



ARTICLE

Interactive Dynamic Graph Convolution with Temporal Attention for Traffic Flow Forecasting

Zitong Zhao¹, Zixuan Zhang² and Zhenxing Niu^{3,*}

¹Mechanical Engineering College, Xi'an Shiyou University, Xi'an, 710065, China

²School of Transportation Engineering, Chang'an University, Xi'an, 710064, China

³School of Highway, Chang'an University, Xi'an, 710064, China

*Corresponding Author: Zhenxing Niu. Email: niuzhenxing@chd.edu.cn

Received: 30 June 2025; Accepted: 21 August 2025; Published: 10 November 2025

ABSTRACT: Reliable traffic flow prediction is crucial for mitigating urban congestion. This paper proposes Attention-based spatiotemporal Interactive Dynamic Graph Convolutional Network (AIDGCN), a novel architecture integrating Interactive Dynamic Graph Convolution Network (IDGCN) with Temporal Multi-Head Trend-Aware Attention. Its core innovation lies in IDGCN, which uniquely splits sequences into symmetric intervals for interactive feature sharing via dynamic graphs, and a novel attention mechanism incorporating convolutional operations to capture essential local traffic trends—addressing a critical gap in standard attention for continuous data. For 15- and 60-min forecasting on METR-LA, AIDGCN achieves MAEs of 0.75% and 0.39%, and RMSEs of 1.32% and 0.14%, respectively. In the 60-min long-term forecasting of the PEMS-BAY dataset, the AIDGCN out-performs the MRA-BGCN method by 6.28%, 4.93%, and 7.17% in terms of MAE, RMSE, and MAPE, respectively. Experimental results demonstrate the superiority of our pro-posed model over state-of-the-art methods.

KEYWORDS: Traffic flow prediction; interactive dynamic graph convolution; graph convolution; temporal multi-head trend-aware attention; self-attention mechanism

1 Introduction

The rapid expansion of urban transportation places escalating pressure on road systems, resulting in intermittent congestion [1,2]. Consequently, researchers have focused on developing intelligent transportation systems (ITS) [3–5]. Precise traffic prediction forms a vital basis for ITS, offering benefits such as enhanced transportation efficiency, alleviated congestion, reduced accidents, minimized energy consumption, and mitigation of environmental pollution [6]. Traffic spatiotemporal data represents a time series recorded at fixed intervals by sensors deployed on road nodes, providing an intuitive reflection of the traffic network's actual situation [7,8]. Urban road traffic conditions are influenced by numerous external factors, resulting in increased uncertainty in the temporal dimension of traffic flow [9]. Resolving this issue is essential for precise traffic prediction. Consequently, effective time series analysis techniques are commonly applied to predict traffic flow.

In general, traffic flow forecasting methods can be broadly classified into three categories: statistics, machine learning, and deep learning. The statistical approach aims to comprehend the temporal characteristics of traffic flow based on the theoretical foundations of statistical analysis. Statistical models include the historical average model, the autoregressive moving average model, and the vector autoregressive



model [10,11]. For instance, Hamed et al. [12] analyzed traffic flow on individual roads and developed an Auto Regressive Integrated Moving Average (ARIMA) model suitable for traffic flow forecasting. Given the challenges faced by statistical forecasting models in handling nonlinear traffic flow data, machine learning models have gained prominence in recent decades. Approaches such as k-nearest neighbor models and Support Vector Regression (SVR) models have emerged. For instance, Mathew and Ali Rawther [13] incorporated data correlation between traffic flows into the classification process using optimized k-nearest neighbor classifiers and sinusoidal k-nearest neighbor classifiers. Wu et al. [14] employed SVR for travel time prediction, offering stronger generalization capabilities and ensuring global optimization. However, machine learning models heavily rely on prior knowledge, making it difficult to determine initial parameters for optimal forecasting accuracy. To overcome these limitations, Huang et al. [15] proposed a network architecture that combines a deep belief network with a regression model, achieving higher accuracy in traffic forecasting. Moreover, Recurrent Neural Network (RNN) and its variants, such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), have been widely utilized in time series prediction tasks leveraging their memory capabilities.

While these deep learning methods have enhanced the capability to capture dynamic spatiotemporal features in traffic flow, they exhibit limited interactive learning ability within the spatiotemporal module. This limitation hampers the traffic flow forecasting model's perception of time series periodicity, trend changes, and the ability to adequately capture dynamic spatiotemporal features. Consequently, these methods still inadequately capture traffic flow's latent spatial features.

Addressing these challenges, we propose Attention-based spatiotemporal Interactive Dynamic Graph Convolutional Network (AIDGCN), modeling dynamic spatiotemporal patterns in traffic data. First, we propose a Dynamic Graph Convolutional Network (DGCN) that utilizes prior knowledge to construct dynamic graphs, capturing latent spatial features in traffic flow. These graphs are then embedded into an Interactive Learning framework, forming the Interactive Dynamic Graph Convolutional Network (IDGCN). The IDGCN analyzes traffic flow periodicity, divides sequences into intervals, and facilitates the capture of deep dynamic spatiotemporal features through interactive learning among these segmented sub-sequences. Moreover, adaptive and dynamic neighbor matrices are proposed to model temporal node relationship evolution. Additionally, a Temporal Trend-Aware Multi-Head Self-Attention module captures local context and integrates nonlinear traffic flow temporal features. AIDGCN overcomes existing limitations through accounting for traffic flow periodicity and dynamics, fully leveraging dynamic spatiotemporal features, and demonstrating accurate forecasting potential.

Our primary contributions are:

1. We propose AIDGCN, the first framework to combine IDGCN with time trend-aware attention. IDGCN's interleaved interval partitioning mechanism enables bidirectional feature sharing between temporally symmetric sub-sequences.
2. We introduce a dynamic graph generator that collaborates with an adaptive adjacency matrix and a learnable adjacency matrix. This fusion mechanism is enhanced through Gumbel-softmax reparameterization, addressing the rigid issues of predefined graphs and the instability of purely adaptive methods, enabling robust modeling of spatiotemporal heterogeneity.
3. We design a temporal multi-head trend-aware self-attention mechanism that integrates 1D convolutions into the attention scoring process. This directly addresses the limitation of standard attention mechanisms in failing to capture local flow trends in continuous data.

The paper is organized as follows: [Section 2](#) presents the related works of traffic flow forecasting, graph convolution network, and attention mechanism. [Section 3](#) elaborates the AIDGCN framework and methodology. [Section 4](#) discusses the results and analysis of the comparative experiment. [Section 5](#) summarizes research findings.

2 Related Works

2.1 Traffic Flow Forecasting

Traffic flow forecasting, a core element of transportation systems, garners substantial research focus. In recent years, it has garnered significant attention as a crucial problem in the development of traffic information systems. Researchers have investigated diverse methodologies, from classical mathematical models to modern data-driven techniques, for developing precise traffic prediction systems. Conventional traffic analysis relies on statistical-mathematical approaches for historical data processing. These models assume that the future forecast data possesses similar attributes to the historical data, thus enabling effective traffic flow forecasting [16].

Progress in computational power and artificial intelligence has revived academic interest in traffic flow forecasting. Non-stationary or complex time series challenge conventional models like ARIMA [17], VAR [18], SVR [19], and FNN [20], resulting in subpar performance. Deep Recurrent Neural Networks, particularly LSTM [21] and GRU [22], demonstrate stronger temporal correlation capture than traditional models. These networks excel at memorizing information from extensive sequential data and learning intricate patterns. However, simple RNNs fail to capture spatial aspects of traffic data, creating a fundamental constraint. To overcome this, CNNs capture spatial variations in Euclidean space but remain constrained by standard grid data formats [23]. Recent studies increasingly employ GCNs to model road networks' non-Euclidean relationships [24].

2.2 Graph Convolution Network

Graph Convolutional Networks have gained prominence as a research field, extending traditional convolution to graph-structured data. Two distinct graph convolution methodologies exist. The first type focuses on generalizing the spatial neighborhood through convolutional filtering, with node neighborhood selection being a key aspect. Notably, Veličković et al. [25] introduced graph-based attention to allocate distinct weights to neighboring nodes. The second type involves extending convolution to the spectral domain of graphs by searching for the appropriate Fourier bases.

The initial work by Bruna et al. [26] established the foundational graph convolution framework using graph Laplacian operators. Later, Defferrard et al. [27] obtained initial results using Chebyshev polynomial approximation for graph eigenvalue decomposition, eliminating Laplacian eigenvector computation. Zhao et al. [28] introduced T-GCN, combining graph convolutions with traffic forecasting. Though Yu et al. [29] proposed a gated GCN for traffic prediction, though it insufficiently modeled dynamic spatiotemporal patterns. Chen et al. [30] proposed MRA-BGCN, an innovative deep learning framework with cutting-edge traffic forecasting performance.

2.3 Attention Mechanism

Attention mechanisms are prevalent in NLP applications, traffic forecasting, and speech recognition due to its effective dependency modeling. Its core function is to identify and extract task-critical information from extensive datasets.

Liang et al. [31] designed a multilayer attention mechanism for modeling dynamic spatiotemporal correlations in geo-sensors. Nevertheless, training separate models per time series incurred excessive computational costs. In a different approach, Guo et al. [32] developed a graph CNN incorporating attention, yielding initial traffic forecasting outcomes.

Diverging from prior methods, this work explicitly models the traffic network's graph structure and dynamic spatiotemporal data patterns. A novel Temporal Multi-Head Trend-Aware Self-Attention mechanism is developed, effectively capturing the dynamic dependencies within the network structure for improved feature representation. Prior works utilized GCNs for spatial modeling. However, IDGCN advances these by integrating interleaved sampling to split sequences into symmetric intervals, enabling feature sharing and dynamically fusing adaptive adjacency matrices and learnable adjacency matrices. Simultaneously, the Temporal Multi-Head Trend-Aware Self-Attention uniquely incorporates convolutional operations to capture localized temporal trends, addressing a gap in traditional attention for continuous data, and IDGCN introduces a graph generator with Gumbel reparameterization to simulate dynamic correlations.

3 Methodology

3.1 Problem Definition

As a core time series prediction problem, traffic flow forecasting receives widespread study [33]. The traffic network is modeled as a weighted directed graph $G = (V, E, A)$. The equation specifies $|V| = N$ as the traffic network's sensor nodes, and E as the set of edges that describe the connection strength among nodes within V . The spatial adjacency matrix $A \in \mathbb{R}^{N \times N}$ encodes the graph G , with each entry A_{ij} indicating either the similarity in node features or the Euclidean distance between road sensors v_i and v_j . A_{ij} is set to 1 if both $v_i, v_j \in V$ and $(v_i, v_j) \in E$ are satisfied, otherwise, A_{ij} is set to 0. The traffic network G is associated with a graph signal $X_G^{(t)} \in \mathbb{R}^{N \times C}$ at time step t , where each element of $X_G^{(t)} \in \mathbb{R}^{N \times C}$ corresponds to traffic attributes (e.g., speed, flow) recorded by individual sensors. Specifically, the historical speed observations of N nodes over T time steps are denoted as $P = [X_G^{(t-T+1)}, X_G^{(t-T+2)}, \dots, X_G^{(t)}] \in \mathbb{R}^{C \times N \times T}$. The traffic speeds of all nodes for the next T' time steps are to be predicted, denoted as $Q = [X_G^{(t+1)}, X_G^{(t+2)}, \dots, X_G^{(t+T')}] \in \mathbb{R}^{N \times C \times T'}$.

Based on this representation, the traffic prediction task is formulated as:

$$Q = \ell(P; G) \quad (1)$$

where ℓ is the function to be learned. This research focuses on traffic speed prediction, an effective indicator of both density and volume.

3.2 Framework of AIDGCN

This work proposes AIDGCN, a new framework that captures traffic flow's dynamic spatiotemporal relationships simultaneously. Fig. 1 shows the AIDGCN framework, and it primarily comprises two key components: IDGCN and a Temporal Multi-Head Trend-Aware Self-Attention mechanism.

Firstly, the Start Convolution layer transforms raw data into high-dimensional spatial features, extracting deeper dependencies. Next, the IDGCN is applied on top of the DGCN using an interactive learning strategy. IDGCN recursively generates two equal-sized subsequences through interleaved sampling, which are then interactively learned and share their respective features. This interactive learning structure embeds the DGCN to capture dynamic spatial features while considering temporal dependencies. After extracting spatiotemporal features with IDGCN, the subsequences are organized in time-indexed order using the serial fusion module. Diffusion Graph Convolution processes these subsequences to extract traffic flow's global

spatiotemporal dynamics. The resulting features then pass through Temporal Multi-Head Trend-Aware Self-Attention and an MLP for sequence prediction.

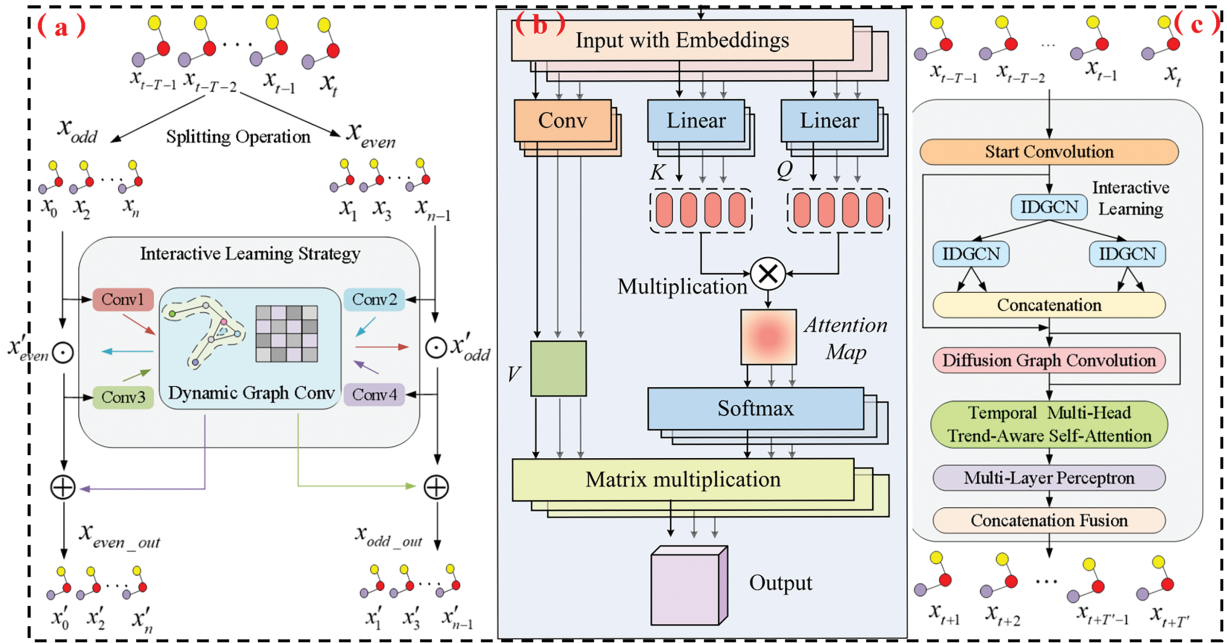


Figure 1: The overall framework of AIDGCN (a) Structure of IDGCN (b) Structure of DGCN (c) AIDGCN Architecture

3.2.1 Interactive Learning

To enable interactive learning, this paper employs a combination of CNN and GCN. GCN is particularly suitable for processing non-Euclidean data, allowing for more effective learning of spatiotemporal dependencies in traffic flow. Compared to CNN and TCN-based methods, GCN is better equipped to model spatiotemporal dynamics. Given traffic flow's periodic, trending nature, the interleaved subsequence preserves most original sequence information. Therefore, we adopt interleaved sampling in this study to perform multi-resolution analysis and expand the sensory field.

The Interactive Learning Framework incorporates three IDGCNs, where the IDGCN module serves as the central element, as depicted in Fig. 1c [34]. This module enables two subsequences to interactively acquire dynamic spatiotemporal features. Prior to feature extraction, convolutional preprocessing expands each subsequence's receptive field. Shared DGCN parameters between subsequences further improve complementary pattern learning.

This paper defines the initial input of IDGCN as $X \in \mathbb{R}^{C \times N \times T}$, with 1D convolution operations denoted as $Conv1$, $Conv2$, $Conv3$, and $Conv4$. Through interleaving sampling, X splits into two subsequences, $X_{odd} \in \mathbb{R}^{C \times N \times T/2}$ and $X_{even} \in \mathbb{R}^{C \times N \times T/2}$, which are then outputted as $X'_{odd} \in \mathbb{R}^{C \times N \times T/2}$ and $X'_{even} \in \mathbb{R}^{C \times N \times T/2}$ after the first interactive learning phase. Subsequent iterations yield the final outputs $X_{odd_out} \in \mathbb{R}^{C \times N \times T/2}$ and $X_{even_out} \in \mathbb{R}^{C \times N \times T/2}$. The detailed steps of interactive dynamic graph convolution are outlined below:

$$X_{even}, X_{odd} = Split(X) \quad (2)$$

$$X'_{odd} = \tanh(DGCN(Conv1(X_{even}))) \odot X_{odd} \quad (3)$$

$$X'_{even} = \tanh(DGCN(Conv2(X_{odd}))) \odot X_{even} \quad (4)$$

$$X_{odd_out} = X'_{odd} + \tanh(DGCN(Conv3(X'_{even}))) \quad (5)$$

$$X_{even_out} = X'_{even} + \tanh(DGCN(Conv4(X'_{odd}))) \quad (6)$$

where $\odot$ represents the Hadamard product, $\tanh$ is the activation function, and $DGCN$ refers to the dynamic graph convolution operation in IDGCN.

3.2.2 Dynamic Graph Convolution Network

The proposed DGCN architecture integrates two key components: a diffusion graph convolution network and a graph generation module [35]. The DGCN employs diffusion graph convolution combined with dynamic graph generation to effectively capture deep spatial dynamics, thereby enhancing the ability of AIDGCN to capture spatial variations. The DGCN processes both $H \in \mathbb{R}^{C \times N \times T}$ and $A \in \mathbb{R}^{N \times N}$ through its Diffusion Graph Convolutional Network. This is followed by the Graph Generator and MLP layer, which generate a discrete matrix containing spatiotemporal information, denoted as $A' \in \mathbb{R}^{N \times N}$. A' is represented as:

$$A' = \text{SoftMax}(\text{MLP}(\text{GCN}(H, A))) \quad (7)$$

where GCN combines Diffusion Convolution with Graph Generator operations, and MLP represents the Multilayer Perceptron. The training utilizes Gumbel reparameterization for feasible sampling.

$$\begin{aligned} A_{learn} &= \text{GumbelSoftmax}(A') \\ &= \text{SoftMax}((\log(A') - \log(-\log(g))) / \tau) \end{aligned} \quad (8)$$

where $g \sim \text{Gumbel}(0, 1)$ is a random variable, τ denotes SoftMax temperature and fixed at 0.5, and A_{learn} represents a dynamic adjacency matrix generated by the Graph Generator to capture node dependencies. Furthermore, $A_{apt} \in \mathbb{R}^{N \times N}$ constructs an adaptive adjacency matrix, denoted as follows:

$$A_{apt} = \text{SoftMax}(\text{Relu}(E_1 E_2^T)) \quad (9)$$

where $E_1 \in \mathbb{R}^{N \times c}$, $E_2^T \in \mathbb{R}^{N \times c}$ denote trainable parameters, and A_{apt} is a predefined connection matrix $A \in \mathbb{R}^{N \times N}$.

The hidden dynamic spatiotemporal correlations in traffic roads are captured by combining A_{learn} and A_{apt} using a dynamic fusion module. The generated dynamic connection matrix $A_{dyn} \in \mathbb{R}^{N \times N}$ is then fed into a Diffusion Graph Convolutional Network. The fusion process is:

$$A_{dyn} = \alpha A_{apt} + (1 - \alpha) A_{learn} \quad (10)$$

where α represents the learnable adaptive parameter factor.

The Diffusion Graph Convolution is employed in various components, including the graph generator network, fusion graph convolution, and concatenation fusion modules. In these cases, the input to the Diffusion Graph Convolution is uniformly defined as $X_{in} \in \mathbb{R}^{C \times N \times T}$. The Diffusion Graph Convolution in graph generator networks is defined as follows:

$$\text{GCN}(X_{in}, A_{apt}) = \sum_{k=0}^K A_{apt}^k X_{in} W \quad (11)$$

where k denotes the diffusion step size, K represents the maximum diffusion steps, and W corresponds to the parameter matrix.

For the fusion graph convolution module with input adjacency matrix A_{dyn} , the diffusion graph convolution is computed as:

$$GCN(X_{in}, A_{dyn}) = \sum_{k=0}^K A_{dyn}^k X_{in} W \quad (12)$$

AIDGCN reassembles dynamic spatiotemporal features from the Interactive Learning Mechanism of the Concatenation module, following a time index order. These features are subsequently processed by the Diffusion Graph Convolution to identify and rectify the entire temporal features. Unlike previous approaches, our method employs both a predefined initial adjacency matrix $A \in \mathbb{R}^{N \times N}$ and a dynamic connection matrix $A_{dyn} \in \mathbb{R}^{N \times N}$ generated through adaptive learning. The initial connection matrix A is represented by a directed graph, with $P_f = A / \text{rowsum}(A)$ and $P_b = A^T / \text{rowsum}(A^T)$ denoting the progressive and regressive transition matrices, respectively. The diffusion graph convolution operation within the concatenation fusion module is formally expressed as:

$$GCN(X_{in}, A, A_{dyn}) = \sum_{k=0}^K (A_f^k X_{in} W_1 + A_b^k X_{in} W_2 + A_{dyn}^k X_{in} W_3) \quad (13)$$

The DGCN module captures latent spatial dependencies between traffic network nodes to extract high-order structural features, while simultaneously modeling dynamic data correlations through traffic flow simulations. By incorporating DGCN into an interactive learning paradigm, the framework leverages these adaptive spatial representations to improve temporal dependency learning across traffic sequences during training.

3.2.3 Temporal Multi-Head Trend-Aware Self-Attention

Self-attention operates as an attention mechanism utilizing the same sequence to generate queries, keys, and values. In practice, the widely used multi-head self-attention mechanism enables simultaneous attention to information from multiple subspaces (Fig. 1b). The multi-head attention mechanism performs these key operations:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_{model}}}\right) V \quad (14)$$

where Q , K , and V denote the query, key, and value, respectively.

In multi-head self-attention, the query, keys, and values are initially projected into separate subspaces. The attention functions are then executed in parallel. The resulting outputs are concatenated and further projected to obtain the final output, represented as follows:

$$SelfAttention(Q, K, V) = \oplus (head_1, \dots, head_h) W^O \quad (15)$$

$$head_j = Attention(QW_j^Q, KW_j^K, VW_j^V) \quad (16)$$

The variable h denotes the quantity of attention heads, W_j^Q , W_j^K , and W_j^V are the projection matrices used on Q , K , and V , respectively. As the terminal projection matrix, W^O enables multi-head self-attention to flexibly model intricate traffic data relationships, resulting in precise long-term predictions.

Originally developed for discrete markers, the conventional multi-head self-attention mechanism fails to effectively represent local trend patterns in continuous datasets. Direct application to traffic signal sequence transformation could lead to compatibility problems. To overcome this limitation and incorporate

local contextual patterns in numerical forecasting, we introduce Temporal Trend-Aware Multi-Head Self-Attention, a novel Convolutional Self-Attention variant. By incorporating context-aware convolutional layers, the model learns latent local traffic patterns. The Temporal Trend-Aware Multi-Head Attention is formally defined as:

$$\text{Trendhead}_j = \text{Attention} \left(Q * \Phi_j^Q, K * \Phi_j^K, V W_j^V \right) \quad (17)$$

where $*$ represents the convolution operation, while Φ_j^Q and Φ_j^K represent the parameters of the convolution kernel.

In the l -th layer, applying the Temporal Trend-Aware Multi-Head Self-Attention Mechanism to the input $\chi^{(l-1)}$ yields an intermediate sequence representation denoted as $Z^{(l-1)} = \left(z_{t-T_h+1}^{(l-1)}, z_{t-T_h+2}^{(l-1)}, \dots, z_t^{(l-1)} \right) \in \mathbb{R}^{N * d_{model} * T_h}$ across all nodes.

3.2.4 Other Components

As a widely-used loss function, Huber Loss [36] integrates the linear components from both Mean Square Error (MSE) and Absolute Error. When predictions approximate the actual values, it demonstrates characteristics similar to Squared Loss, while exhibiting Absolute Loss behavior for significant deviations. Such characteristics significantly reduce outlier influence while maintaining training stability. Therefore, this study employs Huber Loss for optimization.

$$\xi = \frac{1}{N \times Q} \sum_{i=1}^Q \sum_{j=1}^N R(\hat{x}_{i,j}, x_{i,j})$$

$$R(\hat{x}, x) = \begin{cases} \frac{1}{2} (\hat{x} - x)^2, & |\hat{x} - x| \leq \delta \\ \delta |\hat{x} - x| - \frac{1}{2} \delta^2, & |\hat{x} - x| > \delta \end{cases} \quad (18)$$

where δ serves as the hyperparameter that balances the Squared Error, $\hat{x}$ and x denote the predicted and true values, respectively.

4 Experiment

4.1 Datasets

Our evaluation employs the widely-used METR-LA and PEMS-BAY transportation datasets to measure the AIDGCN predictive capabilities [37–40]. Additional details regarding the experimental datasets can be found in Table 1.

Table 1: Description of the traffic dataset

Data	METR-LA	PEMS-BAY
Type	Sequentially	Sequentially
Attribute	Speed	Speed
Location	Highways of Los Angeles	The Bay Area
Edges	1515	2369
Time steps	34,272	52,116
Nodes	207	325

Prior to model input, data underwent min-max normalization (0–1 scaling). Applied normalization equation:

$$Z_i^{norm} = \frac{z_i - \mu}{\vartheta} \quad (19)$$

where z_i denotes the i -th raw data point, with μ and ϑ indicating the dataset's mean standard deviation, while Z_i^{norm} signifies the normalized value.

4.2 Parameter Settings

All models were developed in PyTorch using PyCharm. Using 60 min of past observations ($P = 12$) to forecast subsequent 60 min ($Q = 12$), we partitioned the data into 60% training, 20% validation, and 20% test based on actual needs. Traffic flow prediction is performed for 15, 30, and 60 min. For comprehensive performance evaluation, this study employs multiple metrics to quantify the discrepancy between predicted and actual traffic flow speeds:

$$MAE = \frac{1}{F} \sum_{i=1}^F |\hat{\partial}_i - \partial_i| \quad (20)$$

$$MAPE = \frac{1}{F} \sum_{i=1}^F \left| \frac{\hat{\partial}_i - \partial_i}{\hat{\partial}_i} \right| \times 100\% \quad (21)$$

$$RMSE = \sqrt{\frac{1}{F} \sum_{i=1}^F (\hat{\partial}_i - \partial_i)^2} \quad (22)$$

where F represents the sample size, with $\hat{\partial}_i$ and ∂_i representing the i -th sample's ground truth and prediction. The AIDGCN model demonstrates superior forecasting capability when MAE, RMSE, and MAPE values are minimized.

4.3 Baselines

The AIDGCN model is evaluated against 15 baseline models: basic methods like HA [10], VAR [11], SVR [41], and ARIMA [10]; neural networks FNN [42] and FC-LSTM [43]; and advanced spatiotemporal graph networks including DCRNN [40], STGCN [29], ASTGCN [32], STSGCN [44], T-GCN [28], Graph WaveNet [45], MRA-BGCN [30], DGRN [46], and D2STGNN [47], which capture dependencies using techniques like diffusion convolution, Chebyshev polynomials, attention mechanisms, synchronous modeling, GCN-GRU fusion, dilated convolutions, dual graphs, adaptive adjacency, and decoupled spatiotemporal modeling.

4.4 Experimental Results

4.4.1 Performance Comparison

AIDGCN was compared with 15 baseline models for 15, 30, and 60-min traffic flow forecasting. In most scenarios, AIDGCN outperformed all baseline methods across all datasets.

Table 2 demonstrates the proposed model's superior forecasting accuracy compared to baselines on METR-LA. For instance, in 15- and 60-min forecasting, AIDGCN outperforms MRA-BGCN by 0.75% and 0.39% in terms of MAE and by 1.32% and 0.14% in terms of RMSE, respectively. Specifically, in the 60-min forecast, AIDGCN reduced MAE by 8.3% compared to DGRN and reduced RMSE by 6.1% compared to

D2STGNN. Predictions for other time intervals also exhibited notable improvements compared to the best baseline results.

Table 2: Performance comparison of different prediction models on METR-LA and PEMS-BAY datasets

Datasets	Methods	15 min			30 min			60 min		
		MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
METR-LA	HA	4.16	7.80	13.00%	4.16	7.80	13.00%	4.16	7.80	13.00%
	VAR	4.42	7.89	10.20%	5.41	9.13	12.7%	6.52	10.11	15.80%
	SVR	3.99	8.45	9.30%	5.05	10.87	12.10%	6.72	13.76	16.70%
	ARIMA	3.99	8.21	9.60%	5.15	10.45	12.70%	6.90	13.23	17.40%
	FNN	3.99	7.94	9.90%	4.23	8.17	12.90%	4.49	8.69	14.00%
	FC-LSTM	3.44	6.30	9.60%	3.77	7.23	10.90%	4.37	8.69	13.20%
	DCRNN	2.77	5.38	7.30%	3.15	6.45	8.80%	3.60	7.59	10.50%
	STGCN	2.88	5.74	7.62%	3.47	7.24	9.57%	4.59	9.40	12.70%
	ASTGCN	4.86	9.27	9.21%	5.43	10.61	10.13%	6.51	12.52	11.64%
	STSGCN	3.31	7.62	8.06%	4.13	9.77	10.29%	5.06	11.66	12.91%
	T-GCN	3.03	5.26	7.81%	3.52	6.12	9.45%	4.30	7.31	11.8%
	Graph WaveNet	2.69	5.15	6.90%	3.07	6.22	8.37%	3.53	7.37	10.01%
	MRA-BGCN	2.67	5.12	6.80%	3.06	6.17	8.30%	3.49	7.30	10.00%
	DGRCN	2.71	5.18	7.12%	3.11	6.41	8.62%	3.56	7.38	10.35%
PEMS-BAY	D2STGNN	2.68	5.15	7.05%	3.08	6.35	8.53%	3.51	7.33	10.22%
	AIDGCN	2.65	5.10	6.95%	3.02	6.26	8.41%	3.44	7.29	10.11%
	HA	2.88	5.59	6.80%	2.88	5.59	6.80%	2.88	5.59	6.80%
	VAR	1.74	3.16	3.60%	2.32	4.25	5.00%	2.93	5.44	6.50%
	SVR	1.85	3.59	3.80%	2.48	5.18	5.50%	3.28	7.08	8.00%
	ARIMA	1.62	3.30	3.50%	2.33	4.76	5.40%	3.38	6.50	8.30%
	FNN	2.20	4.42	5.19%	2.30	4.63	5.43%	2.46	4.98	5.89%
	FC-LSTM	2.05	4.19	4.80%	2.20	4.55	5.20%	2.37	4.96	5.70%
	DCRNN	1.38	2.95	2.90%	1.74	3.97	3.90%	2.07	4.47	4.90%
	STGCN	1.36	2.96	2.90%	1.81	4.27	4.17%	2.49	5.69	5.79%
	ASTGCN	1.52	3.13	3.22%	2.01	4.27	4.48%	2.61	5.42	6.00%
	STSGCN	1.44	3.01	3.04%	1.83	4.18	4.17%	2.26	5.21	5.40%
	T-GCN	1.50	2.83	3.14%	1.73	3.40	3.76%	2.18	4.35	4.94%
	Graph WaveNet	1.30	2.74	2.73%	1.63	3.70	3.67%	1.95	4.52	4.63%
	MRA-BGCN	1.29	2.72	2.90%	1.61	3.67	3.80%	1.91	4.46	4.60%
	DGRCN	1.27	2.78	2.75%	1.62	3.68	3.92%	1.95	4.50	4.70%
	D2STGNN	1.25	2.75	2.68%	1.59	3.63	3.65%	1.90	4.51	4.50%
	AIDGCN	1.22	2.71	2.63%	1.55	3.58	3.53%	1.79	4.24	4.27%

In the case of PEMS-BAY, AIDGCN achieves the best forecasting performance. Compared to MRA-BGCN, it achieves 6.28% lower MAE, 4.93% lower RMSE and 7.17% lower MAPE for 60-min long-term forecasting. Our approach achieves cutting-edge results compared to existing methods.

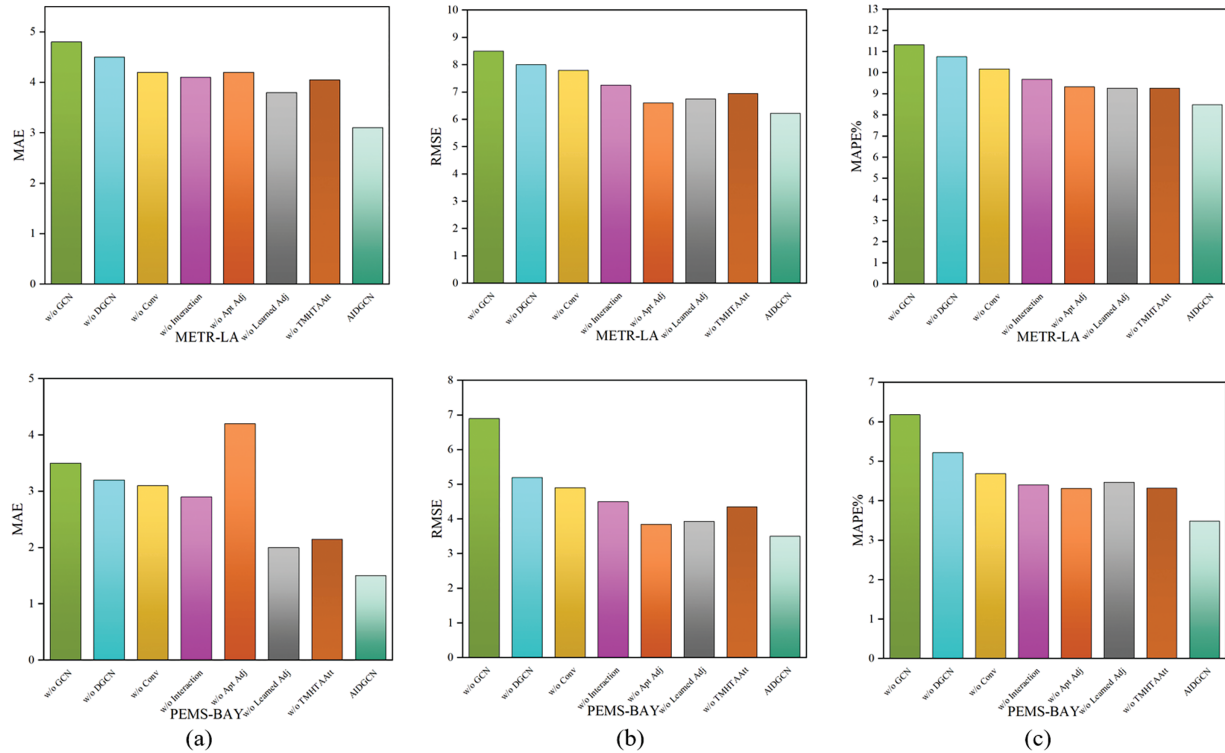
Table 3 evaluates computational efficiency of spatiotemporal models on METR-LA/PEMS-BAY datasets. Among models with advanced capabilities, AIDGCN outperforms MRA-BGCN in both size and speed. While Graph WaveNet is faster than similarly-sized DCRNN, AIDGCN's inference approaches DCRNN's despite greater complexity. AIDGCN thus offers a superior accuracy-efficiency trade-off for demanding tasks, delivering high representational power with manageable latency compared to peers like MRA-BGCN.

Table 3: Computational Efficiency on METR-LA and PEMS-BAY datasets

Model	METR-LA		PEMS-BAY	
	Params (M)	Inference time (ms)	Params (M)	Inference time (ms)
DCRNN	1.8	42	2.0	53
STGCN	1.5	28	1.7	36
Graph WaveNet	1.9	32	2.1	41
MRA-BGCN	2.3	41	2.5	52
AIDGCN	2.1	38	2.3	48

4.4.2 Ablation Experiment

To evaluate the performance contributions of its modules, seven variants of the AIDGCN model were developed and compared in Fig. 2. These variants systematically remove or modify key components: w/o GCN eliminates the diffusion GCN; w/o DGCN replaces the DGCN with a standard diffusion GCN using the predefined adjacency matrix; w/o Conv removes the 1D convolution from the interactive learning framework; w/o Interaction substitutes the interactive learning component with a sequentially connected temporal convolutional network (TCN); w/o Apt Adj removes the adaptive adjacency matrix from the DGCN module; w/o Learned Adj eliminates the graph generator while retaining the adaptive matrix and replaces the fusion GCN with diffusion GCN; and w/o TMHTAAtt removes the Temporal Multi-Head Trend-Aware Self-Attention component.

**Figure 2:** Metrics comparison on the two datasets. (a) MAE; (b) RMSE; (c) MAPE

Ablation studies reveal the dynamic graph structure learning mechanism as AIDGCN's most critical component, particularly for spatial adaptation, where removing the adaptive adjacency matrix caused a

180% MAE surge in PEMS-BAY. The spatiotemporal interaction module substantially impacts performance (93.3% MAE increase when removed in PEMS-BAY), while the 1D convolution is vital for temporal features (19.9% MAPE rise in METR-LA). Temporal attention shows dataset-specific efficacy but universal value in pattern capture (24.1% MAPE improvement in PEMS-BAY). Cross-dataset analysis highlights PEMS-BAY's $5\times$ greater sensitivity to graph structures vs. METR-LA's emphasis on temporal modeling. By impact, adaptive adjacency and interaction structures rank highest, accounting for over 50% of gains, with optimization priorities favoring adaptive graph learning and scenario-specific customization of temporal components.

4.4.3 Visual Analysis

Fig. 3 demonstrates AIDGCN's superior forecasting performance against baseline models (FNN, FC-LSTM, Graph WaveNet, STGCN) across all prediction steps on PEMS-BAY.

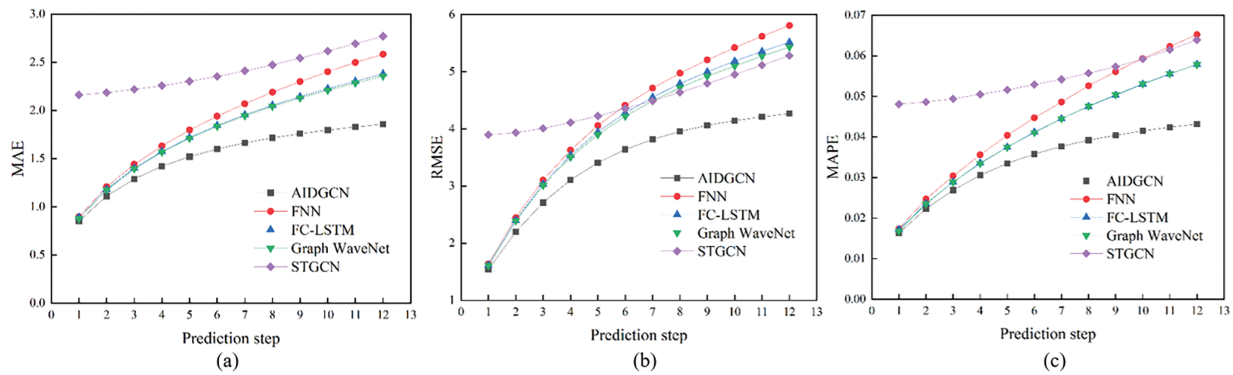


Figure 3: Metrics of different models on PEMS-BAY dataset. (a) MAE; (b) RMSE; (c) MAPE

The significantly slower growth rate of MAE, RMSE, and MAPE for AIDGCN stems directly from its core innovations. The adaptive graph convolution dynamically captures evolving spatial dependencies, preventing error propagation from outdated connections. Simultaneously, the temporal attention mechanism focuses on crucial historical states, effectively mitigating error accumulation over long horizons—a key limitation of baseline models. This synergy results in the observed superior stability beyond 15 min.

The widening performance gap for longer steps underscores AIDGCN's strength in modeling non-linear, time-evolving traffic patterns. Unlike baselines with static graphs, its adaptive module continuously refines spatial relationships based on real-time data, enabling more accurate capture of complex dynamics like congestion propagation. AIDGCN's robust long-term accuracy is vital for proactive ITS applications, where reliable forecasts >15 min are essential for effective decision-making.

5 Conclusion

This paper introduced AIDGCN, a novel deep learning model designed to overcome key limitations in traffic forecasting, including static spatial modeling, insufficient capture of complex spatiotemporal dependencies, and degraded long-term prediction accuracy. AIDGCN demonstrably achieves state-of-the-art performance, significantly outperforming existing baselines.

(1) AIDGCN combines a dynamic graph structure and a Temporal Multi-Head Trend-Aware Self-Attention mechanism for learning evolving traffic patterns comprehensively. The dynamic graph structure

simulates associations among nodes, uncovering hidden spatial correlations. The interactive learning framework of IDGCN incorporates a DGCN module to simultaneously model periodic patterns, temporal trends, and spatiotemporal relationships.

(2) The Temporal Multi-Head Trend-Aware Self-Attention mechanism leverages dynamic temporal features to improve forecasting accuracy, addressing the insensitivity of traditional attention mechanisms to local trends in continuous data.

(3) On the challenge ng PEMS-BAY dataset for 60-min predictions, AIDGCN reduces MAE, RMSE, and MAPE by 6.28%, 4.93%, and 7.17% respectively compared to the best-performing baseline (MRA-BGCN). Ablation studies further confirm the indispensable contribution of each proposed module to this performance gain. The model's robust performance across varying prediction horizons highlights its effectiveness for both short and long-term forecasting tasks.

Future research should focus on:

1. Integration of External Factors: Investigating specific mechanisms to incorporate fine-grained weather, events, and holidays.
2. Scalability and Efficiency Optimization: Exploring techniques like model compression and distributed computing for large-scale, real-time deployment.
3. Enhanced Dynamic Graph Learning: Developing more sophisticated and interpretable methods for graph construction.
4. Generalization to Diverse Scenarios: Evaluating and enhancing robustness across varied networks, data conditions, and disruptive events using adaptation techniques.

Acknowledgement: Not applicable.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization, Zitong Zhao; methodology, Zixuan Zhang; software, Zhenxing Niu; validation, Zitong Zhao; formal analysis, Zixuan Zhang; investigation, Zhenxing Niu; resources, Zitong Zhao; data curation, Zitong Zhao; writing—original draft preparation, Zhenxing Niu; writing—review and editing, Zixuan Zhang; visualization, Zitong Zhao; supervision, Zhenxing Niu; project administration, Zitong Zhao; funding acquisition, Zhenxing Niu. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are available from the Corresponding Author upon reasonable request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Zhang K, Chu Z, Xing J, Zhang H, Cheng Q. Urban traffic flow congestion prediction based on a data-driven model. *Mathematics*. 2023;11(19):4075. doi:10.3390/math11194075.
2. Kohan M, Ale JM. Discovering traffic congestion through traffic flow patterns generated by moving object trajectories. *Comput Environ Urban Syst*. 2020;80(6):101426. doi:10.1016/j.compenvurbsys.2019.101426.
3. Zhang J, Wang FY, Wang K, Lin WH, Xu X, Chen C. Data-driven intelligent transportation systems: a survey. *IEEE Trans Intell Transp Syst*. 2011;12(4):1624–39. doi:10.1109/TITS.2011.2158001.

4. Zhang X, Liu J, Zhang X, Lu Y. Self-supervised graph feature enhancement and scale attention for mechanical signal node-level representation and diagnosis. *Adv Eng Inform.* 2025;65:103197. doi:10.1016/j.aei.2025.103197.
5. Zhang X, Liu J, Zhang X, Lu Y. Multiscale channel attention-driven graph dynamic fusion learning method for robust fault diagnosis. *IEEE Trans Ind Inform.* 2024;20(9):11002–13. doi:10.1109/TII.2024.3397401.
6. Long W, Xiao Z, Wang D, Jiang H, Chen J, Li Y, et al. Unified spatial-temporal neighbor attention network for dynamic traffic prediction. *IEEE Trans Veh Technol.* 2023;72(2):1515–29. doi:10.1109/TVT.2022.3209242.
7. Wang T, Chen J, Lu J, Liu K, Zhu A, Snoussi H, et al. Synchronous spatiotemporal graph transformer: a new framework for traffic data prediction. *IEEE Trans Neural Netw Learn Syst.* 2023;34(12):10589–99. doi:10.1109/TNNLS.2022.3169488.
8. Zhang X, Zhang X, Liu J, Wu B, Hu Y. Graph features dynamic fusion learning driven by multi-head attention for large rotating machinery fault diagnosis with multi-sensor data. *Eng Appl Artif Intell.* 2023;125(6):106601. doi:10.1016/j.engappai.2023.106601.
9. Samaan SS, Korial AE, Sarra RR, Humaidi AJ. Multilingual web traffic forecasting for network management using artificial intelligence techniques. *Results Eng.* 2025;26(1):105262. doi:10.1016/j.rineng.2025.105262.
10. Williams BM, Hoel LA. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: theoretical basis and empirical results. *J Transp Eng.* 2003;129(6):664–72. doi:10.1061/(asce)0733-947x(2003)129:6(664).
11. Wang L, Ding S. Vector autoregression and envelope model. *Stat.* 2018;7(1):e203. doi:10.1002/sta4.203.
12. Hamed MM, Al-Masaeid HR, Said ZMB. Short-term prediction of traffic volume in urban arterials. *J Transp Eng.* 1995;121(3):249–54. doi:10.1061/(asce)0733-947x(1995)121:3(249).
13. Mathew A, Ali Rawther F. Hadoop based short—term traffic flow prediction on D² its using correlation model and KNN HSSine. In: 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI); 2017 Sep 21–22; Chennai, India. Piscataway, NJ, USA: IEEE; 2017. p. 1123–9. doi:10.1109/ICPCSI.2017.8391885.
14. Wu CH, Wei CC, Su DC, Chang MH, Ho JM. Travel time prediction with support vector regression. In: Proceedings of the 2003 IEEE International Conference on Intelligent Transportation Systems; 2003 Oct 12–15; Shanghai, China. Piscataway, NJ, USA: IEEE; 2003. p. 1438–42. doi:10.1109/ITSC.2003.1252721.
15. Huang W, Song G, Hong H, Xie K. Deep architecture for traffic flow prediction: deep belief networks with multitask learning. *IEEE Trans Intell Transp Syst.* 2014;15(5):2191–201. doi:10.1109/TITS.2014.2311123.
16. Abdelhafid Z, Harrou F, Sun Y. An efficient statistical-based approach for road traffic congestion monitoring. In: 2017 5th International Conference on Electrical Engineering—Boumerdes (ICEE-B); 2017 Oct 29–31; Boumerdes, Algeria. Piscataway, NJ, USA: IEEE; 2017. p. 1–5. doi:10.1109/ICEE-B.2017.8192228.
17. Yao R, Zhang W, Zhang L. Hybrid methods for short-term traffic flow prediction based on ARIMA-GARCH model and wavelet neural network. *J Transp Eng Part A Syst.* 2020;146(8):04020086. doi:10.1061/jtepbs.0000388.
18. Song X, Guo Y, Li N, Zhang L. Online traffic flow prediction for edge computing-enhanced autonomous and connected vehicles. *IEEE Trans Veh Technol.* 2021;70(3):2101–11. doi:10.1109/TVT.2021.3057109.
19. Hu W, Yan L, Liu K, Wang H. A short-term traffic flow forecasting method based on the hybrid PSO-SVR. *Neural Process Lett.* 2016;43(1):155–72. doi:10.1007/s11063-015-9409-6.
20. Yang H, Du L, Zhang G, Ma T. A traffic flow dependency and dynamics based deep learning aided approach for network-wide traffic speed propagation prediction. *Transp Res Part B Methodol.* 2023;167(2):99–117. doi:10.1016/j.trb.2022.11.009.
21. Zhang Z, Yang H, Yang X. A transfer learning-based LSTM for traffic flow prediction with missing data. *J Transp Eng Part A Syst.* 2023;149(10):04023095. doi:10.1061/jtepbs.teeng-7638.
22. Yi X, Zhou H, Zhong S. Real-time adaptive traffic flow prediction based on a GE-GRU-KNN model. *Promet-Traffic Transport.* 2025;37(3):754–72. doi:10.7307/ptt.v37i3.775.
23. Yao H, Wu F, Ke J, Tang X, Jia Y, Lu S, et al. Deep multi-view spatial-temporal network for taxi demand prediction. *Proc AAAI Conf Artif Intell.* 2018;32(1):2388–95. doi:10.1609/aaai.v32i1.11836.
24. Lv Z, Xu J, Zheng K, Yin H, Zhao P, Zhou X. LC-RNN: a deep learning model for traffic speed prediction. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence; 2018 Jul 13–19; Stockholm, Sweden. p. 3470–6. doi:10.24963/ijcai.2018/482.

25. Veličković P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y. Graph attention networks. arXiv:1710.10903. 2017. doi:10.48550/arXiv.1710.10903.
26. Bruna J, Zaremba W, Szlam A, LeCun Y. Spectral networks and locally connected networks on graphs. arXiv:1312.6203. 2013. doi:10.48550/arxiv.1312.6203.
27. Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering. In: Lee DD, Sugiyama M, Luxburg UV, Guyon I, Garnett R, editors. Advances in neural information processing systems 29 (NIPS 2016). Vol. 29. La Jolla, CA, USA: Neural Information Processing Systems (NIPS); 2016.
28. Zhao L, Song Y, Zhang C, Liu Y, Wang P, Lin T, et al. T-GCN: a temporal graph convolutional network for traffic prediction. IEEE Trans Intell Transp Syst. 2020;21(9):3848–58. doi:10.1109/TITS.2019.2935152.
29. Yu B, Yin H, Zhu Z. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence; 2018 Jul 13–19; Stockholm, Sweden. San Jose, CA, USA: International Joint Conferences on Artificial Intelligence Organization; 2018. p. 3634–40. doi:10.24963/ijcai.2018/505.
30. Chen W, Chen L, Xie Y, Cao W, Gao Y, Feng X. Multi-range attentive bicomponent graph convolutional network for traffic forecasting. In: Proceedings of the AAAI Conference on Artificial Intelligence; 2020 Feb 7–12; New York, NY, USA. p. 3529–36.
31. Liang Y, Ke S, Zhang J, Yi X, Zheng Y. GeoMAN: multi-level attention networks for geo-sensory time series prediction. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence; 2018 Jul 13–19; Stockholm, Sweden. San Jose, CA, USA: International Joint Conferences on Artificial Intelligence Organization; 2018. p. 3428–34. doi:10.24963/ijcai.2018/476.
32. Guo S, Lin Y, Feng N, Song C, Wan H. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. Proc AAAI Conf Artif Intell. 2019;33(1):922–9. doi:10.1609/aaai.v33i01.3301922.
33. Yao Y, Chen L, Wang X, Wu X. Traffic flow forecasting based on augmented multi-component recurrent graph attention network. Transp Lett. 2025;2025:1–9. doi:10.1080/19427867.2025.2450577.
34. Chen L, Chen L, Wang H, Zhang H. Traffic flow prediction based on interactive dynamic spatio-temporal graph convolution with a probabilistic sparse attention mechanism. Transp Res Rec J Transp Res Board. 2024;2678(9):837–53. doi:10.1177/03611981241230545.
35. Zhang H, Zhu S, Zhang X, Gong L. Research on traffic flow forecasting based on interactive dynamic meta-graph learning. Proc Inst Mech Eng Part D J Automob Eng. 2024;33(1):09544070241292392. doi:10.1177/09544070241292392.
36. Xie J, Liu S, Chen J, Jia J. Huber loss based distributed robust learning algorithm for random vector functional-link network. Artif Intell Rev. 2023;56(8):8197–218. doi:10.1007/s10462-022-10362-7.
37. Wang T, Ni S, Qin T, Cao D. TransGAT: a dynamic graph attention residual networks for traffic flow forecasting. Sustain Comput Inform Syst. 2022;36(3):100779. doi:10.1016/j.suscom.2022.100779.
38. Chen H, Wang H, Chen Z. Traffic flow prediction based on dynamic time slot graph convolution. Transp Res Rec J Transp Res Board. 2025;2679(5):785–99. doi:10.1177/03611981241308868.
39. Ounoughi C, Ben Yahia S. Sequence to sequence hybrid Bi-LSTM model for traffic speed prediction. Expert Syst Appl. 2024;236:121325. doi:10.1016/j.eswa.2023.121325.
40. Li Y, Yu R, Shahabi C, Liu Y. Diffusion convolutional recurrent neural network: data-driven traffic forecasting. arXiv:1707.01926. 2017. doi:10.48550/arXiv.1707.01926.
41. Wu Y, Tan H. Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework. arXiv:1612.01022. 2016. doi:10.48550/arXiv.1612.01022.
42. Li HX, Lee ES. Interpolation functions of feedforward neural networks. Comput Math Appl. 2003;46(12):1861–74. doi:10.1016/s0898-1221(03)90242-2.
43. Hu Z, Zhou J, Huang K, Zhang E. A data-driven approach for traffic crash prediction: a case study in Ningbo, China. Int J Intell Transp Syst Res. 2022;20(2):508–18. doi:10.1007/s13177-022-00307-3.

44. Song C, Lin Y, Guo S, Wan H. Spatial-temporal synchronous graph convolutional networks: a new framework for spatial-temporal network data forecasting. In: Proceedings of the AAAI Conference on Artificial Intelligence; 2020 Feb 7–12; New York, NY, USA.
45. Wu Z, Pan S, Long G, Jiang J, Zhang C. Graph WaveNet for deep spatial-temporal graph modeling. arXiv:1906.00121. 2019. doi:10.48550/arXiv.1906.00121.
46. Li F, Feng J, Yan H, Jin G, Yang F, Sun F, et al. Dynamic graph convolutional recurrent network for traffic prediction: benchmark and solution. ACM Trans Knowl Discov Data. 2023;17(1):1–21. doi:10.1145/3532611.
47. Shao Z, Zhang Z, Wei W, Wang F, Xu Y, Cao X, et al. Decoupled dynamic spatial-temporal graph neural network for traffic forecasting. Proc VLDB Endow. 2022;15(11):2733–46. doi:10.14778/3551793.3551827.