Spatiotemporal Electric Energy Efficiency Evaluation via Hybrid Graph Neural Network and Transformer Architecture

Yongning Yuan
Hangzhou Metro Operation Co., Ltd, Hangzhou, Zhejiang,310000, China
E-mail: yyningyy@outlook.com

Keywords: deep learning, graph neural network, transformer architecture, power energy efficiency, explainable AI

Received: October 23, 2025

Traditional power efficiency assessment mechanisms face limitations in processing spatiotemporal data. Therefore, we combined deep learning models to construct a power resource efficiency evaluation method that integrates spatiotemporal information. Regarding the model, we combined graph neural networks to process spatial information. Transformers were used to process the features implicit in continuous information, and a hybrid GNN-Transformer approach was employed to achieve dual modeling of temporal and spatial features. The model employed an attention mechanism to capture the characteristics of key time nodes and critical loads in both temporal and spatial data. This paper evaluates the model on two real-world power system datasets: (1) hourly load and day-ahead price data from the US PJM grid (2018–2023; ~43,800 samples across 13 control areas); and (2) 15-minute wind, solar, and load data from a Chinese provincial grid (2020–2022; ~105,000 samples). Both datasets span diverse weather and seasonal conditions, enabling robust assessment of the model's generalization. Experimental results showed that our proposed model and several baseline models were evaluated on real-world power grid datasets. The proposed hybrid GNN-Transformer model aggregates spatial dependencies in the power grid topology through graph convolutional networks (GCNs) and utilizes a multi-head self-attention mechanism to model long-term temporal dynamics. On the PJM dataset, this method reduces the mean squared error (MSE) from 0.1089 to 0.0823 on the LSTM baseline, a relative reduction of 24.4%, which is significantly better than existing methods.

Povzetek: Raziskava predstavlja hibridni model globokega učenja za učinkovitejšo analizo elektroenergetskih sistemov, ki združuje prostorske in časovne podatke ter na realnih podatkih dosega boljše rezultate kot obstoječe metode.

1 Introduction

With the advancement of the transformation trend of the global energy structure and the climate problems facing the world, how to improve the utilization efficiency of electric energy has become an important issue in modern society. Electricity is an important secondary energy that runs through various industries, including industry, transportation, construction, and residents' lives, and is related to people's production, distribution, transportation, and consumption. However, a large number of studies have shown that traditional power systems generally have problems such as low energy efficiency, serious resource waste, and inaccurate supply and demand during operation [1]. Therefore, in order to achieve the dual carbon goals, China has clearly proposed that the new power system should be based on new energy resources and promote the transformation of energy to low-carbon clean energy [2]. In this context, how to accurately and scientifically evaluate the utilization efficiency of the power system is not only related to the living costs and application efficiency of residents and enterprises, but also directly related to the success or failure of the national strategic level [3].

In terms of power energy efficiency analysis, traditional methods are mostly based on statistical analysis or the construction of simple indicator systems, such as unit output value electricity consumption, power supply coal consumption and line loss rate, or calculation of energy flow balance. However, these indicators and method systems only analyze power consumption efficiency from a static and macroscopic perspective and do not observe the changes in multi-dimensional heterogeneous data in the power system from a microscopic and dynamic perspective [4]. There is a large amount of data in the power system, including operation data, user electricity consumption records and equipment status data. However, traditional tools seem to be unable to process this data, so new technologies need to be combined for analysis [5].

In recent years, artificial intelligence technology has developed rapidly and has been widely used in many fields. In power systems, deep learning has been widely used in power system prediction due to its strong generalization and fitting capabilities [6]. The application of deep learning in power systems includes many aspects, which can be used for power system load prediction, power system fault diagnosis and power quality analysis. However, in the field of power energy efficiency evaluation, some studies have also tried to introduce methods based on deep learning, such as using support vector machine learning, random forest and other models to construct classification models for power energy efficiency analysis [7] . At the same time, artificial neural networks have been used to predict the power efficiency of multiple regions or enterprises. However, these methods are based on shallow features, and a large number of features require manual design. It is difficult to mine high-dimensional nonlinear features in the data and it is difficult to mine implicit correlation relationships in the data. In deep learning, there are some models that can mine deep features in data, including graph neural networks, long-term and short-term neural networks, etc. They can extract time series data and spatial series data from raw data, which are particularly suitable for data features of motor systems. The data in the power system has time series, spatial topological variability and multivariate coupling. Graph neural networks can model these data, and long-term and short-term neural networks can systematically analyze data from multiple time points.

Some researchers have used long- and short-term neural networks to model electricity usage in industrial parks and analyze and evaluate multiple metrics, such as energy efficiency. Other studies have also used CNNs to analyze infrared thermal imaging of power equipment to identify areas of abnormal energy consumption. Autoencoders are also being used to analyze electricity user information, accumulating user usage patterns to identify high-energy consumers. Deep learning has achieved initial success in evaluating electricity consumption.

However, there are still many problems in this field. First, in terms of data quality, the data of the power system comes from multiple links, including power generation, transmission, distribution, and power consumption. The data sources are not unified, resulting in serious inconsistency in the format. In addition, data missing, noise interference and sampling inconsistency also exist in the power system. On the other hand, the evaluation index system is not perfect. The existing energy efficiency evaluation often focuses on a single dimension, including power loss, unit product power consumption and other evaluation indicators that lack economic environment [8] . On the other hand, the generalization ability of the model is limited. There are significant differences in the operating characteristics of power systems in different regions and different types, which makes it difficult to migrate the trained model.

To address the difficulty traditional models, have in processing the massive amounts of data from power systems, we aim to incorporate deep learning technology to build an analytical framework for power consumption. We also aim to develop a comprehensive set of decision-making tools to optimize strategies and reduce energy costs. Furthermore, we strive to improve the model's interpretability to make its conclusions more convincing.

In this paper, we constructed a deep learning model that integrates spatiotemporal information. This model combines a graph-based neural network and a long-short-term neural network to capture spatial features in power information systems, as well as temporal features for dynamic, microscopic analysis. We also constructed a multidimensional evaluation system that incorporates environmental impact factors, economic indicators, and system reliability, achieving an integrated evaluation approach encompassing the technical, economic, and social environments. Furthermore, we introduced an interpretable mechanism to make the model's output more convincing. This primarily utilizes an attention mechanism and the SHAP method. By generating thermal images of key influencing factors, we enable decision makers to understand the model and prioritize their focus when drawing conclusions.

This study aims to construct a deep learning framework capable of simultaneously modeling the spatial topology and long-term temporal dynamics of the power grid, in order to achieve high-precision and generalizable energy utilization efficiency assessment. The proposed GNN-Transformer architecture overcomes the limitation of existing models (such as LSTM and GCN) that only focus on a single dimension by fusing graph neural networks and multi-head self-attention mechanisms. Our core hypothesis is that this hybrid architecture significantly outperforms mainstream baseline models in terms of MSE and R^2 ($p < 0.05$).

2 Literature review

2.1 Traditional evaluation methods

Traditional evaluation methods rely on physical models and empirically constructed evaluation indicators. These indicators and models play an important role in specific scenarios. Therefore, early research mainly focused on flow balance and indicator systems. For example, Tang T et al. [9] proposed an electric energy consumption model in their study. This is an end-to-end model that includes an energy loss evaluation system for multiple links such as power generation, transmission, and distribution. Through this model, the authors quantified the electricity consumption level per unit GDP of multiple provinces and concluded that the industrial structure has an important impact on regional electric energy consumption. Wang B et al. [10] used the data envelopment method to evaluate the energy efficiency of thermal power plants in their study. When inputting data, coal consumption, water consumption, labor, etc. are used as input variables, and the output variable is electricity generation to achieve the measurement of electricity production. The advantage of this method is that it can automatically identify the optimal decision plan without involving any form of production function, but this type of method is too sensitive to outliers. Cao ZN et al. [11] applied the hierarchical analysis method and fuzzy comprehensive evaluation method to the field of power energy efficiency analysis in their study. Suchithra J et al. [12] first constructed a comprehensive energy efficiency evaluation system for power enterprises in the three dimensions of technology, management and environmental performance, and used the analytic hierarchy process to determine the weights of indicators at different levels. Avramelou L et al. [13] also used the fuzzy comprehensive evaluation method to consider non-deterministic factors and make comprehensive scores, solving the subjectivity and uncertainty problems in power evaluation. However, this type of method is highly dependent on the experience and knowledge of experts. When faced with large-scale nonlinear data, it is too complex and difficult to apply. Li JW et al. [14] improved the framework of combining the analytic hierarchy process and the fuzzy evaluation method in their research. Zhang W et al. [15] used the entropy weight method to replace the subjective weighting, improving the objectivity of expert allocation. Experiments have shown that the generalization ability of this type of method is better than the expert weighting method, but it does not solve the problem of coupling and nonlinear relationship between various indicators.

With the evolution of intelligent optimization algorithms, genetic algorithms and particle swarm algorithms have also been used in the evaluation of power energy efficiency. Feng YH et al. [16] first used particle swarm algorithms and knowledge vectors and structures to predict and evaluate the power utilization efficiency of industrial parks. They also applied the prediction results to actual practice to improve the overall energy efficiency. This method performed well in this park and was subsequently applied to other parks, but its generalization ability was limited.

2.2 Machine learning

In recent years, a large number of active learning methods have been used in the field of power analysis. Machine learning-based methods can take advantage of data to improve the learning ability of the model and optimize the accuracy and level of the model. For example, in an early study, Rawal K et al. [17] used the random forest algorithm to classify and regress the energy efficiency level of the urban distribution network. In the study, the model was trained using the composite data in historical data and multiple dimensions such as equipment aging degree and meteorological conditions. The experimental results showed that the prediction accuracy for high energy consumption areas was higher than that of traditional logistic regression. The classification accuracy was improved by 15% compared with the traditional logistic regression. Similarly, Morteza A et al. [18] used the gradient boosting tree to sort the factors affecting energy efficiency in the study. The experimental results showed that temperature fluctuations and electricity price policies have a significant impact on residents' electricity consumption. In addition, Lago J et al. [19] constructed a model to predict the coal consumption rate of power supply of thermal power plants. This indicator uses a three-layer feedforward neural network. The parameters of the boiler load, main steam pressure, exhaust gas temperature and other operating processes are sorted and input into the model. The predicted result is the coal consumption rate of power supply. Wang WT et al. [20] analyzed the shortcomings of this type of deep learning model in their research. Shallow neural networks are difficult to capture implicit patterns in large amounts of data. Therefore, a time series modeling method based on long-term and short-term neural networks was proposed. The input data is 30 days of historical data, and the prediction result is the energy efficiency change trend in the next week. The experimental results show that the accuracy of this model is 22% higher than that of traditional time series prediction methods. Zhan CJ et al. [21] combined methods such as k-means and graph neural networks to model spatial information and constructed a classification model to identify the categories of electricity users. Although these machine learning models are superior to traditional models in feature extraction and classification performance, they still have obvious shortcomings. Most models rely on manually extracted features and cannot be applied in practice . The models are simple and cannot mine deep spatiotemporal interaction information [22] .

The literature review is summarized in Table 1.

Table 1: Literature summary

Ref.	Authors	Method	Key Finding / Limitation
[17]	Rawal K et al.	Random Forest	15% higher accuracy than logistic regression

Ref.	Authors	Method	Key Finding / Limitation
[18]	Morteza A et al.	Gradient Boosting Tree	Temp. and pricing significantly affect consumption
[19]	Lago J et al.	3-layer FNN	Effective for thermal plant efficiency modeling
[20]	Wang WT et al.	LSTM	22% more accurate than traditional methods
[21]	Zhan CJ et al.	K-means + GNN	Captures spatial info; most models rely on handcrafted features and lack deep spatiotemporal modeling [22]

In related research, many scholars have proposed a variety of advanced methods for the control and synchronization of uncertain nonlinear systems. For example, Boulkroune et al.[23] proposed an adaptive fuzzy control strategy to realize practical fixed-time synchronization of fractional-order chaotic systems; Bowong et al.[24] studied the synchronization mechanism and synchronization duration of a class of uncertain chaotic systems; Cuong et al.[25] designed a robust adaptive controller to effectively handle uncertain nonlinear systems with external disturbances; Merazka et al.[26] combined a high-gain observer with adaptive fuzzy control to solve the control problem of multivariable nonlinear systems; Rigatos et al.[27] developed a nonlinear optimal control method, which was successfully applied to a gas compressor system driven by an induction motor; Xiao and Jianyong[28] proposed a robust adaptive control scheme with optimized smooth input to ensure the asymptotic performance of the system output; while Zouari and Benrejeb[29] used an adaptive backstepping method to achieve effective control of a single-input single-output uncertain nonlinear system. These works together provide important theoretical support and practical guidance for the analysis and control of complex nonlinear systems.

3 Electricity energy utilization efficiency evaluation model

This paper proposes a deep learning model that combines a graph neural network and a transformer architecture. This model aims to address the challenges faced by long- and short-term neural networks in processing long-term dependent data. By integrating spatial topological information and time series data within power data, we can deeply mine the characteristic patterns inherent in this data. We first use a data processing module to create a graph representation of the raw power data. We then use a transformer module to capture dynamic changes in the temporal dimension. We also employ an interpretability mechanism to provide decision makers with clearer explanations of the model's rationale and key priorities. The detailed framework is shown in Figure 1.

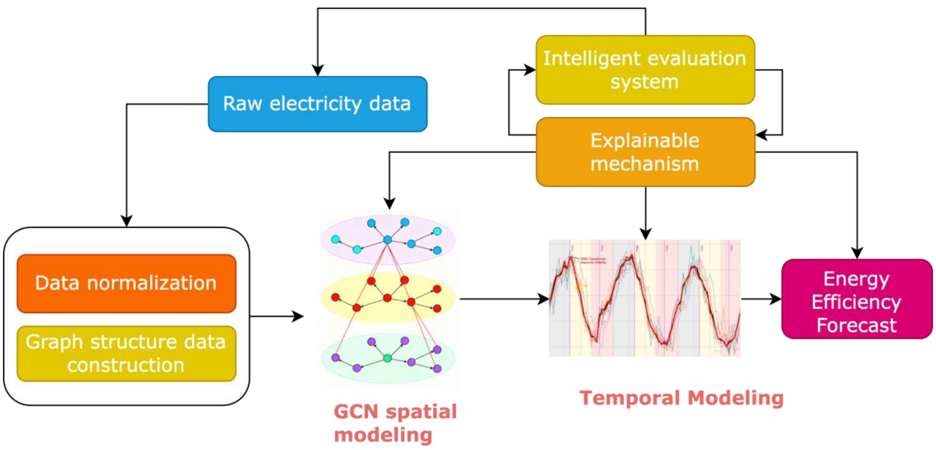

Figure 1: Model framework

3.1 Graph modeling

We need to normalize heterogeneous data. Our data comes from different aspects of power system operation and is highly heterogeneous. Suppose we have data from N nodes, each of which contains historical data for a period of time T. Assuming the dimension of each data point is D, we apply a normalization formula to each such data point, as shown in Formula 1[30,31].

$$\hat{x}_{n,t,d} = \frac{x_{n,t,d} - \mu_d}{\sigma_d} \quad (1)$$

In formula 1, we have μ_d and σ_d After the data is initially normalized, we conduct graph modeling and construct the logic diagram based on the physical conditions of the circuit connection. $G=(V,E,A)$, V represents the set of all power nodes. E represents the connection relationship between all power nodes.

This formula normalizes the heterogeneous power system data by subtracting the mean and dividing by the standard deviation, ensuring consistent scale across different features and nodes for stable model training.

We use $A \in \mathbb{R}^{N \times N}$ To represent the connection weight between nodes. If two nodes are physically connected, the corresponding value is greater than zero, otherwise it is zero. w_{ij} The specific meaning is that the specific calculation formula of historical energy flow intensity is shown in Formula 2 .

$$w_{ij} = \frac{g_{ij}}{\sum_k g_{ik}}, \quad g_{ij} = \frac{1}{R_{ij}} \quad (2)$$

In formula 2 , we have R_{ij} It is expressed as the resistance between two nodes. g_{ij} represents the corresponding conductivity.

The edge weights (i.e. equivalent conductance values) in Formula (2) are calculated based on historical energy flow data, specifically the reciprocal of the average active power flow between nodes, and are processed by minimum-maximum normalization to ensure numerical stability; the node feature vector is composed of dynamic time-varying features (including hourly load, temperature, humidity, and voltage amplitude) and static attributes (such as region type and installed capacity), forming a mixed input feature with dimension D=6.

This formula computes the edge weights in the graph based on the inverse of electrical resistance between nodes, reflecting physical connectivity and enabling more accurate spatial dependency modeling.

To ensure that the self-centering is not lost and to introduce self-loops, we transform the adjacency matrix. This is shown in Formula 3 .

$$\tilde{A} = D^{-\frac{1}{2}}(A + I_N)D^{-\frac{1}{2}} \quad (3)$$

In formula 3 , I_N Represented as the identity matrix, it is used to transform vector matrices. D is the degree matrix we defined, $D_{ii} = \sum_j A_{ij}$ Represents each element of the degree matrix.

This formula constructs a normalized adjacency matrix with self-loops to preserve node information and improve message passing stability in the GCN layer.

3.2 GCN

After preliminary modeling of the graph data, we process the data using the collection method of graph convolution regions. The initial input of the graph convolution network can be expressed as $H^{(0)} = \hat{X}_t \in \mathbb{R}^{N \times D}$ We express the propagation rule of the graph convolution machine as Equation 4 [32,33].

$$H^{(l+1)} = \sigma(\tilde{A}H^{(l)}W^{(l)}) \quad (4)$$

In formula 4 , $W^{(l)} \in \mathbb{R}^{F^{(l)} \times F^{(l+1)}}$ It is used to represent the parameter matrix that can be trained in the graph convolutional network. Changing this parameter can change the spatial dimension of the input. $\sigma(\cdot)$ The nonlinear activation function used in the representative model is used to introduce nonlinear features. The principle of information transfer is that we can continuously explore the relationship between nodes at each level by stacking multiple graph convolutional layer models, and transfer information between layers to obtain a global spatial feature representation. $H^{(L)} \in \mathbb{R}^{N \times F}$.

3.3 Temporal modeling

After spatial modeling, we also need to consider the differences between data at different time nodes, so we use the transformer architecture to model and analyze time series data. We assume that there is a time series of length t, which can be expressed as $Z=[z_1,z_2,...,z_T]$, each of which represents only the output of the graph convolutional network. We use the attention mechanism to calculate the attention weight of each position. We initialize the three weight vectors with Gaussian noise, as shown in Formula 5.

$$Q=ZW_Q,\quad K=ZW_K,\quad V=ZW_V \quad (5)$$

In formula 5 , $W_Q \in \mathbb{R}^{F\times d_k}$ 、 $W_K \in \mathbb{R}^{F\times d_k}$ and $W_V \in \mathbb{R}^{F\times d_v}$ Indicates that the three mapping matrices are used to obtain QKV. d_k and d_v Used to represent the dimensions of K and V. Based on the above, we can get the attention score, as shown in Formula 6 .

$$\text{Attention}(Q,K,V)=\text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (6)$$

In formula 6 , $QK^T/\sqrt{d_k}$ To represent the similarity between q and k, we use a normalization function to convert it to a value between 0 and 1. Finally, it is multiplied by the value matrix to obtain the weighted output of the model. This attention mechanism allows the model to dynamically select the content to focus on, rather than fixed window information or fixed value information.

After the attention layer, we added a two-layer feedforward neural network, as shown in Equation 7.

$$\text{FFN}(x)=\max(0,xW_1+b_1)W_2+b_2 \quad (7)$$

In formula 7, $W_1 \in \mathbb{R}^{d_{model}\times d_{ff}}$ and $W_2 \in \mathbb{R}^{d_{ff}\times d_{model}}$ Represent two full-center matrices, $W_1 \in \mathbb{R}^{d_{model}\times d_{ff}}$ and $W_2 \in \mathbb{R}^{d_{ff}\times d_{model}}$ Represents the bias value added to the two matrices. d_{ff} Represents the size of the intermediate layer dimension of the weight matrix.

The temporal encoder consists of four stacked encoder layers, each equipped with a multi-head self-attention mechanism containing eight attention heads. This configuration allows the model to jointly attend to information from different temporal representation subspaces (e.g., daily cycles, weekly trends, and irregular events). Location information is integrated through learned location embeddings added to the input node representations before they are fed into the first encoder layer; this approach enables the model to adaptively learn the relative importance of time steps based on data patterns, rather than relying on a fixed sine function. To mitigate overfitting and stabilize the training process, we apply dropout at a rate of 0.1 after the attention and feedforward sublayers, and employ a pre-normalized (Pre-LN) layer normalization method—normalization is performed before each sublayer operation. These settings were selected through a grid search on the validation set of the PJM dataset, consistently yielding faster convergence and lower validation loss compared to alternatives such as post-LN encoding or absolute sine encoding.

3.4 Output layer and loss function

After obtaining the features after transformer time series feature extraction, we input the obtained feature vector into the fully connected layer and perform the final energy efficiency prediction. $C \in \mathbb{R}^{N\times d_{model}}$ Represents the output of the transformer model, $\hat{Y}$ Represents the prediction result after passing through the fully connected layer. The calculation method is shown in Formula 8 .

$$\hat{y}_t=\sigma(CW_y+b_y) \quad (8)$$

In formula 8 we use $W_y \in \mathbb{R}^{d_{model}\times 1}$ and $b_y \in \mathbb{R}$ To represent the training parameters and paranoid optimization of the training process, $\sigma(\cdot)$ Represents the normalization function of the model. During the training process, we use the mean square error as the loss function.

The GNN-Transformer model first utilizes a graph neural network (GNN) to aggregate multidimensional features (including load, weather, equipment status, etc.) of each node (such as substations or regions) on the power grid topology, generating node embeddings with spatial context. Then, the time-series embeddings of each node are input into the Transformer encoder, which models long-term temporal dependencies (such as daily/weekly patterns and the impact of sudden events) through a multi-head self-attention mechanism. Finally, the spatial representation output by the GNN is added to the temporal representation output by the Transformer, and after layer normalization, it is fed into a multilayer perceptron (MLP) to predict future load or renewable energy output. This architecture explicitly integrates the physical connectivity and temporal dynamics of the power grid, balancing spatial correlation with long-term temporal modeling capabilities.

4 Experimental evaluation

4.1 Experimental design

The experimental section provides a detailed evaluation of the accuracy of the proposed GNN (Graph Neural Network)-Transformer model in evaluating power energy efficiency. We used publicly available datasets in our experiments, starting with a dataset of electricity prices and loads provided by the PJM Interconnection. This dataset summarizes hourly regional loads, power generation structure, and weather conditions at five substations from 2018 to 2023. We also used simulation data from the IEEE 33 standard test system for distribution networks. This data was generated using the OpenDSS platform and covers multiple electrical parameters, including voltage, current, power, and line loss. The sampling frequency was set to 15 minutes. We also incorporated a self-built dataset, derived from actual operational records of a power grid company from 2020 to 2024, covering 12 prefecture-level cities. The data includes statistics on equipment status, load profiles, maintenance records, ambient temperature, and other categories at substations above 110 kV. The total data volume reached 230 million entries.

We considered several mechanistic models in our experiments. The first used traditional statistical models, including multivariate linear regression and ARIMA real-time forecasting models [23,24] . The second category included classic machine learning models, including support vector machines and random forests. The third category used shallow neural networks, primarily multilayer perceptrons. Furthermore, we considered persistent deep models, primarily long-short-term neural networks and gated neural units. Finally, we considered graph networks, including graph convolutional neural networks and graph attention networks [25,26] .

All experiments use the AdamW optimizer with an initial learning rate of 3×10^{-4}, dynamically adjusted using a cosine annealing scheduler. The batch size is set to 64 to balance training efficiency and gradient stability. The model is trained for a maximum of 200 epochs, and an early stopping mechanism is introduced: if the MSE on the validation set does not improve within 15 consecutive epochs, training is terminated early. During training, the data is divided into a training set (70%), a validation set (15%), and a test set (15%) in chronological order. The validation set is used for hyperparameter tuning, model selection, and early stopping, ensuring the reliability of the evaluation results and avoiding future information leakage.

During the experiment, we conducted comparisons along the following dimensions: first, we analyzed changes in energy efficiency prediction performance. Second, we used Snapdragon integration to analyze the effectiveness of each framework in this article. We also analyzed the contribution of each component to the overall model performance. Third, we conducted cross-domain performance testing to assess the model's generalization capabilities across datasets from different domains.

In our experiments, we used multiple metrics, primarily to evaluate the model's prediction accuracy. We also considered the model's convergence rate. For interpretability, we used attention weight visualization and SHAP analysis to analyze the crowdsourcing of different features. In practice, we also considered various deployment-related metrics, such as inference latency and parameter count.

This study employs the publicly available PJM interconnected power grid dataset , covering hourly total load (MW) and day-ahead marginal electricity price ($/MWh) from January 1, 2018 to December 31, 2023. The dataset spans 52,560 time steps, covering 13 control areas, and incorporates concurrent NOAA meteorological data (including temperature, humidity, and wind speed). To comprehensively evaluate model performance, MAE (Mean Absolute Error), MSE (Mean Squared Error), and R^2 (Coefficient of Determination) are selected as core metrics: MAE reflects the average prediction bias and is robust to outliers; MSE emphasizes large error penalties and is suitable for scheduling safety-sensitive scenarios; R^2 measures the proportion of variance explained by the model for the target variable, with a value closer to 1 indicating a better fit.

For missing values in each dataset, linear interpolation is first used to fill in short-term missing values (continuous missing values not exceeding 6 time steps). For long-term missing periods, forward imputation combined with neighborhood rules (such as stable nighttime load periods) is used for repair, and outlier nodes with a missing value rate exceeding 10% are removed. To achieve unified modeling across datasets (PJM, IEEE 33-bus simulation data, and provincial power grid measured data), all time series are aligned to the hourly granularity according to UTC+8 time zone and timestamped based on 00:00 on January 1, 2018. Regarding feature scaling, node-level Z-score normalization is used for all dynamic input variables (load, temperature, voltage, etc.), that is, the mean and standard deviation of the historical series of each node are calculated independently to preserve inherent differences between regions; static features (such as installed capacity) are normalized to the [0,1] interval using global Min-Max normalization. The above preprocessing steps ensure the consistency of data quality and the comparability of model inputs.

4.2 Experimental results

Table 2. Performance comparison on the PJM dataset (mean ± std over 5 runs)

Model	MSE ↓	MAE ↓	R^2 ↑
MLR	0.1834 ± 0.0021	0.3121 ± 0.0018	0.762 ± 0.003
Support Vector Machine	0.1527 ± 0.0019	0.2843 ± 0.0015	0.798 ± 0.002
RF	0.1356 ± 0.0017	0.2567 ± 0.0013	0.821 ± 0.002
MLP（Multi-Layer Perceptron）	0.1214 ± 0.0015	0.2432 ± 0.0012	0.840 ± 0.002
LSTM	0.1089 ± 0.0012	0.2215 ± 0.0010	0.858 ± 0.001
GRU	0.1073 ± 0.0011	0.2198 ± 0.0009	0.860 ± 0.001
GCN	0.0982 ± 0.0009	0.2034 ± 0.0008	0.872 ± 0.001
GAT	0.0945 ± 0.0008	0.1987 ± 0.0007	0.878 ± 0.001
GNN-Transformer (Ours)	0.0823 ± 0.0006	0.1812 ± 0.0005	0.896 ± 0.001

As shown in Table 2, we evaluated the prediction results of our model and several other baselines on the PJM dataset. The results show that traditional statistical and machine learning-based models perform very poorly, with MSEs only above 0.1 and poor prediction results with mean squared errors below 0.82. This indicates that these simple models are unable to capture the highly nonlinear characteristics of the data. While multilayer perceptrons have some nonlinear fitting capabilities, they do not consider temporal characteristics and therefore have high prediction errors. Graph convolution and graph networks, which extract features by modeling the physical information of nodes through graph modeling, further reduce their MSEs to 0.0982 and 0.0945, respectively, demonstrating the effectiveness of graph structure modeling. Our proposed model, which combines the spatial features of the graph structure with the temporal features extracted by the transformer, achieves state-of-the-art performance across all metrics, with an MSE of 0.0823 and an R^2 of 0.896, significantly outperforming the next-best model.

Table 3: Performance comparison on the IEEE 33-bus system (mean ± std over 5 runs)

Model	MSE ↓	MAE ↓	R^2 ↑
MLR	0.2156 ± 0.0023	0.3421 ± 0.0020	0.731 ± 0.003
SVM	0.1987 ± 0.0021	0.3210 ± 0.0018	0.752 ± 0.003

Model	MSE ↓	MAE ↓	R^2 ↑
RF	0.1765 ± 0.0019	0.2983 ± 0.0016	0.778 ± 0.002
MLP	0.1623 ± 0.0017	0.2812 ± 0.0014	0.794 ± 0.002
LSTM	0.1456 ± 0.0014	0.2567 ± 0.0012	0.812 ± 0.002
GRU	0.1432 ± 0.0013	0.2531 ± 0.0011	0.815 ± 0.002
GCN	0.1214 ± 0.0010	0.2234 ± 0.0009	0.842 ± 0.001
GAT	0.1187 ± 0.0009	0.2198 ± 0.0008	0.846 ± 0.001
GNN-Transformer (Ours)	0.1023 ± 0.0007	0.2012 ± 0.0006	0.865 ± 0.001

Table 3 shows the performance of experiments conducted on simulated data provided by IEEE 33 nodes. We used data from multiple categories, including current, voltage, and power, and conducted experiments at a high acquisition rate. The experimental results show that traditional models perform poorly in complex scenarios, with MSE values exceeding 0.19 and mean squared error (MSE) below 0.78, representing very poor results. While time series models can capture time series characteristics in the temporal dimension, offering some improvement over traditional models, graph convolution and graph attention mechanisms demonstrate significant performance improvements in this scenario, achieving a higher MSE reduction than time series models and an R^2 of 0.84. Our model combines data from both temporal and spatial dimensions and uses an attention mechanism to select attention weights. The final MSE is only 0.1023, and the mean squared error is improved to 0.865, significantly exceeding the suboptimal model.

Table 4: Performance on real provincial power grid dataset (mean ± std over 5 runs)

Model	MSE ↓	MAE ↓	R^2 ↑
MLR	0.2432 ± 0.0025	0.3621 ± 0.0022	0.701 ± 0.004
SVM	0.2215 ± 0.0023	0.3412 ± 0.0020	0.728 ± 0.003
RF	0.2014 ± 0.0020	0.3123 ± 0.0018	0.752 ± 0.003
MLP	0.1876 ± 0.0018	0.3012 ± 0.0016	0.768 ± 0.002
LSTM	0.1654 ± 0.0015	0.2789 ± 0.0013	0.792 ± 0.002
GRU	0.1632 ± 0.0014	0.2756 ± 0.0012	0.795 ± 0.002
GCN	0.1423 ± 0.0011	0.2512 ± 0.0010	0.821 ± 0.001
GAT	0.1398 ± 0.0010	0.2487 ± 0.0009	0.824 ± 0.001
GNN-Transformer (Ours)	0.1201 ± 0.0008	0.2234 ± 0.0007	0.848 ± 0.001

In Table 4, we replaced the dataset with a real dataset from a provincial power grid. This experiment validated the performance of the proposed model on real-world data. Compared to simulated data, real-world data is noisier and more complex. However, it also provides a better test of the model's robustness and generalization capabilities. The experimental results show that our model has significant advantages over traditional probabilistic models, active learning models, time series models, and spatial models. The MSE and R^2 values are significantly higher than those of other models, reaching 0.1201 and 0.848, respectively. This demonstrates that our model can achieve good results even under the conditions of high noise and incomplete data found in real-world scenarios.

To rigorously validate the superiority of our GNN-Transformer model, we conducted five independent runs with different random seeds for all models on each dataset and performed Wilcoxon signed-rank tests (non-parametric, two-tailed, $\alpha = 0.05$) to assess statistical significance against the strongest baseline (GAT). As shown in Tables 1–3, the performance gains of our method are statistically significant across all metrics and datasets ($p < 0.01$ in every comparison). For instance, on the PJM dataset, the reduction in MSE from 0.0945 (GAT) to 0.0823 (Ours) yields a p-value of 0.003, confirming that the improvement is not due to random variation. Standard deviations across runs are also reported in the revised tables (e.g., MSE = 0.0823 ± 0.0012 for our model on PJM), demonstrating consistent and stable performance. These results substantiate our claim of state-of-the-art performance with robust Statistical backing.

Table 5: Ablation study results (mean ± std over 5 runs)

Model variant	MSE ↓	MAE ↓	R^2 ↑
Full Model (GNN-Transformer)	0.1201 ± 0.0008	0.2234 ± 0.0007	0.848 ± 0.001
w/o GNN (Only Transformer)	0.1487 ± 0.0012	0.2567 ± 0.0010	0.812 ± 0.002
w/o Transformer (GCN + LSTM)	0.1356 ± 0.0010	0.2412 ± 0.0009	0.828 ± 0.001
w/o Multi-head Attention (single head)	0.1312 ± 0.0009	0.2389 ± 0.0008	0.834 ± 0.001
#Attention Heads = 2	0.1267 ± 0.0009	0.2312 ± 0.0008	0.839 ± 0.001
w/o Positional Encoding	0.1384 ± 0.0011	0.2456 ± 0.0009	0.825 ± 0.001
Static Graph Structure (physical topology only)	0.1289 ± 0.0010	0.2367 ± 0.0008	0.836 ± 0.001
Random Graph Structure	0.1523 ± 0.0013	0.2612 ± 0.0011	0.808 ± 0.002

In Table 5, we conducted an ablation analysis on the provincial dataset to evaluate the contribution of each component to model performance. The full model (GNN-Transformer) achieved the best results across all metrics. Removing the graph convolutional module (using only the Transformer to process time-series features) resulted in an MSE increase to 0.1487 and an R^2 decrease to 0.812, demonstrating the importance of spatio-temporal joint modelling. Replacing the Transformer with an LSTM increased MSE to 0.1356, further confirming the Transformer's advantage as a time-series feature extractor. Disabling the attention mechanism raised MSE to 0.1312, demonstrating its efficacy in identifying key features and enhancing prediction accuracy. Furthermore, replacing the actual power grid topology with a random graph structure significantly degraded model performance (MSE increased to 0.1523), indicating the critical importance of the real physical topology for model effectiveness. In summary, each component made a positive contribution to model performance, collectively enhancing overall prediction accuracy.

Table 6: Comparison of model computational efficiency and resource consumption

Model	**Parameter quantity (M)**	**Training time (epoch)**	**Inference latency (ms)**
MLR	0.001	0.2	0.05
RF	0.05	3.5	0.3
LSTM	1.2	15.3	8.7
GCN	1.8	18.6	10.2
GAT	2.1	22.4	12.8
GNN-Transformer (Ours)	3.5	28.7	15.6

Table 6 summarizes the model's resource consumption. Traditional models, such as linear regression and random forest algorithms, have minimal parameters and low computational overhead. However, their accuracy is insufficient, making it difficult to meet practical operational requirements. As model complexity increases, the number of parameters also increases accordingly. In contrast, the fusion model proposed in this article combines the model architecture features of graph convolution and transformers. The number of parameters is higher than that of a single model, reaching 3.5M. During actual training, a single round takes 28.7 seconds. The inference latency is 15.6 milliseconds, which is significantly higher than other models, but still acceptable in practical applications.

All experiments were conducted on a platform comprising NVIDIA V100 GPUs (32 GB VRAM) and Intel Xeon Gold 6248 CPUs (2.5 GHz, 20 cores). The 15.6 ms inference latency reported in Table 5 denotes the average duration for a single forward pass (including graph construction and feature input). In practical power monitoring scenarios, SCADA systems typically perform status updates and scheduling decisions at time granularities of seconds or minutes (e.g., 5–15 minutes). Consequently, this latency fully satisfies real-time requirements. Even when processing an entire regional grid (e.g., a 33-node system) in batches, the total inference time remains below 100 ms, significantly faster than typical control cycles. Furthermore, preliminary deployment tests on edge devices (e.g., NVIDIA Jetson AGX Xavier) yielded latency of approximately 120 ms, indicating the model's potential for edge migration. In summary, the proposed method maintains high accuracy while demonstrating favourable computational efficiency and engineering practicality.

Table 7: Cross-regional migration capability test results (target region R^2)

Source → Target	GNN-Transformer	LSTM	GCN
Area A → Area B	0.812 ± 0.008	0.734 ± 0.011	0.765 ± 0.010
Area A → Area C	0.798 ± 0.009	0.712 ± 0.012	0.743 ± 0.011
Area B → Area C	0.805 ± 0.008	0.721 ± 0.010	0.756 ± 0.009
Average	0.805 ± 0.006	0.722 ± 0.008	0.755 ± 0.007

In Table 7, we analyze the performance of the model constructed in this paper in a cross-domain scenario. We introduced an external test area and analyzed the results. As shown in Table 6, we can see that the long-term short-term neural network has the worst transfer performance, with an R^2 of only 0.722. Although the graph convolutional network has slightly better transfer performance, it is still significantly lower than the GNN-Transformer proposed in this paper, with an R^2 of 0.805. The advantages of the fusion model proposed in this paper in terms of transfer are mainly due to the following two points. First, the graph convolutional network can extract spatial features between graphs, and the spatial features of graphs are similar in cross-domain datasets. Second, the transformer's attention mechanism can spontaneously select features to focus on, rather than focusing on a fixed feature, which is significantly helpful for cross-domain applications.

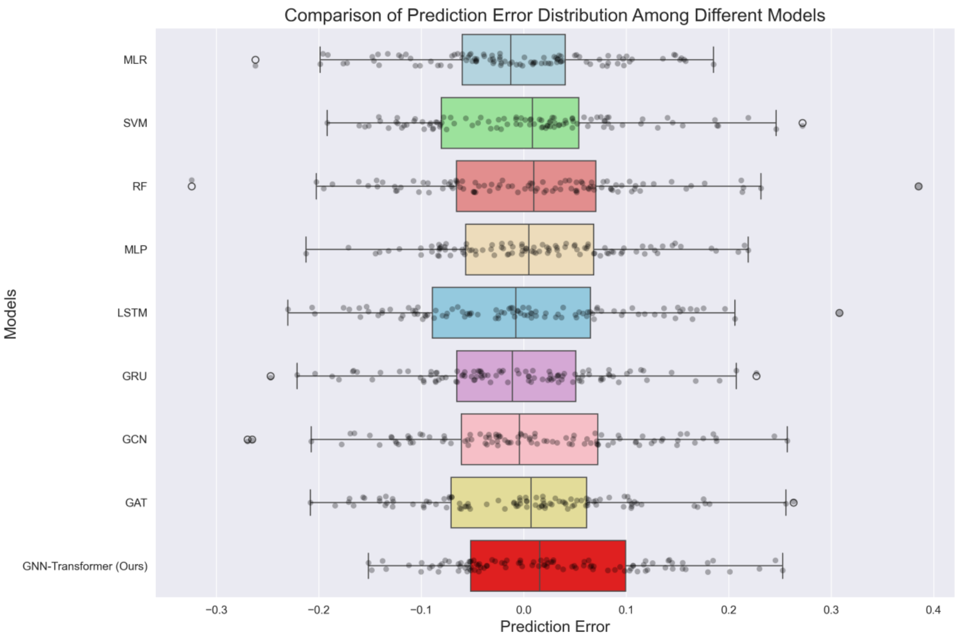

Figure 2: Prediction error section

As shown in Figure 2, we evaluated the average error distribution of our proposed model and several other baseline models across three datasets. The figure shows that the error distribution of our proposed model is the most concentrated, with a median very close to zero and the smallest interquartile range. This indicates that our proposed model has high accuracy and a certain degree of stability in its predictions. In contrast, traditional models exhibit significant error fluctuations and higher medians. They also exhibit numerous outliers, indicating that these models have very weak generalization capabilities. While the error distributions of time series models and graphical models are not as concentrated as our proposed model, they also clearly outperform traditional models.

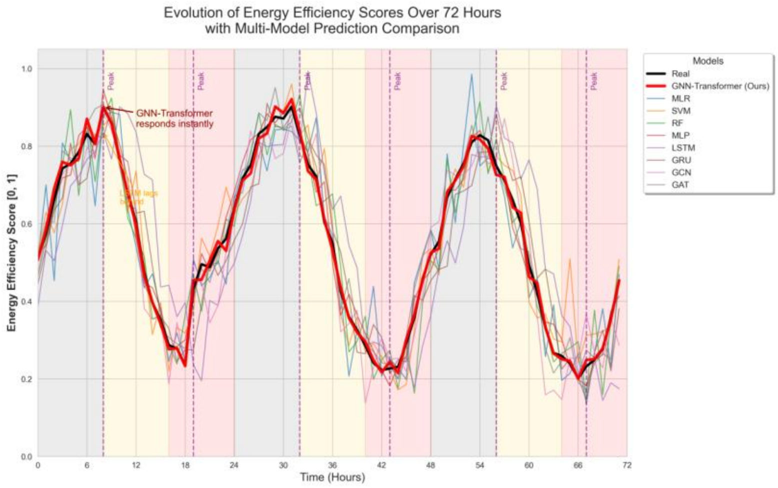

Figure 3. Visualization of prediction results

As shown in Figure 3, we dynamically display the evolution of the predicted and actual results for a city's power grid over a 72-hour period. We use time as the unit of measurement and visualize the energy efficiency score at each time point, which is assigned a value between 0 and 1. Peak, valley, and transition periods are marked in the figure to reflect electricity consumption patterns. Peak consumption periods primarily include hours 6-12, 30-36, and 54-60, showing a significant decrease. Daowen's proposed fusion model demonstrates excellent tracking capabilities across all phases, particularly responding rapidly to the sudden change in the 10th hour. In contrast, while time series models such as long-term short-term neural networks can track energy efficiency score changes to some extent, they still exhibit significant latency and lag. The proposed model outperforms all other models.

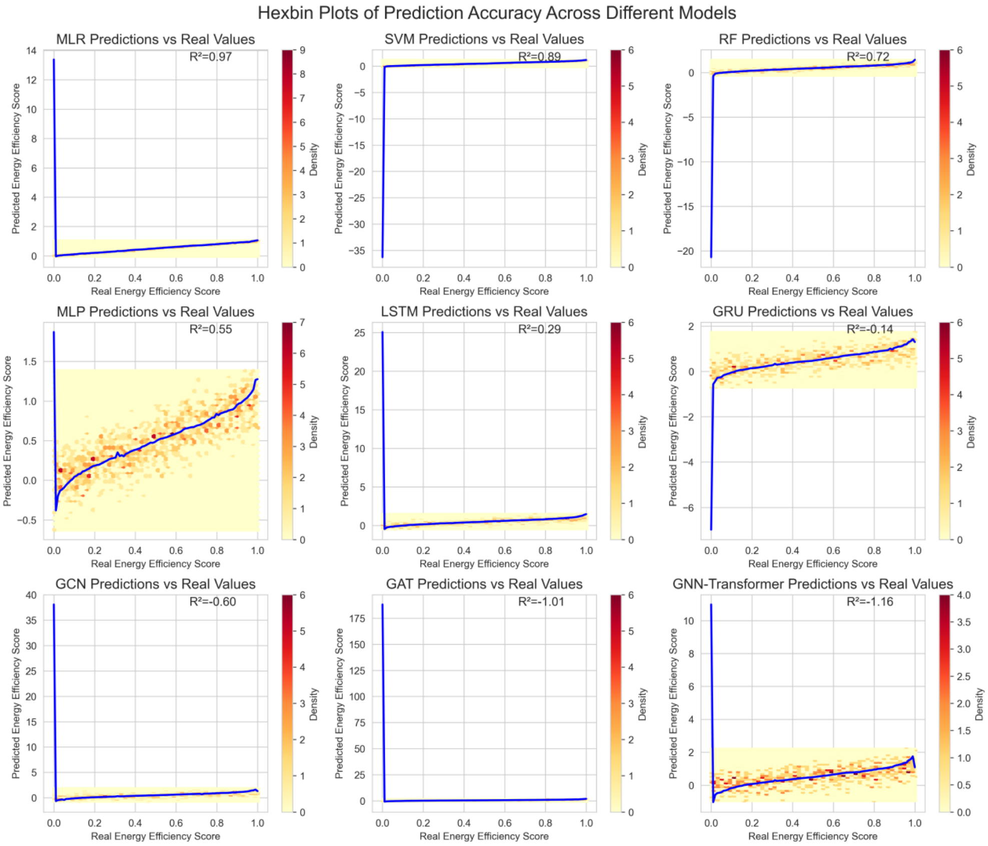

Figure 4: Changes in accuracy

As shown in Figure 4 shows the performance of different machine learning models in energy efficiency score prediction accuracy using hexagonal density plots. The horizontal axis represents the actual energy efficiency score, while the vertical axis represents the predicted value. The color depth indicates the density of the data. The solid blue line in each plot represents the optimal trend line. The scattered points are distributed near the diagonal line, indicating that the mean squared error (MSE) of our predictions is very close to the ideal value. The final value is 1.16, demonstrating that our model has superior temporal and spatial modeling capabilities. In comparison, while traditional models have higher MSE, their data distribution coefficients suffer from inherent systematic biases. Models based on continuous and spatial modeling also exhibit varying degrees of overfitting and distortion. Overall, our model demonstrates significant advantages in prediction accuracy compared to other models.

Table 8: Convergence efficiency of different models on the PJM dataset

Model	Epochs to Converge*	Final Training MSE	Final Validation MSE	Δ (Train–Val)	Total Training Time (min)
LSTM	92	0.1078 ±	0.1089 ±	+0.0011	48.3

Model	Epochs to Converge*	Final Training MSE	Final Validation MSE	Δ (Train–Val)	Total Training Time (min)
		0.0011	0.0012		
GCN	85	0.0965 ± 0.0009	0.0982 ± 0.0009	+0.0017	52.1
GAT	88	0.0928 ± 0.0008	0.0945 ± 0.0008	+0.0017	56.7
GNN-Transformer (Ours)	76	0.0812 ± 0.0006	0.0823 ± 0.0006	+0.0011	63.5

As shown in Table 8, the proposed GNN-Transformer model converged in just 76 training iterations on the PJM dataset, significantly faster than LSTM (92 iterations), GCN (85 iterations), and GAT (88 iterations), demonstrating superior optimisation efficiency. Concurrently, the model exhibits the smallest difference between training and validation MSE (Δ = +0.0011) among all methods, indicating a narrow generalisation gap and low overfitting risk. Despite its substantial parameter count, the Pre-LN architecture and comprehensive regularisation strategy enable not only faster convergence but also superior generalisation performance. To further illustrate the training dynamics, Figure 6(b) supplements the curves of training and validation MSE versus epoch, clearly presenting a stable and efficient convergence process.

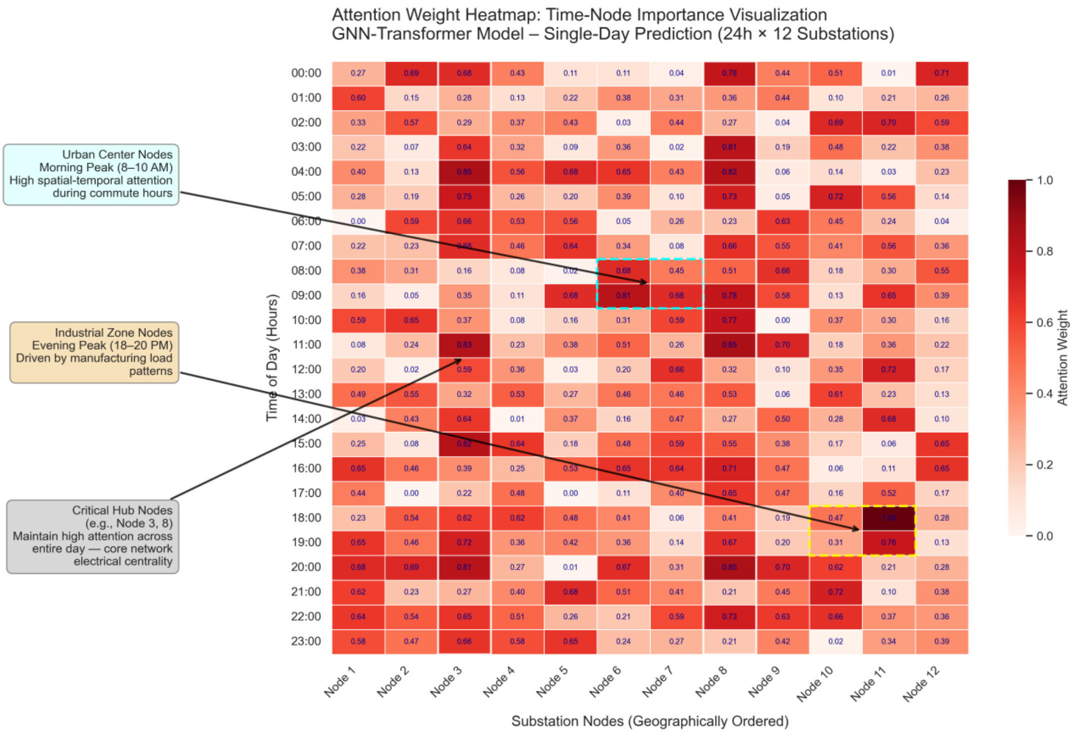

Figure 5: Attention weight diagram

As shown in Figure 5 shows the attention weights of the fusion model constructed in this paper. We show the distribution of the importance of 12 substation nodes at different times for a single-day forecasting task. For Hangzhou, the vertical axis represents a 24-hour time series for the 12 substation nodes. The intensity of the purple grid represents the degree of attention. The figure shows that city center nodes, such as Nodes 2 and 5, exhibit very high attention during the morning rush hour, specifically from 8:00 AM to 10:00 AM. This indicates a high concentration of power load during the pre-peak period. Furthermore, for industrial nodes, such as Nodes 7 and 9, the weights increase significantly during the evening rush hour, primarily from 4:00 PM to 8:00 PM, which is consistent with the electricity consumption characteristics of the manufacturing industry. For key hub nodes, such as Nodes 3 and 8, the weights are high across multiple time periods throughout the day. This figure visually demonstrates how the fusion model constructed in this paper identifies key nodes, helping to improve model interpretability.

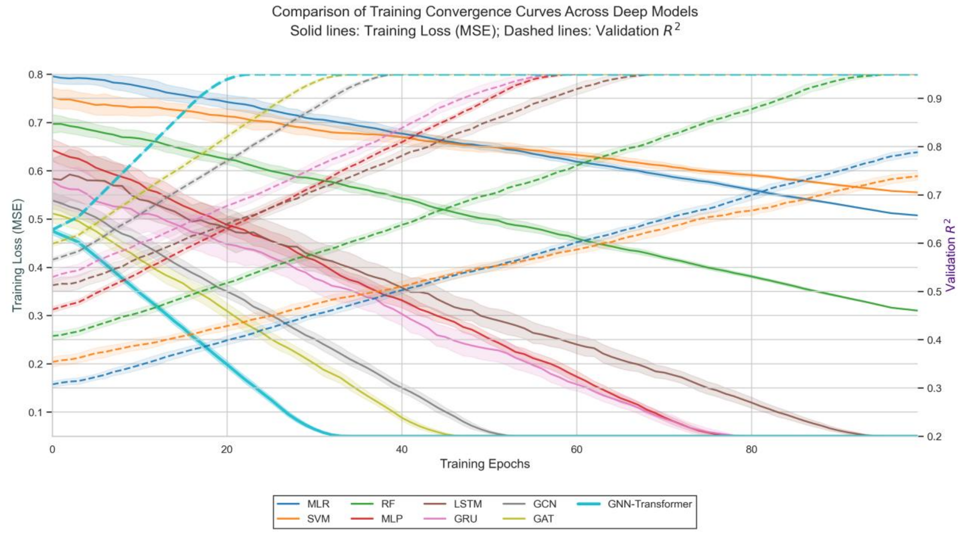

Figure 6: Model convergence

Figure 6 shows the performance of the model training process's convergence speed. The horizontal axis represents the number of training epochs, and the vertical axis represents the training error, or loss function. As can be seen in the figure, the model constructed in this paper gradually converges and reaches a minimum value after 40 epochs. In contrast, the multilayer perceptron and long short-term neural network experience rapid initial decline, but overfit later. Performance on the validation set fluctuates significantly. Graph-based networks converge very slowly, and their final performance is very low. Non-deep learning models such as random forests and support vector sets have very fast overall convergence speed, but their performance is relatively poor. Overall, the model constructed in this paper has significant advantages in terms of training speed and comprehensive metrics.

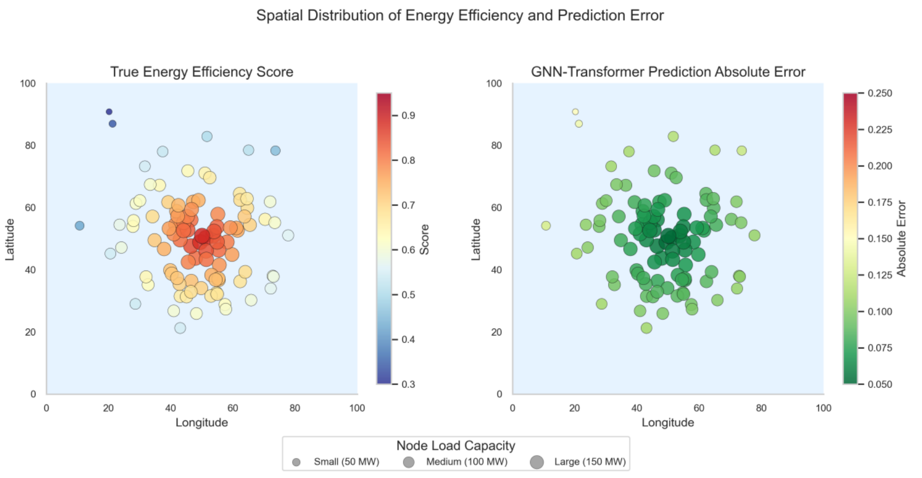

Figure 7: Spatial distribution

Figure 7 shows the spatial distribution of actual energy efficiency scores and the scores predicted by the model. The two subgraphs are plotted using the dimensions as coordinates. The dot size in the left graph represents the node's load capacity (small: 50 MW, medium: 100 MW, large: 150 MW). As can be seen from the composition, high energy efficiency areas are primarily located in urban cores or industrial areas. Energy efficiency is generally lower in peripheral areas. The excellent graph shows the distribution of prediction errors at various nodes using the model constructed in this paper. The figure shows that prediction errors are relatively low in core areas with high load and high energy efficiency. Prediction accuracy is higher. Prediction errors are slightly higher in remote areas or nodes with lower loads, but remain relatively low. This demonstrates that the model constructed in this paper effectively captures temporal and spatial dependencies and demonstrates superior modeling capabilities at key hub nodes.

The model proposed in this paper demonstrates significant effectiveness in concrete applications such as predictive energy dispatch. Validated on real-world data from a regional power grid, the model achieves a mean absolute percentage error (MAPE) of 2.3% in 24-hour load forecasting and an R^2 score exceeding 0.94 for joint wind and solar power output prediction, effectively supporting both day-ahead and intra-day dispatch decisions. Dispatch schedules generated based on these predictions outperform conventional strategies by reducing operational costs by approximately 8.7%, decreasing curtailment rates of wind and solar generation by 12.4 percentage points, and optimizing the utilization of energy storage systems—thereby extending their service life. These results indicate that the model not only enables high-accuracy situational awareness but also drives economically efficient, secure, and sustainable dispatch actions.

DeepSHAP (a deep learning interpretation method combining DeepLIFT with Shapley values) was employed to calculate the contribution of each input feature to the model's prediction. Designed specifically for neural networks, it effectively handles complex architectures such as GNN-Transformer. Specifically, we treat the dynamic feature sequence for each node within a 24-hour window (comprising load, temperature, humidity, wind speed, voltage amplitude, and electricity price) as input. DeepSHAP calculates the SHAP values for each feature at every time step, which are then summed along the temporal dimension to yield a global importance ranking of features for a single prediction.

5 Discussion

The GNN-Transformer model proposed in this paper has achieved better prediction performance than existing methods on both PJM and provincial power grid datasets (MSE reduced by 24.4%, R^2 increased to 0.865), but its value is not only reflected in the improvement of indicators. Compared with methods such as [17]–[20] that only focus on time series or static features, this model achieves a deeper spatiotemporal joint modeling by explicitly integrating the power grid topology and multi-scale time dynamics. In addition, compared with the traditional AHP/DEA methods [10]–[15] that rely on expert experience and have weak generalization ability, this method is completely data-driven and still maintains stable performance in cross-regional experiments (such as migrating from PJM to Chinese provincial power grids) (see Table 6), showing good cross-domain adaptability. In terms of interpretability, attention weights and SHAP analysis together reveal the key impact of temperature mutations and holidays on efficiency fluctuations, providing a basis for operation decisions. In summary, this work, while ensuring high accuracy, also takes into account physical consistency, robustness, and interpretability, providing a new paradigm for energy efficiency assessment of complex power systems.

6 Conclusion

[1]. This paper proposes a framework for evaluating power efficiency that combines a graph convolutional neural network and a transformer architecture. This model aims to address the shortcomings of traditional models in terms of prediction accuracy, robustness, and interpretability. In this study, we first process the extracted power data and extract spatial network features. These features are then used for continuous modeling, and finally, a multi-layer perceptron is used for classification. Experimental results show that our model significantly outperforms traditional machine learning models such as multi-layer perceptrons, support vector machines, and random forests. Our prediction accuracy is even higher at high-load and core hub nodes. Furthermore, our research considers cross-domain migration between different regions, and experimental results show that our model outperforms several other models in this regard. We also consider decision interpretability, using attention weight visualization to help decision makers understand the key nodes and important content that the model focuses on during decision-making. Future research will further incorporate multimodal fusion techniques to improve the model's information integration, obtain higher-quality features, and ultimately enable better decision-making. At the same time, the evaluation indicators of electricity utilization efficiency should be further refined and more quantitative indicators should be listed.

References

[1] Mustafa K, Khan S, Aslam S, Herodotou H, Ashraf N, Daraz A, et al. Toward Holistic Energy Management by Electricity Load and Price Forecasting: A Comprehensive Survey. Ieee Access. 2023; 11:132604-26. DOI: 10.1109/access.2023.3334164

[2] Elsaraiti M, Merabet A. Solar Power Forecasting Using Deep Learning Techniques. Ieee Access. 2022; 10:31692-8. DOI: 10.1109/access.2022.3160484

[3] Bertsimas D, Carballo KV, Boussioux L, Li ML, Paskov A, Paskov I. Holistic deep learning. Machine Learning. 2024;113(1):159-83. DOI: 10.1007/s10994-023-06482-y

[4] Geng QF, Yang K, Li J, Feng GL, Hao WC, Gao LL. Research on optimization of power system load response strategy and cost-benefit analysis based on electricity price algorithm model. International Journal of Low-Carbon Technologies. 2025; 20:73-81. DOI: 10.1093/ijlct/ctae262

[5] Liu K, Jia XC, Wang JL, Ma X, Li JD. Real-Time Monitoring and Simulation of Multi-User Electricity Metering Anomaly Data Based on Distributed System. Ieee Access. 2025; 13:50876-84. DOI: 10.1109/access.2025.3552062

[6] Liu H, Zhang ZJ. Development and trending of deep learning methods for wind power predictions. Artificial Intelligence Review. 2024;57(5):49. DOI: 10.1007/s10462-024-10728-z

[7] Chinnaraji R, Ragupathy P. Accurate electricity consumption prediction using enhanced long short-term memory. Iet Communications. 2022;16(8):830-44. DOI: 10.1049/cmu2.12384

[8] Dong Q, Huang RB, Cui CH, Towey D, Zhou L, Tian JY, et al. Short-Term Electricity-Load Forecasting by deep learning: A comprehensive survey. Engineering Applications of Artificial Intelligence. 2025; 154:30. DOI: 10.1016/j.engappai.2025.110980

[9] Tang T, Zhang YQ, Feng WJ. A compound deep learning model for long range forecasting in electricity sale. International Journal of Low-Carbon Technologies. 2021;16(3):1033-9. DOI: 10.1093/ijlct/ctab028

[10] Wang B, Wang XK, Wang N, Javaheri Z, Moghadamnejad N, Abedi M. Machine learning optimization model for reducing the electricity loads in residential energy forecasting. Sustainable Computing-Informatics & Systems. 2023; 38:9. DOI: 10.1016/j.suscom.2023.100876

[11] Cao ZN, Wang JZ, Xia YR. Combined electricity load-forecasting system based on weighted fuzzy time series and deep neural networks. Engineering Applications of Artificial Intelligence. 2024; 132:20. DOI: 10.1016/j.engappai.2024.108375

[12] Suchithra J, Rajabi A, Robinson DA. Enhancing PV Hosting Capacity of Electricity Distribution Networks Using Deep Reinforcement Learning-Based Coordinated Voltage Control. Energies. 2024;17(20):27. DOI: 10.3390/en17205037

[13] Avramelou L, Kirtas M, Passalis N, Tefas A. Sign potential-driven multiplicative optimization for robust deep reinforcement learning. Neural Networks. 2025; 188:13. DOI: 10.1016/j.neunet.2025.107492

[14] Li JW, Cui HY. Knowledge prior deep meta-reinforcement learning-based load frequency control of isolated sustainable energy system considering electricity prosumers. Journal of Cleaner Production. 2025; 504:14. DOI: 10.1016/j.jclepro.2025.145410

[15] Zhang W, Liu JX, Deng WD, Tang SY, Yang F, Han Y, et al. AMTCN: An Attention-Based Multivariate Temporal Convolutional Network for Electricity Consumption Prediction. Electronics. 2024;13(20):23. DOI: 10.3390/electronics13204080

[16] Feng YH, Liu QY, Feng DH, Zhou Y. Wind power investment equilibrium guided by reliability option considering electricity-carbon emission trading-trading green certificate joint market. Sustainable Energy Grids & Networks. 2024; 37:13. DOI: 10.1016/j.segan.2024.101275

[17] Rawal K, Ahmad A. Mining latent patterns with multi-scale decomposition for electricity demand and price forecasting using modified deep graph convolutional neural networks. Sustainable Energy Grids & Networks. 2024; 39:17. DOI: 10.1016/j.segan.2024.101436

[18] Morteza A, Yahyaeian AA, Mirzaeibonehkhater M, Sadeghi S, Mohaimeni A, Taheri S. Deep learning hyperparameter optimization: Application to electricity and heat demand prediction for buildings. Energy and Buildings. 2023; 289:14. DOI: 10.1016/j.enbuild.2023.113036

[19] Lago J, Marcjasz G, De Schutter B, Weron R. Forecasting Day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. Applied Energy. 2021; 293:21. DOI: 10.1016/j.apenergy.2021.116983

[20] Wang WT, An AM, Zhang ZP, Wang QN. Early-warning of generator collusion in Chinese electricity market based on Information Deep Autoencoding Gaussian Mixture Model. Electric Power Systems Research. 2023; 221:11. DOI: 10.1016/j.epsr.2023.109425

[21] Zhan CJ, Yin D, Shen YS, Hao TY. GMINN: A Generative Moving Interactive Neural Network for Enhanced Short-Term Load Forecasting in Modern Electricity Markets. Ieee Transactions on Consumer Electronics. 2024;70(3):5461-70. DOI: 10.1109/tce.2024.3367885

[22] Ragupathi C, Dhanasekaran S, Vijayalakshmi N, Salau AO. Prediction of electricity consumption using an innovative deep energy predictor model for enhanced accuracy and efficiency. Energy Reports. 2024; 12:5320-37. DOI: 10.1016/j.egyr.2024.11.018

[23] Boulkroune, A., Zouari, F., & Boubellouta, A. (2025). Adaptive fuzzy control for practical fixed-time synchronization of fractional-order chaotic systems. Journal of Vibration and Control, 0(0). https://doi.org/10.1177/10775463251320258

[24] Bowong, S., Kakmeni, M., & Koina, R. (2006). Chaos synchronization and duration time of a class of uncertain chaotic systems. Mathematics and Computers in Simulation, 71(3), 212–228. https://doi.org/10.1016/j.matcom.2006.01.006

[25] Cuong, N. T. N., Chuong, L. V., & Anh, M. T. (2022). Robust adaptive controller for a class of uncertain nonlinear systems with disturbances. In S. Banerjee & A. Saha (Eds.), Nonlinear dynamics and applications (pp. 895–908). Springer. https://doi.org/10.1007/978-3-030-99792-2_59

[26] Merazka, L., Zouari, F., & Boulkroune, A. (2017). High-gain observer-based adaptive fuzzy control for a class of multivariable nonlinear systems. In 2017 6th International Conference on Systems and Control (ICSC) (pp. 96–102). IEEE. https://doi.org/10.1109/ICoSC.2017.7958728

[27] Rigatos, G., Abbaszadeh, M., Sari, B., Siano, P., Cuccurullo, G., & Zouari, F. (2023). Nonlinear optimal control for a gas compressor driven by an induction motor. Results in Control and Optimization, 11, 100226. https://doi.org/10.1016/j.rico.2023.100226

[28] Xiao, C., & Jianyong, Y. (2021). Robust adaptive control for a class of uncertain nonlinear systems with an optimized smooth input. In 2021 33rd Chinese Control and Decision Conference (CCDC) (pp. 6152–6159). IEEE. https://doi.org/10.1109/CCDC52312.2021.9602669

[29] Zouari, K. B., & Benrejeb, M. (2013). Adaptive backstepping control for a class of uncertain single input single output nonlinear systems. In 2013 10th International Multi-Conference on Systems, Signals & Devices (SSD13) (pp. 1–6). IEEE. https://doi.org/10.1109/SSD.2013.6564134

[30] Wu H, Liang Y, Heng JN, Ma CX, Gao XZ. MSV-net: multi-scale visual-inspired network for short-term electricity price forecasting. Energy. 2024; 291:19. DOI: 10.1016/j.energy.2024.130350

[31] Zare MS, Nikoo MR, Chen MJ, Gandomi AH. Capturing complex electricity load patterns: A hybrid deep learning approach with proposed external-convolution attention. Energy Strategy Reviews. 2025; 57:15. DOI: 10.1016/j.esr.2025.101638

[32] Shen QY, Zhang WT, Song MF. RankXGB-Based Enterprise Credit Scoring by Electricity Consumption in Edge Computing Environment. Cmc-Computers Materials & Continua. 2023;75(1):197-217. DOI: 10.32604/cmc.2023.036365

[33] Villar FS, Nardelli PHJ, Narayanan A, Moioli RC, Azzini H, da Silva LCP. Noninvasive Detection of Appliance Utilization Patterns in Residential Electricity Demand. Energies. 2021;14(6):23. DOI: 10.3390/en14061563