



# A Prediction Model of Significant Wave Height in the South China Sea Based on Attention Mechanism

Peng Hao<sup>1</sup>, Shuang Li<sup>1\*</sup>, Chengcheng Yu<sup>1</sup> and Gengkun Wu<sup>2</sup>

## OPEN ACCESS

### Edited by:

Lichuan Wu,  
Uppsala University, Sweden

### Reviewed by:

Chenhua Ni,  
National Marine Technology Center,  
China  
Mingxiang He,  
Shandong University of Science and  
Technology, China  
Antonio Ricchi,  
University of L'Aquila, Italy

### \*Correspondence:

Shuang Li  
lshuang@zju.edu.cn

### Specialty section:

This article was submitted to  
Physical Oceanography,  
a section of the journal  
Frontiers in Marine Science

**Received:** 13 March 2022

**Accepted:** 23 May 2022

**Published:** 16 June 2022

### Citation:

Hao P, Li S, Yu C and Wu G (2022) A  
Prediction Model of Significant Wave  
Height in the South China Sea Based  
on Attention Mechanism.  
Front. Mar. Sci. 9:895212.  
doi: 10.3389/fmars.2022.895212

<sup>1</sup> Institute of Physical Oceanography and Remote Sensing, Ocean College, Zhejiang University, Zhoushan, China, <sup>2</sup> College of Computer Science and Engineering, Shandong University of Science and Technology, Tsingtao, China

Significant wave height (SWH) prediction plays an important role in marine engineering fields such as fishery, exploration, power generation, and ocean transportation. Traditional SWH prediction methods based on numerical models cannot achieve high accuracy. In addition, the current SWH prediction methods are largely limited to single-point SWH prediction, without considering regional SWH prediction. In order to explore a new SWH prediction method, this paper proposes a deep neural network model for regional SWH prediction based on the attention mechanism, namely CBA-Net. In this study, the wind and wave height of the ERA5 data set in the South China Sea from 2011 to 2018 were used as input features to train the model to evaluate the SWH prediction performance at 1 h, 12 h, and 24 h. The results show that the single use of a convolutional neural network cannot accurately predict SWH. After adding the Bi-LSTM layer and attention mechanism, the prediction of SWH is greatly improved. In the 1 h SWH prediction using CBA-Net, SARMSE, SAMAPE, SACC are 0.299, 0.136, 0.971 respectively. Compared with the CNN + Bi-LSTM method that does not use the attention mechanism, SARMSE and SAMAPE are reduced by 43.4% and 48.7%, respectively, while SACC is increased by 5%. In the 12 h SWH prediction, SARMSE, SAMAPE, and SACC of CBA-Net are 0.379, 0.177, 0.954 respectively. In the 24 h SWH prediction, SARMSE, SAMAPE, and SACC of CBA-Net are 0.500, 0.236, 0.912 respectively. Although with the increase of prediction time, the performance is slightly lower than that of 12 h, the prediction error is still maintained at a small level, which is still better than other methods.

**Keywords:** CBA-Net, significant wave height (SWH), deep learning, South China Sea, attention mechanism

## INTRODUCTION

Wave disasters are the most common marine disasters in the world. When huge waves reach the coast, the waves will cause huge losses to people's lives and property (Hsiao et al., 2020; Gao et al., 2021). Therefore, accurate significant wave height (SWH) prediction can effectively improve the safety of marine activities and the efficiency of marine operations, reduce the occurrence of marine accidents, and is of great significance in marine engineering such as fishery, exploration, power generation, and marine transportation (Young and Ribal, 2019; Fan et al., 2020; Zhang et al., 2021).

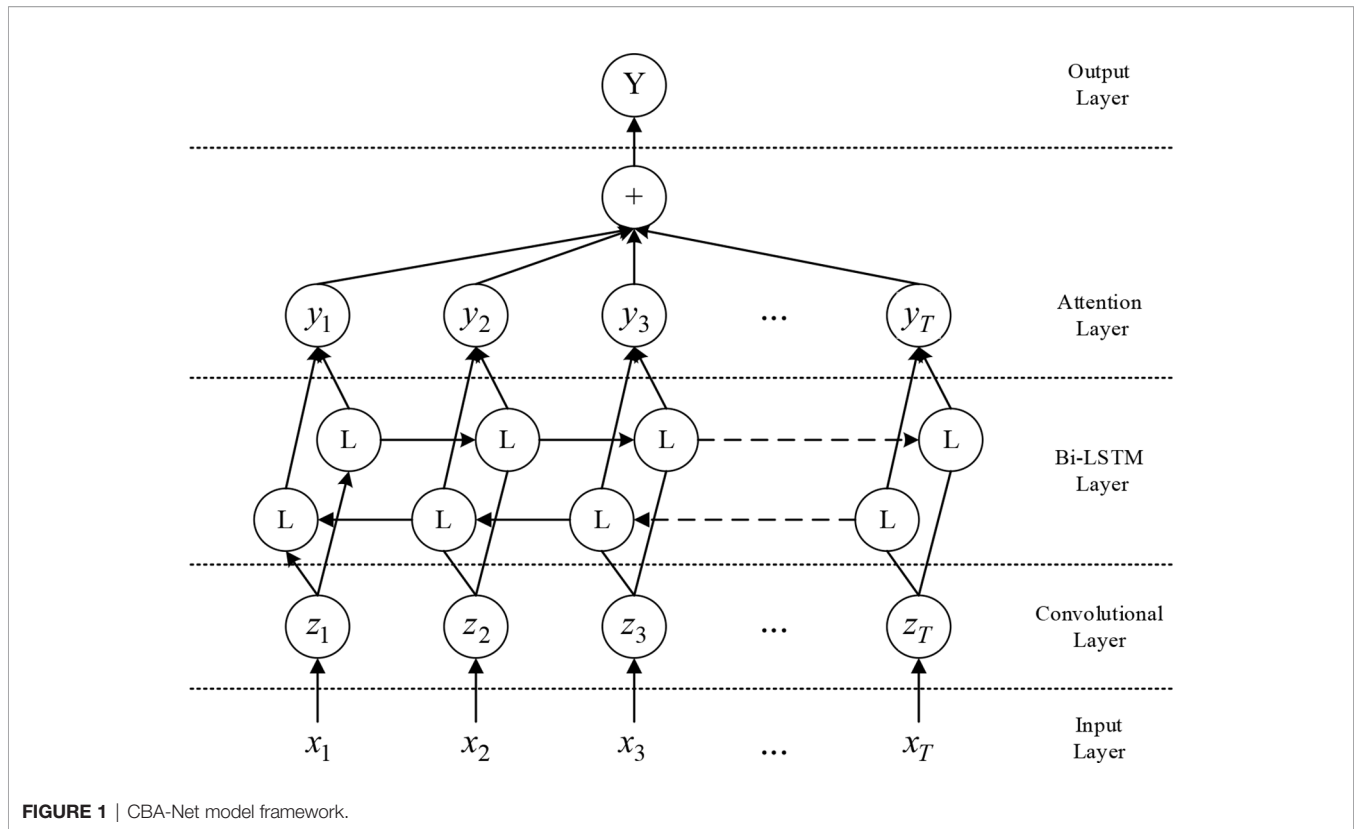
Due to the importance and application value of SWH prediction, SWH prediction methods have been continuously developed in recent decades. Numerical and statistical models (Méndez et al., 2008; Vanem, 2016; Wu et al., 2019; Emmanouil et al., 2020; Wu, 2021; Wu and Qiao, 2022) have been widely used in global sea state prediction. Among them, the common numerical models mainly include such as WAM (Group, 1988; Umesh and Swain, 2018; Swain et al., 2019), WAVEWATCH (Kazeminezhad and Siadatmousavi, 2017; Liu et al., 2019; Li et al., 2020), and SWAN (Akpınar et al., 2017; Liang et al., 2019; Lin et al., 2019). Both numerical model methods and statistical methods try to predict SWH by approximating mathematical relational models. However, due to the strong nonlinearity of the physical processes and mechanisms of ocean waves, especially in extreme cases (e.g., typhoons), such methods may largely fail to achieve high prediction accuracy and need to be improved (Huang and Dong, 2021). In addition, the numerical model requires expensive meteorological and oceanographic data and a large amount of calculation work, and the long-running time is an important bottleneck restricting the development of rapid and accurate SWH prediction (Zhou et al., 2021).

With the rapid development of artificial intelligence (AI), due to its advantages of fast calculation speed, low computational cost, and strong nonlinear learning ability, in recent years, the SWH prediction method based on deep learning has been highly valued by researchers. The deep learning method only needs to know which factors are related to the target physical quantity, establish an input-output prediction model, and predict the SWH for a while in the future. (Panchang and Londhe, 2006) used Artificial Neural Networks (ANN) based on existing wave data sets to predict the wave heights of six geographically separated buoy positions and found that this method has a better prediction effect in the future short-term time range. (Berbić et al., 2017) used ANN and Support Vector Machines (SVM) to predict significant wave heights between 0.5 and 5.5 h. Experiments have verified that ANN and SVM are better than numerical models in this interval. However, the above method can only be applied to forecasts in a relatively short time under normal conditions, while the forecasts under extreme conditions are not ideal. In addition, with the increase in the number of inputs and the increase in complexity, the accuracy of the ANN may drop sharply because the model cannot extract enough features (Ni and Ma, 2020).

Recently, due to the limitations of ANN in SWH, the recurrent neural network (RNN) (Zaremba et al., 2014) has gradually become a more popular SWH prediction model. (Mandal and Prabakaran, 2006) introduced an artificial neural network RNN with a rprop update algorithm and applied it to SWH prediction. (Sadeghifar et al., 2017) used RNN to predict the correlation coefficients of SWH at 3 h, 6 h, 12 h, and 24 h to be 0.96, 0.90, 0.87, and 0.73, respectively. (Miky et al., 2021) integrated neural network-based nonlinear autoregressive network and RNN network for SWH prediction. The experimental results show that the use of RNN for SWH prediction has better results than previous ANN methods. However, the optimization algorithm faces a big problem during RNN training, that is, the problem of long-term dependence-due to the deepening of the network structure, the model loses the ability to learn previous information.

In response to the above problems, the researchers designed a variant of RNN, namely Long Short-Term Memory (LSTM) (Hochreiter and Schmidhuber, 1997). Compared with RNN, it designed a ring structure with two gated units. It can effectively solve the long-term dependence of information, avoid the disappearance or explosion of gradients, thereby significantly improving the accuracy of SWH prediction. (Gao et al., 2021) used LSTM neural network to establish a wave height prediction model at three stations in the Bohai Sea. The model uses sea surface wind and wave height as training samples to evaluate the prediction performance of the model and perform error analysis. It is found that for SWH in the range of 3 to 5 m, the prediction accuracy of the LSTM model is significant. (Zhang et al., 2021) proposed the Numerical Long Short-Term Memory method. This method takes the measured wave height value at the current moment and the combined wave height of the simulated nearshore wave prediction value as input, and generates the corrected numerical prediction as output. Experimental results show that this method effectively improves the SWH prediction accuracy of the Bohai Sea and Wheat Island. (Raj and Brown, 2021) developed and applied a high-precision bidirectional long-term and short-term memory (Bi-LSTM) algorithm to predict SWH, and conducted overall analysis and evaluation of wave characteristics at two coastal locations in Queensland.

However, the application of predicting SWH using AI methods is currently still mainly limited to single points, rather than regions. (Li et al., 2021) First, SWH prediction models are usually a mixture of short-term and long-term dependencies. A successful SWH prediction model should capture these two dependencies to make accurate predictions. Long-term dependence considers the differences between different seasons, and short-term dependence considers wave height fluctuations caused by wind direction and wind direction changes in a short time. If these two dependencies are not considered, it is impossible to make accurate SWH predictions. Secondly, the situation of each site is different, only considering the predictive performance of a single point, without measuring the overall area, the generalization of the model is often



relatively poor. Solving the limitations of existing methods in SWH prediction is the focus of this work.

This paper proposes a deep learning model for SWH prediction of regional multivariate time series, that is, a convolutional bidirectional long-term time series network based on the attention mechanism. As shown in **Figure 1**, it uses a convolutional layer to find local dependencies between multi-dimensional input variables; uses a Bi-LSTM layer to capture complex long-term dependencies; finally, it combines the attention mechanism with the nonlinear neural network part to make the model is more robust. To better demonstrate the effectiveness of the various components of the model, we have carried out an Ablation Study on the model, specifically, we remove each component one at a time in our CBA-Net model framework.

The remainder of this paper is structured as follows. In section 2, we describe our proposed CBA-Net. In section 3, the experimental design details such as the experimental data set, metrics, and parameter settings are introduced. In section 4, we discussed and analyzed the results of SWH prediction. Finally, in section 5, we summarized our findings.

## PROPOSED METHOD

In this section, we introduce the details of the various components of the proposed CBA-Net architecture.

### Convolutional Neural Network Module

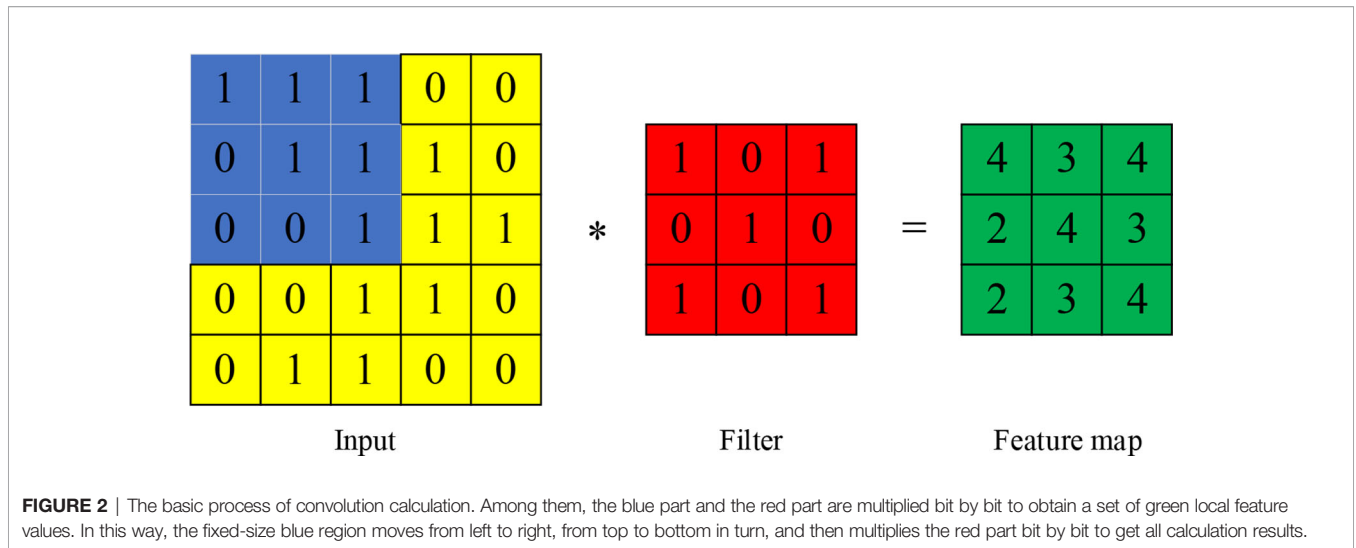
Traditional neural network layers are fully connected. If the number of network layers deepens, this connection method may have an astronomical number of parameters. Convolutional neural network (CNN) has fewer learning parameters than neural networks, which contributes to trainability; in addition, CNN also shows excellent performance in successfully extracting local and translation invariant features (LeCun and Bengio, 1995; Goodfellow et al., 2016).

The first layer of CBA-Net is a CNN without pooling, whose purpose is to extract the local dependencies between variables in the time dimension. This function is mainly accomplished by the filter in the convolution layer. CNN regards the filter as a scanner with specified window size, and extracts feature information by repeatedly scanning the input time series data from left to right and from top to bottom. The convolution calculation process is shown in **Figure 2**.

In this paper, the convolutional layer we built is composed of a filter with a depth of 48 and a width of 3. The  $k$ -th filter sweeps the input time series matrix  $X$  and produces the corresponding calculation results. The calculation formula is as follows,

$$z_k = \text{RELU}(W_k * X + b_k) \quad (1)$$

where  $*$  denotes the convolution operation and the output  $z_k$  would be a vector, the  $\text{RELU}$  function is  $\text{RELU}(x) = \max(0, x)$ ,  $W$  is the weight matrix, and  $b_k$  is the bias.



### Bi-Directional Long Short-Term Memory Module

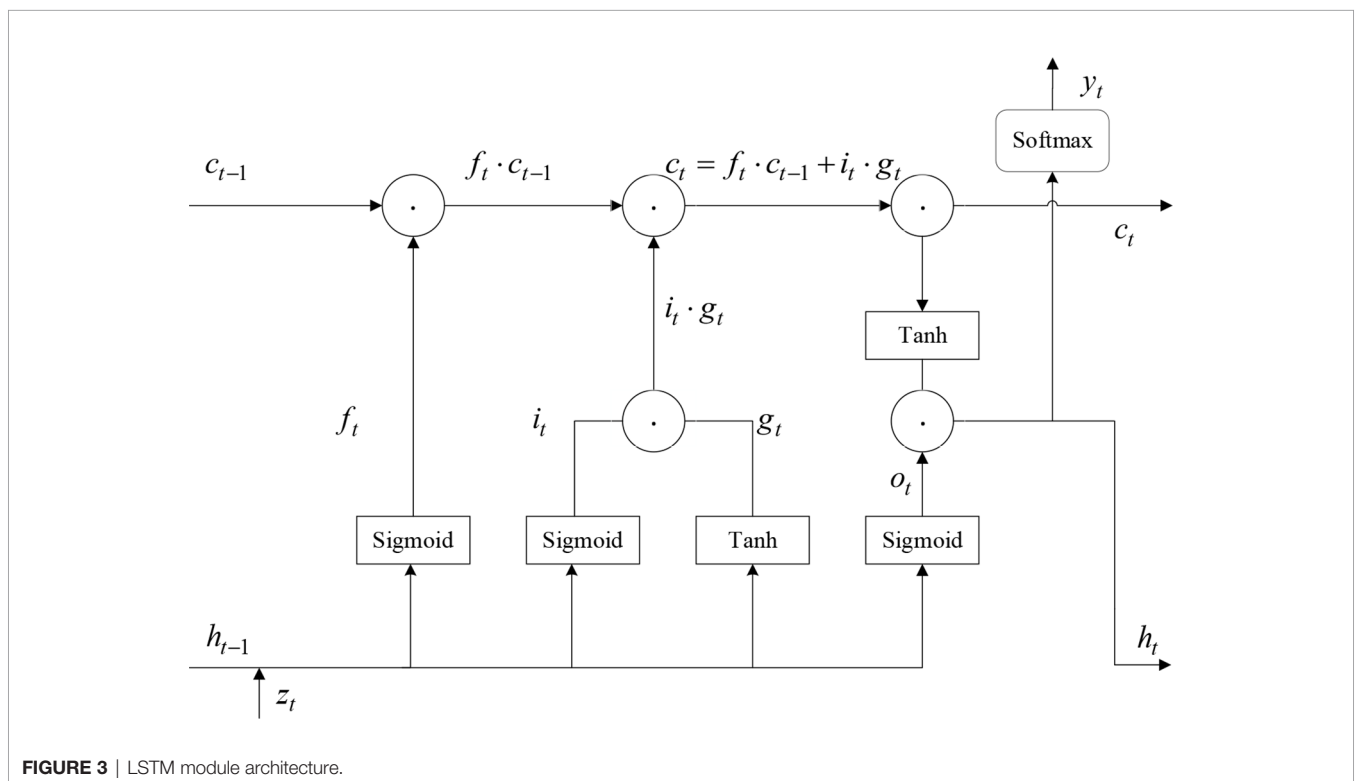
The output of the convolutional layer is input to the Bi-LSTM module in **Figure 1**. Bi-LSTM is a combination of forward LSTM and backward LSTM. As shown in **Figure 3**, LSTM uses two gates to control the content of the cell state  $c$ : one is the forget gate, which determines how much the cell state  $c_{t-1}$  from the previous moment is retained to the current moment  $c_t$ , the other is the input gate, which determines how much of the input  $z_t$  of the network at the current moment is saved in the unit state  $c_t$ .

the LSTM uses an output gate to control how much of the unit state  $c_t$  is input to the current output value  $h_t$  of the LSTM.

This module uses the  $\tanh$  function as the activation function, and the information state transfer formula of the unit at time  $t$  in LSTM is as follows,

$$f_t = \sigma(w_f[h_{t-1}, z_t]) \tag{2}$$

$$i_t = \sigma(w_i[h_{t-1}, z_t]) \tag{3}$$



$$g_t = \tanh(w_g[h_{t-1}, z_t]) \tag{4}$$

$$c_t = f_t c_{t-1} + i_t \cdot g_t \tag{5}$$

$$o_t = \sigma(w_o[h_{t-1}, z_t]) \tag{6}$$

$$h_t = o_t \cdot \tanh(c_t) \tag{7}$$

where  $f_t$  represents the processing formula of the forget gate,  $i_t$  represents the processing formula of the input gate,  $g_t$  represents the new state candidate vector,  $o_t$  represents the processing formula of the output gate,  $w$  represents the given weight coefficient,  $\sigma$  represents the sigmoid function, and  $\cdot$  represents the element-wise product.

Using the LSTM model can better capture long-term dependencies because LSTM can learn what information to remember and what information to forget through the training process, but there is a problem when only building a model with LSTM: it cannot code from back to front information. Therefore, as shown in **Figure 4**, Bi-LSTM is used in this work to better capture the two-way dependency.

At this stage, for the given Bi-LSTM input  $Z = z_1, \dots, z_T$ , where  $T$  is the length of the input time series, the model needs to continuously predict SWH from the input time series data, that is  $H = h_1, h_2, \dots, h_T$ .

### Attention Module

Attention Mechanism originated from the study of human vision (Yang, 2020; Guo et al., 2021). In cognitive science, due to the bottleneck of information processing, humans will selectively focus on part of all information while ignoring other visible information. The attention mechanism has two main aspects:

decide which part of the input needs to be paid attention to; allocate limited information processing resources to the important part. Models without an attention mechanism tend to lose a lot of detailed information when the input data is relatively large-scale. This is the main reason for introducing an attention mechanism in this work.

In this module, multiple dimensions are used to predict one-dimensional data. To fully extract data features and improve the accuracy of predicting SWH, we use the attention mechanism to determine which dimensions play a key role in predicting the dimension.

At time  $t$ , the predicted output  $y_t$  is,

$$y_t = \text{softmax}(W_0 s_t + b_0) \tag{8}$$

Among them,  $W_0$  and  $b_0$  are trainable parameters,  $s_t$  is the hidden state of LSTM at time  $t$ , the formula is as follows,

$$s_t = \text{LSTM}(y_{t-1}, p_t, s_{t-1}) \tag{9}$$

$p_t$  is a context vector, which is calculated from the weighted sum  $H = h_1, h_2, \dots, h_t$  of in the previous stage. The formula is as follows,

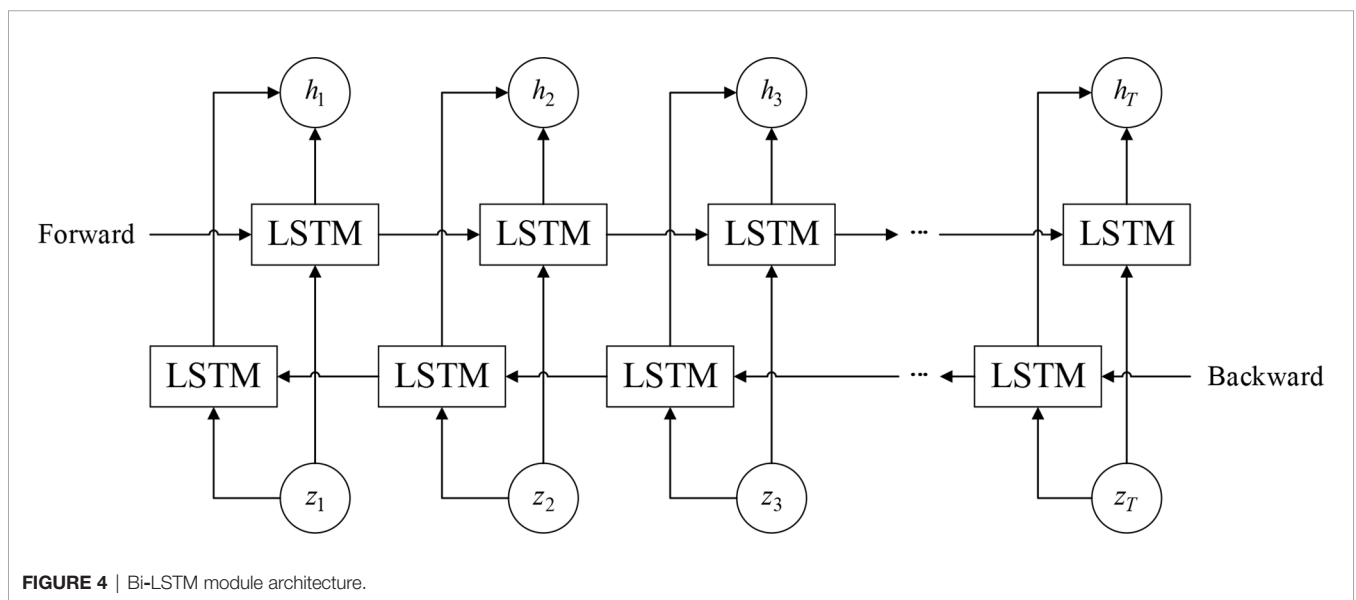
$$p_t = \sum_{i=1}^I \alpha_{ti} h_i \tag{10}$$

Among them,  $\alpha_{ti}$  is called attention weight, and the calculation formula is as follows,

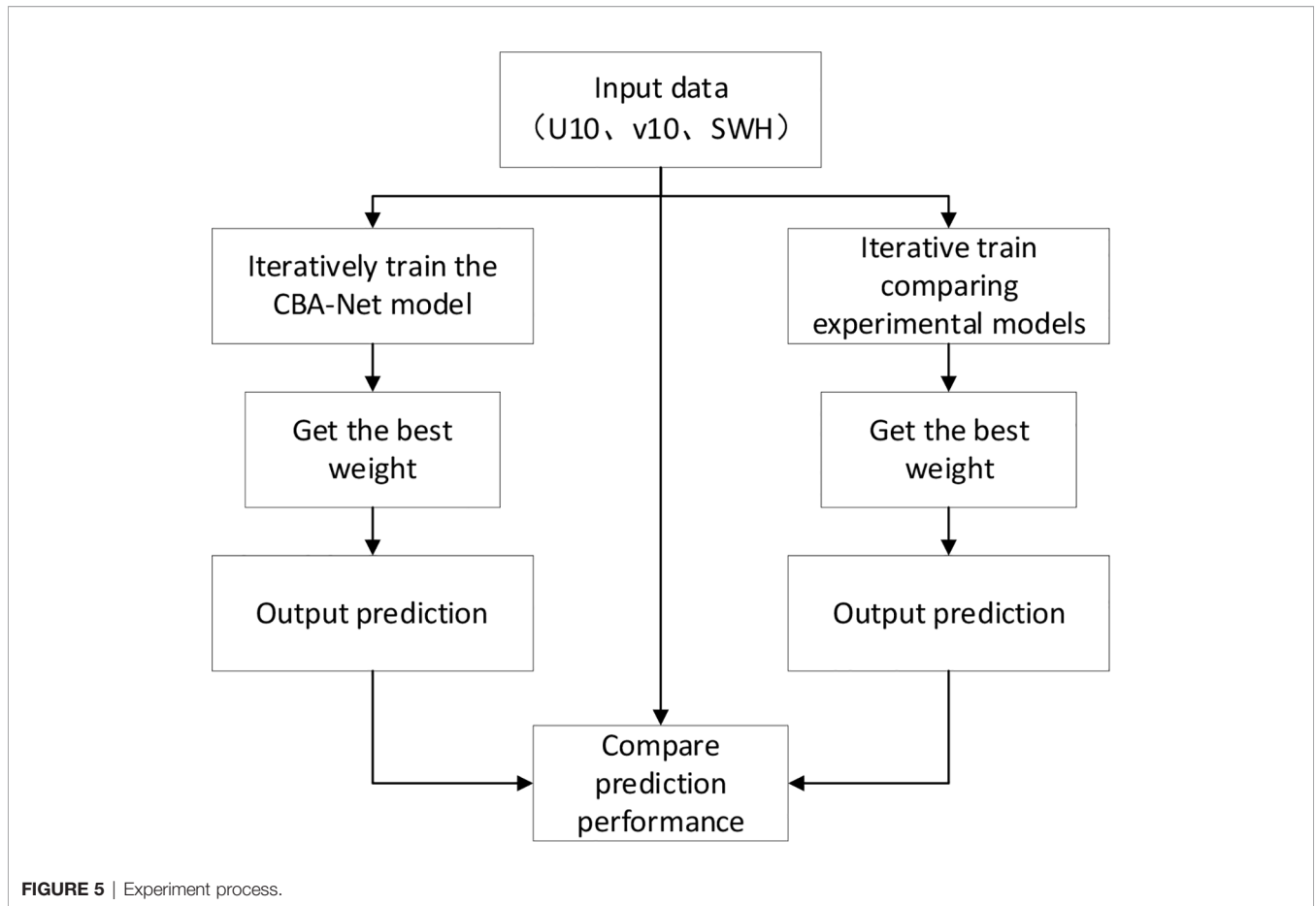
$$\alpha_{ti} = \frac{\exp(\tau_{ti})}{\sum_{k=1}^I \exp(\tau_{tk})} \tag{11}$$

The calculation formula of  $\tau_{ti}$  is as follows,

$$\tau_{ti} = v^T \tanh(Ws_{t-1} + Vh_i + \delta) \tag{12}$$



**FIGURE 4** | Bi-LSTM module architecture.



Among them,  $v$ ,  $W$ ,  $V$  and  $\delta$  are trainable parameters, and the number of LSTM hidden states is set to 256.

## EVALUATION

In this section, we will explain the specific details of the experiment. In order to better understand the experimental process, **Figure 5** shows the flow chart of the overall experiment.

### Dataset

ERA5 is the fifth generation ECMWF reanalysis for the global climate and weather for the past 4 to 7 decades. We select (0°~25°N, 105°~124.75°E) as the study area. This area is dominated by wind and waves and is greatly affected by the monsoon. The time resolution of the data is hours, and the spatial resolution is 0.5°×0.5°.

For the prediction of SWH, we use the data from 2011 to 2018 to generate the corresponding training set and the last 720 hours of data in 2020 as the test set. To ensure the relative independence of training and test data sets, the test data is excluded from model training.

### Metrics

To evaluate the performance of the model, we use the following three metrics, namely Spatial Average Root Mean Square Error (SARMSE), Spatial Average Mean Absolute Percent Error (SAMAPE), and Spatial Average Correlation Coefficient (SACC). The calculation formulas for the above three metrics are as follows,

$$SARMSE = \frac{1}{m} \sum_{j=1}^m \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (13)$$

$$SAMAPE = \frac{1}{mn} \sum_{j=1}^m \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \times 100\% \quad (14)$$

$$SACC = \frac{1}{m} \sum_{j=1}^m \frac{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}}_i)(y_i - \bar{y}_i)}{\sqrt{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}}_i)^2 \sum_{i=1}^n (y_i - \bar{y}_i)^2}} \quad (15)$$

In the formula,  $m$  as the sum of the number of stations in the entire study area  $n$  is the total number of test samples, and  $y_1$  and  $\hat{y}_i$  are the true and predicted values, respectively. Note that the lower the SARMSE and SAMAPE values, the better the



consistency between the measurement and the prediction, but the higher the SACC value, the more accurate the prediction.

## Experimental Details

We use Intel Gold 6330 processor and Nvidia GeForce RTX 3090 graphics card for experiments in the Ubuntu20.04 system. The methods mentioned in the experiment are all implemented by Tensorflow 2.x in the Python environment.

To prevent overfitting, we add a dropout layer after the convolutional layer and Bi-LSTM layer and set the parameter to 0.3. In addition, the model uses the Adam algorithm (Kingma and Ba, 2014) to optimize model parameters. Adam algorithm improves the quality and speed of optimization by obtaining an adaptive learning rate for each parameter.

## RESULTS

We conducted multiple sets of experiments to verify the performance of the model to predict SWH and analyzed the experimental results. To prove the effectiveness of each component of our model, we conducted a careful ablation study.

### One-Hour SWH Prediction

**Table 1** lists the one-hour prediction results of the different models after training, verification, and testing. The optimal results are marked in bold in the table. It is worth noting that we sequentially add a module to perform an ablation study to verify the effectiveness of each component of our proposed model.

It can be seen from the results that the error of using CNN alone to predict SWH is too large, and it may not be able to accurately predict SWH. However, after using Bi-LSTM based on CNN, the effect has been significantly improved, and the correlation of prediction reaches 0.925. The error is also greatly reduced, SARMSE is reduced from 1.039 to 0.528, a reduction of 49.2%. SAMAPE is also maintained at a low level, indicating that CNN only considers local dependencies to predict SWH is unreliable. After applying Bi-LSTM, long-term dependencies can be captured, which greatly improves the accuracy of SWH prediction. Based on the first two, after introducing the attention mechanism, SARMSE is reduced to 0.299, SAMAPE is reduced to 0.136, which is a reduction of 48.7%, and the error has reached a very low level. At the same time, the correlation of prediction can reach 0.971, The reliability of the model prediction SWH is greatly improved.

From the experimental results in **Table 1**, it can be seen that CBA-Net can maintain better prediction performance when

**TABLE 1** | One-hour performance results.

Metrics	CNN	CNN+Bi-LSTM	CNN+Bi-LSTM+ATTENTION
SARMSE	1.039	0.528	<b>0.299</b>
SAMAPE	0.371	0.265	<b>0.136</b>
SACC	0.745	0.925	<b>0.971</b>

predicting SWH for 1 h. To display the SWH prediction results more intuitively, **Figure 6** show the prediction results of 1 h SWH by different methods. Because CNN only pays attention to the local dependencies, it is easy to fall into the local minimum point, which leads to the under-fitting of the prediction model. After adding the Bi-LSTM module based on CNN, the predictive ability of the model has been improved. Although the resulting error is still at a relatively large level, it has a similar change trend with the real data. After introducing the attention mechanism based on the first two, the predictive ability of the model is further improved. Only some areas have an error of about 0.4m, and the overall error is maintained at a relatively low level.

### Twelve-Hour SWH Prediction

**Table 2** lists the twelve-hour prediction results of different algorithms after training, verification, and testing. The optimal results are marked in bold in the table.

It can be seen from the table that CNN's SWH prediction performance indicators are further reduced. It can be concluded that a single CNN model is not suitable for time series SWH prediction. As the SWH prediction period increases, the correlation between data decreases. But Bi-LSTM can fully extract the dependency between data and data through the ingenious design of bi-directional LSTM. After introducing Bi-LSTM, the degree of data error is slightly higher than the 1 h prediction result under the same conditions, indicating that Bi-LSTM The application of the algorithm is meaningful. After introducing the attention mechanism on the basis of the first two, compared with the 1 h SWH prediction results under the same conditions, the correlation decreases from 0.971 to 0.954. Due to the reduced correlation, both SARMSE and SAMAPE increased slightly, from 0.299 to 0.379 and 0.136 to 0.177 respectively; however, the error was within an acceptable range.

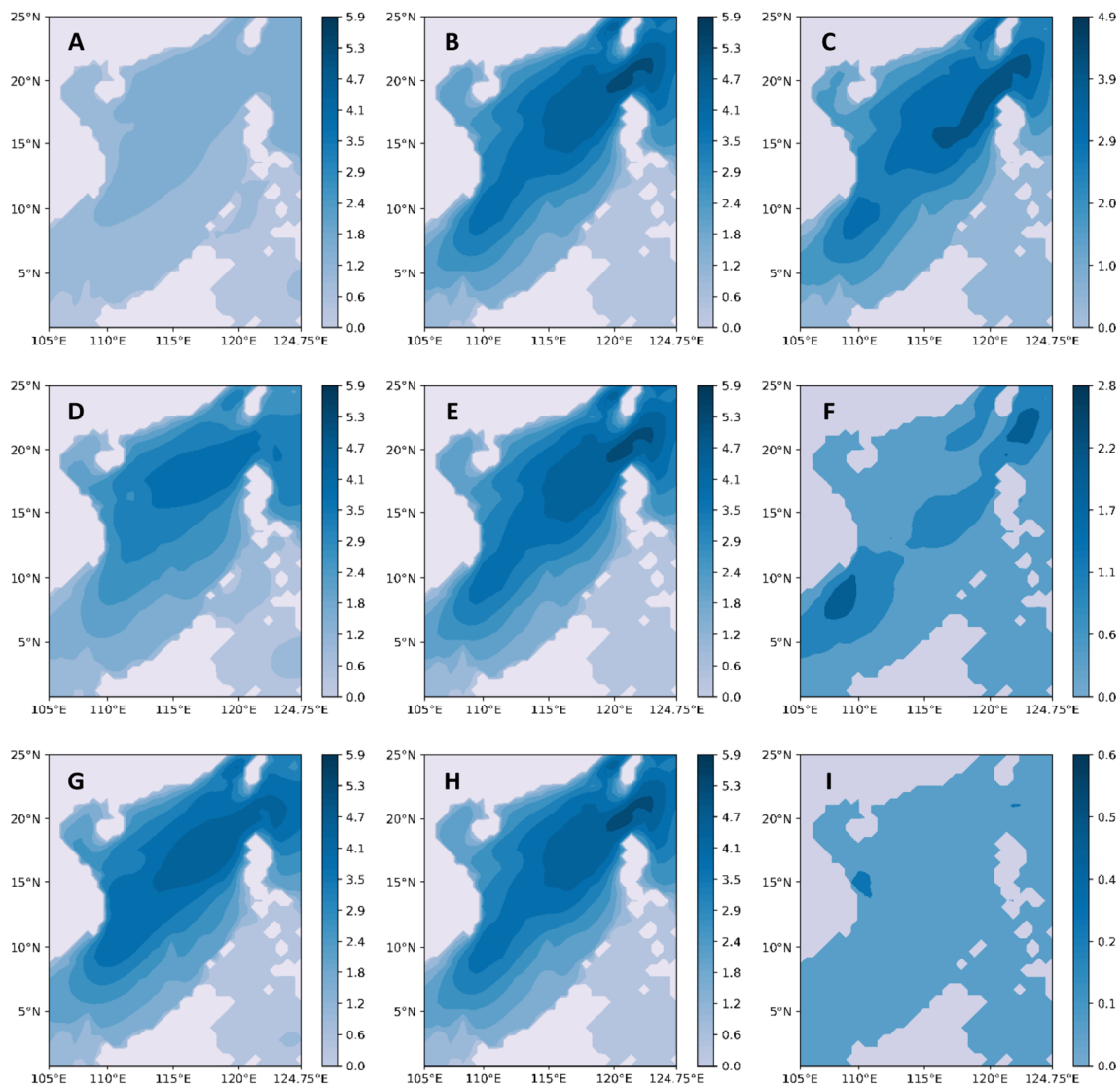
Compared with the prediction performance of the 1 h SWH model under the same conditions, the 12 h SWH prediction index is slightly lower. At present, the possible reason is that the forecast period is relatively large.

**Figure 7** show the prediction results of different methods for 12 h SWH. Although the forecast period increases, the results of CBA-Net are in good agreement with the original data, indicating that the method proposed in this paper has strong generalization ability and long-term prediction ability. In a small part of the area, the 12 h prediction results have a slightly larger error, with an error of about 0.5m, but the overall prediction is the same as the actual measured data, which shows that CBA-Net is feasible in the 12 h SWH prediction.

### Longer-Term SWH Prediction

**Table 3** lists the One-day prediction results of different algorithms after training, verification, and testing. The optimal results are marked in bold in the table.

As the complexity of marine engineering increases, so does the demand for long-term SWH forecasts. It can be seen that,



**FIGURE 6** | Continuous prediction for one-hour prediction. (A–C) are the prediction results of the CNN model, the ERA5 data results, and the prediction errors, respectively. (D–F) are the prediction results of the CNN + Bi-LSTM model, the ERA5 data results, and the prediction errors, respectively. (G–I) are CBA-Net model prediction results, ERA5 data results, and prediction errors, respectively.

**TABLE 2** | Twelve-hour performance results.

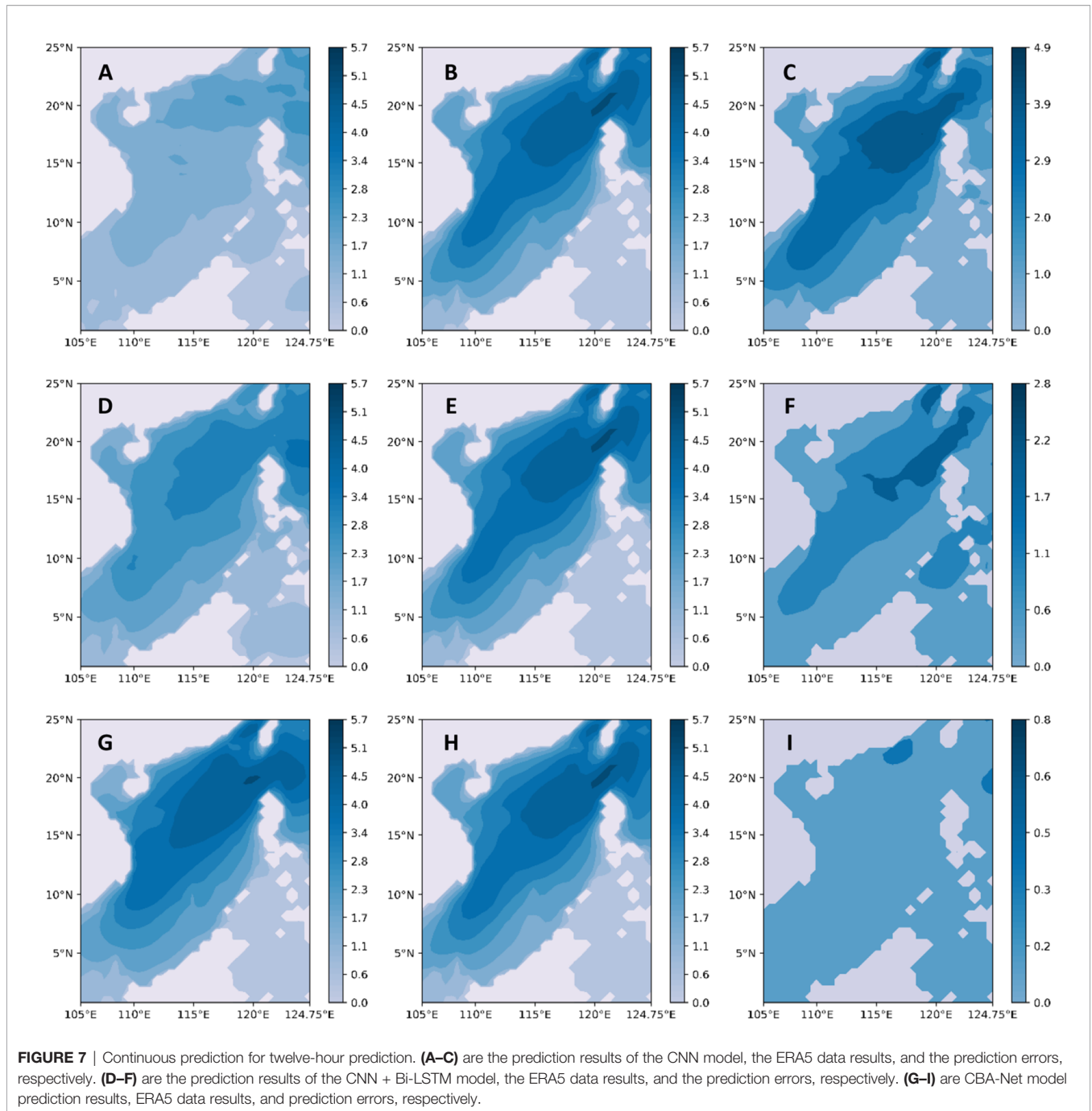
Metrics	CNN	CNN+Bi-LSTM	CNN+Bi-LSTM+ATTENTION
SARMSE	1.137	0.558	<b>0.379</b>
SAMAPE	0.374	0.272	<b>0.177</b>
SACC	0.715	0.917	<b>0.954</b>

compared with the 1 h and 12 h SWH prediction, the 24 h SWH prediction SARMSE and SAMAPE are 0.500 and 0.236, respectively, which slightly increase, while SACC is 0.912, which slightly decreases. SAMAPE is an important indicator that reflects the effect of prediction. CBA-Net’s SAMAPE has

always been better than other methods. Although the increase in the prediction time interval will reduce the correlation coefficient and the accuracy of the prediction, this drawback can be well alleviated by adding a Bi-LSTM layer and an attention mechanism.

Figure 8 show the prediction results of different methods for 24 h SWH. The result confirmed our judgment once again, that a single CNN is not suitable for SWH prediction in time series. CBA-Net’s 24 h SWH prediction has large errors in only a small part of the area, but it is still an acceptable level, and the overall prediction effect is better. It also shows that CBA-Net’s 24 h SWH prediction is feasible.





**FIGURE 7** | Continuous prediction for twelve-hour prediction. **(A–C)** are the prediction results of the CNN model, the ERA5 data results, and the prediction errors, respectively. **(D–F)** are the prediction results of the CNN + Bi-LSTM model, the ERA5 data results, and the prediction errors, respectively. **(G–I)** are CBA-Net model prediction results, ERA5 data results, and prediction errors, respectively.

**TABLE 3** | One-day performance results.

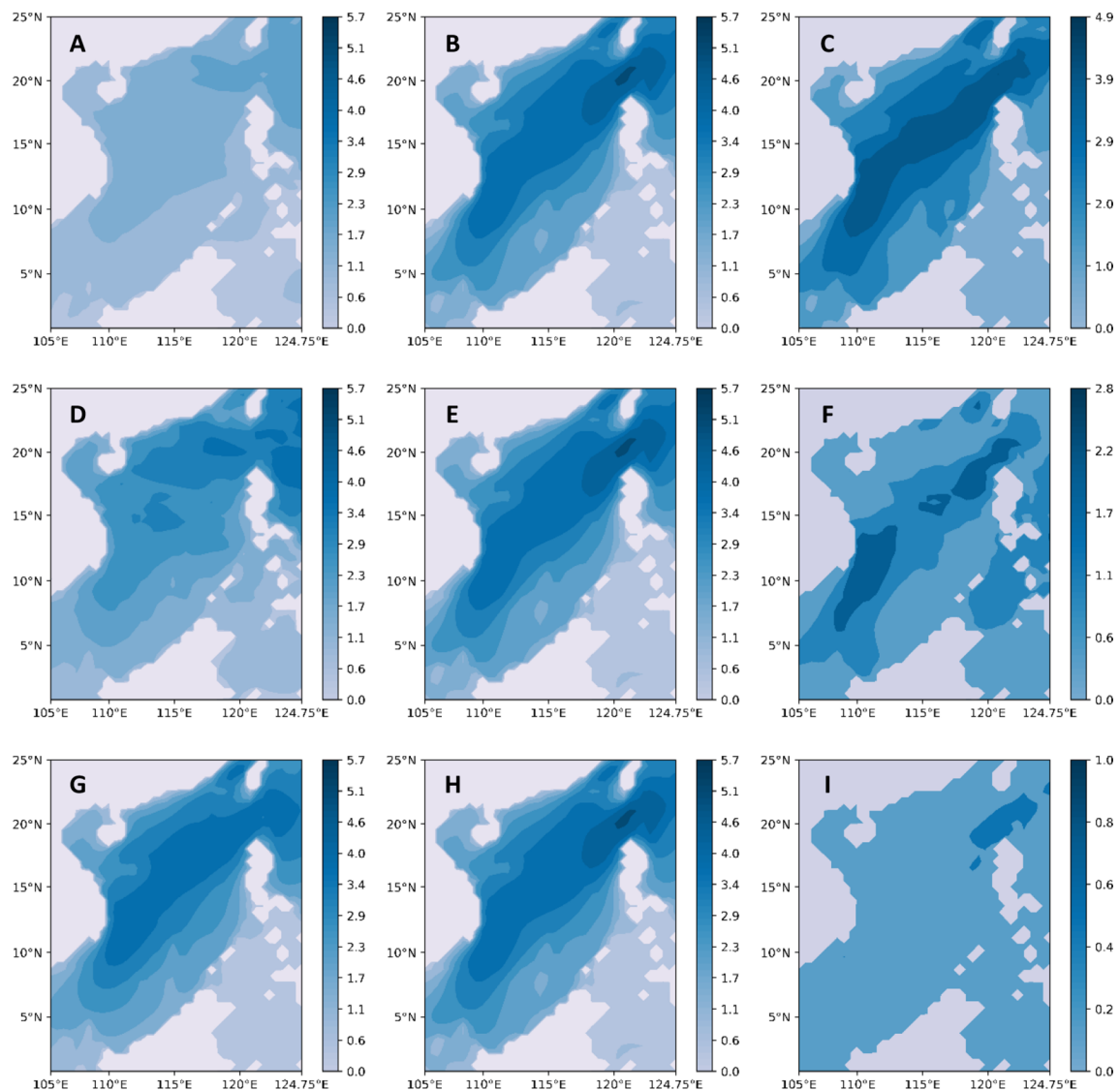
Metrics	CNN	CNN+Bi-LSTM	CNN+Bi-LSTM+ATTENTION
SARMSE	1.182	0.736	<b>0.500</b>
SAMAPE	0.390	0.336	<b>0.236</b>
SACC	0.707	0.820	<b>0.912</b>

## CONCLUSIONS

In this paper, a deep learning-based SWH prediction model for the South China Sea region is proposed and named CBA-Net.

The model is trained to predict SWH with U10, V10, and SWH of the ERA5 dataset as input. The model first uses the convolutional layer to find the local dependencies between the multi-dimensional input variables; then uses the Bi-LSTM layer to capture the complex long-term dependencies; finally, the attention mechanism is combined with the nonlinear neural network part to make the model have stronger robustness.

In order to prove the effectiveness of the method proposed in this paper, we used three different methods to predict the SWH in the South China Sea. We use the 2011–2018 ERA5 data set to train the model, and use the three indicators of SARMSE,



**FIGURE 8** | Continuous prediction for one-day prediction. (A–C) are the prediction results of the CNN model, the ERA5 data results, and the prediction errors, respectively. (D–F) are the prediction results of the CNN + Bi-LSTM model, the ERA5 data results, and the prediction errors, respectively. (G–I) are CBA-Net model prediction results, ERA5 data results, and prediction errors, respectively.

SAMAPE, and SACC to evaluate the accuracy and stability of the prediction results. The results show that the CBA-Net method can obtain more accurate results in the predictions of 1 h, 12 h, and 24 h. The ablation study also shows that each component of the method proposed in this paper is effective.

It can be seen that the SWH prediction technology based on CBA-Net can make full use of the important information of sea wind and significant wave height, establish a prediction model, and realize business applications. For future research, there are several promising directions to extend this work. Due to the complexity of the actual marine environment, it is a challenging task to extend the CBA-Net method to all global domains. In

addition, the number of input features directly determines the prediction results, such as wind speed, water depth, terrain, etc., which need to be considered and added to the training of the model. This general deep learning SWH prediction model deserves more attention in future work.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://cds.climate.copernicus.eu/>.

## AUTHOR CONTRIBUTIONS

PH designed the experiment. CY and GW collected and analyzed the data. SL provided critical revision of the article. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by the National Natural Science Foundation of China (Grant Number 41876003), the National

Key Research and Development Plan of China (Grant Number 2017YFA0604101), and the National Natural Science Foundation of China (Grant Number 41830533).

## ACKNOWLEDGMENTS

We would like to acknowledge the organizations that provided the sources of the data used in this work - namely, the European Centre for Medium-Range Weather Forecasts. We also thank the HPC Center of Zhejiang University (Zhoushan Campus) for computational support.

## REFERENCES

- Akpınar, A., Bingölbali, B., and Vledder, G. P. V. (2017). Long-Term Analysis of Wave Power Potential in the Black Sea, Based on 31-Year SWAN Simulations. *Ocean. Eng.* 130, 482–497. doi: 10.1016/j.oceaneng.2016.12.023
- Berbić, J., Ocvirk, E., Carević, D., and Lončar, G. (2017). Application of Neural Networks and Support Vector Machine for Significant Wave Height Prediction. *Oceanologia* 59 (3), 331–349. doi: 10.1016/j.oceano.2017.03.007
- Emmanouil, S., Aguilari, S. G., Nane, G. F., and Schouten, J.-J. (2020). Statistical Models for Improving Significant Wave Height Predictions in Offshore Operations. *Ocean. Eng.* 206, 107249. doi: 10.1016/j.oceaneng.2020.107249
- Fan, S., Xiao, N., and Dong, S. (2020). A Novel Model to Predict Significant Wave Height Based on Long Short-Term Memory Network. *Ocean. Eng.* 205, 107298. doi: 10.1016/j.oceaneng.2020.107298
- Gao, S., Huang, J., Li, Y., Liu, G., Bi, F., and Bai, Z. (2021). A Forecasting Model for Wave Heights Based on a Long Short-Term Memory Neural Network. *Acta Oceanol. Sin.* 40 (1), 62–69. doi: 10.1007/s13131-020-1680-3
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning* (MIT press).
- Group, T. W. (1988). The WAM Model—A Third Generation Ocean Wave Prediction Model. *J. Phys. Oceanogr.* 18 (12), 1775–1810. doi: 10.1175/1520-0485(1988)018<1775:TWMTO>2.0.CO;2
- Guo, M.-H., Xu, T.-X., Liu, J.-J., Liu, Z.-N., Jiang, P.-T., Mu, T.-J., et al. (2021). Attention Mechanisms in Computer Vision: A Survey[J]. *Computational Visual Media*. 2022, 1–38. doi: 10.1007/s41095-022-0271-y
- Hochreiter, S., and Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Comput.* 9 (8), 1735–1780. doi: 10.1162/neco.1997.9.8.1735
- Hsiao, S.-C., Chen, H., Wu, H.-L., Chen, W.-B., Chang, C.-H., Guo, W.-D., et al. (2020). Numerical Simulation of Large Wave Heights From Super Typhoon Nepartak, (2016) in the Eastern Waters of Taiwan. *J. Mar. Sci. Eng.* 8 (3), 217. doi: 10.3390/jmse8030217
- Huang, W., and Dong, S. (2021). Improved Short-Term Prediction of Significant Wave Height by Decomposing Deterministic and Stochastic Components. *Renewable Energy* 177, 743–758. doi: 10.1016/j.renene.2021.06.008
- Kazeminezhad, M. H., and Siadatmousavi, S. M. (2017). Performance Evaluation of WAVEWATCH III Model in the Persian Gulf Using Different Wind Resources. *Ocean. Dynam.* 67 (7), 839–855. doi: 10.1007/s10236-017-1063-2
- Kingma, D. P., and Ba, J. (2014). Adam: A Method for Stochastic Optimization. *LeCun, Y., and Bengio, Y. (1995). Convolutional Networks For Images, Speech, and Time Series[J]. The Handbook of Brain Theory and Neural Networks*. 3361, (10), 1995.
- Liang, B., Gao, H., and Shao, Z. (2019). Characteristics of Global Waves Based on the Third-Generation Wave Model SWAN. *Mar. Struct.* 64, 35–53. doi: 10.1016/j.marstruc.2018.10.011
- Li, N., Cheung, K. F., and Cross, P. (2020). Numerical Wave Modeling for Operational and Survival Analyses of Wave Energy Converters at the US Navy Wave Energy Test Site in Hawaii. *Renewable Energy* 161, 240–256. doi: 10.1016/j.renene.2020.06.089
- Li, S., Hao, P., Yu, C., and Wu, G. (2021). CLTS-Net: A More Accurate and Universal Method for the Long-Term Prediction of Significant Wave Height. *J. Mar. Sci. Eng.* 9 (12), 1464. doi: 10.3390/jmse9121464
- Lin, Y., Dong, S., Wang, Z., and Guedes Soares, C. (2019). Wave Energy Assessment in the China Adjacent Seas on the Basis of a 20-Year SWAN Simulation With Unstructured Grids. *Renewable Energy* 136, 275–295. doi: 10.1016/j.renene.2019.01.011
- Liu, Q., Rogers, W. E., Babanin, A. V., Young, I. R., Romero, L., Zieger, S., et al. (2019). Observation-Based Source Terms in the Third-Generation Wave Model WAVEWATCH III: Updates and Verification. *J. Phys. Oceanogr.* 49 (2), 489–517. doi: 10.1175/jpo-d-18-0137.1
- Mandal, S., and Prabaharan, N. (2006). Ocean Wave Forecasting Using Recurrent Neural Networks. *Ocean. Eng.* 33 (10), 1401–1410. doi: 10.1016/j.oceaneng.2005.08.007
- Méndez, F. J., Menéndez, M., Luceño, A., Medina, R., and Graham, N. E. (2008). Seasonality and Duration in Extreme Value Distributions of Significant Wave Height. *Ocean. Eng.* 35 (1), 131–138. doi: 10.1016/j.oceaneng.2007.07.012
- Miky, Y., Kaloop, M. R., Elnabwy, M. T., Baik, A., and Alshouny, A. (2021). A Recurrent-Cascade-Neural Network- Nonlinear Autoregressive Networks With Exogenous Inputs (NARX) Approach for Long-Term Time-Series Prediction of Wave Height Based on Wave Characteristics Measurements. *Ocean. Eng.* 240, 109958. doi: 10.1016/j.oceaneng.2021.109958
- Ni, C., and Ma, X. (2020). An Integrated Long-Short Term Memory Algorithm for Predicting Polar Westerlies Wave Height. *Ocean. Eng.* 215, 107715. doi: 10.1016/j.oceaneng.2020.107715
- Panchang, V., and Londhe, S. N. (2006). One-Day Wave Forecasts Based on Artificial Neural Networks. *J. Atmosph. Ocean. Technol.* 23 (11), 1593–1603. doi: 10.1175/jtech1932.1
- Raj, N., and Brown, J. (2021). An EEMD-BiLSTM Algorithm Integrated With Boruta Random Forest Optimiser for Significant Wave Height Forecasting Along Coastal Areas of Queensland, Australia. *Remote Sens.* 13 (8), 1456. doi: 10.3390/rs13081456
- Sadeghifar, T., Nouri Motlagh, M., Torabi Azad, M., and Mohammad Mahdizadeh, M. (2017). Coastal Wave Height Prediction Using Recurrent Neural Networks (RNNs) in the South Caspian Sea. *Mar. Geodesy.* 40 (6), 454–465. doi: 10.1080/01490419.2017.1359220
- Swain, J., Umesh, P., and Balchand, A. (2019). WAM and WAVEWATCH-III Intercomparison Studies in the North Indian Ocean Using Oceansat-2 Scatterometer Winds. *J. Ocean. Climate* 9, 2516019219866569. doi: 10.1177/2516019219866569
- Umesh, P. A., and Swain, J. (2018). Inter-Comparisons of SWAN Hindcasts Using Boundary Conditions From WAM and WWIII for Northwest and Northeast Coasts of India. *Ocean. Eng.* 156, 523–549. doi: 10.1016/j.oceaneng.2018.03.029
- Vanem, E. (2016). Joint Statistical Models for Significant Wave Height and Wave Period in a Changing Climate. *Mar. Struct.* 49, 180–205. doi: 10.1016/j.marstruc.2016.06.001
- Wu, L. (2021). Effect of Atmosphere-Wave-Ocean/Ice Interactions on a Polar Low Simulation Over the Barents Sea. *Atmosph. Res.* 248, 105183. doi: 10.1016/j.atmosres.2020.105183
- Wu, L., Breivik, Ø., and Rutgersson, A. (2019). Ocean-Wave-Atmosphere Interaction Processes in a Fully Coupled Modeling System. *J. Adv. Model. Earth Syst.* 11 (11), 3852–3874. doi: 10.1029/2019MS001761
- Wu, L., and Qiao, F. (2022). Wind Profile in the Wave Boundary Layer and Its Application in a Coupled Atmosphere-Wave Model. *J. Geophys. Res.: Ocean.* 127 (2), e2021JC018123. doi: 10.1029/2021JC018123
- Yang, X. (2020). “An Overview of the Attention Mechanisms in Computer Vision,” in *Journal of Physics: Conference Series* (IOP Publishing), 012173.

- Young, I. R., and Ribal, A. (2019). Multiplatform Evaluation of Global Trends in Wind Speed and Wave Height. *Science* 364 (6440), 548–552. doi: 10.1126/science.aav9527
- Zaremba, W., Sutskever, I., and Vinyals, O. (2014). Recurrent Neural Network Regularization. *ArXiv. Prepr. ArXiv.* 1409, 2329. doi: 10.48550/arXiv.1409.2329
- Zhang, X., Li, Y., Gao, S., and Ren, P. (2021). Ocean Wave Height Series Prediction With Numerical Long Short-Term Memory. *J. Mar. Sci. Eng.* 9 (5), 514. doi: 10.3390/jmse9050514
- Zhou, S., Xie, W., Lu, Y., Wang, Y., Zhou, Y., Hui, N., et al. (2021). ConvLSTM-Based Wave Forecasts in the South and East China Seas. *Front. Mar. Sci.* 8. doi: 10.3389/fmars.2021.680079

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer MH declared a shared affiliation with the author GW to the handling editor at the time of review.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright © 2022 Hao, Li, Yu and Wu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*