applied sciences

MDPI

Article

DHGAR: Multi-Variable-Driven Wind Power Prediction Model Based on Dynamic Heterogeneous Graph Attention Recurrent Network

Mingrui Xu [1], Ruohan Zhu [2], Chengming Yu [3,*] and Xiwei Mi [4]

[1] College of Science, China University of Petroleum, Qingdao 266580, China; 2209060405@s.upc.edu.cn
[2] Institute of Logistics Science and Engineering, Shanghai Maritime University, Shanghai 201306, China; ruohan_zhu@163.com
[3] Institute of Artificial Intelligence & Robotics (IAIR), Key Laboratory of Traffic Safety on Track of Ministry of Education, School of Traffic and Transportation Engineering, Central South University, Changsha 410075, China
[4] Collaborative Innovation Center of Railway Traffic Safety, School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China; mixiwei@bjtu.edu.cn
* Correspondence: 1109150133@csu.edu.cn

Abstract: Accurate and stable wind power prediction is essential for effective wind farm capacity management and grid dispatching. Wind power generation is influenced not only by historical data, but also by turbine conditions and external environmental factors, such as weather. Although deep learning has made significant progress in the field of wind power forecasting, it often fails to account for two key characteristics of the data: dynamic variability and heterogeneity. Specifically, the influence of external variables on wind power changes over time, and due to the diverse nature of the information carried by different variables, simple weighted fusion approaches are insufficient to fully integrate heterogeneous data. To address these challenges, this paper introduces a dynamic heterogeneous graph attention recurrent network (DHGAR), which incorporates dynamic graphs, heterogeneous graph attention mechanisms, and gated recurrent units. Dynamic graphs capture real-time associations between wind power and external variables, while heterogeneous graph attention allows for more effective aggregation of diverse information. These two components are integrated into the gated recurrent units, replacing traditional fully connected layers to better capture temporal dependencies in the wind power time series. Experimental results on three real-world datasets demonstrate the superior performance and practical applicability of the proposed model.

Keywords: wind power prediction; multiple variates; flow-aware dynamic graph; heterogeneous graph attention; gated recurrent unit

check for updates

Academic Editor: Rodolfo Dufo-López

Received: 29 October 2024
Revised: 26 January 2025
Accepted: 10 February 2025
Published: 11 February 2025

Citation: Xu, M.; Zhu, R.; Yu, C.; Mi, X. DHGAR: Multi-Variable-Driven Wind Power Prediction Model Based on Dynamic Heterogeneous Graph Attention Recurrent Network. *Appl. Sci.* **2025**, *15*, 1862. https://doi.org/10.3390/app15041862

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Due to a shortage of fossil fuels and concerns about environmental pollution, governments around the world have begun to vigorously promote the development of clean energy [1]. Wind and solar energy are currently among the most important clean energy sources [2]. However, wind power generation is characterized by unstable amplitude and frequency and the presence of harmonics, and its output is significantly affected by environmental conditions [3]. These factors make it difficult for wind power to be directly integrated into the grid and it imposes additional burdens on grid operations, including

peak regulation, frequency adjustment, and voltage control [4]. Therefore, in order to ensure grid security and improve power supply quality, it is of great significance to effectively regulate wind power output currents and mitigate their negative impact on the grid.

Currently, power distribution networks mainly regulate wind power frequency and output through methods such as generator adjustments, substation control, and energy storage systems [5]. Wind power forecasting plays a crucial role in this process by providing real-time data for the formulation of regulation strategies [6]. With the predicted wind power data, the power system can develop load balancing and frequency regulation plans in advance, and wind turbines can also adjust in real time [7]. Developing a stable and accurate wind power forecasting method is of utmost importance.

Given that accurate wind power forecasting methods significantly impact the power regulation process, how we can achieve further improvements in this area has become a deeply researched question. Unlike general time series forecasting problems, wind power series exhibit nonlinear characteristics and are strongly influenced by environmental factors [8]. Thus, prediction accuracy improvement cannot rely solely on historical data and model performance. Instead, it requires a comprehensive approach that also considers environmental factors and the characteristics of wind turbines. Additionally, environmental variables such as wind speed and wind direction dynamically influence wind power over time [9]. Capturing this dynamic feature correlation in the model is essential for enhancing forecasting stability.

Many scholars have conducted extensive research, and various models, such as Long Short-Term Memory (LSTM) [10], hybrid models [9], ensemble models [11], and graph models [12], have been applied in this field. These studies mainly focus on two aspects: data preprocessing and model improvement. Data preprocessing enhances prediction accuracy by transforming data into a format that is easier to learn, like Variational Mode Decomposition (VMD) [13], or by extracting and selecting relevant features, like convolutional neural networks (CNNs) [14]. It is also crucial to improve the model structure and make it more appropriate for wind power prediction. For instance, ensemble models enhance robustness by combining multiple predictors [15] and graph models incorporate information about the spatial relationships between wind turbine locations to improve prediction accuracy [16]. Current research has made advancements in wind power prediction. However, to our knowledge, few studies have examined the dynamic correlation among different features and the heterogeneity in how different environmental factors influence wind power has not been sufficiently addressed. Therefore, to fill this gap, it is necessary to develop a wind power prediction model that makes full use of environmental features and achieves dynamic learning of data correlations, which is the focus of this paper.

2. Related Works

Wind power forecasting is significant for grid regulation, but the nonlinear and fluctuating nature of wind power presents challenges for modeling [17]. Moreover, it is highly influenced by environmental factors, which further increases the difficulty of prediction. Plenty of scholars have conducted in-depth research on how to improve the accuracy of wind power forecasting. It has been found that the proper utilization of multiple environmental factors and the design of effective model structures are key to enhancing the accuracy and stability of wind power forecasting [18], and these are the two main aspects emphasized in this study. In addition, this research also considers the dynamic impact of data flow changes on model performance.

Initially, single-factor forecasting methods were applied to the field of wind power prediction which extract temporal variation information from historical wind power sequences and work out prediction results. After extensive research, many single-factor forecasting

models suitable for wind power prediction have been proposed. Ban [19] and his team modified Autoformer with Mixture of Experts (MOE) decomposition to adapt it to long-term univariate wind speed forecasting, which proved to be useful. Akbal and Ünlü [20] proposed a univariate sequence-to-sequence model to achieve wind production prediction. It was effective in short-term and mid-term forecasts, but the accuracy decreased when the forecast steps were increased. Ahn and Hur [21] utilized enhanced wavelet transform to preprocess the wind power series, and then adopted the autoregressive integrated moving average with the explanatory variable (ARIMAX) method to obtain prediction results. Although the transformed wind power data were easier to model, the performance of the whole model was still limited due to a lack of external factors and the linear predictor. Several other univariate models [22,23] have been proposed in this field, but since the environmental features have not been taken into consideration, the performance of these methods is restricted.

Therefore, the multi-variate wind power forecast is developed to alleviate this bottleneck. By taking external features into account, like wind speed, wind direction, generator conditions, and so on, the basic input information is more abundant, thereby improving the model performance [24]. Chen et al. [25] employed wind turbine monitoring data and numerical weather prediction (NWP) data and designed a new wind power prediction framework. The NWP data proved to be effective in wind power forecasting. Zhao et al. [12] proposed a new neural network model that utilized spatial information and took several wind power series from the different stations as inputs. The performance of this spatial–temporal model was satisfying, but there is still room for enhancement in terms of model structure. In Sulaiman's research [26], generator conditions like temperature, generator speed, and rotor speed were considered and proved to be helpful. Plenty of other studies [27–29] on discovering and extracting useful external features have also been conducted, and it is evident that the use of environmental factors allows for promotion in the whole model's performance. However, more related features always lead to more information redundancy [30]. It is crucial to explore ways of rationally and carefully using the information contained in those features.

It is urgent to design an appropriate model structure that can make full use of the information provided in the features and get rid of redundant characteristics. Plenty of scholars have put effort into the improvement of multi-variate prediction models and achieved some positive results. Firstly, due to the changeable dimensions of neural networks, many single-factor forecasting models have been directly applied to the multi-factor forecasting domain, such as LSTM [31], recurrent neural networks (RNNs) [32], and temporal convolutional networks (TCNs) [33]. However, these approaches can increase model complexity to some extent, tend to overfit, and capture the relationships between features inadequately. Therefore, some studies attempt to improve overall model efficiency and performance by employing feature preprocessing or pre-training methods. Bhukya used fractional order neural networks to preprocess and fill in the missing data, which enhanced the final prediction results. Yang et al. [34] employed multi-variate VMD and elastic net to preprocess the original features and select useful ones, and then performed CNN and LSTM model to obtain prediction results. The experiments demonstrate its effectiveness, but feature selection methods are usually unstable when the importance of features varies with the data flow. Peng et al. [35] proposed an interpretable AI solution with interactive features, which improved the interpretability of the prediction model and could effectively employ the information contained in various features. Tan et al. [36] proposed a convolution attention fusion block to achieve feature fusion, and also used a deep temporal LSTM model for prediction, which reached satisfying results. Like Tan's models, in [37,38], the authors have also explored the attention mechanism in wind data analysis and demonstrated its

effectiveness. However, the above methods may ignore the basic connections among the features and increase model complexity.

To address this issue, graph attention can be a good choice, which is more efficient and interpretable. In common sense, graph learning is always used to deal with spatial correlation or process features of graph form [39]. Yang [40] and his team utilized graph attention to realize large-scale wind power prediction. By taking different grid nodes as graph nodes, and distance as edges, the method gave a good framework for large-scale wind power prediction. Arrieta-Prieto and Schell [41] proposed a spatially transferable model to realize wind power prediction and achieved good performance, but it ignored the correlation among environmental features. Qiu et al. [42] combined a graph neural network with a gated dilated inception network, which was useful for spatial wind power forecasting, but the feature analysis was ignored in the study and restricted the model performance. To the best of our knowledge, most of the studies using graph learning in wind power prediction concentrate on spatial correlation modeling [43,44], while very few studies have been conducted on using graph learning to achieve multiple feature analysis. Due to the efficiency and practicality of graph learning, it is meaningful and effective to explore the value of graph attention learning on multi-variate wind power forecasting. In addition, among all wind power-related features, relationships can be complex and heterogeneous. For example, relationships between the features of wind turbines must be different from those among the weather features. Therefore, to fill this gap, a heterogeneous graph attention mechanism is designed in this paper.

As for the final predictor, there have been many effective models proposed by researchers around the world. As discussed in the references above, LSTM, TCN, and RNN have been widely accepted. Those methods have advantages in long-term relationship extraction, but some of them are too complicated [45]. Therefore, some scholars adopted Gated Recurrent Unit (GRU) as the predictor, which is more efficient and also has the ability to deal with long time series [46]. Liu [47,48] and his team designed explainable models for sensitivity analysis and efficient learning, which could discover the changing trend in the time series. But their performance on wind power analysis still remains to be validated. Recently, many models based on an encoder–decoder structure, also known as transformer-based models, have become popular in time series forecasts. For example, the double sampling transformer (DSformer) [49] and the multi-granularity spatiotemporal fusion transformer (MGSFformer) [50], which are variants of the transformer model, were proposed for multi-variate time series prediction, and proved to be satisfactory and accurate. Since the core part of those models, the attention mechanism, was embedded in the proposed feature analysis module, it would be redundant and too complex to reuse such a model. Taking the above analysis into account, this paper modifies the GRU model as the external forecasting framework and embeds the heterogeneous graph attention mechanism into GRU, which is efficient and effective.

In addition, except for the impact of multiple related features and the model structure, the changing data flow also has an effect on the prediction accuracy. The question of how to dynamically regulate the model in real time and make it more suitable for the current data also appeals to various researchers [51]. Sun et al. [52] utilized data augmentation to process acoustic features related to wind power so as to realize real-time modeling. However, the data augmentation method is not stable enough and is seldomly ineffective, resulting in large deviations in the results. Faruque et al. [53] proposed a real-time wind power forecasting model based on a modified LSTM model and a random searching optimization method, but, in fact, it could not dynamically and sufficiently adapt the model to the changing data flow. Mo [54] and his team proposed Powerformer to achieve a dynamic wind power generation forecast, which addressed the complexity of the issue and enhanced

feature extraction capability. However, the dynamic property of the model is embodied in the weather prediction and system operation, and the changing relationship among features resulting from the changing data flow has been not valued. Therefore, it is urgent to propose a method that is able to address and model the changeable relationship of the data features. In this paper, a flow-aware dynamic graph learning method is designed to solve this problem which directly employs the input data to adjust and generate the graph structure dynamically.

In order to realize dynamic wind power prediction and further enhance the accuracy and stability of the forecasting model, this study raises a new DHGAR model that has the ability to make rational use of the related features and dynamically regulate the model parameters according to data flow. **The detailed innovation and contributions of the proposed model can be summarized as follows:**

(1) In addition to weather features such as wind speed, wind direction, and temperature, the proposed model also incorporates generator-related conditions, including generator speed, rotor speed, and current. This ensures that the prediction model is supplied with comprehensive and sufficient information.

(2) Aiming at making full use of the multi-variate data, a heterogeneous graph attention mechanism is proposed. By forming the relationship among features into a directed graph, the model complexity is reduced, while the connections become more interpretable. The graph attention mechanism effectively learns the weighting of different features, uncovering their interrelationships. Additionally, a heterogeneous function is integrated into the graph attention mechanism, enabling the analysis of heterogeneous characteristics from diverse sources.

(3) A novel flow-aware dynamic graph learning method is proposed to link the generated graph with real-time data. Since the relationships among features can be influenced by changing data, it is essential to dynamically adjust the forecasting model in response to current data.

(4) Taking GRU as the external predictor and replacing its linear layer by the dynamic heterogeneous graph attention mechanism, the proposed wind power forecasting model achieves greater efficiency and effectiveness in practice and can effectively utilize the abundant information embedded in the features. Furthermore, experimental results show that the proposed model outperforms 10 state-of-the-art baselines across three real-world datasets, demonstrating its superiority in wind power forecasting.

The rest of this paper is organized as follows: Section 3 introduces the whole structure of the proposed dynamic heterogeneous graph attention recurrent network (DHGAR) model and explains the basic theory in detail. Section 4 conducts several meaningful experiments, including comparison experiments, ablation experiments, and hyperparameter experiments, and illustrates and discusses the results. Finally, Section 5 summarizes the current research work and analyzes some valuable aspects of future work.

3. Methodology

3.1. Preliminaries

This section gives an explanation for and problem reconstruction of wind power prediction. General wind power prediction focuses on the historical inherent features of the power series, giving regular prediction results:

Definition 1. *Wind Power Prediction. Taking the historical wind power series* $P = (P_{t-S+1}, P_{t-S+2}, \ldots, P_t)$ *of S points as inputs, the wind power prediction model attempts to provide the forecast values* $\hat{P} = (\hat{P}_{t+1}, \hat{P}_{t+2}, L, \hat{P}_{t+T_f})$ *of the next wind power in the next* T_f *time slice. The process can be denoted as below:*

$$(P_{t-S+1}, P_{t-S+2}, \cdots, P_t) \rightarrow (\hat{P}_{t+1}, \hat{P}_{t+2}, \cdots, \hat{P}_{t+T_f}). \quad (1)$$

However, wind power prediction is not only affected by historical sequence characteristics, but also by environmental factors such as wind speed, wind direction, and weather conditions. Thus, an input matrix containing features of external factors can be expressed as below:

$$X = \begin{bmatrix} P \\ X_1 \\ X_2 \\ \cdots \\ X_n \end{bmatrix} = \begin{bmatrix} P_{t-S+1}, P_{t-S+2}, \cdots, P_t \\ x_{1,t-S+1}, x_{1,t-S+2}, \cdots, x_{1,t} \\ x_{2,t-S+1}, x_{2,t-S+2}, \cdots, x_{2,t} \\ \cdots \\ x_{n,t-S+1}, x_{n,t-S+2}, \cdots, x_{n,t} \end{bmatrix} \quad (2)$$

where X represents the feature matrix and X_n is the vector of the *n*-th feature.

In this study, we develop a dynamic graph to explore the associative information of features. This approach enables the incorporation of various external environmental characteristics by treating different features as nodes and dynamically learning the connections between these nodes. Consequently, we define the power feature graph as follows:

Definition 2. *Power feature graph: The feature matrix is reformed as an undirected graph* G = (V, E), *in which* V *is the set of features, also called nodes, and* E *is the set of edges. Each node is a factor related to wind power, and the main node represents the wind power. Each edge represents the relationship between two nodes.*

Definition 3. *Heterogeneous graph: a heterogeneous graph can be expressed as* $G = (V, E, \phi, \psi)$, *where* ϕ *gives the type of a node and* ψ *gives the type of an edge.*

3.2. The Overall Structure of DHGAR

To address the significant impact of environmental factors and the complex feature associations in wind power forecasting, the DHGAR model is proposed in this study. It learns the relationships between features by constructing a feature graph, which effectively utilizes diverse feature information. Additionally, since the association patterns between each feature and others may vary with the data, the proposed model employs dynamic graphs and node injection to dynamically learn the adjacency matrix parameters, thereby enhancing the stability of wind power predictions. Furthermore, the model's heterogeneous graph attention mechanism effectively filters information, further improving the prediction accuracy of the external GRU network.

As shown in Figure 1, the proposed DHGAR model consists of three main components: a dynamic graph learning mechanism, heterogeneous graph attention, and a modified GRU network. First, the flow-aware dynamic graph learning mechanism improves the Data-Adaptive Graph Generation (DAGG) method by concatenating the current data stream into the node embedding matrix E and transforming the weights through a linear layer. This enhancement allows the generated dynamic graph to better reflect node relationships under real-time data. Then, the generated dynamic graph and adjacency matrix are fed into the heterogeneous graph attention module to update and iterate node weights. The heterogeneous graph attention mechanism achieves the learning of heterogeneous attention weights by embedding an edge-type matrix, which better captures the influence of different feature types on wind power. Finally, the GRU network is used to predict wind power,

with its multi-layer perceptual (MLP) hidden layer replaced by heterogeneous dynamic graph attention. The designed DHGAR model is able to effectively extract historical wind power characteristics and environmental factors and achieve stable and accurate wind power predictions.

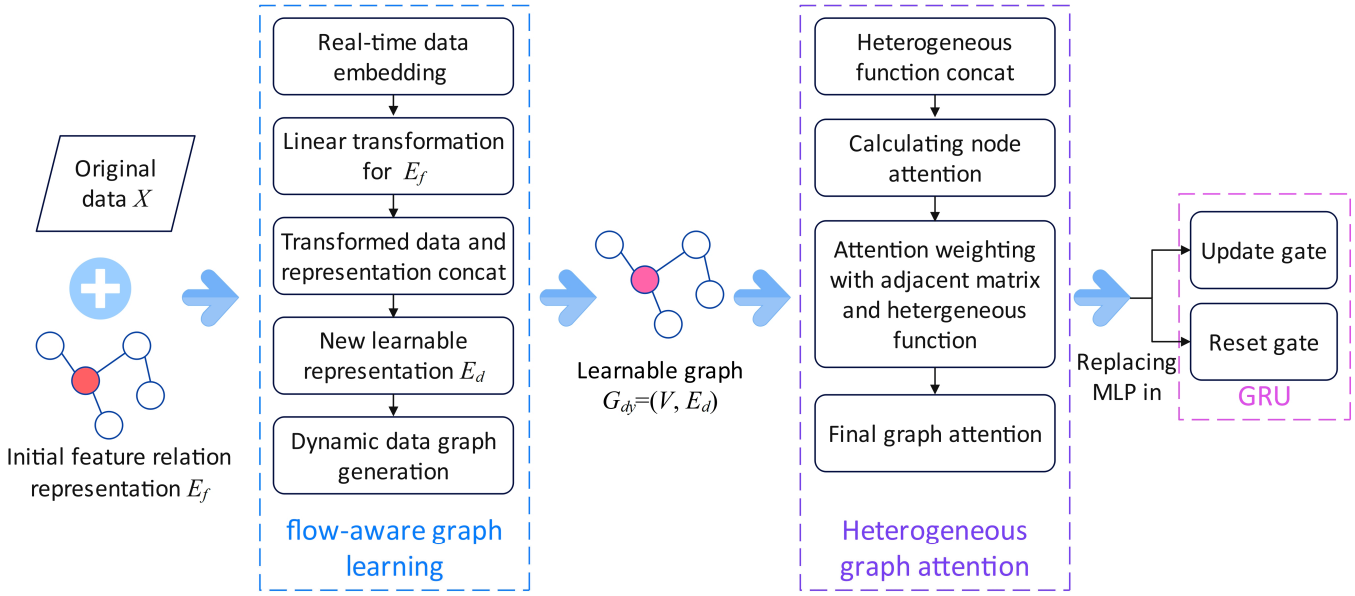


Figure 1. The structure of the proposed DHGAR model.

3.3. Flow-Aware Dynamic Graph Learning

As the relationships between features may change with varying input data, the Flow-aware dynamic graph learning method is proposed to establish a connection between the generated graph and real-time data. As shown in Figure 2, at first, an initial adjacency matrix is generated using Kaiming initialization to represent the relationships among features. Then, two linear transformation layers are designed to separately transform the real-time data input X and the adjacency matrix E_f. The transformed vectors are subsequently concatenated and subjected to an additional linear transformation, resulting in a new embedding, E_d. The transformation process can be expressed by the following formula:

$$E_d = W_1[W_2X+b_2; W_3X+b_3]+b_1 \quad (3)$$

where W_1, W_2, and W_3 are learnable weight matrices.

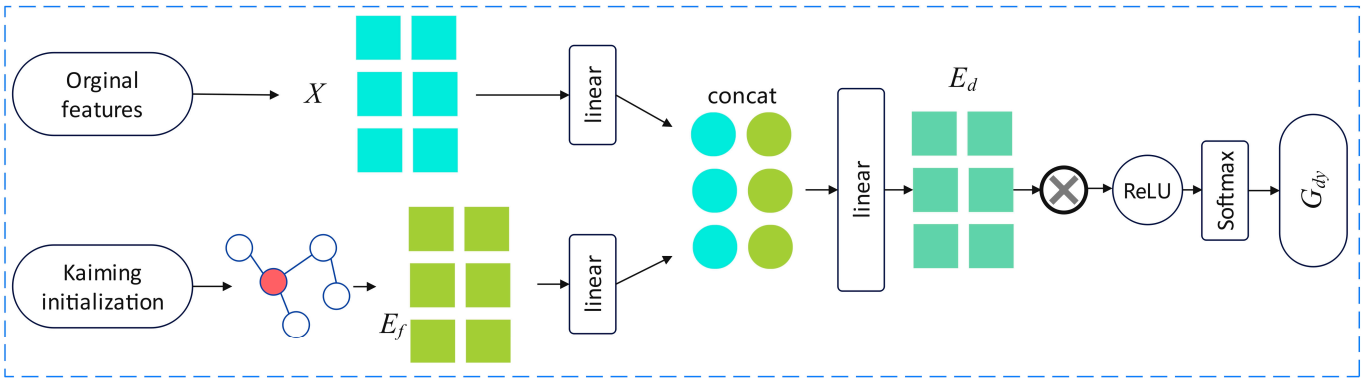


Figure 2. Structure of the flow-aware dynamic graph learning.

To generate dynamic graphs based on the current input data, a Data-Adaptive Graph Generation (DAGG) method is employed. This method was first introduced in the Diffusion Convolutional GRU (DCGRN) model and is capable of generating and refining graph structures according to the data patterns. The general DAGG method primarily consists of dot product operations, a ReLU layer, and a softmax function. It takes an adjacency matrix as the input and outputs the generated graph. In this study, since the input adjacency matrix E_d is learnable and incorporates a real-time data stream, the generated graph possesses the ability to dynamically reflect changes in the data. As shown in Figure 2, E_d is first subjected to a dot product operation and then processed by a ReLU layer and a softmax function. Ultimately, the dynamic graph is obtained. This process can be expressed as follows:

$$G_{dynamic} = \text{softmax}\left(\text{ReLU}\left(E_d E_d^T\right)\right) \tag{4}$$

It is worth noting that the dynamic graph adjusts in real time according to different input data, not only during training, but also during testing, generating varying graph structures. Therefore, this method can effectively capture changes in the impact of different environmental features on wind power.

3.4. Heterogeneous Graph Attention

With the issue of graph structure generation addressed, methods of effectively training the graph network parameters and obtaining appropriate feature weights is crucial for enhancing the accuracy of wind power predictions. The graph attention mechanism can autonomously learn the relationships between nodes, assigning more attention to strongly correlated features, and has demonstrated strong performance in graph learning. However, conventional graph attention methods often overlook node or edge types and may fail to accurately identify intrinsic data relationships when the graph structure contains heterogeneous nodes and edges. In wind power forecasting, the types of associations among different features may vary from each other, and these relationships may change with the data. Therefore, this study introduces heterogeneous functions to the graph attention structure, enabling heterogeneous learning of wind power feature graphs. The specific principles of the learnable heterogeneous graph attention model are as follows:

As shown in Figure 3a, a typical graph attention structure calculates the bidirectional weights between connected nodes i and j to obtain the corresponding graph attention coefficients α_{ij} and α_{ji}, which is expressed in Equation (5).

$$\alpha_{ij} = \frac{\exp\left(\text{LeakyReLU}\left(\vec{a}^T\left[W\vec{h}_i \middle\| W\vec{h}_j\right]\right)\right)}{\sum_{l \in N_i} \exp\left(\text{LeakyReLU}\left(\vec{a}^T\left[W\vec{h}_i \middle\| W\vec{h}_l\right]\right)\right)} \tag{5}$$

where LeakyReLU is an activation function, W is the weight matrix, $\vec{a}^T$ is an attention kernel that is a one dimension tensor with twice size of the output and used to generate the original attention rate from the concatenated tensor $\left[W\vec{h}_i \middle\| W\vec{h}_j\right]$, $\vec{h}$ represents a node feature vector, and N_i is the number or nodes directly connected to node i.

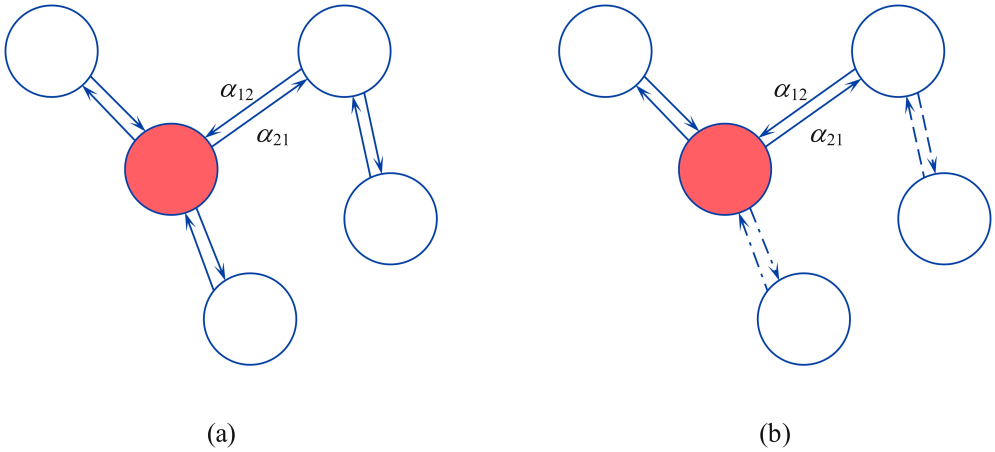


Figure 3. The schematic diagram of traditional graph attention and heterogeneous graph attention, where different line types stand for various edge types. (**a**) Traditional graph attention which line types are the same. (**b**) Heterogeneous graph attention with different line types.

In this computation process, the types of edges among all connected nodes are assumed to be the same, which is suitable for cases where node feature associations are similar. However, in wind power forecasting, features include various types such as weather conditions and wind turbine characteristics, leading to complex relationships among features. Therefore, by introducing an edge-type function, we modify the aforementioned formula to Equation (6):

$$\alpha_{ij} = \frac{\exp\left(\text{LeakyReLU}\left(\vec{a}^{T}\left[W\vec{h}_i \middle\| W\vec{h}_j \middle\| W_E E_{\psi(\langle i,j\rangle)}\right]\right)\right)}{\sum_{l \in N_i} \exp\left(\text{LeakyReLU}\left(\vec{a}^{T}\left[W\vec{h}_i \middle\| W\vec{h}_l \middle\| W_E E_{\psi(\langle i,l\rangle)}\right]\right)\right)} \tag{6}$$

where W_E is the weight matrix of various edge types, $\psi(\langle i,j\rangle)$ is the edge type function, and $E_{\psi(\langle i,j\rangle)}$ is the embedding of edge types.

As shown in Figure 3b, the types of edges between different nodes may vary. For example, during the data learning process, edge-type parameters between wind speed and wind direction typically differ significantly from those between engine speed and current. By designing the aforementioned heterogeneous graph attention layer, the graph of features can be generated dynamically, considering the real-time data property. Therefore, the learning capability of the model can be further enhanced, improving the robustness and accuracy of the proposed wind power prediction model.

3.5. Modified GRU Model

GRU is one of the most widely used models for time series forecasting, known for its efficiency, high accuracy, and simple structure. In this study, we enhance GRU by replacing its linear hidden layers with dynamic heterogeneous graph attention layers to develop the final prediction model. As illustrated in Figure 4, the linear hidden layers of the update gate and reset gate in the basic GRU structure are substituted with dynamic heterogeneous graph attention layers. Through the modification, the dynamic learning ability of GRU is strengthened and the efficiency and accuracy of the proposed model are enhanced. The underlying principles of the improved GRU model are as follows.

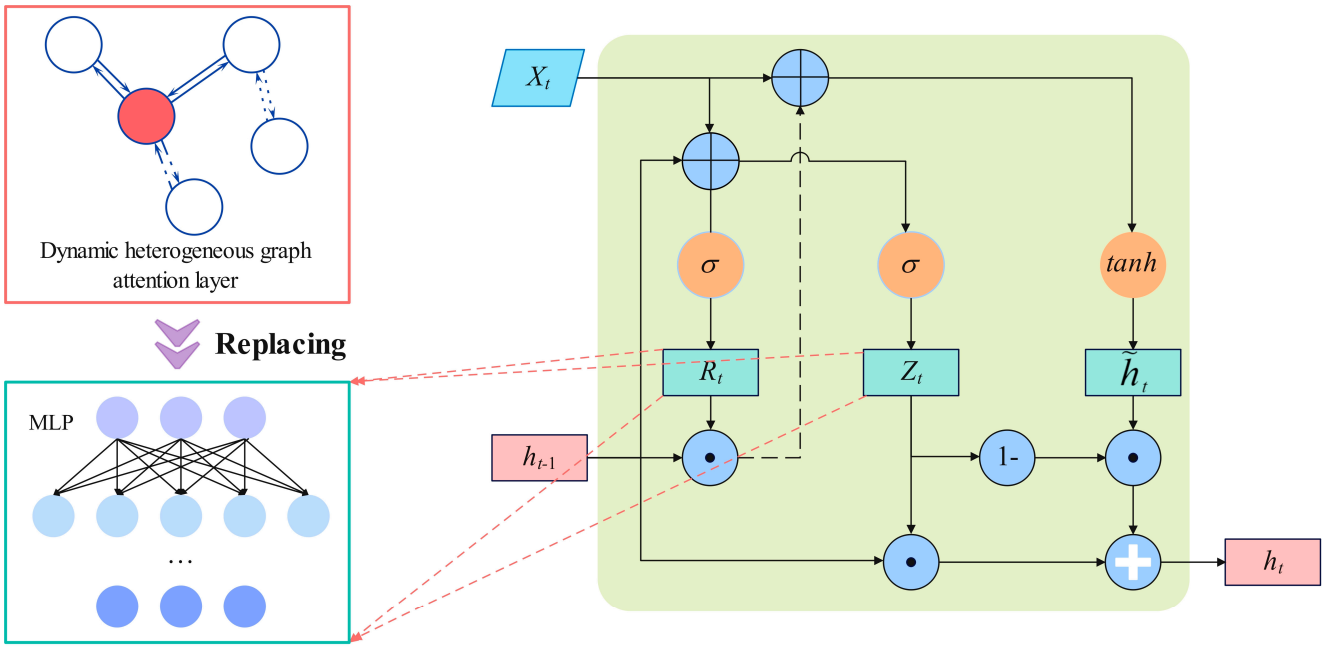


Figure 4. The structure of the modified GRU model.

As shown in Figure 4, the improved GRU model retains the basic structure of the standard GRU model, primarily consisting of the update gate and reset gate. The update gate Z controls the amount of historical information used, while the reset gate R determines the extent to which the historical state influences the current state. The update gate and reset gate in the classic GRU model are represented as follows:

$$R_t = \sigma(x_t W_{rx} + h_{t-1} W_{rh} + b_r) \tag{7}$$

$$Z_t = \sigma(x_t W_{zx} + h_{t-1} W_{zh} + b_z) \tag{8}$$

where R_t and Z_t represent the reset gate value and update gate value at time t, respectively, and σ is the sigmoid activation function, x_t is the input vector, W_{rx} and W_{rh} are weight matrixes of the reset gate, W_{zx} and W_{zh} are weight matrixes of the update gate, and b_r and b_z are biases.

In addition, the content in the brackets represents the MLP hidden layer of GRU. In this paper, the dynamic heterogeneous graph attention is used to replace this layer. The update gate and reset gate after substitution are described as follows:

$$\begin{aligned} \widetilde{HA} &= \text{HeterAttention}\left(G_{dynamic}\right) \\ &= \text{HeterAttention}\left(\text{softmax}\left(\text{ReLU}\left(E_d E_d^T\right)\right)\right) \end{aligned} \tag{9}$$

$$R_t = \sigma\left(\widetilde{HA}[x_t] + h_{t-1} W_{rh} + b_r\right) \tag{10}$$

$$Z_t = \sigma\left(\widetilde{HA}[x_t] + h_{t-1} W_{zh} + b_z\right) \tag{11}$$

where HeterAttention is the heterogeneous attention process, and $\widetilde{HA}$ is an intermediate variable representing the results given by the dynamic heterogeneous attention layer.

Given that the expression of Z_t and R_t are obtained, the candidate hidden state $\tilde{h}$ and the final hidden state h can be calculated as below:

$$\tilde{h}_t = \tanh\left(\widetilde{HA}[x_t] + (R_t \odot h_{t-1}) W_h + b_h\right) \tag{12}$$

$$h_t = (1-Z_t) \odot \widetilde{h}_t + Z_t \odot h_{t-1} \quad (13)$$

where $\odot$ means matrix dot product, that is, the multiplication of the corresponding position elements of two matrices.

3.6. Benefit Analysis

To clarify the function of the model in wind power prediction, it is necessary to analyze the benefits of each part of the designed model. Firstly, not only the common weather features, but also the generator-related conditions, are adopted to supply comprehensive and sufficient information. Then, the design of flow-aware dynamic graph learning effectively links the generated graph with real-time data, which can adjust the model learning ability in real time and influence the changing relationships among features. In addition, the heterogeneous graph attention mechanism can model the relationship among features as a directed graph, which is more interpretable and efficient, and its heterogeneous function enables the analysis of heterogeneous characteristics from diverse sources. Finally, taking GRU as the external predictor and replacing its linear layer by the dynamic heterogeneous graph attention mechanism, the proposed wind power forecasting model achieves greater efficiency and effectiveness in practice and can effectively utilize the abundant information embedded in the features. In summary, the proposed model provides a novel and valuable framework for wind power prediction, thoroughly accounting for environmental factors and demonstrating strong stability and accuracy.

4. Results and Discussion

4.1. Experiment Settings

4.1.1. Datasets

To fully and comprehensively evaluate the predictive performance and practical application effects of the proposed model, this paper conducts comparative experiments using three datasets collected from three different wind farms in Xinjiang, China. The basic statistics for these three datasets are shown in Table 1. Each dataset contains 10,000 samples with a time interval of 10 min between each sample. Other variables include wind speed, wind direction, angle against wind, ambient temperature, generator speed, rotor speed, voltage, current, blade angle, engine room temperature, oil temperature, winding temperature, shaft temperature, and bearing temperature. In this paper, these datasets are divided into training sets, validation sets, and test sets according to the ratio 7:1:2. All of the experiments are performed on the Intel(R) Xeon(R) Gold 5217 CPU @ 3.00 GHz 128 G RAM (Intel Corporation, Santa Clara, CA, USA) computing server with an RTX 3090 graphics card (NVIDIA Corporation, Santa Clara, CA, USA). The batch size is set to 32.

Table 1. The basic statistics of these data.

Wind Power Data	#1	#2	#3
Minimum	0.215	0.228	0.198
Mean	325.796	416.174	289.531
Maximum	1632.475	1589.344	1602.128
Standard deviation	315.188	299.148	345.179

4.1.2. Baselines

To thoroughly validate the predictive performance of the proposed model, this study selects baselines from five perspectives: (1) forecasting models based on spatiotemporal graph neural networks, like dynamic graph convolutional recurrent networks (DGCRNs) and adaptive graph convolutional recurrent networks (AGCRNs); (2) forecasting models

based on transformers, such as inverted transformers (iTransformer) and cross-dimension transformers (Crossformer); (3) forecasting models based on MLPs, like time series mixer (TSmixer) and Rlinear; (4) hybrid forecasting models based on ensemble learning, such as Mi's model and Yu's model; and (5) classic forecasting models, like ARIMA and support vector machine (SVM). The introductions of all baselines are as follows:

- DGCRN [55]: it introduces dynamic graph convolutional networks to improve the performance of recurrent neural networks.
- AGCRN [56]: it introduces adaptive graph learning to improve the forecasting power of GRUs.
- iTransformer [57]: it uses multi-head attention to mine the correlations between different variables, and then uses the full connection layer to mine the context of the time series.
- Crossformer [58]: it uses transformer encoders to mine the context information of the time series, and then uses routing attention to mine the correlations between different variables.
- TSmixer [59]: it uses the MLP framework to mine both variable correlations and temporal information
- Rlinear [60]: it uses Revin and channel independence to improve the forecasting performance of MLP models.
- Mi's model [61]: it combines deep reinforcement learning and multi-objective optimization algorithms to propose dynamic ensemble learning strategies.
- Yu's model [62]: it uses reinforcement theory to integrate two different spatiotemporal graph neural networks to improve the forecasting performance of the model.
- SVM [63]: it is a classical kernel-based machine learning model.
- ARIMA [64]: it is a classical statistical model based on autoregression.

4.1.3. Evaluation Metrics

To evaluate the model's performance more objectively and rationally, this paper adopts three common indicators for time series forecasting, as well as a special indicator in the field of PV power prediction. The three common indicators include Mean Absolute Error (MAE) [65], Root Mean Square Error (RMSE) [62], and R Squared index (CORR) [66], which are represented in Equation (8) below:

$$\begin{cases} \mathrm{MAE} = \left(\sum_{n=1}^{N} \left| P_n - \hat{P}_n \right| \right) / N \\ \mathrm{RMSE} = \sqrt{\left(\sum_{n=1}^{N} \left| P_n - \hat{P}_n \right|^2 \right) / N} \\ \mathrm{CORR} = 1 - \left(\sum_{n=1}^{N} \left| P_n - \hat{P}_n \right|^2 \right) / \left(\sum_{n=1}^{N} \left| P^{-} - P_n \right|^2 \right) \end{cases} \tag{14}$$

where N represents the length of the wind power series, P_n is the wind power at the n-th point, $\hat{P}_n$ is the predicted wind power at the n-th point, and $\bar{P}$ is the average value of the wind power series. To express the advancement of the proposed model more intuitively, the promoting percentages of MAE and RMSE are also calculated as follows:

$$\begin{cases} p_{\mathrm{MAE}} = (\mathrm{MAE}_1 - \mathrm{MAE}_2) / \mathrm{MAE}_1 \times 100\% \\ p_{\mathrm{RMSE}} = (\mathrm{RMSE}_1 - \mathrm{RMSE}_2) / \mathrm{RMSE}_1 \times 100\% \end{cases} \tag{15}$$

where p_{MAE} and p_{RMSE} represent the promoting percentages of MAE and RMSE, respectively.

4.2. Comparison Experiments with Advanced Baseline Models

Table 2 shows the experimental results of the proposed model and all baselines on three datasets. Table 3 shows the training time of the proposed model and several baselines. Based on the experimental results, it can be seen that DHGAR can achieve the best experimental results on all the datasets. In addition, to directly demonstrate the forecasting performance of the proposed model, this section visualizes the forecasting results of the proposed model and several baselines. Figures 5–7 show the visualizations on all datasets. Based on Tables 2 and 3 and Figures 5–7, the following conclusions can be drawn:

(1) Compared to other deep learning-based models, classical models have the largest forecasting errors. The main reason is that classical models rely on stationarity assumptions, which limits their effectiveness in modeling non-stationary wind power sequences. In addition, it is difficult for these models to effectively model the implicit correlations between different variables.

(2) Compared with other baselines, the spatiotemporal graph neural network can obtain more accurate forecasting results. Specifically, the spatiotemporal graph neural network uses adaptive graph learning to effectively mine the correlations between different variables. In this way, the model can effectively use other variables to help improve the accuracy of wind power prediction.

(3) DHGAR achieves the most satisfactory forecasting accuracy on all datasets. The experimental results fully prove the application value of the proposed model in the field of wind power prediction. First, different from the traditional single-variable wind power time series prediction model, we introduce other variables to improve the prediction effect of the model. Then, we introduce the dynamic graph structure to construct the correlation between other variables and wind power, aiming to dynamically select other features to assist wind power prediction. Next, we propose heterogeneous graph attention to further explore the implicit correlations between heterogeneous variables. Finally, we replace all fully connected layers in the GRU with the proposed component to realize the mining of the wind power time series context. Through the above method, DHGAR can dynamically and deeply explore the spatiotemporal correlation between other variables and wind power, which ensures the DHGAR's effectiveness.

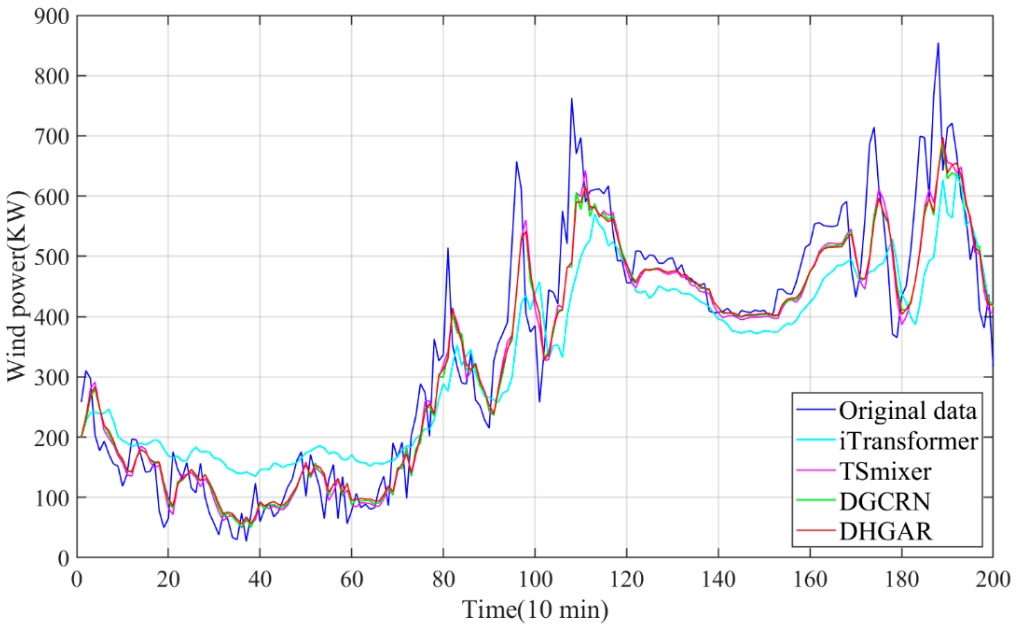


Figure 5. Experimental results of different models for dataset #1.

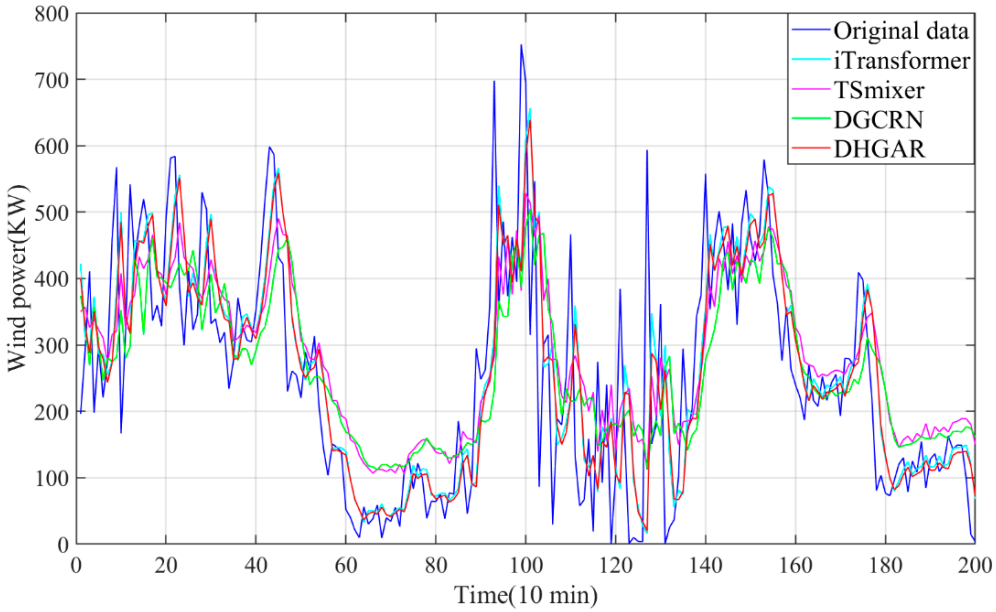


Figure 6. Experimental results of different models for dataset #2.

Table 2. The time series statistical analysis indexes of all contrast predictors.

Dataset	Prediction Models	MAE (KW)	RMSE (KW)	CORR
#1	**DHGAR**	**76.135**	**114.295**	**0.915**
	DGCRN	80.146	122.548	0.894
	AGCRN	82.719	125.675	0.885
	iTransformer	86.429	131.178	0.842
	Crossformer	87.995	133.784	0.839
	TSmixer	83.443	127.342	0.878
	RLinear	84.137	128.795	0.867
	Mi's model	83.715	127.159	0.876
	Yu's model	85.632	130.544	0.861
	SVM	95.314	145.199	0.787
	ARIMA	94.178	144.328	0.795
#2	**DHGAR**	**67.424**	**93.418**	**0.924**
	DGCRN	70.295	99.145	0.902
	AGCRN	71.342	101.697	0.897
	iTransformer	75.638	108.245	0.823
	Crossformer	77.199	110.334	0.802
	TSmixer	75.117	109.121	0.821
	RLinear	72.612	103.447	0.874
	Mi's model	74.253	107.889	0.837
	Yu's model	73.942	105.244	0.855
	SVM	82.166	121.749	0.765
	ARIMA	83.194	123.586	0.743
#3	**DHGAR**	**84.216**	**121.547**	**0.901**
	DGCRN	86.147	125.334	0.874
	AGCRN	88.322	131.556	0.862
	iTransformer	95.745	144.624	0.786
	Crossformer	96.947	146.154	0.774
	TSmixer	92.547	137.014	0.807
	RLinear	93.326	139.984	0.799
	Mi's model	91.177	135.647	0.819
	Yu's model	90.894	135.024	0.836
	SVM	102.322	155.789	0.714
	ARIMA	101.479	154.324	0.709

Table 3. The training time of the proposed model and several baselines.

Model	Training Time (s)
DHGAR	**157.24**
DGCRN	263.14
AGCRN	134.25
iTransformer	100.34
Crossformer	189.46

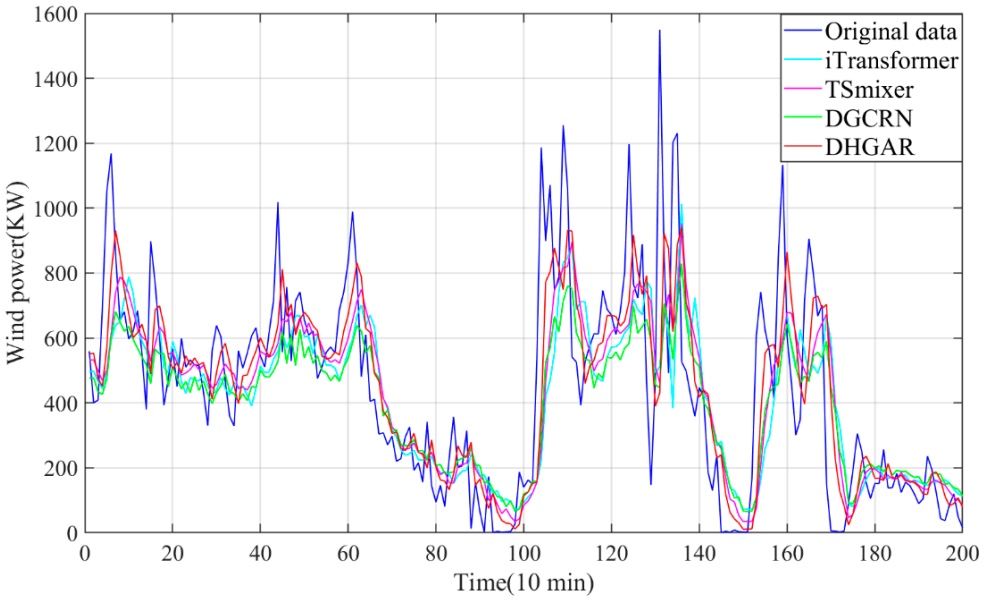


Figure 7. Experimental results of different models for dataset #3.

4.3. Ablation Experiments

Considering that DHGAR has two important components, the dynamic graph and heterogeneous association function, the ablation experiments are conducted from the following perspectives:

(1) W/O DG: we removed the dynamic graph structure and replaced it with a static graph structure based on Pearson coefficients.
(2) W/O HA: We removed the heterogeneous association function. In this case, the model only uses dynamic graphs to model variable correlations.
(3) W/O GA: We removed all the graph structures. In this case, the model uses only GRU to achieve wind power forecasting.

Table 4 shows the results of the ablation experiment. Based on the experimental results, the following conclusions can be drawn:

(1) The dynamic graph has a great influence on the prediction results of DHGAR. If only a static graph is used to establish the correlation between wind power and other variables, the model ignores the difference in the correlation between different variables at different time points, which limits the prediction effect.
(2) The heterogeneous association function also has a great influence on the prediction results of DHGAR. Specifically, if the heterogeneous association function is deleted, the model can only achieve multi-variate information aggregation simply using weighted fusion. In this case, DHGAR cannot fully capture deeper correlations between heterogeneous variables, which limits its effectiveness.
(3) When the graph structure is removed, the model is equivalent to directly modeling the original data using the GRU. In this case, the prediction accuracy of the model

decreases significantly due to the inability to accurately establish the correlation between the different variables.

Table 4. The time series statistical analysis indexes of the ablation experiments.

Dataset	Models	MAE (KW)	RMSE (KW)	P_{MAE}	P_{RMSE}
#1	**DHGAR**	**76.135**	**114.295**	-	-
	W/O DG	78.143	119.257	2.570	4.161
	W/O HA	78.225	118.933	2.672	3.900
	W/O GA	92.154	140.243	17.383	18.502
#2	**DHGAR**	**67.424**	**93.418**	-	-
	W/O DG	69.314	98.145	2.727	4.816
	W/O HA	69.759	99.023	3.347	5.660
	W/O GA	79.433	115.423	15.118	19.065
#3	**DHGAR**	**84.216**	**121.547**	-	-
	W/O DG	86.158	125.741	2.254	3.335
	W/O HA	86.345	125.644	2.466	3.261
	W/O GA	98.324	150.29	14.348	19.125

4.4. Hyperparameter Experiments

The main hyperparameters that affect the predictive performance of the model include historical length, embedding size, and node embedding size. In this paper, the hyperparameters are adjusted mainly by the raster search method [67]. The setting range of hyperparameters mainly refers to the existing literature [68]. In addition, to fully evaluate the influence of hyperparameters on experimental results, this paper carried out hyperparameter analysis experiments. Table 5 shows the values of the model hyperparameters. Figures 8–10 shows the MAE values of hyperparameter experiments. Based on the experimental results, the following conclusions can be drawn:

(1) The node embedding size has the least influence on the prediction results of DHGAR. The main reason for this is that the dynamic graph learning component does not require a particularly large number of parameters to obtain a graph structure that accurately models variable correlations.

(2) The historical length has the greatest influence on the prediction accuracy of DHGAR. On one hand, a too-large historical length can lead to model overfitting. On the other hand, a too-small historical length may not provide enough information. Therefore, finding a proper balance in historical length is crucial.

Table 5. The main hyperparameters of DHGAR.

Config	Values
optimizer	Adam
learning rate	0.002
dropout	0.3
historical length	5
embedding size	128
node embedding size	64
learning rate schedule	MultiStepLR
milestone	1, 25, 50, 75
gamme	0.5
epoch	100
batch size	32

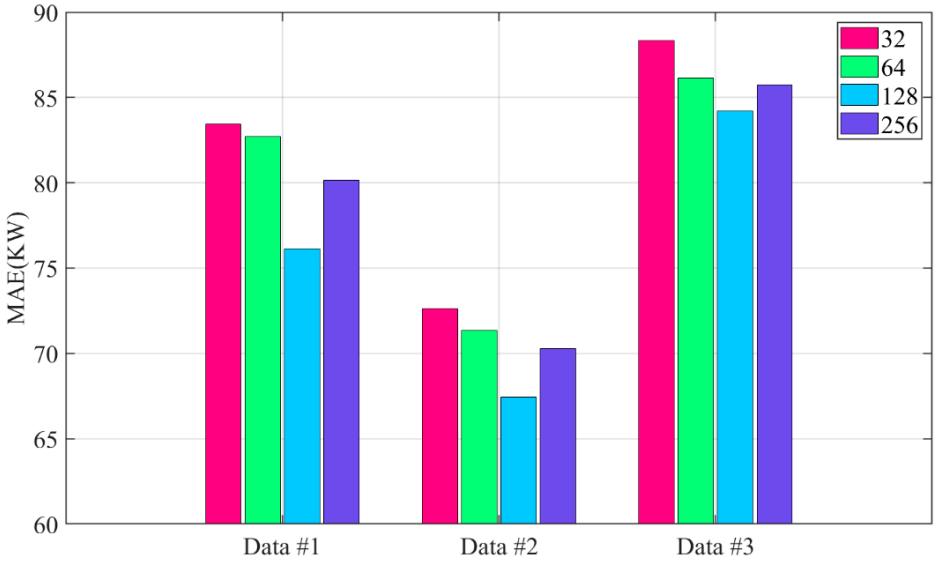


Figure 8. MAE values of different embedding sizes.

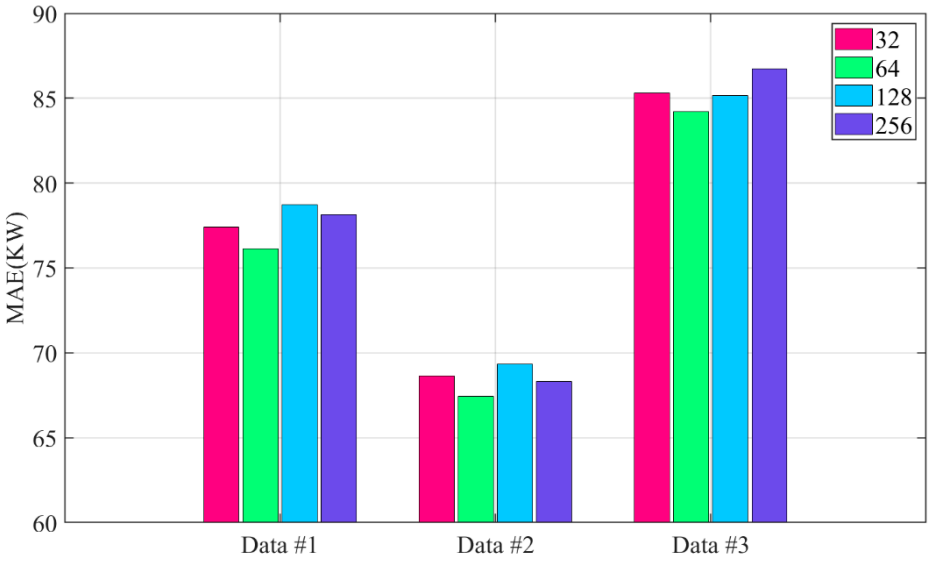


Figure 9. MAE values of different node embedding sizes.

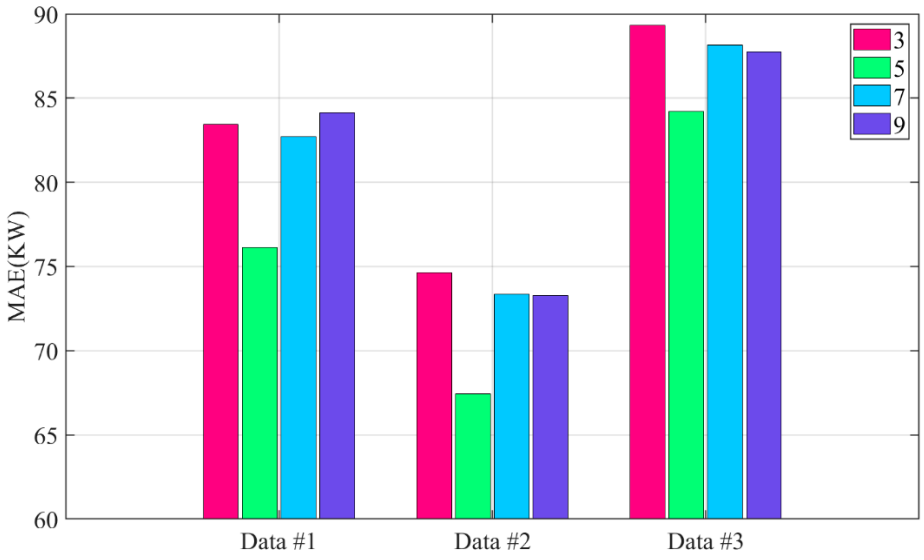


Figure 10. MAE values of different historical lengths.

5. Conclusions and Future Work

Accurate wind power prediction is of great significance for energy management and scheduling in wind farms. To achieve high-precision wind power forecasting, this paper proposes improvements from the perspectives of the dynamic and heterogeneous relationships between wind power and its covariates, and proposes the dynamic heterogeneous graph attention recurrent network, including the following enhancements:

(1) In addition to weather characteristics such as wind speed, wind direction, and temperature, the proposed model incorporates conditions related to the generator, including generator speed, rotor speed, and current. This provides the prediction model with comprehensive and sufficient information.
(2) To fully utilize multi-variate data, a heterogeneous graph attention mechanism is proposed. This method not only effectively learns the interrelationships among different features, but also introduces heterogeneous functions to capture diverse features from different sources.
(3) A novel flow-aware dynamic graph learning method is proposed, which connects the generated graph with real-time data.
(4) The proposed model adopts GRU as an external predictor and replaces its linear layer with a dynamic heterogeneous graph attention mechanism. This allows the model to effectively utilize the rich information embedded within the features.
(5) Experimental results demonstrate that the proposed model can effectively exploit the heterogeneity and dynamics of the wind power dataset, leading to significantly improved performance compared to existing SOTA baselines.

Although the proposed model has achieved promising prediction results, it still has some limitations. In the future, we will further improve it from the following perspectives:

(1) In addition to heterogeneous variables, the wind power data from different monitoring stations also influence each other. In the future, we will further consider establishing a spatiotemporal wind power prediction model.
(2) Due to the combination of graph neural networks and recurrent neural networks, the computational complexity of DHGAR is $O(N^2L^2)$. In the future, we will focus on optimizing the model's efficiency.

Author Contributions: Conceptualization, C.Y. and M.X.; methodology, M.X.; software, R.Z. and M.X.; validation, R.Z. and C.Y.; formal analysis, M.X.; investigation, M.X.; resources, R.Z.; data curation, R.Z.; writing—original draft preparation, M.X.; writing—review and editing, R.Z.; visualization, R.Z. and X.M.; supervision, X.M.; project administration, X.M.; funding acquisition, X.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset used in this study has been released on GitHub at https://github.com/ChengqingYu/MTS_dataset, accessed on 28 October 2024.

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

References

1. Karijadi, I.; Chou, S.-Y.; Dewabharata, A. Wind power forecasting based on hybrid CEEMDAN-EWT deep learning method. *Renew. Energy* **2023**, *218*, 119357. [CrossRef]
2. Yu, C.; Qiao, J.; Chen, C.; Yu, C.; Mi, X. TFEformer: A new temporal frequency ensemble transformer for day-ahead photovoltaic power prediction. *J. Clean. Prod.* **2024**, *448*, 141690. [CrossRef]

3. Zehtabiyan-Rezaie, N.; Justsen, J.P.; Abkar, M. Wind-farm power prediction using a turbulence-optimized Gaussian wake model. *Wind Energy Eng. Res.* **2024**, *2*, 100007. [CrossRef]
4. Kim, D.; Ryu, G.; Moon, C.; Kim, B. Accuracy of a short-term wind power forecasting model based on deep learning using LiDAR-SCADA integration: A case study of the 400-MW Anholt offshore wind farm. *Appl. Energy* **2024**, *373*, 123882. [CrossRef]
5. Li, J.; Meng, F.; Zhang, Z.; Zhang, Y. Prediction of wind power ramp events via a self-attention based deep learning approach. *Energy Rep.* **2024**, *12*, 1488–1502. [CrossRef]
6. Chen, Y.; Lin, C.; Zhang, Y.; Liu, J.; Yu, D. Proactive failure warning for wind power forecast models based on volatility indicators analysis. *Energy* **2024**, *305*, 132310. [CrossRef]
7. Lee, H.-C.; Liu, H.-Y.; Teng, S.-Y. Distributed energy strategy using renewable energy transformation in Kinmen Island: Virtual power plants that take the military camps as the mainstay. *Energy Strategy Rev.* **2022**, *44*, 100993. [CrossRef]
8. Camal, S.; Girard, R.; Fortin, M.; Touron, A.; Dubus, L. A conditional and regularized approach for large-scale spatiotemporal wind power forecasting. *Sustain. Energy Technol. Assess.* **2024**, *65*, 103743. [CrossRef]
9. Ge, Y.; Zhang, C.; Wang, Y.; Chen, J.; Wang, Z.; Shahzad Nazir, M.; Peng, T. A novel hybrid model based on multiple influencing factors and temporal convolutional network coupling ReOSELM for wind power prediction. *Energy Convers. Manag.* **2024**, *313*, 118632. [CrossRef]
10. Mahata, S.; Harsh, P.; Shekher, V. Comparative study of time-series forecasting models for wind power generation in Gujarat, India. *e-Prime—Adv. Electr. Eng. Electron. Energy* **2024**, *8*, 100511. [CrossRef]
11. de Azevedo Takara, L.; Teixeira, A.C.; Yazdanpanah, H.; Mariani, V.C.; dos Santos Coelho, L. Optimizing multi-step wind power forecasting: Integrating advanced deep neural networks with stacking-based probabilistic learning. *Appl. Energy* **2024**, *369*, 123487. [CrossRef]
12. Zhao, B.; He, X.; Ran, S.; Zhang, Y.; Cheng, C. Spatial correlation learning based on graph neural network for medium-term wind power forecasting. *Energy* **2024**, *296*, 131164. [CrossRef]
13. Dai, G.; Tang, J.; Luo, W. Short-term traffic flow prediction: An ensemble machine learning approach. *Alex. Eng. J.* **2023**, *74*, 467–480. [CrossRef]
14. Anthony, K.; Arunachalam, V. Application of cascaded neural network for prediction of voltage stability margin in a solar and wind integrated power system. *Eng. Appl. Artif. Intell.* **2024**, *138*, 109368. [CrossRef]
15. Malakouti, S.M. Improving the prediction of wind speed and power production of SCADA system with ensemble method and 10-fold cross-validation. *Case Stud. Chem. Environ. Eng.* **2023**, *8*, 100351. [CrossRef]
16. Sarangi, S.; Dash, P.K.; Bisoi, R. Short-term prediction of wind power using an improved kernel based optimized deep belief network. *Energy Convers. Manag.* **2024**, *316*, 118821. [CrossRef]
17. Wang, Y.; Zhao, K.; Hao, Y.; Yao, Y. Short-term wind power prediction using a novel model based on butterfly optimization algorithm-variational mode decomposition-long short-term memory. *Appl. Energy* **2024**, *366*, 123313. [CrossRef]
18. Chen, Y.; Yu, S.; Lim, C.P.; Shi, P. A novel interval estimation framework for wind power forecasting using multi-objective gradient descent optimization. *Sustain. Energy Grids Netw.* **2024**, *38*, 101363. [CrossRef]
19. Ban, G.; Chen, Y.; Xiong, Z.; Zhuo, Y.; Huang, K. The univariate model for long-term wind speed forecasting based on wavelet soft threshold denoising and improved Autoformer. *Energy* **2024**, *290*, 130225. [CrossRef]
20. Akbal, Y.; Ünlü, K.D. A univariate time series methodology based on sequence-to-sequence learning for short to midterm wind power production. *Renew. Energy* **2022**, *200*, 832–844. [CrossRef]
21. Ahn, E.; Hur, J. A short-term forecasting of wind power outputs using the enhanced wavelet transform and arimax techniques. *Renew. Energy* **2023**, *212*, 394–402. [CrossRef]
22. Rahman, N.H.A.; Hussin, M.Z.; Sulaiman, S.I.; Hairuddin, M.A.; Saat, E.H.M. Univariate and multivariate short-term solar power forecasting of 25MWac Pasir Gudang utility-scale photovoltaic system using LSTM approach. *Energy Rep.* **2023**, *9*, 387–393. [CrossRef]
23. Arslan Tuncar, E.; Sağlam, Ş.; Oral, B. A review of short-term wind power generation forecasting methods in recent technological trends. *Energy Rep.* **2024**, *12*, 197–209. [CrossRef]
24. Cheng, R.; Yang, D.; Liu, D.; Zhang, G. A Reconstruction-based Secondary Decomposition-Ensemble Framework for Wind Power Forecasting. *Energy* **2024**, *308*, 132895. [CrossRef]
25. Chen, B.; Li, Z.; Li, S.; Zhao, Q.; Liu, X. A Wind Power Prediction Framework for Distributed Power Grids. *Energy Eng.* **2024**, *121*, 1291–1307. [CrossRef]
26. Sulaiman, M.H.; Mustaffa, Z. Enhancing wind power forecasting accuracy with hybrid deep learning and teaching-learning-based optimization. *Clean. Energy Syst.* **2024**, *9*, 100139. [CrossRef]
27. Xin, Z.; Liu, X.; Zhang, H.; Wang, Q.; An, Z.; Liu, H. An enhanced feature extraction based long short-term memory neural network for wind power forecasting via considering the missing data reconstruction. *Energy Rep.* **2024**, *11*, 97–114. [CrossRef]
28. Yang, M.; Che, R.; Yu, X.; Su, X. Dual NWP wind speed correction based on trend fusion and fluctuation clustering and its application in short-term wind power prediction. *Energy* **2024**, *302*, 131802. [CrossRef]

29. Wang, S.; Li, B.; Li, G.; Li, B.; Li, H.; Jiao, K.; Wang, C. A comprehensive review on the development of data-driven methods for wind power prediction and AGC performance evaluation in wind–thermal bundled power systems. *Energy AI* **2024**, *16*, 100336. [CrossRef]
30. Zhao, H.; Xu, P.; Gao, T.; Zhang, J.J.; Xu, J.; Gao, D.W. CPTCFS: CausalPatchTST incorporated causal feature selection model for short-term wind power forecasting of newly built wind farms. *Int. J. Electr. Power Energy Syst.* **2024**, *160*, 110059. [CrossRef]
31. Cao, W.; Wang, G.; Liang, X.; Hu, Z. A STAM-LSTM model for wind power prediction with feature selection. *Energy* **2024**, *296*, 131030. [CrossRef]
32. Harrou, F.; Dairi, A.; Dorbane, A.; Sun, Y. Enhancing wind power prediction with self-attentive variational autoencoders: A comparative study. *Results Eng.* **2024**, *23*, 102504. [CrossRef]
33. Samal, K.K.R.; Panda, A.K.; Babu, K.S.; Das, S.K. Multi-output TCN autoencoder for long-term pollution forecasting for multiple sites. *Urban Clim.* **2021**, *39*, 100943. [CrossRef]
34. Yang, T.; Yang, Z.; Li, F.; Wang, H. A short-term wind power forecasting method based on multivariate signal decomposition and variable selection. *Appl. Energy* **2024**, *360*, 122759. [CrossRef]
35. Peng, Q.; Liu, Y.; Jin, Y.; Yang, X.G.; Wang, R.; Liu, K. Coating Feature Analysis and Capacity Prediction for Digitalization of Battery Manufacturing: An Interpretable AI Solution. *IEEE Trans. Syst. Man Cybern. Syst.* **2025**, *55*, 284–294. [CrossRef]
36. Tan, L.; Chen, Y.; Xia, J.; Wang, Y. Research on the short-term wind power prediction with dual branch multi-source fusion strategy. *Energy* **2024**, *291*, 130402. [CrossRef]
37. Elmousalami, H.; Elmesalami, H.H.; Maxi, M.; Farid, A.A.K.M.; Elshaboury, N. A comprehensive evaluation of machine learning and deep learning algorithms for wind speed and power prediction. *Decis. Anal. J.* **2024**, *13*, 100527. [CrossRef]
38. Ozbak, M.; Ghazizadeh-Ahsaee, M.; Ahrari, M.; Jahantigh, M.; Mirshekar, S.; Mirmozaffari, M.; Aranizadeh, A. Improving power output wind turbine in micro-grids assisted virtual wind speed prediction. *Sustain. Oper. Comput.* **2024**, *5*, 119–130. [CrossRef]
39. Qu, Z.; Hou, X.; Li, J.; Hu, W. Short-term wind farm cluster power prediction based on dual feature extraction and quadratic decomposition aggregation. *Energy* **2024**, *290*, 130155. [CrossRef]
40. Yang, M.; Han, C.; Zhang, W.; Wang, B. A short-term power prediction method for wind farm cluster based on the fusion of multi-source spatiotemporal feature information. *Energy* **2024**, *294*, 130770. [CrossRef]
41. Arrieta-Prieto, M.; Schell, K.R. Spatially transferable machine learning wind power prediction models: V−logit random forests. *Renew. Energy* **2024**, *223*, 120066. [CrossRef]
42. Qiu, H.; Shi, K.; Wang, R.; Zhang, L.; Liu, X.; Cheng, X. A novel temporal–spatial graph neural network for wind power forecasting considering blockage effects. *Renew. Energy* **2024**, *227*, 120499. [CrossRef]
43. Lai, Z.; Wu, T.; Fei, X.; Ling, Q. BERT4ST:: Fine-tuning pre-trained large language model for wind power forecasting. *Energy Convers. Manag.* **2024**, *307*, 118331. [CrossRef]
44. Zhao, Y.; Liao, H.; Pan, S.; Zhao, Y. Interpretable multi-graph convolution network integrating spatio-temporal attention and dynamic combination for wind power forecasting. *Expert Syst. Appl.* **2024**, *255*, 124766. [CrossRef]
45. Yu, C.; Wang, F.; Shao, Z.; Qian, T.; Zhang, Z.; Wei, W.; Xu, Y. GinAR: An End-To-End Multivariate Time Series Forecasting Model Suitable for Variable Missing. In Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Barcelona, Spain, 25–29 August 2024; pp. 3989–4000.
46. Ayene, S.M.; Yibre, A.M. Wind power prediction based on deep learning models: The case of Adama wind farm. *Heliyon* **2024**, *10*, e39579. [CrossRef]
47. Liu, K.; Liu, Y.; Zhao, S.; Li, X.; Peng, Q. An Ultrasonic Wave-Based Method for Efficient State-of-Health Estimation of Li-Ion Batteries. *IEEE Trans. Ind. Electron.* **2024**, *99*, 1–11. [CrossRef]
48. Liu, K.; Peng, Q.; Liu, Y.; Cui, N.; Zhang, C. Explainable Neural Network for Sensitivity Analysis of Lithium-Ion Battery Smart Production. *IEEE/CAA J. Autom. Sin.* **2024**, *11*, 1944–1953. [CrossRef]
49. Yu, C.; Wang, F.; Shao, Z.; Sun, T.; Wu, L.; Xu, Y. DSformer: A Double Sampling Transformer for Multivariate Time Series Long-term Prediction. In Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, Birmingham, UK, 21–25 October 2023; pp. 3062–3072.
50. Yu, C.; Wang, F.; Wang, Y.; Shao, Z.; Sun, T.; Yao, D.; Xu, Y. MGSFformer: A Multi-Granularity Spatiotemporal Fusion Transformer for air quality prediction. *Inf. Fusion* **2025**, *113*, 102607. [CrossRef]
51. Qiu, X.; Wu, X.; Lin, Y.; Guo, C.; Hu, J.; Yang, B. Duet: Dual clustering enhanced multivariate time series forecasting. *arXiv* **2024**, arXiv:2412.10859.
52. Sun, B.; Su, M.; He, J. Wind power prediction through acoustic data-driven online modeling and active wake control. *Energy Convers. Manag.* **2024**, *319*, 118920. [CrossRef]
53. Faruque, M.O.; Hossain, M.A.; Islam, M.R.; Alam, S.M.M.; Karmaker, A.K. Very short-term wind power forecasting for real-time operation using hybrid deep learning model with optimization algorithm. *Clean. Energy Syst.* **2024**, *9*, 100129. [CrossRef]
54. Mo, S.; Wang, H.; Li, B.; Xue, Z.; Fan, S.; Liu, X. Powerformer: A temporal-based transformer model for wind power forecasting. *Energy Rep.* **2024**, *11*, 736–744. [CrossRef]

55. Li, F.; Feng, J.; Yan, H.; Jin, G.; Yang, F.; Sun, F.; Jin, D.; Li, Y. Dynamic graph convolutional recurrent network for traffic prediction: Benchmark and solution. *ACM Trans. Knowl. Discov. Data* **2023**, *17*, 1–21. [CrossRef]
56. Bai, L.; Yao, L.; Li, C.; Wang, X.; Wang, C. Adaptive graph convolutional recurrent network for traffic forecasting. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 17804–17815.
57. Liu, Y.; Hu, T.; Zhang, H.; Wu, H.; Wang, S.; Ma, L.; Long, M. iTransformer: Inverted Transformers Are Effective for Time Series Forecasting. In Proceedings of the Twelfth International Conference on Learning Representations, Vienna, Austria, 7–11 May 2024.
58. Zhang, Y.; Yan, J. Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. In Proceedings of the Eleventh International Conference on Learning Representations, Kigali, Rwanda, 1–5 May 2023.
59. Ekambaram, V.; Jati, A.; Nguyen, N.; Sinthong, P.; Kalagnanam, J. Tsmixer: Lightweight mlp-mixer model for multivariate time series forecasting. In Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Long Beach, CA, USA, 6–10 August 2023; pp. 459–469.
60. Li, Z.; Qi, S.; Li, Y.; Xu, Z. Revisiting long-term time series forecasting: An investigation on linear mapping. *arXiv* **2023**, arXiv:2305.10721.
61. Mi, X.; Yu, C.; Liu, X.; Yan, G.; Yu, F.; Shang, P. A dynamic ensemble deep deterministic policy gradient recursive network for spatiotemporal traffic speed forecasting in an urban road network. *Digit. Signal Process.* **2022**, *129*, 103643. [CrossRef]
62. Chengqing, Y.; Guangxi, Y.; Chengming, Y.; Yu, Z.; Xiwei, M. A multi-factor driven spatiotemporal wind power prediction model based on ensemble deep graph attention reinforcement learning networks. *Energy* **2023**, *263*, 126034. [CrossRef]
63. Singh, S.N.; Mohapatra, A. Data driven day-ahead electrical load forecasting through repeated wavelet transform assisted SVM model. *Appl. Soft Comput.* **2021**, *111*, 107730. [CrossRef]
64. ArunKumar, K.E.; Kalaga, D.V.; Mohan Sai Kumar, C.; Kawaji, M.; Brenza, T.M. Comparative analysis of Gated Recurrent Units (GRU), long Short-Term memory (LSTM) cells, autoregressive Integrated moving average (ARIMA), seasonal autoregressive Integrated moving average (SARIMA) for forecasting COVID-19 trends. *Alex. Eng. J.* **2022**, *61*, 7585–7603. [CrossRef]
65. Kiyoki, S.; Yoshida, S.; Rushdi, M.A. Machine Learning-Based Prediction of 2 MW Wind Turbine Tower Loads During Power Production Based on Nacelle Behavior. *Energies* **2025**, *18*, 216. [CrossRef]
66. Huang, J.; Chen, Q.; Yu, C. A New Feature Based Deep Attention Sales Forecasting Model for Enterprise Sustainable Development. *Sustainability* **2022**, *14*, 12224. [CrossRef]
67. Qiu, X.; Hu, J.; Zhou, L.; Wu, X.; Du, J.; Zhang, B.; Guo, C.; Zhou, A.; Jensen, C.S.; Sheng, Z. Tfb: Towards comprehensive and fair benchmarking of time series forecasting methods. *arXiv* **2024**, arXiv:2403.20150. [CrossRef]
68. Shao, Z.; Wang, F.; Xu, Y.; Wei, W.; Yu, C.; Zhang, Z.; Yao, D.; Sun, T.; Jin, G.; Cao, X. Exploring progress in multivariate time series forecasting: Comprehensive benchmarking and heterogeneity analysis. *IEEE Trans. Knowl. Data Eng.* **2025**, *37*, 291–305. [CrossRef]

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