



Article

Short-Term Prediction Model of Wave Energy Converter Generation Power Based on CNN-BiLSTM-DELA Integration

Yuxiang Zhang ^{1,*} , Shihao Liu ¹ , Qian Shen ¹ , Lei Zhang ¹, Yi Li ¹, Zhiwei Hou ¹ and Renwen Chen ²

¹ Faculty of Automation, Huaiyin Institute of Technology, Huaian 223003, China; 212309601013@hyit.edu.cn (S.L.); qianshen@hyit.edu.cn (Q.S.); 212309611057@hyit.edu.cn (L.Z.); 212309611056@hyit.edu.cn (Y.L.); zhiweih@hyit.edu.cn (Z.H.)

² State Key Laboratory of Mechanics and Control of Mechanical Structures, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China; rwchen@nuaa.edu.com

* Correspondence: yxzhang@hyit.edu.cn

Abstract: Wave energy is a promising source of sustainable clean energy, yet its inherent intermittency and irregularity pose challenges for stable grid integration. Accurate forecasting of wave energy power is crucial for reliable grid management. This paper introduces a novel approach that utilizes a Bidirectional Gated Recurrent Unit (BiGRU) network to fit the power matrix, effectively modeling the relationship between wave characteristics and energy output. Leveraging this fitted power matrix, the wave energy converter (WEC) output power is predicted using a model that incorporates a Convolutional Neural Network (CNN), a Bidirectional Long Short-Term Memory (BiLSTM) network, and deformable efficient local attention (DELA), thereby improving the accuracy and robustness of wave energy power prediction. The proposed method employs BiGRU to transform wave parameters into power outputs for various devices, which are subsequently processed by the CNN-BiLSTM-DELA model to forecast future generation. The results indicate that the CNN-BiLSTM-DELA model outperforms BiLSTM, CNN, BP, LSTM, CNN-BiLSTM, and GRU models, achieving the lowest mean squared error (0.0396 W) and mean absolute percentage error (3.7361%), alongside the highest R^2 (98.69%), underscoring its exceptional forecasting accuracy. By enhancing power forecasting, the method facilitates effective power generation dispatch, thereby mitigating the adverse effects of randomness on the power grid.



Citation: Zhang, Y.; Liu, S.; Shen, Q.; Zhang, L.; Li, Y.; Hou, Z.; Chen, R. Short-Term Prediction Model of Wave Energy Converter Generation Power Based on CNN-BiLSTM-DELA Integration. *Electronics* **2024**, *13*, 4163. <https://doi.org/10.3390/electronics13214163>

Academic Editor: Fabio Corti

Received: 8 October 2024

Revised: 17 October 2024

Accepted: 18 October 2024

Published: 23 October 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: attention mechanism; convolutional neural network; BiLSTM; power matrix; prediction; wave energy converter

1. Introduction

The oceans cover 71% of the Earth's surface, and wave energy is a major part of ocean energy, with many advantages, including large reserves, high power density, and low pollution. In the 20th century, the large-scale use of fossil fuels, such as coal and oil, has led to global warming and the depletion of natural resources. Consequently, renewable energy, such as wave power, has come into focus [1]. The global potential output of ocean wave energy is calculated to be 337 gigawatts [2], with an energy density several times higher than traditional green energies such as solar and wind power [3]. Against this backdrop, interest in harnessing electricity from waves using wave energy converters has surged. Wave energy has become a formidable contender among renewable energies, playing a crucial role in the industrial development of coastal cities. However, weather, seasons, sea surface temperature, and atmospheric pressure affect the output power of wave energy devices, leading to instability. The resulting randomness and intermittency pose challenges to the stable operation of the power grid [4,5]. Therefore, accurate power output prediction is of great significance for ensuring the safe and stable operation of wave energy devices once integrated into the grid.

Currently, researchers have conducted extensive studies on the prediction of wave energy converter, with the primary technical approaches categorized into three main types [6]: physical modeling [7], statistical methods [1,8–10], and machine learning techniques [6,11–13]. Among them, physical modeling methods rely on validated physical models to establish mathematical models without depending on extensive historical data, making them suitable for large-scale wave energy predictions. For example, Zheng et al. [14] developed a wave forecast product for Chinese seas based on the WAVEWATCH-III (WW3) physical model. Similarly, Shao et al. [15] developed a hexagonal array wave energy converter product based on wave power and predicted wave power generation efficiency using physical models.

Statistical methods use parameter estimation and historical data curve fitting to establish predictive models, offering robust, versatile, and fast modeling techniques. Wu et al. [16] employed an Auto Regressive Integrated Moving Average (ARIMA) model to predict wave heights. However, the effectiveness of statistical methods is closely related to the quality of historical data. During the actual sampling process, some of the historical wave data may be missing or incorrect, resulting in significant degradation of the model performance. For this purpose, Discrete Wavelet Transform (DWT) is combined with ARMA to predict the true output power.

With advances in computer science and hardware, machine learning techniques are increasingly recognized for their advantages in renewable energy prediction. Forms of artificial intelligence and machine learning methods can be autonomous in learning the association between input features and output results [17]. Accurate predictions can be achieved by using deep learning methods on target objects. Yan et al. [5,18] reduced network energy consumption and minimized error through behavioral critique and deep reinforcement learning. These methods possess strong resilience to interference and high prediction accuracy and are now widely applied in solar [19] and wind energy [18,20] forecasting. The development of ocean observation technology has led to the availability of abundant datasets of sea waves and oceanic meteorological data, providing a foundation for machine learning-based wave energy prediction. For instance, Elbisy et al. [21] used a Support Vector Machine (SVM) optimized by the Licorne algorithm to predict wave parameters, demonstrating strong generalization ability and minimal prediction error. Feng et al. [16] compared the performance of Recurrent Neural Networks (RNNs), Gated Recurrent Unit (GRU) networks, and Long Short-Term Memory (LSTM) networks in predicting wave parameters, finding that LSTM and GRU models significantly outperformed traditional RNN models. Additionally, Yang et al. [22] applied a Convolutional Neural Network (CNN) combined with Seasonal-Trend Decomposition (STL) and Positional Encoding (PE) to predict significant wave heights. Building on these findings, this paper proposes a short-term wave power prediction method that integrates CNN, BiLSTM, and Deformable Efficient Local Attention (DELA) mechanisms, as well as a power-fitting matrix based on BiGRU. The main contributions of this paper include the following:

1. Designing a BiGRU power fitting matrix that can convert wave parameters into wave energy power output using the wave power generation matrix obtained from simulation software when wave information is input.
2. Utilizing CNN and BiLSTM to extract feature maps from multidimensional inputs, which are derived from time series data.
3. Designing a new attention mechanism to enhance the feature extraction capability of BiLSTM, and comparing it with seven mainstream attention mechanisms, with results showing that the DELA attention mechanism has strong feature extraction capabilities.
4. Comparing the proposed CNN-BiLSTM-DELA model with other established wave prediction models and generalization experiments using January–June 2024 data, and the results show that the model proposed in this paper outperforms the benchmark model in wave power prediction.

2. Power Conversion Module

This chapter analyzes the relationship between significant wave height, wave period, and power through simulation software, thereby constructing the power matrix for the point absorber wave energy converter. The BiGRU model was used for fitting to achieve high-accuracy power prediction.

2.1. Mathematical Model of Wave Energy Converter Device

This study utilizes a point absorber wave energy converter device, the structure of which is shown in Figure 1. The floater is rigidly linked to a permanent magnet synchronous motor, and both perform synchronous heaving motion, with the motor rotor cutting through the magnetic field to generate electricity.

According to Newton’s second law, analyzing the vertical forces on the direct-drive wave energy converter system allows for the derivation of the floater’s time-domain hydrodynamic model:

$$m\ddot{x}(t) = F_d(t) + F_r(t) + F_s(t) + F_g(t). \tag{1}$$

In the equation, m is the total mass of the system’s moving parts; \ddot{x} is the heave acceleration of the floater; $F_d(t)$ is the wave excitation force; $F_r(t)$ is the radiation force; $F_s(t)$ is the hydrostatic restoring force; and $F_g(t)$ is the counter-electromotive force.

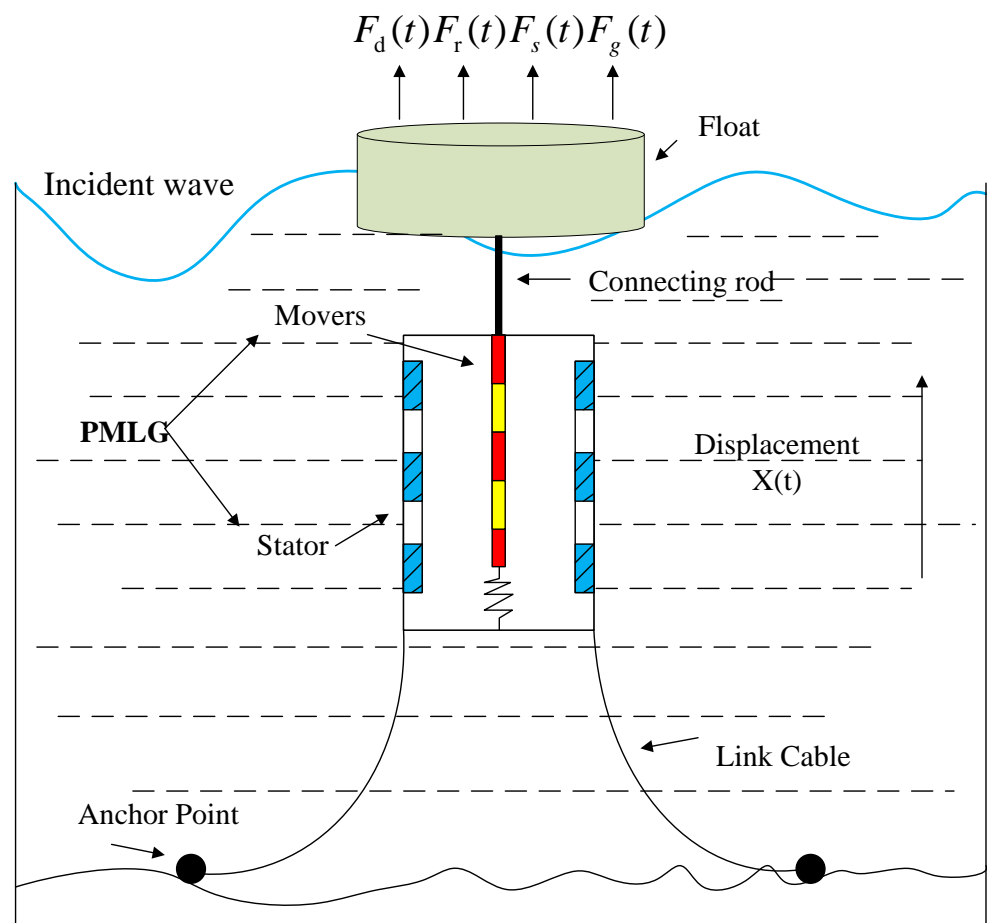


Figure 1. Direct drive generator patterns in CNN-BILSTM-DELA based short-term prediction model structure.

When the float is in an equilibrium position, it can be expressed as:

$$F_s = Kz(t) = \rho g \pi r^2 x(t). \tag{2}$$

In the equation, r is the float radius; K is the hydrostatic spring coefficient of the float. The radiation force can be expressed as:

$$F_r = m_a \ddot{x}(t) + B \dot{x}(t). \tag{3}$$

In the equation, m_a is the additional mass of the system in the infinite frequency domain, and B is the damping coefficient caused by the radiation force. The wave excitation force and radiation force can be expressed as:

$$\begin{cases} F_d = i\omega\rho \iint_S (\phi_I + \phi_D) n dS. \\ F_r = i\omega\rho \iint_S \phi_R n dS. \end{cases} \tag{4}$$

In the equation, ρ is the density of seawater; n is the unit vertical normal of the floater; S is the wetted surface area of the floater; and Φ_R, Φ_I , and Φ_D are the incident wave radiation potential, incident wave velocity potential, and incident wave diffraction potential, respectively. The wave excitation force F_d is the sum of the Froude–Krylov force and diffraction force derived from the incident and diffraction potentials. The electromagnetic damping force acting on the floater can be expressed as:

$$F_g(t) = R_g \dot{x}(t). \tag{5}$$

In the equation, R_g is the electromagnetic damping coefficient.

By substituting (2) to (5) into (1), the motion equation of the entire system can be expressed as:

$$M \ddot{x}(t) + B \dot{x}(t) + Kx(t) + F_g(t) = F_d(t). \tag{6}$$

In the equation, $M = m + m_a$. The average electromagnetic power \bar{p} generated by the direct driven point absorption wave energy converter system can be calculated by the following formula:

$$\bar{P} = \frac{1}{T} \int_0^T (R_d \dot{x}^2) dt. \tag{7}$$

where T is the wave period.

The three power generation powers generated by the permanent magnet linear generator can be calculated using the following formula:

$$P_{gen} = u_a i_a + u_b i_b + u_c i_c. \tag{8}$$

In the equation, u_a, u_b , and u_c are the three-phase output voltages generated by the permanent magnet linear generator; i_a, i_b , and i_c are the three-phase currents generated.

The average power generation can be calculated using the following formula:

$$\bar{P} = \frac{1}{T} \int_0^T P_{gen} dt. \tag{9}$$

2.2. Pearson Correlation Analysis of Factors Affecting Wave Energy Converter

Factors affecting power generation include significant wave heights, seawater temperatures, sea surface temperatures, wind speeds, atmospheric pressure, and wave periods. To determine the input variables for the model, the Pearson correlation coefficient is used to establish the correlation between the power generation of the wave energy converter and other factors. The formula is:

$$r_{x,y} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}}. \tag{10}$$

In the equation, x_i and y_i are the variables for correlation analysis, \bar{x} and \bar{y} are the means of these variables, n is the sample size, x is wave height, and y represents other factors besides wave height.

A heatmap of the wave feature correlations is shown in Figure 2. Power generation (power) has the highest correlation with significant wave height (swh) and mean wave period (mwp). It also has a correlation greater than 0.4 with the wave drag coefficient (cdww), air density above the ocean (p140209), and mean sea level pressure (msl), while other factors are negatively correlated. Therefore, swh and mwp are selected as input features for the prediction model.

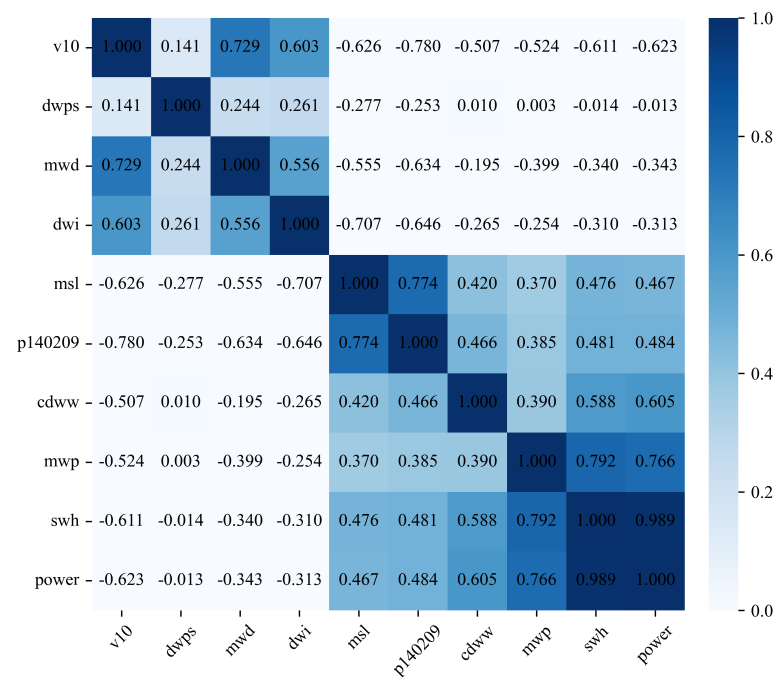


Figure 2. Wave feature correlation analysis heatmap.

2.3. Simulation Model Establishment

Wave energy converter devices generally consist of three energy conversion parts: the wave energy capture system, the mechanical transmission system, and the generator [23]. However, as this study uses a direct-drive wave energy generator, it eliminates the energy loss associated with the intermediate mechanical transmission part. To obtain the power conversion matrix for the direct-drive point absorber wave energy converter device, a simulation approach combining ANSYS AQWA and COMSOL was used, with the floater simulation model established first, as shown in Figure 3a. The capture device has an outer diameter of 6 m, a height of 1.6 m, a draft of 0.68 m in a stationary state, and a model weight of 17 tons.

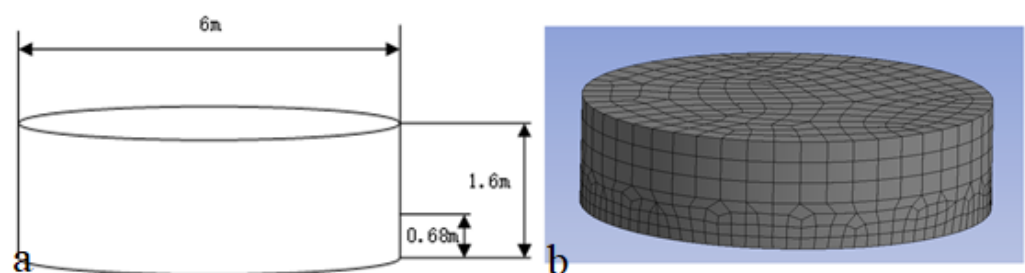


Figure 3. Floating body: (a) Floating body size. (b) Simulation network diagram.

In the modeling process, the X-Y plane of the global coordinate system was aligned with the still water surface, defining the waterline height as 0.68 m below and 0.92 m above the waterline. The ANSYS AQWA simulation mesh tool was subsequently employed to generate separate grid elements for regions below and above the waterline. Given that AQWA calculations predominantly occur beneath the waterline, the maximum mesh size was set to 0.4 m above the waterline and 0.2 m below it, leading to a total of 3015 grid elements, as illustrated in Figure 3b. The wave conditions were defined as regular waves, with Figure 4a depicting the response curves of the floater subjected to waves of varying amplitudes and periods. Subsequently, a direct-drive generator model was developed in COMSOL, utilizing the response curves as input for the generator rotor. Finite element simulations were then conducted to derive the three-phase voltages. Figure 4b displays the three-phase voltage waveforms generated by COMSOL, and the power output within a cycle was calculated using (9), as detailed in Table 1.

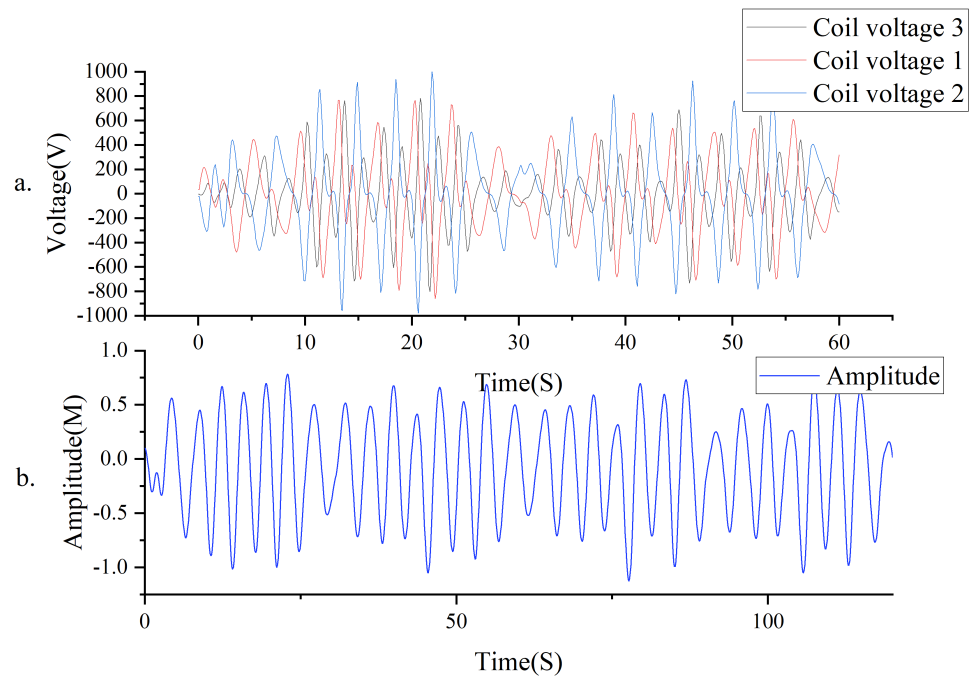


Figure 4. (a) The three-phase voltage output generated by COMSOL for a wave amplitude of 1 m and a period of 4 s. (b) The vertical displacement amplitude of the floater calculated by ANSYS AQWA for a wave period of 4 s and an amplitude of 1 m.

Table 1. Power conversion matrix of the wave energy converter system (W).

Wave Height (m)	Wave Period (s)												
	4	5	6	7	8	9	10	11	12	13	14	15	16
1.0	1.2	1.3	1.2	1.2	1.1	1.0	0.9	0.8	0.7	1.7	0.7	1.6	1.7
1.5	2.6	2.5	2.3	2.2	2.3	2.0	1.9	1.7	1.4	1.5	1.2	1.2	1.2
2.0	4.4	4.0	3.7	3.6	3.5	3.1	2.8	2.5	2.3	2.2	2.0	1.8	1.7
2.5		4.0	5.2	4.5	4.6	4.3	3.9	3.6	3.0	2.8	2.5	2.7	2.6
3.0		7.4	6.7	6.2	5.7	5.4	4.7	4.1	4.1	3.7	3.3	3.3	3.2
3.5			8.4	7.3	6.9	5.8	5.4	4.9	4.4	4.2	3.7	3.4	3.6
4.0			8.9	8.6	7.6	6.8	6.2	5.6	5.0	4.6	4.5	4.3	3.6

Table 1. Cont.

Wave Height (m)	Wave Period (s)												
	4	5	6	7	8	9	10	11	12	13	14	15	16
4.5			10.6	9.5	8.7	7.6	7.0	6.1	5.9	5.4	5.1	5.0	4.7
5.0			12.2	10.8	9.2	8.6	7.3	7.2	6.3	5.9	5.7	5.4	5.0
5.5				11.1	10.1	8.9	8.1	7.5	6.8	6.4	6.1	5.5	5.8
6.0				13.1	11.3	10.1	9.1	8.3	7.5	6.7	6.9	6.4	5.8
6.5				13.5	11.6	10.4	9.8	9.0	7.6	7.3	7.5	6.2	6.4
7.0				15.0	12.9	10.9	10.0	8.8	8.6	8.2	7.6	7.3	6.8

2.4. Power Conversion Matrix

The power conversion matrix can be obtained by performing simulation experiments through the above steps, as shown in Table 1. The unit of power is watts (W). Table 1 illustrates the relationship between the wave energy conversion system, significant wave height, wave period, and power generation, where the empty entry indicates that the wave height exceeds the maximum value that can be calculated by the simulation software at this period and cannot be calculated. The relationship between these three variables can be observed more visually in Figure 5.

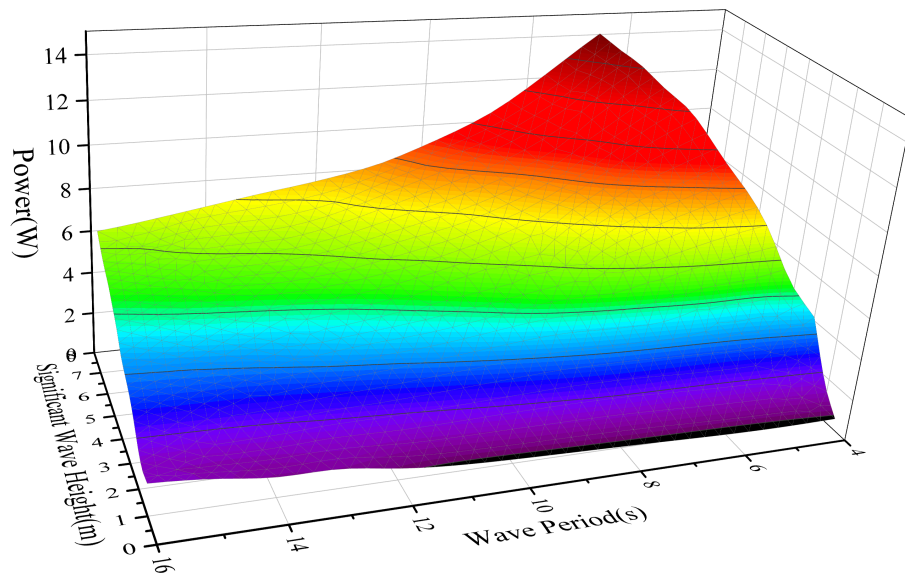


Figure 5. Mathematical relationship between wave period, significant wave height, and power.

As shown in Figure 5, the relationship between the three variables is discrete, and when a polynomial is used for fitting, there are 146 discrete parameters in the table, which are fitted with a large error and do not accurately reflect the actual power generated by the current equipment. In this study, a deep learning method is used to fit the power conversion matrix.

2.5. Power Matrix Fitting Model

In this study, a Bidirectional Gated Recurrent Unit (BiGRU), a variant of Recurrent Neural Networks (RNN), was employed to address issues such as gradient vanishing or explosion when processing long sequence data. It also demonstrated superior performance in scenarios with limited data. By introducing update and reset gates, the BiGRU effectively enhanced accuracy under small data conditions. The core concept of the GRU is to use update and reset gates at each time step to determine which information should be passed to the next time step.

The equations for the BiGRU model are presented as follows [24]:

$$\begin{cases} z_t = \sigma(W_z \cdot [h_{t-1}, x_t]). \\ r_t = \sigma(W_r \cdot [h_{t-1}, x_t]). \\ \tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t]). \\ h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t. \end{cases} \quad (11)$$

In the equations, z_t represents the update gate, r_t denotes the reset gate, \tilde{h}_t is the candidate activation value, and h_t is the output, where σ denotes the sigmoid activation function. The BiGRU further enhances GRU information processing capabilities by simultaneously handling both forward and backward information. Compared to traditional RNN models, it significantly improves prediction accuracy. The structure of the BiGRU is illustrated in Figure 6.

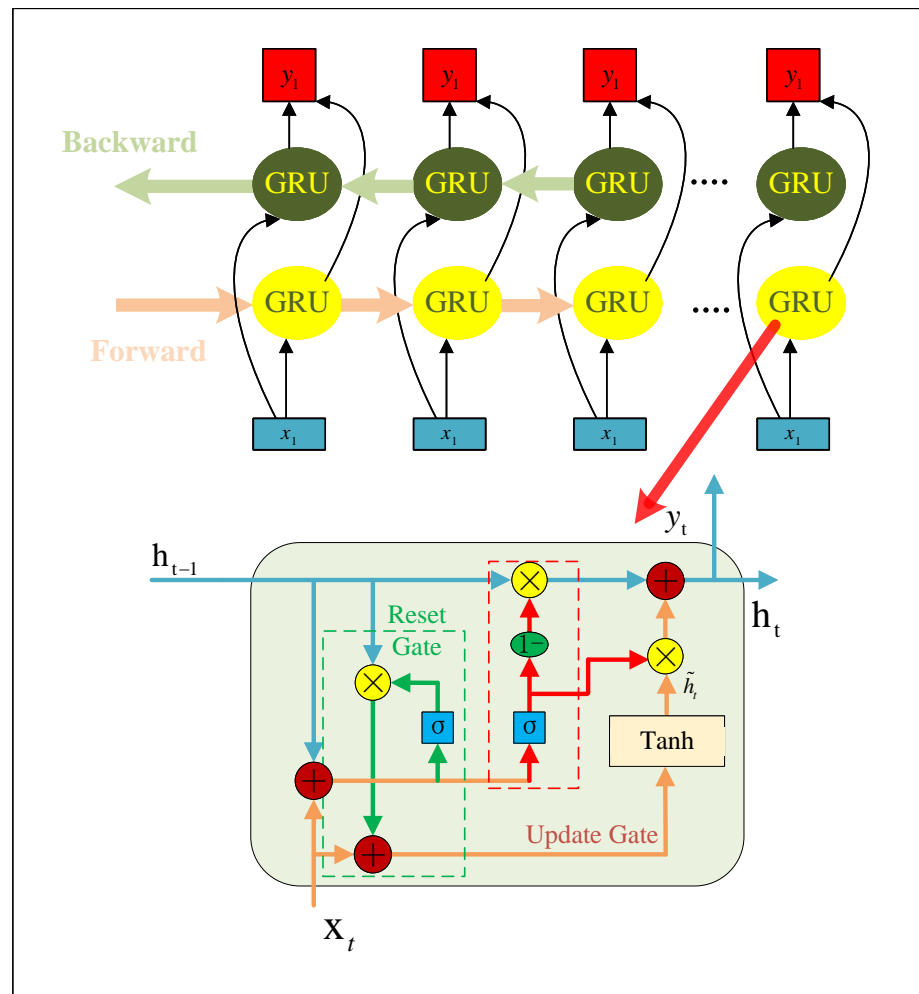


Figure 6. Structure of the Bidirectional Gated Recurrent Unit network.

The calculation formulas are as follows [25]:

$$\begin{cases} \vec{h}_t = GRU(X_t, \vec{h}_{t-1}). \\ \overleftarrow{h}_t = GRU(X_t, \overleftarrow{h}_{t-1}). \\ h_t = \omega_t \vec{h}_t + \nu_t \overleftarrow{h}_t + b_t. \end{cases} \quad (12)$$

In the equations, \vec{h}_t and \overleftarrow{h}_t represent the forward and backward outputs of the hidden layer at time t , respectively; ω_t and ν_t denote the weights for the forward and backward state layers; and b_t represents the bias term. Figure 7 illustrates the predictions obtained from the BiGRU model when the relationships between significant wave height, wave period, and the effective power of the wave energy conversion device are input. The blue curve represents the actual values, while the red triangles indicate the predicted values for each point.

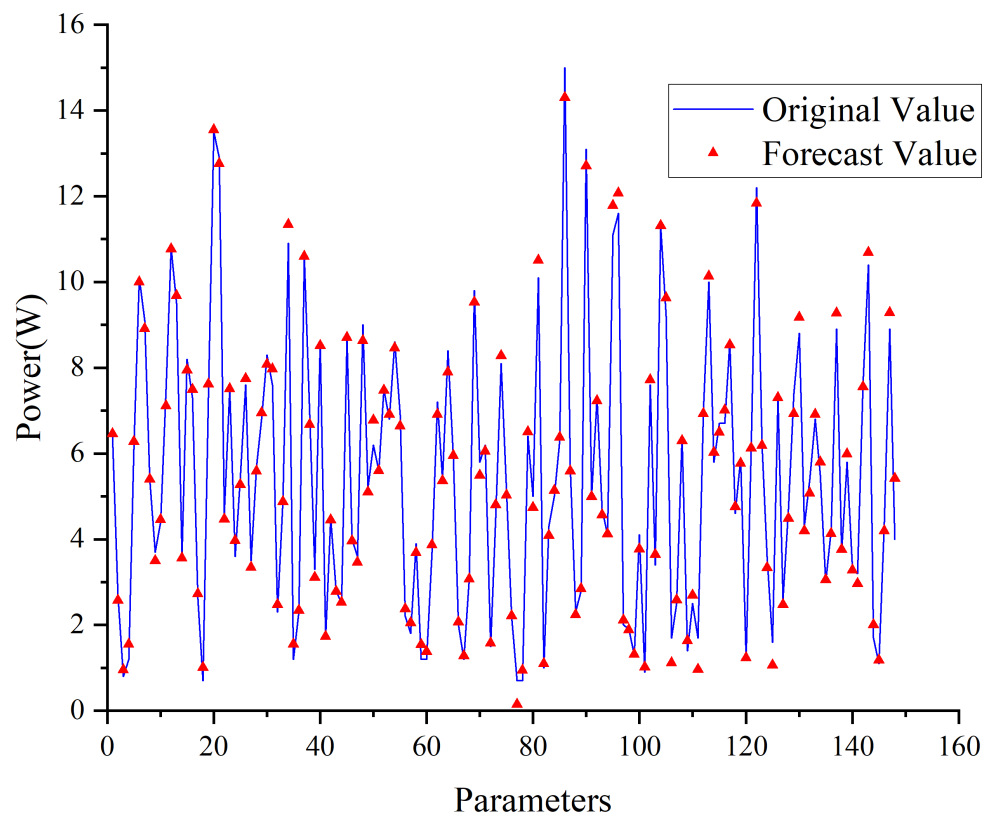


Figure 7. Wave energy conversion matrix fitting.

3. Wave Energy Converter Power Prediction Model Based on CNN-BiLSTM-DELA

In the previous section, the BiGRU model was utilized to fit the power conversion matrix of the point absorber wave energy device. This section further explored the optimization of the power prediction method. A CNN-BiLSTM-DELA model incorporating a deformable efficient local attention (DELA) was proposed. This model integrated CNN and BiLSTM, enhancing the ability of BiLSTM to identify nonlinear local features through the DELA attention mechanism. The model took time series as input. Since parameters such as significant wave height and wave period were independent data sequences, the fitted wave energy parameters were input into the power prediction module. The power at each time step was represented by the associated wave factors.

In the CNN-BiLSTM-DELA prediction model, the CNN and BiLSTM-DELA modules were configured in a parallel structure. Wave parameters were separately input into the two frameworks. After a series of transformations, features from both modules were fused,

and the final prediction for the wave energy converter device was obtained through a fully connected layer. A structure diagram is shown in Figure 8.

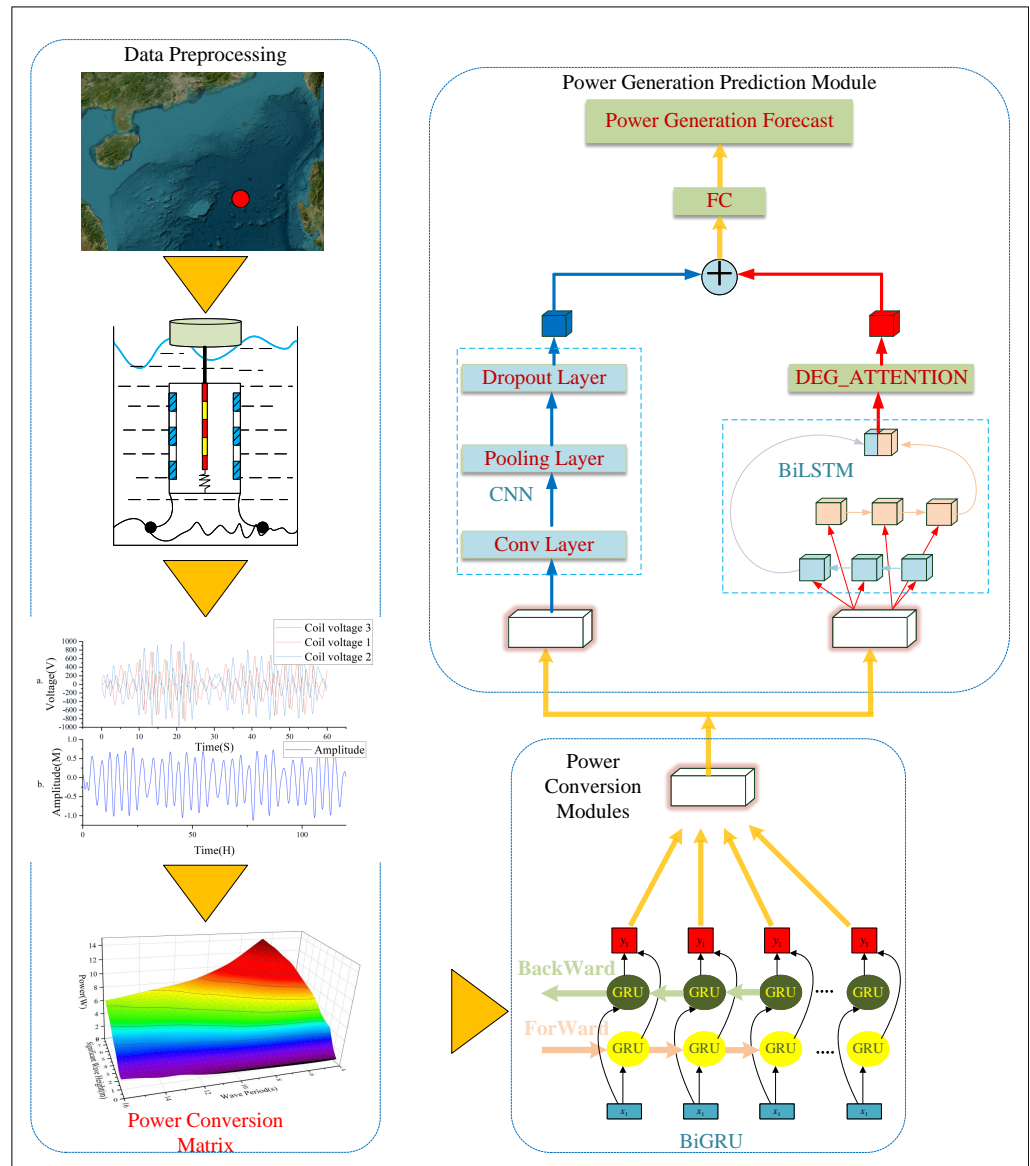


Figure 8. The overall operation process in the short-term prediction model structure based on CNN-BiLSTM-DELA. The red dot in the upper left corner is the selected position of the experimental target data set, which is located at 116° E and 16° N.

3.1. Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are a type of feedforward neural network designed to extract features from data. They primarily consist of an input layer, convolutional layers, pooling layers, and dropout layers. The convolutional layers perform convolution operations between the input and weights, and the formula for this operation is as follows:

$$y_1 = \sigma(b_1 + w * x_0). \tag{13}$$

In the equation, y_1 represents the output of the convolutional layer; b_1 denotes the bias; w stands for the CNN module weights; and x_0 is the input to the convolutional layer.

The pooling layer optimizes the input from the convolutional layer by reducing its dimensionality, enabling effective coupling with the features' output by BiLSTM-DELA. Dropout, on the other hand, involves randomly removing a portion of neurons during

training. Through backpropagation, the weights of the removed neurons are updated and retrained, helping to prevent overfitting in the model.

3.2. Bidirectional Long Short-Term Memory Networks

LSTMs are also a variant of Recurrent Neural Networks (RNNs), designed to address the limitations of RNNs in predicting long sequences. RNNs are constrained by their structure, which can lead to issues such as vanishing and exploding gradients during backpropagation, and they can learn only short-term dependencies [26]. To overcome these limitations, LSTMs introduce memory cells, which consist of a cell state c_t , an input gate i_t , an output gate o_t , and a forget gate f_t . The structure of the LSTM is illustrated in Figure 9.

In the figure, x_t represents the input at time t , while h_t and h_{t-1} denote the hidden states at time t and $t - 1$, respectively. Here, h_t is computed based on the output of the previous hidden state and the current input. When x_t is fed into the LSTM, it interacts with the previous hidden state h_{t-1} through the forget gate, input gate, and output gate. The formulas for the input gate, output gate, forget gate, and hidden state are as follows :

$$\begin{cases} f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f). \\ i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i). \\ o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o). \\ h_t = o_t \otimes \tanh(C_t). \end{cases} \quad (14)$$

In the equations, W represents the weight matrix, U denotes the output matrix, and b is the bias vector. The subscripts f , i , and o correspond to the forget gate, input gate, and output gate, respectively. The symbol \otimes indicates element-wise multiplication.

The cell state c_t is the core component of the LSTM, represented by the horizontal line at the top of Figure 9. It features a minimal branch conveyor belt structure design, allowing the input information to flow through the cell with minimal alterations, and is regulated by the input, output, and forget gates. The formulas for the candidate cell state and cell state are as follows:

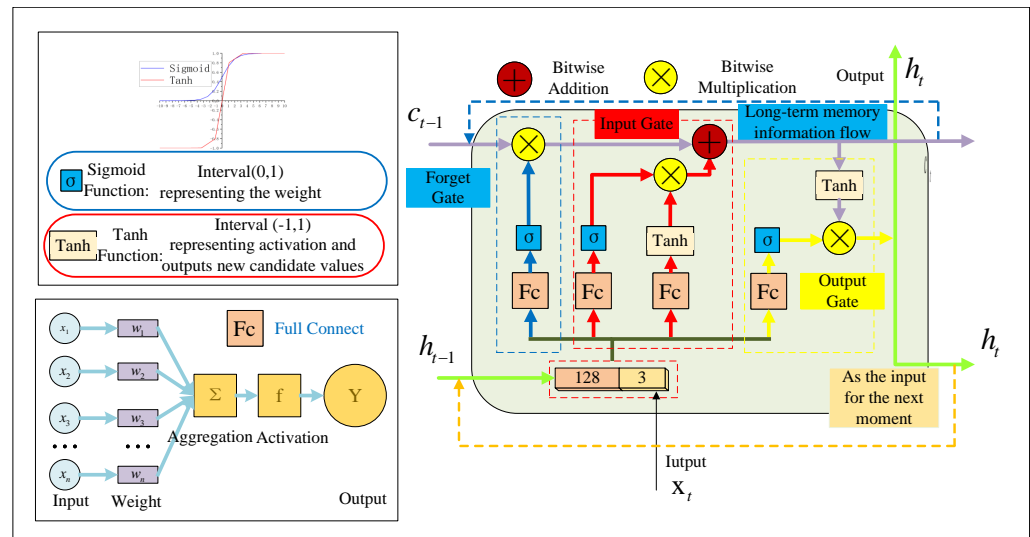


Figure 9. Structure of the Long Short Term Memory network.

$$\begin{aligned} \tilde{C}_t &= \tanh(W_c x_t + U_c h_{t-1} + b_c). \\ C_t &= f_t \otimes C_{t-1} + i_t \otimes \tilde{C}_t. \end{aligned} \quad (15)$$

To enhance the nonlinearity of the network, the sigmoid and tanh functions are chosen as activation functions. Their formulas are as follows:

$$\sigma(x) = \frac{1}{1 + \exp(-x)}. \tag{16}$$

$$\tanh(x) = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)}. \tag{17}$$

In time series prediction tasks, it is essential to fully consider both forward and backward temporal information patterns to significantly enhance prediction accuracy. The CNN-BiLSTM-DELA model primarily consists of three forward LSTMs and three backward LSTMs. Unlike standard LSTM, which transmit states unidirectionally from past to future, the BiLSTM framework incorporates both forward and backward data patterns, demonstrating significantly enhanced performance.

As illustrated in Figure 9, the BiLSTM model comprises both forward and backward computations. The orange horizontal arrows represent the forward flow of time series information within the model, while the light blue arrows denote the backward flow of the same information. Additionally, the data information flows unidirectionally through the input layer, hidden layer, and output layer.

3.3. Attention Mechanism

The attention mechanism mimics the human brain’s focus on specific regions at particular moments, allowing for the selective acquisition of more pertinent information while disregarding irrelevant data [27]. It achieves this by assigning different probabilistic weights to the hidden layer units of the neural network, thereby emphasizing the impact of critical information and improving the accuracy of the model’s predictions. To address this issue, this paper proposes a novel Deformable Efficient Local Attention (DELA) mechanism to tackle model overfitting caused by extended input sequences, which impedes the accurate learning of appropriate weights. DELA is a three-channel attention mechanism comprising spatial attention modules, channel attention modules, and local attention modules, as depicted in Figure 10.

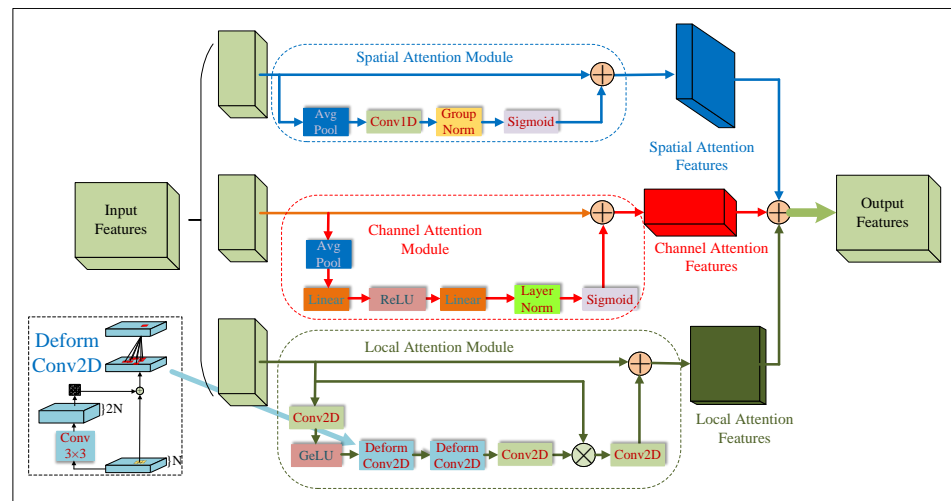


Figure 10. Attention Mechanism in the Structure of Short-term Prediction Model Based on CNN-BiLSTM-DELA.

3.3.1. Spatial Attention Module

Given that the wave energy parameters involve long temporal sequences, identifying relevant features within these sequences is challenging when relying solely on LSTM. Consequently, a spatial attention module is introduced, which primarily focuses on weighting features across extended time steps. It also determines the contribution of different features at each time step to the power output of the generator. Through average pooling, all feature values across the time dimension are averaged to obtain global information about each

feature across the entire time series. This step outputs a vector of size n , representing the average value of each feature in the time series. Subsequently, a one-dimensional convolutional layer (Conv1D) generates a temporal attention map of the same size as the input features. This map is used to weight each time step according to the importance of different features at that step. Finally, a sigmoid function is applied to normalize the generated temporal attention map. The module's formula is as follows:

$$y_{Si} = x_i + \sigma(G_n(W_1(x_i))). \quad (18)$$

In the equation, G_n denotes the GroupNorm operation, while W_1 represents the Conv1D convolution operation. The spatial attention features are derived from the equation. The spatial attention mechanism identifies which features at specific time steps hold higher priority, thereby enhancing the ability of the model to capture temporal sequences more effectively.

3.3.2. Channel Attention Module

Predicting wave energy power involves handling various parameters such as wave period and significant wave height. To enhance the ability of the model to extract the relative importance of different features, this study introduces a channel attention mechanism. The mechanism determines the contribution of each parameter to the overall model and applies appropriate weighting to these parameters. Initially, global pooling is applied to obtain the global average of each feature channel, providing an overall representation of each feature across the entire time series. The input is first reduced by a factor of 1/8 through two linear layers and then restored. This dimensionality reduction and expansion process effectively captures the nonlinear relationships between feature channels. Finally, a sigmoid function is employed to generate the weights for each feature channel. The formula is as follows:

$$y_{Di} = \mathbf{x}_i + \sigma(W_2 \text{LN}(W_1 \mathbf{x}_i)). \quad (19)$$

In the equation, LN denotes the LayerNorm operation, while W_1 and W_2 represent two linear layers, with W_1 performing dimensionality reduction and W_2 performing dimensionality expansion. The channel attention mechanism highlights the most critical features in the prediction task while reducing the influence of irrelevant features. This approach enhances the ability of the model to accurately capture valuable information when dealing with complex temporal data, thereby improving prediction accuracy and robustness.

3.3.3. Local Attention Module

In the processing of temporal data, recognizing the importance of long sequences is crucial; however, fine-grained attention to the combinations of specific time steps and features is equally important. Therefore, a local attention module is proposed to capture the interrelationships among significant wave height, wave period, and power generation across different time steps. Initially, an offset generation network is used to predict the offsets in convolution along both the time and feature dimensions, enabling dynamic capture of significant changes and anomalies at specific time steps. These offsets are generated through Conv2D, followed by a deformable convolution operation. As shown by the arrow in Figure 10.

Unlike standard convolution, deformable convolution allows the kernel to dynamically adjust sampling locations along both the time and feature axes, thus capturing finer temporal features. When combinations of significant wave height and wave period at specific time steps significantly impact power generation, the local attention mechanism can focus specifically on these combinations, thereby enhancing the capture of critical time steps and features. Finally, a standard convolution layer integrates the local features to generate the final output feature, as illustrated by the following formula:

$$y_{Li} = \mathbf{x}_i + W_3(x_i \otimes (W_2 D W_1 D W_2(W_1 \mathbf{x}_i))). \quad (20)$$

In the equation, DW denotes DeformConv, with subscripts indicating the first and second dimensions. The equation highlights the importance of weighting features to emphasize critical information regarding time steps and feature combinations, especially in the context of extreme weather conditions affecting wave height and period on power generation.

The DELA module integrates spatial, channel, and local attention mechanisms, as described by the following formula:

$$Y = y_{Li} + y_{Si} + y_{Di}. \quad (21)$$

In the equation, y_{Li} , y_{Si} , and y_{Di} represent the outputs of the spatial, channel, and local attention modules, respectively. These outputs enhance the model's ability to handle and predict complex temporal data, thereby improving the overall prediction performance of the module.

4. Simulation Results Analysis

4.1. Data Sources

The wave height and wave period data are from the European meteorological dataset ERA5 located at the latitudes of 116° E and 16.5° N. Its geographical location is shown in Figure 11. Details of the three stations are listed. Figure 12 shows the raw significant wave height and wave period size, and Figure 1 shows the characteristic distribution of raw significant wave heights and wave periods. The dataset covers 8593 wave observations for the period from 1 January to 24 December 2023. The dataset includes significant wave heights and wave periods, and the data were sampled at one-hour intervals, with the first 80% selected as the training set, the last 168 pieces of data as the test set, and the remaining data as the validation set.

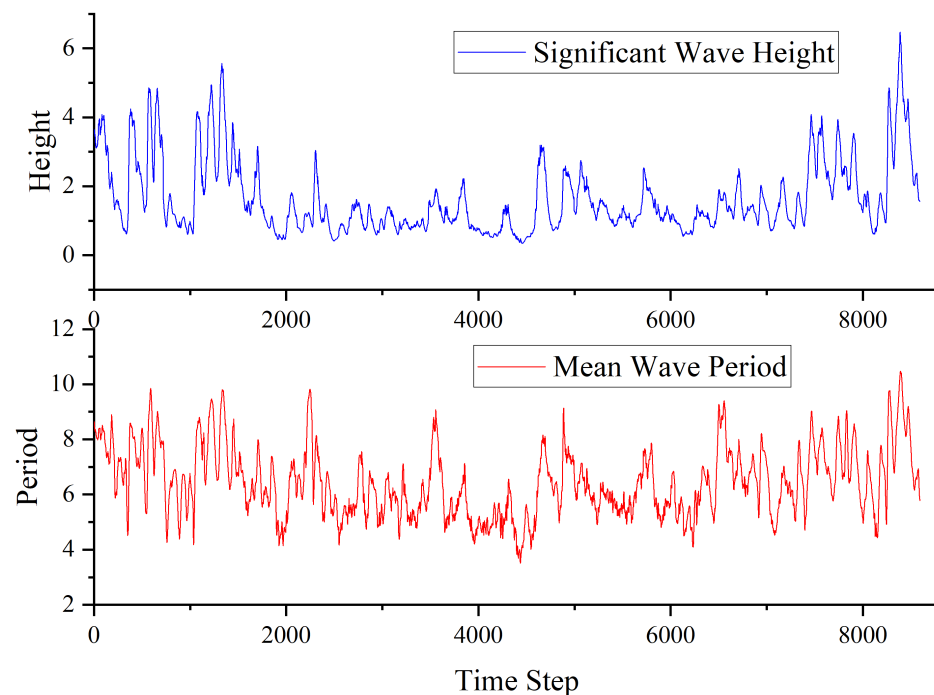


Figure 11. Original effective wave height and wavelength data for longitude 116° E, latitude 16.5° N.

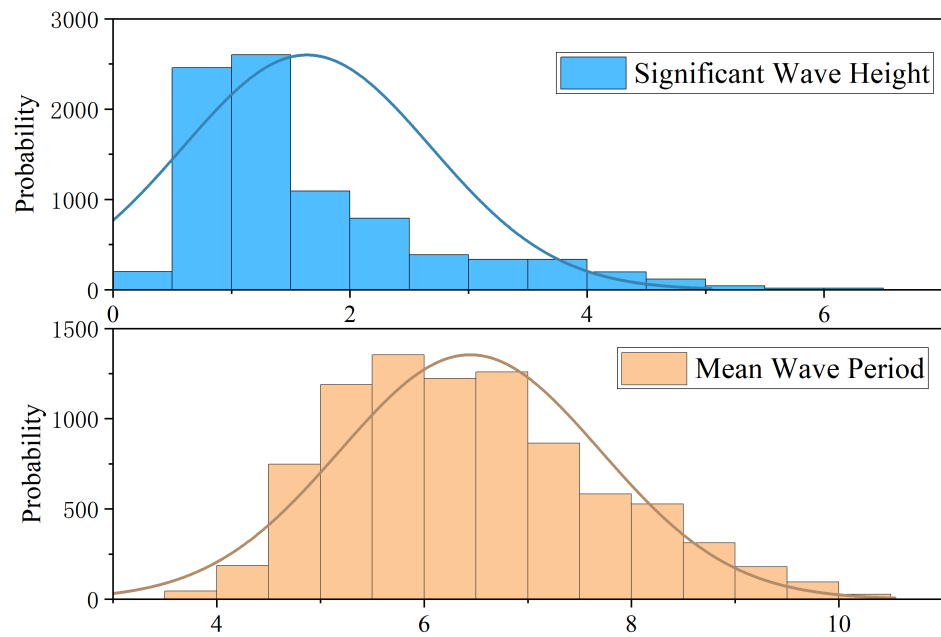


Figure 12. Distribution of significant wave heights and wave periods corresponding to longitude 116° E and latitude 16.5° N.

4.2. Model Evaluation Metrics

This study employed standard evaluation metrics in wave energy prediction, including Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), Mean Absolute Error (MAE), and the Coefficient of Determination (R^2), to assess the accuracy of the prediction results.

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right|. \quad (22)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n \left(y_i - \hat{y}_i \right)^2. \quad (23)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n | \hat{y}_i - y_i |. \quad (24)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (\bar{y}_i - y_i)^2}. \quad (25)$$

In the equation, \hat{y}_i , y_i , and \bar{y}_i represent the predicted values, true values, and mean values, respectively. n denotes the number of samples.

4.3. Attention Comparison Experiment

In order to assess the efficacy of both the attention mechanism and the model, comparative simulation experiments were designed and executed with a focus on validating their performance. The detailed parameter configurations for these experiments are outlined in Table 2. The tests were performed in a Windows 11 environment, leveraging an NVIDIA 4070ti graphics card for enhanced computational capabilities.

Table 2. Power conversion module comparison experiment parameter settings.

Parameter Name	Parameter Settings
Prediction Time	168
Batch Size	128
Input Length	24
Output Length	1
Epochs	200
Learning Rate	0.01
Weight Decay	0.0032
Optimizer	Adam
Scheduler	Cosine

To evaluate the effectiveness of the proposed attention mechanism in this model, this study compared six different attention mechanisms. The baseline model, CNN-BiLSTM-DELA, had the DELA component replaced with Efficient Local Attention (ELA) [28], Efficient Multiscale Attention (EMA) [29], Global Context (GC) [30], Squeeze Excitation Attention (SE) [31], Simple Attention Module (SimAM) [32], Convolutional Block Attention Module (CBAM) [33], and Deformable Large Kernel Attention (DLKA) [34]. The results are illustrated in Figure 13, and the error metric calculations are provided in Table 3.

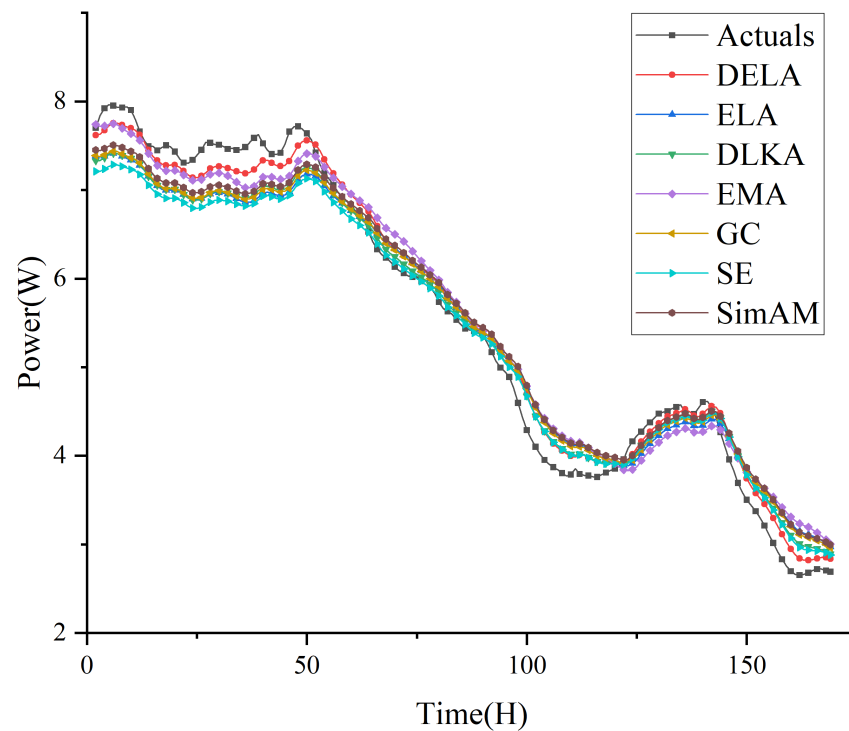


Figure 13. Short-term prediction model structure based on CNN-BiLSTM-DELA.

Table 3. Error results for Deformable Efficient Local Attention and six attention mechanisms.

Attention Models	MSE (W)	MAPE (%)	MAE (W)	R ² (%)
DELA	0.0396	3.7361	0.1809	98.69
ELA	0.1305	6.3364	0.3148	95.70
EMA	0.0977	6.1926	0.2813	96.78
GC	0.1172	5.9948	0.2973	96.14
SE	0.1321	5.3625	0.2902	95.65
SimAM	0.1081	6.0582	0.2906	96.44
CBAM	0.0831	5.5292	0.2578	97.27
DLKA	0.0997	5.0635	0.2614	96.72

4.4. Model Comparison Experiment

To validate the effectiveness of the CNN-BiLSTM-DELA improved model, comparative experiments were conducted with BiLSTM, CNN, BP, LSTM, CNN-BiLSTM, and GRU. The model prediction curves are illustrated in Figure 14, and the results of the comparative error analysis are presented in Table 4.

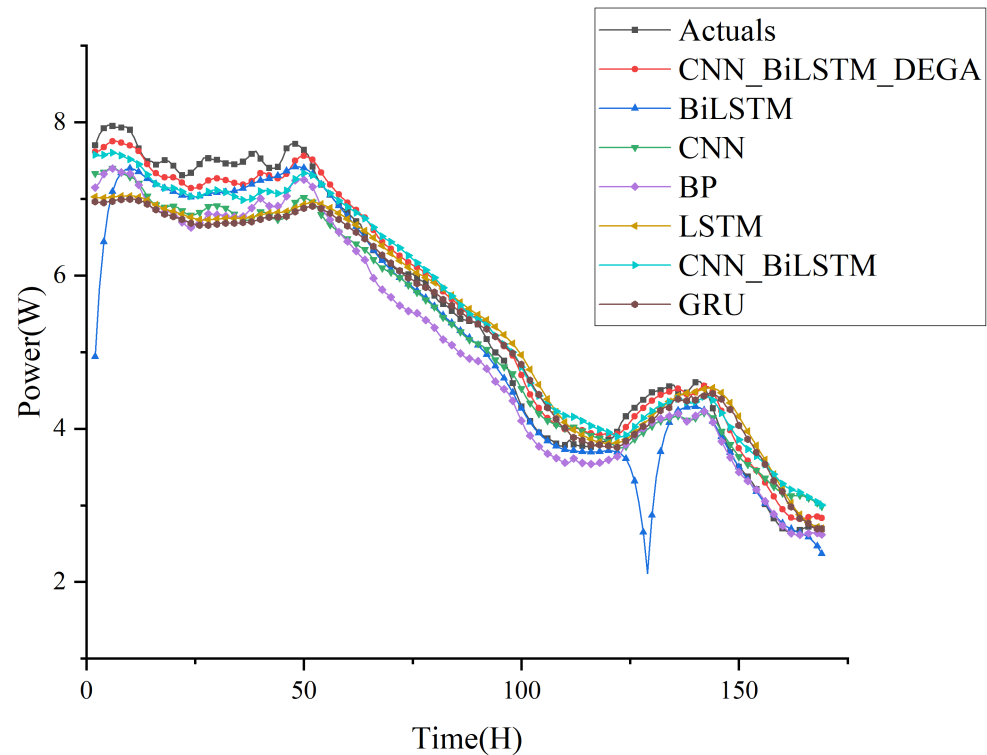


Figure 14. Short-term prediction model structure based on CNN-BiLSTM-DELA.

Table 4. Error results for CNN-BiLSTM-DELA and six common models.

Attention Models	MSE (W)	MAPE (%)	MAE (W)	R ² (%)
CNN-BiLSTM-DELA	0.0396	3.7361	0.1809	98.69
BiLSTM	0.2533	5.6679	0.2957	91.66
CNN	0.1699	6.6819	0.3566	94.01
BP	0.1799	6.3790	0.3737	94.08
LSTM	0.2290	7.2830	0.3908	92.47
CNN-BiLSTM	0.1054	6.2621	0.2920	96.53
GRU	0.2345	6.8091	0.3787	92.28

As shown in Table 4, the CNN-BiLSTM-DELA model outperformed all seven comparative models. It achieved a significantly lower Mean Squared Error (MSE) of 0.0396 W, representing reductions of 84.3%, 76.7%, 77.9%, 82.8%, 62.5%, and 83.1% compared to BiLSTM, CNN, BP, LSTM, CNN-BiLSTM, and GRU, respectively. In terms of Mean Absolute Percentage Error (MAPE), the CNN-BiLSTM-DELA model also demonstrated superior performance, with a value of 3.7361%, which was 1.6802%, 2.9458%, 2.6429%, 3.5469%, 2.5250%, and 3.0730% lower than BiLSTM, CNN, BP, LSTM, CNN-BiLSTM, and GRU, respectively. Additionally, the CNN-BiLSTM-DELA model achieved the highest Coefficient of Determination (R²) at 98.69%, indicating the strongest fit for wave energy power prediction. In summary, the CNN-BiLSTM-DELA model demonstrated superior accuracy and fitting capability compared to the other models, showcasing the highest precision and best fitting performance in wave energy power prediction.

4.5. Generalization Experiments

In order to verify the generalizability of the CNN-BiLSTM-DELA improved model, we conducted comparative generalization experiments using BiLSTM, CNN, LSTM, CNN-BiLSTM, and GRU. The dataset from January to June 2024 is used for the experiments, the model prediction curves are shown in Figure 15, and the results of the comparative error analysis are shown in Table 5.

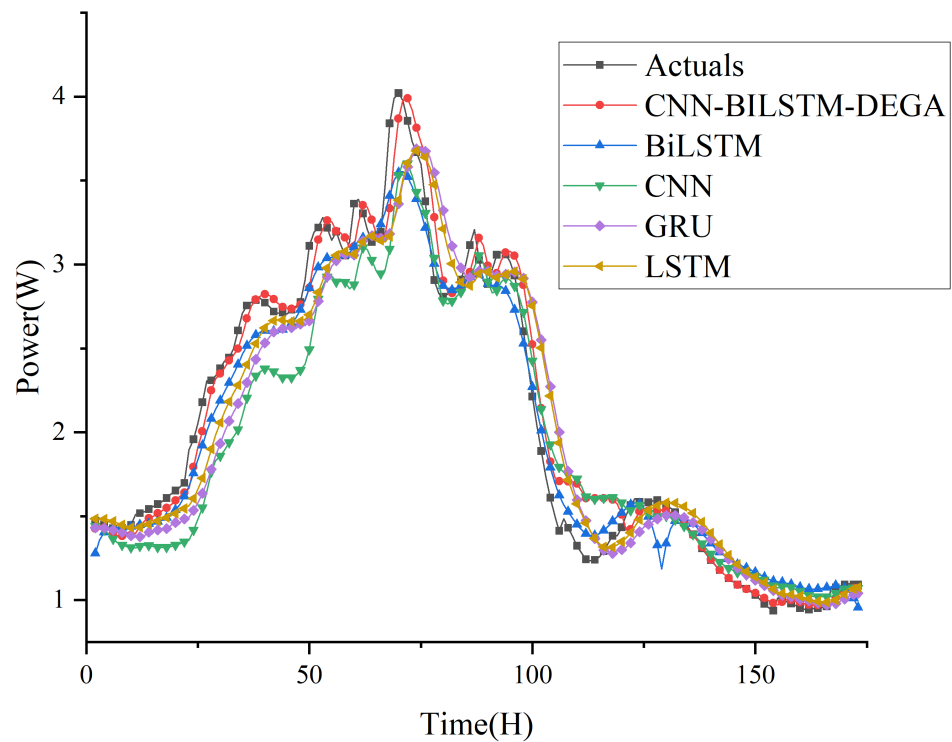


Figure 15. Error results for CNN-BiLSTM-DELA with common models for January–June 2024 wave data.

Table 5. Error results for CNN-BiLSTM-DELA with common models for January–June 2024 wave data.

Attention Models	MSE (W)	MAPE (%)	MAE (W)	R ² (%)
CNN-BiLSTM-DELA	0.0223	5.3593	0.1026	97.12
BiLSTM	0.0361	7.7835	0.1540	95.34
CNN	0.0816	10.8808	0.2217	89.49
LSTM	0.0591	8.8645	0.1781	92.38
CNN-BiLSTM	0.0474	8.5832	0.1689	93.90
GRU	0.0765	10.0181	0.2051	90.14

5. Conclusions

To mitigate the intermittency and stochastic nature of wave energy, which poses challenges to grid stability, a novel short-term wave energy forecasting model is proposed, integrating CNN, BiLSTM, and an innovative DELA mechanism. This model can accurately predict short-term wave power generation and support decision-makers in optimizing power dispatch, thereby enhancing the efficiency of wave energy conversion. This study achieved the following results:

1. The relationship between wave height, wave period, and power output of point absorber wave energy converters was simulated. A power matrix was developed and optimized using a BiGRU model, allowing for the rapid estimation of power outputs across various marine environments.

2. DELA is a three-channel attention mechanism that processes BiLSTM outputs through spatial, channel, and local attention mechanisms, merging their outputs. This mechanism outperformed seven established attention mechanisms in a comparative analysis.
3. Outputs from the BiGRU model, derived from direct-drive wave energy converters and buoy-based wave parameters in the South China Sea, are fed into the CNN-BiLSTM-DELA model. Operating in parallel, the CNN component primarily identifies extreme wave conditions, while the BiLSTM-DELA component forecasts wave energy based on temporal data.
4. Through comparative studies, the CNN-BiLSTM-DELA model showed the highest accuracy and goodness of fit, surpassing alternative models and demonstrating superior predictive performance.

In summary, the short-term wave energy forecasting model offers enhanced accuracy and adaptability, supporting decision-makers in optimizing scheduling strategies to maintain power system stability and economic efficiency.

Author Contributions: Writing—original draft preparation, S.L.; writing—review and editing, Y.Z., Q.S., Z.H. and R.C.; supervision, L.Z. and Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the National Natural Science Foundation of China (grant No. 51675265), the Natural Science Foundation of Jiangsu Higher Education Institutions of China (grant No. 23KJB460005), and the Natural Science Foundation of Huaian (grant No. HAB202226).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The authors can provide the raw data in this work upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Falcão, A.F.d.O. Wave energy utilization: A review of the technologies. *Renew. Sustain. Energy Rev.* **2010**, *14*, 899–918. [[CrossRef](#)]
2. Yang, Y.; Han, L.; Qiu, C.; Zhao, Y. A short-term wave energy forecasting model using two-layer decomposition and LSTM-attention. *Ocean Eng.* **2024**, *299*, 117279. [[CrossRef](#)]
3. Neshat, M.; Nezhad, N.; Sergiienko, Y.; Mirjalili, G.; Garcia, D.A. Wave power forecasting using an effective decomposition-based convolutional bi-directional model with equilibrium nelder-mead optimiser. *Energy* **2022**, *256*, 124623. [[CrossRef](#)]
4. Zhu, G.; Chen, Z.; Dai, Y.; Feng, Y. Development of an overtopping wave absorption device for a wave-making circulating water channel. *Ocean Eng.* **2023**, *273*, 113945. [[CrossRef](#)]
5. Feng, Z.; Hu, P.; Li, S.; Mo, D. Prediction of Significant Wave Height in Offshore China Based on the Machine Learning Method. *J. Mar. Sci. Eng.* **2022**, *110*, 836. [[CrossRef](#)]
6. Song, D.; Yu, M.; Wang, Z.; Wang, X. Wind and wave energy prediction using an AT-BiLSTM model. *Ocean Eng.* **2023**, *281*, 115008. [[CrossRef](#)]
7. Ni, C.; Peng, W. An integrated approach using empirical wavelet transform and a convolutional neural network for wave power prediction. *Ocean Eng.* **2023**, *276*, 114231. [[CrossRef](#)]
8. Pinson, P.; Reikard, G.; Bidlot, J.-R. Probabilistic forecasting of the wave energy flux. *Appl. Energy* **2012**, *93*, 364–370. [[CrossRef](#)]
9. Li, D. Experimental study of a floating two-body wave energy converter. *Renew. Energy* **2023**, *218*, 119351. [[CrossRef](#)]
10. Ghimire, S.; Deo, R.C.; Wang, H.; Al-Musaylh, M.S.; Casillas-Pérez, D.; Salcedo-Sanz, S. Stacked LSTM Sequence-to-Sequence Autoencoder with Feature Selection for Daily Solar Radiation Prediction: A Review and New Modeling Results. *Energies* **2022**, *15*, 1061. [[CrossRef](#)]
11. Shadmani, A.; Nikoo, M.R.; Gandomi, A.H.; Wang, R.-Q.; Golparvar, B. A review of machine learning and deep learning applications in wave energy forecasting and WEC optimization. *Energy Strategy Rev.* **2023**, *49*, 101180. [[CrossRef](#)]
12. Raj, L.N.; Prakash, R. Assessment and prediction of significant wave height using hybrid CNN-BiLSTM deep learning model for sustainable wave energy in australia. *Sustain. Horizons* **2024**, *11*, 100098. [[CrossRef](#)]
13. Le, X.-H.; Ho, H.V.; Lee, G.; Jung, S. Application of Long Short-Term Memory (LSTM) Neural Network for Flood Forecasting. *Water* **2019**, *11*, 1387. [[CrossRef](#)]
14. Zheng, C.W.; Li, C.Y.; Chen, X.; Pan, J. Numerical forecasting experiment of the wave energy resource in the China sea. *Adv. Meteorol.* **2016**, *2016*, 5692431. [[CrossRef](#)]
15. Shao, X.; Yao, H.-D.; Ringsberg, J.W.; Li, Z.; Johnson, E. Performance analysis of two generations of heaving point absorber WECs in farms of hexagon-shaped array layouts. *Ships Offshore Struct.* **2024**, *19*, 687–698. [[CrossRef](#)]

16. Wu, F.; Jing, R.; Zhang, X.-P.; Wang, F. A combined method of improved grey BP neural network and MEEMD-ARIMA for day-ahead wave energy forecast. *IEEE Trans. Sustain. Energy* **2021**, *12*, 2404–2412. [[CrossRef](#)]
17. Nezhad, M.M.; Neshat, M.; Sylaios, G.; Garcia, A.D. Marine energy digitalization digital twin's approaches. *Renew. Sustain. Energy Rev.* **2024**, *191*, 114065. [[CrossRef](#)]
18. Mousavi, S.M.; Ghasemi, M.; Dehghan Manshadi, M.; Mosavi, A. Deep Learning for Wave Energy Converter Modeling Using Long Short-Term Memory. *Mathematics* **2021**, *9*, 871. [[CrossRef](#)]
19. Neshat, M.; Nezhad, M.M.; Mirjalili, S.; Garcia, A.D.; Dahlquist, E.; Gandomi, A.H. Short-term solar radiation forecasting using hybrid deep residual learning and gated LSTM recurrent network with differential covariance matrix adaptation evolution strategy. *Energy* **2023**, *278*, 127701. [[CrossRef](#)]
20. Chen, G.; Tang, B.; Zeng, X.; Zhou, P.; Kang, P.; Long, H. Short-term wind speed forecasting based on long short-term memory and improved BP neural network. *Int. J. Electr. Power* **2022**, *134*, 107365. [[CrossRef](#)]
21. Elbisy, M.S. Sea wave parameters prediction by support vector machine using a genetic algorithm. *J. Coast. Res.* **2015**, *31*, 892–899. [[CrossRef](#)]
22. Yang, S.; Deng, Z.; Li, X.; Zheng, C.; Xi, L.; Zhuang, J.; Zhang, Z.; Zhang, Z. A novel hybrid model based on STL decomposition and one-dimensional convolutional neural networks with positional encoding for significant wave height forecast. *Renew. Energy* **2021**, *173*, 531–543. [[CrossRef](#)]
23. Guo, B.; Wang, T.; Jin, S.; Duan, S.; Yang, K.; Zhao, Y. A Review of Point Absorber Wave Energy Converters. *JMSE* **2022**, *10*, 1534. [[CrossRef](#)]
24. Allam, J.P.; Sahoo, S.P.; Ari, S. Multi-stream bi-GRU network to extract a comprehensive feature set for ECG signal classification. *Biomed. Signal Process. Control* **2024**, *92*, 106097. [[CrossRef](#)]
25. Wu, K. Named entity recognition of rice genes and phenotypes based on BiGRU neural networks. *Comput. Biol. Chem.* **2024**, *108*, 107977. [[CrossRef](#)]
26. Bengio, Y.; Simard, P.; Frasconi, P. Learning long-term dependencies with gradient descent is difficult. *IEEE Trans. Neural Netw.* **1994**, *5*, 157–166. [[CrossRef](#)]
27. Cahuantzi, R.; Chen, X.; Güttel, S. A Comparison of LSTM and GRU Networks for Learning Symbolic Sequences. Available online: <http://arxiv.org/abs/2107.02248> (accessed on 4 January 2023).
28. Xu, W.; Wan, Y. ELA: Efficient Local Attention for Deep Convolutional Neural Networks. Available online: <http://arxiv.org/abs/2403.01123> (accessed on 3 September 2024).
29. Ooyang, D.; He, S.; Zhang, G.; Luo, M.; Guo, H.; Zhang, J.; Huang, Z. Efficient multi-scale attention module with cross-spatial learning. In Proceedings of the ICASSP 2023—2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 4–10 June 2023; pp. 1–5.
30. Cao, Y.; Xu, J.; Lin, S.; Wei, F.; Hu, H. GCNet: Non-Local Networks Meet Squeeze-Excitation Networks and Beyond. Available online: <http://arxiv.org/abs/1904.11492> (accessed on 25 April 2019).
31. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E. Squeeze-and-Excitation Networks. Available online: <http://arxiv.org/abs/1709.01507> (accessed on 16 May 2019).
32. Yang, L.; Zhang, R.; Li, L.; Xie, X. SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks. *Int. Conf. Mach. Learn.* **2021**, *139*, 11863–11874.
33. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. CBAM: Convolutional block attention module. *Comput. Vis.* **2018**, *11211*, 3–19.
34. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E. Beyond Self-Attention: Deformable Large Kernel Attention for Medical Image Segmentation. Available online: <http://arxiv.org/abs/2309.00121> (accessed on 31 August 2023).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.