

Research Article

Power Generation Prediction Method of Offshore Wind Turbines Based on Cascaded Deep Learning

Zhen'ao Sun ¹ and Zhe Chen ^{1,2}

¹School of Electrical Engineering, Shenyang University of Technology, Shenyang 110870, China

²Department of Energy Technology, Aalborg University, Aalborg, DK-9220, Denmark

Correspondence should be addressed to Zhen'ao Sun; sunzhenao@mjcedu.cn

Received 6 June 2022; Revised 13 September 2022; Accepted 14 September 2022; Published 23 September 2022

Academic Editor: Nick Papanikolaou

Copyright © 2022 Zhen'ao Sun and Zhe Chen. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Aiming at the problems of low prediction accuracy, long time, and poor results in current wind turbine generation power prediction methods, an offshore wind turbine generation power prediction method based on cascaded deep learning is proposed. Using deep belief networks, stacked autoencoding networks, and long short-term memory networks, a cascaded deep learning method is proposed to predict the power generation of offshore wind turbines. Multiple feature extractors are used to extract and fuse high-level features to form a unified feature with richer information to predict the power generation sequence of offshore wind turbines. According to the modeling strategy and port design strategy, using the stacked autoencoding networks as the basic unit, a cascaded deep learning model for generating power prediction of offshore wind turbines is established. Through the selection of input variables, the variables that have a great correlation with wind power are obtained. The layer-by-layer greedy algorithm is used for training from bottom to top, and supervised learning is used to fine-tune the network parameters from top to bottom to realize the generation power prediction of the offshore wind turbine. The experimental results show that the proposed method is effective in predicting the power generation of offshore wind turbines, which can effectively improve the prediction accuracy and shorten the prediction time.

1. Introduction

Due to the vigorous exploitation of nonrenewable energy such as coal mines and oil, today's society is facing various environmental problems such as environmental pollution, the greenhouse effect, and the depletion of nonrenewable energy [1]. In order to solve a series of problems, a large number of renewable energies, such as solar energy and wind energy, will occupy a more and more important position in the energy field as important strategic energy development objects. With the gradual depletion of global fossil energy and the deterioration of Earth's overall environment, the development and utilization of renewable energy have become a consensus all over the world. With the characteristics of fast growth and high technical maturity, the proportion of wind energy in power supply all over the world has increased year by year. Wind energy is a pollution-free, renewable, and

relatively mature renewable energy source. The cost of power generation is decreasing day by day, and the world is rich in reserves. At present, the utilization of wind energy can effectively alleviate the shortage of fossil fuels [2–4]. At present, with the continuous growth of wind power installed capacity, the impact of large-scale wind power access on the power grid is becoming more and more obvious. Accurate and effective prediction of wind power can reduce the adverse impact of wind power access on the power grid and optimize power grid dispatching [5]. Therefore, the research on wind power prediction has important practical significance.

At present, scholars in related fields have studied wind power prediction and achieved some theoretical results. Peng et al. [6] proposed a wind power prediction method for wind farms based on multifeature similarity matching. The multifeature similarity matching method is used to optimize and analyze the key parameters. The influence of each key

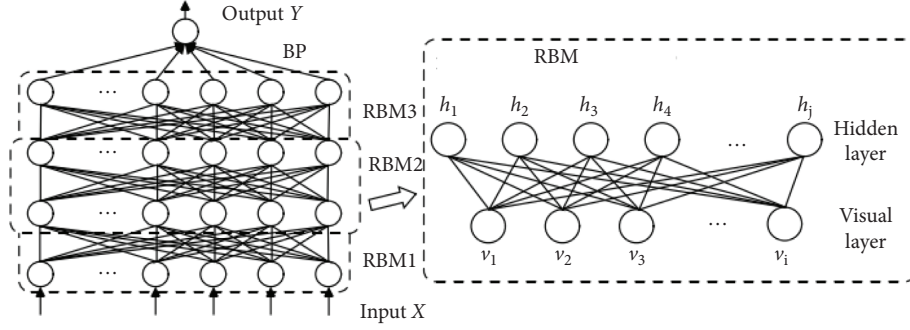


FIGURE 1: Structure diagram of a deep belief network.

parameter on the prediction error and the applicability of this method at different regional scales are analyzed. The new method and optimization analysis process are verified by the wind farm group, composed of wind farms. This method has been proven to be an effective wind power prediction method for wind farms and has potential industrial application prospects. Sun et al. [7] proposed an artificial neural network wind turbine power modeling and optimization method based on wind farm experimental data. The power prediction model is constructed and the yaw angle is optimized to minimize the impact on the overall wake of the wind turbine. The artificial neural network considering the wake effect is used to estimate the total power generation of a wind turbine under a given wind speed, wind direction, and yaw angle. The model is trained and evaluated using the experimental data of five wind turbines operating on a wind farm. It can effectively improve the total power of wind turbines in all directions. However, the above methods still have the problems of low prediction accuracy, long waiting times, and poor effect.

To solve the above problems, a power generation prediction method of offshore wind turbines based on cascade deep learning is proposed. A cascade deep learning method is used to predict the power generation of offshore wind turbines [8–10]. Multiple feature extractors are used to extract and fuse high-order features to predict the generation power sequence of offshore wind turbines. According to the modeling strategy and port design strategy, the stacked autoencoding network is used as the basic unit to build a cascaded deep learning model for offshore wind turbine generation power prediction. Through the selection of input variables, variables that have a greater correlation with wind power are obtained, and the power generation prediction of offshore wind turbines is realized. The forecasting effect of the offshore wind turbine generating power of this method is good, which can effectively improve the forecasting accuracy and shorten the forecasting time.

2. Cascaded Deep Learning Network

2.1. Deep Belief Network. A deep belief network (DBN) is formed by stacking several Restricted Boltzmann Machines (RBM) with powerful unsupervised learning capabilities [11]. The structure of the deep belief network is shown in Figure 1.

In Figure 1, each RBM contains a visible layer and a hidden layer, and the last layer is superimposed on the BP neural network. The training of DBN consists of two processes: unsupervised pretraining of RBM and supervised fine-tuning using the BP algorithm. Suppose that the network input is X , the label is Y , the hidden layer of RBM is h and its corresponding j neuron is h_j , the visible layer is v and its corresponding i neuron is v_i . The training process is as follows:

2.1.1. Unsupervised Pretraining. Pretraining is to input the original data into the lowest RBM visual layer, and then train RBM1. After training, take the hidden layer of RBM1 as the visual layer of RBM2, continue to train Rbm2, then take the hidden layer of RBM2 as the visual layer of RBM3, and so on until the training is completed. RBM is an energy model based on thermodynamics. Its energy function is defined as $E(v, h)$, and the joint probability distribution of the hidden layer and the visible layer is $P(v, h)$, then,

$$E(v, h) = - \sum_{i=1}^{n_v} \sum_{j=1}^{n_h} w_{ij} v_i h_j - \sum_{i=1}^{n_v} a_i v_i - \sum_{j=1}^{n_h} b_j h_j. \quad (1)$$

In Formula (1), w_{ij} represents the weight between the i neuron in the hidden layer and the j neuron in the visible layer, a_i represents the threshold of the i neuron in the visible layer, and b_j represents the hidden layer. The threshold of j neurons, n_v represents the number of visible layers, and n_h represents the number of hidden layers. From Formula (1), the joint probability density of $P(v, h)$ can be obtained as follows:

$$P(v, h) = \frac{1}{Z} e^{-E(v, h)}. \quad (2)$$

In Formula (2), Z represents the normalized numerator. Because the structural feature of RBM is that there is a connection between layers and there is no connection inside each layer, if the state value of visual neurons is given, whether each hidden neuron is activated is conditionally independent. When the state of the visible layer is known, then the activation probability of the j neuron in the hidden layer is given as follows:

$$P(h_j = 1 | v) = f \left(\sum_{i=1}^{n_v} w_{ij} v_i + a_j \right). \quad (3)$$

When the state of the hidden layer is known, the activation probability of the i neuron in the visible layer is given as follows:

$$P(v_i = 1|h) = f\left(\sum_{j=1}^{n_h} w_{ij}h_j + b_i\right). \quad (4)$$

In Formula (4), $f(x)$ is the activation function, and the training process of RBM is unsupervised. Assuming parameter $\theta = (w, b, a)$, the training process of RBM is to find the value of the parameter θ . θ can be obtained by maximizing the log-likelihood function of RBM on the training set. Let $v(t)$ be the known t input sample, then,

$$\begin{aligned} L(\theta) &= \sum_{t=1}^T \ln P(v^{(t)}, h), \\ &= \sum_{t=1}^T \left(\ln \sum_{n_h} e^{-[-E(v^{(t)}, h)]} - \ln \sum_{n_v} e^{-[-E(v, h)]} \right), \end{aligned} \quad (5)$$

$$\theta = \arg \max L(\theta) = \arg \max \sum_{t=1}^T P(v^{(t)}, h).$$

In order to obtain the optimal solution of θ , the gradient descent method is used to find the maximum value of $L(\theta)$ [12]. Then, the gradient of the log-likelihood function with respect to θ is given as follows:

$$\begin{aligned} \frac{\delta L}{\delta \theta} &= \sum_{t=1}^T \frac{\delta}{\delta \theta} \left(\ln \sum_{n_h} e^{-[-E(v^{(t)}, h)]} - \ln \sum_{n_v} e^{-[-E(v, h)]} \right), \\ &= \sum_{t=1}^T \left(\left\{ \frac{\delta[-E(v^{(t)}, h)]}{\delta \theta} \right\}_{P(h|v^{(t)})} \right). \end{aligned} \quad (6)$$

In Formula (6), $\{\cdot\}_P$ represents the mathematical expectation about the distribution P , and $P(h|v^{(t)})$ represents the probability distribution of the hidden layer when the visible neuron is a known training sample $v^{(t)}$. Considering that the structure of the RBM model is symmetrical and the neuron state conditions are independent, the contrast divergence algorithm (CD) is used to solve the problem.

2.1.2. Supervised Fine-Tuning. The BP network is superimposed on the last layer of DBN, and the output eigenvector of RBM is used as its input. Because, after RBM network training, it can only ensure that the weight in its own layer is optimal for the feature vector mapping of this layer and not for the whole DBN network, the BP network is also used to propagate the error signal from the top to bottom to each RBM layer to fine-tune the parameters of the whole DBN network.

The training of the RBM network can be regarded as the initialization of the weights of a deep BP network, which gives DBN the advantage of strong feature learning ability, but it does not have the disadvantages of the BP network that it is easy to fall into the local optimum and the training is slow [13–15].

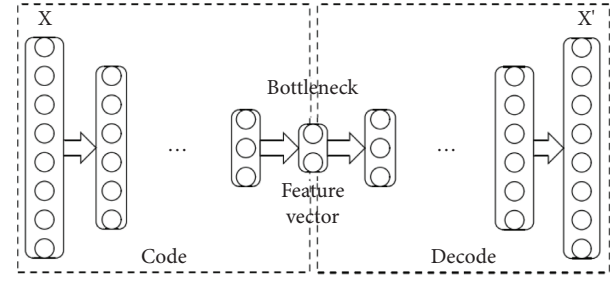


FIGURE 2: Stacked self-encoding network structure diagram.

2.2. Stacked Self-Encoding Network. Stacked Autoencoder (SAE) is one of the most commonly used deep learning methods at present. It is composed of several AEs stacked to realize the gradual abstraction of information features. Similar to DBN, it is a generative model [16]. The structure of the stacked self-encoding network is shown in Figure 2.

According to Figure 2, the SAE network is symmetrical left and right. The side where the number of neurons is reduced layer by layer from the leftmost input layer is called the coding side, and then these layers form the other side through a mirror image, which is called the decoding side, and the middle layer is called the bottleneck of SAE (i.e., the characteristic direction of data). SAE has been widely used in image classification, data analysis, audio analysis, and other fields because of its flexible structure, simple training, and strong feature extraction ability. Firstly, a brief introduction to the self-encoder is given.

2.2.1. Self-Encoder. The structure and principle of the self-encoder are shown in Figure 3.

According to Figure 3(a), a simple AE is a three-layer symmetric MLP in which the number of neurons in the hidden layer is less than that in the input layer and the output layer. From the input layer to the hidden layer is called the encoding process, and from the hidden layer to the output layer is called the decoding process. The purpose of training AE is to minimize the reconstruction error of input data. At this time, the output of a hidden layer is another expression of input data. Its dimension is less than that of original data, so as to achieve the effect of removing redundancy and extracting the characteristics of original data. Its function is similar to principal component analysis (PCA).

According to Figure 3(b), assuming that x is the training data set, which contains N data vectors, and $L(X, Z)$ is the reconstruction error, then the encoding process and decoding process of AE can be expressed as follows:

$$\begin{aligned} y(x) &= f(\mathbf{W}x + \mathbf{b}), \\ z(x) &= g(\mathbf{V}y(x) + \mathbf{c}). \end{aligned} \quad (7)$$

In Formula (7), \mathbf{W} represents a coding matrix, \mathbf{V} represents a decoding matrix, \mathbf{b} represents a coding threshold vector, \mathbf{c} represents a decoding threshold vector, and f and g represent an activation function. The essence of training AE is to find the parameter that minimizes $L(X, Z)$ on the data

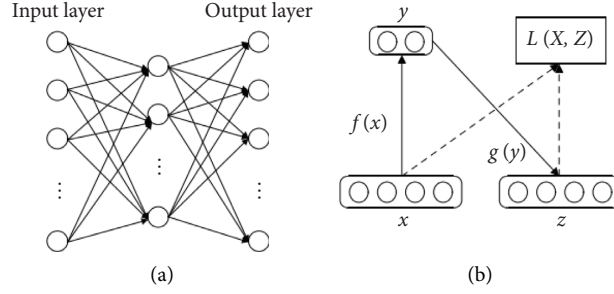


FIGURE 3: Autoencoder structure and principle diagram. (a) AE basic structure, (b) AE fundamentals.

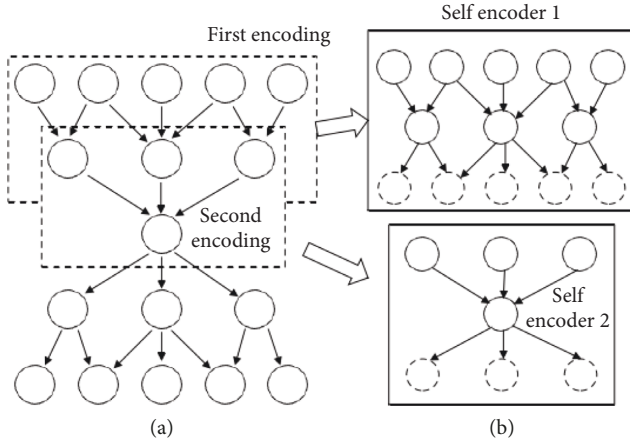


FIGURE 4: Schematic diagram of the stacked autoencoding layer-by-layer training method.

set x , and set the optimal parameter to θ^* , which can be expressed as follows:

$$\theta^* = \arg \min_{\theta} L(X, Z) = \arg \min_{\theta} \frac{1}{2N} \sum_{i=1}^N \|x^{(i)} - z(x^{(i)})\|^2. \quad (8)$$

2.2.2. Stacked Autoencoding Network Training Process. The layer-by-layer training method of stacked autoencoding is shown in Figure 4.

The training process of SAE is similar to that of DBN. It also needs pretraining and fine-tuning. First, reconstruct the input layer of each AE layer until the whole network is trained, that is, the pretraining stage is completed, and then use the BP algorithm to realize supervised fine-tuning and adjust all network parameters from top to bottom.

2.3. Long Short-Term Memory Network. The long short-term memory network (LSTM) is an improved form of a cyclic neural network. It has most of the excellent characteristics of a cyclic neural network model. At the same time, it solves the problem that the gradient of a cyclic neural network is easy to disappear in the training process and cannot continue to improve the accuracy of the model [17–19]. Other neural networks are composed of independent neuron connections, and LSTM networks are similar to them, which are formed

by connecting independent LSTM blocks. It contains three gates: input, forget, and output, as well as a block input, a block output, a memory cell, an output activation function, and a peephole connection, where the block output is repeatedly connected back to the block input and all the gates. The training process of the LSTM network includes the forward transmission of information and the back propagation of error through time.

2.3.1. Information Forwarding. Suppose the input vector at time t is x^t , the number of LSTM blocks is N , and the number of inputs is M . The weight vectors of an LSTM network layer are input weight: $\mathbf{W}_z, \mathbf{W}_i, \mathbf{W}_f, \mathbf{W}_o \in R^{N \times N}$, loop weight: $\mathbf{R}_z, \mathbf{R}_i, \mathbf{R}_f, \mathbf{R}_o \in R^{N \times N}$, peephole connection weight: $\mathbf{P}_i, \mathbf{P}_f, \mathbf{P}_o \in R^N$, and threshold: $\mathbf{b}_z, \mathbf{b}_i, \mathbf{b}_f, \mathbf{b}_o \in R^N$, then the calculation formula for each vector is as follows:

The block input is given as follows:

$$z^t = g(W_z x^t + R_z y^{t-1} + b_z). \quad (9)$$

The input gate is given as follows:

$$\tilde{i}^t = \sigma(W_i x^t + R_i y^{t-1} + p_i e c^{t-1} + b_i). \quad (10)$$

The forget door is given as follows:

$$f^t = \sigma(W_f x^t + R_f y^{t-1} + p_f e c^{t-1} + b_f). \quad (11)$$

The cell status is given as follows:

$$c^t = z^t e i^t + c^{t-1} e f^t. \quad (12)$$

The output gate is given as follows:

$$o^t = \sigma(W_o x^t + R_o y^{t-1} + p_o e c^t + b_o). \quad (13)$$

The block output is given as follows:

$$y^t = h(c^t) e o^t. \quad (14)$$

2.3.2. The Error Propagates Back through Time. The increment δ inside the LSTM block is calculated as follows:

$$\begin{aligned}
\delta y^t &= \Delta^t + R_z^T \delta z^{t+1} + R_i^T \delta i^{t+1} + R_f^T \delta f^{t+1} + R_o^T \delta o^{t+1}, \\
\delta \bar{o}^t &= \delta y^t e h(c^t) e \sigma'(\bar{o}^t), \\
\delta c^t &= \delta y^t e o^t e h'(c^t) + p_o e \delta \bar{o}^t + p_i e \delta i^{t+1} + p_f e \delta \bar{f}^{t+1} + \delta c^{t+1} e f^{t+1}, \\
\delta \bar{f}^t &= \delta c^t e c^{t-1} e \sigma'(\bar{f}^t), \\
\delta \bar{i}^t &= \delta c^t e z^t e \sigma'(\bar{i}^t), \\
\delta \bar{z}^t &= \delta c^t e i^t e g'(\bar{z}^t).
\end{aligned} \tag{15}$$

In Formula (15), Δ^t represents the incremental vector passed down from the previous layer. If E is a loss function, then it is generally equal to $\partial E / \partial y^t$ but it does not include cyclic dependence. The input increment only needs to be calculated when the next layer needs training. The calculation formula is as follows:

$$\delta x^t = W_z^T \delta \bar{z}^t + W_i^T \delta \bar{i}^t + W_o^T \delta \bar{f}^t + W_o^T \delta \bar{o}^t. \tag{16}$$

Finally, the gradient calculation of each weight is as follows:

$$\begin{aligned}
\delta W_* &= \sum_{t=0}^T (\delta_{*t}, x^t), \\
\delta R_* &= \sum_{t=0}^{T-1} (\delta_{*t+1}, y^t), \\
\delta b_* &= \sum_{t=0}^T \delta_{*t}, \\
\delta p_i &= \sum_{t=0}^{T-1} c^t e \delta i^{t+1}, \\
\delta p_f &= \sum_{t=0}^{T-1} c^t e \delta \bar{f}^{t+1}, \\
\delta p_o &= \sum_{t=0}^T c^t e \delta \bar{o}^t.
\end{aligned} \tag{17}$$

In Formula (17), $**$ represents any one of $\{\bar{z}, \bar{i}, \bar{f}, \bar{o}\}$. Compared with MLP, the biggest difference in training LSTM is that if you want to predict a value p_t at time t , the previous n samples $\{p_{t-n}, \dots, p_{t-1}\}$ need to be propagated through the network, and the number of time steps n is defined when the network is set up. These memory cells will store transient information according to their training state and output a predicted value p_t .

3. Prediction Method for Power Generation of Offshore Wind Turbines

Using a stacked autoencoding network and the long short-term memory network, aiming at the problem of generating power prediction of offshore wind turbines with a variety of heterogeneous data, a cascaded deep learning offshore wind turbine generating power prediction model is proposed with the help of multimodal learning and multitask learning

strategies. The model is a comprehensive prediction framework composed of multiple feature extractors, a feature fusion layer, and a prediction terminal. Each feature extractor automatically extracts features from each variable and then sends them to the feature fusion layer for data fusion, and finally, the prediction terminal gives the prediction result.

3.1. Power Generation Prediction of Multimodal Offshore Wind Turbines

3.1.1. Problem Description. From the perspective of pattern recognition, the power generation prediction of offshore wind turbines can be converted into a mapping process between multiple objects, namely $y = f(x)$. Specifically, y is the power generation sequence of the offshore wind turbine to be predicted, x is the input variable, and $f(x)$ is the implicit function of the prediction model. At present, most offshore wind turbine generation power prediction models are univariate models, that is, only the historical measurement values of the offshore wind turbine generation power are used for prediction [20]. However, in actual engineering, a wind farm often records multiple measurements, and these data may also serve as a forecast for the power generation of offshore wind turbines. From a physical point of view, the power generated by offshore wind turbines is a reflection of large aerodynamic energy, which is directly or indirectly related to a variety of factors, including wind turbine parameters, geographic conditions, and meteorological information. In fact, different quantitative measurements belong to the typical multisource heterogeneous data. They come from different sensors and have different physical meanings and dimensions. The machine learning community refers to them as multimodal data, and the learning task for them is called multimodal learning [21, 22]. Essentially, these quantitative measurements describe the different attributes of air kinetic energy and all contain the knowledge required for the power generation prediction of offshore wind turbines. From a philosophical point of view, they are different aspects of the same thing. They are different from each other, but they are internally unified.

3.1.2. Mathematical Modeling. The multistep prediction of the power generation of offshore wind turbines is of great significance to the power system and is widely used in equipment maintenance, energy storage management, and power market operations [23–25]. Multistep offshore wind power generation forecasting mainly includes two methods, namely direct forecasting and iterative forecasting. In this paper, direct prediction is used to reduce the cumulative error more effectively, so as to achieve a more accurate forecast of the power generation of offshore wind turbines.

Given S types of measurement data, predict the time series $y_t = (y_{t+p}, y_{t+p+1}, \dots, y_{t+q})$ of the offshore wind turbine generating power in a certain period of time in the future at time t . Among them, y_t represents the power generation value of the offshore wind turbine at time t , and p and q represent the minimum and maximum predicted step

sizes, respectively. The prediction problem is a typical sequence-to-sequence prediction problem, which can be expressed as follows:

$$y_t = f(\mathbf{x}_t | \theta) = f(\left(\mathbf{x}_t^{(1)}, \mathbf{x}_t^{(2)}, \dots, \mathbf{x}_t^{(S)} | \theta\right)). \quad (18)$$

In Formula (18), \mathbf{x}_t represents the input vector, θ and $f(\cdot)$ represent the parameter vector, and the implicit function of the prediction model, respectively, and $\mathbf{x}_t^{(i)}$ represents the subvector corresponding to the measurement i at time t . Generally, $\mathbf{x}_t^{(i)}$ can be defined as follows:

$$\mathbf{x}_t^{(i)} = (\mathbf{x}_{t-n_i+1}^{(i)}, \mathbf{x}_{t-n_i+2}^{(i)}, \dots, \mathbf{x}_t^{(i)}) \in R^{n_i}. \quad (19)$$

In Formula (19), n_i represents the dimension of $\mathbf{x}_t^{(i)}$, and $\mathbf{x}_t^{(i)}$ represents the value of the quantity measurement i at time t . Naturally, the dimension of \mathbf{x}_t is $\sum_{i=1}^S n_i$. The training of predictive models is a supervised learning problem. Given the training set P , which contains N input/output pairs, namely $P = \{(\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2), \dots, (\mathbf{x}_N, \mathbf{y}_N)\}$, then the optimal parameter $\hat{\theta}$ of the prediction model can be obtained by minimizing the loss function, namely,

$$J = \frac{1}{N} \sum_{t=1}^N \|y_t - \hat{y}_t\|_F, \quad (20)$$

$$\hat{y}_t = f(\left(\mathbf{x}_t^{(1)}, \mathbf{x}_t^{(2)}, \dots, \mathbf{x}_t^{(S)} | \hat{\theta}\right)).$$

In Formula (20), $\|\cdot\|_F$ represents the F norm. It is worth noting that when $S = 1$, the prediction problem degenerates into a univariate prediction problem, that is, only the historical sequence of offshore wind power generation is used for prediction. When $p = q$, the prediction problem degenerates into a single-step offshore wind turbine generating power prediction problem. Therefore, compared to a single-step and a univariate offshore wind turbine generation power prediction, a multistep offshore wind turbine generation power prediction using multi-modal information is a more general problem.

3.2. Constructing an Offshore Wind Turbine Generating Power Prediction Model

3.2.1. Modeling Strategy. A variety of heterogeneous data is used to predict the power generation of offshore wind turbines, and multiple feature extractors are used to extract high-level features from various measurements. The high-level features from different measurements are fused to form a unified feature with richer information. Predict the power generation sequence of offshore wind turbines based on unified characteristics. In this way, the power generation prediction problem for offshore wind turbines can be described as follows:

$$\begin{aligned} y_t N &= f(\mathbf{x}_t) = f(\mathbf{x}_t^{(1)}, \mathbf{x}_t^{(2)}, \dots, \mathbf{x}_t^{(S)}) \approx u(h_t), \\ &= u(v(\mathbf{r}_t^{(1)}, \mathbf{r}_t^{(2)}, \dots, \mathbf{r}_t^{(S)})), \\ N &= u(v(g_1(\mathbf{x}_t^{(1)}), g_2(\mathbf{x}_t^{(2)}), \dots, g_S(\mathbf{x}_t^{(S)}))). \end{aligned} \quad (21)$$

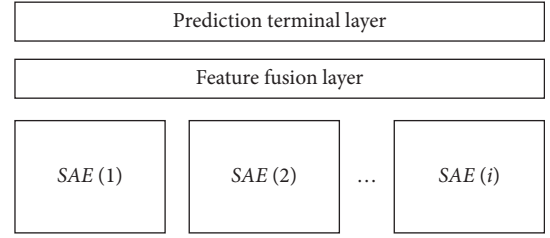


FIGURE 5: Cascade deep learning model for power generation prediction of offshore wind turbines.

In Formula (21), \mathbf{h}_t represents the unified feature, $\mathbf{r}_t^{(i)}$ represents the high-order feature corresponding to the quantity measurement i , $g_i(\cdot)$ represents the implicit function of the feature extractor corresponding to the quantity measurement i in the first stage, and $v(\cdot)$ and $u(\cdot)$ respectively represent the second stage and the implicit function of the third-stage function. According to this strategy, a variety of quantitative measurements have been integrated under the same learning framework, which has the potential to provide richer information for offshore wind turbine generation power prediction.

3.2.2. Model Structure. According to the modeling strategy and port design strategy, a cascaded deep learning model for offshore wind turbine generation power prediction is proposed. SAE is used as the basic unit to build, and its structure is shown in Figure 5.

In Figure 5, $SAE(i)$ represents the feature extractor corresponding to the quantity measurement i , L_i represents the number of layers of $SAE(i)$, L_i^j represents the dimension of the j layer in $SAE(i)$, and D_F and D_P represent the dimensions of the fusion layer and the prediction terminal layer, respectively. The model consists of S feature extractors, a feature fusion layer, and a prediction terminal layer. Among them, the feature extractor is SAE, the feature fusion layer is an ordinary fully connected network, and the prediction terminal layer is a regression layer.

3.2.3. Input Variable Selection. The input variable selection aims to select a part of the candidate variables for model building, and it plays a very important role in the performance of the model. Theoretically, introducing more input variables will provide richer information for wind power forecasting. However, in practice, this may cause problems with noise and excessive model scale [26]. Therefore, in practical applications, it is necessary to make a purposeful selection of input variables, so as to achieve the purpose of controlling the scale of the model, reducing the computational complexity, and improving the performance of the model. In the task of wind power forecasting, the variables that have a greater correlation with wind power are selected through input variable selection. For the cascaded deep learning model of offshore wind turbine generation power prediction, the purpose of input variable selection is to select K ($K \geq S$) variables from S candidate variables for wind power prediction.

3.2.4. Training Algorithm. The training of the cascaded deep learning model for power generation prediction of offshore wind turbines is as follows: the SAE at the bottom is trained from bottom to top through the layer-by-layer greedy algorithm [27–29]. In order to enhance the feature extraction capability of SAE, sparsity constraints are added to hidden layer units. At the same time, in order to reduce over-fitting, L2 regularization is performed on the network. For each AE, its loss function is as follows:

$$L(\mathbf{x}, \mathbf{z}) = \frac{1}{n} \sum_{t=1}^n \|\mathbf{x}_t - \mathbf{z}_t\|^2 + \alpha \sum_{j=1}^D KL(\rho|\hat{\rho}_j) + \beta \|\mathbf{W}\|_2^2. \quad (22)$$

In Formula (22), the second term is KL divergence, and the third term is the L2 regular term. All SAE top layers and feature fusion layers are regarded as a new AE, and its input variable \mathbf{r}_t is paralleled by the input of the SAE top layer, namely,

$$\mathbf{r}_t = (\mathbf{r}_t^{(1)}, \mathbf{r}_t^{(2)}, \dots, \mathbf{r}_t^{(S)}). \quad (23)$$

After this step, all layers except the top layer are initialized. Use supervised learning to adjust the parameters of the network from the top to the bottom [30]. After fine-tuning the parameters, the entire network has better prediction performance. Through the above steps, the power generation prediction for offshore wind turbines is realized.

4. Experimental Analysis

4.1. Experimental Environment and Dataset. In order to verify the effectiveness of the generation power prediction method for offshore wind turbines based on cascade deep learning, a wind turbine with an offshore installed capacity of 160 MW is used as the research object. The ground clearance of the wind turbine hub is 80 m, the average altitude is 5.8 m, and the annual average wind speed is 7.2 m/s. Observe the power generation-related data of the offshore wind turbines from January 2021 to December 2021 and form a wind dataset containing data for more than 126,000 stations. The data set is divided into two sub-datasets: the training set and the confirmation set, including the first 60% and the last 40% of the whole dataset. The training set and the test set are used for model training and testing, respectively, and the validation set is used for model selection and overfitting prevention. Predict the generated power of offshore wind turbines in January 2022 and compare it with the actually observed generated power data of offshore wind turbines in January 2022 to determine the prediction performance of this method. In order to ensure that the many-to-many mapping structure can reach the final convergence, the learning rate is reduced each time the loss decreases slowly, and each time it is divided by 2 on the basis of the previous one. Set the model parameters of the stacked noise reduction automatic encoder as shown in Table 1.

4.2. Power Prediction Evaluation Index. The root mean square error (RMSE) and the mean absolute error (MAE) are used as evaluation indexes. MAE is the average value of

TABLE 1: Model parameters of the stacked noise reduction automatic encoder.

| Network type | Parameter |
|--|------------------|
| Activation function | Sigmoid function |
| Batch size | 100 |
| Noise rate | 0.2 |
| Number of training iterations layer by layer | 50 |
| Layer by layer training learning rate | 0.1 |

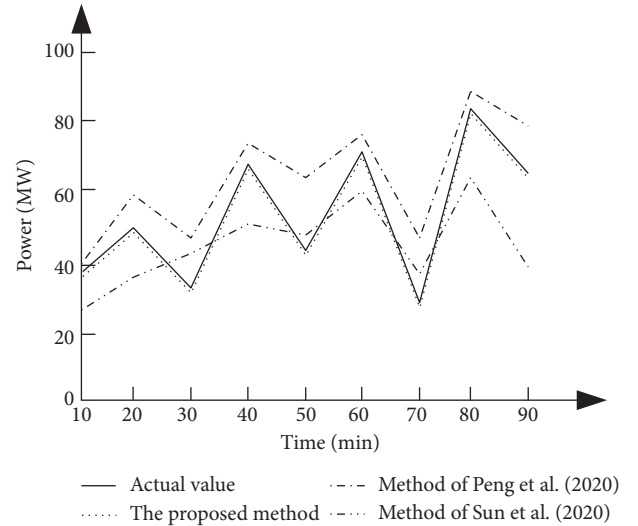


FIGURE 6: Comparison results for different methods of offshore wind turbine generating power prediction results.

absolute error. RMSE reflects the dispersion and reliability of actual value and predicted value. The smaller its value is, the higher the power prediction accuracy is. It is defined as follows:

$$\begin{aligned} \varepsilon_{\text{MAE}} &= \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i|, \\ \varepsilon_{\text{RMSE}} &= \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2}. \end{aligned} \quad (24)$$

In Formula (24), y_i and \hat{y}_i are the actual and predicted values of wind power, respectively, and N is the number of predicted data.

4.3. Comparison of Power Generation Prediction Effects of Offshore Wind Turbines. In order to verify the prediction effect of the proposed method, the methods of Peng et al. [6] and Sun et al. [7] are used to compare with the proposed method, and the comparison results of the power generation power prediction results of different methods of offshore wind turbines are shown in Figure 6.

It can be seen from Figure 6 that under different data sample collection intervals, the offshore wind turbine generation power prediction result of the method of Peng et al. [6] is relatively large, and there is a certain deviation from

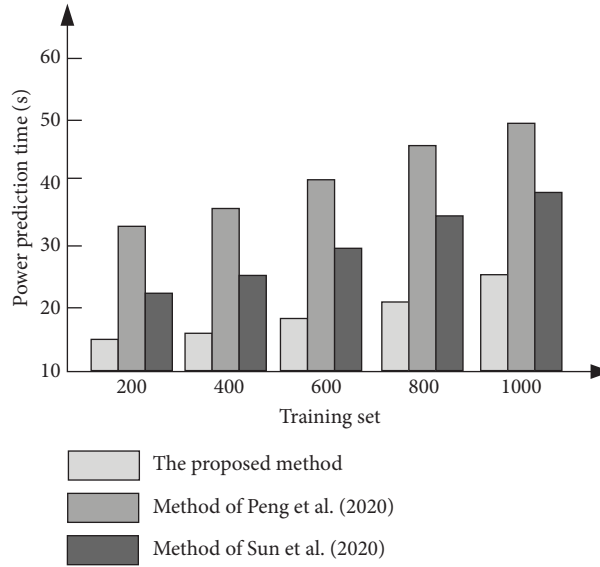


FIGURE 7: Comparison results of different methods of offshore wind turbine generating power prediction time.

TABLE 2: Comparison results of different methods of offshore wind turbine generating power prediction errors.

| Training set | The proposed method (s) | | The method of Peng et al. [6] (s) | | The method of Sun et al. [7] (s) | |
|--------------|-------------------------|---------|-----------------------------------|---------|----------------------------------|---------|
| | RMSE (%) | MAE (%) | RMSE (%) | MAE (%) | RMSE (%) | MAE (%) |
| 200 | 3.34 | 5.98 | 8.76 | 9.89 | 6.76 | 7.88 |
| 400 | 5.18 | 8.87 | 12.8 | 15.4 | 9.24 | 12.7 |
| 600 | 8.97 | 15.7 | 19.6 | 19.7 | 13.6 | 18.2 |
| 800 | 10.2 | 19.8 | 26.9 | 24.6 | 16.3 | 22.8 |
| 1000 | 13.3 | 26.9 | 31.5 | 31.9 | 17.8 | 29.5 |

the actual value. The prediction result of the offshore wind turbine power generation by the method of Sun et al. [7] is relatively small, and the deviation from the actual value is the largest. The power generation power prediction results of the proposed method are basically consistent with the actual power generation fluctuations of the offshore wind turbines. It can be seen that, compared with the method of Peng et al. [6] and the method of Sun et al. [7], the proposed method has a better effect on predicting the power generation of offshore wind turbines.

4.4. Comparison of Power Generation Forecast Time of Offshore Wind Turbines. To further verify the prediction time of the proposed method, the method of Peng et al. [6], the method of Sun et al. [7], and the proposed method were used to compare, and the comparison results of the power generation power prediction time of offshore wind turbines were obtained as shown in Figure 7.

It can be seen from Figure 7 that with the increase of data sample training set data, the generation power prediction time of offshore wind turbines using different methods increases accordingly. When the training set is 1000, the prediction time of offshore wind turbine power by the method of Peng et al. [6] is 49.9 s, and the prediction time of offshore wind turbine power by the method of Sun et al. [7]

is 47.5 s. The prediction time of offshore wind turbine power by the proposed method is only 25 s. It can be seen that, compared with the method of Peng et al. [6] and the method of Sun et al. [7], the proposed method has a shorter time to predict the power generation of offshore wind turbines.

4.5. Comparison of Power Generation Prediction Accuracy of Offshore Wind Turbines. On this basis, the prediction accuracy of the proposed method is further verified, and the method of Peng et al. [6], the method of Sun et al. [7], and the proposed method are used to compare, and the comparison results of the power generation power prediction errors of different methods of offshore wind turbines are obtained in Table 2.

According to the data in Table 2, as the data in the training set of data samples increases, the prediction error of the power generation of offshore wind turbines using different methods also increases. When the training set is 1000, the RMSE and MAE values of the method of Peng et al. [6] are 31.5% and 31.9%, respectively, and the RMSE and MAE values of the method of Sun et al. [7] are 17.8% and 29.5%, respectively. The RMSE and MAE values of the proposed method are 13.3% and 26.9%, respectively. It can be seen that, compared with the methods of Peng et al. [6] and Sun et al. [7], the RMSE and MAE values of the proposed method

are smaller, which can effectively reduce the power generation prediction error of offshore wind turbines and improve the power generation prediction accuracy of offshore wind turbines.

5. Conclusion

The generation power prediction method of offshore wind turbines based on cascade deep learning proposed in this paper gives full play to the advantages of deep learning algorithms. It has high generation power prediction accuracy, can effectively shorten the generation power prediction time of offshore wind turbines, and has a good generation power prediction effect. However, due to the limitation of data acquisition channels, the prediction effect of wind farm groups has not been considered in this study. Therefore, in the next research, it is necessary to further broaden the historical data range of wind turbine power generation to the wind farm group, in order to realize the wide area of wind power big data prediction technology.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

References

- [1] M. Ahmad, C. Işık, G. Jabeen, T. Ali, I. Ozturk, and D. W. Atchike, "Heterogeneous links among urban concentration, non-renewable energy use intensity, economic development, and environmental emissions across regional development levels," *Science of the Total Environment*, vol. 765, Article ID 144527, 2021.
- [2] T. T. Cai, M. Y. Dong, H. N. Liu, and S. Nojavan, "Integration of hydrogen storage system and wind generation in power systems under demand response program: a novel p-robust stochastic programming," *International Journal of Hydrogen Energy*, vol. 47, no. 1, pp. 443–458, 2022.
- [3] C. Guo, C. Ye, Y. Ding, and P. Wang, "A multi-state model for transmission system resilience enhancement against short-circuit faults caused by extreme weather events," *IEEE Transactions on Power Delivery*, vol. 36, no. 4, pp. 2374–2385, 2021.
- [4] X. Li, Y. Peng, W. Wang et al., "A method for optimizing installation capacity and operation strategy of a hybrid renewable energy system with offshore wind energy for a green container terminal," *Ocean Engineering*, vol. 186, Article ID 106125, 2019.
- [5] A. Ahmadpour and S. Gholami Farkoush, "Gaussian models for probabilistic and deterministic wind power prediction: wind farm and regional," *International Journal of Hydrogen Energy*, vol. 45, no. 51, pp. 27779–27791, 2020.
- [6] X. Peng, Y. Chen, K. Cheng et al., "Wind power prediction for wind farm clusters based on the multi-feature similarity matching method," *IEEE Transactions on Industry Applications*, vol. 56, no. 5, pp. 4679–4688, 2020.
- [7] H. Sun, C. Qiu, L. Lu, X. Gao, J. Chen, and H. Yang, "Wind turbine power modelling and optimization using artificial neural network with wind field experimental data," *Applied Energy*, vol. 280, no. 6, Article ID 115880, 2020.
- [8] J. Hu, C. Ye, Y. Ding, J. Tang, and S. Liu, "A distributed MPC to exploit reactive power V2G for real-time voltage regulation in distribution networks," *IEEE Transactions on Smart Grid*, vol. 13, no. 1, pp. 576–588, 2022.
- [9] H. Wang, K. Hou, J. Zhao, X. Yu, H. Jia, and Y. Mu, "Planning-oriented resilience assessment and enhancement of integrated electricity-gas system considering multi-type natural disasters," *Applied Energy*, vol. 315, Article ID 118824, 2022.
- [10] M. Zhang, X. Wang, and W. Mi, "Marine current generators multimachine system based on high order sliding mode control," *Journal of Coastal Research*, vol. 103, no. 1, pp. 373–377, 2020.
- [11] T. E. Lin and H. Xu, "A post-processing method for detecting unknown intent of dialogue system via pre-trained deep neural network classifier," *Knowledge-Based Systems*, vol. 186, Article ID 104979, 2019.
- [12] Y. Chen and D. Miao, "Granular regression with a gradient descent method," *Information Sciences*, vol. 537, no. 4, pp. 246–260, 2020.
- [13] C. Lu, H. Zhou, L. Li et al., "Split-core magnetoelectric current sensor and wireless current measurement application," *Measurement*, vol. 188, Article ID 110527, 2022.
- [14] S. Lu, Y. Ban, X. Zhang et al., "Adaptive control of time delay teleoperation system with uncertain dynamics," *Frontiers in Neurobotics*, vol. 16, Article ID 928863, 2022.
- [15] X. L. Qi, D. G. Zhang, H. Y. Guo, and Y. X. Chen, "The fairing arrangement for vortex induced vibration suppression effect in soliton current," *Journal of Coastal Research*, vol. 103, no. 1, pp. 293–298, 2020.
- [16] Z. Deng, Y. Li, H. Zhu, K. Huang, Z. Tang, and Z. Wang, "Sparse stacked autoencoder network for complex system monitoring with industrial applications," *Chaos, Solitons & Fractals*, vol. 137, Article ID 109838, 2020.
- [17] T. Chen, H. Yin, X. Yuan, Y. Gu, F. Ren, and X. Sun, "Emotion recognition based on fusion of long short-term memory networks and SVMs," *Digital Signal Processing*, vol. 117, no. 1, Article ID 103153, 2021.
- [18] Z. Miao, X. Meng, and L. Liu, "Analyzing and optimizing the power generation performance of thermoelectric generators based on an industrial environment," *Journal of Power Sources*, vol. 541, Article ID 231699, 2022.
- [19] X. Xu, D. Niu, B. Xiao, X. Guo, L. Zhang, and K. Wang, "Policy analysis for grid parity of wind power generation in China," *Energy Policy*, vol. 138, Article ID 111225, 2020.
- [20] H. H. Tsao, Y. G. Leu, and L. F. Chou, "A center-of-concentrated-based prediction interval for wind power forecasting," *Energy*, vol. 237, no. 15, Article ID 121467, 2021.
- [21] M. Behmanesh, P. Adibi, J. Chanussot, C. Jutten, and S. M. S. Ehsani, "Geometric multimodal learning based on local signal expansion for joint diagonalization," *IEEE Transactions on Signal Processing*, vol. 69, pp. 1271–1286, 2021.
- [22] S. Park, S. Yoon, B. Lee, S. Ko, and E. Hwang, "Probabilistic forecasting based joint detection and imputation of clustered bad data in residential electricity loads," *Energies*, vol. 14, no. 1, p. 165, 2020.
- [23] N. Mehmood, Z. Liang, and J. Khan, "Diffuser augmented horizontal axis tidal current turbines," *Research Journal of Applied Sciences Engineering & Technology*, vol. 4, no. 18, pp. 3522–3532, 2012.

- [24] Y. Wang and N. Bianchi, "Modeling and investigation of self-excited reluctance generators for wind applications," *IEEE Transactions on Industry Applications*, vol. 55, no. 6, pp. 5809–5817, 2019.
- [25] J. Xu, L. Jin, L. Liang, Z. Feng, and D. Xie, "A new humanlike facial attractiveness predictor with cascaded fine-tuning deep learning model," *Computer Science*, vol. 70, no. 1, pp. 45–79, 2015.
- [26] J. Ma, X. Shi, L. Wang, and T. Li, "Short-term prediction of wind power with consideration of information fusion," *Computer Simulation*, vol. 37, no. 7, pp. 137–143, 2020.
- [27] Z. Chang, Y. Chen, W. Yang, and Z. Zhou, "Mission planning problem for optical video satellite imaging with variable image duration: a greedy algorithm based on heuristic knowledge," *Advances in Space Research*, vol. 66, no. 11, pp. 2597–2609, 2020.
- [28] H. Li, K. Hou, X. Xu, H. Jia, L. Zhu, and Y. Mu, "Probabilistic energy flow calculation for regional integrated energy system considering cross-system failures," *Applied Energy*, vol. 308, Article ID 118326, 2022.
- [29] Y. Zhang, X. Shi, H. Zhang, Y. Cao, and V. Terzija, "Review on deep learning applications in frequency analysis and control of modern power system," *International Journal of Electrical Power & Energy Systems*, vol. 136, Article ID 107744, 2022.
- [30] M. Dong and L. S. Grumbach, "Combining unsupervised and supervised learning for asset class failure prediction in power systems," *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 5033–5043, 2019.