



# LSTM-TCN: dissolved oxygen prediction in aquaculture, based on combined model of long short-term memory network and temporal convolutional network

Wenshu Li<sup>1,2,3,4</sup> · Yaoguang Wei<sup>1,2,3,4</sup> · Dong An<sup>1,2,3,4</sup> · Yisha Jiao<sup>1,2,3,4</sup> · Qiong Wei<sup>1,2,3,4</sup>

Received: 18 June 2021 / Accepted: 24 January 2022 / Published online: 1 February 2022  
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

## Abstract

Dissolved oxygen (DO) is an important water quality monitoring parameter of great significance in aquaculture. Accurate prediction of dissolved oxygen can help farmers to take necessary measures in advance to ensure the healthy growth of cultured species. The characteristics of multivariate and long-term correlation of water quality time series in the traditional methods make it difficult to achieve the expected prediction accuracy. To solve this problem, we propose the combined prediction method LSTM-TCN (long short-term memory network and temporal convolutional network). After the preprocessing of time series, the LSTM extracts the features of the series in time dimension, and then combines with TCN to build the fusion prediction model. In this study, we have carried out the DO predictions of LSTM and TCN algorithms separately, followed by the analysis of DO prediction, based on CNN-LSTM and LSTM-TCN combined models. The effects of attention mechanism and window size of historical time on the prediction results were also investigated. The experimental results show that the combined method has high accuracy in dissolved oxygen prediction, and can capture better characteristics of historical data with increasing time window of the historical dissolved oxygen sequence. The LSTM-TCN method achieves better prediction performance, with evaluation index values of MAE = 0.236, MAPE = 3.10%, RMSE = 0.342, and  $R^2 = 0.94$ .

**Keywords** Dissolved oxygen prediction · Long short-term memory network · Temporal convolutional network · Time window · Attention mechanism · Combined model

## Introduction

The future direction of rapidly developing aquaculture focuses on intensive and intelligent solutions. In the complex, dynamic, and nonlinear aquaculture systems, the quality of water is affected by many factors. Human management alone is not enough to respond to the rapid change of water quality in time that restricts the sustainable development of aquaculture. As one of the key indicators of water quality, the concentration of dissolved oxygen affects the healthy growth of cultured species directly. Keeping the dissolved oxygen content in a proper range improves feed efficiency, reduces the stress response of aquaculture species, and ensures the economic benefits of rearing. Therefore, it is of great practical significance to establish a prediction model of dissolved oxygen, based on historical data, and to predict the trend of dissolved oxygen content accurately.

There are many water quality variables in aquaculture, while the water quality information exists in the form of multivariate time series datasets. Water quality changes

---

Responsible Editor: Marcus Schulz

---

✉ Yaoguang Wei  
wyg@cau.edu.cn

- <sup>1</sup> National Innovation Center for Digital Fishery, China Agricultural University, Beijing 100083, China
- <sup>2</sup> Engineering and Technology Research Center for Internet of Things in Agriculture, China Agricultural University, Beijing 100083, China
- <sup>3</sup> Precision Agricultural Technology Integration Research Base (Fishery), Ministry of Agriculture and Rural Affairs, Beijing 100083, China
- <sup>4</sup> College of Information and Electrical Engineering, China Agricultural University, Haidian District, Tsinghua East Road 17#, Beijing 100083, China

periodically in time. Many scholars carried out relevant exploration and in-depth research to solve the problem of water quality prediction worldwide. The common methods for water quality prediction include artificial neural networks (ANN), regression analyses (RA), grey model (GM), and support vector regressions (SVR) (Zhou et al. 2018), as well as other methods (Rahman et al. 2020). Zhu et al. (2017) used the model of least squares support vector regression (LSSVR) with fruit fly optimization algorithm (FOA) to predict dissolved oxygen. Girija and Mahanta (2010) applied and compared artificial neural network and Mamdani's fuzzy logic control for the prediction of dissolved oxygen in an effluent-impacted urban river. The results showed that the neural network had better prediction performance. Xiao et al. (2017) used the traditional method to predict the dissolved oxygen in the pond and compared the algorithms according to the prediction accuracy. The results showed that the neural network was better than auto-regression (AR), grey model (GM), support vector machines (SVM), and curve fitting (CF). Ji et al. (2017) used support vector machine (SVM) to predict the dissolved oxygen concentration in a hypoxic river in South-Eastern China, and this method provided the optimal performance. The traditional models have some shortcomings, such as local optimal solutions, poor stability, and generalization ability. The self-learning characteristics of neural networks and their ability to process nonlinear information make up for the shortcomings of traditional dissolved oxygen time series models. Accordingly, they are widely used for dissolved oxygen prediction.

In recent years, deep learning method has made significant progress in improving prediction accuracy in time series prediction (Qin 2019; Dabrowski et al. 2020). For the prediction of dissolved oxygen in aquaculture, the long short-term memory (LSTM) has good performance in deep learning models (Liu et al. 2019; Dabrowski et al. 2018). LSTM can memorize the previous data; thus, it has some advantages in dealing with time series prediction. It is widely used in machine translation (Cui et al. 2015; Wang et al. 2016), natural language processing (Yao and Huang 2016; Yao and Guan 2018), time series prediction (Shi et al. 2015; Li et al. 2019), and other fields. Chen et al. (2018) proposed a prediction model of dissolved oxygen in aquaculture based on principal component analysis (PCA) and long short-term memory (LSTM) neural network, which has better prediction performance and generalization ability than traditional prediction methods. Hu et al. (2019a, b) proposed a water quality prediction method, based on the deep LSTM learning network for maricultural environment. Yang et al. (2018) established an end-to-end and trainable LSTM neural network model, which combined temporal and spatial information to realize the prediction of sea surface temperature. Its improvement and combination with other algorithms are more and more widely used in water quality prediction, and

the experimental results show that the combined model has more reliable performance and higher prediction accuracy, than the single one. Accordingly, Li et al. (2018) proposed a hybrid model based on sparse auto-encoder (SAE) and long short-term memory (LSTM) network. Barzegar et al. (2020) established a hybrid model of long short-term memory (LSTM) and convolutional neural network (CNN). Yuan et al. (2018) investigated the accuracy of hybrid long short-term memory neural network and ant lion optimizer model (LSTM-ALO) in prediction of monthly runoff.

Recently, temporal convolutional network has been proposed to develop further time series prediction. The results show that convolution architectures can outperform recurrent networks for various datasets and tasks, and the temporal convolutional network has better performance in the sequence modeling task (Bai et al. 2018). Ta and Wei (2018) proposed a simplified reverse understanding convolutional neural network (CNN) to predict dissolved oxygen, which had faster convergence rate and better stability, than BP network. Zhao et al. (2019) proposed a new method combining graph convolutional network (GCN) and gated recurrent unit (GRU), which can capture the spatial and temporal dependences to achieve accurate and real-time traffic prediction. Deng et al. (2019) proposed a novel knowledge-driven temporal convolutional network (KDTCN) to predict the future trend of stocks.

Related experiments show that RNNs and CNNs can simulate complex nonlinear feature interactions and have good performance for time series processing (Chen et al. 2019). How to utilize the historical data of water quality time series effectively is still a challenge.

In this paper, a dissolved oxygen prediction model, based on the combination of long short-term memory network and temporal convolutional network, is proposed. Long short-term memory network has the ability to deal with long-term dependence of complex time series. Temporal convolutional network solves the problem of concurrency in the LSTM network and improves the flexibility of model structure. The structure of dilations and causal convolutions in the network make the model suitable for time series data of large receiving field with advanced calculation. After preprocessing, the water quality data pass through the LSTM and TCN network model stepwise, and finally the tool calculates the prediction results. In parallel, the comparative experiments of LSTM, CNN-LSTM, and TCN algorithms are carried out. The influence of attention mechanism on prediction and the change of prediction accuracy caused by the size of historical time window are discussed. The experimental results demonstrate that the combined LSTM-TCN has better dissolved oxygen prediction performance, than other algorithms.

The rest of the paper is organized as follows: The basic principles of LSTM and TCN and the overall framework of our proposed combined method are summarized in the

“Methodology” section. “Experiments” presents the application and results of the model to the prediction of dissolved oxygen in the actual industrial aquaculture, comparing it with other algorithms. Finally, the “Conclusion” section summarizes the findings of the work.

### Methodology

This section introduces the basic principles of LSTM and TCN, and shows the framework and overview of the proposed model. Next, the combined algorithm of LSTM-TCN is introduced in detail.

#### Long short-term memory

Recurrent neural network (RNN) has a network structure in which the nodes of hidden layer are connected with each other along multiple times. Therefore, the network can generate the memory state of historical data, can establish the dependence relationship between the input data at different times, and is applicable to the current output calculation. Based on this feature, RNN is suitable for mining time series information (Hu et al. 2019a, b). The network structure is shown in Fig. 1, illustrating how RNN expands by time series. RNN network architecture includes only three parts, namely input layer, hidden layer, and output layer. For a given time series  $x = (x_1, x_2, \dots, x_n)$  the expected output series  $y = (y_1, y_2, \dots, y_n)$  can be obtained by calculation. For example, at time  $t$  we apply the calculation formulas, as follows.

$$h^t = f(W_{xh}x^t + W_{hh}h^{t-1} + d_h) \tag{1}$$

$$y^t = W_{hy}h^t + d_y \tag{2}$$

Here  $W_{xh}$  is the weight coefficient matrix from the input layer to the hidden layer;  $W_{hh}$  is the weight coefficient matrix for the hidden layer;  $W_{hy}$  is the weight coefficient matrix from the hidden layer to the output layer;  $d_h$  and  $d_y$  are the bias vectors of hidden layer and output layer respectively;  $f$  is the nonlinear activation function.

With increasing input time series, RNN has a long-term dependence problem, and will gradually forget the input information appeared a long time ago. In gradient back-propagation, a gradient explosion appears in the training process. To solve this problem, LSTM algorithm was first proposed by Hochreiter and Schmidhuber (1997). As an improved RNN, the gating mechanism is added to improve the phenomenon of gradient disappearance. Compared with the RNN algorithm, LSTM has better performance in longer time series. The input of the current hidden layer consists of the output of the previous hidden layer and the input layer of the current time.

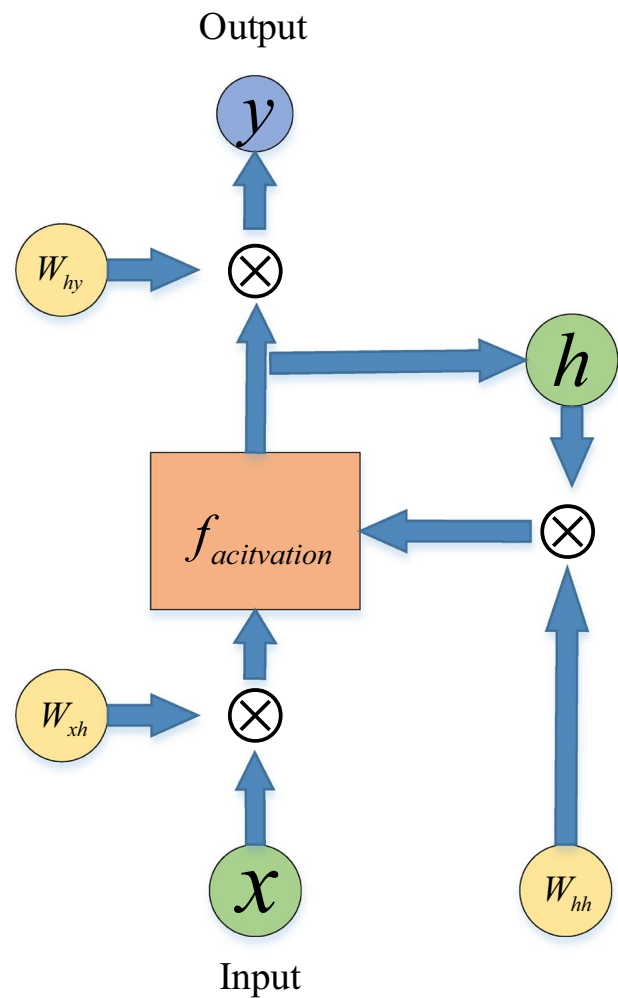


Fig. 1 Network structure diagram of RNN unit

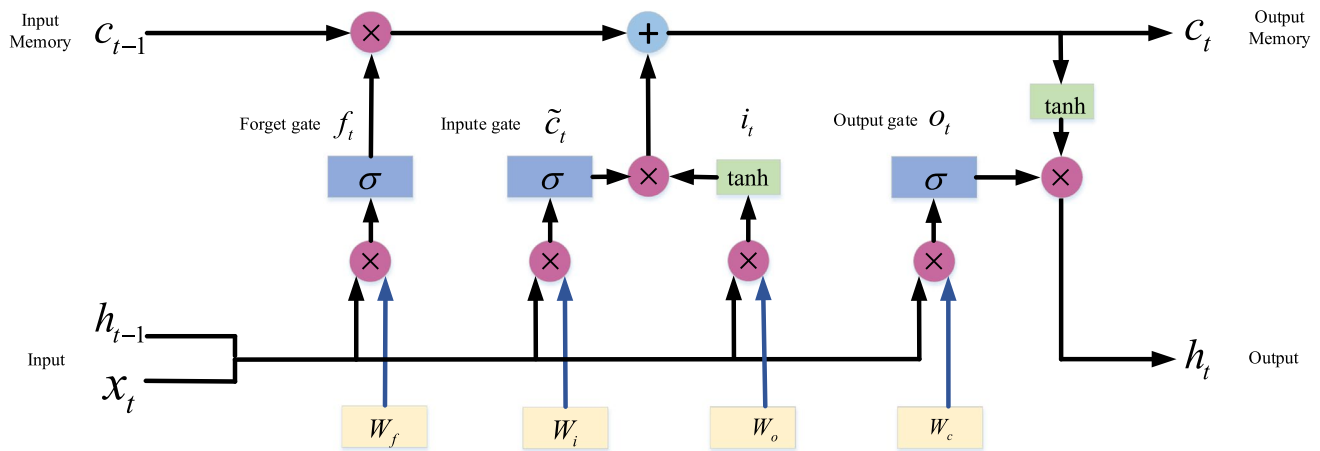
Figure 2 shows the structure of LSTM hidden layer neurons, illustrating the calculation process of current hidden layer neurons. Each hidden layer neuron contains input gate, forget gate, output gate, and current state.

Actually, the input gate determines how much input of the current time network is saved into the unit state, i.e., how much information is updated into the neuron. The calculation formula is as follows:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{3}$$

The forgetting gate determines which information is discarded from the cell state, i.e., how much information of the last cell state  $c_{t-1}$  is retained in the current time  $c_t$ . The calculation formula is as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{4}$$



**Fig. 2** Structure diagram of LSTM hidden layer neurons

The output gate controls the output at the current time  $c_t$  and determines the final output information. The calculation formula is as follows:

$$\begin{aligned} o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\ h_t &= o_t * \tanh(c_t) \end{aligned} \quad (5)$$

The calculation formula of cell state  $c_t$  is the following:

$$\begin{aligned} \tilde{c}_t &= \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \\ c_t &= f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t \end{aligned} \quad (6)$$

Here,  $\sigma$  is the activation function;  $W_f$ ,  $W_i$ ,  $W_c$  and  $W_o$  are the weight matrices of the forget gate, update gate, and output gate, while  $b_f$ ,  $b_i$ ,  $b_c$ , and  $b_o$  are the corresponding bias vectors.

The memory unit structure is added to the LSTM to forget and remember historical data by combining with the gating unit. It can effectively deal with the complex long-term time series dynamic dependence, so it is suitable for the prediction of dissolved oxygen in aquaculture.

## Temporal convolutional network

As a variant of the convolutional neural network (CNN), temporal convolutional network (TCN) performs better in time series prediction than the recurrent neural network (RNN) and long short-term memory (LSTM) network. It is more suitable for processing sequential data with large receptive fields and temporality (Yan et al. 2020), as it employs causal convolutions and dilations.

The model of temporal convolutional network (TCN) (Bai et al. 2018) adopts convolution form. Its network architecture is mainly composed of fully convolutional network (FCN),

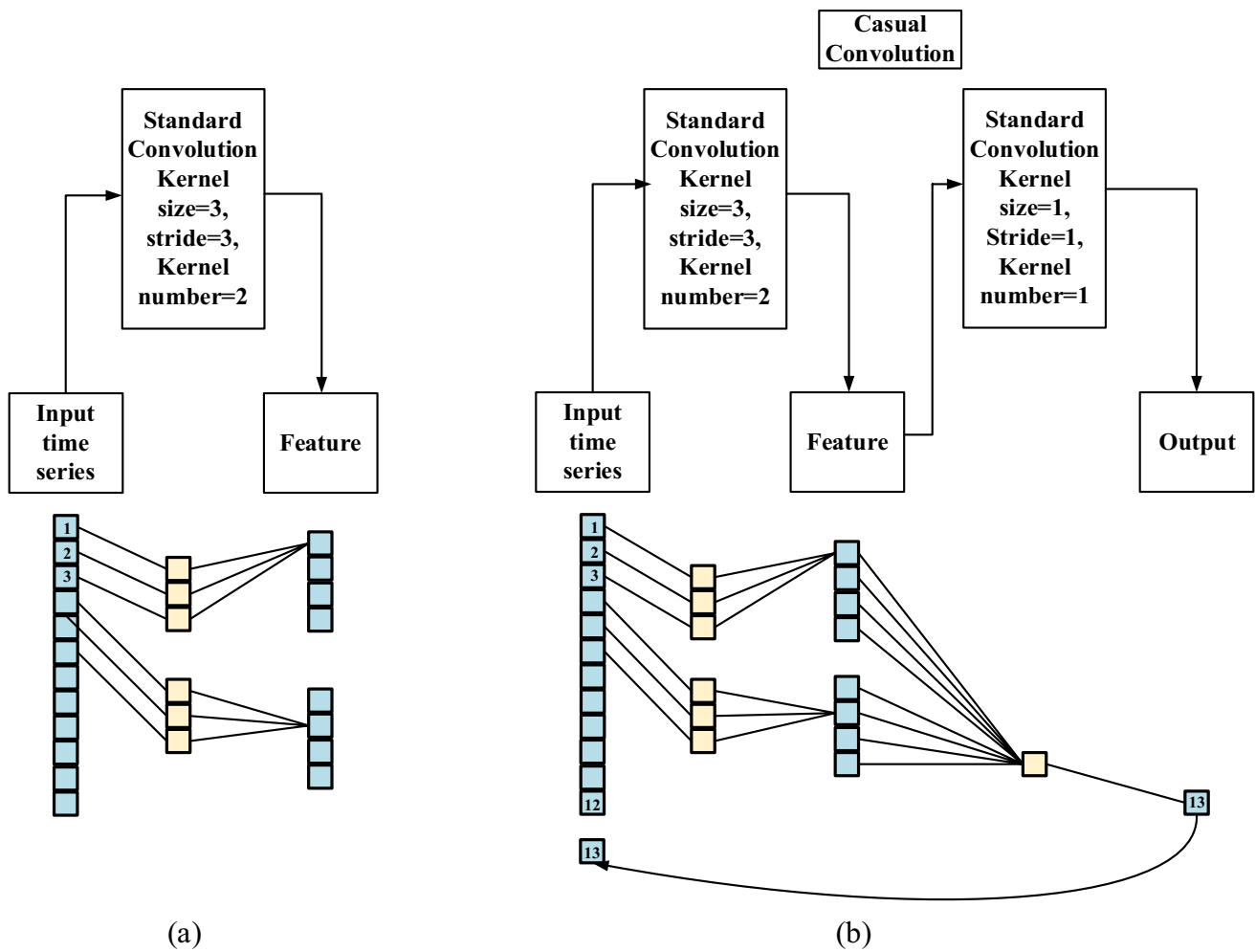
causal convolution, dilated convolutions, and residual block. Corresponding to each module, the network has the following characteristics:

1. The fully convolutional network can guarantee the same output and input sequence length;
2. The network architecture of causal convolution makes no information “leakage” from future to past;
3. The dilated convolutions and residual layer modules are used to build a deep network model, and accordingly, more historical information can be obtained for prediction.

Figure 3 is a comparison of standard convolution (left) and causal convolution (right). Different from the traditional convolutional neural network, for the prediction of time  $t$ , only the observed sequence  $(x_1, x_2, \dots, x_{t-1}, x_t)$  will be used instead of the future information in TCN.

Figure 4 adds the dilated convolution to the causal convolution, which is the dilated causal convolution with dilation factors  $d=1, 2, 4$  and filter size  $k=3$  in TCN. The use of dilated convolution makes it necessary to use padding in the TCN network to ensure that the input and output length of hidden layer and input layer are equal, while the padding size is  $(k-1) \times d$ . Actually, causal convolution ensures that the prediction at time  $t$  will not use future information, because the output of time  $t$  will only be obtained by convolution of time  $t$  and the elements from previous layers. For the sequence tasks that need to acquire a long history, the dilated convolution is used to increase the receptive field of the network. There are two ways to increase the receptive field, namely.

1. To increase the dilation factor  $d$ ;
2. To select larger filter sizes  $k$ .



**Fig. 3** Standard convolution and causal convolution with kernel size 3. **a** For standard convolution, the kernel size and stride are 3, and the kernel number is 2. **b** For causal convolution, two standard convolution layers are composed. Each figure is described in **b**

The calculation formula is as follows:

$$f(s) = (x \times_d f)(s) = \sum_{i=0}^{k-1} f(i) \cdot x_{s-d \times i} \tag{7}$$

where  $x$  is the one-dimensional sequence input and  $x \in R^n$ ,  $f$  is the filter and,  $f : \{0, \dots, k - 1\} \rightarrow R$   $s$  is the element in the sequence, and  $F$  is the dilated convolution operation.

With the increase of the length of time series, TCN needs a larger receptive field to receive high-dimensional time series. The increase of network depth leads to network degradation, increasing the consumption of computing resources, gradient explosion, and gradient disappearance. The solution of this problem is to add residual module into the network. The residual network consists of a series of residual blocks, each of them consists of two parts: direct mapping part and residual part. Adding jump links to realize residual learning and using identity

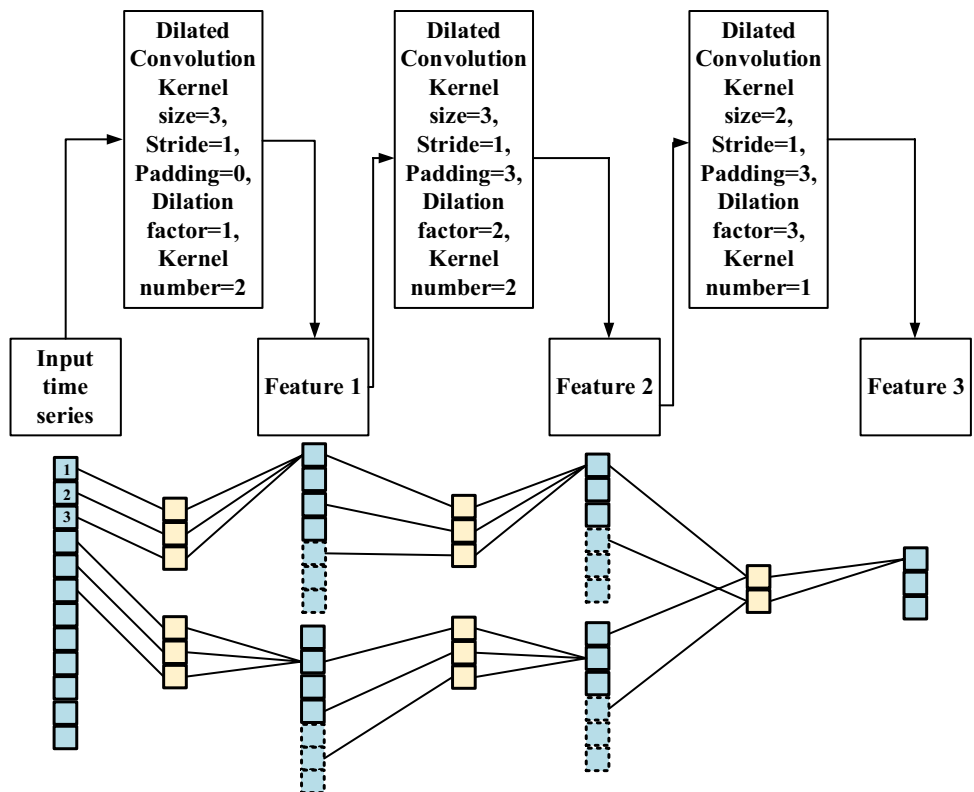
map as shortcut link reduce the complexity of residual network, making the deep network easier to be trained and optimized. It can be expressed as follows:

$$x_{1+1} = h(x_1) + F(x_1, W_1) \tag{8}$$

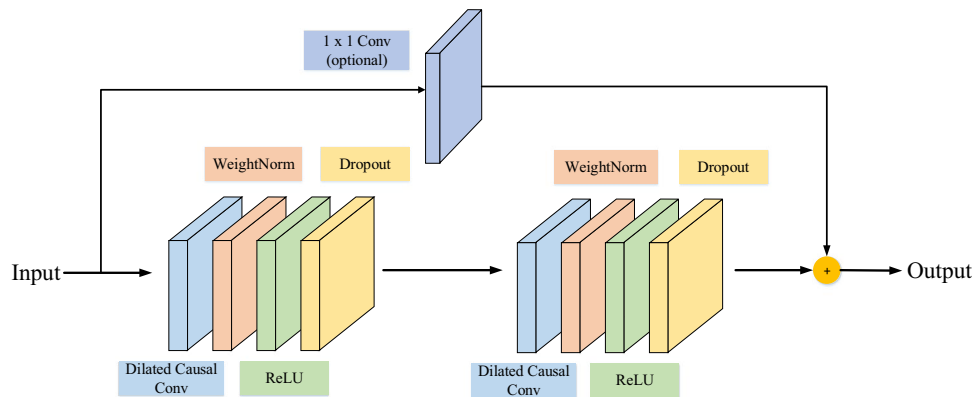
where  $F(x_1, W_1)$  is the residual part, and  $W_1$  is the weight matrix.  $h(x_1)$  is the direct mapping part, including  $1 \times 1$  convolution operation.  $x_1$  is the input of the residual block, and  $x_{1+1}$  is the output of the residual block.

Figure 5 shows the residual block structure in TCN. The residual block has two layers, including dilated causal convolution, weight normalization, non-linearity (ReLU), and dropout layer for regularization. If the input and output of residual have different dimensions, then it is usually solved by adding a  $1 \times 1$  convolution. The residual network adds the identity map of cross layer connection. Instead of modifying the entire transformation, it allows the layer to modify the identity mapping. It is beneficial for TCN to

**Fig. 4** Dilated causal convolution with dilation factors 1, 2, and 3 for three convolution layers



**Fig. 5** The structure of TCN residual block



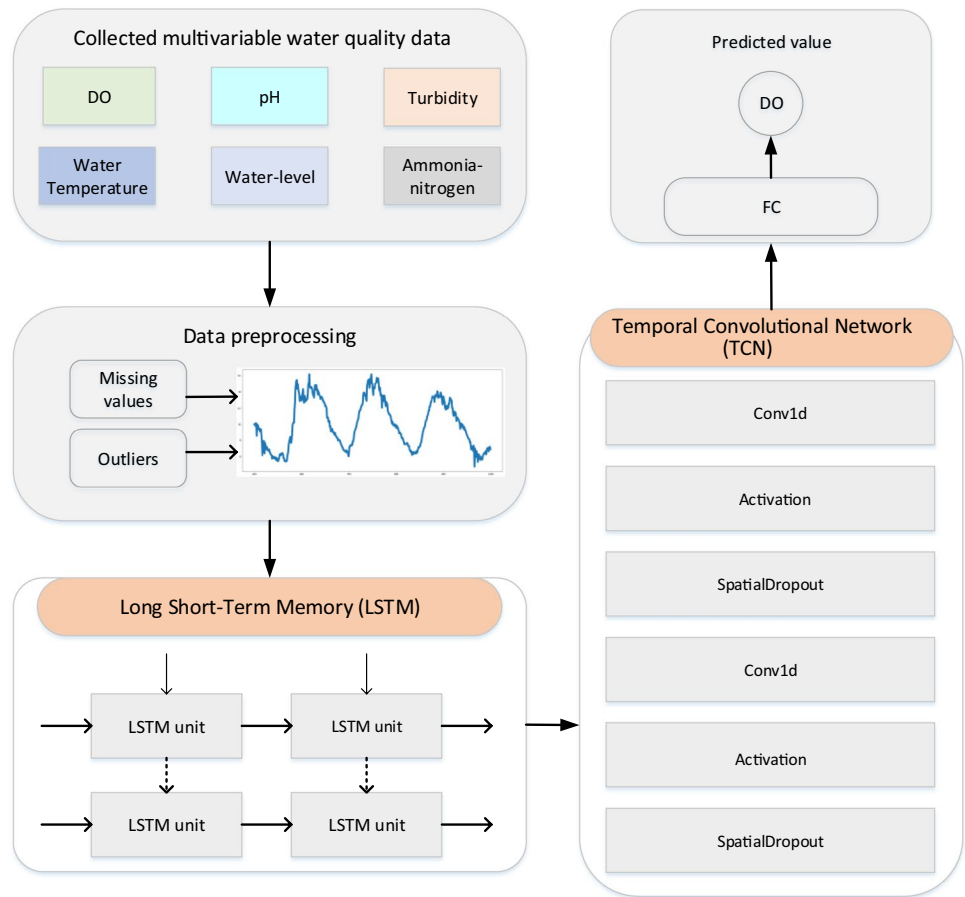
build a deeper network structure that increases the stability of the TCN network.

**The combined model (LSTM-TCN)**

Figure 6 shows the main network architecture of our proposed LSTM-TCN prediction model. After data preprocessing, the multivariable water quality data enter the LSTM network through the input layer first to extract the correlation of datasets in time dimension. Then, the time series prediction model is established by combining with TCN network. Finally, the prediction DO value goes through the full connection layer to realize the prediction of dissolved

oxygen in the future. Long short-term memory (LSTM) network solves the long-term dependence problem of recurrent neural network that can capture the short-term and long-term dependence, and can be widely applied for time series problems. As a new algorithm for time series prediction, temporal convolutional network (TCN) has been proved to be better than RNN in some datasets. It can deal with large-scale data and with a larger sensing field, which enables it to retain the information memory of a longer time ago. By comparison, it uses less memory when processing the same long sequence. Therefore, based on the unique advantages of the two algorithms, we propose a fusion prediction model LSTM-TCN that combines the two algorithms.

**Fig. 6** Network architecture of LSTM-TCN



## Experiments

### Dataset and data preprocessing

The dataset used in this paper is the real water quality data collected from industrial recirculating aquaculture workshop by using multi-parameter water quality sensor. The sampling time is from January 1, 2019, to July 9, 2019, and the data is collected in every 10 min. The collected water quality data include six characteristics of dissolved oxygen, including water temperature, pH, turbidity, ammonia nitrogen, and water level. The collected six water quality parameters are used as input data to predict the dissolved oxygen value in the future from the historical water quality data. The dataset contains 27,077 samples: 24,370 samples are used for training and the remaining for testing. The verification set accounts for 10% of the training set.

Due to the inevitable influence of the sensor’s behavior and complex external environment, there may be missing data and abnormal values. Considering abnormal data, the method of mean inputting is used in the experiment for the improvement. Regarding missing data, we use the temporal regularized matrix factorization (TRMF) method to predict and repair the time series data, considering the time

dependence. The expression of the respective formula is as follows:

$$x_t \approx \sum_{l \in L} \theta_l \otimes x_{t-l} \tag{9}$$

where  $L$  is a lagged set that stores the associated distance between columns in  $X$ ;  $\theta_l$  is the time series coefficient vector of  $x_{t-l}$ .

### Evaluation criteria

In the experiment, four evaluation indexes are used to evaluate the prediction results of different models, which are mean absolute error (MAE), mean absolute percentage error (MAPE), root mean square error (RMSE), and coefficient of determination ( $R^2$ ). The smaller the value of the first three indexes is, the better the prediction, while the closer the  $R^2$  value is to 1, the more accurate the prediction of the model. Their respective expressions are the followings:

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i| \tag{10}$$

$$\text{MAPE} = \frac{1}{m} \sum_{i=1}^m \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (11)$$

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (12)$$

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y}_i)^2} \quad (13)$$

where  $y_i$  is the actual value observed,  $\hat{y}_i$  is the predicted value,  $m$  is the number of test data, and  $\bar{y}_i$  is the mean value of the original observation data.

## Experimental details

### Prediction results of the combined model LSTM-TCN

In order to verify the better performance of the combined LSTM-TCN model for the prediction of dissolved oxygen, experiments with LSTM, TCN single model forecast, and CNN-LSTM combined model forecast were also performed. On the basis of these results, contrast analysis was conducted. In each algorithm, the historical time window was set to 2 to predict 20 min in advance. Figures 7, 8, 9 and 10 are the prediction results of models LSTM, TCN, CNN-LSTM, and LSTM-TCN, respectively. Among them, the blue curve represents the real value, while the orange curve represents the predicted value of dissolved oxygen.

For the prediction of dissolved oxygen, the curve of LSTM-TCN algorithm is more consistent with the actual

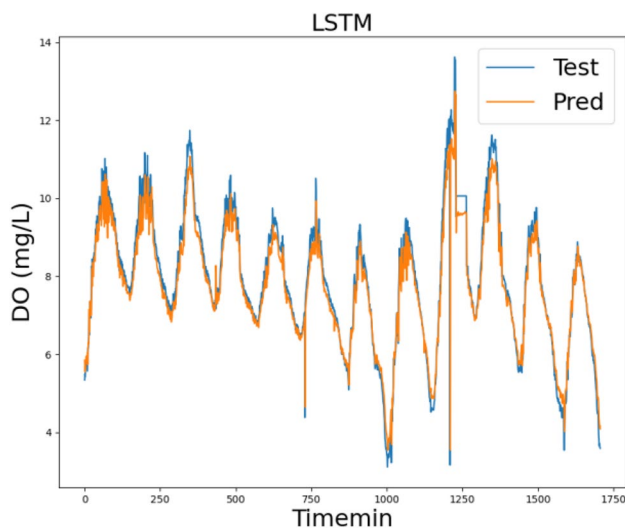


Fig. 7 Prediction results of the LSTM model

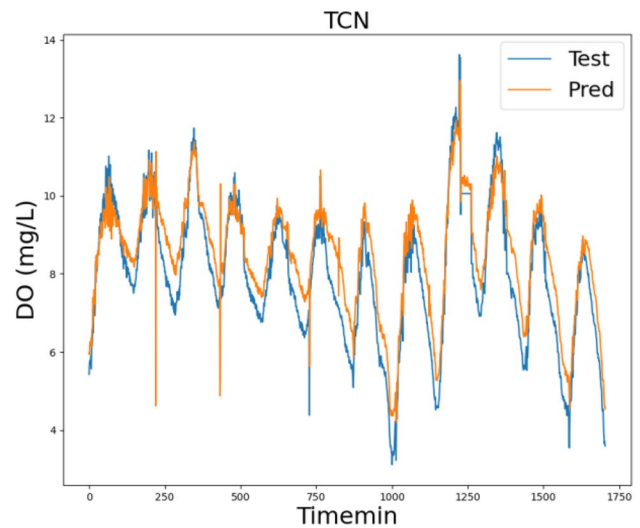


Fig. 8 Prediction results of the TCN model

data, and the prediction result is more accurate, as it can be seen in Figs. 7, 8, 9 and 10.

Table 1 shows the comparison of error results of different prediction models with four evaluation indexes. The pattern of the predicted result curves and evaluation index values can be summarized as follows:

Combined algorithm LSTM-TCN has better prediction than LSTM, TCN, and CNN-LSTM. The final evaluation index values of LSTM-TCN are as follows: for MAE 0.236, which improves the prediction accuracy by 36% compared with LSTM; for MAPE 0.031, the error value decreases by 93% compared with CNN-LSTM; for RMSE 0.342, which improves the prediction accuracy by 33% compared with LSTM;  $R^2$  is 0.94, which is a little bit less than in the case of

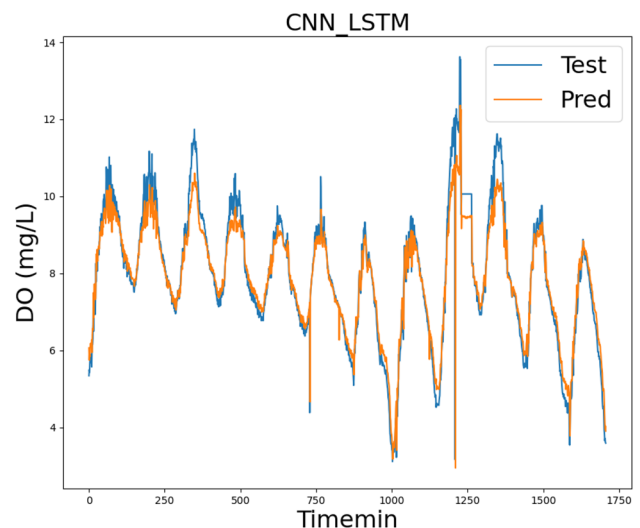


Fig. 9 Prediction results of the CNN-LSTM model



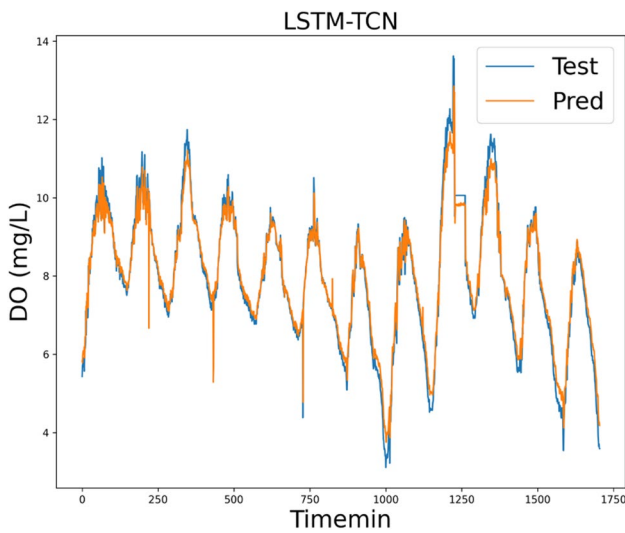
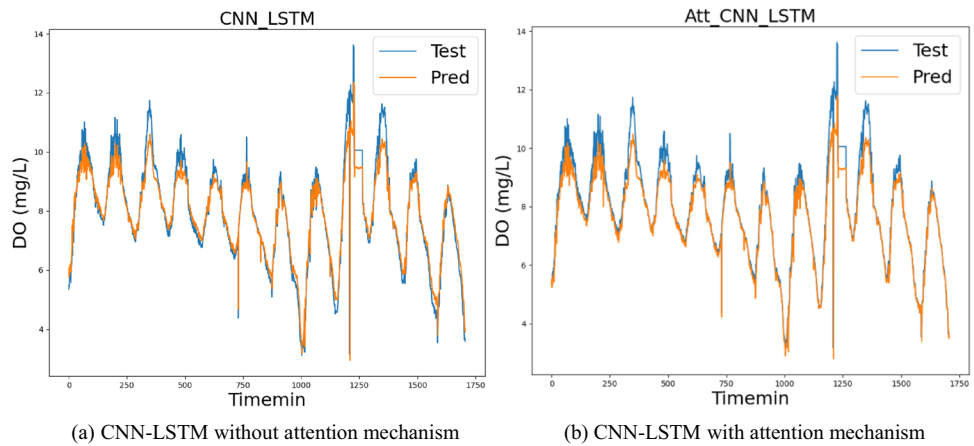


Fig. 10 Prediction results of the LSTM-TCN model

algorithms (0.97). In conclusion, the combined algorithm of LSTM-TCN has better prediction and improves the prediction accuracy of dissolved oxygen to a certain extent.

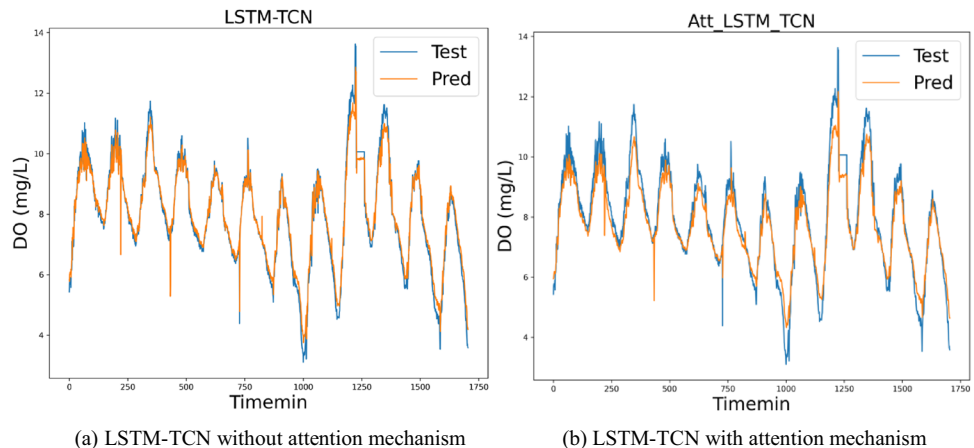
Fig. 11 Comparison of prediction results of CNN-LSTM with and without attention mechanism. a CNN-LSTM without attention mechanism. b CNN-LSTM with attention mechanism



(a) CNN-LSTM without attention mechanism

(b) CNN-LSTM with attention mechanism

Fig. 12 Comparison of prediction results of LSTM-TCN with or without attention mechanism. a LSTM-TCN without attention mechanism. b LSTM-TCN with attention mechanism



(a) LSTM-TCN without attention mechanism

(b) LSTM-TCN with attention mechanism

### Experiment with different lookback window sizes

To study how temporal convolutional network can be used for long historical data, we carried out experiments with different historical time windows to realize the prediction of dissolved oxygen. We changed time window values from 2, 4, 8, 16, 32 to 64 with the same other conditions. Table 2 shows the error results under different time window sizes.

It can be seen from Table 2 that with increasing of time window, the error evaluation index value of dissolved oxygen fluctuates up and down. The prediction results are better when the time windows are 2 and 16. The experimental results show that the TCN network has the ability to obtain more historical data, and the prediction model after adding the TCN network can also have better prediction for the larger historical time window values.

### Attention mechanism

In 2014, attention model was applied to machine translation as part of the RNN framework (Bahdanau et al. 2014). With the development of deep learning, it has become an important concept in the field of neural networks, and it is widely

**Table 1** The result value of evaluation indexes of different algorithms

	Algorithms			
	LSTM	CNN_LSTM	TCN	LSTM_TCN
Evaluation indexes				
MAE	0.366	0.319	0.648	0.236
MAPE	0.256	0.470	0.082	0.031
RMSE	0.514	0.495	0.750	0.342
$R^2$	0.97	0.97	0.73	0.94

used in image classification, speech, and machine translation (Ran et al. 2019). In recent years, more and more attention mechanism modules have been introduced into the prediction of time series. By giving weight values for different time and space data, the efficiency and accuracy of prediction may be improved.

In this section, the attention mechanism experiments are explained. The attention mechanism layer is added between the fusion model network layer and the full connection layer. The aim is to explore whether attention mechanism has an effect on the prediction of dissolved oxygen for specific scenarios and data. The fusion models CNN-LSTM and LSTM-TCN are compared in terms of attention mechanism, and the predicted results are analyzed and compared. Figure 11 shows the comparison of prediction results of CNN-LSTM combined model with and without attention mechanism. Figure 12 shows the comparison of prediction results of LSTM-TCN combined model with and without attention mechanism.

Table 3 shows the error evaluation index results of CNN-LSTM and LSTM-TCN algorithms with and without attention mechanism. It can be seen from Table 3 that after adding attention mechanism to the CNN-LSTM fusion algorithm, the error evaluation indexes MAE and MAPE decreased. For the LSTM-TCN fusion algorithm, the addition of attention mechanism does not improve the final prediction accuracy.

**Table 3** The influence of attention mechanism on prediction results

	CNN-LSTM		LSTM-TCN	
	Without attention mechanism	With attention mechanism	Without attention mechanism	With attention mechanism
MAE	0.319	0.317	0.236	0.484
MAPE	0.470	0.379	0.031	0.064
RMSE	0.495	0.496	0.342	0.575
$R^2$	0.97	0.97	0.94	0.84

## Conclusion

In this work, we propose a fusion prediction model of LSTM and TCN to predict dissolved oxygen in multivariable aquaculture environment. LSTM is used to extract the sequence features in time dimension, dealing with the long-term dependence in complex time series. We built the fusion prediction model of time series with the TCN network. In addition, we also studied the influence of the size of historical time window, as well as the effect of attention mechanism on the prediction. The experiments show that the fusion model still has a good prediction performance when the time window is large. The addition of attention mechanism improves the prediction of the combined model CNN-LSTM, but does not improve the combined model LSTM-TCN. Finally, the performance of the LSTM-TCN prediction model was verified with the dissolved oxygen data in real aquaculture by comparing with the LSTM, TCN, and CNN-LSTM models. The model established in this paper has better prediction performance and higher prediction accuracy, and it can be applied to practical aquaculture, with error measures of MAE = 0.236, MAPE = 0.031, RMSE = 0.342, and  $R^2 = 0.94$ . This study considers only the correlation between water quality data characteristics in time. In the future work, we will combine the water quality environmental data with

**Table 2** Evaluation index results of dissolved oxygen prediction for different lookback window size

	Lookback window					
	2	4	8	16	32	64
Evaluation indexes						
MAE	0.236	0.367	0.407	0.296	0.453	0.667
MAPE	0.031	0.048	0.054	0.039	0.060	0.088
RMSE	0.342	0.568	0.496	0.401	0.563	0.815
$R^2$	0.94	0.87	0.89	0.92	0.83	0.67

the behavior data of aquaculture species, as well as consider the prediction of spatial characteristics of water quality data.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11356-022-18914-8>.

**Author contribution** W.S L: methodology, writing-original draft. Y.G W and DA: conceptualization, supervision, writing-reviewing and editing. Y.S J and QW: software, investigation.

**Funding** This work was supported by the National Key Research & Design Program (2017YFD0701702), and the National Key Research and Development Program of China (2019YFD0901000).

the national key research and development program of china,2017YFD0701702,Yaoguang Wei ,2019YFD0901000,Dong An

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

## Declarations

**Ethics approval and consent to participate** Not applicable.

**Consent for publication** Not applicable.

**Competing interests** The authors declare no competing interests.

## References

- Bahdanau D, Cho K, Bengio Y (2014) Neural machine translation by jointly learning to align and translate [J]. arXiv preprint arXiv:1409.0473
- Bai S, Kolter JZ, Koltun V (2018) An empirical evaluation of generic convolutional and recurrent networks for sequence modeling [J]. arXiv preprint arXiv:1803.01271
- Barzegar R, Aalami MT, Adamowski J (2020) Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model[J]. *Stoch Env Res Risk A* 2020:1–19
- Chen Y, Cheng Q, Fang X, Yu H, Li D (2018) Principal component analysis and long short-term memory neural network for predicting dissolved oxygen in water for aquaculture. *Nongye Gongcheng Xuebao/transactions of the Chinese Society of Agricultural Engineering* 34(17):183–191
- Chen Y, Kang Y, Chen Y, Wang Z (2020) Probabilistic forecasting with temporal convolutional neural network[J]. *Neurocomputing* 399:491–501. <https://doi.org/10.1007/s11356-022-18914-8>
- Cui Y, Wang S, Li J (2015) LSTM neural reordering feature for statistical machine translation[J]. arXiv preprint arXiv:1512.00177
- Dabrowski JJ, Rahman A, George A (2018) Prediction of dissolved oxygen from pH and water temperature in aquaculture prawn ponds[C]//Proceedings of the Australasian joint conference on artificial intelligence-workshops pp 2–6 <https://doi.org/10.1145/3314487.3314488>
- Dabrowski JJ, Zhang YF, Rahman A (2020) ForecastNet: a time-variant deep feed-forward neural network architecture for multi-step-ahead time-series forecasting[C]. International Conference on Neural Information Processing. Cham: Springer pp 579–591
- Deng S, Zhang N, Zhang W, Chen J, Pan JZ, Chen H (2019) Knowledge-driven stock trend prediction and explanation via temporal convolutional network[C]. Companion Proceedings of The 2019 World Wide Web Conference pp 678–685
- Girija TR, Mahanta C (2010) Fuzzy logic and artificial neural network approaches for dissolved oxygen prediction. *Int J Environ Waste Manag* 6(3–4):237–254
- Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9(8):1735–1780
- Hu P, Tong J, Wang J, Yang Y, Turci LDO (2019) A hybrid model based on CNN and Bi-LSTM for urban water demand prediction[C]. 2019 IEEE Congress on evolutionary computation (CEC). IEEE pp 1088–1094
- Hu Z, Zhang Y, Zhao Y, Xie M, Zhong J, Tu Z, Liu J (2019) A water quality prediction method based on the deep LSTM network considering correlation in smart mariculture[J]. *Sensors* 19(6):1420
- Ji X, Shang X, Dahlgren RA, Zhang M (2017) Prediction of dissolved oxygen concentration in hypoxic river systems using support vector machine: a case study of Wen-Rui Tang River, China. *Environ Sci Pollut Res* 24:16062–16076. <https://doi.org/10.1007/s11356-017-9243-7>
- Li Z, Peng F, Niu B, Li G, Wu J, Miao Z (2018) Water quality prediction model combining sparse auto-encoder and LSTM network. *IFAC-PapersOnLine* 51(17):831–836
- Li Y, Zhu Z, Kong D, Han H, Zhao Y (2019) EA-LSTM: Evolutionary attention-based LSTM for time series prediction[J]. *Knowledge-Based Systems* 181:104785
- Liu P, Wang J, Sangaiah AK, Xie Y, Yin X (2019) Analysis and prediction of water quality using LSTM deep neural networks in IoT environment[J]. *Sustainability* 11(7):2058
- Qin H (2019) Comparison of deep learning models on time series forecasting: A case study of dissolved oxygen prediction[J]. arXiv preprint arXiv:1911.08414
- Rahman A, Dabrowski J, McCulloch J (2020) Dissolved oxygen prediction in prawn ponds from a group of one step predictors, *Information Processing in Agriculture*, 7(2), 2020. ISSN 307–317:2214–3173. <https://doi.org/10.1016/j.inpa.2019.08.002>
- Ran X, Shan Z, Fang Y, Lin C (2019) An LSTM-based method with attention mechanism for travel time prediction. *Sensors* 19(4):861
- Shi X, Chen Z, Wang H, Yeung D, Wong WK, Woo W (2015) Convolutional LSTM network: A machine learning approach for precipitation nowcasting[C]. *Advances in neural information processing systems*. 802–810
- Ta X, Wei Y (2018) Research on a dissolved oxygen prediction method for recirculating aquaculture systems based on a convolution neural network[J]. *Comput Electron Agric* 145:302–310
- Wang C, Yang H, Bartz C, Meinel C (2016) Image captioning with deep bidirectional LSTMs[C]. Proceedings of the 24th ACM international conference on Multimedia 988–997
- Xiao Z, Peng L, Chen Y, Liu H, Wang J, Nie Y (2017) The dissolved oxygen prediction method based on neural network[J]. *Complexity* 1–6
- Yan J, Mu L, Wang L, Ranjan R, Zomaya AY (2020) Temporal convolutional networks for the advance prediction of ENSO[J]. *Scientific reports* 10(1):1–15. <https://doi.org/10.1038/s41598-020-65070-5>
- Yang Y, Dong J, Sun X, Lima E, Mu Q, Wang X (2018) A FCCLSTM model for sea surface temperature prediction. *IEEE Geosci Remote Sens Lett* 15(2):207–211

- Yao Y, Huang Z (2016) Bi-directional LSTM recurrent neural network for Chinese word segmentation[C]//International Conference on Neural Information Processing. Cham: Springer pp 45–353
- Yao L, Guan Y (2018) An improved LSTM structure for natural language processing[C]. 2018 IEEE International Conference of Safety Produce Informatization (IICSPI). IEEE 565–569
- Yuan X, Chen C, Lei X, Yuan Y, Adnan RM (2018) Monthly runoff forecasting based on LSTM–ALO model[J]. *Stoch Env Res Risk A* 32(8):2199–2212
- Zhao L, Song Y, Zhang C, Liu Y, Wang P, Lin T, Deng M, Li H (2019) T-gcn: A temporal graph convolutional network for traffic prediction[J]. *IEEE Trans Intell Transp Syst* 21(9):3848–3858
- Zhou J, Wang Y, Xiao F, Wang Y, Sun L (2018) Water quality prediction method based on IGRA and LSTM[J]. *Water* 10(9):1148
- Zhu C, Liu X, Ding W (2017) Prediction model of dissolved oxygen based on FOA-LSSVR[C]. 2017 36th Chinese Control Conference (CCC). IEEE 9819–9823

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.