ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

MOE-STSGKAN for electric vehicle charging demand forecasting based on hierarchical clustering

Meiying Yang[1,2,*], *Fang Wang*[1], *Yajing Zhang*[1], *Cheng Gong*[1,2], *Hao Ma*[1,2], *Yi Li*[2], *Xinzhi Lin*[2] *and Zhao Zhang*[2]

[1] State Grid Beijing Electric Power Company Electric Power Science Research Institute, Beijing, 100031, China.
[2] Beijing Dingcheng Hong'an Technology Development Co. Ltd, Beijing, 100075, China.

SUMMARY: *With the deepening global focus on sustainable development and carbon emission reduction, electric vehicles (EVs) have become a key component of clean energy transportation. To address the challenges of computational inefficiency in largescale node scenarios and the difficulty of synchronous spatiotemporal feature extraction in traditional graph neural networks, this paper proposes a Mixture-ofExperts Spatiotemporal Synchronous Graph Kolmogorov-Arnold Network based on hierarchical clustering. First, the global charging pile network is partitioned into 4 spatiotemporally correlated subgraphs using an agglomerative hierarchical clustering algorithm, and temporal patching strategy is employed for feature reconstruction. Second, a Spatiotemporal Synchronous Graph KolmogorovArnold Network is designed as expert models, where Fourier coefficients replace traditional spline functions to achieve frequency-domain fusion of spatiotemporal features. Finally, a gating network integrates expert outputs through transfer learning. Validation on real-world data from 24,761 charging piles in 247 Shenzhen districts demonstrates: For 15/30/45-minute prediction tasks, the proposed model achieves MAEs of 1.63/3.02/3.69, outperforming state-of-the-art baseline PAG by 4.7%/4.4%/7.3% respectively. The training time is reduced to 20.43 minutes. Experimental results confirm the method effectively balances prediction accuracy and computational efficiency in large-scale charging infrastructure scenarios.*

KEYWORDS: *Electric vehicles; Predicting charging demands; Hierarchical clustering; Graph Kolmogorov-Arnold network; Mixture expert models*

1 Introduction

The development and popularization of electric vehicles have attracted widespread attention, and by 2023, global sales of electric vehicles will exceed 10 million units. Promoting electric vehicles not only helps reduce global carbon emissions, but also is an important way to achieve energy system transformation, upgrading, and sustainable development [1, 2]. However, the explosive growth of electric vehicle sales has brought huge challenges to the construction of charging infrastructure [3]. Under the current conditions of charging infrastructure construction, accurate prediction of charging demand is crucial for the stability of the power grid system, the optimal layout of charging stations, and the improvement of user experience [4]. For example, the prediction error of charging demand may lead to a sharp increase in peak load of the power grid, resulting in localized power outages. Creating efficient and accurate charging demand

*yangmeiying@bj.sgcc.com.cn
https://doi.org/10.65102/is2026835
Published: 30 April, 2026

forecasting models has become a research focus of modern smart grids [5].

In traditional electric vehicle (EV) charging demand prediction algorithms, charging station terminals are often used as the prediction center. For example, [6] proposes a two-stage Poisson distribution model for the distribution of electric vehicle charging stations, reflecting the clustering characteristics of their distribution areas. In [7], scholars analyzed the charging demand of electric vehicles on the highway network and modeled it using the spatiotemporal distribution characteristics of dynamic traffic flow to achieve accurate prediction of the arrival rate of electric vehicles at charging stations. [8] constructs a charging load model based on daily load data from charging stations, and uses mathematical analysis to validate the charging process of electric vehicles.

These studies have achieved certain results in estimating charging demand. However, the research process fails to sufficiently consider the diversity of EVs and the randomness of user travel. Moreover, these studies fix the charging demand at specific locations and cannot objectively reflect the diversity of charging demand and the complexity of its spatiotemporal distribution. The common flaws of these methods are: inability to effectively integrate the characteristics of spatial and temporal dimensions. When the node scale expands, such as in city-level charging networks, the computational efficiency plummets [9]. Furthermore, these methods are not good at capturing the diverse behaviors of EV users, such as charging habits and travel patterns, which limits the prediction accuracy [10].

The rise of graph based neural networks, particularly graph convolutional networks (GCN) [11] and graph attention networks (GAT) [12], has shown great ability in handling graph structured data. For example, [13] designed a spatiotemporal graph convolutional network (STGCN) for traffic forecasting. It combined gated convolutional and graph convolutional structures to extract spatiotemporal correlations in road networks, surpassing combined CNN and RNN models in prediction accuracy.

[14] developed an adaptive module for spatial nodes in a graph convolutional recursive network (AGCRN), enabling automatic capture of fine grained traffic spatiotemporal correlations. [15] introduced an attention based spatiotemporal graph attention network (AST-GAT) for link level traffic speed prediction. Its multi-head graph attention blocks captured spatial dependencies between links, offering a new view on traffic flow dynamics. [16] treated the traffic network as a multi-graph structure. Using an interpretable spatiotemporal graph convolutional network (STGMN), it mined hidden features and added layers in a residual structure. Its prediction results were competitive with existing advanced models.

The aforementioned graph neural networks have been successfully applied to traffic prediction with high accuracy. In the field of EV charging demand prediction, graph neural networks can capture the complex relationships between charging stations and integrate topological information in node feature learning, which is a natural advantage for understanding and predicting charging demand patterns. Some recent studies have also used graph neural networks for EV charging demand. For example [17] presented a heterogeneous spatial-temporal graph convolutional network for short term EV charging demand prediction. It constructed a heterogeneous graph to model spatial correlations between charging areas and designed a region specific prediction module to boost accuracy. [18] proposed a physics informed and attention-based graph learning method (PAG) for regional EV charging demand prediction. It integrated graph attention networks with temporal pattern attention mechanisms, enhancing prediction accuracy and interpretability. [19] introduced a dynamic adaptive graph neural network-based approach for EV charging load forecasting. By creating a power grid graph structured matrix, it captured the spatial information of user charging characteristics. Also, a time convolutional network was used to perceive temporal information in charging data, and

an improved graph temporal convolutional network algorithm with attention mechanisms was applied for better fusion of spatial and temporal information.

Existing GNN methods, despite enhancing prediction accuracy, haven't solved two key issues:

(1) Computational efficiency bottleneck: As nodes increase, traditional GNNs' adjacency matrix operations lead to $O(N^2)$ computational complexity. Training time stretches from hours to tens of hours, failing to meet real time prediction needs [20]. For example, in scenarios with millions of nodes, STGCN's training time exceeds 48 hours, surpassing practical application thresholds [21].

(2) Asynchronous mining of spatial-temporal features: Most models use a serial architecture, processing spatial features first and then temporal dependencies [22]. This disconnects spatial-temporal dynamics, causing significant prediction deviations, such as in the morning rush hour tidal effect, due to feature disconnection. Synchronous fusion methods like STSGCN, limited by high computational costs of spline functions, converge slowly [23]. These issues are particularly prominent in large scale charging networks, restricting the practicality and scalability of prediction models [24].

To address the issues of plummeting computational efficiency when node scale expands and the difficulty of synchronous mining of spatial temporal features in traditional GNNs, this paper proposes a novel mixture-of-experts spatiotemporal synchronous graph Kolmogorov-Arnold network based on hierarchical clustering. Current methods show exponentially growing computation time with increasing nodes. For example, in large scale charging networks, training time of traditional GNNs may stretch from hours to tens of hours, hindering real time prediction and fast response. Also, traditional GNNs fall short in capturing dynamic changes in the time dimension while extracting spatial features, and vice versa, leading to considerable deviations in the predicted results' spatial-temporal accuracy. To tackle these problems, this paper innovatively integrates a hierarchical clustering algorithm with a mixture-of-experts model. By dividing the complex charging network into multiple subgraphs with close spatiotemporal links, constructing expert models for each subgraph, and fusing them via a gating network, this approach ensures high prediction accuracy while greatly boosting computational efficiency. The innovation of its structure lies in:

(1) Integrating agglomerative hierarchical clustering with temporal patching to split the overall charging pile layout network into multiple local spatiotemporal subgraphs. Then, pre-training datasets for each subgraph to get multiple expert models. Next, transferring these expert models and fine-tuning the expert gating network. This process yields a prediction model for the entire road network graph. It ensures high prediction accuracy while effectively cutting down prediction time.

(2) Secondly, a graph Kolmogorov-Arnold network is used in each expert model to extract spatiotemporal features synchronously. Fourier coefficients replace the spline functions used in the standard KAN. This change helps to improve the expressive power of the network model and reduce the training complexity of the algorithm model. Therefore, during the training process, the algorithm model that integrates GCN network and KAN network has more efficient performance.

2 Problem definition

Let X_t denote the feature vector of a node at timestamp *t*. The forecasting objective is to learn a complex nonlinear function *F* from historical observations, which computes future EV charging occupancy density metrics over τ steps:

$$\left(y_{t+1},\cdots,y_{t+\tau}\right)=F\left(X_{t-k;t},\vartheta\right) \tag{1}$$

where y_{t+i} represents predicted charging occupancy at $t+i$, $X_{t-k;t}$ represents historical feature sequence from $t-k$ to t, θ represents model parameters capturing spatiotemporal dependencies.

Moreover, this paper defines some concepts used in the approach as follows. The data of charging pile distribution networks can be represented by an undirected graph structured data $G(V,E)$, where V is the set of nodes and E represents the set of edges. The existence of edges between nodes is indicated by an adjacency matrix A. If $A(i,j)=1$, there is an edge between node i and node j; if $A(i,j)=0$, there is no edge connecting the two nodes.

3 Prediction model

To guarantee high prediction accuracy for each node and address the feature extraction challenges posed by numerous spatial nodes, this paper devises the MOE-STSGKAN method tailor made for EV charging demand prediction, as depicted in Figure 1.

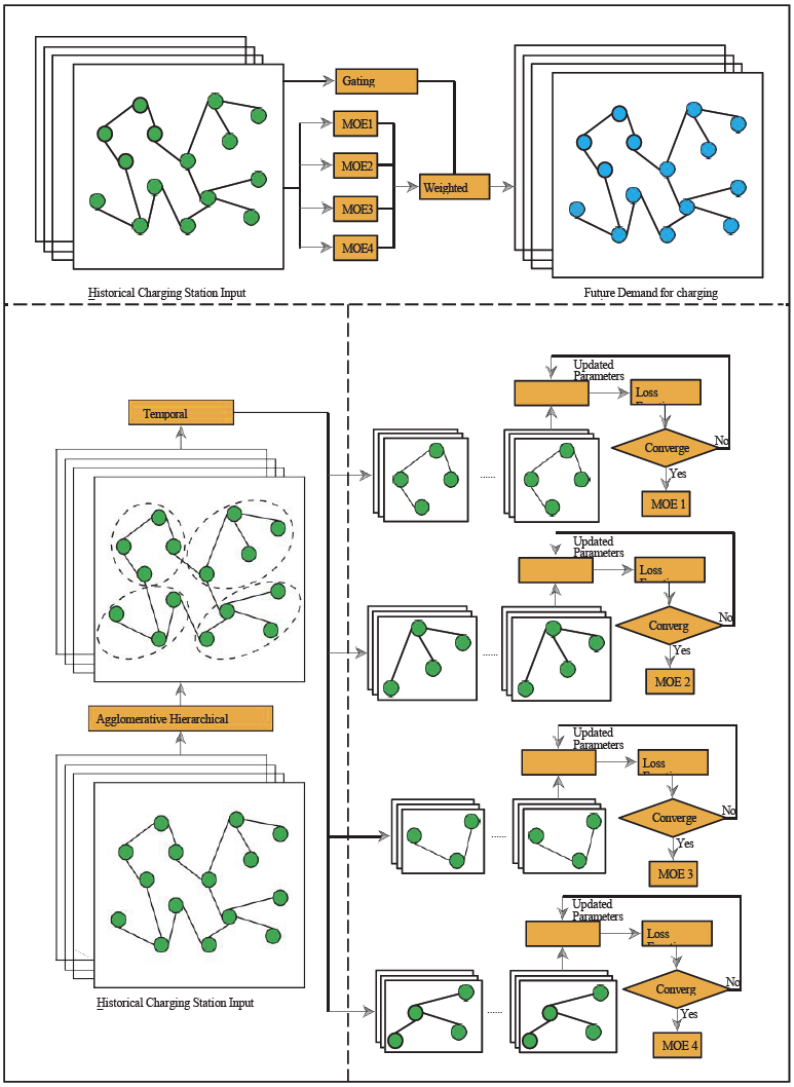

Figure 1: Prediction structure of MOE-STSGKAN based on hierarchical clustering

The method comprises three key modules: constructing local spatiotemporal subgraphs, pretraining expert models, and fine-tuning expert models. Initially, agglomerative hierarchical clustering and temporal patching are integrated to reconstruct historical input features into multiple local spatiotemporal subgraphs. Subsequently, STSGKAN is employed for pretraining, with each model being saved and designated as an expert model. Finally, combining all fixed parameter expert models and a fine-tuned gating network, the historical input features are trained to obtain the final prediction. The model's framework will be elaborated on below.

3.1 Constructing local spatiotemporal subgraphs

Agglomerative hierarchical clustering, a bottom-up method, is first applied to divide a large scale graph into subgraphs in the spatial dimension. Initially, each node is a separate subgraph, and they are sequentially merged based on proximity until a termination condition is met. Utilizing modularity as a similarity measure, this approach ensures the subgraphs are highly correlated in space-time. Consequently, the charging network is partitioned into multiple relatively independent subgraphs, with nodes in each subgraph exhibiting similar charging demand patterns. This strategy not only reduces computational complexity but also enhances the ability of subsequent expert models to capture local spatiotemporal features.

1. Initialization: Treat each node as a distinct subgraph C_i, with the subgraph count set to *n* (the total number of nodes).

2. Similarity Calculation: Compute the similarity $S\left(C_i, C_j\right)$ between all pairs of subgraphs. Here, modularity is adopted as the similarity metric:

$$S\left(C_i, C_j\right) = \frac{1}{2m} \sum_{a,b} \left[A_{ab} - \frac{k_a k_b}{2m} \right] \delta\left(c_a, c_b\right) \tag{2}$$

where A is the network's adjacency matrix. k_a and k_b are the degrees of nodes a and b, respectively. m is the total number of edges in the network. $\delta\left(c_a, c_b\right)$ is an indicator function that equals 1 if c_a and c_b are in the same subgraph, and 0 otherwise.

3. Merge Subgraphs: Identify the pair of subgraphs $\left(C_i, C_j\right)$ with the highest similarity and merge them into a new subgraph.

4. Update Similarity: Update the similarity between subgraphs based on the information subgraph structure.

5. Repeat steps 2-4: until the termination condition (a preset number of subgraphs) is met.

Assuming that the input features of the *q-th* subgraph are $X^q = \left[X^q_{t-\tau} \right]$ and the number of time channels for the time sliding window used for feature reconstruction is 3, with a stride of 1, then the reconstructed input features can be represented as follows:

$$ReconstructedFeatures = \left\{ X_{q,t} \middle| t = 1, 2, 3 \right\} \tag{3}$$

As the count of original input feature nodes is N^q, the count of newly input feature nodes after reconstruction becomes $3N^q$. Post-reconstruction, the new adjacency matrix A signifies the graphical connectivity for each channel, as shown in figure2.

As can be observed, it consists of the original adjacency matrix, an identity matrix, and a zero matrix, with its dimension being $3N^q \times 3N^q$.

3.2 Expert Model Pre-training

After obtaining multiple subgraph features, this paper employs a patio-emporal ynchronous raph Kolmogorov-Arnold etwork (STSGKAN) to perform deep feature extraction on each subgraph and forecast the future charging stations electricity demand for each node within the subgraphs.

STSGKAN is a graph convolutional network based on KAN. It takes node features and an adjacency matrix as input. Information is passed through multiple KAN layers, and a linear layer gives the output. KAN, rooted in the Kolmogorov-Arnold theorem, approximates complex high dimensional functions by combining nonlinear ones. In STSGKAN, this KAN feature is used to explore nonlinear relationships in spatiotemporal features. KAN maps input data to a higher dimensional space for better nonlinear relationship capture. Then, through multi-layer KAN information passing, node features are integrated and updated in space and time, mining spatio-temporal synchronous features. This design lets STSGKAN excel in handling complex spatiotemporal data. Compared to traditional graph neural networks, it grasps the spatiotemporal patterns of charging demand more accurately.

Suppose the input is *X*, the adjacency matrix is *A*, and there are *L* KAN layers. The entire forward propagation of STSGKAN can be expressed as follows:

$$\begin{cases} H^{(0)} = X \\ H^{(l)} = \sigma\left(\mathrm{KAN}\left(H^{(l-1)}, A\right)\right), l = 1, \cdots, L-1 \\ Y = W_{L-1} H^{(L-1)} + b_{L-1} \end{cases} \tag{4}$$

where $H^{(l)}$ denotes the hidden state of the *l*-th layer, W_{L-1} are the weight matrices of the output layers, respectively, b_l represents the bias of the *l*-th layer, *Y* corresponds to the final output, $\sigma(\cdot)$ is the internal nonlinear activation function of KAN. It interacts with input features using sine and cosine functions, and combines their results through predefined Fourier coefficients. Further details follow.

Within the KAN layer, input data is transformed into a higher dimensional space where each dimension corresponds to a distinct frequency component. This process can be approximated via a Fourier series. Firstly, a parameter *F* is initialized to store Fourier coefficients, structured as $\left(2, d_{\text{out}}, d_{\text{in}}, g_{\text{size}}\right)$. The initial dimension (sized 2) contains coefficients for cosine and sine terms; the second (sized d_{out}) aligns with the output feature dimension; the third (sized d_{in}) matches the input feature dimension; and the last (sized g_{size}) indicates the count of frequency components. Subsequently, for each input feature z_i, an integer sequence *k* ranging from 1 to g_{size} is generated. Finally, the Fourier coefficients are multiplied by the corresponding sine and cosine values and aggregated through summation.

$$y_i = \sum_{m=1}^{d_{\text{in}}} \sum_{n=1}^{d_{\text{out}}} \sum_{k=1}^{K} \left[c_{m,n,k} \cdot \cos\left(k \cdot z_{i,m}\right) + s_{m,n,k} \cdot \sin\left(k \cdot z_{i,m}\right) \right] \tag{5}$$

where $c_{m,n,k}$ and $s_{m,n,k}$ are coefficients extracted from $F[0:1]$ and $F[1:2]$, respectively. By doing this, KAN transforms input features into frequency domain representations. This enables more effective capture of periodicity and trends in data. This frequencydomain fusion method enhances the model's feature representation capability. It also eases training and allows the model to converge more quickly.

After building the STSGKAN forecasting model for subgraphs through the above process, the model is trained using the Mean Squared Error (MSE) as the loss function and the Adam

optimization algorithm for parameter updates. The trained parameters are then saved. Each trained model serves as an expert model for subsequent transfer learning. Figure 2 shows the adjacency matrix of the reconstructed subgraph.

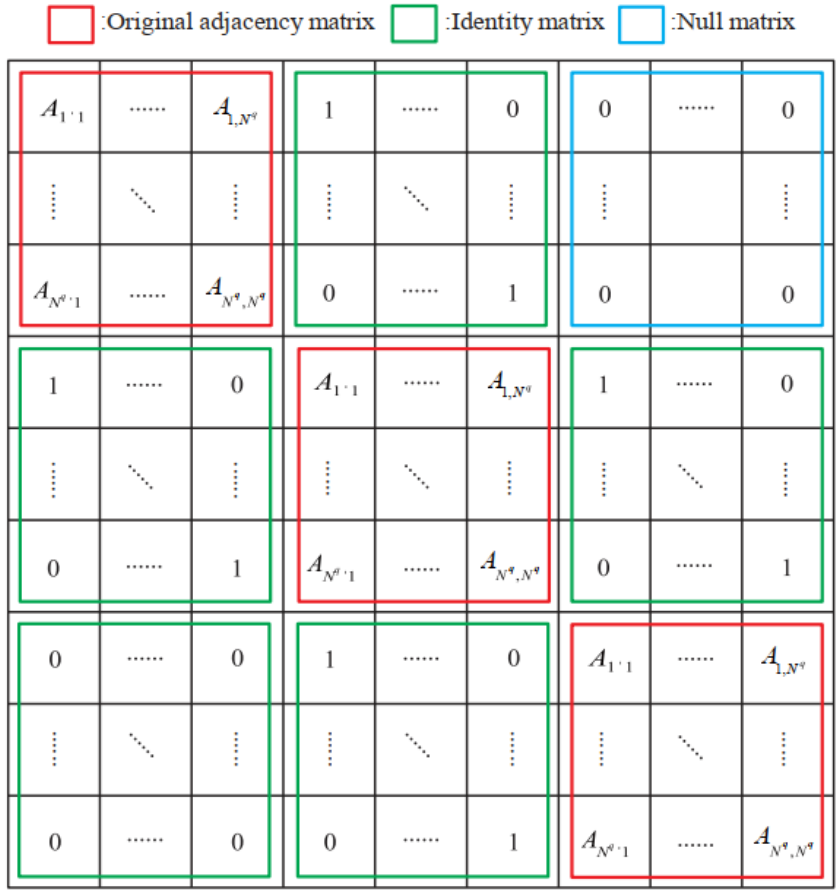

Figure 2: Reconstruct the adjacency matrix of the subgraph

3.3 Fine-tuning the Expert Model

Once the whole historical dataset is fed as input features into each pre-trained expert model, a gating network will be employed to weigh the output features of every expert model. The gating network consists of two fully connected layers. The first layer uses linear mapping to reduce the input feature channels in the time dimension to 1. In the second layer, linear mapping cuts the input node count to match the number of expert models. The calculation is as follows:

$$H^{(G)} = \mathrm{Softmax}\left(W_2 \times W_1\right) \tag{6}$$

where $H^{(G)}$ represents the output sequence of the gating network, while W_1 and W_2 denote the weights of the two fully connected layers.

4 Experiments and analysis

All simulations were performed on a computer with an NVIDIA GeForce RTX 4090 GPU, using the PyTorch open source framework for model building.

4.1 Data description

This study used data from a public mobile app that shows real time charging station availability. It focuses on Shenzhen, China, covering June 19 July 18, 2022 (30 days). The study involves 18,061 public charging stations, with data sampled every five minutes across 8,640 time points.

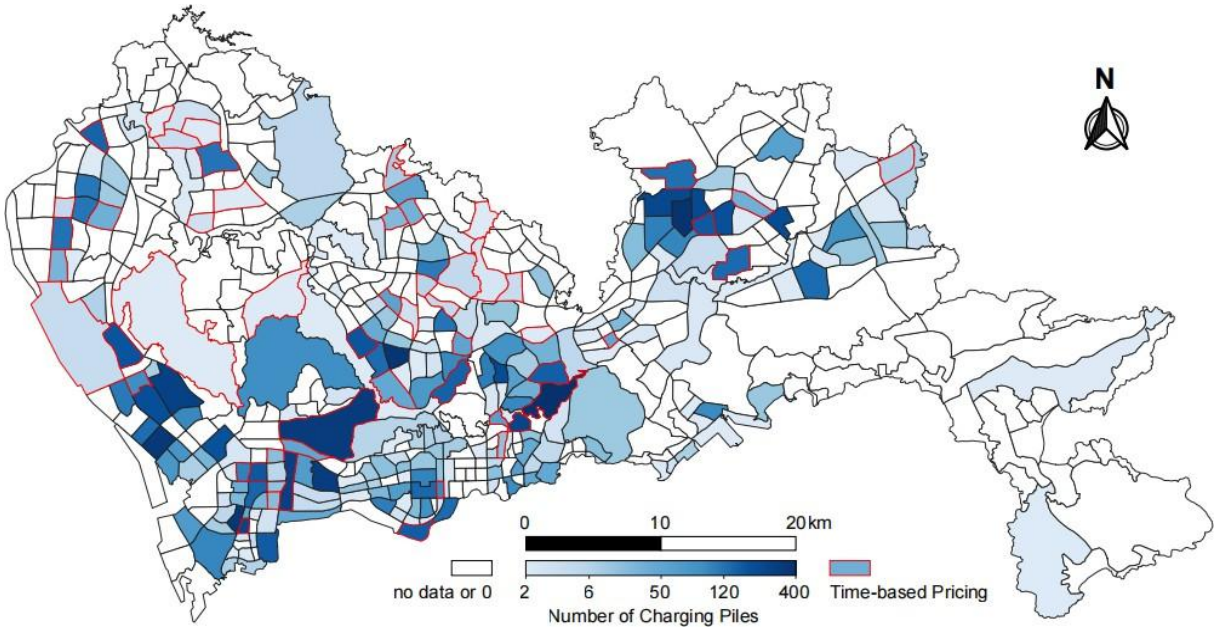

Figure 3: Spatial distribution characteristic map of public electric vehicle charging piles in Shenzhen City

As shown in Figure 3, Shenzhen is modeled as a graph with 247 nodes (areas) and 1,006 edges (adjacent relationships). The first 80% of the dataset is the training set, and the last 20% is the test set. Before the experiment, data preprocessing was conducted. To address the noise and missing values in the raw data: 1) The mean value imputation method was applied to fill in the missing charging demand data. 2) Data standardization was performed to scale the charging demand values to the [0, 1] range. This improves the convergence speed and stability of model training.

4.2 Model parameter settings

Through repeated training and testing experiments, the optimal MOE-STSGKAN model parameters were selected as follows: 1) The model's input feature historical time window length is set to 1 hour, with prediction time horizons of 15, 30, or 45 minutes. Given the 5 minutes interval between adjacent time points, the time parameters are $\tau_1 = 12$ and $\tau_2 \in [3,6,9]$. 2) The dimension of parameter F, which stores Fourier coefficients in KAN, is set to (2, 10, 10, 300). 3) During iterative optimization, the batch size per sample is 64, and the learning rate is 1e-4.

During model training and testing, the following steps were adopted:The dataset was divided into a training set and a test set. The training set was used for model training and hyperparameter tuning, while the test set was reserved for final model evaluation. In the training phase, early stopping was implemented. Training ceased if the validation loss remained unchanged for 10 consecutive epochs to prevent overfitting. During testing, the model's predictions on the test set were recorded, and relevant evaluation metrics were computed.

4.3 Performance Advantage Analysis

Other baseline parameters are set per the references. First, the prediction accuracy of each model is compared using two metrics: Mean Absolute Error (MAE) and Root Mean Square Error (RMSE):

$$\begin{cases} MAE = \dfrac{1}{N \cdot T_{\text{te}}} \sum_{i=1}^{N} \sum_{j=1}^{T_{\text{te}}} \left| \hat{Y}_{i,j} - Y_{i,j} \right| \\ \text{RMSE} = \sqrt{\dfrac{1}{N \cdot T_{\text{te}}} \sum_{i=1}^{N} \sum_{j=1}^{T_{\text{te}}} \left(\hat{Y}_{i,j} - Y_{i,j} \right)^2} \end{cases} \tag{7}$$

where T_{te} denotes the number of channels in the time dimension of the test set, $\hat{Y}$ represents the predicted values, and Y represents the truth values. Generally, smaller MAE and RMSE values indicate higher model prediction accuracy.

In model comparison experiments, all models underwent identical preprocessing and data division. LSTM, GCN, STGCN, AST-GAT, STGMN, and PAG models were trained per literature parameters, with loss changes and validation performance tracked. For reliability, each model was trained and tested five times independently, averaging results for the final outcome.

As indicated in Table 1, it presents the prediction results of different models for domain time lengths of 15, 30, and 45 minutes. It can be seen that the MAE and RMSE of LSTM and GCN are the highest. This is because LSTM can only extract the temporal correlation of historical data, while GCN can only extract the spatial correlation of historical data, so the prediction accuracy is poor. STGCN, which can extract both temporal and spatial correlations, has better prediction accuracy than LSTM and GCN. AST-GAT and STGMN further enhance prediction accuracy by integrating attention mechanisms with graph neural networks. PAG demonstrates the best baseline performance by combining physical information with attention mechanisms. However, the proposed model, which uses STSGKAN to simultaneously extract spatiotemporal correlations from each subgraph and integrates multiple STSGKAN through a mixture-of-experts approach, shows the highest prediction accuracy.

Table 1: Error evaluation metrics of different prediction models

Model	15min		30min		45min	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
LSTM	2.24	3.87	3.65	5.89	4.81	8.46
GCN	4.41	5.64	6.87	8.57	7.74	9.52
STGCN	1.99	3.74	3.21	5.48	4.56	7.93
AST-GAT	1.83	3.41	3.37	5.44	4.42	7.74
STGMN	1.79	3.28	3.33	5.41	4.25	7.27
PAG	1.71	3.04	3.16	5.06	3.98	6.85
Ours	1.63	2.81	3.02	4.77	3.69	6.42

Model training and testing times are crucial for performance evaluation. Figure 4 and Figure 5 shows the computational times for the designed model and baseline models. The designed model takes longer than LSTM, GCN, and STGCN, which have simpler structures but lower prediction accuracy. However, its computational time is shorter than that of the complex AST-GAT, STGMN, and PAG models. This indicates the designed model effectively addresses the issue of increased prediction time caused by constructing local spatiotemporal graphs for

simultaneous correlation mining.

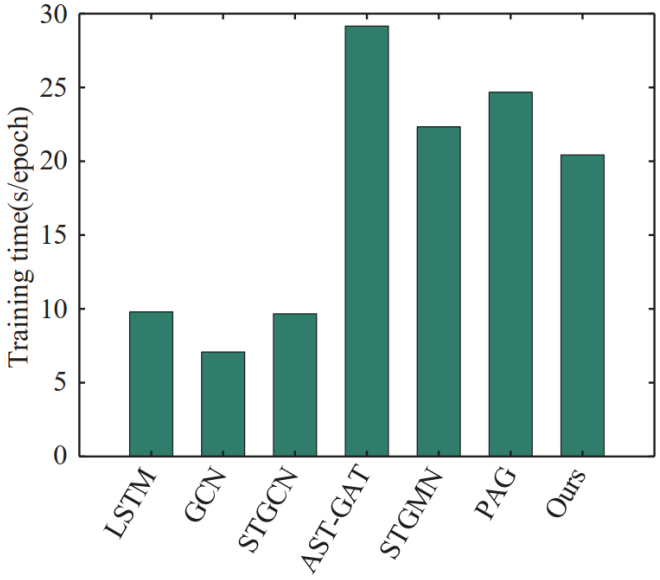

Figure 4: The training time of different models

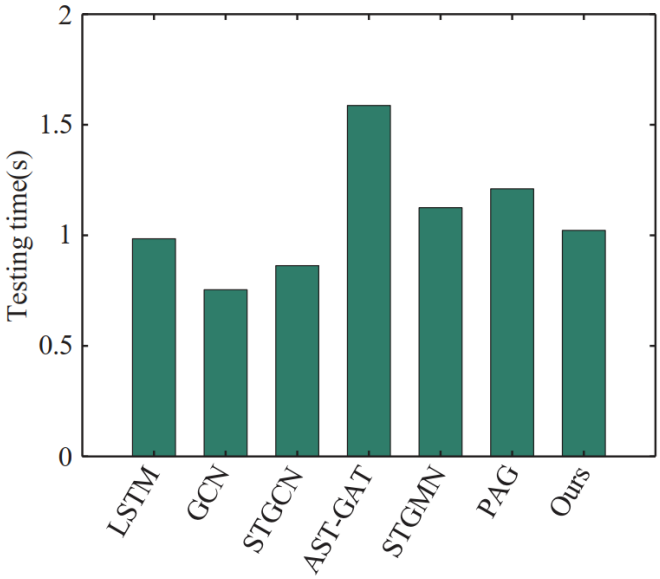

Figure 5: The testing time of different models

In charging station networks, many spatial nodes may be heterogeneous. To verify the designed prediction model's higher accuracy across different spatial node types, this study compares predicted and actual values for high, medium and low occupancy spatial points, as illustrated in Figure 6. The model is shown to adapt effectively to data from these spatial points with varying occupancy levels.

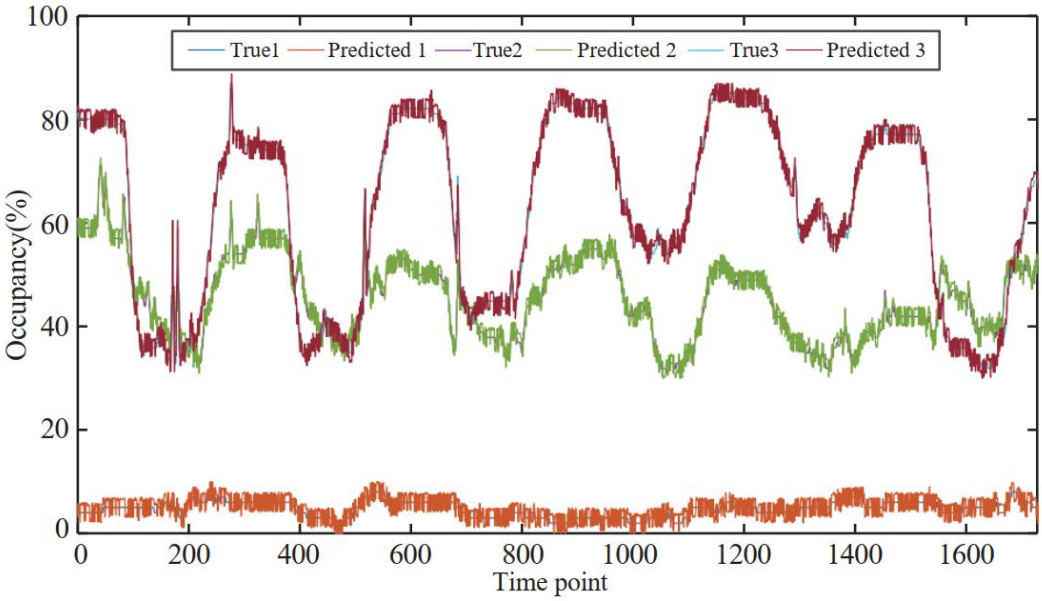

Figure 6: Visualization of the real and predicted occupancy rate of charging piles at different spatial points

4.4 Performance Verification of Key Modules

The model designed in this paper has two key modules: the mixture of expert's module and the GKAN network. To check if pre-training multiple expert models and using a gating system can reduce training time for large node prediction models, we trained a STSGKAN model on the entire spatial graph data. Then, by adjusting the preset number of subgraphs in the agglomerative hierarchical clustering algorithm, we built MOE-STSGKAN models with different numbers of subgraphs (3, 4, 5, and 6). Each model underwent full training and testing, with its MAE, RMSE, training time, and testing time recorded. The results are in Table 2.

From Table 2, the MOE-STSGKAN model has lower prediction accuracy but shorter computation time than the STSGKAN model. This shows that the MOE approach reduces accuracy slightly but cuts computation time significantly. By choosing an appropriate number of subgraphs, we can balance prediction accuracy and computation time. For example, setting the number of subgraphs to 4, as in this study, achieves a good trade off.

To verify the performance improvement of the GKAN network over other graph neural networks in prediction models, this study conducted comparative experiments with prediction models using GCN and GAT networks. The experiments only replaced the graph neural network module while keeping other conditions constant. Each model underwent full training and testing, with evaluation metrics recorded, as shown in Table 3.

Table 2: Verify the comparison of different algorithms in the performance of hybrid expert model modules

Model	MAE	RMSE	Training time	Testing time
STSGKAN	1.57	2.58	33.35	2.94
MOE-STSGKAN(3 subgraphs)	1.61	2.73	23.41	1.87
MOE-STSGKAN(4 subgraphs)	1.63	2.81	20.43	1.03
MOE-STSGKAN(5 subgraphs)	1.69	3.01	19.17	0.95
MOE-STSGKAN(6 subgraphs)	1.75	3.16	17.52	0.87

Table 3: Verify the comparison of different GKAN network performance algorithms

Model	MAE	RMSE	Training time	Testing time
MOE-STSGCN	1.78	3.34	22.65	1.34
MOE-STSGAT	1.72	3.10	21.09	1.16
MOE-STSGKAN	1.63	2.81	20.43	1.03

As shown in Table 3, the prediction model with the GKAN network outperforms those with GCN and GAT in prediction accuracy and reduces computation time.

4.5 Key process validation experiment

To comprehensively verify the effectiveness of the key processes in the proposed method, including the advantages of the agglomerative hierarchical clustering algorithm in subgraph partitioning, the boosting effect of the time patching strategy on model performance, and the advantages of Fourier coefficients in fusing spatiotemporal features in the frequency domain, various combinations of conditions are designed to evaluate the impact of different key process combinations on the overall model performance.

As shown in Table 4, the sub graphs divided by the agglomerative hierarchical clustering algorithm have the highest modularity, increasing by 19.7% and 12.3% over K-means and DBSCAN respectively. Under identical other conditions, the model with this algorithm outperforms others in MAE and RMSE, indicating its effectiveness in dividing closely spatiotemporally correlated sub graphs and providing a better basis for local feature mining.

When comparing models with and without the time patching strategy under the same clustering algorithm, those with the strategy have significantly lower MAE and RMSE and reduced training convergence time. For instance, with K-means clustering plus spline functions, adding time patching lowers MAE by 6.3%, RMSE by 6.4%, and cuts training time by 4.7%. This shows the strategy enhances the model's ability to capture time dimensional features, improving prediction accuracy and training speed. Moreover, under the same clustering algorithm and time patching strategy, replacing spline functions with Fourier coefficients further lowers MAE and RMSE and reduces training time. Take K-means clustering plus time patching as an example; using Fourier coefficients reduces MAE by 4.7%, RMSE by 6.9%, and slashes training time by 6.1%. This confirms that Fourier coefficients excel in fusing spatiotemporal features in the frequency domain. By transforming input features into the frequency domain, the model better captures periodicity and trends, boosting feature representation, easing training, and accelerating convergence.

Table 4: Comparison of model performance under different combination conditions

Experimental conditions	Modularity	MAE	RMSE	Training time
K-means+No time patching+Spline functions	0.61	2.05	3.58	27.43
K-means+Time patching+Spline functions	0.61	1.92	3.35	26.15
K-means+Time patching+Fourier coefficient	0.61	1.83	3.12	24.56
DBSCAN+No time patching+Spline functions	0.65	1.98	3.41	26.78
DBSCAN+Time patching+Spline functions	0.65	1.85	3.24	25.34
DBSCAN+Time patching+Fourier coefficient	0.65	1.75	3.01	23.89
Agglomerative hierarchical+ No time patching+Spline functions	0.73	1.87	3.28	25.47
Agglomerative hierarchical+ Time patching+Spline functions	0.73	1.76	3.10	24.23
Agglomerative hierarchical+ Time patching+Fourier coefficient	0.73	1.63	2.81	20.43

5 Conclusion

This paper presents a MOE-STSGKAN model for predicting EV charging station demand, designed to address long computation times and high hardware demands in networks with numerous nodes. The model uses an agglomerative hierarchical clustering algorithm based on optimal modularity to divide the charging station network into subgraphs. Each subgraph's data is restructured by a time Patching module. Multiple expert models are pre-trained and then fused through fine-tuning to produce the final prediction. Simulation results highlight the model's high accuracy and short training and testing times. They also show that selecting the optimal subgraph preset value can reduce computation time while maintaining prediction accuracy.

References

[1] Acharige S S G, Haque M E, Arif M T, et al. Review of electric vehicle charging technologies, standards, architectures, and converter configurations[J]. IEEE access, 2023, 11: 41218-41255.

[2] Hopkins E, Potoglou D, Orford S, et al. Can the equitable roll out of electric vehicle charging infrastructure be achieved?[J]. Renewable and Sustainable Energy Reviews, 2023, 182: 113398.

[3] Hemavathi S, Shinisha A. A study on trends and developments in electric vehicle charging technologies[J]. Journal of energy storage, 2022, 52: 105013.

[4] LaMonaca S, Ryan L. The state of play in electric vehicle charging services–A review of infrastructure provision, players, and policies[J]. Renewable and sustainable energy reviews, 2022, 154: 111733.

[5] Barakat S, Osman A I, Tag-Eldin E, et al. Achieving green mobility: Multi-objective optimization for sustainable electric vehicle charging[J]. Energy Strategy Reviews, 2024, 53: 101351.

[6] Dimitriadou K, Rigogiannis N, Fountoukidis S, et al. Current trends in electric vehicle charging infrastructure; opportunities and challenges in wireless charging integration[J]. Energies, 2023, 16(4): 2057.

[7] Garofalaki Z, Kosmanos D, Moschoyiannis S, et al. Electric vehicle charging: A survey on the security issues and challenges of the open charge point protocol (OCPP)[J]. IEEE Communications Surveys & Tutorials, 2022, 24(3): 1504-1533.

[8] Rachid A, El Fadil H, Gaouzi K, et al. Electric vehicle charging systems: comprehensive review[J]. Energies, 2022, 16(1): 255.

[9] Gilleran M, Bonnema E, Woods J, et al. Impact of electric vehicle charging on the power demand of retail buildings[J]. Advances in Applied Energy, 2021, 4: 100062.

[10] Mazhar T, Asif R N, Malik M A, et al. Electric vehicle charging system in the smart grid using different machine learning methods[J]. Sustainability, 2023, 15(3): 2603.

[11] Yuvaraj T, Devabalaji K R, Kumar J A, et al. A comprehensive review and analysis of the allocation of electric vehicle charging stations in distribution networks[J]. IEEE access, 2024, 12: 5404-5461.

[12] Sagar A, Kashyap A, Nasab M A, et al. A comprehensive review of the recent development of wireless power transfer technologies for electric vehicle charging systems[J]. Ieee Access, 2023, 11: 83703-83751.

[13] Unterluggauer T, Rich J, Andersen P B, et al. Electric vehicle charging infrastructure planning for integrated transportation and power distribution networks: A review[J]. ETransportation, 2022, 12: 100163.

[14] Borlaug B, Yang F, Pritchard E, et al. Public electric vehicle charging station utilization in the United States[J]. Transportation Research Part D: Transport and Environment, 2023, 114: 103564.

[15] Rajendran G, Vaithilingam C A, Misron N, et al. A comprehensive review on system architecture and international standards for electric vehicle charging stations[J]. Journal of energy storage, 2021, 42: 103099.

[16] Karmaker A K, Hossain M A, Pota H R, et al. Energy management system for hybrid renewable energy-based electric vehicle charging station[J]. IEEE Access, 2023, 11: 27793-27805.

[17] Nezamuddin O N, Nicholas C L, dos Santos E C. The problem of electric vehicle charging: State-of-the-art and an innovative solution[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 23(5): 4663-4673.

[18] Güven A F, Yücel E. Sustainable energy integration and optimization in microgrids: enhancing efficiency with electric vehicle charging solutions[J]. Electrical Engineering, 2025, 107(2): 1541-1573.

[19] Bilal M, Ahmad F, Rizwan M. Techno-economic assessment of grid and renewable powered electric vehicle charging stations in India using a modified metaheuristic technique[J]. Energy Conversion and Management, 2023, 284: 116995.

[20] Khalid M R, Khan I A, Hameed S, et al. A comprehensive review on structural topologies, power levels, energy storage systems, and standards for electric vehicle charging stations and their impacts on grid[J]. IEEE access, 2021, 9: 128069-128094.

[21] Rahman S, Khan I A, Khan A A, et al. Comprehensive review & impact analysis of integrating projected electric vehicle charging load to the existing low voltage distribution system[J]. Renewable and Sustainable Energy Reviews, 2022, 153: 111756.

[22] Allouhi A, Rehman S. Grid-connected hybrid renewable energy systems for supermarkets with electric vehicle charging platforms: Optimization and sensitivity analyses[J]. Energy Reports, 2023, 9: 3305-3318.

[23] Mukherjee S, Ruiz J M, Barbosa P. A high power density wide range DC–DC converter for universal electric vehicle charging[J]. IEEE Transactions on Power Electronics, 2022, 38(2): 1998-2012.

[24] Rane N L, Achari A, Saha A, et al. An integrated GIS, MIF, and TOPSIS approach for appraising electric vehicle charging station suitability zones in Mumbai, India[J]. Sustainable Cities and Society, 2023, 97: 104717.