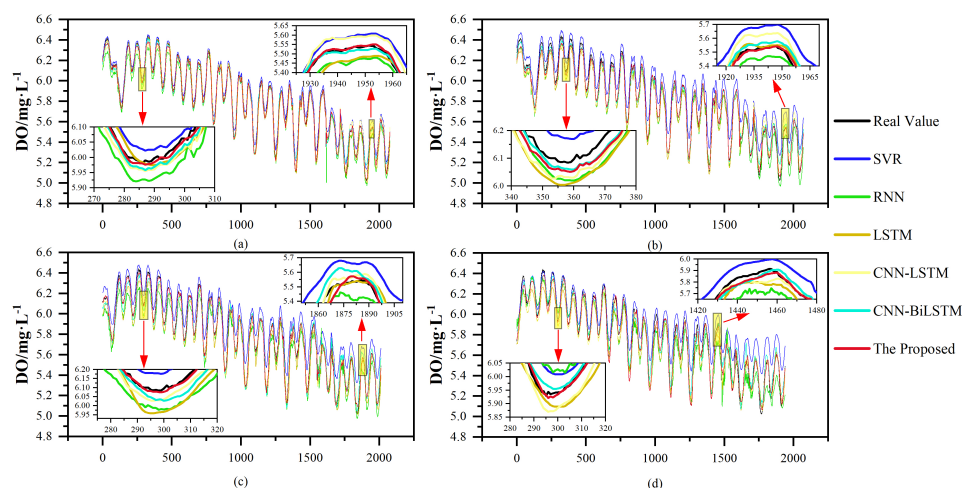
### 3.2. Model development

In the experiments of this paper, several classical prediction models were trained and predicted to compare and evaluate the prediction performance of CNN-BiLSTM-AM models. They are SVR, RNN, LSTM, CNN-LSTM and CNN-BiLSTM models.

Meanwhile, to ensure the fairness of the experiments, we use the same training set, validation set and test set for all the models and choose the same values for the same hyperparameters in these models. The MSE is used as the loss function for model training, and the weights of the models are optimized using the adaptive moment estimation method. The learning rate was 0.001 and the training period was 100 times. All models were done in a Python programming environment and implemented on the Keras framework.

### 3.3. Performance analysis on periodic datasets

In order to visualize the prediction results, Figure 6 shows the short-term prediction fitting curves of different comparison models and the CNN-BiLSTM-AM model proposed in this paper on the test set. The four subplots (a)–(d) represent 3-step prediction, 6-step prediction, 36-step prediction and 72-step prediction, respectively. It can be seen from the Figure 6 that the dissolved oxygen concentration in the water environment of this aquaculture farm shows a cyclic variation, fluctuating up and down in a particular range. Although all models could predict the overall trend of DO correctly, the prediction accuracy of different models varied considerably at the peaks and valleys of different time steps. We can see that the prediction curves of the SVR model deviate the farthest from the actual values at the peaks and valleys, and the rest of the deep learning models and the combined models are closer to the actual values.



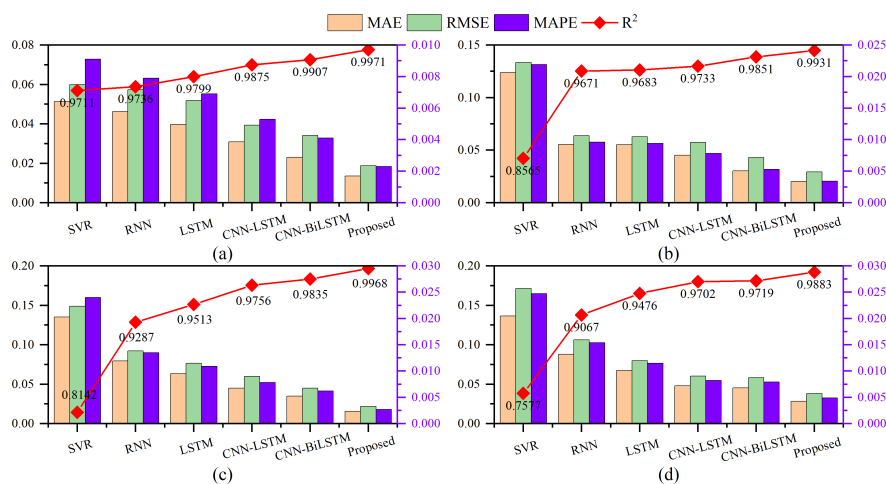
**Figure 6.** Fitted curve of dissolved oxygen prediction for the first data set.

Table 3 collates the short-term prediction results of the CNN-BiLSTM-Attention model proposed

in this paper for dissolved oxygen concentration in an aquaculture water environment, including three steps ahead (30 minutes) and six steps ahead (1 hour). Based on the prediction results of 3 steps ahead (30 minutes), it can be seen that the MAE, RMSE and MAPE of the deep learning RNN model are reduced by 10.12, 4.34 and 13.19%, respectively, compared with the machine learning SVR model. The three metrics of LSTM were reduced by another 13.85, 9.59 and 12.66%, respectively, compared to the RNN model. It is verified that LSTM alleviates the long-term dependence problem of RNN by increasing the threshold mechanism and improving the model's prediction accuracy. Compared with the combined model, none of the three single prediction models has higher prediction accuracy than the combined model at 30 minutes of prediction. It reflects that the combined model can utilize different network structures to improve the overall prediction of short-term forecasts.

**Table 3.** Short-term prediction performance metrics for the first dataset.

Methods	3 steps				6 steps			
	MAE	RMSE	MAPE	$R^2$	MAE	RMSE	MAPE	$R^2$
SVR	0.0514	0.0599	0.0091	0.9711	0.1238	0.1333	0.0219	0.8565
RNN	0.0462	0.0573	0.0079	0.9736	0.0555	0.0638	0.0096	0.9671
LSTM	0.0398	0.0518	0.0069	0.9799	0.0552	0.0627	0.0094	0.9683
CNN-LSTM	0.0309	0.0394	0.0053	0.9875	0.0452	0.0575	0.0078	0.9733
CNN-BiLSTM	0.0231	0.0342	0.0041	0.9907	0.0303	0.0429	0.0053	0.9851
<b>Proposed</b>	<b>0.0135</b>	<b>0.0187</b>	<b>0.0023</b>	<b>0.9971</b>	<b>0.0202</b>	<b>0.0293</b>	<b>0.0034</b>	<b>0.9931</b>



**Figure 7.** Performance comparison of different models on the first dataset.

To verify whether the model's prediction performance is improved after adding AM, a CNN-BiLSTM comparison model is introduced in this paper. The experimental results show that the prediction performance of the CNN-BiLSTM-AM model after adding AM improves by 41.56, 45.32 and 43.9%, respectively, when predicting 30 minutes in the short term. It demonstrates that focusing the

model's attention on the time step can improve the short-term prediction performance of the model. In the prediction results of 6 steps ahead (1 hour), SVR reduced 11.43% in  $R^2$  metric compared to LSTM in deep learning models. The CNN-BiLSTM-AM model proposed in this paper also improved 33.33, 31.7 and 35.85% over the best-performing CNN-BiLSTM model in the combined neural network.

To verify whether the CNN-BiLSTM-AM model proposed in this paper still has excellent prediction performance for long-term prediction, we take 36 steps ahead (6 hours) and 72 steps ahead (12 hours) for multiple time steps of DO concentration prediction. According to Table 4, the prediction accuracy of the SVR model decreases sharply as the prediction time step increases. In the  $R^2$  index, its accuracy decreased by 21.97%, from 97.11% for 3-step prediction to 75.77% for 72-step prediction. RNN and LSTM closely follow this.

**Table 4.** Long-term prediction performance metrics for the first dataset.

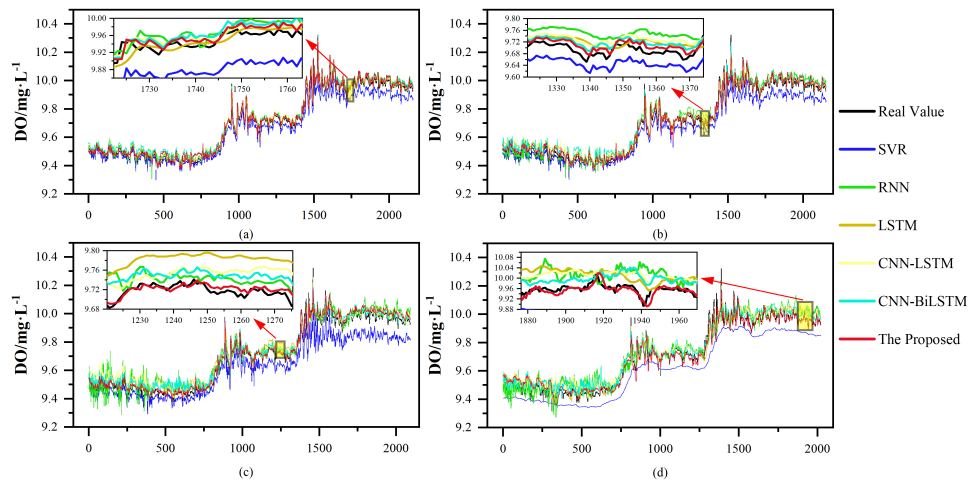
Methods	36 steps				72 steps			
	MAE	RMSE	MAPE	$R^2$	MAE	RMSE	MAPE	$R^2$
SVR	0.1352	0.1490	0.0240	0.8142	0.1365	0.1713	0.0247	0.7577
RNN	0.0794	0.0923	0.0135	0.9287	0.0881	0.1063	0.0154	0.9067
LSTM	0.0633	0.0765	0.0109	0.9513	0.0675	0.0797	0.0115	0.9476
CNN-LSTM	0.0448	0.0599	0.0078	0.9756	0.0479	0.0601	0.0082	0.9702
CNN-BiLSTM	0.0347	0.0451	0.0062	0.9835	0.0452	0.0584	0.0079	0.9719
<b>Proposed</b>	<b>0.0156</b>	<b>0.0219</b>	<b>0.0027</b>	<b>0.9968</b>	<b>0.0283</b>	<b>0.0381</b>	<b>0.0049</b>	<b>0.9883</b>

The 36-step predictions of RNN and LSTM decrease by 3.97 and 1.75%, respectively, in  $R^2$  compared to their 6-step predictions for short-term prediction of DO concentration. Although the accuracy of long-term prediction was reduced, LSTM had higher stability than RNN, verifying that LSTM alleviated the long-term dependence problem by its unique gating mechanism. The combined neural network models all outperformed the single model in long-term prediction. Compared with LSTM, CNN-BiLSTM improved by 45.18, 41.06 and 43.12% in 36-step prediction, and 33.04, 26.73 and 31.3% in 72-step prediction.

Happily, the performance of the CNN-BiLSTM-AM model used in this paper is 55.04, 51.44 and 56.45% higher than that of CNN-BiLSTM without attention mechanism in 36-step prediction. In the 72-step prediction, each index improved by 37.39, 34.76 and 37.97%, respectively. The above results indicate that the CNN-BiLSTM-AM model has more robust prediction performance on longer prediction steps and has higher stability than similar models.

#### 3.4. Performance analysis on non-periodic datasets

Based on the above, this paper explores the generalization ability of CNN-BiLSTM-AM model on different datasets, and we use the water environment data monitored by another farm in Yantai. As shown in Figure 8, it can be seen that the periodicity of this dataset is not apparent, and the overall dissolved oxygen content is higher than that of the first dataset.



**Figure 8.** Fitting curve of dissolved oxygen prediction for the second data set.

**Table 5.** Short-term prediction performance metrics for the second data set.

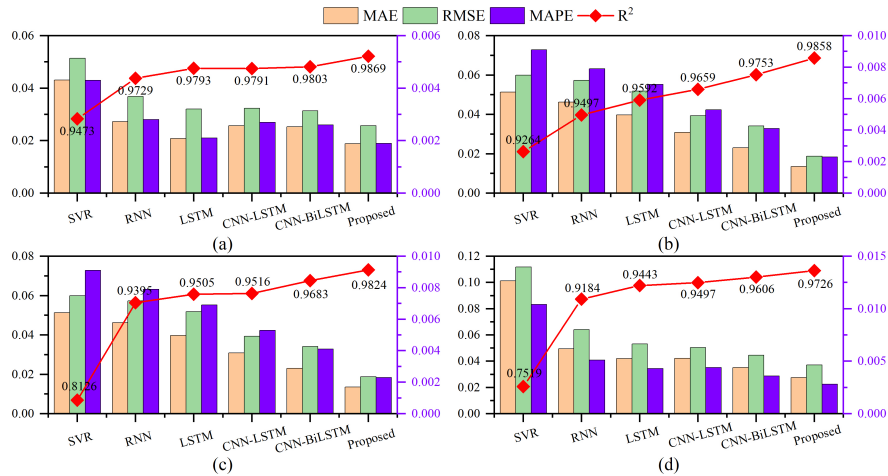
Methods	3 steps				6 steps			
	MAE	RMSE	MAPE	$R^2$	MAE	RMSE	MAPE	$R^2$
SVR	0.0431	0.0514	0.0043	0.9473	0.0518	0.0607	0.0053	0.9264
RNN	0.0273	0.0368	0.0028	0.9729	0.0393	0.0502	0.0040	0.9497
LSTM	0.0208	0.0321	0.0021	0.9793	0.0334	0.0452	0.0034	0.9592
CNN-LSTM	0.0257	0.0323	0.0027	0.9791	0.0329	0.0413	0.0034	0.9659
CNN-BiLSTM	0.0253	0.0314	0.0026	0.9803	0.0272	0.0352	0.0028	0.9753
<b>Proposed</b>	<b>0.0189</b>	<b>0.0257</b>	<b>0.0019</b>	<b>0.9869</b>	<b>0.0201</b>	<b>0.0266</b>	<b>0.0021</b>	<b>0.9858</b>

**Table 6.** Long-term prediction performance metrics for the second data set.

Methods	36 steps				72 steps			
	MAE	RMSE	MAPE	$R^2$	MAE	RMSE	MAPE	$R^2$
SVR	0.0849	0.0971	0.0086	0.8126	0.1013	0.0118	0.0104	0.7519
RNN	0.0436	0.0555	0.0045	0.9395	0.0495	0.0642	0.0051	0.9184
LSTM	0.0409	0.0499	0.0042	0.9505	0.0420	0.0532	0.0043	0.9443
CNN-LSTM	0.0394	0.0494	0.0041	0.9516	0.0422	0.0504	0.0044	0.9497
CNN-BiLSTM	0.0332	0.0402	0.0034	0.9683	0.0352	0.0446	0.0036	0.9606
<b>Proposed</b>	<b>0.0221</b>	<b>0.0298</b>	<b>0.0023</b>	<b>0.9824</b>	<b>0.0276</b>	<b>0.0372</b>	<b>0.0028</b>	<b>0.9726</b>

Tables 5 and 6 summarize the performance metrics of the different models for short-term and long-

term forecasting on this dataset. We can clearly see that the CNN-BiLSTM model with the added attention mechanism has excellent performance in all metrics. In terms of MAE metrics, it improves 25.29%, 26.10%, 33.43% and 21.59% in short-term and long-term predictions, respectively, over the CNN-BiLSTM model without the added attention mechanism. It improved 30.76%, 48.85%, 49.31% and 44.24% in short-term and long-term predictions, respectively, over the RNN model. The growth trend of the prediction steps shows that adding the attention mechanism to the CNN-BiLSTM model in this paper can effectively enhance the long-term prediction ability of the model and has higher prediction accuracy compared with the benchmark model.



**Figure 9.** Performance comparison of different models on the second dataset.

Both Figure 7 and Figure 9 show a combination of bar and line graphs that visualize the prediction accuracy of the different prediction models on the two data sets, respectively. Subplots (a) to (d) in the figures represent 3-step forecasting, 6-step forecasting, 36-step forecasting, and 72-step forecasting, respectively. Starting from the legend, MAE and RMSE correspond to the orange and green bars, whose values correspond to the left y-axis. MAPE corresponds to the purple beads, whose values correspond to the right y-axis. The red line graphs represent the coefficients of determination of the indicators, and higher values represent better fitting results of the model.

In summary, we can conclude that both short-term prediction (30 minutes, 1 hour) and long-term prediction (6 hours, 12 hours). The CNN-BiLSTM-AM model proposed in this paper has excellent prediction performance and better stability in long-term prediction.

#### 4. Conclusions

This paper proposed a multi-step prediction model of dissolved oxygen concentration based on the combination of attention mechanisms and a combined neural network to improve the prediction accuracy of dissolved oxygen concentration in an aquaculture water environment. The model incorporates an attention mechanism in the BiLSTM network to focus the model’s attention on the time step of prediction, which enhances the long-term prediction ability of the model. By comparing the SVR, RNN, LSTM, CNN-LSTM and CNN-BiLSTM prediction models, the CNN-BiLSTM-AM model established



in this paper has higher prediction accuracy. As the time step increases, the model's superiority becomes more evident.

In the study process, this paper only used four water quality parameters as the input of the model. Future studies can consider on-farm weather factors and more water quality factors for the model.

## Acknowledgments

This work was supported by the Yantai Science and Technology Innovation Development Plan Project (Grant No. 2022XDRH015).

## Conflict of interest

The authors declare that there is no conflict of interests regarding the publication of this paper.

## References

1. Agriculture Organization of the United Nations, Fisheries Department, *The State of World Fisheries and Aquaculture*, Food & Agriculture Org, 2000.
2. X. Li, J. Li, Y. Wang, L. Fu, Y. Fu, B. Li, et al., Aquaculture industry in china: Current state, challenges, and outlook, *Rev. Fish. Sci.*, **19** (2011), 187–200. <https://doi.org/10.1080/10641262.2011.573597>
3. J. Huan, H. Li, M. Li, B. Chen, 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** (2020), 105530. <https://doi.org/10.1016/j.compag.2020.105530>
4. S. Midilli, D. Çoban, M. Güler, S. Küçük, Gas bubble disease in Nile tilapia and hybrid red tilapia (Cichlidae, *Oreochromis* spp.) under culture conditions, *J. Fish. Aquat. Sci.*, **36** (2019), 285–291. <http://dx.doi.org/10.12714/egejfas.2019.36.3.09>
5. C. E. Boyd, E. L. Torrans, C. S. Tucker, Dissolved oxygen and aeration in ictalurid catfish aquaculture, *J. World Aquacult. Soc.*, **49** (2018), 7–70. <https://doi.org/10.1111/jwas.12469>
6. A. Sentas, L. Karamoutsou, N. Charizopoulos, T. Psilovikos, A. Psilovikos, A. Loukas, The use of stochastic models for short-term prediction of water parameters of the thesaurus dam, River Nestos, Greece, *Proceedings*, **2** (2018), 634. <https://doi.org/10.3390/proceedings2110634>
7. M. Valera, R. K. Walter, B. A. Bailey, J. E. Castillo, Machine learning based predictions of dissolved oxygen in a small coastal embayment, *J. Mar. Sci. Eng.*, **8** (2020), 1007. <https://doi.org/10.3390/jmse8121007>
8. C. Xu, X. Chen, L. Zhang, Predicting river dissolved oxygen time series based on stand-alone models and hybrid wavelet-based models, *J. Environ. Manage.*, **295** (2021), 113085. <https://doi.org/10.1016/j.jenvman.2021.113085>
9. A. Sorjamaa, J. Hao, N. Reyhani, Y. Ji, A. Lendasse, Methodology for long-term prediction of time series, *Neurocomputing*, **70** (2007), 2861–2869. <https://doi.org/10.1016/j.neucom.2006.06.015>

10. M. Långkvist, L. Karlsson, A. Loutfi, A review of unsupervised feature learning and deep learning for time-series modeling, *Pattern Recognit. Lett.*, **42** (2014), 11–24. <https://doi.org/10.1016/j.patrec.2014.01.008>
11. L. Karamoutsou, A. Psilovikos, Deep learning in water resources management: The case study of Kastoria lake in Greece, *Water*, **13** (2021), 3364. <https://doi.org/10.3390/w13233364>
12. Q. Ye, X. Yang, C. Chen, J. Wang, River water quality parameters prediction method based on LSTM-RNN model, in *2019 Chinese Control And Decision Conference (CCDC)*, (2019), 3024–3028. <https://doi.org/10.1109/CCDC.2019.8832885>
13. J. Yan, Y. Gao, Y. Yu, H. Xu, Z. Xu, A prediction model based on deep belief network and least squares svr applied to cross-section water quality, *Water*, **12** (2020), 1929. <https://doi.org/10.3390/w12071929>
14. L. Sheng, J. Zhou, X. Li, Y. Pan, L. Liu, Water quality prediction method based on preferred classification, *IET Cyber-Phys. Syst. Theory Appl.*, **5** (2020), 176–180. <https://doi.org/10.1049/iet-cps.2019.0062>
15. J. Wu, Z. Wang, A hybrid model for water quality prediction based on an artificial neural network, wavelet transform, and long short-term memory, *Water*, **14** (2022), 610. <https://doi.org/10.3390/w14040610>
16. B. Lim, S. Zohren, Time-series forecasting with deep learning: A survey, *Philos. Trans. A Math. Phys. Eng. Sci.*, **379** (2021), 20200209. <https://doi.org/10.1098/rsta.2020.0209>
17. H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, P. A. Muller, Deep learning for time series classification: A review, *Data Min. Knowl. Discovery*, **33** (2019), 917–963. <https://doi.org/10.1007/s10618-019-00619-1>
18. P. Shi, G. Li, Y. Yuan, G. Huang, L. Kuang, Prediction of dissolved oxygen content in aquaculture using clustering-based softplus extreme learning machine, *Comput. Electron. Agric.*, **157** (2019), 329–338. <https://doi.org/10.1016/j.compag.2019.01.004>
19. W. Li, H. Wu, N. Zhu, Y. Jiang, J. Tan, Y. Guo, Prediction of dissolved oxygen in a fishery pond based on gated recurrent unit (GRU), *Inf. Process. Agric.*, **8** (2021), 185–193. <https://doi.org/10.1016/j.inpa.2020.02.002>
20. C. Li, Z. Li, J. Wu, L. Zhu, J. Yue, A hybrid model for dissolved oxygen prediction in aquaculture based on multi-scale features, *Inf. Process. Agric.*, **5** (2018), 11–20. <https://doi.org/10.1016/j.inpa.2017.11.002>
21. Q. Ren, X. Wang, W. Li, Y. Wei, D. An, Research of dissolved oxygen prediction in recirculating aquaculture systems based on deep belief network, *Aquacult. Eng.*, **90** (2020), 102085.
22. J. Huang, S. Liu, S. G. Hassan, L. Xu, C. Huang, A hybrid model for short-term dissolved oxygen content prediction, *Comput. Electron. Agric.*, **186** (2021), 106216. <https://doi.org/10.1016/j.compag.2021.106216>
23. H. Liu, R. Yang, Z. Duan, Wind speed forecasting using a new multi-factor fusion and multi-resolution ensemble model with real-time decomposition and adaptive error correction, *Energy Convers. Manage.*, **217** (2020), 112995. <https://doi.org/10.1016/j.enconman.2020.112995>

24. J. Bi, Y. Lin, Q. Dong, H. Yuan, M. Zhou, An improved attention-based lstm for multi-step dissolved oxygen prediction in water environment, in *2020 IEEE International Conference on Networking, Sensing and Control (ICNSC)*, (2020), 1–6. <https://doi.org/10.1109/ICNSC48988.2020.9238097>
25. Y. Liu, Q. Zhang, L. Song, Y. Chen, Attention-based recurrent neural networks for accurate short-term and long-term dissolved oxygen prediction, *Comput. Electron. Agric.*, **165** (2019), 104964. <https://doi.org/10.1016/j.compag.2019.104964>
26. X. Yang, B. Liu, Uncertain time series analysis with imprecise observations, *Fuzzy Optim. Decis. Making*, **18** (2018), 263–278. <https://doi.org/10.1007/s10700-018-9298-z>
27. T. Wang, M. Wang, Communication network time series prediction algorithm based on big data method, *Wireless Pers. Commun.*, **102** (2017), 1041–1056. <https://doi.org/10.1007/s11277-017-5138-7>
28. Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, et al., Back-propagation applied to handwritten zip code recognition, *Neural Comput.*, **1** (1989), 541–551. <https://doi.org/10.1162/neco.1989.1.4.541>
29. Á. Zarándy, C. Rekeczky, P. Szolgay, L. O. Chua, Overview of cnn research: 25 years history and the current trends, in *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*, (2015), 401–404. <https://doi.org/10.1109/ISCAS.2015.7168655>
30. L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, et al., Review of deep learning: concepts, cnn architectures, challenges, applications, future directions, *J. Big Data*, **8** (2021), 53. <https://doi.org/10.1186/s40537-021-00444-8>
31. M. Sahu, R. Dash, A survey on deep learning: convolution neural network (CNN), in *Intelligent and Cloud Computing*, Springer, (2021), 317–325.
32. G. Ortac, G. Ozcan, Comparative study of hyperspectral image classification by multidimensional convolutional neural network approaches to improve accuracy, *Expert Syst. Appl.*, **182** (2021), 115280. <https://doi.org/10.1016/j.eswa.2021.115280>
33. X. Song, Y. Liu, L. Xue, J. Wang, J. Zhang, J. Wang, et al., Time-series well performance prediction based on long short-term memory (LSTM) neural network model, *J. Pet. Sci. Eng.*, **186** (2020), 106682. <https://doi.org/10.1016/j.petrol.2019.106682>
34. S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural Comput.*, **9** (1997), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
35. R. Yu, J. Gao, M. Yu, W. Lu, T. Xu, M. Zhao, et al., LSTM-EFG for wind power forecasting based on sequential correlation features, *Future Gener. Comput. Syst.*, **93** (2019), 33–42. <https://doi.org/10.1016/j.future.2018.09.054>
36. X. Wu, J. Li, Y. Jin, S. Zheng, Modeling and analysis of tool wear prediction based on SVD and BiLSTM, *Int. J. Adv. Manuf. Technol.*, **106** (2020), 4391–4399. <https://doi.org/10.1007/s00170-019-04916-3>
37. M. Schuster, K. K. Paliwal, Bidirectional recurrent neural networks, *IEEE Trans. Signal Process.*, **45** (1997), 2673–2681. <https://doi.org/10.1109/78.650093>

38. S. Siami-Namini, N. Tavakoli, A. S. Namin, The performance of LSTM and BiLSTM in forecasting time series, in *2019 IEEE International Conference on Big Data (Big Data)*, (2019), 3285–3292. <https://doi.org/10.1109/BigData47090.2019.9005997>
39. Z. Niu, G. Zhong, H. Yu, A review on the attention mechanism of deep learning, *Neurocomputing*, **452** (2021), 48–62. <https://doi.org/10.1016/j.neucom.2021.03.091>
40. A. De Santana Correia, E. L. Colombini, Attention, please! A survey of neural attention models in deep learning, *Artif. Intell. Rev.*, **2022** (2022), 1–88. <https://doi.org/10.1007/s10462-022-10148-x>



AIMS Press

©2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)