



## OPEN ACCESS

## EDITED BY

Jinran Wu,  
Australian Catholic University, Australia

## REVIEWED BY

Linfei Yin,  
Guangxi University, China  
Du Zongjuan,  
Xi'an Jiaotong University, China

## \*CORRESPONDENCE

Yuze Ji,  
✉ jcyjuzhe@163.com

RECEIVED 30 March 2024

ACCEPTED 02 May 2024

PUBLISHED 14 June 2024

## CITATION

Chen W, Wang X, Ji Y, Zhang Y, Zhu J and Ma W (2024), A physical virtual multi-graph convolutional coordinated prediction method for spatio-temporal electricity loads integrating multi-dimensional information. *Front. Energy Res.* 12:1409647. doi: 10.3389/fenrg.2024.1409647

## COPYRIGHT

© 2024 Chen, Wang, Ji, Zhang, Zhu and Ma. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# A physical virtual multi-graph convolutional coordinated prediction method for spatio-temporal electricity loads integrating multi-dimensional information

Wengang Chen, Xinrui Wang, Yuze Ji\*, Yujuan Zhang, Jianfei Zhu and Weitian Ma

Jincheng Power Supply Company, State Grid Shanxi Electric Power Company Limited, Jincheng, China

Traditional load prediction methods are unable to effectively predict the loads according to the spatial topology of each electricity consumer in neighboring areas and the load dependency correlations. In order to further improve the load prediction accuracy of each consumer in the region, this paper proposes a short-term prediction method of electric load based on multi-graph convolutional network. First, the input data are selected with maximum information coefficient method by integrating multi-dimensional information such as load, weather, electricity price and date in the areas. Then, a gated convolutional network is used as a temporal convolutional layer to capture the temporal features of the loads. Moreover, a physical-virtual multi-graph convolutional network is constructed based on the spatial location of each consumer as well as load dependencies to capture the different evolutionary correlations of each spatial load. Comparative studies have validated the effectiveness of the proposed model in improving the prediction accuracy of power loads for each consumer.

## KEYWORDS

graph convolutional network, short-term load, multidimensional information, spatiotemporal prediction, maximum information coefficient

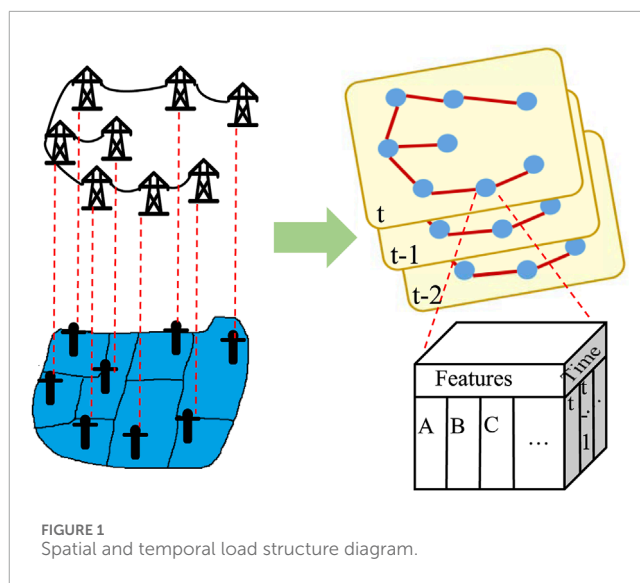
## 1 Introduction

The global electricity demand is experiencing rapid growth, and the structure of urban distribution networks is becoming increasingly complex, which elevates the challenges associated with power grid scheduling and control (Hou et al., 2021). The ongoing expansion of hybrid renewable power systems has led to the integration of a substantial number of variable renewable energy sources, such as wind and solar, transforming the grid into an active distribution network. This transformation has concomitantly increased the volatility and uncertainty inherent to power systems (Cleary et al., 2015). Accurate load prediction is of paramount importance for enhancing the safety, stability, and efficient operation of the power grid (Celebi and Fuller, 2012). Furthermore, as power systems undergo reform, electricity sales companies and virtual power plants participating in the electricity market

must accurately predict the electricity consumption of individual consumers (Aparicio et al., 2012).

The existing research on power load prediction can be broadly categorized into two primary methodological approaches: statistical models and machine learning techniques. The statistical modeling approach offers simplicity and expedient prediction capabilities. Prominent statistical methods include linear regression and exponential smoothing (Shi W. et al., 2023). Saber and Alam (2017) leveraged the autoregressive integrated moving average model to analyze the correlation between load demand and influential factors, and established a non-stationary stochastic prediction framework. However, such statistical techniques generally suffer from limitations in prediction accuracy and robustness. In contrast, machine learning methods possess adaptive and self-learning capabilities that have demonstrated improvements in load prediction precision. These advanced analytical techniques include support vector machines, extreme learning machines, long short-term memory (LSTM) networks, and convolutional neural networks (Li et al., 2020; Samadianfard et al., 2020; Zhang J. et al., 2021; Tang et al., 2021; Roy and Yeafi, 2022; Sun et al., 2023; Deng et al., 2024). Specifically, Li et al. (2020) proposed a power load decomposition and reconstruction prediction approach based on support vector machines. Furthermore, Roy and Yeafi (2022) and Sun et al. (2023) leveraged machine learning theory to establish residual self-attention encoding-decoding networks for electricity consumption and wind power prediction, effectively capturing the coupling relationships within the data. Additionally, Tang et al. (2021), Zhang J. et al. (2021) and Samadianfard et al. (2020) employed echo state networks, LSTM, and multi-layer perceptrons to predict wind direction, speed, and power generation. While the aforementioned methods utilize multi-dimensional information, such as load data and weather factors, to model the temporal correlations in load patterns, they have largely overlooked the potential spatial correlations in electricity consumption among multiple consumers. Neighboring consumers are affected by factors such as weather, electricity prices, and holidays, exhibiting similar electricity consumption behaviors and load profiles (Lin et al., 2021). Fully capturing and leveraging the spatial correlation information among these neighboring consumers has the potential to further improve the accuracy of load prediction. However, the non-Euclidean, interconnected graph structure of the consumer data limits the direct applicability of conventional neural network architectures, and thus necessitates specialized modeling approaches capable of learning from the complex spatial correlation of neighboring consumers.

Graph neural networks have attracted widespread attention because they can learn implicit representations of node data on graph structures and process non-Euclidean spatial data. Currently, graph neural networks have been successfully applied in fields such as transportation and load prediction. Yan et al. (2021) proposed a multi-time scale traffic prediction method based on graph convolutional networks, which treats each road sensor as a node to construct a spatio-temporal module and capture spatio-temporal correlations. Liao et al. (2023) established a three-dimensional Gaussian wake function that represents the relevant information of each wind turbine and used graph neural networks combined with attention mechanisms to predict the output power of non-uniform wind farms, reducing prediction errors. Shi P. et al. (2023) proposed



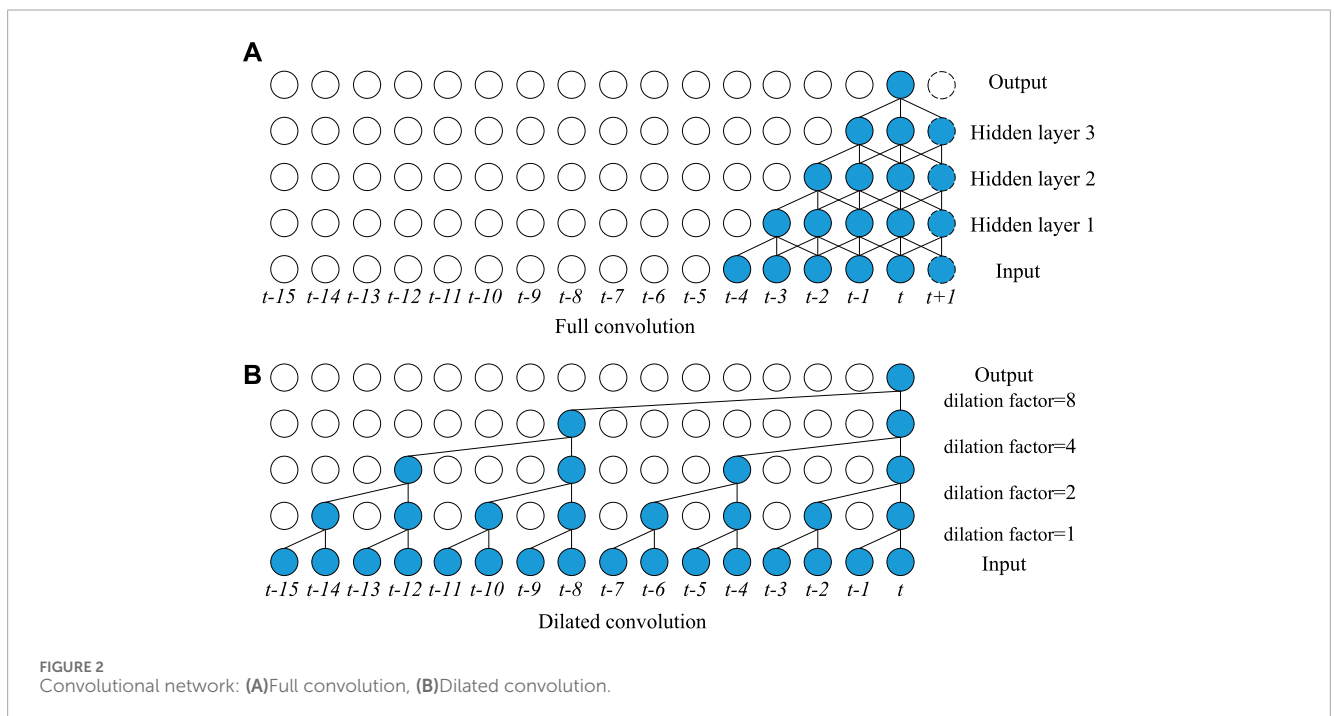
a multi-user short-term power load spatio-temporal prediction method using multi-head attention and adaptive graph theory and compared it with various methods. Zhang L. et al. (2021) used K-means clustering to divide user groups, capture the intrinsic spatio-temporal correlation information of the data using local spatio-temporal graphs, and finally aggregate the calculation results of each part to predict the future spatio-temporal power demand sequence. Fahim et al. (2024) took a load of each charging station as a node, used an adaptive adjacency matrix to reflect the spatial relationship between stations, and proposed a multi-station charging demand prediction method for electric vehicle charging stations based on graph networks. Existing literature has shown that graph neural networks can explore potential relationships between loads and improve prediction accuracy. However, the above literature only considers the fixed spatial connection relationships of consumers and relies on a single graph representation, failing to reflect the various spatial correlations between electricity loads in the neighbors.

Multimodality neural networks have improved prediction accuracy, which has attracted the attention of researchers. Zheng et al. (2023) used virtual dynamic graph and physical road graph to extract heterogeneous, variable, and inherent spatial patterns of the road network. Liu et al. (2020) presented a physical-virtual collaboration graph neural network for passenger flow prediction. The network is a general model that can be directly applied to online pedestrian flow prediction. Xiu et al. (2024) adopted parallel convolutional networks and combines relational data within the metro network to predict ridership. In addition, the train timetable as feature input to the network, improving prediction accuracy. However, there is limited application of literature in load prediction. How to design a proper prediction network based on the characteristics of electricity demands is an important issue.

Multi-graph convolutional networks have been applied in literature for load prediction. Wei et al. (2023) presented a novel multi-graph neural networks for short-term electricity demand prediction, which is embedded with the directed static graph and directed dynamic graph. The results show that the network has

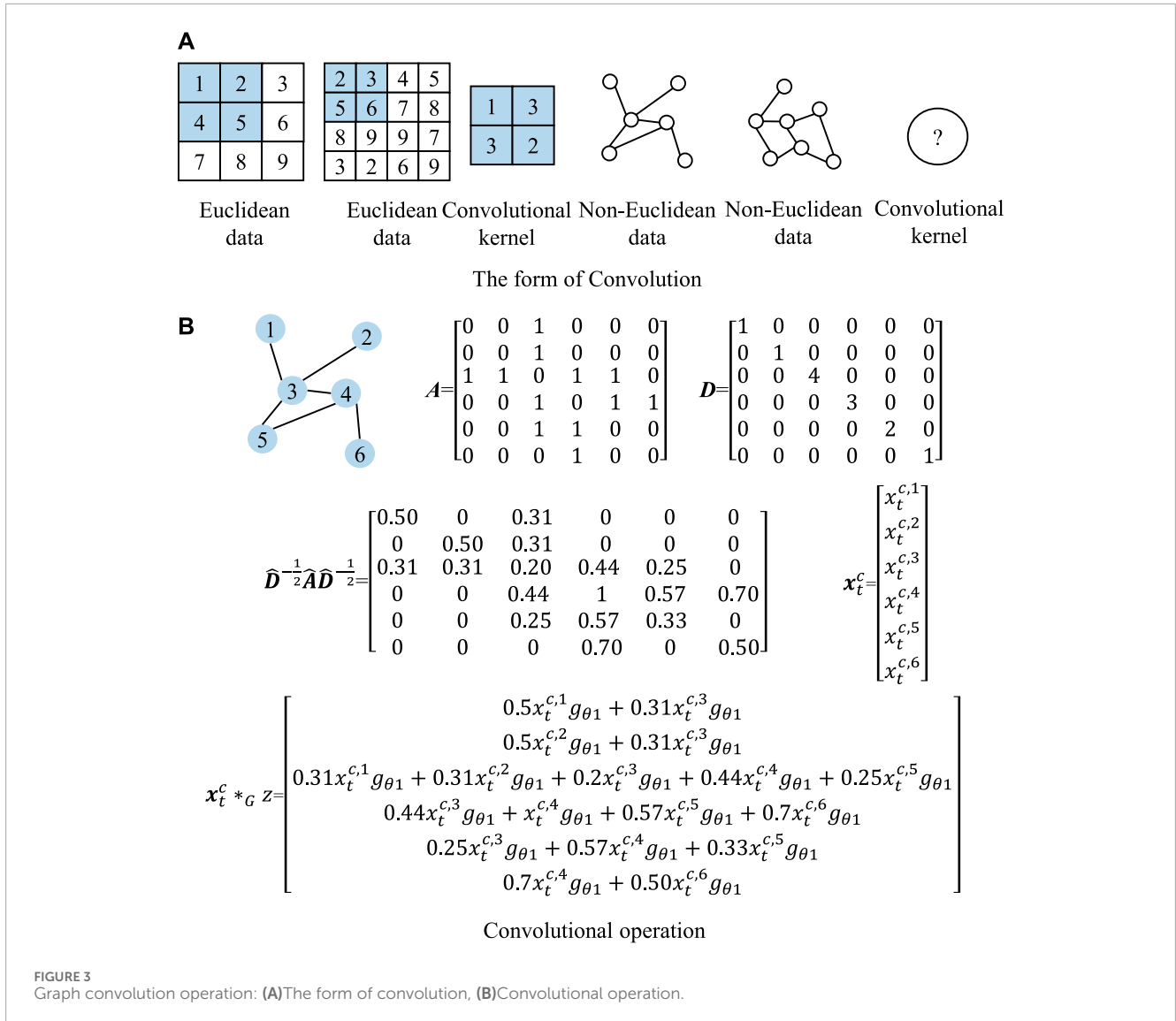
TABLE 1 Influencing factors of electric load.

| Type                     | Factor            | Meaning                               | Label |
|--------------------------|-------------------|---------------------------------------|-------|
| Historical load          | Load              | Historical electricity load           | A     |
| Time                     | Time              | Time                                  | B     |
|                          | Date              | Weekday, weekend, holiday             | C     |
|                          | Temperature       | Temperature at the time of prediction | D     |
| Meteorological condition | Humidity          | Humidity at the time of prediction    | E     |
|                          | Wind speed        | Wind speed at the time of prediction  | F     |
|                          | Rain              | Rain at the time of prediction        | G     |
| Price                    | Electricity price | Peek, flat, valley                    | H     |



a strong ability to capture periodic features. Yanmei et al. (2024) adopted dynamic load knowledge graph to extract the correlation between internal at-tributes and external influencing factors of various loads. Moreover, the attention mechanism enhances the learning ability of load feature representation. To capture complex non-linear correlations of loads, Wang et al. (2023) proposed spatial and temporal graph neural network for residential load prediction. The multiple dependence graphs consists of synchronization graph and causality graph, which can model linear and non-linear dependence. However, the exist research has not fully captured the in-dependence with multidimensional data, and electricity demand is associated with various complex and unknown factors. Therefore, predefined graphs cannot fully reflect load correlations. In addition, the coupling of spatio-temporal multidimensional information and the large amount of data make effective utilization to improve model performance another key issue.

This article proposes a multi-graph convolutional spatio-temporal collaborative prediction method for power load integrating multi-dimensional information. By constructing a multi-graph network, the spatial information of each consumer's load is fully captured to improve prediction accuracy. First, based on historical load, weather, and electricity prices, the maximum information coefficient (MIC) is used to analyze the correlation of load sequence and to construct input data that integrates multi-dimensional information. Then, the network adopts a dilated convolution and gating mechanism to parallelly capture the practical information of temporal loads. Moreover, based on the actual location connection between consumers and the similarity of electricity loads, a physical, virtual multi-graph convolutional model is established to capture various interrelationships between loads in space. T the performance of the proposed model is tested on real



electricity datasets and compared with other baseline models to verify its effectiveness.

The innovation of this paper is as the following:

- The gated casual convolution is adopted to accelerate the temporal convolution, which can capture correlation of time series information.
- We proposed a physical virtual multi-graph convolutional network to fully capture electricity load evolution patterns. The physical graph contains connection and distance data, which is based on realistic grid topology. The virtual graphs are built based on human domain knowledge.

Moreover, the main contribution of this paper is to use MIC to obtain the correlation of nonlinear influencing factors, which reduces input data redundancy. Specifically, this method filters out irrelevant spatio-temporal data and selects high MIC values as input, reducing the interference of input on the prediction results.

## 2 Spatio-temporal network of electricity loads

### 2.1 Spatial and temporal load structure

Stable electricity promotes the development of social production, and electricity is transmitted through power grid lines. The power consumption from different spatial locations is ultimately integrated into the power load of different grid nodes. As shown in Figure 1, different nodes in the grid topology correspond to electricity demand generated in different actual geographic areas. The load in an area corresponding to a grid node is regarded as the information of the nodes in the graph  $G$ , and the connection between grid nodes is regarded as the edges between the nodes in the graph  $G$ . We use the graph  $G(V, E, A)$  to describe the spatial load information, where  $V$  is a node,  $E$  is an edge, and  $A$  is an adjacency matrix, representing the connection between nodes. Each node in the graph  $G$  generates data with a total number of features  $F$  in a time interval. As shown in Figure 1, each time slice is a spatial graph that records the feature information of all nodes in the time interval.

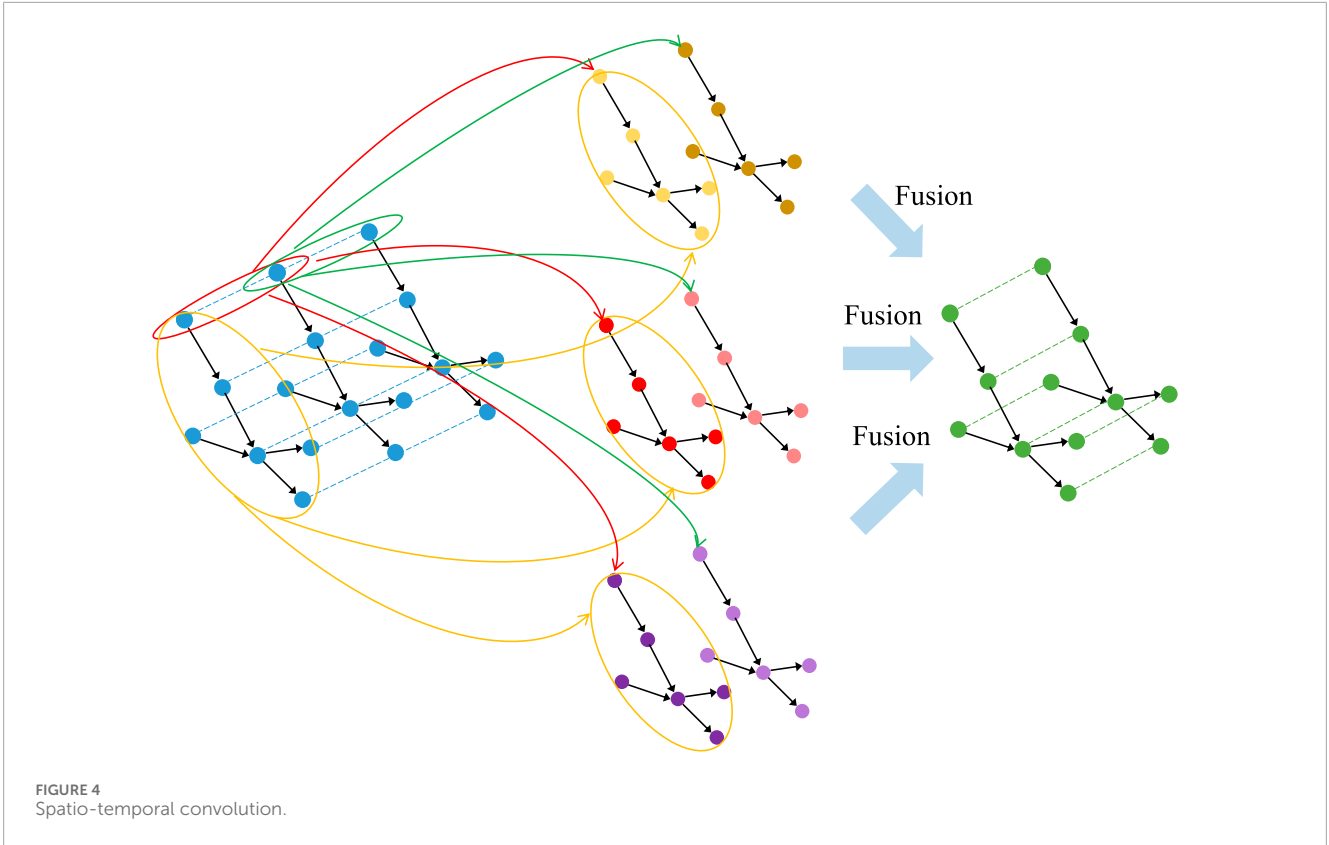


FIGURE 4 Spatio-temporal convolution.

## 2.2 Definition of parameters and sets for electricity load prediction

Let the electricity load generated by each node in the graph in a future time period be the forecast data. Let  $x_t^{f,i}$  be the features  $f$  generated by node  $i$  in the time interval  $t$ , then  $X^c = (x_1^f, \dots, x_\tau^f)$  be the features  $f$  generated by all nodes in time period  $\tau$ , and  $\chi = (X_1, \dots, X_\tau)$  be all features generated by all nodes in time period  $\tau$ . It is assumed that the load  $y_t^i = x_t^{f,i}$  generated by a node  $i$  at a certain time  $t$ , that is, all the characteristics  $\chi$  including the load generated by all nodes within a certain time period are known. Then, the electricity load  $\hat{Y} = (\hat{y}^1, \dots, \hat{y}^N)$  for a certain time period in the future is predicted where  $\hat{y}^j$  is the electricity load during a certain time period of the node.

## 2.3 Maximum information coefficient

The main factors influencing load prediction are historical loads, weather conditions, time, and electricity prices (Quilumba et al., 2014; Sun et al., 2022). Table 1 summarizes the load influencing factors. Although applying influencing factors directly to neural networks as input data can also predict loads, excessive data increases computational complexity and speed. Using proper methods to select input data can improve prediction accuracy and accelerate computational speed. Therefore, this paper applies the maximum information coefficient theory for feature extraction.

The maximum information coefficient was proposed by Reshef based on mutual information theory (Reshef et al., 2011). MIC can analyze the linear and nonlinear correlation between two variables

and screen parameters that affect load. The mutual information between sequences  $X_a$  and  $Y_a$  can be expressed as Eq. 1.

$$I_m(X_a, Y_a) = \sum_{x_a \in X_a} \sum_{y_a \in Y_a} p(x_a, y_a) \log_2 \frac{p(x_a, y_a)}{p(x_a)p(y_a)} \quad (1)$$

where  $I_m(X_a, Y_a)$  represents mutual information, and  $p(\cdot)$  is the probability density function,  $x_a \in X_a$  and  $y_a \in Y_a$ .

Let  $D_a = \{(x_{a,i}, y_{a,i}), i = 1, \dots, n\}$  be the set of binary data, and divide the value domains of  $X_a$  and  $Y_a$  into segments  $p_a$  and  $q_a$  in grid  $G_a$ . Define the maximum mutual information of  $D_a$  in grid  $G_a$  to be  $I_{mi}$  that is calculated using the Eq. 2.

$$I_{mi}(D_a, p_a, q_a) = \max I_m(D_a | G_a) \quad (2)$$

where  $D_a | G_a$  represents the data  $D_a$  divided by grid  $G_a$ .

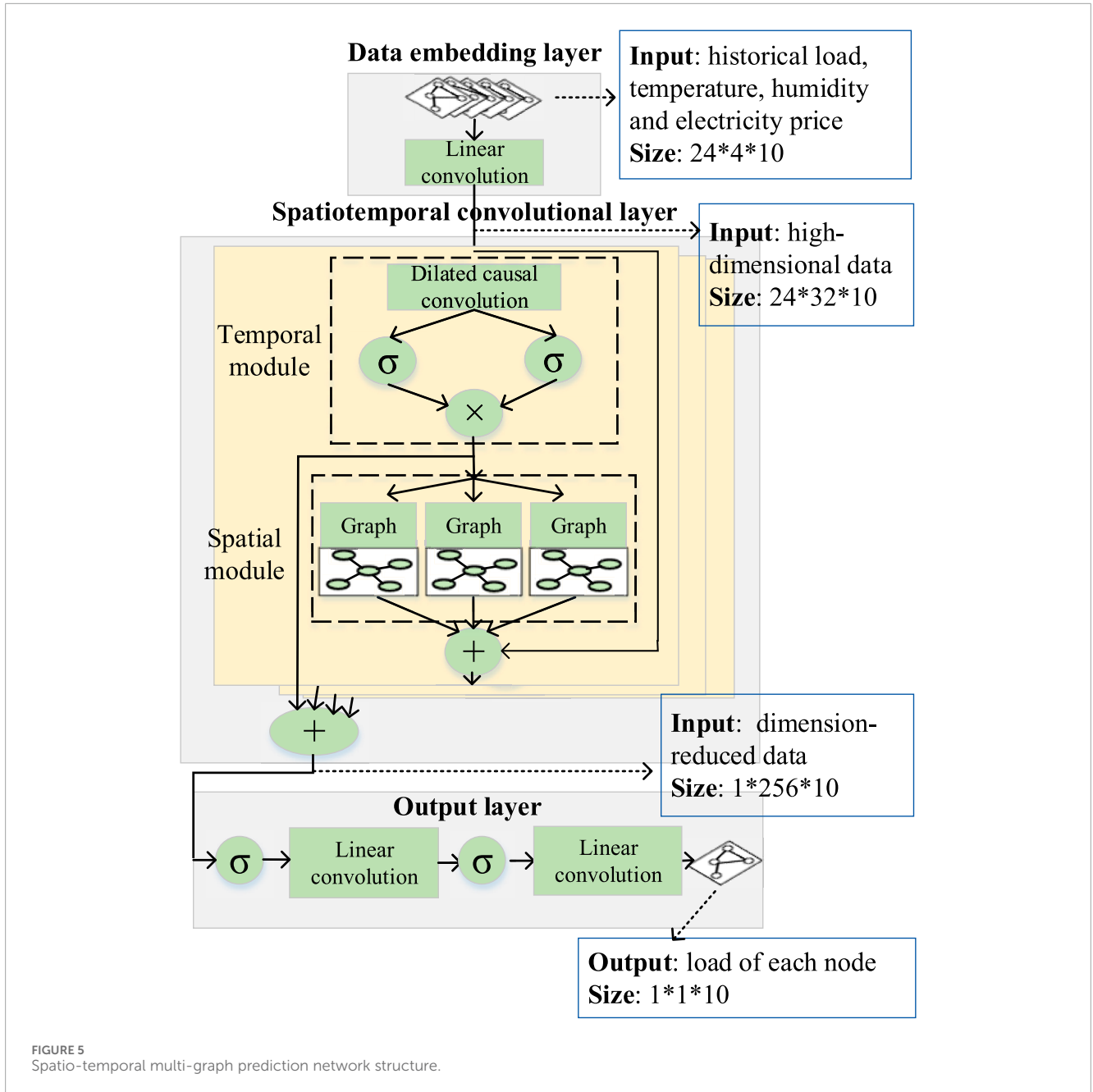
Therefore, the maximum information coefficient is formulated as Eq. 3.

$$I_{mic}(X_a, Y_a) = \max_{p_a, q_a < B(n)} \frac{I_{mi}(D_a, p_a, q_a)}{\log_2(\min(p_a, q_a))} \quad (3)$$

where  $B_n$  is the limit on the number of grid divisions, generally  $B_n = n^{0.6}$  (Reshef et al., 2011).

## 3 Spatio-temporal multi-graph prediction network

The spatio-temporal power prediction model mainly comprises a data embedding layer, a spatio-temporal prediction layer, and an output layer. The spatio-temporal prediction layer contains a



temporal convolution module and a spatial multi-graph convolution module to capture the spatio-temporal correlation features of the data and the spatio-temporal dependencies of the data.

### 3.1 Data embedding layer

The data embedding layer consists of a convolutional network that transforms the input feature volume into high-dimensional data suitable for the spatio-temporal prediction layer. A standard convolutional network consists of three parts: convolutional, pooling, and fully connected layers, where the convolution is defined by Eq. 4

$$y_c = f(x_{in}^* \omega_c + b_c) \tag{4}$$

where  $x_{in}$  is the input to the convolutional layer;  $\omega_c$  is the convolutional kernel, i.e., the weight parameter;  $b_c$  is the bias value;  $f(\cdot)$  is the convolution operation;  $\sigma$  is the activation function;  $y_c$  is the output value. In this paper, linear convolution is used to linearly transform the input data into high dimensional data by convolution operation, i.e., no activation function is used.

### 3.2 Spatio-temporal prediction layer

Mining the dependencies of loads in the time dimension can help improve prediction accuracy, and choosing an appropriate network structure is crucial. Recurrent neural networks have the structure of loops that accept data from themselves and other neurons and are particularly suitable for processing time-series

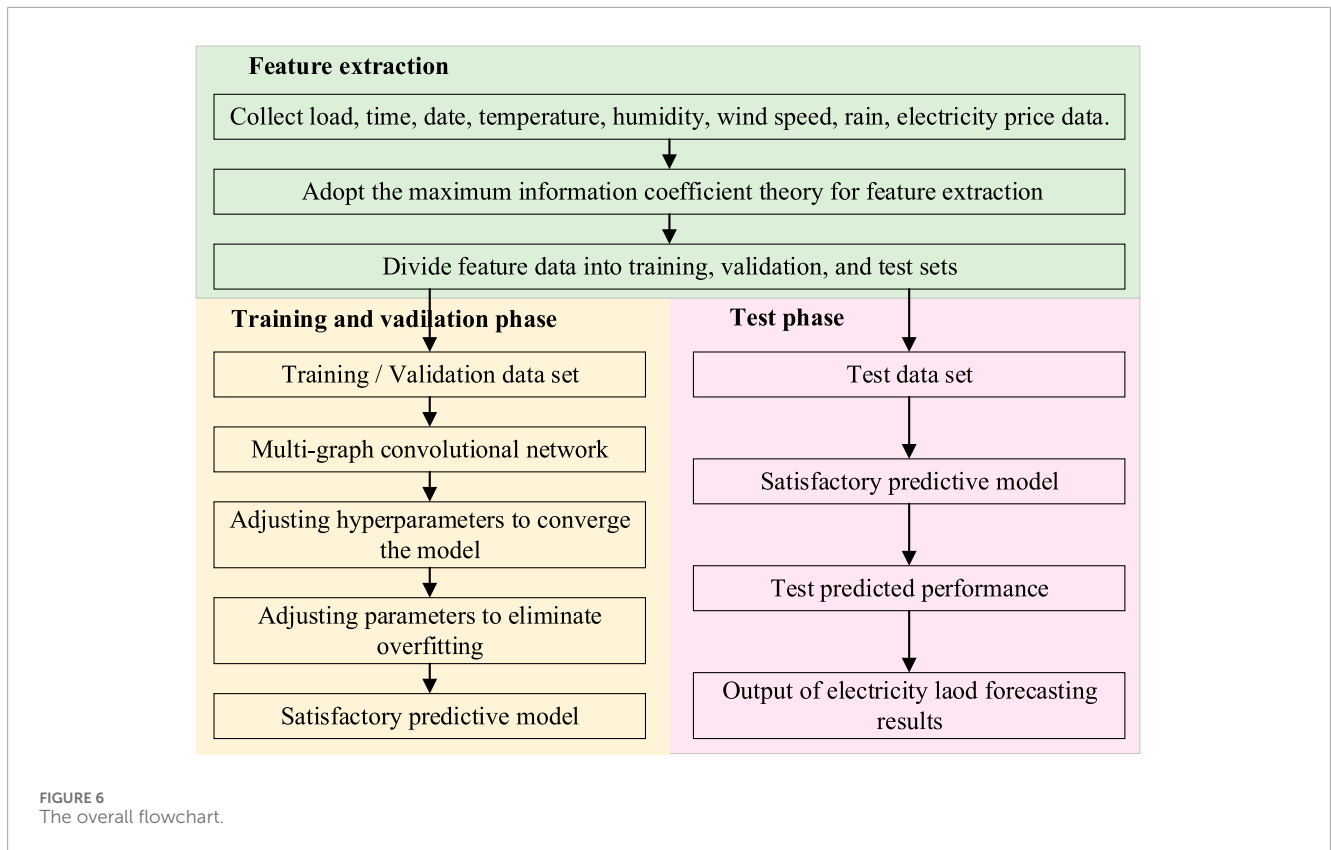


TABLE 2 Parameters of the model.

| Parameters                 | Values | Parameters      | Values     |
|----------------------------|--------|-----------------|------------|
| Chebyshev polynomial order | 3      | Batch size      | 32         |
| Iterations                 | 200    | Learning rate   | 0.0001     |
| Dropout                    | 0.3    | Dilation factor | 1, 2, 4, 8 |
| Convolution kernel         | 2      | Optimizer       | Adam       |

data. However, deeper networks take a long time to compute results, and they are prone to gradient explosion and vanishing problems. In this paper, we choose convolutional neural networks with strong robustness and faster computation to capture the mutual characteristics of data in time.

### 3.2.1 Causal convolutional networks

Causal convolution is a special convolutional neural network that utilizes only past data in its computation. Its expansion factor can be controlled to quickly increase the receptive field, thus capturing load data for a longer period. As shown in Figure 2, causal convolution does not rely on data from future moments for computation compared to ordinary convolutional networks. In addition, stacking more layers of the null convolution can result in an exponential increase in the receptive field, covering more input data and speeding up the computation. The causal convolution can

be expressed by Eq. 5 (Wu et al., 2019).

$$m(t) * x_d = \sum_{s=0}^{K-1} m(s) x_d(t - q_d s) \tag{5}$$

where  $m$  is the convolution kernel of the null convolution;  $s$  is the serial number of the convolution kernel;  $K$  is the size of the convolution kernel;  $x_d \in \mathbf{R}^T$  is the input sequence;  $t$  is the moment;  $q_d$  is the dilation factor, i.e., the interval between two factors.

### 3.2.2 Gated mechanisms

Gating methods can selectively control the rate of data accumulation to avoid memory saturation. Combining the gating mechanism with a casual convolutional network can capture the complex relationship between loads in the temporal dimension, which has a significant advantage in processing sequential data. The output of the gating operation can be expressed as Eq. 6.

$$h_g = g_h(\omega_h * x_g) \odot g_s(\omega_s * x_g) \tag{6}$$

where  $x_g$  is the input data;  $\omega_h$  and  $\omega_s$  are the learnable model parameters;  $g_h(\cdot)$  is the hyperbolic tangent function;  $g_s(\cdot)$  is the Sigmoid function;  $\odot$  is the operator for multiplying elements;  $h_g$  is the output value. The temporal convolution module mined the features and correlations between power loads of the time series using gated null convolution network and fed the processed data into the spatial multi-graph convolution module.

### 3.3 Spatial multi-graph convolution module

In order to exploit the dependencies between electric loads in space, this section proposes a representation and calculation method of spatial loads based on spectral graph theory. Then, a physical-virtual multi-graph convolutional network based on the spatial location of loads and load similarity is built to represent the different dependencies of loads in the spatial dimension.

#### 3.3.1 Spectral convolution

The graph load data is non-Euclidean space data, and each load node has a different connection relationship with other nodes. Moreover, the convolutional network is based on the translation invariant operation of the data, which cannot be directly applied to the non-Euclidean space. Bruna et al. (2014) defined graph convolution operation in spectral space based on graph theory and expressed the graph structure as a mathematical form. As a result, the non-Euclidean space data is transformed into Euclidean data for convolution operation.

In spectral theory, graph information can be represented by a Laplace matrix  $L$ . The equation is  $L = D - A$  and the standard form is  $L = I_N - D^{\frac{1}{2}}AD^{\frac{1}{2}}$  where  $A$  is the adjacency matrix,  $D$  is the degree matrix, and  $I_N$  is the identity matrix. Let  $\alpha_{ij}$  and  $d_{ij}$  be the elements of  $A$  and  $D$ , and  $i$  and  $j$  be the number of rows and columns, then  $d_{ij} = \sum_j \alpha_{ij}$ . Decompose  $L$  into eigenvalues,  $L = U\Lambda U^T$ , where  $U$  is the eigenvector and  $\Lambda$  is the diagonal array of eigenvalues. The graph convolution operation relies on the Fourier transform, defining the Fourier transform of the information on the graph  $x$  as  $\hat{x} = U^T x$  and  $\hat{x}$  as the value of  $x$  in the spectral domain. Accordingly, the Fourier inverse transform of  $x$  is  $x = U\hat{x}$ . By the convolution theorem, the Fourier transform of the convolution of signals is equal to the product of their individual Fourier transforms (Shuman et al., 2013). Therefore, it is possible to multiply the Fourier transform of the information on the graph by the information in the spectral domain and then invert the transformation to obtain the convolution result on the graph, as shown in Eq. 7.

$$x *_{GZ} z = U((U^T x) \odot (U^T z)) = U g_{\theta} U^T x \quad (7)$$

where  $x$  and  $z$  are the signals on the graph;  $*_G$  is the graph convolution;  $g_{\theta}$  is the convolution kernel, and  $g_{\theta} = U^T z$

The graph convolution operation can be realized based on Eq. 7. However, calculating the Laplace matrix is cumbersome when the graph size is large. Therefore, the Chebyshev graph convolution approximation is used to solve the convolution kernel to simplify the operation:

$$x *_{GZ} z = U g_{\theta} U^T x \approx \sum_m^{M-1} \theta_m T_m(\tilde{L}) x \quad (8)$$

where  $\theta_m$  is the Chebyshev polynomial coefficients;  $T_m(\tilde{L})$  is the Chebyshev polynomial,  $T_m(\tilde{L}) = 2\tilde{L}T_{m-1}(\tilde{L}) - T_{m-2}(\tilde{L})$  and  $T_0(\tilde{L}) = 1$ ,  $T_1(\tilde{L}) = \tilde{L}$ ;  $\tilde{L} = 2L/\lambda_{\max} - I_N$ ; and  $m$  is the order of Chebyshev polynomial.

The information in the graph is updated by the order information of itself and its neighboring nodes  $M-1$ , and the depth of the transmitted information can be adjusted by controlling the maximum order  $M$ . In the actual calculation, the value of

$L = I_N - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$  is dispersed, so it is generally replaced  $\widehat{D}^{-\frac{1}{2}}\widehat{A}\widehat{D}^{-\frac{1}{2}}$ , where  $\widehat{A} = A + I_N$  and  $\widehat{D}$  are the degree matrices of  $\widehat{A}$  (Yan et al., 2021).

An example of graph convolution operation is shown in Figure 3. Figure 3A is form of convolution, and the right side shows the convolution on non-Euclidean space. Figure 3B shows the spectral convolutional operation. Given a 6-bus grid, the adjacency matrix  $A$  and degree matrix  $D$  are obtained based on the grid. Then, we can get  $D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ . Given the input data  $x_i^c$  and convolution kernel  $g_{\theta 1}$ , the result of graph convolution  $x_i^c *_{GZ} z$  can be obtained.

The representation of the convolutional results on the network is further enhanced by the activation function ReLU:

$$h_f = g_r \left( \sum_m^{M-1} \theta_m T_m(\tilde{L}) x \right) \quad (9)$$

where  $g_r$  is the activation function ReLU and  $h_f$  is the spatial convolution output.

#### 3.3.2 Multi-graph construction

Different dependencies are implied between loads at different locations in space. The load relationship implied by the different interconnections and distances of electric loads in different regions is called neighborhood dependence, and the relationship implied by the different load similarities due to the different patterns of electricity use is called load correlation. In order to mine the proximity dependence and load correlation of electric loads at each location in space, a physical connectivity map, location distance map, and virtual correlation map are constructed.

1) Physical connection graph: The connection matrix is established based on the connection relationship between the lines where the power loads are located at each location, i.e., the interconnections of the nodes in the grid topology. The element of this matrix can be defined as Eq. 10.

$$a_{ij}^a = \begin{cases} 1, \text{ node } i \text{ is connected to node } j \\ 0, \text{ else} \end{cases} \quad (10)$$

2) Positional distance graph: A distance matrix is created based on the distance of the nodes where each power load is located  $A^d$ . The element of this matrix can be defined as Eq. 11.

$$a_{ij}^d = \begin{cases} \exp\left(-\frac{d_{ij}^2}{\varepsilon_d^2}\right), & d_{ij} \geq \rho_d \\ 0, & d_{ij} < \rho_d \end{cases} \quad (11)$$

where  $a_{ij}^d$  is the element in  $A^d$ ;  $d_{ij}$  is the distance between node  $i$  and node  $j$ ;  $\varepsilon_d$  is the matrix threshold parameter;  $\rho_d$  is the distance threshold parameter.

3) Virtual similarity graph: A similarity matrix is created based on the similarity between the electrical loads at each location  $A^s$  (Shi J. et al., 2023). The element of this matrix can be defined as Eq. 12, and the similarity between the load of node  $i$  and the load of node  $j$  can be calculated by Eq. 13.

$$a_{ij}^s = \exp(-\rho_s c_{ij}^d) \quad (12)$$

$$c_{ij}^d = \sqrt{\sum_{t=1}^T (x_{ti}^L - x_{tj}^L)^2} \quad (13)$$



where  $a_{ij}^s$  is the element in  $A^s$ ;  $\rho_s$  is the parameter controlling the decay rate;  $c_{ij}^d$  is the similarity between the load of node  $i$  and the load of node  $j$ ;  $x_{i,i}^l$  and  $x_{j,j}^l$  are the loads of node  $i$  and node  $j$  at time  $t$ .

The matrices  $A^a$ ,  $A^d$  and  $A^s$  are used to obtain the corresponding  $L^a$ ,  $L^d$  and  $L^s$  by bringing them into the standard computational form of the Laplace matrix, respectively. The Laplace matrices are brought into Eqs 8, 9 to obtain the convolution results of the loadings in each graph.

### 3.3.3 Multi-graph fusion

The graph fusion method is the key to graph neural networks, and a simple average summation of each graph will reduce the prediction performance. In this paper, we use the convolution results of each graph to be fused into a new graph by weighted summation to reflect the degree of influence of each graph in space. The weights of each graph are normalized using the Softmax function formulates as Eq. 14.

$$\mathbf{w}_{as}, \mathbf{w}_{ds}, \mathbf{w}_{ss} = g_{so}(\mathbf{w}_a, \mathbf{w}_d, \mathbf{w}_s) \quad (14)$$

where  $g_{so}$  is the Softmax function;  $\mathbf{w}_a$ ,  $\mathbf{w}_d$  and  $\mathbf{w}_s$  are the learnable weight parameters of the physical connectivity graph, positional distance graph and virtual association graph, respectively;  $\mathbf{w}_{as}$ ,  $\mathbf{w}_{ds}$  and  $\mathbf{w}_{ss}$  are the weights of the graphs after normalization, which indicate the influence degree of each graph in the new graph.

The weight parameters are multiplied with the results of the convolution of each graph and then summed, as shown in Eq. 15.

$$\mathbf{h}_{new} = \mathbf{w}_{as} \odot \mathbf{h}_f^a + \mathbf{w}_{ds} \odot \mathbf{h}_f^d + \mathbf{w}_{ss} \odot \mathbf{h}_f^s \quad (15)$$

where  $\mathbf{h}_{new}$  is the convolution result of the new graph;  $\mathbf{h}_f^a$ ,  $\mathbf{h}_f^d$  and  $\mathbf{h}_f^s$  are the convolution outputs of the physical connectivity graph, the positional distance graph and the virtual association graph, respectively.

The data features are extracted through the spatio-temporal convolution module, and the spatio-temporal convolution process is shown in Figure 4.

## 3.4 Output layer

The output layer is connected to the spatio-temporal prediction layer, which converges and transforms the passed results into the desired dimensions. The use of linear convolution can effectively transform the data dimension, and the selection of an appropriate activation function can extract the nonlinear features of the data. Due to the large degree of nonlinearity and high dimensionality of the data, this paper adopts the ReLU activation function and linear convolution twice in series, i.e., the predicted power load value is finally obtained without losing too much information each time. The prediction step size of this network is adjustable, i.e., the load value can be obtained at one time for more than one moment.

## 3.5 Spatio-temporal multi-graph prediction network structure

Before inputting the data, the resulting data should be blank-filled, outliers removed, and corrected (Azeem et al., 2021). The

**Input:** Data set of  $\{\mathbf{X}_i\}_M$ ; the size of  $\mathbf{X}_i$  is  $i * f * t$ ;  $i$  is node,  $f$  is feature, and  $t$  is time.

**Output:** Multi-graph convolution model result  $\hat{\mathbf{Y}}$ ;

```

1: for each epoch do
2:   for each batch do
3:     Linear convolution:  $Conv(\mathbf{X}_i) \rightarrow X_{start}$ ;
4:     Initial value  $\theta \rightarrow X_{res}$ ;
5:     for each spatio-temporal convolutional
layer do
6:       if first layer then
7:          $X_{start} \rightarrow X_{in}$ 
8:       else
9:         Previous  $X_{res}$  is current  $X_{in}$ :  $X_{res} \rightarrow X_{in}$ ;
10:      end if
11:      Gated casual convolution:
 $Conv(X_{in}) \odot Conv(X_{in}) \rightarrow X_w$ ;
12:      Skip connection:  $Y_{skip} + X_{skip} \rightarrow Y_{skip}$ ;
13:      Graph 1 convolution:  $G1conv(X_{skip}) \rightarrow X_{G1}$ ;
14:      Graph 2 convolution:  $G2conv(X_{skip}) \rightarrow X_{G2}$ ;
15:      Graph 3 convolution:  $G3conv(X_{skip}) \rightarrow X_{G3}$ ;
16:      Graphs fusion:  $X_{G1} + X_{G2} + X_{G3} \rightarrow X_{res}$ ;
17:      Residual connection:  $X_{res} + Y_{res} \rightarrow Y_{res}$ ;
18:    end for
19:    Linear convolutions:  $Conv(Conv(Y_{res})) \rightarrow Y_{out}$ ;
20:    Obtain MAE of network;
21:    Adjust hyperparameters;
22:  end for
23: end for
24: Obtain result of  $\hat{\mathbf{Y}}$ .
```

Algorithm 1. Spatio-temporal multi-graph prediction algorithm.

features such as historical electric load power, weather, and date are filtered using maximum information coefficient analysis to select the most relevant features as input data into the prediction network. The structure of the spatio-temporal multi-graph prediction network is shown in Figure 5, and the corresponding multi-graph convolution algorithm is shown in Algorithm 1.

The prediction network mainly comprises a data embedding layer, a spatio-temporal prediction layer, and an output layer. The spatio-temporal prediction layer consists of multiple temporal and spatial convolutional blocks stacked together, enabling the network to capture data correlations at different temporal levels. Different spatio-temporal convolutional blocks converge different levels of information to the output layer through skip connections. In addition, residual connections are utilized in the blocks to accelerate convergence and to address possible degradation of the deep network (He et al., 2016). The overall flowchart is shown in Figure 6.

## 3.6 Evaluation indicators

The performance of the prediction network is evaluated by applying Mean Absolute Error (MAE)  $I_{MAE}$  calculated by Eq. 16, Mean Absolute Percentage Error (MAPE)  $I_{MAPE}$  calculated by Eq.

17, Mean Squared Error (MSE)  $I_{MSE}$  calculated by Eq. 18, and Root Mean Squared Error (RMSE)  $I_{RMSE}$  calculated by Eq. 19. MAE is the difference between the predicted load and the actual load, which truly reflects the prediction error, and in this paper, we choose the Mean Absolute Error as the loss function of the network.

$$I_{MAE} = \frac{1}{TM} \sum_{t=1}^T \sum_{m=1}^M |y_{t,m} - \hat{y}_{t,m}| \quad (16)$$

$$I_{MAPE} = \frac{1}{TM} \sum_{t=1}^T \sum_{m=1}^M \left| \frac{y_{t,m} - \hat{y}_{t,m}}{y_{t,m}} \right| \times 100\% \quad (17)$$

$$I_{MSE} = \frac{1}{TM} \sum_{t=1}^T \sum_{m=1}^M (y_{t,m} - \hat{y}_{t,m})^2 \quad (18)$$

$$I_{RMSE} = \sqrt{\frac{1}{TM} \sum_{t=1}^T \sum_{m=1}^M (y_{t,m} - \hat{y}_{t,m})^2} \quad (19)$$

where  $y_{t,m}$  and  $\hat{y}_{t,m}$  are the real and predicted values of load at time  $t$  node  $m$  respectively;  $m$  is the node number; and  $M$  is the total number of nodes.

## 4 Case study

### 4.1 Data set and parameters

In this paper, we use the 10 kV voltage level electric load dataset of a region in North China, including loads, weather conditions, date information and electricity prices, as shown in Table 2. All data have been desensitized and normalized to [0, 1]. The dataset contains a total of 10 bus data with a time range of 1 January 2020 to 1 June 2021, with a time interval of 60 min and a total of 24 points per day.

The predictive network model is implemented in Python software's PyTorch learning library. In the debugging process of the network, considering the size of the data volume, the data set is taken as 70% as the training set, 20% as the validation set, and 10% as the test set. After several comparative analyses and comprehensive prediction performance, the model parameters are set as shown in Table 2. Among them, the Dropout means to make the neurons not work in a certain proportion, which can make the model generalization ability stronger. In addition, the model is a single-step prediction, i.e., all bus loads at the next moment are predicted using all bus data of the previous day. The framework of the model of this paper is illustrated in Figure 7.

### 4.2 Feature selection results

The historical characteristics of Table 1 were analyzed by the maximum information coefficient analysis method to calculate the contribution of the influence of each characteristic quantity on the load, and the results are shown in Figure 8.

The meanings of the letter labels in the graph are shown in Table 1, and label I is the predicted day load. It can be seen that the MIC of historical load, time and electricity price with forecast daily load is high, which represents a strong correlation. And the MIC between electricity price and time is 1, which represents a high correlation with cyclical changes in time

and electricity price on a daily basis. Due to the higher MIC between electricity price and load, electricity price was chosen as one of the input features. The MIC for weather conditions is generally between 0.2 and 0.4, with temperature and humidity having a greater impact on load. Too much input data will reduce the computing speed of the model, in order to have better performance of the prediction network, this paper takes the threshold of MIC as 0.3. In summary, the input features are historical load, temperature, humidity and electricity price.

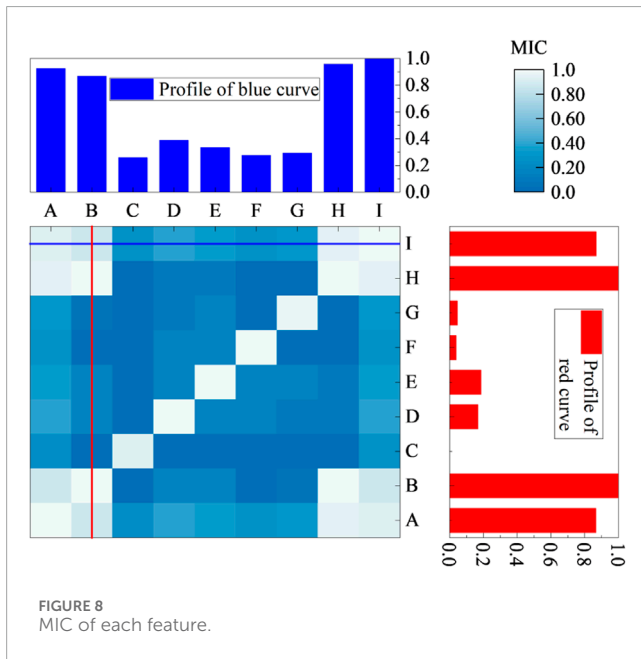
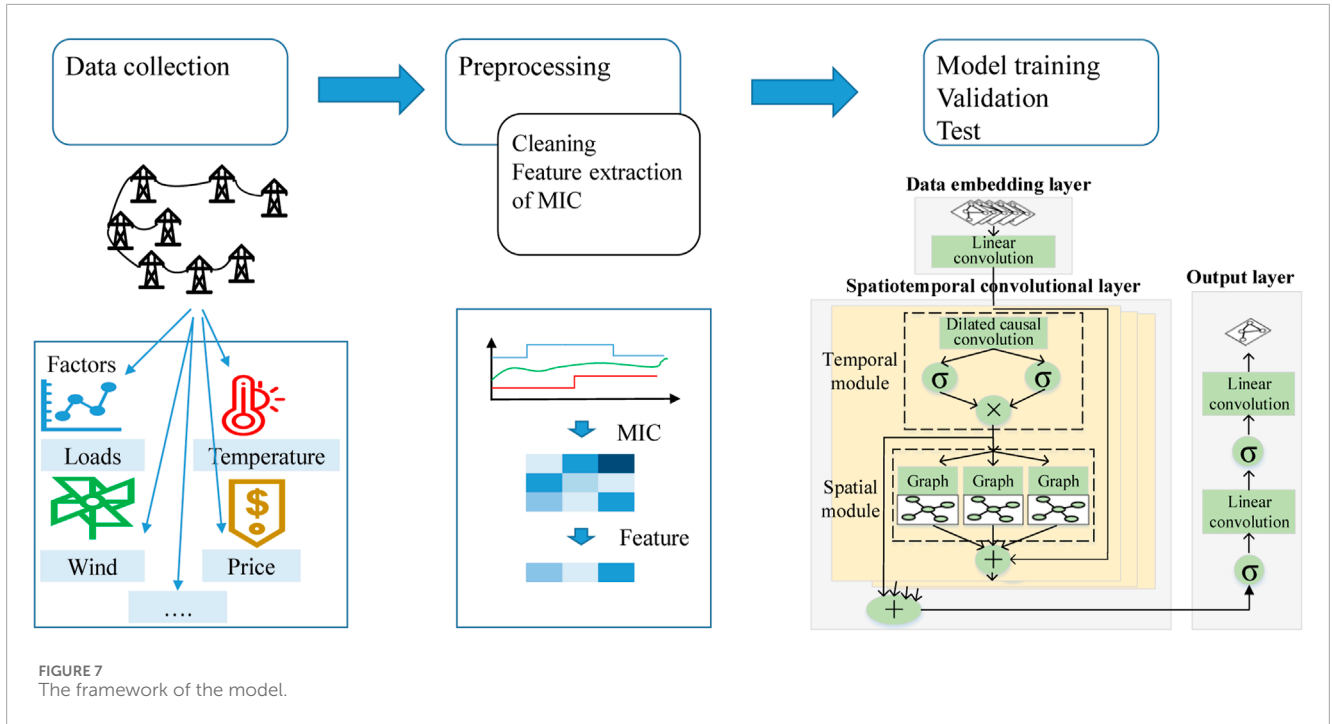
### 4.3 Analysis of forecast results

The neighbor matrices in the physical connection graph, location distance graph and virtual association graph of the prediction model are shown in Figure 9. It can be seen that since some of the nodes are not directly connected to each other, the elements of the connection matrix are 0. To control the sparsity of the graph, the distance matrix elements of the two nodes that are too close to each other are set to 0 to improve the speed of operation. The similarity matrix elements vary as the nodes have different power usage patterns. We adjust the parameters  $\rho_s$  to make the distribution of the adjacency matrix more uniform. Applying the proposed spatio-temporal multi-graph prediction model, the metrics for evaluating power load forecasts at 10 nodes are shown in Table 3. As can be seen from Table 3, the mean absolute error varies from node to node due to their different load characteristics, and the overall MAE is 0.0136. The node 4 has a smaller MAE and a larger MAPE due to its small load power and high degree of fluctuation. However, node 6 and node 7 have irregular daily power loads with high uncertainty, resulting in larger MAPE and MSE. Node 3, node 5, node 8 and node 9 have smooth and distinctly cyclical load variations and have higher predicted MAPE. Node 1 has a higher load MAE and MAPE than node 5, but low MSE and RMSE, which indicates a higher degree of deviation from the individual results of the predictive model at node 5. Overall, the multigraph convolutional model predicted a MAPE of 5.26%.

### 4.4 Model comparison

To further validate the performance of the spatio-temporal multi-graph prediction network, it is compared with the following four widely used prediction networks:

- 1) Historical Average (HA): this model takes the average of the most recent load data as the predicted value and is one of the most classical statistical methods;
- 2) Gated Recurrent Unit (GRU): this network is a type of recurrent neural network that employs a gating mechanism to filter out the information in the long term sequences, thus improving the prediction performance;
- 3) Convolutional Neural Network-Long Short-term Memory Network (CNN-LSTM): this network utilizes a convolutional neural network to extract valid information from the input data. Due to the ability of LSTM to handle longer time series, they are integrated into the CNN for prediction;



4) Spatio-Temporal Convolutional Network (STGCN): this network consists of gated linear units to extract temporal features, graph convolutional networks to extract spatial features, and multiple spatio-temporal blocks superimposed to form a prediction network (Yu et al., 2017).

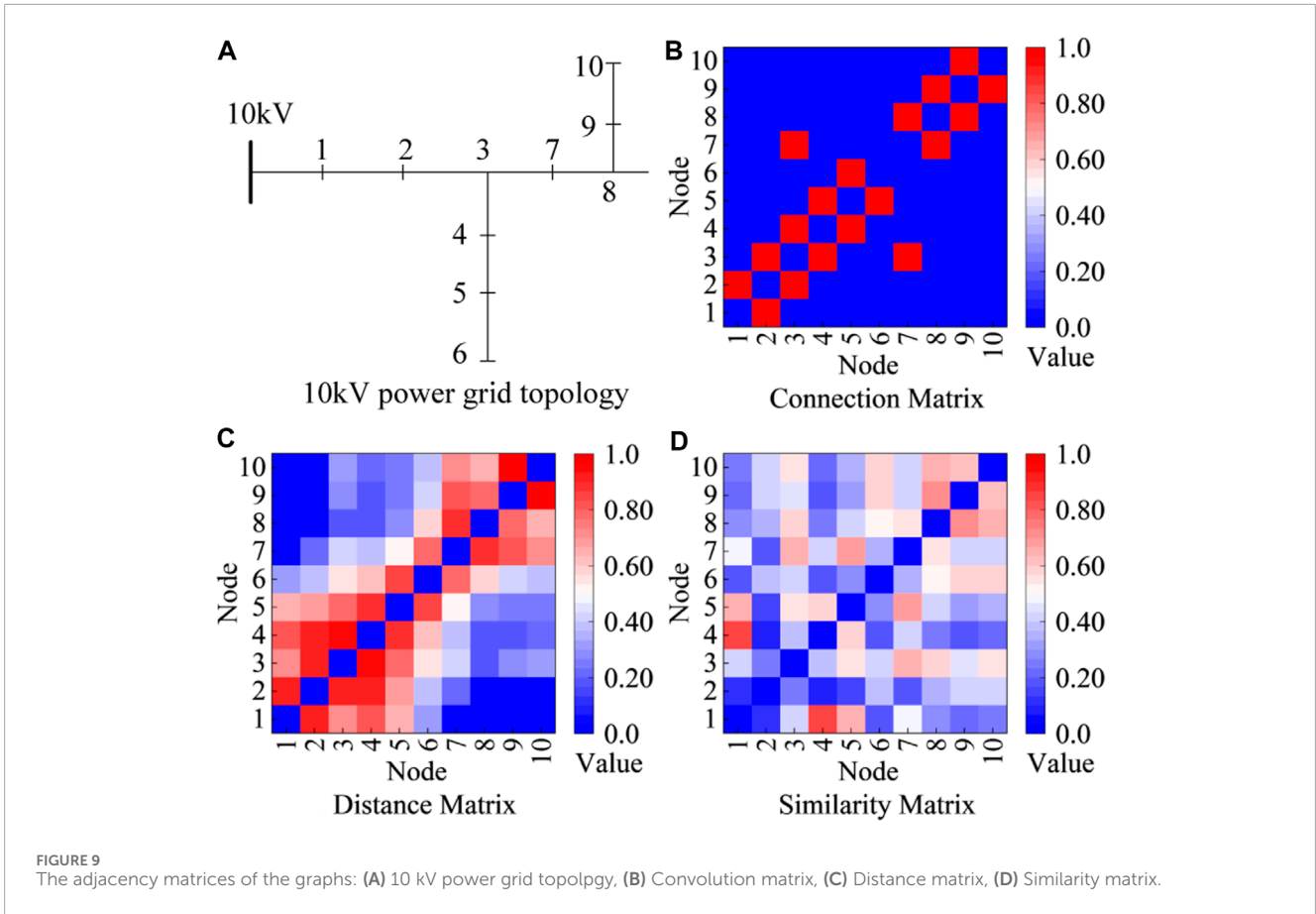
The prediction results are shown in Table 4. It can be seen that the performance of the multi-graph convolutional prediction model proposed in this paper are satisfactory. Since HA relies on simple averaging of historical loads to obtain the results, it is unable to

capture the nonlinear factors of power loads in the time series, and thus has the lowest prediction accuracy. The GRU, a neural network with memory function, captures the correlation features of loads in the time series, with a MAPE of 6.62%, which reduces by 1.44% compared with that of HA, and achieves a better result. CNN-LSTM utilizes the convolutional network to process the feature information of the input load, which further reduces the MAPE by 0.31%. STGCN, a classical graph neural network, predicts a MAE of 0.0152 and a MAPE of 5.63%, which outperforms the traditional neural networks and statistical models. This is due to the fact that graph convolutional networks can process non-Euclidean load information and capture the hidden information of spatial loads. Due to the use of physical-virtual multi-graph structure to mine the different evolutionary relationships of loads in space, the proposed method has a MAPE of 5.26%, which is the best performance. The predicted MAE and MAPE evaluation metrics for each comparison method at each node load are shown in Figure 10.

It can be seen that the MAE and MAPE of each method are different due to the different fluctuation patterns of electric loads at each node. HA is a classical statistical model with large prediction errors in predicting more volatile loads such as nodes 4, 6 and 7. While deep learning models such as GRU and CNN-LSTM have less difference in MAPE at each node. The prediction networks with graph structure such as STGCN and MGCN can learn the potential relationship of each node and can further reduce the prediction error of each load.

### 4.5 Ablation experiments

In order to analyze the contribution of each module in the proposed physical virtual multi-graph network structure, we design ablation experiments. We compare the proposed model with the following variants:



**TABLE 3** Load prediction results of each node.

| Node    | MAE      | MAPE/% | MSE/1e-5 | RMSE    |
|---------|----------|--------|----------|---------|
| Node 1  | 0.00409  | 4.60   | 2.37     | 0.00487 |
| Node 2  | 0.0219   | 4.09   | 75.34    | 0.0274  |
| Node 3  | 0.00835  | 3.00   | 10.35    | 0.0101  |
| Node 4  | 0.006794 | 9.61   | 7.52     | 0.00867 |
| Node 5  | 0.00391  | 2.13   | 2.60     | 0.00510 |
| Node 6  | 0.0359   | 8.13   | 194.23   | 0.0440  |
| Node 7  | 0.0279   | 11.73  | 108.43   | 0.0329  |
| Node 8  | 0.0101   | 2.93   | 15.54    | 0.0124  |
| Node 9  | 0.000963 | 2.48   | 16.93    | 0.0130  |
| Node 10 | 0.0169   | 3.87   | 63.41    | 0.0251  |
| Average | 0.0136   | 5.26   | 49.67    | 0.0222  |

- MGCN: The model is the proposed network, which contains the multi-graph and temporal convolutional network simultaneously.

**TABLE 4** Load forecasting results of different methods.

| Model    | MAE    | MAPE/% | MSE/1e-5 | RMSE   |
|----------|--------|--------|----------|--------|
| HA       | 0.0225 | 8.06   | 104.84   | 0.0323 |
| GRU      | 0.0192 | 6.62   | 86.58    | 0.0294 |
| CNN-LSTM | 0.0171 | 6.31   | 60.76    | 0.0246 |
| STGCN    | 0.0152 | 5.63   | 52.79    | 0.0229 |
| MGCN     | 0.0136 | 5.26   | 49.67    | 0.0222 |

- PC-GCN: In this variant, we retain the physical connection graph and remove the other graphs.
- P-GCN: Similarly, the virtual similarity graph is removed, retaining the positional graph and the physical connection graph.
- VS-GCN: This variant adopts virtual similarity matrix as features of graph, without employing the physical connection and distance graph.
- TCN: Different with above variants that contain the graph network, the variant is constructed with only temporal convolutional network.

The performances of different variants is shown in Table 5.

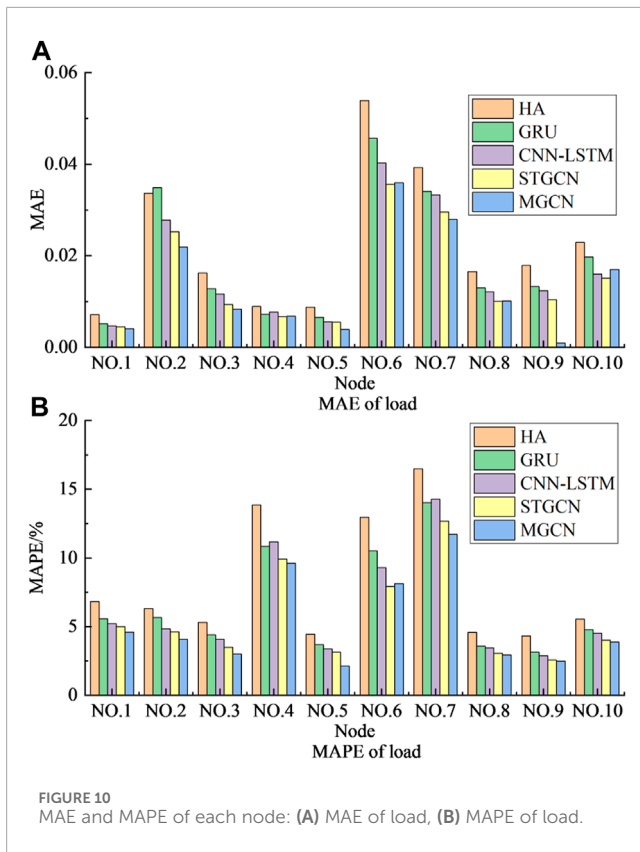


FIGURE 10 MAE and MAPE of each node: (A) MAE of load, (B) MAPE of load.

In Table 5, TCN obtains the MAPE of 6.40, which is similar to the performance of GRU and CNN-LSTM. Due to the lack of graph modules, they can only capture temporal features of data. When there is a physical graph in the model, the error in load prediction decreases significantly. In addition, We can observe that the MAE of P-GCN is very close to that of the STGCN. This indicates that physical graphs can extract hidden patterns of loads in the spatial dimension. We further combine a virtual similarity graph with the convolutional network, which achieves superior performance. Notably, the proposed virtual graph based on human domain knowledge can fully explore the evolution patterns of electricity loads.

### 5 Conclusion

In order to fully explore the correlation of various modes between the power loads of each node, this paper proposes a

multi-graph convolutional spatio-temporal synergistic prediction method for power loads by fusing multi-dimensional information, and the theoretical analysis and the results of the arithmetic examples show that:

- 1) The maximum information coefficient method can effectively analyze load prediction influencing factors, select the most relevant features and reduce the redundancy of input information;
- 2) The non-Euclidean load information is processed by using spectral graph theory, and the constructed physical-virtual multi-graph convolutional network mines multiple spatial relationships between loads at each node, enriches the spatial characteristics of loads and improves the prediction accuracy;
- 3) Compared with statistical models, traditional neural networks and graph convolution models, the multi-graph spatio-temporal prediction network proposed in this paper has high prediction accuracy, which verifies the effectiveness of the method;

Although this paper has made some progress in constructing multi-graph convolution for spatio-temporal load prediction, the graph convolution network needs to be improved further: 1) The superior performance of the graph convolution network requires multiple rounds of manual hyperparameter tuning. More generalized and concise prediction networks can be considered for future adoption to improve the model's quality. 2) The electricity demand periodically changes over a large period. For example, the electricity load during the New Year is usually similar. We can add modules to learn load characteristics if there is continuous electricity data for every year. 3) We will design more general prediction models to achieve robust performance with incomplete data.

### Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

### Author contributions

YJ: Supervision, Writing–original draft. WC: Conceptualization, Methodology, Software, Writing–original draft. XW: Software, Validation, Writing–original draft. YZ: Data curation,

TABLE 5 Load prediction results of different modules.

| Indicators | MGCN   | PC-GCN | P-GCN  | VS-GCN | TCN    |
|------------|--------|--------|--------|--------|--------|
| MAE        | 0.0136 | 0.0159 | 0.0153 | 0.0160 | 0.0178 |
| MAPE/%     | 5.26   | 5.83   | 5.60   | 6.06   | 6.40   |
| MSE/1e-5   | 49.67  | 54.88  | 52.52  | 56.48  | 68.32  |
| RMSE       | 0.0222 | 0.0233 | 0.0229 | 0.0237 | 0.0252 |

Writing–review and editing. JZ: Validation, Writing–review and editing. WM: Software, Writing–review and editing.

## Funding

The authors declare that financial support was received for the research, authorship, and/or publication of this article. This work is supported by the Science and Technology Project of State Grid Shanxi Electric Power Company Limited (no. 5205E0220001).

## Conflict of interest

Authors WC, XW, YJ, YZ, JZ, and WM were employed by State Grid Shanxi Electric Power Company Limited.

## References

- Aparicio, N., MacGill, I., Abbad, J. R., and Beltran, H. (2012). Comparison of wind energy support policy and electricity market design in Europe, the United States, and Australia. *IEEE Trans. Sustain. Energy* 3, 809–818. doi:10.1109/tste.2012.2208771
- Azeem, A., Ismail, I., Jameel, S. M., and Harindran, V. R. (2021). Electrical load forecasting models for different generation modalities: a review. *IEEE Access* 9, 142239–142263. doi:10.1109/access.2021.3120731
- Bruna, J., Zaremba, W., Szlam, A., and LeCun, Y. (2014). “Spectral networks and deep locally connected networks on graphs, 2nd int,” in Conf. on Learning Representations, ICLR 2014—Conference Track Proceedings, Banff, AB, Canada, April 14–16, 2014.
- Celebi, E., and Fuller, J. D. (2012). Time-of-use pricing in electricity markets under different market structures. *IEEE Trans. Power Syst.* 27, 1170–1181. doi:10.1109/tpwrs.2011.2180935
- Cleary, B., Duffy, A., Oconnor, A., Conlon, M., and Fthenakis, V. (2015). Assessing the economic benefits of compressed air energy storage for mitigating wind curtailment. *IEEE Trans. Sustain. Energy* 6, 1021–1028. doi:10.1109/tste.2014.2376698
- Deng, J., Wu, J., Zhang, S., Li, W., and Wang, Y.-G. (2024). Physical informed neural networks with soft and hard boundary constraints for solving advection-diffusion equations using Fourier expansions. *Comput. Math. Appl.* 159, 60–75. doi:10.1016/j.camwa.2024.01.021
- Fahim, S. R., Atat, R., Kecici, C., Takiddin, A., Ismail, M., Davis, K. R., et al. (2024). “Forecasting EV charging demand: a graph convolutional neural network-based approach,” in 2024 4th International Conference on Smart Grid and Renewable Energy (SGRE), Doha, Qatar, 8–10 January 2024, 1–6.
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). “Deep residual learning for image recognition,” in Proceedings of the IEEE conference on computer vision and pattern recognition, Seattle, WA, USA, June 14–19, 2020, 770–778.
- Hou, H., Chen, Y., Liu, P., Xie, C., Huang, L., Zhang, R., et al. (2021). Multisource energy storage system optimal dispatch among electricity hydrogen and heat networks from the energy storage operator prospect. *IEEE Trans. Industry Appl.* 58, 2825–2835. doi:10.1109/tia.2021.3128499
- Li, G., Li, Y., and Roozitalab, F. (2020). Midterm load forecasting: a multistep approach based on phase space reconstruction and support vector machine. *IEEE Syst. J.* 14, 4967–4977. doi:10.1109/jsyst.2019.2962971
- Liao, W., Wang, S., Bak-Jensen, B., Pillai, J. R., Yang, Z., and Liu, K. (2023). Ultra-short-term interval prediction of wind power based on graph neural network and improved bootstrap technique. *J. Mod. Power Syst. Clean Energy* 11, 1100–1114. doi:10.35833/mpce.2022.000632
- Lin, W., Wu, D., and Boulet, B. (2021). Spatial-temporal residential short-term load forecasting via graph neural networks. *IEEE Trans. Smart Grid* 12, 5373–5384. doi:10.1109/tsg.2021.3093515
- Liu, L., Chen, J., Wu, H., Zhen, J., Li, G., and Lin, L. (2020). Physical-virtual collaboration modeling for intra- and inter-station metro ridership prediction. *IEEE Trans. Intelligent Transp. Syst.* 23, 3377–3391. doi:10.1109/tits.2020.3036057
- Quilumba, F. L., Lee, W.-J., Huang, H., Wang, D. Y., and Szabados, R. L. (2014). Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities. *IEEE Trans. Smart Grid* 6, 911–918. doi:10.1109/tsg.2014.2364233
- Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., et al. (2011). Detecting novel associations in large data sets. *science* 334, 1518–1524. doi:10.1126/science.1205438
- Roy, A. D., and Yeafi, A. (2022). “Implementation of encoder-decoder based long short-term memory network for short-term electrical load forecasting,” in 2022 4th International Conference on Sustainable Technologies for Industry 4.0 (STI) (IEEE), Dhaka, Bangladesh, Dec. 17 2022 to Dec. 18 2022, 1–6.
- Saber, A. Y., and Alam, A. R. (2017). “Short term load forecasting using multiple linear regression for big data,” in 2017 IEEE symposium series on computational intelligence (SSCI) (IEEE), Honolulu, Hawaii, USA, 27 November - 1 December 2017, 1–6.
- Samadianfard, S., Hashemi, S., Kargar, K., Izadyar, M., Mostafaeipour, A., Mosavi, A., et al. (2020). Wind speed prediction using a hybrid model of the multi-layer perceptron and whale optimization algorithm. *Energy Rep.* 6, 1147–1159. doi:10.1016/j.egyrs.2020.05.001
- Shi, J., Zhang, W., Bao, Y., Gao, D. W., and Wang, Z. (2023a). Load forecasting of electric vehicle charging stations: attention based spatiotemporal multi-graph convolutional networks. *IEEE Trans. Smart Grid* 15, 3016–3027. doi:10.1109/tsg.2023.3321116
- Shi, J., Geng, L., Zhang, M., Xu, D., and Li, H. (2023b). “A short-term regional net load prediction method based on parallel fragment attention-bi- lstm,” in 2023 IEEE 6th International Conference on Automation, Electronics and Electrical Engineering (AUTEEE) (IEEE), Shenyang, China, December 15, 2023 - December 17, 2023, 624–629.
- Shi, W., Li, Y., Dong, S., Lu, X., Ye, H., and Hu, B. (2023c). “Short term load forecasting for holidays based on exponential smoothing of correlative correction,” in 2023 International Conference on Power System Technology (PowerCon) (IEEE), Kuala Lumpur, Malaysia, 12–14 September 2022, 1–4.
- Shuman, D. I., Narang, S. K., Frossard, P., Ortega, A., and Vandergheynst, P. (2013). The emerging field of signal processing on graphs: extending high-dimensional data analysis to networks and other irregular domains. *IEEE signal Process. Mag.* 30, 83–98. doi:10.1109/msp.2012.2235192
- Sun, S., Du, Z., Jin, K., Li, H., and Wang, S. (2023). Spatiotemporal wind power forecasting approach based on multi-factor extraction method and an indirect strategy. *Appl. Energy* 350, 121749. doi:10.1016/j.apenergy.2023.121749
- Sun, S., Li, M., Wang, S., and Zhang, C. (2022). Multi-step ahead tourism demand forecasting: the perspective of the learning using privileged information paradigm. *Expert Syst. Appl.* 210, 118502. doi:10.1016/j.eswa.2022.118502
- Tang, Z., Zhao, G., and Ouyang, T. (2021). Two-phase deep learning model for short-term wind direction forecasting. *Renew. Energy* 173, 1005–1016. doi:10.1016/j.renene.2021.04.041
- Wang, Y., Rui, L., Ma, J., and Jin, Q. (2023). A short-term residential load forecasting scheme based on the multiple correlation-temporal graph neural networks. *Appl. Soft Comput.* 146, 110629. doi:10.1016/j.asoc.2023.110629
- Wei, C., Pi, D., Ping, M., and Zhang, H. (2023). Short-term load forecasting using spatial-temporal embedding graph neural network. *Electr. Power Syst. Res.* 225, 109873. doi:10.1016/j.epsr.2023.109873
- Wu, Z., Pan, S., Long, G., Jiang, J., and Zhang, C. (2019). *Graph wavenet for deep spatial-temporal graph modeling*. arXiv preprint arXiv:1906.00121.

Xiu, C., Zhan, S., Pan, J., Peng, Q., Lin, Z., and Wong, S. (2024). Correlation-based feature selection and parallel spatiotemporal networks for efficient passenger flow forecasting in metro systems. *Transp. A Transp. Sci.*, 1–37. doi:10.1080/23249935.2024.2335244

Yan, B., Wang, G., Yu, J., Jin, X., and Zhang, H. (2021). Spatial-temporal Chebyshev graph neural network for traffic flow prediction in iot-based its. *IEEE Internet Things J.* 9, 9266–9279. doi:10.1109/jiot.2021.3105446

Yanmei, J., Mingsheng, L., Yangyang, L., Yaping, L., Jingyun, Z., Yifeng, L., et al. (2024). Enhanced neighborhood node graph neural networks for load forecasting in smart grid. *Int. J. Mach. Learn. Cybern.* 15, 129–148. doi:10.1007/s13042-023-01796-8

Yu, B., Yin, H., and Zhu, Z. (2017). *Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting*. arXiv preprint arXiv:1709.04875.

Zhang, J., Liu, D., Li, Z., Han, X., Liu, H., Dong, C., et al. (2021a). Power prediction of a wind farm cluster based on spatiotemporal correlations. *Appl. Energy* 302, 117568. doi:10.1016/j.apenergy.2021.117568

Zhang, L., Xie, D., Luo, G., Qian, G., Song, M., and Chen, S. (2021b). “Research on short-term load forecasting based graph computation in power supply areas,” in 2021 IEEE 11th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER) (IEEE), Jiaxian, China, July 27–31, 2021, 638–642.

Zheng, G., Chai, W. K., Zhang, J., and Katos, V. (2023). Vdgcnet: a novel network-wide virtual dynamic graph convolution neural network and transformer-based traffic prediction model. *Knowledge-Based Syst.* 275, 110676. doi:10.1016/j.knosys.2023.110676