



Multi-input multi-output temporal convolutional network for predicting the long-term water quality of ocean ranches

Xuan Zhang^{1,2,3} · Dashe Li^{1,2,3}

Received: 4 April 2022 / Accepted: 12 August 2022 / Published online: 1 September 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

The prediction of water quality parameters is of great significance to the control of marine environments and provides a scientific decision-making basis for maintaining the stability of water environments and ensuring the normal survival and growth of marine aquatic products. However, the water quality in ocean ranches is affected by the complex, dynamic, and changeable environments of open water, which have complex nonlinear relationships, poor accuracy, high time complexity, and poor long-term predictability. Therefore, in this paper, a multi-input multi-output end-to-end prediction model based on a temporal convolutional network (MIMO-TCN) is proposed to predict water quality. A ConvNeXt module and TCN module were used as the model encoder and decoder, respectively. ConvNeXt was used to extract the features of the input data, and the TCN used the extracted feature data to achieve improved prediction accuracy. The model adds skip connections between its modules to solve the gradient disappearance problem as the number of network layers increases. To prove the effectiveness of the proposed method, a model robustness and prediction ability evaluation was conducted in this paper based on the dissolved oxygen in multiple ocean pasture validation samples. Compared with other learning models, the mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE) of the MIMO-TCN prediction results were reduced by 60.77%, 30.88%, and 52.45% on average, respectively, and the R^2 improved by 6.07% on average over those of other models. The experimental results show that the proposed method has higher forecasting accuracy than competing approaches.

Keywords Water quality · Deep learning · Temporal convolutional network · Prediction · Long term · Multi-input multi-output

Introduction

A large number of pollutants produced by human activities enter the ocean, exceeding the self-purification capacity of the ocean and causing the destruction of the marine ecosystem. Pollution caused by harmful substances entering the marine environment will damage biological resources, endanger human health, hinder fishing and other human activities at sea, and damage the quality of seawater and the environment. (Kisi et al. 2020; Al-Ghazawi and Alawneh 2021). The water quality parameter is one of the important indices used to evaluate the degree of pollution in an aquatic environment. Keeping the water quality within the normal range is an important means for monitoring the water environment. For example, the normal range of pH in mariculture is 7.0~8.5 (Nong et al. 2020; Huang et al. 2021), and the dissolved oxygen is kept at 5~12mg/L. In traditional mariculture, controlling water

Responsible Editor: V.V.S.S. Sarma

✉ Dashe Li
201513101@sdtbu.edu.cn

¹ School of Computer Science and Technology, Shandong Technology and Business University, Binhai Middle Road, Yantai 264005, Shandong, China

² Key Laboratory of Intelligent Information Processing, Shandong Technology and Business University, Binhai Middle Road, Yantai 264005, Shandong, China

³ Co-innovation Center of Shandong Colleges and Universities: Future Intelligent Computing, Shandong Technology and Business University, Binhai Middle Road, Yantai 264005, Shandong, China

quality based only on aquaculture experience may lead to uncontrollable water quality deterioration, resulting in a series of problems, such as decreased aquatic output. Extensive open aquaculture causes great damage to the ocean environment. With the development of modern fisheries, the number of marine ranches is increasing, and the associated sea area is also expanding. The current data show that China's marine ranches cover an area of approximately 1500 square kilometers, and approximately 178 marine ranch demonstration areas are under construction. Therefore, we should scientifically and effectively predict the water environmental quality parameters of marine ranches and grasp the change trends exhibited by the quality parameters over time, which can help a large number of farmers take countermeasures before the environment seriously deteriorates, ensure the survival and production of fish in the most suitable environment, and improve the quality and output of aquatic products. At the same time, real-time water quality prediction can not only provide an early warning function but also provide a scientific decision-making basis for water environment protection and governance (Cao et al. 2021; Liu et al. 2013; Yan et al. 2021; Hadgu et al. 2014). This will help promote the modernization of marine fisheries and enable the establishment of intelligent aquaculture mechanisms. Therefore, the scientific and effective prediction and timely control of water quality can help farmers adjust their breeding strategies according to the best water quality content, ensure the survival and growth of fish in the most suitable environment, and improve the quality and yield of aquatic products. At the same time, real-time water quality prediction can not only provide an early warning function but also provide a scientific decision-making basis for water environment protection and governance (Cao et al. 2021; Liu et al. 2013; Yan et al. 2021; Hadgu et al. 2014).

Due to the uncertainty of ocean ranch environments and the fact that water quality is influenced by multiple factors, complex nonlinear relationships are involved. The prediction of water quality is a challenging task considering these complex behaviors and the interactions among these factors (Zhu et al. 2020). As the detailed mechanisms of a marine environment cannot all be considered, it is difficult to accurately describe the complex variation trends of such an environment through only mechanism modelling. The application of data-driven methods to predict water quality has achieved remarkable results (Baek et al. 2020). The main methods include the time series method (Zhang et al. 2019), interval- and fuzzy number-based time series prediction method (Liu et al. 2020), traditional machine learning model-based prediction method (Liu et al. 2018; Sun et al. 2021; Li et al. 2022), and neural network prediction method (Deng et al. 2021; Lin et al. 2020; Xu et al. 2022; Shen et al. 2017). These approaches are no longer able to adapt to

complex water quality changes. They possess some defects, such as their poor generalization abilities and limited prediction accuracies.

In recent years, sequence modelling based on deep learning has attracted increasing attention. Deep learning has achieved the most advanced performance in many areas, including image processing (Luo et al. 2020; Liang et al. 2022), language recognition, and text classification. Hinton proposed a restricted Boltzmann machine (RBM) and a deep belief network (DBN) (Hinton and Salakhutdinov 2006; Hinton et al. 2006) in 2006, pointing out that deep networks have strong feature extraction capabilities. The prediction accuracy of deep neural networks (DNNs) in time series problems has been significantly improved (Feng et al. 2019). Some specific DNNs have also been widely used in time series prediction. Du et al. (2023) proposed a deep learning model called Deep Air to predict the surface PM_{2.5} concentration of Shanghai. Sun et al. (2022) proposed a new mixed ship motion and posture prediction model based on long short-term memory (LSTM) and Gaussian process regression (GPR). Yu et al. (2021) proposed a hybrid convolutional neural network-gated recurrent unit (CNN-GRU) deep learning method to predict soil moisture content. Ansari et al. (2022) proposed a network intrusion alarm prediction method based on a GRU. These deep learning models also have many applications in hydrology. Ren et al. (2020) used a multilayer perceptron (MLP) and a RNN to construct hydrological data for water level prediction. Zheng et al. (2021) used LSTM to predict harmful gases in a whole water body. Pu et al. (2019) designed a hierarchical CNN to represent the relationships between Landsat 8 images and in situ water quality levels. In the field of deep learning, recurrently developed neural structures are commonly used in sequence modelling. Including RNNs, LSTM, and GRUs, recurrent networks can meet most prediction accuracy requirements, but they have unsolvable length limitations and gradient problems, and they are difficult to train.

To solve these problems, the convolutional structure of TCN can be used for time series modelling; this type of network accepts an input with an arbitrary length through a sliding 1D convolution, and the gradient problem is not a concern (Bai et al. 2018). In addition, the TCN module has the advantage of a convolution operation. It can accurately capture temporal feature changes; this has been effectively verified by Zhao et al. (2019) in a short-term urban traffic prediction scenario. The accuracy rate can reach a level as high as 95%. Second, each layer of a TCN can utilize convolution operations in parallel instead of sequential processing, as in an RNN, which is helpful for extracting features from complex data (Samal et al. 2021). TCNs have good applications in different fields. Ma et al. (2022) used a TCN and support vector regression based on particle swarm optimization to predict ultrashort-term traction loads. The absolute

error did not exceed 0.3 MW. TCNs have also been effectively applied in the field of wind forecasting (Li et al. 2022). Meka et al. (2021) established a robust deep learning model for the short-term prediction of wind power generation in wind farms by using a TCN. Bian et al. (2022) proposed a short-term power load prediction method based on a TCN-DNN hybrid deep learning model. Fan et al. (2021) applied a TCN in the field of health data detection and achieved improved long-term prediction accuracy. Overall, TCN models are feasible and precise prediction methods in the field of sequence prediction, but TCNs are seldom studied with respect to the prediction of water quality.

In this paper, a TCN-based multi-input multioutput (MIMO) end-to-end prediction model is proposed to represent the nonlinear mapping of water quality. The whole model adopts a fully convolutional network. The MIMO-TCN model consists of an encoder and a decoder. A ConvNeXt module acts as the encoder to read data at the input stage. A stacked TCN acts as the decoder for internal data information processing. As the number of network layers increases, skip connections are added between each pair of modules to solve the gradient disappearance problem. The model can accomplish the tasks of single-step prediction and multistep prediction. The main contributions of this model include the following.

- 1) A deep learning water quality prediction method based on an MIMO end-to-end architecture is proposed; this approach can solve the problem that traditional machine learning models have difficulty conducting prediction across multiple time steps.
- 2) A novel ConvNeXt module and a TCN module form a combined network with the ability to understand and process complex marine water quality scenarios.
- 3) A prediction method based on full convolution is proposed; this technique adapts to the size of the input window, achieves compatibility between multisize inputs and outputs, and reduces the computational burden of the model.
- 4) A deep learning method with a skip connection structure is proposed to solve the gradient explosion, gradient vanishing, and network degradation problems.

Methodology

Problem statement

In marine water quality prediction problems, the main objective of the forecasting task is to predict the multiple water quality values in a future period of time given historical data. As expressed by Eq. (1), X represents water quality

and contains more than one historical value. Here, N represents the sequence dimension, and T represents the time dimension.

$$X = \{X^{(0)}, X^{(1)}, X^{(2)} \dots X^{(N)}\} \in R^{N \times T} \quad (1)$$

The sequence of each dimension $X^{(n)}$ is represented by Eq. (2), where t represents the current timestamp.

$$X^{(n)} = \{x_1^{(n)}, x_2^{(n)}, \dots, x_t^{(n)}\} (n \in [0, N]) \quad (2)$$

When sufficient historical feature data are provided, feature capture and inertia prediction can be realized for these data by training the model. Both single-step prediction objectives can be expressed as $P(X_{t+d}|X_{0:t})$. The multistep prediction case can be expressed as $P(X_{t+1:t+d}|X_{0:t})$, where d represents the future time step.

The output sequence obtained after model calculation is expressed as \hat{Y} , as shown in Eq. (3). \hat{y}_{t+d} represents the future data at time $t+d$. $\hat{y}_{t:t+d}$ represents the future data from time t to $t+d$.

$$\hat{Y} = \{\hat{y}_{T+1}, \hat{y}_{T+2}, \dots, \hat{y}_{T+d}\} \quad (3)$$

The main task of this paper is to establish a mapping from X to \hat{Y} according to Eq. (4) while minimizing the *loss*(\hat{Y}, Y) between the observed value Y and the predicted value \hat{Y} .

$$\hat{Y} = F(X^{(0)}, X^{(1)}, X^{(2)} \dots X^{(N)}) \quad (4)$$

End-to-end strategy

Most prediction models employ single-step prediction, where a sequence is used to train the model to predict the value at the next time step. Considering the nonlinear and complex characteristics of water quality in the ocean, the water quality prediction task requires MIMO water quality prediction.

In essence, traditional machine learning algorithms cannot properly deal with MIMO problems, so a direct multistep forecasting strategy (Taieb and Hyndman 2012) is generally adopted to solve such problems. In other words, multiple models are constructed to predict data in different time periods. Neural networks can overcome such limitations. A structure with a large number of neurons can flexibly support multiple input forms and predict multiple values at once. However, if the given data are directly input into the model, a problem arises: time series relationship is not considered. An end-to-end structure based on deep learning can solve this problem when processing time series. Performing processing with a particular structure such as an encoder or a decoder (Vaswani et al. 2017), thus realizing multiple inputs and multiple

outputs, can also be combined with time series relationship, improving the performance of the resulting model.

In this paper, a novel TCN autoencoder is designed via an end-to-end deep learning strategy, which can realize MIMO water quality data. It can be used to predict the water quality of different ocean ranches. The architecture of MIMO-TCN is shown in Fig. 1, which represents the training process of the model in detail. In Fig. 1a, the input sequence is expressed as $X = \{X^{(0)}, X^{(1)}, X^{(2)} \dots X^{(N)}\} \in R^{N \times T}$. The sequence passes through the encoder and decoder of the model. After training, the model can capture the trends of various changes. The output sequence is expressed as $Y = \{Y^{(0)}, Y^{(1)}, Y^{(2)}, \dots, Y^{(N)}\} \in R^{N \times T}$. Figure 1b shows the movement of the sliding window. The trained network can accurately capture the similar changes exhibited by the time series in the current period and the historical series to predict future data according to historical inertia. This strategy considers the temporal causality in a deep learning network.

The value $X_{t:t+d}$ in the future period only relies on the historical data before time t . We predict $\hat{y}_{t:t+d}$ from $x_1 \dots x_t$, so that the value of $\hat{y}_{t:t+d}$ is close to the actual value. This step of predicting results through historical data is expressed by Eq. (5) below.

$$p(\hat{y}_{t:t+d}) = \prod_{t=0}^T p(y_{t:t+d} | x_1, \dots, x_{t-1}) \tag{5}$$

MIMO-TCN model

The MIMO-TCN model is shown in Fig. 2. The whole model uses a network structure with full convolution and causal convolution. This approach can read a long sequence without limitations. The input is related only to the previous

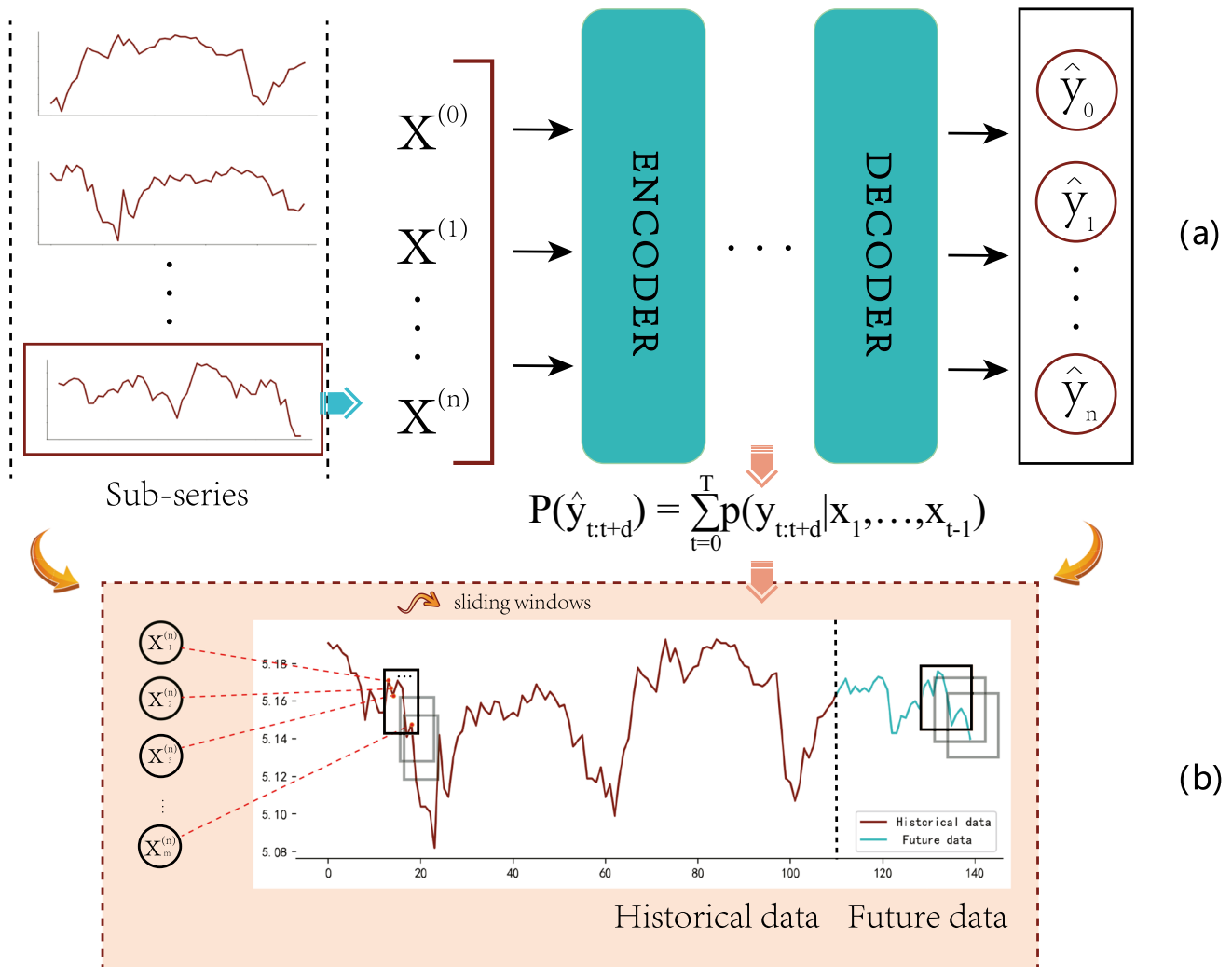
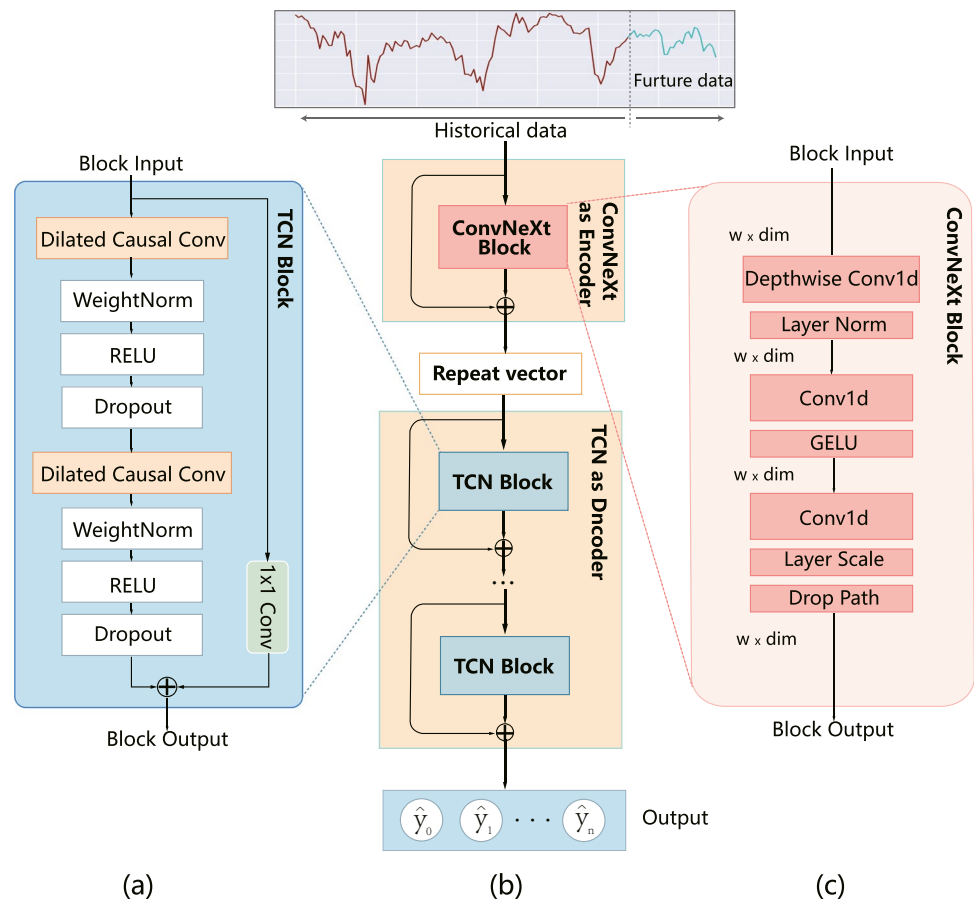


Fig. 1 Deep learning-based MIMO prediction strategy. (a) The model’s strategy. (b) Historical data is similar to future data

Fig. 2 MIMO-TCN structure. **(a)** TCN Block; **(b)** The MIMO-TCN's whole structure; **(c)** ConvNeXt Block



sequence. The proposed model follows this overall architecture and forms a kind of deep learning-based end-to-end forecasting method through a series containing an encoder and a decoder. The encoder and decoder are used to construct the convolutional network. Such networks can adapt to the input window size, support end-to-end training, better learn sequence information, and realize the sizes of inputs and outputs required for purposes. Additionally, the computational cost of the model is reduced. An input sequence $X \in R^{N \times T}$ is mapped to a continuous sequence $Z = \{Z^{(0)}, Z^{(1)}, Z^{(2)} \dots Z^{(N)}\} \in R^{N \times T}$ by the encoder. In the next operation, the previous mapping $Z \in R^{N \times T}$ passes through the repeat vector as input. The decoder then generates an output sequence represented by $\hat{Y} = \{\hat{y}_{T+1}, \hat{y}_{T+2}, \dots, \hat{y}_{T+t'}\}$.

ConvNeXt block

In this paper, a ConvNeXt block (Liu et al. 2022) is used as the encoder of the model to capture multi-input feature information, as shown in Fig. 2c. The ConvNeXt Block first uses a depthwise convolution structure, which calculates each channel of the input data separately. The number of convolution kernels must be equal to the number of input channels. Thus, the number of output channels

is equal to the number of input channels, which greatly reduces the required numbers of computations and parameters. The ConvNeXt model increases the convolution kernel size of depthwise convolution. Second, the ConvNeXt block borrows an important design from the transformer network; that is, the hidden layer adds a bottleneck structure as the implementation of nonlinear transformation. Its memory efficiency is higher than that of the conventional expansion structure, and the ConvNeXt module adjusts the inversion bottleneck structure according to the number of floating point operations. Furthermore, the module uses Gaussian error linear unit (GELU) for activation functions and uses fewer normalization functions, reducing the normalization layer and retains only the normalization layer after performing depthwise convolution. The paper replaces batch normalization with layer normalization, which is commonly used in CNNs. Third, the ConvNeXt network uses the two above convolution layers with convolutions, and a GELU activation function is added between the two layers. Finally, the output size is adjusted to the module input size through layer scaling and path dropping. In this study, the ConvNeXt block is applied for time series prediction. The data to be processed are a one-dimensional sequence, so the two-dimensional convolution

layer is replaced by a one-dimensional convolution layer in the ConvNeXt block.

TCN

MIMO-TCN uses TCN as its decoder, as shown in Fig. 2a. MIMO-TCN makes use of the advantages of causal convolution, dilated convolution, and residual connections in the TCN for time series prediction. On the basis of traditional convolution, a strict time constraint is added to the causal convolution step. The value at time t of the next layer only depends on the value at time T and those before time T in the last layer. The main process is that we predict a future value y_t from $x_1 \dots x_{t-1}$, to make \hat{y}_t close to the observed value y_t . Sequence modelling tasks are accomplished by using the past and current information in each layer. The dilated convolution operation can solve the problem that the longer the length of the input sequence of the causal convolution is, the greater the computational cost. In the calculation of the dilated convolution kernel, the original sequence is discontinuous. That is, some position values are skipped in the continuous sequence to calculate the maximum amount of field information with the limited size of the convolution kernel.

The TCN can achieve unlimited data coverage through the accumulative calculation process of the multilayer convolution and dilated convolution. Finally, the input sequence is determined by the size of the convolution kernel and the number of network layers. The dilation factor $d = (k - 1)^{i-1}$. The length of the input sequence S_X is expressed in Eq. (6), where k is the size of the convolution kernel and L is the number of convolution layers.

$$S_X = \sum_{i=1}^L (k - 1)^i + 1 \tag{6}$$

The more hidden layers there are, the longer the sequence length is. However, to process longer input sequences, the model needs an extremely deep network or a very large convolution kernel. The addition of residual blocks ensures that as the network layers deepen, the accuracy is improved, making the TCN structure more stable than other networks. In other words, equal mappings are added between different layers; this is expressed in Eq. (7), where O represents the output sequence. This module is an overly complex linear regression model, but an activation function is added on top of the convolution layer to introduce nonlinearity, and a rectified linear unit (ReLU) activation function is added after the two convolution layers. Weight normalization is applied to each convolution layer. Regularization is introduced after each convolution layer after the dropout operation to prevent overfitting.

$$O = Activation(X + F(X)) \tag{7}$$

Each layer of the TCN module has 100 nodes, and the output layer is fully connected. Multiple TCN modules are used to form the decoder in MIMO-TCN.

Skip connections

In this paper, skip connections (Long et al. 2015) are added pairs of modules. This enables a single module to contain more detailed information. The model can obtain precise results through step-by-step sampling of the information possessed by a single module. During the process of increasing the number of layers, the network realizes multiple branches to ensure that the degradation of some layers does not affect the overall performance. The skip connection output represents the mapping of the output as a superposition of a nonlinear function $F(X)$ and the original input X . Skip connections are added because they address many problems that are encountered with the deepening of the network, such as the gradient vanishing problem and gradient explosion problem. Taking the vanishing gradient problem as an example, deep learning relies on chain backpropagation to update parameters. During the propagation process, if one of the derivatives is very small, the gradient may decrease after repeated multiplication. However, if residuals are used, an identity term 1 is added to each derivative, as shown in Eq. (8). Even if the original $\frac{df}{dx}$ is small, the derivatives can still be effectively propagated back. At the same time, the reason that the training process of a DNN may fail involves not only the disappearance of the gradient but also the degradation of the weight matrix. Skip connections break the symmetry of the neural network and improve the neuron utilization rate at each layer. The original degraded weight matrix can recover its expression ability after adding skip connections (Orhan and Pitkow 2018). MIMO-TCN breaks the symmetry of the network and improves the representation ability of the network through skip connections.

$$\frac{dh}{dx} = \frac{d(f + x)}{dx} = 1 + \frac{df}{dx} \tag{8}$$

Experiments and discussion

To evaluate the proposed MIMO-TCN method, prediction experiments are performed on 8 real-world ocean ranch dissolved oxygen datasets. First, the dissolved oxygen prediction results obtained for one marine ranch are compared with those of other methods in the “Single-pasture prediction evaluation with other prediction models” section. Second, an ablation experiment and a multistep prediction experiment are carried out on the same experimental data in the “Ablation experiment” and “Analysis of multi-output

results” sections, respectively. Finally, the prediction results are verified by the dissolved oxygen data of eight ocean ranches in the “Prediction accuracy evaluation on data from multiple ranches” section.

Experimental settings

Study areas and data description

Data were collected from 8 ocean ranches along the coast of Shandong Province, China (shown in Fig. 3). Covering approximately 1806 nm of coastline in the Western Pacific Ocean, these locations have longitudes of $114^{\circ}47'E \sim 122^{\circ}42'E$, and latitudes of $34^{\circ}22'N \sim 38^{\circ}24'N$. The depth of the coastal ocean was less than 200 m. Therefore, the water quality change pattern of the coast was different from that of the ocean, which is greatly affected by human interference, as well as terrestrial and marine ecosystems. The variation trends are different in different seasons, and these variations are complex.

The water quality measure studied in this paper was dissolved oxygen. The interval of the collected dissolved oxygen data was 10 min. A total of 144 consecutive samples were collected every day. The time period of data collection was 2020.12.1–2021.8.1. The dissolved oxygen data of 8 marine pastures were collected, and 34,000 samples of data were divided into a training set and test set in time order at a ratio of 15:2. Among them, 30,000 samples of data were input for training, and 4000 samples of data were predicted from the test set. The statistical summary of the dissolved

oxygen water quality levels of the 8 marine pastures is shown in Table 1.

Model forecasting performance metrics

To more clearly evaluate the prediction performance of the model and analyze the errors between the predicted values and the observed values, four performance metrics are adopted in this paper: mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), and coefficient of determination (R^2).

Their mathematical expressions are shown in Eqs. (9), (10), (11), and (12), respectively. The MSE calculates the expectation of the squared difference between the predicted values and the observed values; it is used to detect the deviation between the predicted values of the model and the observed values. The objective of the MAE is to determine the absolute value of the differences between the observed values and the predicted values. The RMSE adds the square root sign on the basis of the MSE, which is more intuitive for comparison purposes. The range of the MAPE is $[0, \infty)$, and a MAPE of 0% represents a perfect model. A MAPE greater than 100% indicates an inferior model; R^2 is the coefficient of determination, which is a measure of how well the regression fits the predicted values. The lower the MAE, MAPE, and RMSE values are, the smaller the error in the predicted values. The higher the R^2 value is, the higher the fitting degree between the predicted values and the observed values.

Fig. 3 Geographical location of the study area

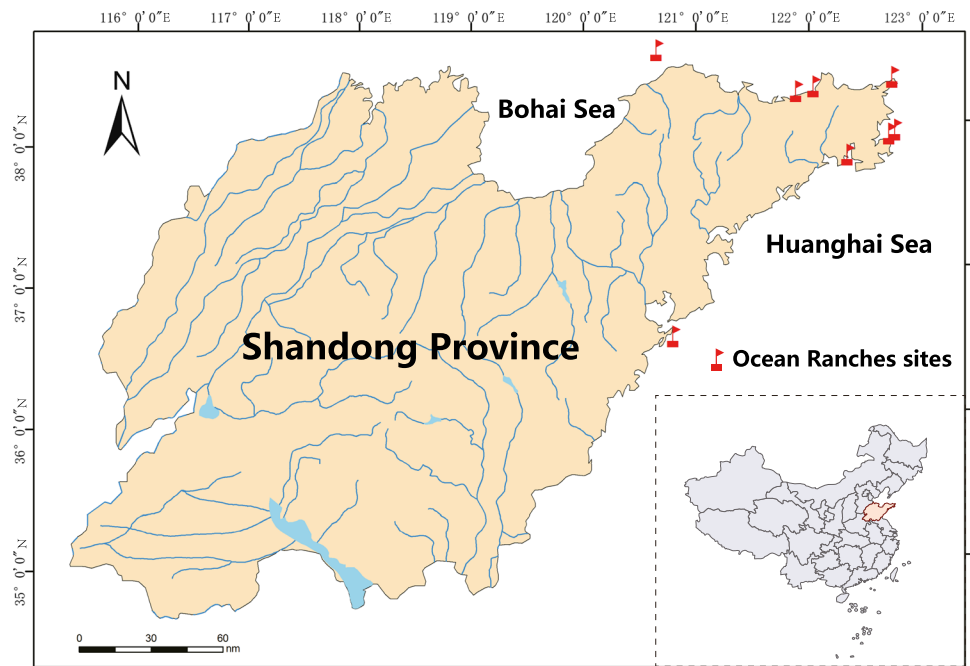


Table 1 Statistical values of each ocean ranch

Ranch number	Dataset	Time	Number	Mean (mg/L)	Max (mg/L)	Min (mg/L)	Std (mg/L)
Ranch 1	All samples	2020.12.1–2021.5.14	24000	3.574737	8.456000	0.134000	1.660393
	Training	2020.12.1–2020.4.16	20000	3.169594	8.456000	0.134000	1.448307
	Testing	2020.4.16–2021.5.14	4000	5.600453	7.348000	2.503000	1.062709
Ranch 2	All samples	2020.12.1–2021.5.14	24000	9.757871	13.851000	1.580000	1.680185
	Training	2020.12.1–2020.4.16	20000	9.676054	13.851000	1.580000	1.739037
	Testing	2020.4.16–2021.5.14	4000	10.166952	12.829000	5.985000	1.271421
Ranch 3	All samples	2020.12.1–2021.5.14	24000	9.883776	13.943000	6.285000	1.136831
	Training	2020.12.1–2020.4.16	20000	10.145956	13.943000	7.641000	1.004064
	Testing	2020.4.16–2021.5.14	4000	8.572872	10.682000	6.285000	0.807165
Ranch 4	All samples	2020.12.1–2021.5.14	24000	9.883776	13.943000	6.285000	1.136831
	Training	2020.12.1–2020.4.16	20000	10.145956	13.943000	7.641000	1.004064
	Testing	2020.4.16–2021.5.14	4000	8.572872	10.682000	6.285000	0.807165
Ranch 5	All samples	2020.12.1–2021.5.14	24000	4.379465	8.697000	0.066000	1.986030
	Training	2020.12.1–2020.4.16	20000	4.204133	8.306000	0.066000	2.114054
	Testing	2020.4.16–2021.5.14	4000	5.256127	8.697000	3.606000	0.630669
Ranch 6	All samples	2020.12.1–2021.5.14	24000	7.332944	12.379000	0.097000	3.956146
	Training	2020.12.1–2020.4.16	20000	8.263106	12.379000	0.097000	3.356914
	Testing	2020.4.16–2021.5.14	4000	2.682135	10.166000	0.143000	3.407042
Ranch 7	All samples	2020.12.1–2021.5.14	24000	6.461706	8.330000	2.794000	0.739697
	Training	2020.12.1–2020.4.16	20000	6.337740	8.330000	5.376000	0.720741
	Testing	2020.4.16–2021.5.14	4000	7.081536	8.076000	2.794000	0.473912
Ranch 8	All samples	2020.12.1–2021.5.14	24000	7.235580	15.196000	0.080000	2.460767
	Training	2020.12.1–2020.4.16	20000	7.814602	15.196000	0.096000	1.953174
	Testing	2020.4.16–2021.5.14	4000	3.412390	7.302000	0.080000	2.004755

$$MAE = \frac{1}{n} \sum_{i=1}^n |o_i - \hat{o}_i| \tag{9}$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (o_i - \hat{o}_i)^2} \tag{10}$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{\hat{o}_i - o_i}{o_i} \right| \tag{11}$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{o}_i - o_i)^2}{\sum_{i=1}^n (\bar{o}_i - o_i)^2} \tag{12}$$

where \hat{o}_i denotes the predicted values, o_i represents the observed values, and \bar{o}_i is the mean of the observed values.

Parameter settings

This study is run on a Windows 10 operating system and compiled with the Python 3.8 language; the specific hardware parameters are an Intel (R) Core (TM) i5-8265 CPU at 1.80 GHz and 8 GB of RAM. The grid search optimization

method is used to set the hyperparameters to obtain the best prediction performance. The optimizer, training loss function, learning rate, dropout rate, number of epochs, batch size, and other hyperparameters for model performance are set. The Adam optimizer is used in this study because of its advantages, such as its fast convergence speed and ease of parameter tuning. The MSE is used as a loss function because the MSE can quickly converge to the minimum loss value during the training process. The dropout rate is set to 0.2 to prevent model overfitting. The learning rate is set to 0.001. The number of epochs is set to 20, and the batch size is 32. The proposed deep learning model achieves the best predictive performance through these hyperparameter settings.

Single-pasture prediction evaluation with other prediction models

To test MIMO-TCN, 34,000 dissolved oxygen samples obtained from the Qingdao Luhai Feng ocean ranch are used as the experimental dataset in this section, and the same experimental set is used in the “[Ablation experiment](#)” and “[Analysis of multi-output results](#)” sections. The training set and test set are divided chronologically. A total of

30,000 data samples are used for training, and 4000 data samples are predicted for the test set. To verify the accuracy of MIMO-TCN, the above predicted dissolved oxygen samples are compared with those of six other similar algorithms according to the performance indicators; these comparison methods include a support vector machine (SVM), a decision tree (DT), an LSTM neural network, a back propagation neural network (BP), an RNN, and a hidden Markov network (HMM). For a fair experimental comparison, all the parameters of the comparative models are searched in a grid. To find the best hyperparameter settings, the same training set and test set are used for comparison. In addition, to clearly display the predicted results and facilitate analysis and understanding, the performance indicators and the predicted results of various models are visually presented in Table 2 and Fig. 4. Table 2 presents the obtained prediction results according to the evaluation indicators, where smaller values and larger values represent more accurate model prediction results. The MAE of MIMO-TCN is 60.77% lower than that of the other algorithms. The RMSE of MIMO-TCN is 30.88% lower than that of the other algorithms. The MAPE of MIMO-TCN is 52.45% lower than that of the other algorithms. The R^2 of MIMO-TCN is 6.07% higher than that of the other algorithms. Compared with these similar algorithms, MIMO-TCN achieves the best prediction performance in terms of all performance indicators with a high fitting degree and a small prediction error. The results show that MIMO-TCN is the best approach. Therefore, the prediction accuracy of MIMO-TCN is higher than that of similar algorithms.

In Fig. 4, an error graph and a line graph are adopted. Figure 4a to f show the error comparison diagrams of the predicted values produced by the SVM, DT, BP, HMM, LSTM, RNN, and MIMO-TCN approaches, respectively. Figure 4g contains the resulting line graphs of all models. The red line in each error graph represents the prediction error of MIMO-TCN, which can be observed in (a)–(f). Compared with the other model errors, the error fluctuation degree of MIMO-TCN is the smallest, and the error value fluctuates slightly near 0. In Fig. 4g, the green line denotes

the observed values, and the red line represents the predicted values. The predicted values of MIMO-TCN have the highest coincidence with the observed values in Fig. 4g. In addition, it can be seen from the analysis of the error graphs and resulting line graphs that when sudden increases or drops in the values occur, the predicted values of the SVM, DT, BP, HMM, LSTM, and RNN models greatly deviate from the real values, and the prediction errors are large; in contrast, the error of MIMO-TCN remains at a small level. As the deep learning model proposed in this paper can learn complex time series, all kinds of changes can still be learned, so high prediction accuracy can still be guaranteed in cases with large data fluctuations.

Ablation experiment

To explore the contribution of each module in MIMO-TCN, the CNN prediction model and its variants are used for dissolved oxygen prediction experiments. Figure 5 shows the prediction results of various CNN model variants. The prediction model based on a CNN has good performance in terms of predicting water quality, as shown in Table 3. The MIMO-TCN model has considerable advantages over the TCN and CNN. Compared with the other models, MIMO-TCN has better performance in terms of the MAE, RMSE, MAPE, and R^2 values of its prediction results. In addition, the performance of the Multi-TCN and MIMO-TCN models has more obvious index value advantage than the CNN, TCN, and TCANs (temporal convolutional attention-based networks). Multi-TCN means that multiple TCN modules are accumulated and skip connections are added. Figure 5 shows the comparison results in terms of the three aspects. To analyze and explore the contribution of each module, the comparison results with respect to these three aspects show the fitting degree of the model at each stage and its measurement index and error distribution performances. Comparison 1, presented in the form of a density correlation graph, is the comparison of the fitting degrees between the predicted values and observed values of the CNN and the variant models. In Fig. 5, the closer the scattered points are to the function $y = x$, the closer the predicted values are to the observed values. In Fig. 5, the slope of the line fitted with the scattered points is closer to 1, and the intercept is closer to 0. The more concentrated the scattered points are, the closer the predicted values are to the observed values and the better the fitting degree is. By comparison 1, it can be clearly observed that the fitting degrees of the two graphs on the right are significantly better than those of the three graphs on the left. The fitting line is closer to the function $y = x$. The aggregation of the scattered points is more concentrated. The fitting degree of the TCAN model is not greatly improved. The MIMO-TCN has the best fitting

Table 2 Performance indices of the experimental results obtained by various comparative models

Model	MAE	RMSE	MAPE	R^2
MIMO-TCN	0.06850	0.131094	1.543634	0.978771
DT	0.131619	0.208057	2.8842917	0.962102
SVM	0.290330	0.444596	6.192001	0.809486
LSTM	0.137658	0.186130	2.268093	0.969227
BP	0.334318	0.261728	3.662923	0.913651
RNN	0.182222	0.237631	3.406464	0.962042
HMM	0.132040	0.231963	2.498296	0.943705

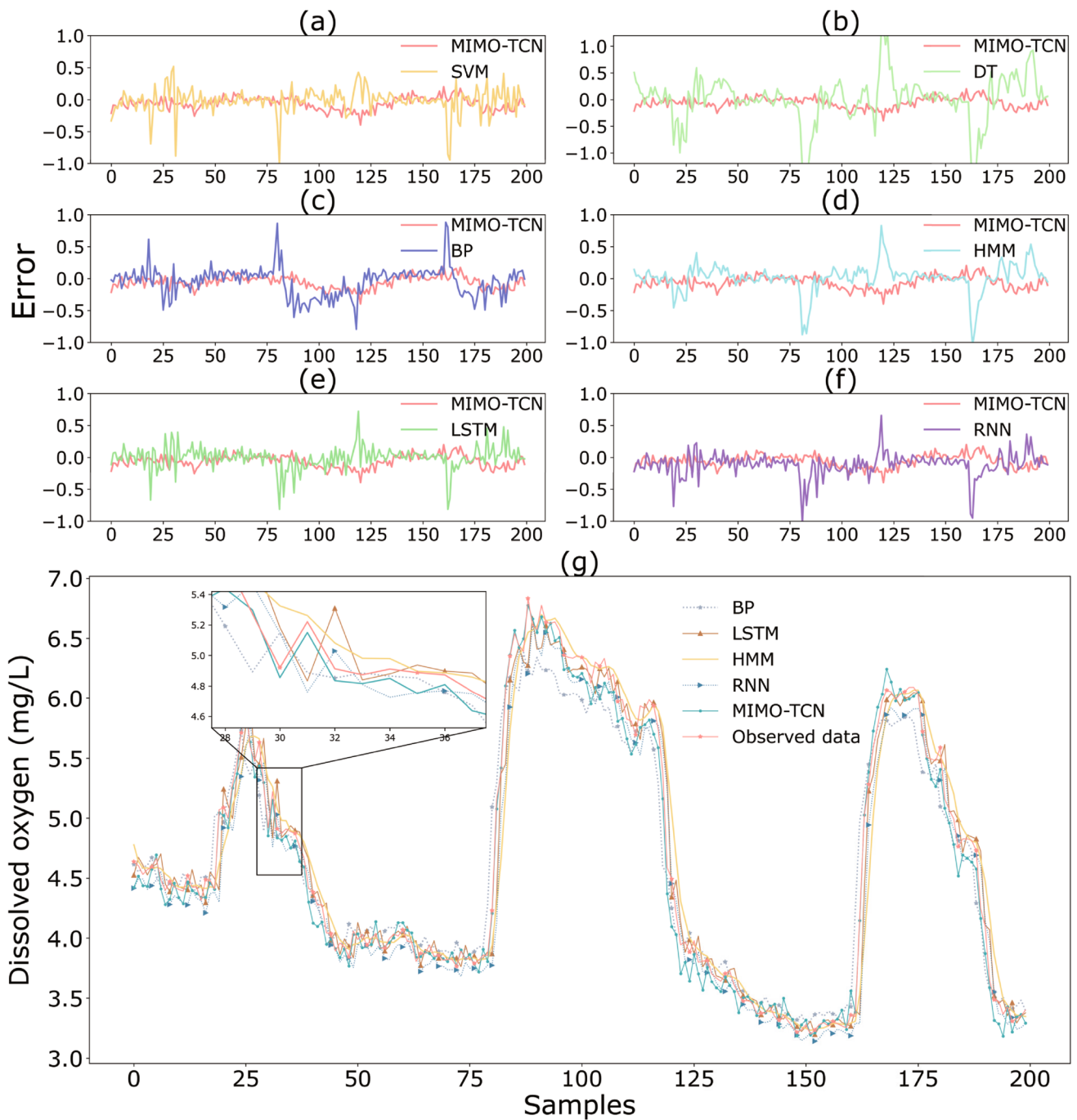


Fig. 4 Comparison of the single-output model forecasting results of dissolved oxygen obtained for a single pasture site. **(a)** are predicted error of MIMO-TCN and SVM; **(b)** are predicted error of MIMO-TCN and DT; **(c)** are predicted error of MIMO-TCN and BP;

(d) are predicted error of MIMO-TCN and HMM; **(e)** are predicted error of MIMO-TCN and LSTM; **(f)** are predicted error of MIMO-TCN and RNN; **(g)** are predicted results of all models

effect, which is better than that of the Multi-TCN model (second from the right in the figure). In comparison 2, different performance metrics are used to compare the prediction results of various CNN model variants, which are displayed in the form of polar axis diagrams. The

MAE figure shows that the MAE value of the MIMO-TCN model is the smallest. The MAE value of the Multi-TCN model decreases to a lesser extent. The MAE value of the TCAN model is not good. Similarly, the MAPE and RMSE values of MIMO-TCN are the lowest, the R^2

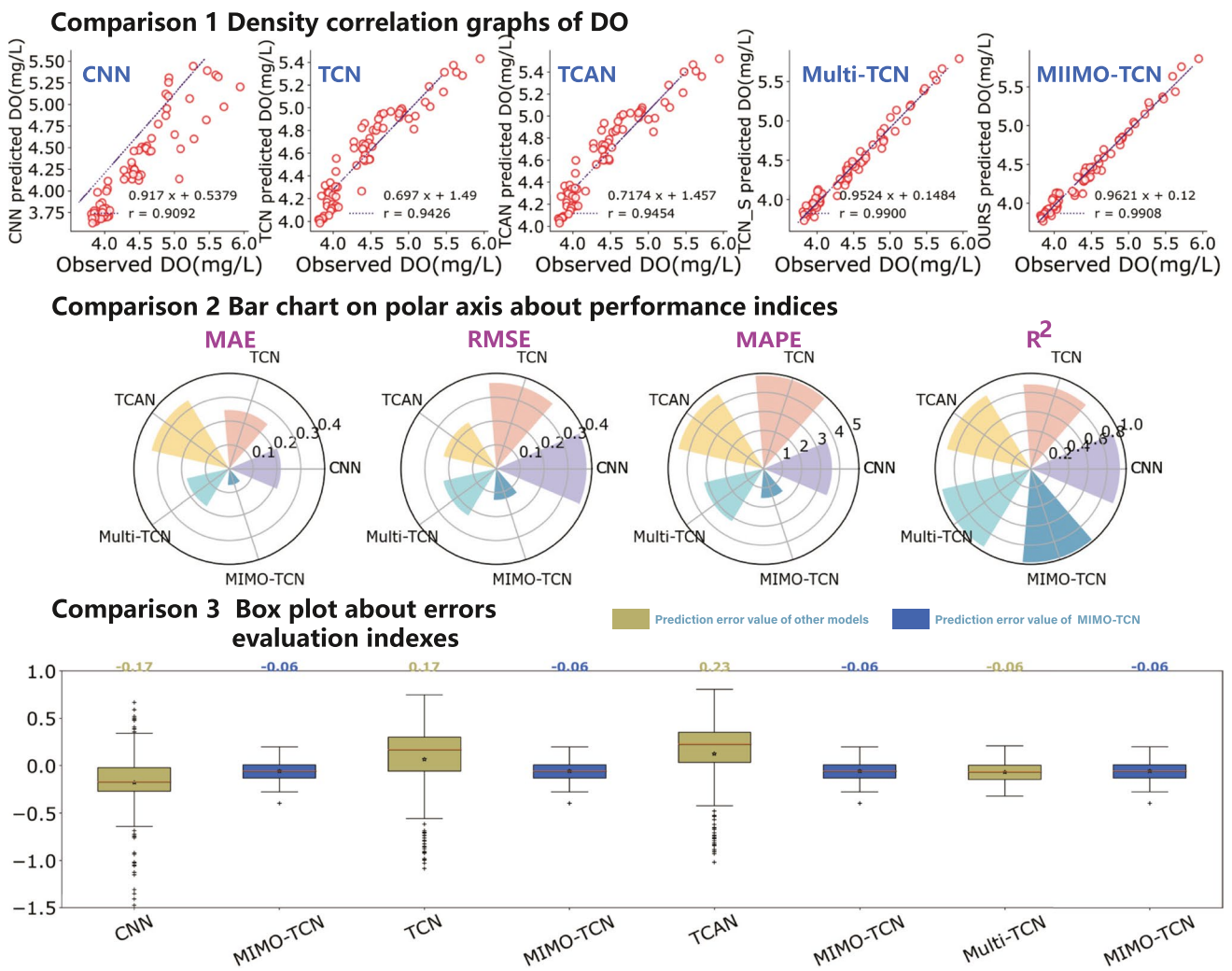


Fig. 5 Forecasting results of the proposed forecasting system and the other benchmark models

increases counterclockwise, and the R^2 value of MIMO-TCN is the largest. In comparison 3, boxplots are used to judge the prediction results the CNN and variant models. The brownish-yellow part shows the model errors, and the blue part shows the MIMO-TCN error. MIMO-TCN has the smallest error and fewer outliers.

Table 3 Performance indices of the ablation results

Model	MAE	RMSE	MAPE	R ²
MIMO-TCN	0.06850	0.131094	1.543634	0.978771
Multi-TCN	0.182464	0.228215	3.199373	0.953146
TCAN	0.334318	0.226904	4.573971	0.899450
TCN	0.245960	0.359642	4.872510	0.883640
CNN	0.216870	0.378171	3.588370	0.935471

This experimental analysis shows that the performance of the proposed model is greatly improved compared with that of the traditional CNN model after the module structure and optimization parameters are adjusted, indicating that the CNN has optimization potential. A TCN is a CNN variant used for sequence prediction, and it has advantages in long sequence prediction tasks. The TCAN increases the attention mechanism, but its experimental performance is barely improved over that of the base model. However, the above experiments prove that the deepening of the TCN model and the combination of the ConvNeXt module are the main reasons for the improved model accuracy. The model proposed in this paper is constructed and compared with various CNN benchmark prediction models and their variants. The above three aspects all prove the effectiveness of the model built in this paper.

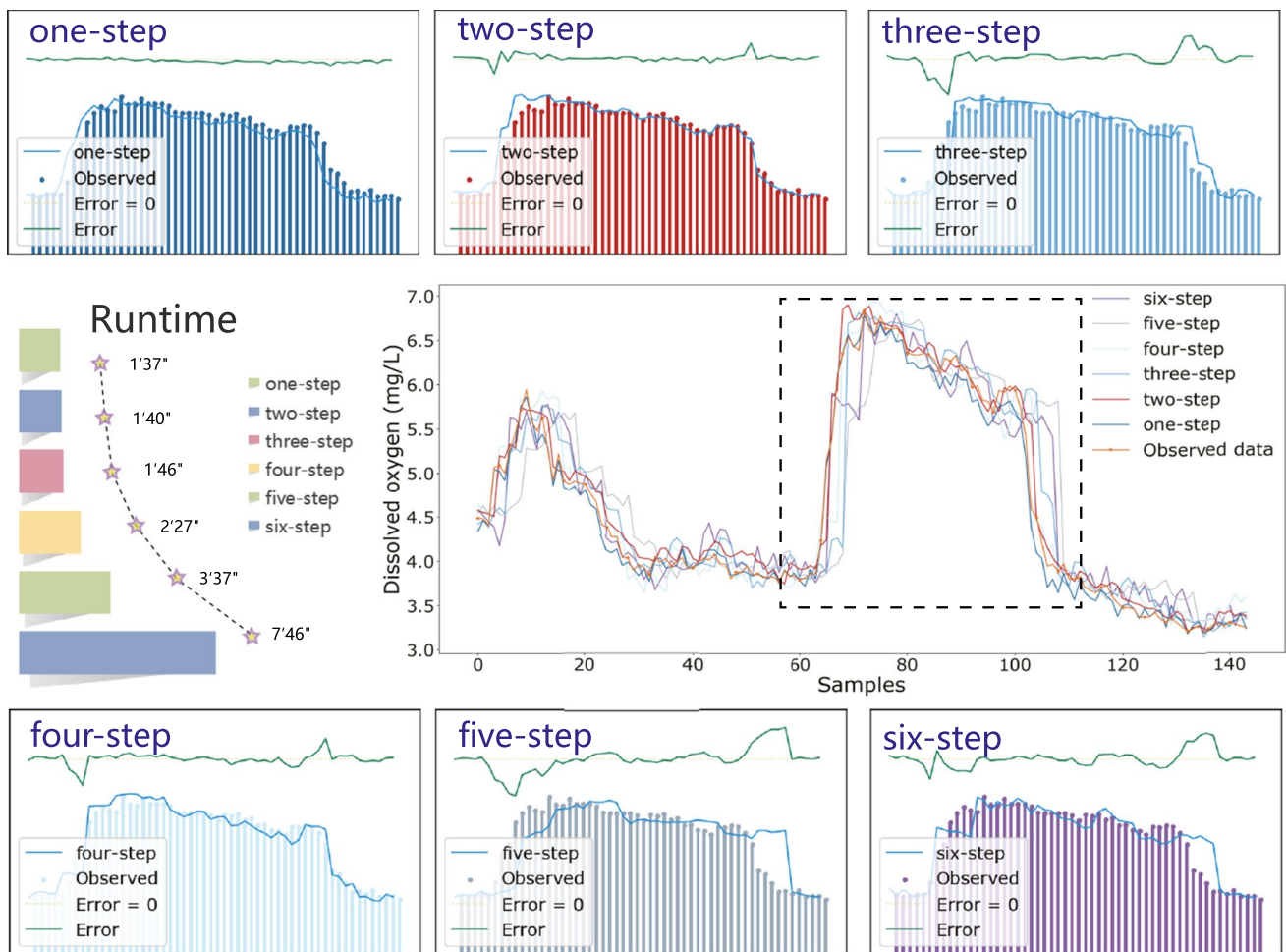


Fig. 6 Comparing the one-step to six-step forecasting accuracy and runtime results

Analysis of multi-output results

On the basis of MIMO prediction, experiments are carried out. In this section, the results of different output steps are compared to verify the prediction results of the model constructed under the MIMO strategy. The number of steps represents the number of results acquired at one time. Figure 6 shows all the obtained prediction results, and the middle part presents multistep prediction. The upper and lower parts of Fig. 6 magnify the results of different steps, while the green line shows the error. The predicted results are basically consistent with the observed values. The error is close to 0 in a single step. With the increase in the number of steps, the error fluctuates slightly. The reason for this result may be that with the increase in the number of steps, the complexity of the data increases. When predicting a step, the MAE, RMSE, MAPE, and R^2 of the model are 0.0685, 0.1310, 1.5436 (%), and 0.9787, respectively. When simultaneously predicting two steps, the MAE, RMSE, MAPE,

and R^2 of the model are 0.1499, 0.2156, 2.8687 (%), and 0.9580, respectively. When simultaneously predicting the three steps, the MAE, RMSE, MAPE, and R^2 of the model are 0.1635, 0.2502, 3.0728 (%), and 0.9434, respectively. When predicting four steps at the same time, the MAE, RMSE, MAPE, and R^2 of the model are 0.1921, 0.2864, 3.6286 (%), and 0.9261, respectively. When predicting five steps, the MAE, RMSE, MAPE, and R^2 of the model are 0.1970, 0.3190, 3.7746 (%), and 0.9079, respectively. When predicting all six steps, the MAE, RMSE, MAPE, and R^2 of the model are 0.2235, 0.3616, 4.2112 (%), and 0.8804, respectively. In addition, with the increase in the number of predicted steps, the running time decreases exponentially. When predicting one step, it takes 7 min and 46 s to train 30,000 iterations. When the number of predicted steps is increased to 6, the training time decreases to 1 min and 37 s. After meeting certain accuracy requirements, the prediction speed can be appropriately increased to provide the latest prediction results in time.

Fig. 7 Performance indicators of the multiranch results

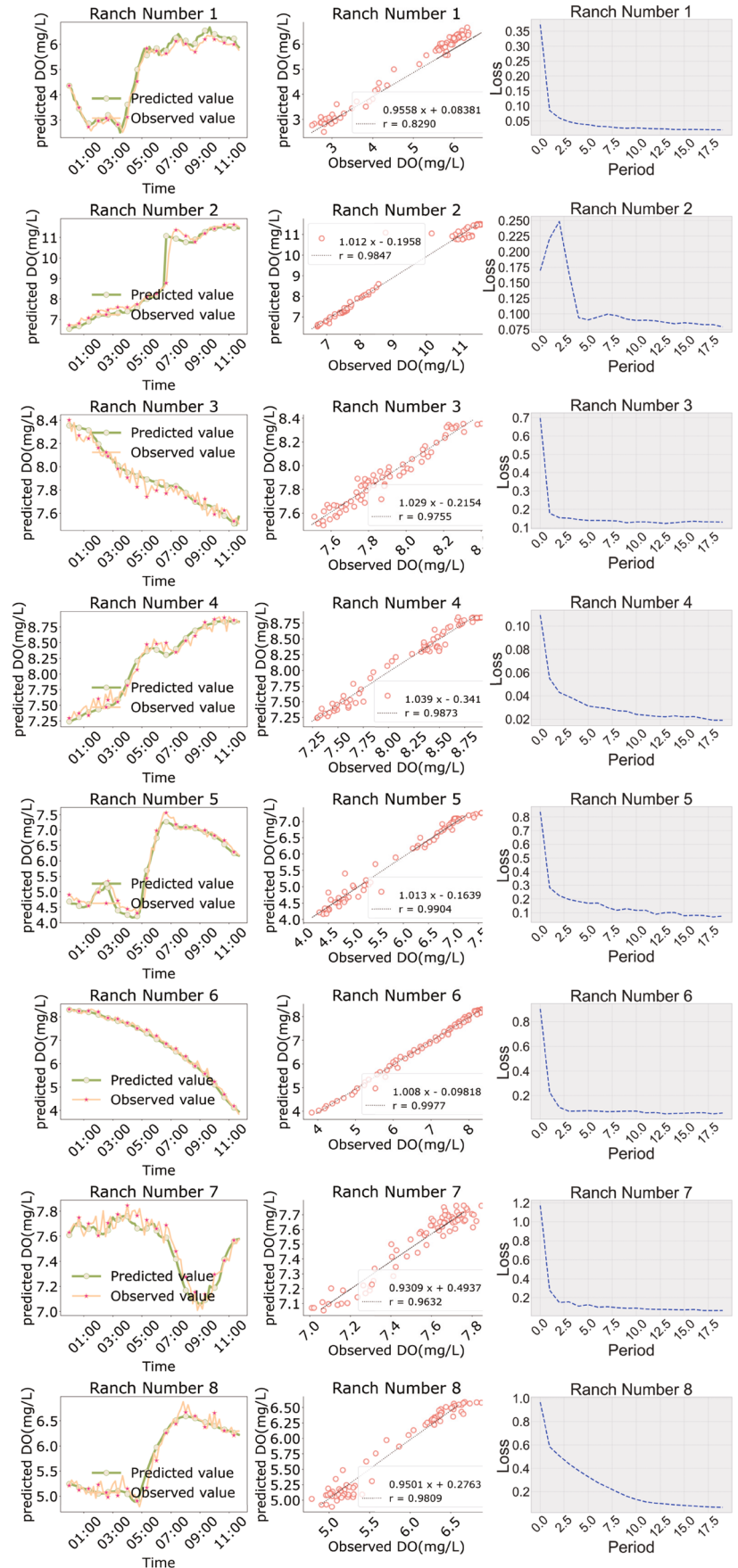


Table 4 Prediction error evaluation indexes of 8 ranch models

Ranch number	MAE	RMSE	MAPE	R^2
Ranch 1	0.185218	0.217589	3.965392	0.976605
Ranch 2	0.202176	0.348654	2.255334	0.967395
Ranch 3	0.048574	0.057009	0.613122	0.946485
Ranch 4	0.077442	0.093889	0.963082	0.972331
Ranch 5	0.142713	0.179730	2.690702	0.974877
Ranch 6	0.075245	0.098661	1.185899	0.994279
Ranch 7	0.055730	0.067810	0.743977	0.908148
Ranch 8	0.099632	0.129570	1.769197	0.959074

Prediction accuracy evaluation on data from multiple ranches

To verify the performance of the MIMO-TCN prediction system in practical applications and to evaluate the accuracy, stability, and generalization ability of the model, prediction experiments are carried out based on 8 real ocean ranches in China. These 8 ranches are distributed in the northern and southern ends of the peninsula, and their environmental climates are different. The near-land ocean areas are greatly affected by the terrestrial climate. The dissolved oxygen prediction experiment involving these 8 ocean ranches can reflect the applicability of MIMO-TCN. The model adopts the same parameter settings as before and adopts the same settings when training different ocean ranches. The prediction results are shown in Table 4 and Fig. 7. The performance of MIMO-TCN is evaluated based on the MAE, MAPE, RMSE, and R^2 indices in Table 4. The MAE, MAPE, and RMSE values of all ranch plots in the table are very low, with R^2 values greater than 0.9. The 12-h (72 samples) prediction results are presented as curves and density correlation plots in Fig. 7. These plots show the predicted and observed values for each ranch and displays their fitting degrees; the dissolved oxygen values in the line chart greatly overlap with the predicted values and the observed values. The density plots enable a correlation analysis between the predicted and observed values; all points located near the $y = x$ line are shown, and the points on the linear regression line are almost covered by the $y = x$ line. This shows that MIMO-TCN can predict dissolved oxygen data with high accuracy. It not only predicts well overall but also performs well in predicting peaks and valleys. Figure 7 also shows the different dissolved oxygen training loss curves of the sea ranches; at the beginning of the training phase, the loss values decrease significantly. After learning until a certain stage, the curves are steady loss curves. The loss change is not as obvious as it is at first. This suggests that the hyperparameters of the model are set appropriately and are stable under the condition of multi-ranch data. The above

experiments show that MIMO-TCN has high prediction stability and a strong generalization ability. MIMO-TCN can be applied to dissolved oxygen data obtained from different environments with high accuracy.

Conclusion

To improve the accuracy of water quality prediction, it is necessary to adapt to increasingly complex water quality changes. In this paper, an end-to-end prediction model called MIMO-TCN is proposed to predict ocean ranch water quality. The encoder and decoder of the model consist of a ConvNeXt module and a TCN, respectively. The skip connection between each pair of modules ensures that the overall model performance is not negatively affected by network degradation as the network deepens. Simultaneously, the ConvNeXt module enlarges the learning limit of the CNN. The whole model not only solves the gradient disappearance problem of the neural network but also achieves improved prediction accuracy. MIMO-TCN is evaluated on a water quality dataset derived from real marine ranches and compared with SVM, DT, BP, HMM, LSTM, and RNN models. The prediction results of the MIMO-TCN model produce error rates that are 30.88% and 52.45% lower than those of the other models, and its R^2 is 6.07% higher on average. Compared with other models, the MIMO-TCN model has better performance in different prediction ranges, with small prediction errors, high fitting degrees, and very good prediction effects in the data segments with large changes and fluctuations. In addition, it has been effectively applied to the water quality data of eight marine pastures with large environmental differences, proving that the method proposed in this paper is robust. Although MIMO-TCN achieves good water quality prediction results, many aspects of the network can still be improved. For example, multidimensional input prediction is realized on the basis of this study. In the future, the prediction model built by the fully convolutional layer in this paper will be adjusted and improved.

Author contribution All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Xuan Zhang and Dashe Li. The first draft of the manuscript was written by Xuan Zhang and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This work was supported in part by the National Natural Science Foundation of China (61772319, 61472227, and 61602277) and the Yantai Science Technology Innovation Development Project (2020YT06000970, 2021XDHZ062, 2022JCYJ032).

Data availability Not applicable.

Declarations

Ethics approval This study follows all ethical practices during writing.

Consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

- Al-Ghazawi Z, Alawneh R (2021) Use of artificial neural network for predicting effluent quality parameters and enabling wastewater reuse for climate change resilience – a case from Jordan. *J Water Process Eng* 44:102423. <https://doi.org/10.1016/j.jwpe.2021.102423>
- Ansari MS, Bartoš V, Lee B (2022) GRU-based deep learning approach for network intrusion alert prediction. *Futur Gener Comput Syst* 128:235–247. <https://doi.org/10.1016/j.future.2021.09.040>
- Baek SS, Pyo J, Chun JA (2020) Prediction of water level and water quality using a CNN-LSTM combined deep learning approach. *Water* 12(12):3399. <https://doi.org/10.3390/w12123399>
- Bai S, Kolter JZ, Koltun V (2018) An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv:1803.01271 [cs] <https://doi.org/http://arxiv.org/abs/1803.01271>
- Bian H, Wang Q, Xu G et al (2022) Load forecasting of hybrid deep learning model considering accumulated temperature effect. *Energy Rep* 8:205–215. <https://doi.org/10.1016/j.egyr.2021.11.082>
- Cao X, Ren N, Tian G et al (2021) A three-dimensional prediction method of dissolved oxygen in pond culture based on Attention-GRU-GBRT. *Comput Electron Agric* 181(105):955. <https://doi.org/10.1016/j.compag.2020.105955>
- Deng Y, Zhou X, Shen J, Xiao G, Hong H, Lin H, Wu F, Liao B-Q (2021) New methods based on back propagation (BP) and radial basis function (RBF) artificial neural networks (ANNs) for predicting the occurrence of haloalkenes in tap water. *Sci Total Environ* 772:145534. <https://doi.org/10.1016/j.scitotenv.2021.145534>
- Du W, Chen L, Wang H et al (2023) Deciphering urban traffic impacts on air quality by deep learning and emission inventory. *J Environ Sci* 124:745–757. <https://doi.org/10.1016/j.jes.2021.12.035>
- Fan J, Zhang K, Huang Y et al (2021) Parallel spatio-temporal attention-based TCN for multivariate time series prediction. *Neural Comput & Applic*. <https://doi.org/10.1007/s00521-021-05958-z>
- Feng S, Zhou H, Dong H (2019) Using deep neural network with small dataset to predict material defects. *Mater Des* 162:300–310. <https://doi.org/10.1016/j.matdes.2018.11.060>
- Hadgu L, Nyadawa M, Mwangi J et al (2014) Assessment of pollution in ndarugu river due to runoff and agro-industrial wastewater disposal. *J Agric Sci Technol* 16(2):110–122
- Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. *Science* 313(5786):504–507. <https://doi.org/10.1126/science.1127647>
- Hinton GE, Osindero S, Teh YW (2006) A fast learning algorithm for deep belief nets. *Neural Comput* 18(7):1527–1554. <https://doi.org/10.1162/neco.2006.18.7.1527>
- Huang J, Liu S, Hassan SG, Xu L, Huang C (2021) A hybrid model for short-term dissolved oxygen content prediction. *Comput Electronics Agric* 186:106216. <https://doi.org/10.1016/j.compag.2021.106216>
- Kisi O, Alizamir M, Docheshmeh Gorgij A (2020) Dissolved oxygen prediction using a new ensemble method. *Environ Sci Pollut Res* 27(9):9589–9603. <https://doi.org/10.1007/s11356-019-07574-w>
- Li D, Jiang F, Chen M et al (2022) Multi-step-ahead wind speed forecasting based on a hybrid decomposition method and temporal convolutional networks. *Energy* 238(121):981. <https://doi.org/10.1016/j.energy.2021.121981>
- Li D, Sun Y, Sun J, et al (2022a) An advanced approach for the precise prediction of water quality using a discrete hidden Markov model. *J Hydrol* p 127659. <https://doi.org/10.1016/j.jhydrol.2022.127659>
- Lin H, Dai Q, Zheng L, Hong H, Deng W, Wu F (2020) Radial basis function artificial neural network able to accurately predict disinfection by-product levels in tap water: taking haloacetic acids as a case study. *Chemosphere* 248:125999. <https://doi.org/10.1016/j.chemosphere.2020.125999>
- Liu G, Xiao F, Lin CT et al (2020) A fuzzy interval time-series energy and financial forecasting model using network-based multiple time-frequency spaces and the induced-ordered weighted averaging aggregation operation. *IEEE Trans Fuzzy Syst* 28(11):2677–2690. <https://doi.org/10.1109/TFUZZ.2020.2972823>
- Liu S, Xu L, Li D et al (2013) Prediction of dissolved oxygen content in river crab culture based on least squares support vector regression optimized by improved particle swarm optimization. *Comput Electron Agric* 95:82–91. <https://doi.org/10.1016/j.compag.2013.03.009>
- Liu Y, Ye L, Qin H et al (2018) Monthly streamflow forecasting based on hidden Markov model and Gaussian Mixture Regression. *J Hydrol* 561:146–159. <https://doi.org/10.1016/j.jhydrol.2018.03.057>
- Liang Y, Rong H, Liu L, et al (2022) Gas-phase catalytic hydration of I2O5 in the polluted coastal regions: reaction mechanisms and atmospheric implications. *J Environ Sci p* S1001074221004009. <https://doi.org/10.1016/j.jes.2021.09.028>
- Liu Z, Mao H, Wu CY, et al (2022) A ConvNet for the 2020s. arXiv:2201.03545 [cs] <https://doi.org/10.48550/arXiv.2201.03545>
- Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3431–3440. <https://doi.org/10.1109/CVPR.2015.7298965>
- Luo Z, Huang F, Liu H (2020) PM2.5 concentration estimation using convolutional neural network and gradient boosting machine. *J Environ Sci* 98:85–93. <https://doi.org/10.1016/j.jes.2020.04.042>
- Ma Q, Wang H, Luo P et al (2022) Ultra-short-term Railway traction load prediction based on DWT-TCN-PSO_svr combined model. *Int J Electr Power Energy Syst* 135(107):595. <https://doi.org/10.1016/j.ijepes.2021.107595>
- Meka R, Alaeddini A, Bhaganagar K (2021) A robust deep learning framework for short-term wind power forecast of a full-scale wind farm using atmospheric variables. *Energy* 221(119):759. <https://doi.org/10.1016/j.energy.2021.119759>
- Nong X, Shao D, Zhong H, Liang J (2020) Evaluation of water quality in the South-to-North Water Diversion Project of China using the water quality index (WQI) method. *Water Res* 178:115781. <https://doi.org/10.1016/j.watres.2020.115781>
- Orhan AE, Pitkow X (2018) Skip connections eliminate singularities. *ICLR p* and 22. <https://doi.org/10.48550/arXiv.1701.09175>
- Pu F, Ding C, Chao Z et al (2019) Water-quality classification of inland lakes using Landsat8 images by convolutional neural networks. *Remote Sens* 11(14):1674. <https://doi.org/10.3390/rs11141674>
- Ren T, Liu X, Niu J et al (2020) Real-time water level prediction of cascaded channels based on multilayer perception and recurrent neural network. *J Hydrol* 585(124):783. <https://doi.org/10.1016/j.jhydrol.2020.124783>

- Samal KKR, Panda AK, Babu KS et al (2021) Multi-output TCN autoencoder for long-term pollution forecasting for multiple sites. *Urban Clim* 39(100):943. <https://doi.org/10.1016/j.uclim.2021.100943>
- Shen L, Wang X, Li R, Yu H, Hong H, Lin H, Chen J, Liao B-Q (2017) Physicochemical correlations between membrane surface hydrophilicity and adhesive fouling in membrane bioreactors. *J Colloid Interface Sci* 505:900–909. <https://doi.org/10.1016/j.jcis.2017.06.090>
- Sun J, Li D, Fan D (2021) A novel dissolved oxygen prediction model based on enhanced semi-naïve Bayes for ocean ranches in north-east China. *PeerJ Comput Sci* 7:e591. <https://doi.org/10.7717/peerj-cs.591>
- Sun Q, Tang Z, Gao J et al (2022) Short-term ship motion attitude prediction based on LSTM and GPR. *Appl Ocean Res* 118(102):927. <https://doi.org/10.1016/j.apor.2021.102927>
- Taieb SB, Hyndman RJ (2012) Recursive and direct multi-step forecasting: the best of both worlds. *Monash Econometrics & Business Stats Working Papers*
- Vaswani A, Shazeer N, Parmar N et al (2017) Attention is all you need. *Advances in neural information processing systems* 30. <https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf>
- Xu Z, Shen J, Qu Y, Chen H, Zhou X, Hong H, Sun H, Lin H, Deng W, Wu F (2022) Using simple and easy water quality parameters to predict trihalomethane occurrence in tap water. *Chemosphere* 286:131586. <https://doi.org/10.1016/j.chemosphere.2021.131586>
- Yan J, Liu J, Yu Y et al (2021) Water quality prediction in the Luan river based on 1-DRCNN and BiGRU hybrid neural network model. *Water* 13(9):1273. <https://doi.org/10.3390/w13091273>
- Yu J, Zhang X, Xu L et al (2021) A hybrid CNN-GRU model for predicting soil moisture in maize root zone. *Agric Water Manag* 245(106):649. <https://doi.org/10.1016/j.agwat.2020.106649>
- Zhang YF, Thorburn PJ, Xiang W et al (2019) SSIM-A deep learning approach for recovering missing time series sensor data. *IEEE Internet Things J* 6(4):6618–6628. <https://doi.org/10.1109/JIOT.2019.2909038>
- Zhao W, Gao Y, Ji T, et al (2019) Deep temporal convolutional networks for short-term traffic flow forecasting. *IEEE Access* 7:114,496–114,507. <https://doi.org/10.1109/ACCESS.2019.2935504>
- Zheng L, Wang H, Liu C et al (2021) Prediction of harmful algal blooms in large water bodies using the combined EFDC and LSTM models. *J Environ Manag* 295(113):060. <https://doi.org/10.1016/j.jenvman.2021.113060>
- Zhu S, Ptak M, Yaseen ZM et al (2020) Forecasting surface water temperature in lakes: a comparison of approaches. *J Hydrol* 585(124):809. <https://doi.org/10.1016/j.jhydrol.2020.124809>

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

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.