Energy Engineering

Tech Science Press

Doi:10.32604/ee.2026.074698

ARTICLE

Multi-Source Fusion with Patch-Guided Multi-Task Learning for Power Prediction of Offshore Wind Farm Clusters

Weijia Tang, Qiang Li* and Ningyu Zhang

Research Institute, State Grid Jiangsu Electric Power Co., Ltd., Nanjing, China

*Corresponding Author: Qiang Li. Email: liq4@js.sgcc.com.cn

Received: 16 October 2025; Accepted: 05 January 2026; Published: 18 June 2026

ABSTRACT: Large-scale offshore wind farm clusters (OWFCs) have been increasingly connected to the power grid, and requires advanced forecasting models to enhance the prediction accuracy of OWFC's power output. This paper proposes a multi-source fusion with patch-guided multi-task learning for power prediction of offshore wind farm clusters. Unlike traditional graph-based approaches that rely on predefined topological relationships, which are limited in capturing the highly similar but rapidly changing meteorological conditions among closely spaced offshore farms, the proposed model employs a parameter-sharing multi-task learning network to achieves both independence and correlation among offshore wind farm clusters, followed by utilizing a dynamically weighted multi-task loss function to gradually optimize the network parameters. Moreover, the proposed model applies the patch-guided feature learning module further enables natural alignment and fusion of multi-source data. To demonstrate the performance of the proposed model, experiments were conducted on offshore wind farm clusters in three different regions. The results show that the proposed model can obtain an average accuracy improvement of around 24.31% for MAE and 19.14% for RMSE, ensuring prediction accuracy and robustness.

KEYWORDS: Wind farm clusters; patch-based feature learning; dynamic weight loss function

1 Introduction

Large-scale offshore wind farm clusters (OWFCs) have gradually been integrated into the power grid on a large scale. However, due to the complex and variable meteorological conditions at sea, their power output presents significant uncertainty and randomness [1]. Accurate forecasting of the power output of offshore wind farm clusters is essential to ensure the secure and stable operation of the power system [2]. Consequently, it is critical to develop more advanced wind power forecasting models for the power prediction of OWFCs.

Existing wind power forecasting models can be categorized into physical models and data-driven models [3]. Among them, physical models rely on how wind turbines work and the principles of wind energy conversion, using basic physics and aerodynamics to explain how wind energy is turned into electricity [4]. However, due to the influence of various factors such as weather and geographical location, physical models have limitations in capturing the nonlinear and dynamic variations of wind power [5]. Consequently, recent research has increasingly focused on data-driven approaches for wind power forecasting. Data-driven models mainly contains traditional statistical models [6] and deep learning (DL) models [7]. Specifically, statistical models, generally consisting of Autoregressive (AR), Moving Average (MA), Autoregressive Integrated Moving Average (ARIMA), et al., employ the inherent patterns and regularities within the time

series to construct models, thereby capturing the trends of wind power and enabling the prediction of future wind power generation [8]. Nonetheless, due to the rapid growth of data from wind farms, traditional statistical models exhibit significant limitations in analyzing large and high-dimensional datasets [9]. They often struggle to capture complex nonlinear correlations and temporal patterns, resulting in reduced forecasting accuracy.

DL-based models, due to its exceptional feature extraction capabilities and its strength in handling large-scale, high-dimensional, and complex data, has become a key tool for wind power forecasting. Earlier deep learning models for wind power forecasting primarily include Recurrent Neural Networks (RNN) [10], Long Short-Term Memory networks (LSTMs) [11], and Gated Recurrent Units (GRU) [12]. For example, Ref. [13] optimized the weights of RNNs using the SpCro (Sparsity-Promoting Cross-Entropy Optimization) algorithm and attained accurate wind power forecasting by considering the past 24 h of wind power data; Although RNN-based models can meet the accuracy requirements of wind power prediction, their strong dependence on long sequences often leads to vanishing or exploding gradient problems, which can degrade predictive performance [14]. To address this proposed, the LSTM-based models are used to controls information flow and gradient propagation [15], enhancing the ability to capture long-range dependencies in time series data. Ref. [16] constructed a multivariate-input wind power forecasting model based on LSTM. Ref. [17] further accounted for the temporal autocorrelation of wind power by using high-resolution numerical weather prediction data and a bidirectional LSTM to model the nonlinear relationship between features and power output. GRU-based models combine the LSTM's forget and input gates into a single update gate and feeds the output back as the memory state, thereby gaining wide use in wind power forecasting [18]. In recent years, transformer models, equipped with multi-head attention mechanisms [19], have demonstrated superior capability in handling long-range dependencies. Consequently, some studies have applied Transformers to wind power forecasting to capture long-term sequential features and enhance prediction performance [20]. Ref. [21] used the multi-head attention mechanism of transformer to capture key information for wind power prediction, it shows more superior than the LSTM model; to address the memory and computational efficiency limitations of Transformer, the Informer model introduces Prob-Sparse Attention, that focusing only on the important time steps in the sequence to reduce model complexity, thereby ensuring computational efficiency while maintaining accuracy in wind power forecasting [22]. In conclusion, transformer models demonstrably ensure a robust technical foundation, not only markedly enhancing the capacity to model long-sequence data but also facilitating innovations in model architecture and performance optimization.

In the power forecasting of wind farm clusters, a variety of data-driven approaches have been employed, including recurrent neural networks such as LSTMs, convolutional neural networks (CNNs), and more recently, graph neural networks (GNNs). For example, Ref. [23] employed a CNN-LSTM integrated forecasting model to capture spatiotemporal feature representations of wind farm clusters, and then output future predictions; Ref. [24] used GNNs to capture dynamic spatio-temporal correlation between adjacent stations, boosting the prediction accuracy of wind power clusters. CNN-based approaches, typically organize the power series of multiple farms into multi-channel inputs and apply convolutional filters to extract shared spatial-temporal features. GNN-based methods explicitly model each wind farm as a node in a graph, with edges defined by geographical proximity or statistical similarity, and generate forecasts by aggregating information from neighboring nodes.

However, existing methods still face key limitations, especially for offshore wind farm clusters. Offshore farms are often densely spaced, with highly similar yet rapidly changing meteorological conditions. Regarding traditional LSTM-based approaches, they often fail to adequately represent spatial correlations and tends to bias the predictions toward dominant farms with stronger signals. As for CNN methods, convolution

enables parameter sharing and captures local correlations, its receptive fields make it difficult to capture long-range or dynamically changing relationships among geographically dispersed farms. GNNs over-rely on adjacency matrices, which limits in capturing these dynamic and subtle inter-farm dependencies. Moreover, multi-source data fusion remains challenging in GNN frameworks. Although multiple data sources data can be attached as node features, their heterogeneous temporal and spatial resolutions make it difficult for standard GNNs to align and effectively leverage these inputs during aggregation.

To enable simultaneous prediction of power output for multiple offshore wind farms and realize superior multi-source information fusion, this paper proposes a Multi-Source Fusion with Patch-Guided Multi-Task Learning framework. Under this framework, a shared backbone network extracts common spatiotemporal features of offshore wind farm clusters but also maintains task-specific optimizaiton objectives, thus balancing inter-farm correlation and task independence without relying on predefined topological structures. Additionally, a patch-based feature learning module is designed to align heterogeneous data sources at the patch level and dynamically captures both local and global dependencies, providing a unified and adaptive feature space for multi-source fusion.

To sum up, the contributions of this paper can be summarized by:

(1) A multi-task learning framework equipped with a patch-based feature learning module is proposed for offshore wind power forecasting. The framework dynamically models inter-farm dependencies without requiring predefined graph structures, facilitates specialized task-specific representation learning, and enables seamless alignment and fusion of multi-source data.
(2) A dynamically weighted loss function that autonomously balances contributions from different wind farms is designed, thereby mitigating the dominance of any single task during parameter updates and reducing computational complexity.
(3) Experiments were carried out based on actual operational data from offshore wind farm clusters in Jiangsu Province. The results show that the proposed model surpasses existing models in both forecasting accuracy and stability, demonstrating its significant meaning for practical engineering applications.

2 Model Framework

This paper focuses on the power forecasting problem of offshore wind farm clusters within a specific region. Traditional single-task wind power model typically requires training a separate prediction model for each wind farm. In contrast, a multi-task wind power model employs shared parameters and a multi-output decoding, enabling a single unified model to simultaneously generate power forecasts for multiple wind farms in a specific offshore region. Compared with single-task models, multi-task learning enables knowledge sharing across wind farms, boosting data efficiency and overall prediction accuracy with a single unified model. Consequently, this paper proposes a multi-task prediction model with a parameter-sharing mechanism, enabling information exchange among different sites.

Specifically, the power data of each wind farm within the region, along with regional meteorological data, are first fed into the proposed model. Next, these inputs are processed by a parameter-sharing networks for feature learning. Finally, the losses between the actual and predicted values for each wind farm are calculated, and these loss functions are dynamically weighted during training to optimize the model parameters.

3 Preliminary

3.1 Attention Structure

Attention structure [25], represented by the tuple (query, key, value), plays a crucial role in capturing global dependencies within sequential data, leading to enhanced feature representation and improved performance in sequential tasks, like machine translation and time series forecasting. The core idea behind the attention is to enable the model to focus on different parts of the input sequence based on the parts' importance to the current input. Specifically, attention scores are computed by taking the dot product of the query and key. These scores are then used to calculate a weighted sum, which forms the basis of the attention structure. The detailed formulation can be expressed by:

$$Attention\left(\boldsymbol{X}_Q,\boldsymbol{X}_K,\boldsymbol{X}_V\right) = softmax\left(\frac{\boldsymbol{X}_Q \cdot \boldsymbol{X}_K{}^T}{\sqrt{d_k}}\right) \cdot \boldsymbol{X}_V \tag{1}$$

where $\boldsymbol{X}_Q,\boldsymbol{X}_K,\boldsymbol{X}_V$ corresponds to the element of the tuple, respectively.

3.2 Input Encoding

For sequential data, it is common practice to first apply an input encoding operation before feeding it into a network architecture, typically using a trainable linear projection. Specifically, the sequential data is mapped into a latent space using a 1-dimensional convolutional neural network (1D-CNN) [26] with an output dimension of d, resulting in input embeddings. To further capture the temporal order, a positional encoding is applied to each index in the sequence, and the positional encoding is then added to the input embeddings. The calculation of the input encoding is seen in Fig. 1, and the specific formulations can be expressed by,

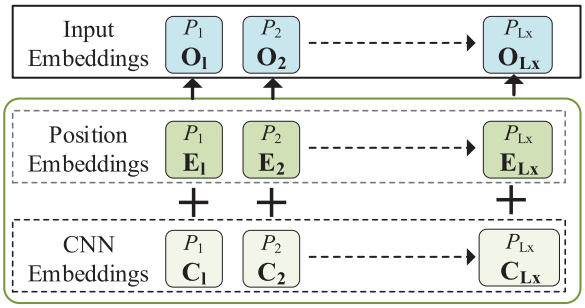

Figure 1: The embedding generations of input encoding

$$PE\left(pos\right) = \begin{cases} sin\left(pos/10000^{2j/d}\right) \\ cos\left(pos/10000^{(2j+1)/d}\right) \end{cases} \tag{2}$$

$$\boldsymbol{O} = 1DCNN\left(X\right) + PE\left(X\right) \tag{3}$$

where d is the output dimension of the 1D-CNN; $j \in \{1,\ldots,d/2\}$; pos is the positional index of the past data; L_x is the length of X; $\boldsymbol{O} \in R^{L_x \times d}$ is the input embeddings of the past data.

4 Model Structure

In Fig. 2, we present the framework of the proposed model in clear. Specifically, the proposed model contains three parts: wind power cluster data preprocessing, patch-based feature learning, and a dynamic multi-tasks decoder. Specifically, the part of wind power cluster data preprocessing introduces the model's inputs from a specific offshore region and explains the rationale behind the selection of these inputs; the patch-based feature learning section describes its underlying principle and explains its benefits in WPC's power forecasting; the dynamic multi-tasks decoder introduces the manner of output predictions for WPC. The specific explanations can be summarized in subsequent contents.

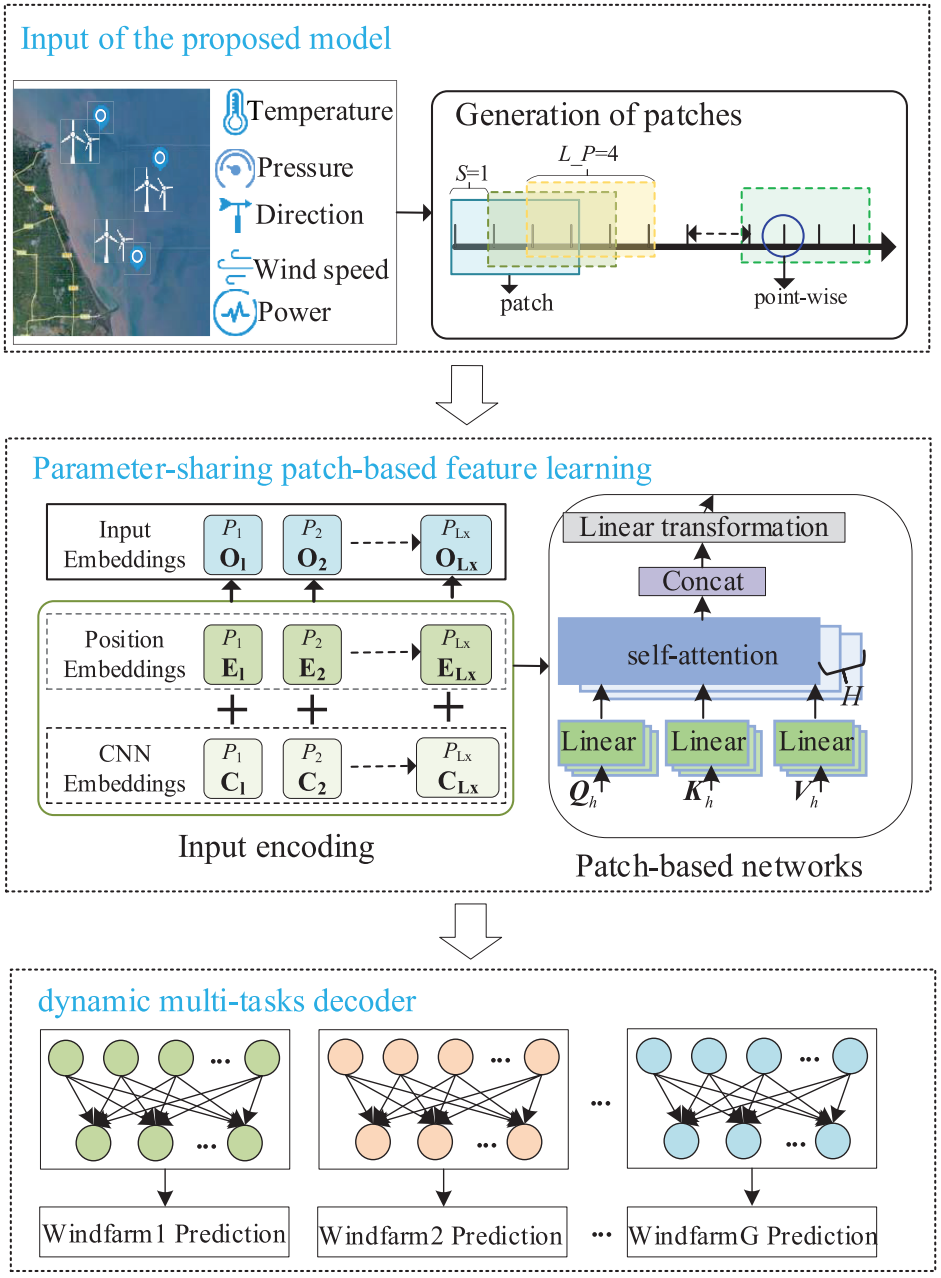

Figure 2: The framework of the proposed model

4.1 Wind Power Cluster Data Preprocessing

Within the same offshore region, wind farms present complex coupling effects, where meteorological factors collectively influence the power output fluctuations across multiple farms. Key variables include wind speed, air pressure, and temperature. Variations in wind speed directly affect the energy capture efficiency of turbines; fluctuations in air pressure alter air density, thereby impacting the stability of wind power generation; temperature changes influence atmospheric stability, which in turn causes wind speed variations that ultimately affect power output. Based on these considerations, the proposed model takes critical meteorological variables and historical power data from each wind farm within the region as inputs. By integrating these diverse data sources, the model aims to accurately forecast the future power fluctuations of multiple wind farms.

Assume that the meteorological data for a specific sea area are denoted as $\boldsymbol{X}_{\mathrm{n}}$, and the power output data of each wind farm within this area are represented by $\boldsymbol{X} = \{X_1, X_2, \ldots, X_{\mathrm{p}}\}$, where n represents the number of meteorological variables and p is the number of Offshore wind power farms in a region. Due to differences in meteorological variables and variations in installed capacities among wind farms, the corresponding power data present inconsistent scales. Therefore, before feeding the data into the prediction model, normalization is necessary to ensure training stability and enhance model performance. The specific calculations can be expressed by,

$$\boldsymbol{X}' = \frac{\boldsymbol{X} - \boldsymbol{X}_{\min}}{\boldsymbol{X}_{\max} - \boldsymbol{X}_{\min}} \tag{4}$$

where $\boldsymbol{X}_{\min}$ and $\boldsymbol{X}_{\max}$ denote the maximum and minimum values of $\boldsymbol{X}$, respectively.

4.2 Patch-Based Parameter-Sharing Feature Learning

From the above modules, it is evident that the model takes as input multiple variables, each presenting a point-wise pattern. Conventional deep learning models generally perform power forecasting by learning these point-wise features. Nonetheless, such point-wise data are often inadequate for capturing local spatial correlations and fine-grained patterns among neighboring regions. To address this limitation, this paper restructures the original point-wise wind power data into patch-based pattern to better extract detailed local features.

Specifically, for a specific variable, this paper segments it into multiple sub-sequences of length P, each referred to as a "Patch", just seen in Fig. 3. We can see that by applying this procedure, we can obtain N continuous patch for wind power, each composed of P data points. The number of patch N can be calculated by,

$$N = \left\lfloor \frac{(L - L_{\mathrm{p}})}{S} \right\rfloor + 2 \tag{5}$$

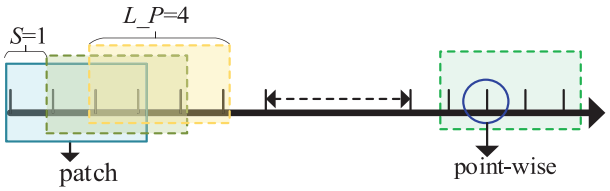

Figure 3: The generation of patches from point-wise time series

By means of the above operations, the original point-wise multivariate input data is transformed into patch-based data, denoted by X. Before feeding the patch-based data into deep learning networks, it is first encoded along the sequence length dimension (as described in Section 3) to generate input embeddings. These embeddings are then feed into the multi-head attention mechanism, with its detailed structure presented in Fig. 3. Building on the single-head attention in Eq. (1), multi-head attention first linearly maps the queries, keys, and values into n separate subspaces, computes scaled dot-product attention in parallel for each subspace, concatenates the resulting heads, and finally applies a linear projection to produce the output. The specific structure can be seen in Fig. 4. In addition, the entire process can be mathematically formulated by,

$$\boldsymbol{Q}_j = \boldsymbol{X} \cdot \boldsymbol{W}_j^Q \tag{6}$$

$$\boldsymbol{K}_j = \boldsymbol{X} \cdot \boldsymbol{W}_j^K \tag{7}$$

$$\boldsymbol{V}_j = \boldsymbol{X} \cdot \boldsymbol{W}_j^V$$

$$\boldsymbol{O}_{atten}^j = Attention\left(\boldsymbol{Q}_j, \boldsymbol{K}_j, \boldsymbol{V}_j\right) = softmax\left(\frac{\boldsymbol{Q}_j \cdot \boldsymbol{K}_j}{\sqrt{d_k}}\right) \cdot \boldsymbol{V}_j \tag{8}$$

where $\boldsymbol{W}$ and $\boldsymbol{b}$ represent the weight and bias, respectively.

Subsequently, the input matrix $\boldsymbol{X}$ is added element-wise to the multi-head attention output through a residual connection and then layer-normalized to yield the sub-layer's final representation O, detailly expressed by:

$$\boldsymbol{O}_L^j = LayerNorm\left(\boldsymbol{X} + \boldsymbol{O}_{atten}^j\right) \tag{9}$$

The output matrix $\boldsymbol{O}_L^j$ is typically fed into a feed-forward network (FFN) sub-layer, again applying a residual connection followed by layer normalization. Finally, the outputs of all sub-layers are aggregated to yield the ultimate representation. The specific equations can be expressed by,

$$\boldsymbol{O}_j = LayerNorm\left(\boldsymbol{O}_L^j + FFN\left(\boldsymbol{O}_L^j\right)\right) \tag{10}$$

$$\boldsymbol{O}_c = Concat\left(\boldsymbol{O}_1, \boldsymbol{O}_2, \ldots, \boldsymbol{O}_J\right) \cdot \boldsymbol{W}_c \tag{11}$$

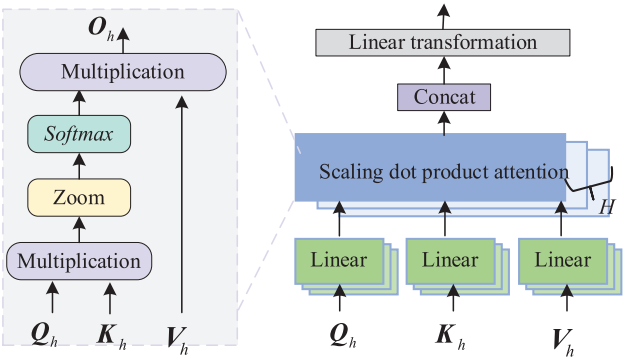

Figure 4: Multi-head attention mechanism structure diagram

4.3 Dynamic Multi-Tasks Decoder

After obtaining the aforementioned feature representations, a fully-connected layer with output dimension equal to the number of wind farms is employed to jointly forecast the future power of every plant. Subsequently, the prediction of each farm is then computed against its true power and aggregated through a weighted loss. The specific equations can be expressed by,

$$\hat{\boldsymbol{Y}} = \boldsymbol{O}_c \cdot \boldsymbol{W} + \boldsymbol{b} \tag{12}$$

where $\hat{\boldsymbol{Y}} \in R^{L \times G}$ represents the predictions for G wind farms with a prediction period of L, (G is the number of wind farms; L is the prediction period); $\boldsymbol{W}$ and $\boldsymbol{b}$ represent the weight and bias, respectively.

For the g-th wind farm, the loss is computed as the root-mean-square error between its predicted and actual values. Subsequently, to ensure adequate training for each task and enhancing overall model performance, dynamic weighting coefficients are employed to calculate the loss functions in different wind farms. The specific calculations can be expressed by,

$$L^{(g)} = MSE\left(\widehat{\boldsymbol{Y}^{(g)}}, \boldsymbol{Y}^{(g)}\right) \tag{13}$$

$$L = \sum_{g=1}^{G} \lambda_g \cdot L^{(g)} \tag{14}$$

where $L^{(g)}$ represents the loss function of g-th wind farm; L is the total loss functions; λ_g represents the coefficient at g-th wind farm, calculated by,

$$\lambda_g = \frac{K_{\text{soft}} exp\left(w_g / T_{\text{dwa}}\right)}{\sum_{g=1}^{G} exp\left(w_g / T_{\text{dwa}}\right)} \tag{15}$$

5 Case Study

5.1 Experimental Environment and Dataset Sources

The multi-wind-farm power prediction model was developed using Python 3.9 within the PyTorch 2.5.1 framework. The experiments were conducted on a hardware platform equipped with an Intel(R) Core(TM) i7-14650HX processor (24 CPUs), running the Windows 10 operating system. The case study utilized historical operational data from three offshore wind farms, with installed capacities of 200, 300, and 400 MW, respectively.

The dataset covers the period from 04 June 2019, to 02 December 2019, with a temporal resolution of 1 h. To ensure temporal continuity and avoids data leakage, enabling rigorous evaluation of the model's generalization capability across seasonal meteorological variations, data partitioning was carried out according to the following rules: (1) training set: 04 June 2019 (00:00)–27 October 2019 (10:00); (2) validation set: 27 October 2019 (11:00)–11 November 2019 (00:00); (3) testing set: 11 November 2019 (01:00)–02 December 2019 (00:00).

5.2 Evaluation Metrics and Model Parameters

The evaluation employs three metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), defined as follows:

$$MAE = \frac{1}{M}\sum_{\mathrm{m=1}}^{M}\left|y_{m,pred} - y_{m,true}\right| \tag{16}$$

$$RMSE = \sqrt{\frac{1}{M}\sum_{m=1}^{M}\left(y_{m,pred} - y_{m,true}\right)^2} \tag{17}$$

$$MAPE = \frac{1}{M}\sum_{m=1}^{M}\left|\frac{y_{m,true} - y_{m,pred}}{y_{m,true}}\right| \times 100\% \tag{18}$$

where M denotes the number of test samples, $y_{m,pred}$ represents the predicted wind power output, and $y_{m,true}$ denotes the actual observed value.

Table 1 presents the key hyperparameter of the proposed model. During training process, the Adam optimizer was used to update the model parameters, enabling to accelerate convergence and enhance overall performance.

Table 1: Key parameters of the DWA-PatchTST-MTL model

Parameter name	Value
Learning rate	1e−5
Epochs	128
Lookback window L	24
Patch length	3
Stride	4
Projection dimension D	512
Attention heads H	8
MLP hidden layers	1
MLP nodes per layer	128
Number of tasks G	3
Task-specific layer dim	128

5.3 Comparison with SoTA Models

To verify the effectiveness of the proposed model proposed in this chapter compared with the single wind power prediction model, typical SoTA models are used for comparison. The SoTA models includes NLinear [16], Transformer [17], Informer [18], RNN-GRU [19], LSTNet [20], Seq2Seq [21]. These models are the most advanced solutions in the field of time series forecasting or wind power prediction in recent years and are widely used in wind power forecasting.

To ensure a fair comparison, the proposed model and the SoTA models use the same training parameters, including learning rate, optimizer, number of epochs, and batch size. In contrast, the parameters of the model network structures differ. The parameter settings of these SoTA models are shown in Table 2. The results of multi-step prediction for different SoTA models and the proposed model are presented in Table 3.

Table 2: The structure parameters of the SoTA models

Model	Network parameters
Seq2Seq	Encoder: {LSTM: output dimensions = 128; Number of layers = 2}; Decoder: {LSTM: output dimensions = 128; Number of layers = 2}
RNN-GRU	RNN: output dimensions = 128; Number of layers = 2; GRU: output dimensions = 128; Number of layers = 2
LSTNet	CNN: output dimensions = 128; convolution kernel size k_{size} = (2,1); GRU: output dimensions = 128; Number of layers = 2
Transformer	Encoder: {output dimensions = 512; Attention Heads = 8; Number of encoders = 1}; Decoder: {output dimensions = 512; Attention Heads = 8; Number of decoders = 1}
Informer	Encoder: {output dimensions = 512; Attention Heads = 8; Number of encoders = 1}; Decoder: {output dimensions = 512; Attention Heads = 8; Number of decoders = 1}
PatchTST	Encoder: {output dimensions = 512; Attention Heads = 8; Number of encoders = 1}; Patch Length = 3, Stride = 4
NLinear	output dimensions = 128

Table 3: Comparison of the evaluation metrics of DWA-PatchTST-MTL and a single model

Model	Windfarm1		Windfarm 2		Windfarm 3	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
RNN-GRU	0.642	0.481	0.610	0.469	0.603	0.457
Seq2Seq	0.520	0.358	0.493	0.347	0.500	0.353
DLinear	0.850	0.688	0.845	0.691	0.845	0.691
Transformer	0.497	0.362	0.479	0.360	0.485	0.379
Informer	0.482	0.341	0.460	0.337	0.456	0.332
PatchTST	0.409	0.278	0.406	0.274	0.441	0.335
Proposed	0.405	0.263	0.401	0.259	0.401	0.269

From the results presented in Table 3, the proposed model consistently outperforms the SoTA counterparts, yielding lower MSE and MAE metrics across all three wind farms. In the comparison of these SoTA models, it can be observed that the latter three models generally produce better prediction accuracy than the former four. This advantage is primarily attributed to their Transformer-based architecture, which offers stronger capabilities in capturing long-term dependencies, modeling complex temporal patterns, and facilitating feature interactions, thereby enhancing the accuracy and stability of wind power forecasting. Consequently, the proposed model is developed based on the Transformer architecture to construct an advanced prediction framework.

When compared with SoTA models based on the Transformer architecture, the proposed model demonstrates a significant improvement in prediction accuracy, primarily attributed to its advantage in capturing common patterns and dynamic trends among different wind farms. Specifically, compared with the PatchTST model, the proposed model obtains reductions in RMSE of 0.9%, 1.0%, and 9.1% for the three wind farms, accompanied by MAE reductions of 5.29%, 5.58%, and 19.6%, respectively. In comparison with the Informer model, the RMSE decreases by 16.04%, 12.89%, and 11.94%; in contrast, compared with the

Transformer model, the RMSE decreases by 18.6%, 16.2%, and 17.4%, with corresponding MAE reductions of 27.1%, 28.4%, and 29%, respectively.

To further demonstrate the superiority of the proposed model, Fig. 5 presents the boxplots of various evaluation metrics for all models. Additionally, the prediction curve of the proposed model and SoTA on these wind farms are shown in Figs. 6–8, respectively. All of these indicate that the proposed model can successfully learn the coupling dependencies among multiple wind farms, significantly capture precise wind power feature representations, and thereby enabling to obtain better prediction accuracy and robustness.

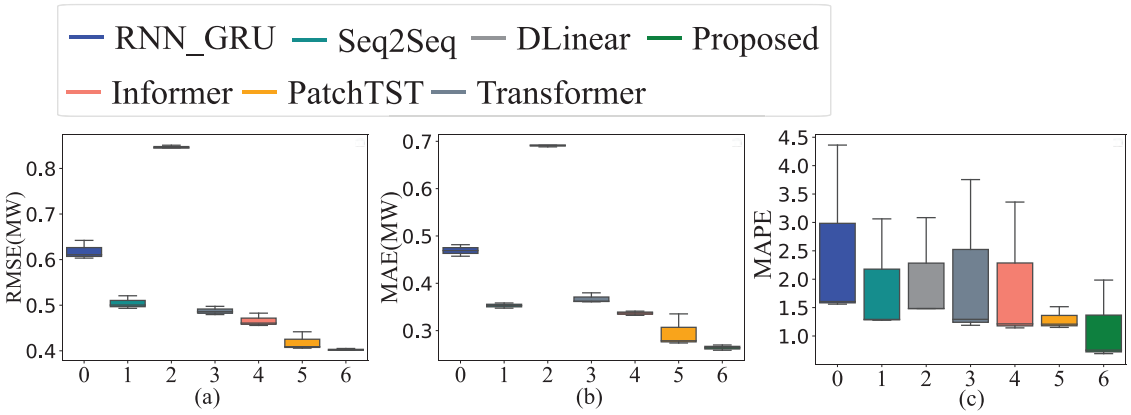

Figure 5: The boxplots of the proposed model and SoTA models. (**a**): RMSE metric; (**b**): MAE metric; (**c**): MAPE metric

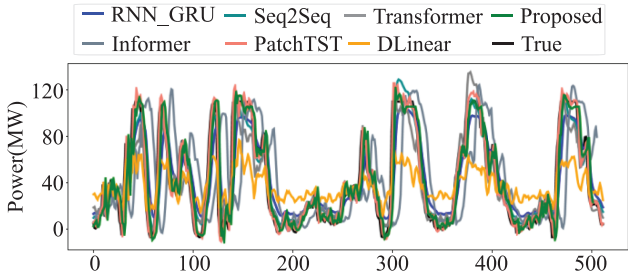

Figure 6: The prediction curve of the proposed model and SoTA models in Windfarm 1

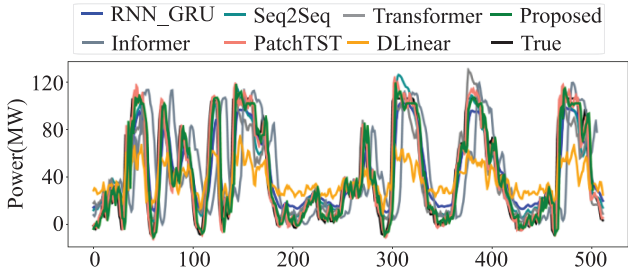

Figure 7: The prediction curve of the proposed model and SoTA models in Windfarm 2

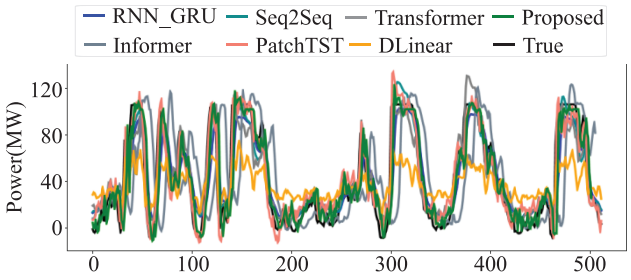

Figure 8: The prediction curve of the proposed model and SoTA models in Windfarm 3

Finally, to compare the performance of the proposed model with the SoTA models in key wind power periods (such as ramping periods), we further calculated the prediction accuracy of each model specifically during the ramping periods. The boxplots of RMSE, MAPE for these models are presented in Fig. 9. It can be seen that the proposed model demonstrates significantly lower error dispersion, approximately 0.16 for RMSE and 0.5 for MAPE, and consistently outperforms all baselines during ramping periods. This gain in prediction performance yields significant implications for grid dispatch: by producing more accurate and stable predictions during rapid wind power fluctuations, the model enables operators to better anticipate sudden ramps, adjust generation schedules more effectively, reduce reliance on costly reserve power, and minimize the risk of imbalance in the power system.

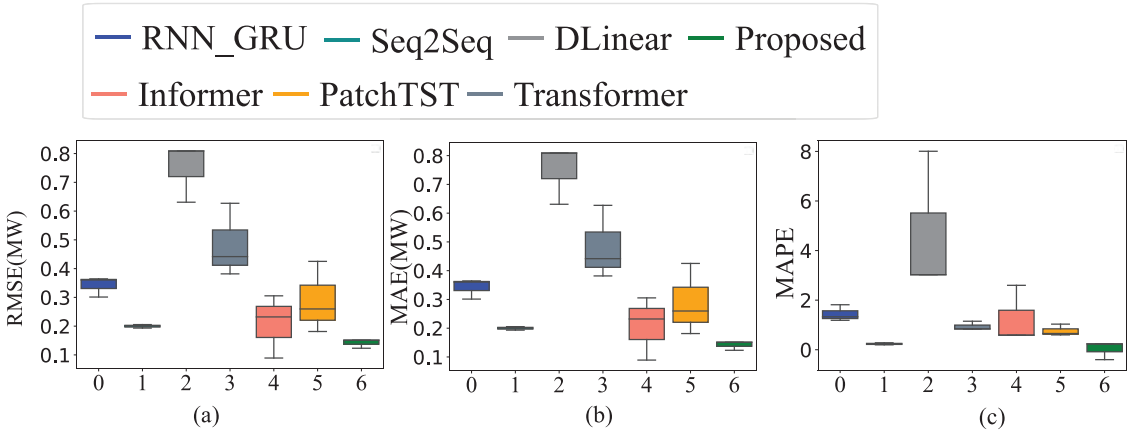

Figure 9: The boxplots of RMSE and MAPE in wind power ramps. (**a**): RMSE metric; (**b**): MAE metric; (**c**): MAPE metric

5.4 Comparison between Single-Task and Multi-Tasks

To systematically assess the applicability and potential advantages of the proposed multi-task modeling approach in wind power forecasting, we compare its forecasting results across different wind farms with those obtained from corresponding single-task models. The change of prediction metrics under 1–4 h prediction periods for both single-task and multi-task models are presented in Figs. 10–12, corresponding to Wind Farm 1, Wind Farm 2, and Wind Farm 3, respectively.

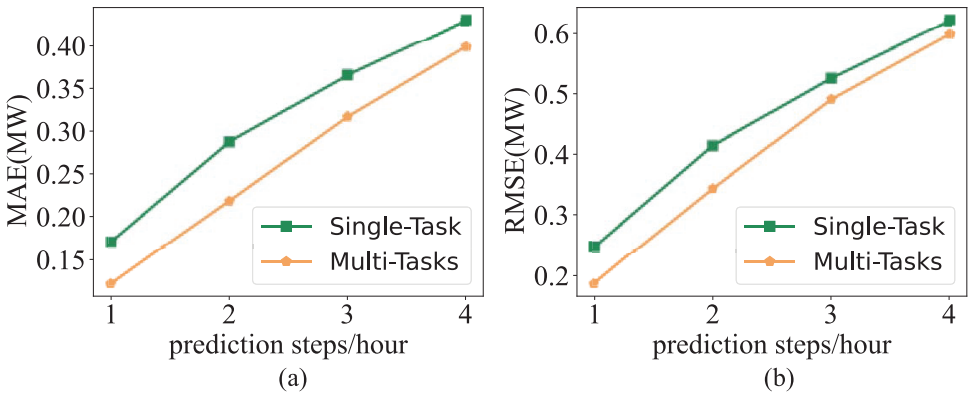

Figure 10: The change of prediction metrics for Wind Farm 1. (**a**): RMSE metric; (**b**): MAE metric

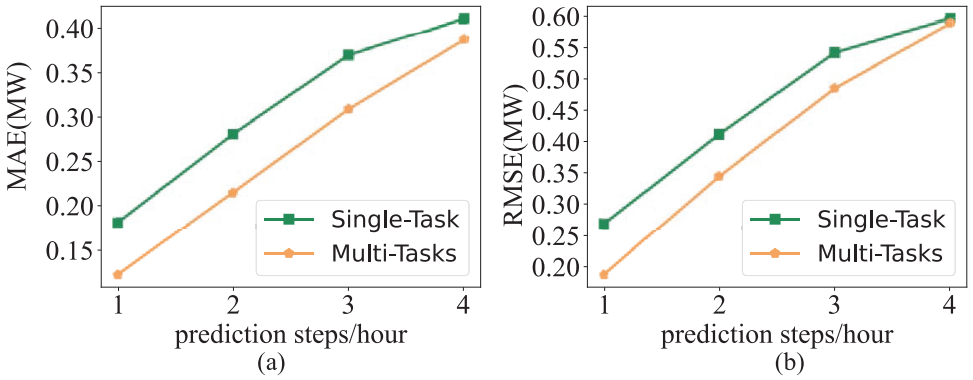

Figure 11: The change of prediction metrics for Wind Farm 2. (**a**): RMSE metric; (**b**): MAE metric

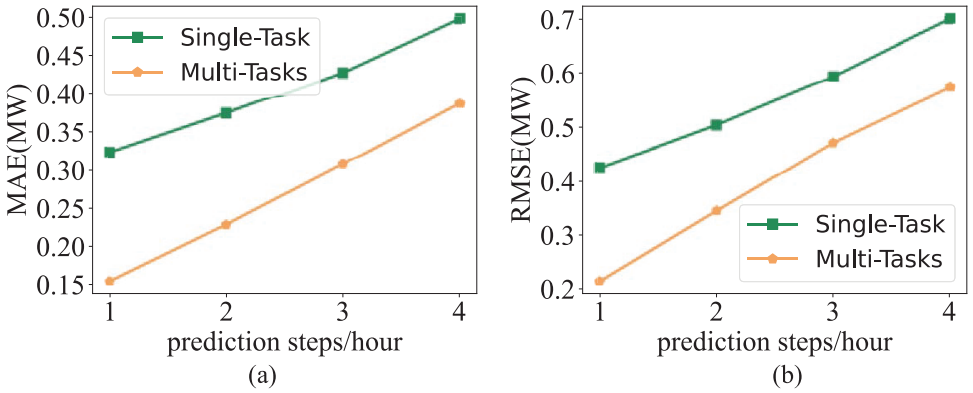

Figure 12: The change of prediction metrics for Wind Farm 3. (**a**): RMSE metric; (**b**): MAE metric

These figures above clearly demonstrate that multi-task modeling of wind farm clusters attains forecasting accuracy significantly superior to that of single-task models, with an average increase of approximately 24.31% for MAE and 19.14% for RMSE over single-task models, underscoring the efficacy of multi-task modeling in capturing wind power dynamics and enhancing model generalization. Specifically, for Wind Farm 1, the proposed multi-task model achieves an average improvement in forecasting accuracy of

approximately 18.19% and 12.94% over the single-task model; for Wind Farm 2, the improvements are around 19.46% and 14.58%; and for Wind Farm 3, multi-task modeling attains the largest percentage improvement, with increases of 35.28% and 29.90%. These results clearly demonstrate the consistent and substantial gains in prediction accuracy based on the proposed multi-task modeling across different wind farms, highlighting its effectiveness in capturing the dynamic patterns of wind power generation.

5.5 Ablation Experiments

To systematically evaluate the specific contributions of each module in the proposed model, we design a series of ablation experiments. In each experiment, a specific module is removed or replaced to analyze changes in model performance. The specific ablation experiments are summarized below.

Ablation 1: using the basic transformer to replace the parameter-sharing patch-based feature learning module, to quantify the contribution of the patch-based temporal modeling.

Ablation 2: using static loss function to replace dynamic weighted loss function, to present the function of the proposed dynamic balanced loss function.

Fig. 13 presents the comparison results between ablation models and the completed model on three wind farms. It can be seen that the complete model consistently outperforms all ablation models in both MAE and RMSE across each wind farm, with average improvements of approximately 43.6% and 34.8%, respectively. Notably, compared with ablation 1, the complete model obtains the most pronounced performance gain, underscoring the significant contribution of the parameter-sharing patch-based feature learning module to feature learning and prediction accuracy.

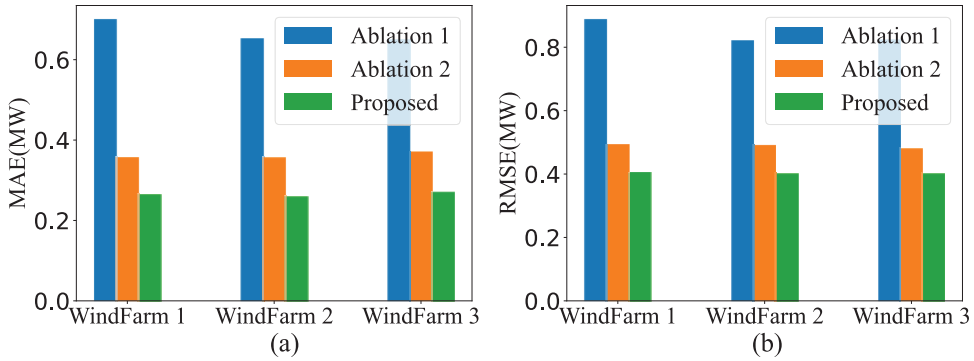

Figure 13: The comparison of prediction metrics between the completed proposed model and its ablation models. (**a**): RMSE metric; (**b**): MAE metric

5.6 Parameter Analysis

This section analyzes the impact of various parameters on the performance of the proposed model in wind power forecasting through a detailed parameter analysis. The goal is to systematically determine the optimal parameter settings that enhance prediction accuracy and improve the stability and reliability of the forecasting results for offshore wind farm clusters. Specifically, the specific parameters studied include the time span of historical wind power data, the learning rate, and the weighting parameters within the multi-task balance framework. These parameters are systematically examined to understand their influence on model performance and to guide the selection of optimal settings for accurate and stable wind power forecasting.

(1) Length of historical wind power data

In wind power forecasting, historical power generation sequences embed essential information on wind speed fluctuations, turbine operating conditions, and meteorological variations. Employing in-depth mining and feature extraction on these data enables the robust capture of temporal dependencies and latent patterns in wind power output, thereby enhancing model performance. The selected time span of historical inputs not only shapes the model's capacity to capture both short- and long-term correlations, but also has a profound impact on forecasting accuracy and generalization. Consequently, determining an appropriate input length is a critical step in developing accurate wind power forecasting models. Fig. 14 presents the change of the average prediction metrics with the length of historical wind power data, including RMSE and MAE, across three offshore wind farms.

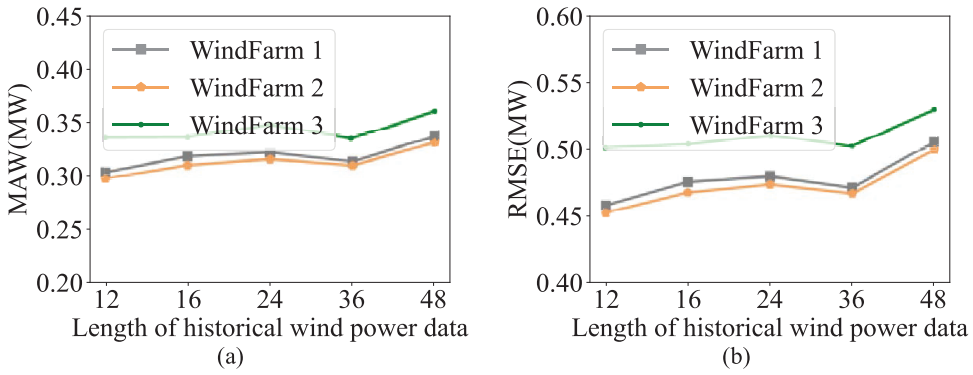

Figure 14: The change of prediction metrics under different length of historical wind power data. (**a**): RMSE metric; (**b**): MAE metric

It can be seen that when the historical wind power data span is less than 36 h, the proposed model yields consistently low RMSE and MAE across all three wind farms. Specifically, RMSE and MAE for wind farms 1 and 2 are approximately 0.3 and 0.45, respectively, and for wind farm 3, they are around 0.35 and 0.5. When the time span exceeds 36 h, prediction errors increase for all sites. This can be attributed to the difficulty deep learning models face in capturing long-range dependencies as the length of historical data grows, thereby limiting prediction performance.

(2) Learning rate

Learning rate is critical in wind power forecasting, as it determines the step size of parameter updates and the convergence speed of the model, in addition to influencing training stability and prediction accuracy. Consequently, properly configuring or dynamically adjusting the learning rate is essential to enhance forecasting performance and generalization ability. In this paper, we examine the impact of different learning rates (0.1, 0.01, 0.001, 0.0001, 0.00001, and 0.000001) on the accuracy of offshore wind power forecasting, just presented in Fig. 15.

It can be seen that the proposed model demonstrates poor performance at a learning rate of 0.1. When the learning rate is set to 0.01 or 0.0001, the prediction accuracy declines, with the mean absolute error (MAE) and root mean square error (RMSE) approximately 0.3 and 0.5, respectively. Notably, reducing the learning rate below 0.0001 does not improve the model's performance, indicating that the model is highly sensitive to the learning rate. Both excessively high and excessively low learning rates result in suboptimal convergence, thereby adversely affecting prediction accuracy.

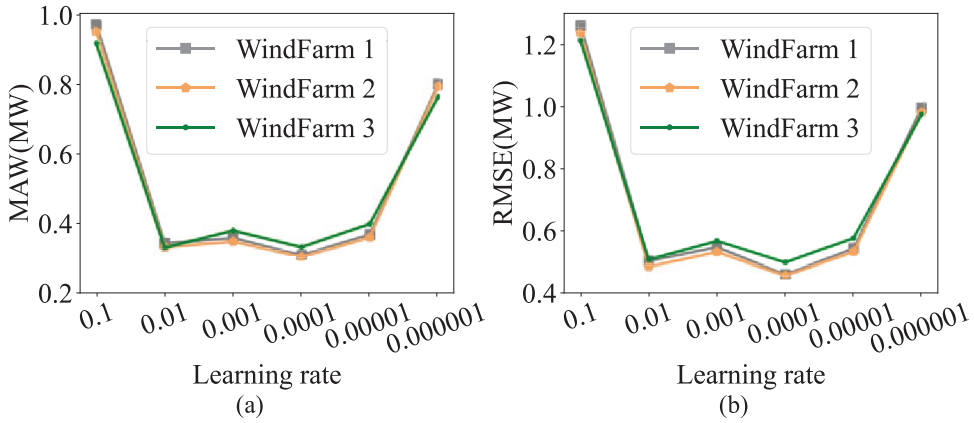

Figure 15: The change of prediction metrics under different learning rate. (**a**): RMSE metric; (**b**): MAE metric

5.7 Computational Complexity

To evaluate the computational cost of the proposed model, we measured both the training time required during the model learning process and the inference time consumed in the prediction proces. These results were then compared with those obtained from a counterpart model without parameter sharing, in order to demonstrate the efficiency benefits introduced by the shared parameter design.

The detailed comparison outcomes are summarized in Table 4.

Table 4: The computational complexity of the proposed model

	Parameter-sharing	No parameter-sharing
Training time	2500 s	6466 s
Prediction time	25 s	61 s

6 Conclusion

Considering the coupling relationship between offshore wind farms, a multi-source fusion with patch-guided multi-task learning is proposed model for power prediction of offshore wind farm clusters. The proposed model not only can balance the contributions from different wind farms by the design of multi-tasks learning, but also keep a balance between shared and differentiated modeling by a shared backbone network. In addition, the shared backbone network can capture dynamic feature representations from the power of wind power clusters but also fuse multi-source data, like meteorological data, power data, and numerical weather prediction data.

We perform multiple comparison analysis on three offshore wind farms to demonstrate the superior of our proposed framework. Specifically, compared with single-task, the multi-task modeling of wind farm clusters attains forecasting accuracy significantly superior, with an average increase of approximately 24.31% for MAE and 19.14% for RMSE. In addition, the results of section “Comparison with SoTA models” indicate that the proposed framework is more powerful than other SoTA schemes. All of these results present that the proposed model can successfully capture precise wind power feature representations, enabling to obtain better prediction accuracy and robustness.

Considering the drastic changes in marine meteorological conditions, in future study, we should focus on the coupling between various wind farms under extreme meteorological conditions and incorporate more meteorological data into prediction models to enhance prediction accuracy.

Acknowledgement: The authors would like to thank for the Science and Technology Project of State Grid Jiangsu Electric Power Co., Ltd. (J2025006) due to offering experimental resources.

Funding Statement: This paper is supported by the Science and Technology project of State Grid Jiangsu Electric Power Co., Ltd. (J2025006).

Author Contributions: Weijia Tang: Methodology, Software, Writing—original draft, Writing—review & editing. Qiang Li: Formal analysis, Writing—review & editing. Ningyu Zhang: Writing—review & editing. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: Data that support the findings of this study are available from the corresponding author, upon reasonable request.

Ethics Approval: Not applicable.

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

References

1. Ayodele TR, Jimoh A, Munda JL, Tehile AJ. Challenges of grid integration of wind power on power system grid integrity: a review. Int J Renew Energy Res. 2012;2(4):618–26.
2. Foley AM, Leahy PG, Marvuglia A, McKeogh EJ. Current methods and advances in forecasting of wind power generation. Renew Energy. 2012;37(1):1–8. doi:10.1016/j.renene.2011.05.033.
3. Tawn R, Browell J. A review of very short-term wind and solar power forecasting. Renew Sustain Energy Rev. 2022;153:111758. doi:10.1016/j.rser.2021.111758.
4. Wu B, Song M, Chen K, He Z, Zhang X. Wind power prediction system for wind farm based on auto regressive statistical model and physical model. J Renew Sustain Energy. 2014;6:013101. doi:10.1063/1.4861063.
5. Tian Z. A state-of-the-art review on wind power deterministic prediction. Wind Eng. 2021;45(5):1374–92. doi:10.1177/0309524x20941203.
6. Ozkan MB, Karagoz P. A novel wind power forecast model: statistical hybrid wind power forecast technique (SHWIP). IEEE Trans Ind Inform. 2015;11(2):375–87. doi:10.1109/TII.2015.2396011.
7. Deng X, Shao H, Hu C, Jiang D, Jiang Y. Wind power forecasting methods based on deep learning: a survey. Comput Model Eng Sci. 2020;122(1):273–301. doi:10.32604/cmes.2020.08768.
8. Aggarwal SK, Gupta M. Wind power forecasting: a review of statistical models-wind power forecasting. Int J Energy Sci. 2013;3(1):1–10.
9. Wang Y, Zou R, Liu F, Zhang L, Liu Q. A review of wind speed and wind power forecasting with deep neural networks. Appl Energy. 2021;304(1):117766. doi:10.1016/j.apenergy.2021.117766.
10. Cali U, Sharma V. Short-term wind power forecasting using long-short term memory based recurrent neural network model and variable selection. Int J Smart Grid Clean Energy. 2019;2019:103–10. doi:10.12720/sgce.8.2.103-110.
11. Liu X, Zhou J. Short-term wind power forecasting based on multivariate/multi-step LSTM with temporal feature attention mechanism. Appl Soft Comput. 2024;150:111050. doi:10.1016/j.asoc.2023.111050.
12. Liu F, Tao Q, Yang D, Sidorov D. Bidirectional gated recurrent unit-based lower upper bound estimation method for wind power interval prediction. IEEE Trans Artif Intell. 2022;3(3):461–9. doi:10.1109/TAI.2021.3123928.
13. Kumar K, Prabhakar P, Verma A. Forecasting wind power using optimized recurrent neural network strategy with time-series data. Optim Control Appl Meth. 2024;45(4):1798–814. doi:10.1002/oca.3122.

14. Kumar D, Mathur HD, Bhanot S, Bansal RC. Forecasting of solar and wind power using LSTM RNN for load frequency control in isolated microgrid. Int J Model Simul. 2021;41(4):311–23. doi:10.1080/02286203.2020.1767840.
15. Joseph LP, Deo RC, Prasad R, Salcedo-Sanz S, Raj N, Soar J. Near real-time wind speed forecast model with bidirectional LSTM networks. Renew Energy. 2023;204:39–58. doi:10.1016/j.renene.2022.12.123.
16. López G, Arboleya P. Short-term wind speed forecasting over complex terrain using linear regression models and multivariable LSTM and NARX networks in the *Andes* Mountains. Ecuador Renew Energy. 2022;183:351–68. doi:10.1016/j.renene.2021.10.070.
17. Li M, Yang M, Yu Y, Li P, Wu Q. Short-term wind power forecast based on continuous conditional random field. IEEE Trans Power Syst. 2024;39(1):2185–97. doi:10.1109/TPWRS.2023.3270662.
18. Zheng W, Chen G. An accurate GRU-based power time-series prediction approach with selective state updating and stochastic optimization. IEEE Trans Cybern. 2022;52(12):13902–14. doi:10.1109/TCYB.2021.3121312.
19. Wen Q, Zhou T, Zhang C, Chen W, Ma Z, Yan J, et al. Transformers in time series: a survey. arXiv:2202.07125. 2022.
20. Tang B, Matteson DS. Probabilistic transformer for time series analysis. In: Proceedings of the 35th International Conference on Neural Information Processing Systems; 2021 Dec 6–14; Virtual. p. 23592–608. doi:10.52202/079017-0415.
21. Qu K, Si G, Shan Z, Kong X, Yang X. Short-term forecasting for multiple wind farms based on transformer model. Energy Rep. 2022;8:483–90. doi:10.1016/j.egyr.2022.02.184.
22. Peng D, Liu Y, Wang D, Luo L, Zhao H, Qu B. Short-term PV-Wind forecasting of large-scale regional site clusters based on FCM clustering and hybrid Inception-ResNet embedded with Informer. Energy Convers Manag. 2024;320:118992. doi:10.1016/j.enconman.2024.118992.
23. Zhang J, Liu D, Li Z, Han X, Liu H, Dong C, et al. Power prediction of a wind farm cluster based on spatiotemporal correlations. Appl Energy. 2021;302:117568. doi:10.1016/j.apenergy.2021.117568.
24. Liu X, Zhang Y, Zhen Z, Xu F, Wang F, Mi Z. Spatio-temporal graph neural network and pattern prediction based ultra-short-term power forecasting of wind farm cluster. IEEE Trans Ind Appl. 2024;60(1):1794–803. doi:10.1109/TIA.2023.3321267.
25. Xiong B, Lou L, Meng X, Wang X, Ma H, Wang Z. Short-term wind power forecasting based on Attention Mechanism and Deep Learning. Electr Power Syst Res. 2022;206:107776. doi:10.1016/j.epsr.2022.107776.
26. Fukuoka R, Suzuki H, Kitajima T, Kuwahara A, Yasuno T. Wind speed prediction model using LSTM and 1D-CNN. J Signal Process. 2018;22(4):207–10. doi:10.2299/jsp.22.207.