HMGNN: Charging Demand Prediction With Hypergraph-Connected Multimodal Temporal-Spatial Data Mining

Haixing Guo[1], Xiaolong Ren[1], Hengbin Si[1], Mengting Wang[1], Qing Wu[1], Xi Chen[1], Shuang Tian[1], and Xuan Zhen[2*]

[1] *State Grid Shaanxi Information and Telecommunication Company, Xi'an, Shaanxi, China*
[2] *School of Informatics, Xiamen University, Xiamen, P.R. China*
[*] *Corresponding author. E-mail: 30920231154331@stu.xmu.edu.cn*

Received: Apr. 20, 2026; Accepted: Jun. 13, 2026

Electric vehicles (EVs) are undergoing rapid expansion, leading to an explosive increase in ownership. This surge, however, poses unprecedented challenges to power grid operation and charging infrastructure planning. In particular, accurately forecasting charging demand has become increasingly difficult. Conventional graph-based or single-modal models often fail to capture the complex and nonlinear nature of charging behaviors because (1) transportation networks exhibit pronounced spatiotemporal heterogeneity, and (2) charging stations are characterized by intricate, multi-scale interdependencies. To overcome these limitations, this study introduces a multimodal hypergraph-based charging demand forecasting framework. The proposed approach constructs multidimensional spatiotemporal hypergraphs from heterogeneous multimodal data, thereby uncovering higher-order correlations among charging stations across different spatial and temporal scales. By dynamically modeling these evolving inter-station relationships, the framework enhances the accuracy of charging demand prediction and provides a robust analytical foundation for the optimal planning and management of EV charging infrastructure. Building upon the proposed methodology, this study validates its effectiveness using a real-world EV charging demand dataset collected from an urban area and further develops an intelligent charging station operation system. Extensive experiments and ablation analyses demonstrate that the proposed multimodal hypergraph based framework consistently outperforms unimodal and conventional graph-based baselines, confirming its superior capability in capturing complex spatiotemporal dependencies and improving forecasting accuracy.

Keywords: Charging Demand Forecasting, Multimodal, Hypergraph, Spatial-Temporal Data Mining

http://dx.doi.org/10.6180/jase.202611_34.052

1. Introduction

As the global energy transition accelerates, electric vehicles (EVs), as a core representative of emerging clean-energy technologies, hold immense potential for reducing carbon emissions and promoting sustainable mobility[1–3]. However, large-scale EV charging imposes substantial stress on power grids, particularly during peak-load periods, leading to greater peak-to-valley disparities that threaten grid stability and operational efficiency. Consequently, accurately forecasting charging demand to alleviate grid stress has become a critical research challenge at the intersection of smart grid management and artificial intelligence [2, 4].

Despite extensive prior research-such as the development of spatiotemporal graph neural network (STGNN) models [5, 6] and other advanced forecasting techniques-most existing studies primarily consider common influencing factors (e.g., charging duration and state of charge [7, 8]) while overlooking the deep spatiotemporal heterogeneity inherent in multimodal data. In reality, charging demand exhibits high complexity across both spatial and temporal dimensions. Spatially, charging stations in dense urban centers display markedly different usage patterns from those in peripheral residential zones, and geographically distant neighborhoods may still share similar behavioral characteristics[9]. Temporally, charging activity within the same area varies dynamically across time periods, such as morning and evening commuting peaks. Moreover, charging behavior is intricately shaped by multimodal contextual factors including holidays, unexpected events, and user mobility habits [10]. Therefore, effectively capturing the spatiotemporal heterogeneity of charging demand through comprehensive multimodal spatiotemporal data integration remains a pressing and unresolved challenge [11, 12].

To address the aforementioned challenges, this study proposes HMGNN, the first framework that integrates hypergraphs with multimodal spatiotemporal data for electric vehicle (EV) charging demand forecasting. Specifically, HMGNN introduces hypergraphs capable of representing higher-order relationships among charging stations, enabling the model to capture the complex spatiotemporal heterogeneity underlying charging behaviors. The proposed method not only leverages historical charging records but also incorporates multimodal contextual information, including urban functional zoning attributes and temporal component similarities. Through this design, HMGNN constructs a multidimensional spatiotemporal hypergraph that connects geographically non-adjacent stations exhibiting similar charging demand patterns, while also linking stations sharing temporal usage similarities. This structure guides the model to learn latent demand dependencies across both spatial and temporal dimensions, thereby substantially enhancing forecasting accuracy. Furthermore, HMGNN employs an interactive learning algorithm to model complex interdependencies among hypergraphs across multiple spatiotemporal scales. The framework focuses on effectively extracting and fusing information from hypergraphs that describe different spatial and temporal granularities. An attention mechanism is incorporated to capture intricate correlations among these spatiotemporal hypergraphs, while a gating mechanism dynamically adjusts the relative importance of each hypergraph dimension across varying operational scenarios. Through this multi-level interaction, HMGNN adaptively learns fine-grained spatiotemporal representations, achieving robust and precise charging demand prediction. The key contributions of our study can be summarized as:

- We address the challenge of accurately forecasting EV charging demand under strong spatiotemporal heterogeneity and multimodal dependencies, which conventional graph-based

models fail to capture.

- We propose HMGNN, a hypergraph-based multimodal spatiotemporal framework that models higher-order spatial and temporal correlations through hypergraph construction, attention-based interaction, and gating-driven adaptation.

- Experiments on real-world urban charging data show that HMGNN significantly outperforms unimodal and traditional graph baselines, validating the effectiveness of its multimodal and multi-scale design.

2. Related works

With the rapid expansion of spatiotemporal data dimensionality-such as urban traffic networks comprising millions of nodes and tens of thousands of time steps-traditional forecasting methods based on differential equations or statistical models (e.g., ARIMA, VAR) have shown clear limitations [7]. In contrast, graph neural networks (GNNs) have emerged as the mainstream paradigm for spatiotemporal forecasting owing to their strong nonlinear representation and feature extraction capabilities [13–15].

The Graph Convolutional Network (GCN) models spatial dependencies through adjacency matrices and employs the graph Laplacian for feature aggregation, effectively handling irregular spatial structures [16, 17]. To further capture temporal dynamics, Spatio-Temporal Graph Convolutional Networks (ST-GCN) extend this framework by introducing two-dimensional convolutional kernels that jointly process spatial and temporal information [18, 19]. ST-GCN has demonstrated remarkable performance across domains-for example, achieving an 18% reduction in MAE compared to LSTM models in urban traffic prediction [20] and effectively modeling infection diffusion patterns in epidemic propagation studies [21]. Similarly, the ST-LSTM model integrates LSTM's temporal modeling with GCN's spatial reasoning, proving effective for short-term renewable energy power forecasting [22]. More recently, Attention-Based Spatio-Temporal Graph Convolutional Networks (ASTGCN) have further enhanced forecasting accuracy by incorporating attention mechanisms into spatiotemporal graph convolutions, reducing mean absolute error by 22% compared to conventional GCNs in traffic prediction tasks [23].

To make the reference scope clearer, the related studies discussed in this work are organized around four categories: EV charging demand forecasting, spatiotemporal graph learning, hypergraph-based modeling, and time-series forecasting. General deep learning and multimodal representation studies are cited only when they provide methodological background for sequence modeling, attention-based fusion, or multimodal feature representation used in the proposed framework.

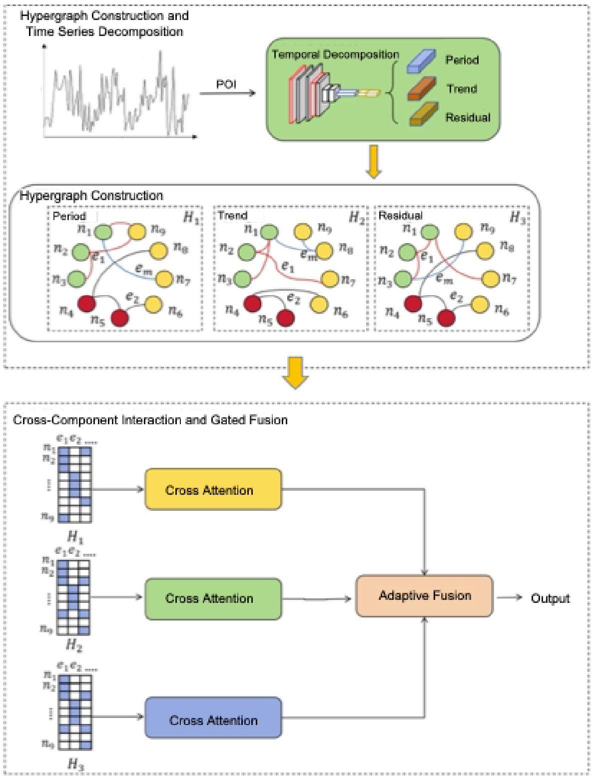

Fig. 1. Framework of HMGNN

3. Methodology

To address spatiotemporal heterogeneity and the limitations of single-modality information, this chapter proposes a multimodal charging demand forecasting framework that integrates hypergraph modeling with spatiotemporal decomposition. In the spatial dimension, the model leverages the similarity of urban functional zones (e.g., residential, commercial, and industrial areas) to construct spatial hyperedges, capturing intrinsic correlations among charging stations across diverse regions. In the temporal dimension, a multi-scale pattern association mechanism is introduced. By employing wavelet transforms to extract feature similarities at multiple time scales (hourly, daily, weekly), the model constructs temporal hyperedges that capture inter-scale dependencies in charging behavior-overcoming the inability of traditional graph methods to represent non-geographic temporal relationships. A learnable wavelet convolutional network further decomposes the charging demand sequence into trend, periodic, and residual components, enabling fine-grained temporal representation learning. To enhance integration across these decomposed components, a Cross-Component Attention (CCA) mechanism is introduced to facilitate information exchange among temporal features, while a dynamic multi-component fusion module adaptively adjusts the contribution of each component to the final prediction. We elaborate the specific designs below.

To clarify the overall workflow of HMGNN, the model pipeline is summarized as follows. Given historical charging-demand records together with multimodal contextual information such as station attributes and POI-based functional features, HMGNN first segments the station-level time series into fixed-length temporal windows and decomposes each sequence into periodic, trend, and event components. It then constructs spatial hyperedges from functional similarity among charging stations and component-specific temporal hyperedges from the decomposed temporal patterns. The resulting component hypergraphs are processed through hypergraph-based representation learning, where cross-component attention models the interactions among periodic, trend, and event representations. Finally, an adaptive fusion module integrates the learned spatial, temporal, and component-level representations to generate the final charging-demand prediction.

To further clarify the role of each modality, HMGNN uses four complementary information sources. Historical charging-demand sequences provide the primary forecasting signal and preserve short-term temporal continuity. POI-based functional features and station attributes describe the urban context of each station and are used to construct spatial hyperedges, allowing stations with similar functions to share information even when they are geographically distant. The decomposition into periodic, trend, and event components separates regular cycles, long-term evolution, and abrupt disturbances, respectively, so that each temporal pattern can be modeled with a dedicated component hypergraph. Finally, cross-component attention and adaptive fusion learn how much each modality and component should contribute under different scenarios, improving the interpretability and predictive accuracy of the framework.

3.1. Hypergraph and Node Definition

Charging demand data at electric vehicle (EV) stations exhibits pronounced spatiotemporal heterogeneity. Traditional graph-based models, which rely on adjacency matrices to describe pairwise spatial relationships, are limited to representing geographic proximity or local neighborhood associations. However, in real-world scenarios, charging demand is shaped not only by spatial distance but also by diverse contextual factors such as urban functional zoning, surrounding Points of Interest (POIs), traffic flow intensity, and user behavioral patterns. For instance, stations located within the same commercial district or residential area may exhibit highly similar demand patterns despite being geographically distant. Likewise, stations that are spatially separated may still share analogous periodic or trend patterns during certain time intervals. Furthermore, even within a single station, charging behaviors can vary significantly across different time slots of the same day due to external environmental influences-for example, increased demand during morning and evening rush hours.

Traditional pairwise relationship graphs cannot simultaneously capture these multidimensional and higher-order correlations, necessitating the use of hypergraphs for more expressive modeling. Unlike conventional graphs that connect nodes through simple edges, hypergraphs introduce hyperedges capable of linking multiple nodes at once, thereby representing joint relationships among multiple stations or time segments. To comprehensively address the spatiotemporal heterogeneity of charging demand, this study constructs two primary categories of hyperedges: spatial hyperedges and temporal hyperedges, which jointly form the theoretical foundation of the proposed hypergraph framework. Building on these categories and the inherent spatiotemporal characteristics of charging behavior, three specialized hypergraphs are designed:

- Periodic hypergraphs, capturing recurring charging patterns over regular intervals,
- Trend hypergraphs, modeling long-term variations in demand, and
- Event hypergraphs, representing transient fluctuations caused by irregular or external factors.

In the initial graph formulation, each node corresponds to a charging station, whose static properties are represented by an T_{total} attribute vector. However, static representation alone cannot reflect temporal variability in charging demand. For instance, the same station may display distinct behavioral patterns across different hours of the day. To capture this dynamic nature, each station's time series is segmented into fixed-length temporal windows, where each sub-sequence is treated as a temporal node. Specifically, the total time steps for each station are divided into T sub-sequences of equal length L :

$$\{v_{i,t} \mid t = 1, 2, \ldots, T\} \tag{1}$$

Based on the time-slice extension method, the constructed temporal nodes more precisely characterize the dynamic evolution of charging station behaviors across different time periods. By connecting these sub-sequences through temporal hyperedges, the dynamic dependencies between adjacent time windows are explicitly modeled within the hypergraph structure. This design enables the model to capture localized temporal variations and short-term fluctuations-for instance, the morning and evening charging peaks at a given station are represented by its corresponding temporal nodes for those periods. Furthermore, spatiotemporal correlations among multiple stations within the same day are captured collectively through the temporal nodes of all stations for that day. Each node's attributes are then dynamically adjusted under different hypergraph configurations to reflect temporal characteristics at multiple scales. The specific feature representations are as follows:

(1) Component Features: The original charging demand sequence is decomposed into periodic, trend, and residual components through temporal decomposition. Nodes within different hypergraphs are assigned component-specific features according to the hypergraph category-for example, nodes in the periodic hypergraph retain periodic component features, while those in the trend hypergraph retain trend components. This ensures that each hypergraph focuses on the corresponding temporal dynamics.

(2) Time Embedding Feature: To incorporate explicit temporal context, sine-cosine positional encoding is employed to map the sub-sequence index into a d-dimensional time embedding vector, allowing the model to perceive sequential ordering and periodicity. The encoding is defined as:

$$\begin{cases} PE_{(pos,2i)} = \sin\left(\frac{pos}{10000^{\frac{2i}{d_t}}}\right) \\ PE_{(pos,2i+1)} = \cos\left(\frac{pos}{10000^{\frac{2i}{d_t}}}\right) \end{cases} \tag{2}$$

After defining nodes and their attribute features in a hypergraph, it is necessary to define the superedge structure representing higher-order relationships among hypergraph nodes. Superedges are elements in hypergraphs used to represent higher-order associative relationships among multiple nodes. While edges in traditional graphs can only represent relationships between two nodes, superedges can directly represent collaborative patterns among three or more nodes. Their proper construction is crucial for uncovering complex patterns in data. A superedge is defined as a set containing any number of nodes with $e \subseteq V$, and $|e| \geq 2$. For example, a hyperedge may encompass multiple charging stations within the same time-subsequence nodes, indicating that these stations share a common load pattern during that period. Simultaneously, to prevent overfitting and reduce complexity, the total number of hyperedges in all hypergraphs is constrained to M, which means the hyperedges E should satisfy:

$$|E| = ME = e_1, e_2, \ldots, e_M \tag{3}$$

In hypergraphs, the association matrix expresses the relationship between hypergraph nodes and hyperedges. While traditional graph models use adjacency matrices $A \in \mathbb{R}^{N\times N}$ to represent pairwise relationships between nodes, hypergraphs employ association matrices to represent the relationship between nodes and hyperedges. However, in traditional supergraph association matrices, the association matrix is defined as an N × M two-dimensional matrix. Here, N denotes the number of charging stations, and M denotes the number of superedge types. This form cannot represent the characteristics of spatiotemporal data. This study extends it to a three-dimensional tensor $H \in \mathbb{R}^{N\times T\times M}$, where T denotes the number of sub sequences. Through this three-dimensional association matrix, it is possible to clearly represent which hyperedges each node belongs to under different time sub-sequences, thereby fully describing the structure of the hypergraph.

3.2. Series Decomposition and Hypergraph Construction

After completing hypergraph and node definitions, this study conducted deeper exploration of the raw data through time series decomposition and component hypergraph construction. The core strategy involved decomposing the original spatiotemporal data into three independent components- periodicity, trend, and events-and constructing hypergraphs for each. This approach aims to capture data patterns across different time scales while effectively isolating noise interference from each component, thereby providing a solid foundation for subsequent precise forecasting and analysis.

Raw spatiotemporal data typically contains multiple intertwined patterns of variation across different timescales, making direct modeling and analysis challenging. The core idea of time series decomposition is to separate these complex patterns for individual processing. The periodic component reflects short-term repetitive patterns in the data, the trend component captures long-term directional changes, while the residual component, acting as an event component, captures sudden anomalous events within the data. By decomposing the data into these three independent components, the characteristics of each component can be observed and analyzed more clearly, enabling the construction of a dedicated hypergraph for each component. Traditional time series decomposition methods (such as STL decomposition) can partially separate these components but lack adaptive learning capabilities for data features. This study employs a learnable wavelet convolutional network for time series decomposition. Its core principle leverages the multi-resolution analysis properties of wavelet transforms combined with the powerful learning capabilities of convolutional neural networks to adaptively extract three independent components-periodicity, trend, and events-from raw data. Wavelet transform enables signal decomposition at multiple scales, making it suitable for capturing information across different time horizons. Convolutional neural networks automatically learn complex patterns and features within data, enhancing the flexibility and accuracy of the decomposition process. The learnable wavelet convolutional network comprises multiple wavelet convolutional layers. Each layer performs convolution on the input signal using learnable wavelet filters, achieving signal decomposition across various scales. Let the input time series be $X = \{x_1, x_2, \ldots, x_n\}$, at the 1-th convolution layer, the initial $X^{(0)} = X$, and the learnable wavelet filters be $W^{(l)}$, this operation can be formulated as:

$$Y^{(l)} = X^{(l-1)} * W^{(l)} \tag{4}$$

where * denotes the convolution operation. The output Y(1) is further separated into the approximation component A(l) and the detail component D(l). The approximation component captures low-frequency information and is used to represent the long-term trend, whereas the detail component captures high-frequency variations related to periodic fluctuations and event-driven changes. In the learnable wavelet convolutional network, the wavelet filter parameters W(1) are optimized through backpropagation, allowing the decomposition process to adapt to the characteristics of charging-demand data. After multiple wavelet-convolution layers, the final approximation component is used as $\boldsymbol{X}_{\text{trend}}$, while the detail components from different layers are combined to obtain the periodic component $\boldsymbol{X}_{\text{period}}$ and the event component $\boldsymbol{X}_{\text{event}}$.

$$X_{\text{period}} = \sum_{j=1}^{m} \text{Reconstruct}\left(D^{(i_j)}\right. \tag{5}$$

$$X_{\text{event}} = X - X_{\text{period}} - X_{\text{trend}} \tag{6}$$

For each component obtained through decomposition, this study employs distinct similarity measures and construction techniques to build dedicated hypergraphs. The hyperedges of each component hypergraph comprise spatial hyperedges and component hyperedges. Spatial hyperedges are constructed based on POI information from each station, serving as static prior information appended to each component hypergraph. Component hyperedges are constructed based on the temporal information represented by each component, characterizing data patterns across different time scales while effectively isolating noise interference from each component.

For reproducibility, the thresholds used in component hyperedge construction are selected on the validation set rather than manually fixed. Specifically, the DTW threshold for periodic similarity and the trend-similarity threshold are searched from candidate grids, and the setting with the lowest validation RMSE is adopted while keeping the total number of component hyperedges consistent with the designed value. The spatial hyperedges are determined by POI-based functional-zone clustering, where the number of spatial hyperedges is fixed according to the functional-zone categories of charging stations.

Next, during the construction of spatial hyperedges, the embedding vector s_i of each charging station is used to measure static similarity. This study employs cosine similarity to effectively capture the similarity of charging stations across functional zone:

$$S_{ij} = \frac{s_i \cdot s_j}{\|s_i\| \left\|s_j\right\|} \tag{7}$$

As S_{ij} approaches 1, it indicates that v_i and v_j are more similar in functional attributes. In spatial hyperedge partitioning, the k-means clustering algorithm is applied to the similarity matrix S to divide all charging stations into M_S clusters. Let the clustering results be C1, C2, ..., CMS. Then, each spatial hyperedge is defined as:

$$s_k^{(S)} = \{v_i \mid v_i \in C_k\}, k = 1, 2, \ldots, M_S \tag{8}$$

To ensure the total number of hyperedges meets design requirements, this study sets the total number of hyperedges M for each component hypergraph to be:

$$M = M_c + M_s \tag{9}$$

Next, we introduce the construction process for each component hypergraph.

Periodic Hypergraph: The objective of the periodic hypergraph is to capture cyclical similarities among nodes within short time periods. In charging station demand data, such periodic patterns may manifest as similar variations in charging demand occurring during the same time intervals each day. By constructing the periodic component hypergraph, clusters of charging stations with similar periodic patterns can be identified, thereby enhancing our understanding and prediction of demand fluctuations at these stations. During periodic hypergraph construction, the Dynamic Time Warping (DTW) algorithm is applied to each time-series node to calculate similarity between periodic timeseries nodes. The DTW algorithm handles scaling and warping of time series along the time axis, enabling more accurate measurement of similarity between two periodic subsequences. Within the periodic patterns of charging stations, local periodic variations in charging demand exist, such as during holidays or distinct morning and evening peak hours. Therefore, it is necessary to calculate the periodic similarity between different time series nodes starting from each charging station's time series node to construct hyperedges. Given two period temporal node i and j, their DTW distance is employed to calculate the distance between sub-sequences:

$$d_{ij} = \min_{w \in \mathcal{W}} \left\{ \sqrt{\sum_{(k,l) \in w} \left(\mathrm{p}_i(k) - \mathrm{p}_j(l)\right)^2} \right. \tag{10}$$

Here, W denotes all possible alignment paths, and (k, l) represents a matching pair within the alignment path $W.k$ indicates the index of a time point in the subsequence p_i, while l denotes the index of the time point in the subsequence p_j that aligns with k. When d_{ij} falls below the preset threshold ep, vi and vj are considered to exhibit high similarity in their periodic behavior and are grouped into the same hyperedge. The construction yields a periodic hyperedge set $E^{(P)} = \left\{e_1^{(P)}, e_2^{(P)}, \ldots, e_{M_c}^{(P)}\right\}$,where Mc denotes the number of periodic hyperedges. To ensure alignment in subsequent cross-attention computations, each hypergraph maintains an identical number of hyperedges.

Notation clarification: for two temporal nodes i and j, D(i,j) denotes their Dynamic Time Warping distance, and $\boldsymbol{d_{ij}}$ is used interchangeably to represent the same pairwise periodic-distance value in the subsequent description. A smaller D (i,j) indicates higher similarity between the periodic subsequences of the two nodes. The threshold epsilon $_{\boldsymbol{p}}$ is selected on the validation set, and two nodes are assigned to the same periodic hyperedge when $\mathrm{D}(\mathrm{i}, \mathrm{j}) <$ epsilon $_p$.

Trend Hypergraph: The objective of the trend hypergraph is to extract long-term trend information from charging load data, capturing the consistency of different charging stations along these trends. Within charging station demand data, long-term trends may be influenced by factors such as urban development and seasonal variations. By constructing a trend component hypergraph, charging stations with potentially similar long-term trends can be identified. During trend hypergraph construction, trend component data typically exhibits smooth and stable characteristics. To reflect the rate of increase or decrease in charging demand over long-term trends, this study employs linear regression to fit the trend component of each charging station. The resulting slope s serves as the trend feature, while an exponential decay function to measure the difference in trend slopes between two stations. The slope differences form a similarity matrix. Finally, by setting a similarity threshold et, two charging stations are assigned to the same superedge when their trend similarity exceeds this threshold. Since trend characteristics manifest over longer time windows when constructing trend superedges. For each sub-sequence of a charging station, linear regression is performed on the timeseries nodes to obtain the trend slope as the overall trend characteristic of the charging station. Hyperedges are partitioned based on the similarity of trend features between different charging stations, assigning all time-series nodes of similar stations to the same hyperedge. Let the overall trend component of each charging station be denoted as Ti, and its corresponding time-series nodes be recorded as:

$$T_i^{(t)} = T_i^{(t)}(1), T_i^{(t)}(2), \ldots, T_i^{(t)}(L) \tag{11}$$

This study employs linear regression fitting on each window to calculate the trend slope $\beta(t)\ i$ within that window, thereby quantifying the local long-term trend and trend:

$$T_i^{(t)}(\tau) \approx \beta_i^{(t)}\tau + c_i^{(t)}\tau, \tau = 1, 2, \ldots, L \tag{12}$$

$$\beta_i^{(t)} = \frac{\sum_{\tau=1}^{L}(\tau - \bar{\tau})\left(T_i^{(t)}(\tau) - \bar{T}_i^{(t)}\right)}{\sum_{\tau=1}^{L}(\tau - \bar{\tau})^2} \tag{13}$$

$$\bar{\tau} = \frac{1}{L}\sum_{\tau=1}^{L}\tau, \bar{T}_i^{(t)} = \frac{1}{L}\sum_{\tau=1}^{L}T_i^{(t)}(\tau) \tag{14}$$

Eventually, we get the feature vector of station i :

$$\beta_i = \left[\beta_i^{(1)}, \beta_i^{(2)}, \ldots, \beta_i^{(W)}\right] \tag{15}$$

Here, W denotes the number of time-series nodes, i.e., the total number of sub-sequences. Since longterm trends within charging stations are typically stable, this study aims to construct features reflecting the overall station trend rather than relying solely on local trends within each window. To achieve this, the study further aggregates the local trend slopes at each time node within the charging station. Ultimately, the global trend representation for charging station i is expressed by a scalar $\beta(i)$, reflecting the long-term rate of change for that station throughout the entire observation period:

$$\beta_i = \frac{1}{W}\sum_{t=1}^{W}\beta_i^{(t)} \tag{16}$$

Since trend superedges are constructed at the charging station level, all temporal subnodes under the same charging station share identical trend superedge information. The construction process is as follows: First, a similarity matrix is constructed by calculating the trend similarity for all charging stations i and j. Then, the trend similarity matrix is grouped based on a preset threshold ϵt, assigning charging stations with similar trends to the same group. Let the resulting set of trend superedges be: $E^T = \left\{e_1^{(T)}, e_2^{(T)\prime}, \ldots, e_{M_c}^{(T)\prime}\right\}$, where Mc is the number of component superedges.

Event Hypergraph: The purpose of constructing an event component hypergraph is to capture sudden abnormal events in the data. In charging station demand data, these abnormal events may be caused by special activities, equipment failures, or other factors. By constructing an event component hypergraph, fluctuations in charging demand caused by short-term anomalies or noise can be captured, isolating the interference of sudden events on the model. This study employs the K-means clustering algorithm to partition set V. The objective of the K-means algorithm is to divide all event subsequences into Mc clusters such that the subsequences within each cluster are as similar as possible. Its objective function is:

Notation clarification: in the K-means objective used for event hypergraph construction, $\boldsymbol{c_j}$ denotes the centroid of the j-th event cluster, and each event subsequence is assigned to the nearest centroid according to Euclidean distance. This notation is used consistently in the subsequent optimization and hyperedge assignment steps.

$$\min_{C_1,\ldots,C_{M_c}} \sum_{j=1}^{M_c}\sum_{E\in C_j}\left\|E - c_j\right\|^2 \tag{17}$$

The component hypergraphs constructed in this section are not independent of each other. They collaborate throughout the model to jointly support final predictions and analysis. The cyclical component hypergraph provides information on short-term periodic patterns, the trend component hypergraph reflects long-term changing trends, while the event component hypergraph captures sudden anomalies. The constructed component hypergraph will furnish rich structured features for subsequent cross-attention mechanisms and dynamic gating fusion, thereby enhancing the overall model's performance and interpretability.

3.3. Cross-Component Attention

After completing the time-series decomposition and constructing the component hypergraphs, we examine the hypergraph structures of the three components: periodic, trend, and event. However, the information contained within these component hypergraphs is not isolated; complex interactions and correlations exist between them. For instance, a sudden anomalous event may influence the pattern of the periodic component in the short term while also potentially perturbing the long-term trend component. Therefore, to fully extract this cross-component information and enhance the model's understanding and predictive capabilities for spatiotemporal data, this study designs an effective cross-component information interaction algorithm. After constructing the periodic, trend, and event hypergraphs, we obtained their respective node feature matrices $X(c)$ and corresponding hyperedge structures $H(c)$, where $c \in \{P, T, E\}$. Each hypergraph, processed independently, effectively captures both local and global relationships within its component. However, information across different components remains relatively isolated. In reality, these three components are complementary and exert spatiotemporal contextual influences upon each other. To address this, this study designs a cross-hyperedge attention mechanism. This mechanism enables the target component (e.g., the periodic component) to dynamically acquire complementary information from other components (e.g., the trend and event components). This process updates its feature representation, allowing it to dynamically adapt to spatiotemporal contextual changes across different scenarios. For each component c, the corresponding node feature matrix $X^{(c)} \in \mathbb{R}^{N\times T\times d}$ (where N is the number of charging stations, T is the number of partitioned temporal nodes, and d is the feature dimension), along with the hyperedge association matrix $H^{(c)} \in \mathbb{R}^{N\times T\times M}$ (where M denotes the number of hyperedges), concatenate $X(c)$ and H(c) along the channel dimension to obtain the concatenated feature representation:

Notation clarification: c in {P, T, E} indexes the periodic, trend, and event components, respectively. X(c) denotes the node feature matrix of component c, H(c) denotes its hyperedge incidence matrix, N is the number of charging stations, T is the number of temporal nodes, d is the feature dimension, and M is the number of hyperedges. These symbols are used consistently in the following equations and attention computations.

$$X^{(c)}_{\text{concat}} = \text{Concat}\left(X^{(c)}, H^{(c)}\right) \in \mathbb{R}^{N\times T\times d_0}, d_0 = d + M \tag{18}$$

For the cross-attention module, the query matrix Q(c) is calculated as:

$$Q^{(c)} = X^{(c)}_{\text{concat}} W^{(c)}_Q \in \mathbb{R}^{N\times T\times d_a} \tag{19}$$

Then, the key matrix K(c) and value matrix V(c) are calculated as:

$$K^{(c')} = X^{(c')}_{\text{concat}} W^{(c')}_K \in \mathbb{R}^{N\times T\times d_a} \tag{20}$$

$$V^{(c')} = X^{(c')}_{\text{concat}} W^{(c')}_V \in \mathbb{R}^{N\times T\times d_a} \tag{21}$$

$$A^{(c)} = \text{softmax}\left(\frac{Q^{(c)} \cdot \left(\tilde{K}^{(\text{others})}\right)^T}{\sqrt{d_a}}\right) \tag{22}$$

For each component c $\in \{\text{P}, \text{T}, \text{E}\}$, the above cross-attention layer is applied to obtain the updated component representation Z(c).

Subsequently, the dynamic gating module adaptively aggregates information from the periodic, trend, and event components.

$$g^{(c)} = \sigma\left(\text{FC}\left(Z^{(c)}_{\text{updated}}\right)\right) \in \mathbb{R}^{N\times T\times d_a} \tag{23}$$

$$Z_{\text{fused}} = \sum_{c\in\{P,T,E\}} g^{(c)} \odot Z^{(c)}_{\text{updated}} \tag{24}$$

In the above equations, c denotes the component type, c $\in \{\text{P}, \text{T}, \text{E}\}$; $\text{Q}(\text{c}), \text{K}(\text{c})$, and $\text{V}(\text{c})$ denote the query, key, and value matrices of component c, respectively; Z(c) represents the component representation after cross-component attention; T$'$ denotes the target prediction time step; and Wout denotes the output-layer weight matrix. The final prediction is compared with the actual charging demand y using Mean-Square Error (MSE) as the training loss:

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2 \tag{25}$$

4. Experiments

This section details the experimental design and analysis of results. The study utilizes charging order data from Xiamen City spanning 2022 to 2024. Subsequently, the paper sequentially introduces the dataset, data cleaning methods, experimental setup, and the design and analysis of relevant experiments.

4.1. Datasets

The primary dataset used in this study "Charging Order Data," containing every charging order record from all charging stations between January 2022 and September 2024. Each record comprises 15 fields. These records provide detailed insights into the usage patterns of charging stations in the city from 2022 to 2024. The 15 fields in the dataset comprehensively cover the specifics of each charging order. The detailed fields and their meanings are listed in Table 1.

During the data processing phase, the primary task is to clean and filter redundant data. The raw dataset contained incomplete charging station information or stations that did not meet public charging station standards, including invalid stations, dedicated stations exclusively for public buses, and private stations. This study focuses on the usage patterns of publicly accessible charging stations.

Based on predefined screening criteria, stations failing to meet the requirements were excluded. Ultimately, 170 invalid stations were removed, leaving 544 public stations. The filtered station data covers all districts of the city, ensuring regional representativeness and data reliability. Subsequently, the temporal granularity of the data required conversion processing. After restructuring and saving in a structured format, this study obtained a dataset integrating charging demand information at hourly intervals, containing only valid public charging stations. Each row represents the total charging demand at a specific time point for each valid public charging station. Furthermore, to characterize the spatiotemporal heterogeneity that transcends geographical distance constraints between charging stations, this study collected functional zoning attributes for each station's district. By calling the Map Application Programming Interface (API), the study obtained POI confidence values within a one-kilometer radius of each charging station, reflecting the functional zoning distribution within the area. By associating the API-obtained POI data with the original order data through charging station IDs and geographic coordinates, this study established the data foundation for constructing spatial superedges and overcoming geographical distance constraints.

4.2. Experimental Settings

In this study, the experimental environment was configured as follows. The hardware setup comprised an Intel(R) Xeon(R) Gold 6148 CPU with 20 cores (1 core utilized for the main application) and an NVIDIA GeForce RTX 3090 GPU with 24 GB of VRAM.

The main hyperparameter settings are summarized as follows. The total number of spatial hyperedges corresponds to the functional zones of the charging stations, totaling 22, while the overall number of hyperedges is set to 50 , including 28 component hyperedges. The Adam optimizer was adopted with an initial learning rate of 0.001 . The wavelet convolutional network comprises three layers ($L = 3$). The projection dimension of the attention mechanism was set to 64 , balancing model capacity and computational cost. Each network layer uses an embedding dimension of 128, and the model was trained for 100 epochs. For temporal modeling, the historical input window spans three consecutive days (72 hours), divided into 12 temporal sub-sequences based on 6 -hour intervals ($T = 12$). The model predicts charging demand for each station over the subsequent 24 hours. To avoid information leakage in time-series forecasting, all samples were sorted chronologically before splitting. The first 80% of the time-ordered samples were used for model training, and the remaining 20% were used for testing. A validation subset was further separated from the tail of the training period for hyperparameter selection and early stopping. The batch size was set to 64, the early-stopping patience was set to 10 epochs, and the random seed was fixed to 2024. Unless otherwise specified, the reported results are averaged over five independent runs.

Implementation details for reproducibility: the batch size was set to 64 , the random seed was fixed to 2024, and early stopping was adopted with a patience of 10 epochs based on validation performance. Unless otherwise specified, each experiment was repeated five times with different random initializations, and the average performance was reported.

4.3. Main Results

To validate the effectiveness of the hypergraph-based charging demand prediction method, this paper designs a series of comparative experiments. The proposed method focuses on constructing component hypergraphs and interacting with cross-component dynamic information to fully capture the periodicity, trends, and event characteristics of charging station demand. It utilizes spatial superedges as static prior supplements to generate global fusion features for prediction, thereby learning the spatiotemporal heterogeneity of charging station demand. To verify this design's effectiveness, this study compares the proposed method with multiple models, including the Heterogeneity-Informed Spatiotemporal Meta-Network (HIMNET), DM-STGCN, GRU, LSTM, Transformer, and the Propagation Delay-Aware Dynamic Long-Range Transformer (PDFormer).

Baseline selection rationale: the comparison set is de-

Table 1. Charging Order Data Field Details

No.	Field	Description
1	Order ID	ID of the charging order
2	Charging gun ID	ID of the charging gun used by the user
3	Charging Pile ID	ID of the charging pile to which the charging gun belongs
4	Charging Station ID	ID of the charging station to which the charging pile belongs
5	Charging Station Name	Name of the charging station on the map
6	Operator ID	ID of the enterprise or institution directly responsible for the overall construction, maintenance,
7	Operator Name	Official name or registered brand name of the operator
8	Region	Municipal district where the charging station is located
9	Agent Operator ID	ID of the third-party agency entrusted by the main operator to manage the charging station's daily operations
10	Agent Operator Name	Official name or registered brand name of the agent operator
11	harging Duration (hours)	Total charging duration for this charging order
12	Start Time	Time when the charging order began recording charging activity
13	End Time	Time when the charging order concluded charging
14	Charge Amount (kWh)	Total electricity consumed during the entire charging process
15	End Time	Time when the charging order concluded charging

signed to cover representative forecasting paradigms, including recurrent sequence models (GRU and LSTM), a vanilla global self-attention model (Transformer), a recent long-range spatiotemporal Transformer (PDFormer), and advanced graph-based spatiotemporal predictors (HIMNET and DM-STGCN). This combination allows us to evaluate whether the gains of HMGNN come from multimodal fusion alone or from the joint modeling of higher-order hypergraph relations and decomposed temporal dynamics. Because publicly available hypergraph forecasting baselines tailored to station-level EV charging demand with multimodal contextual inputs are still limited, we strengthen the discussion against strong Transformer-based and graph-based competitors in the revised manuscript.

We use the standard RMSE, MAPE and MAE as evaluation metrics. The compared results are shown in Table 2.

Table 2. Results of Compared Methods

Model	RMSE	MAPE	MAE
HIMNET	3.426	39.45%	2.391
DM-STGCN	3.615	43.81%	2.557
GRU	5.693	44.45%	3.032
LSTM	5.801	47.11%	3.144
Transformer	6.823	53.85%	3.865
PDFormer	3.557	41.83%	2.457
Our Method	3.253	36.53%	2.197

Based on experimental results, the proposed model demonstrates high predictive accuracy across all metrics. Specifically, the model achieves an RMSE of only 3.253 , the lowest among all models. This indicates that the hypergraph construction module effectively uncovers complex higher-order relationships among multi-source data, accurately captures data trend changes, reduces model prediction bias, and minimizes the mean squared error between predicted and actual values. Other models, such as HIMNET, exhibit limitations in handling the complex spatiotemporal relationships within this dataset. DM-STGCN is influenced by the quality of feature data, resulting in an incomplete characterization of charging demand. GRU and LSTM models, overly focused on time series, fail to adequately mine spatial and other information. Transformers, emphasizing global features, struggle to capture local information effectively, leading to suboptimal prediction performance. However, the proposed model shows relatively minor improvements in RMSE compared to other metrics, suggesting it may be more susceptible to data noise. In noisy scenarios, the hypergraph construction process may be disrupted, causing the model to extract biased data relationships and resulting in reduced stability under noisy conditions. Regarding MAPE, the proposed model achieves significantly reduced errors. Compared to advanced models

Compared with HIMNET, DM-STGCN, and PDFormer, the proposed model reduces RMSE by 2.92%, 7.28%, and 5.3%, respectively, demonstrating its advantage in capturing complex spatiotemporal dynamics at charging stations. The lower errors indicate that component-wise decomposition helps separate periodic, trend, and event-driven patterns, while hypergraph modeling further captures higher-order spatial and temporal correlations. In contrast, DM-STGCN has limited ability to distinguish component-specific temporal patterns, and PDFormer, although effective for spatiotemporal fusion, remains constrained in modeling higher-order station relationships. On the MAE metric, HMGNN achieves an error of 2.197 , reducing the absolute deviation by 0.835 and 0.947 compared with GRU and LSTM, respectively. These results verify the effectiveness of combining temporal decomposition, component hypergraphs, and cross-component interaction for charging-demand forecasting.

4.4. Parameter Analysis

To thoroughly investigate the impact of key hyperparameters on the performance of the proposed charging demand forecasting model and to validate its robustness and stability under different parameter settings, a series of parameter experiments were designed. These experiments primarily focused on two aspects: the configuration of the total number of hyperedges M and the number of learnable layers L in the convolutional network. Through systematic parameter sensitivity experiments, this study aims to assess how different parameters influence model performance.

First, regarding the setting of the total hyperedge count M, this parameter influences the model's ability to capture higher-order spatio-temporal relationships within the data. Since the hyperedge count in this model comprises dynamic component hyperedges Mc and spatial hyperedges Ms, Ms is fixed at 22 based on the number of functional zone codes for charging stations. Mc values of 20, 28, and 36 were tested experimentally, yielding the following results:

Table 3. Performance Under Different Hyperedge Numbers

SuperEdge Number	RMSE	MAPE	MAE
$M_c = 20$	3.521	39.12%	2.456
$M_c = 28$	3.253	36.53%	2.197
$M_c = 36$	3.311	37.20%	2.240

To thoroughly investigate the impact of key hyperparameters on the performance of the proposed charging demand forecasting model and to validate its robustness and stability under different parameter settings, a series of parameter experiments were designed. These experiments primarily focused on two aspects: the configuration of the total number of hyperedges M and the number of learnable layers L in the convolutional network. Through systematic parameter sensitivity experiments, this study aims to assess how different parameters influence model performance.

First, regarding the total hyperedge count M, the results indicate that when Mc = 28, the model achieves the lowest values across RMSE, MAPE, and MAE metrics. This suggests an optimal ratio between dynamic and spatial hyperedges at this setting, effectively balancing the integration of temporal dynamic information and static prior knowledge. Conversely, at Mc = 20, the limited number of dynamic hyperedges restricts the model's ability to capture intricate dynamic relationships. The relatively simple hypergraph structure fails to fully extract complex patterns within the data. In the charging demand forecasting scenario, this limitation impairs the model's capacity to identify subtle variations across different time scales and express certain critical higher-order interactions, thereby increasing the discrepancy between predicted and actual values. At Mc = 36, while the model captures more granular dynamic relationships, the excessive number of hyperedges introduces redundant information and noise. This reduces the model's generalization capability, leading to a slight deterioration in predictive performance.

Secondly, in experiments with learnable L layers for the wavelet convolutional network, structures with fewer than 2 layers are typically too simplistic to adequately capture subtle variations in complex time-series data. Conversely, structures exceeding 4 layers tend to introduce noise, increase model parameters and training time, thereby elevating the risk of overfitting. Therefore, this study compares the performance of 2-layer, 3-layer, and 4-layer architectures. Increasing the number of layers in a wavelet network enables finer decomposition of data, separating more components across different frequencies. Shallower networks may only distinguish between high- and low-frequency information, while deeper networks can further subdivide these components, allowing the model to capture more subtle variations. Simultaneously, wavelet networks with varying layer counts capture features across different temporal scales. As layers increase, the model acquires multiscale information from short to long cycles, enabling a more comprehensive understanding of charging demand patterns. The results of the parameter experiments are shown in Table 4.

Table 4. Performance Under Different Wavelet Network Layers

Layer Number	RMSE	MAPE	MAE
L = 2	3.618	40.27%	2.539
L = 3	3.253	36.53%	2.197
L = 4	3.286	36.61%	2.131

5. Conclusion

This study proposed HMGNN, a multimodal spatiotemporal charging demand forecasting framework that integrates hypergraph modeling with temporal decomposition to address the inherent heterogeneity of electric vehicle (EV) charging data. By constructing spatial and temporal hyperedges to capture higher-order correlations among charging stations and employing waveletbased temporal decomposition with cross-component attention, the model effectively learns complex dependencies across multiple scales. Extensive experiments on real-world urban charging datasets demonstrated that HMGNN consistently outperforms unimodal and traditional graph-based baselines in both forecasting accuracy and stability. The results confirm the effectiveness of combining hypergraph representation learning and multimodal information fusion for modeling dynamic, nonlinear charging behaviors.

This work provides a new perspective for fine-grained, data-driven planning and management of EV charging infrastructure, contributing to the advancement of smart grid intelligence and sustainable energy systems. In practical deployment, more accurate station-level demand forecasts can support charging-station maintenance scheduling, peak-load management, transformer and feeder capacity planning, and dynamic allocation of charging resources across urban functional zones. Such forecasts can also help operators identify stations that may experience demand surges and coordinate charging services with grid operation requirements. Future research will explore adaptive hypergraph evolution mechanisms, cross-city transfer learning, and integration with real-time grid operation data to further enhance scalability and generalization in practical applications.

Acknowledgement

This paper is supported by the Science and Technology Project of State Grid Shaanxi Electric Power Co., LTD. (5226TX25000H).

References

[1] Z. Deng, F. Deng, M. Yang, X. Guo, J. Si, B. Yu, and B. Feng. "Transformer-Based Multi-Industry Electricity Demand Forecasting". In: *International Conference on Applied Intelligence*. Singapore: Springer Nature Singapore, 2023. DOI: 10.1007/978-981-97-0827-7_3.

[2] H. Qu, H. Kuang, Q. Wang, et al., (2024) *"A Physics-Informed and Attention-Based Graph Learning Approach for Regional Electric Vehicle Charging Demand Prediction"* **IEEE Transactions on Intelligent Transportation Systems**: DOI: 10.1109/TITS.2024.3401850.

[3] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan, (2019) *"Attention Based Spatial-Temporal Graph Convolutional Networks for Traffic Flow Forecasting"* **Proceedings of the AAAI Conference on Artificial Intelligence** *33*(1): 922–929. DOI: 10.1609/aaai.v33i01.3301922.

[4] X. L. Li and P. Liang. "Prefix-Tuning: Optimizing Continuous Prompts for Generation". In: *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*. Online: Association for Computational Linguistics, 2021, 4582–4597. DOI: 10.18653/v1/2021.acl-long.353.

[5] D. P. Kingma and J. Ba. *Adam: A Method for Stochastic Optimization*. 2014. arXiv: 1412.6980 [cs.LG].

[6] Y.-C. Chen, L. Li, L. Yu, A. El Kholy, F. Ahmed, Z. Gan, Y. Cheng, and J. Liu. "UNITER: Universal Image-Text Representation Learning". In: *European Conference on Computer Vision*. 2020, 104–120. DOI: 10.1007/978-3-030-58577-8_7.

[7] L. Han, B. Du, L. Sun, Y. Fu, Y. Lv, and H. Xiong. "Dynamic and Multi-Faceted Spatio-Temporal Deep Learning for Traffic Speed Forecasting". In: *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2021, 547–555. DOI: 10.1145/3447548.3467275.

[8] W. Fedus, B. Zoph, and N. Shazeer, (2022) *"Switch Transformers: Scaling to Trillion Parameter Models with Simple and Efficient Sparsity"* **Journal of Machine Learning Research** *23*(120): 1–39.

[9] S. Yu, F. Xia, S. Li, et al., (2023) *"Spatio-Temporal Graph Learning for Epidemic Prediction"* **ACM Transactions on Intelligent Systems and Technology** *14*(2): 1–25. DOI: 10.1145/3579815.

[10] M. Chen, H. Yang, S. Li, et al., (2024) *"STAA: Spatio-Temporal Alignment Attention for Short-Term Precipitation Forecasting"* **IEEE Geoscience and Remote Sensing Letters**: DOI: 10.1109/LGRS.2024.3431099.

[11] B. Lim, S. O. Arik, N. Loeff, and T. Pfister, (2021) *"Temporal Fusion Transformers for Interpretable Multi-Horizon Time Series Forecasting"* **International Journal of Forecasting** *37*(4): 1748–1764. DOI: 10.1016/j.ijforecast.2021.03.012.

[12] W. Liao, B. Bak-Jensen, J. R. Pillai, et al., (2022) *"Short-Term Power Prediction for Renewable Energy Using Hybrid Graph Convolutional Network and Long Short-Term Memory Approach"* **Electric Power Systems Research** ***211***: 108614. DOI: 10.1016/j.epsr.2022.108614.

[13] Q. Liu, Y. Zeng, L. Chen, et al. "Social-Aware Optimal Electric Vehicle Charger Deployment on Road Network". In: *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. 2019. DOI: 10.1145/3347146.3359382.

[14] J. Li, R. R. Selvaraju, A. D. Gotmare, S. Joty, C. Xiong, and S. C. H. Hoi. *Align before Fuse: Vision and Language Representation Learning with Momentum Distillation*. 2021. arXiv: 2107.07651 [cs.CV].

[15] C. Jia, Y. Yang, Y. Xia, Y.-T. Chen, Z. Parekh, H. Pham, Q. V. Le, Y. Sung, Z. Li, and T. Duerig. "Scaling Up Visual and Vision-Language Representation Learning with Noisy Text Supervision". In: *Proceedings of the 38th International Conference on Machine Learning*. 2021, 4904–4916.

[16] L. Liu and L. Chen, (2023) *"Research Progress about Deep Reinforcement Learning"* **Mechatronic Systems and Control *51***(4): 210–217. DOI: 10.2316/J.2023.201-0371.

[17] M. Xu. *Text2Vec: A Tool for Text to Vector*. 2022.

[18] Z. A. Sahili and M. Awad. *Spatio-Temporal Graph Neural Networks: A Survey*. 2023. arXiv: 2301.10569 [cs.LG].

[19] Y. Zhang, Y. Li, L. Cui, et al. *Siren's Song in the AI Ocean: A Survey on Hallucination in Large Language Models*. 2023. arXiv: 2309.01219 [cs.CL].

[20] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. Cambridge, MA: MIT Press, 2016.

[21] A. Graves. "Long Short-Term Memory". In: *Supervised Sequence Labelling with Recurrent Neural Networks*. Berlin: Springer, 2012, 37–45. DOI: 10.1007/978-3-642-24797-2_4.

[22] Y. Li, R. Yu, C. Shahabi, and Y. Liu. *Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting*. 2017. arXiv: 1707.01926 [cs.LG].

[23] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio. *Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling*. 2014. arXiv: 1412.3555 [cs.NE].