



A Multi-Level Dynamic GCN-Transformer Framework with Spatio-Temporal Interaction for Traffic Flow Prediction

Guan LIAN¹, Caihua HUANG², Qi SUN³, Wenyong LI⁴, Yingzi WU⁵

Original Scientific Paper
Submitted: 23 June 2025
Accepted: 9 Sep 2025
Published: 27 May 2026

¹ Corresponding author, lianguan@guet.edu.cn, School of Architecture and Transportation Engineering, Guilin University of Electronic Technology, Guilin, China
² hch_18@mails.guet.edu.cn, School of Architecture and Transportation Engineering, Guilin University of Electronic Technology, Guilin, China
³ sunqi@mails.guet.edu.cn, School of Architecture and Transportation Engineering, Guilin University of Electronic Technology, Guilin, China
⁴ traffic@guet.edu.cn, Guangxi Key Laboratory of Intelligent Transportation, Guilin University of Electronic Technology, Guilin, China
⁵ wuyingzi718@mails.guet.edu.cn, School of Civil Engineering and Transportation, Northeast Forestry University, Harbin, China



This work is licensed under a Creative Commons Attribution 4.0 International Licence.

Publisher:
Faculty of Transport and Traffic Sciences,
University of Zagreb

ABSTRACT

As a core task of intelligent transportation systems, traffic flow prediction is characterised by high spatio-temporal complexity. Due to the limitations of existing methods in modelling complex spatio-temporal dependencies, particularly regarding medium- to long-term prediction accuracy and generalisation capabilities, this paper proposes a combination prediction model based on a multi-level dynamic GCN-Transformer framework (DGTFormer), to enhance the accuracy of short- and long-term traffic flow predictions. DGTFormer adopts a dual-stream architecture to achieve spatio-temporal decoupling modelling. The spatio-temporal dynamic graph convolutional network processes dynamic changes in the road network structure, and the temporal transformer encoder processes temporal information related to traffic flow. A spatio-temporal gated fusion mechanism is introduced to deeply couple spatial and temporal information. Experimental results on three real-world traffic datasets (PeMSD4, PeMSD7 and PeMSD8) demonstrate that DGTFormer significantly outperforms mainstream baseline models on multiple key evaluation metrics. Compared with advanced baseline methods, DGTFormer achieves performance improvements of up to 8.61% and 9.15% in RMSE and MAE, respectively. Furthermore, the coefficient of determination, R^2 , remains stable at an excellent level above 0.9 across different prediction time steps, which fully validates that the DGTFormer model possesses superior predictive performance and generalisation capabilities.

KEYWORDS

traffic flow prediction; graph convolutional network; transformer model; gated fusion; spatio-temporal modelling.

1. INTRODUCTION

As urbanisation accelerates and economic conditions improve, the number of cars in cities increases significantly. This makes the urban road traffic environment increasingly complex, leading to serious traffic problems that restrict sustainable urban development. Efficiently predicting traffic flow conditions to achieve dynamic perception and control of the operational status of the urban road network is an important component of building an intelligent transportation system (ITS) [1]. Urban traffic flow prediction is one of the core tasks and is essentially a complex spatial and temporal sequence modelling problem integrating spatial dependency and temporal dynamics [2]. It requires modelling the dynamic evolution of time series and capturing potential spatial correlations in the traffic network. This places high demands on the model's spatial and temporal feature

extraction capabilities. Accurate and efficient traffic flow prediction enhances the operational efficiency of urban roads, making it crucial for traffic management, route planning and congestion relief [3].

Early researchers primarily used statistical models for traffic flow prediction, such as the autoregressive integrated moving average (ARIMA) model [4] and the Kalman filter model [5-6]. While these models are simple in structure and computationally efficient, and perform stably in short-term predictions, they are constrained by the assumption of stationarity. This makes it difficult to adapt to the dynamic nature and complex spatio-temporal coupling characteristics of traffic flow. These models also perform poorly when dealing with complex nonlinear relationships and long-term temporal dependencies. Subsequently, machine learning methods have been widely applied in traffic forecasting, such as support vector machine (SVM) [7], K-nearest neighbours (KNN) [8], random forest (RF) [9] and gradient boosting decision tree (GBDT) [10]. Although these models have improved the ability to model nonlinear features to some extent, they still struggle to handle spatial dependencies effectively, exhibit poor generalisation capabilities, and remain limited by strong reliance on manual feature extraction and insufficient capture of multi-scale spatio-temporal associations [11].

In recent years, deep learning technology has been widely applied to traffic flow prediction tasks [12]. Compared to traditional statistical models and machine learning methods, its advantages lie in its ability to automatically extract high-level abstract features and adapt to complex dynamic environments, particularly in the context of the continuous growth of large-scale traffic sensing data, demonstrating significant modelling potential. Among these, recurrent neural networks (RNN) [13] and their variants, long short-term memory (LSTM) [14-15] and convolutional neural network (CNN) [16], have performed well in modelling the temporal dependencies of traffic data, establishing themselves as a mainstream approach. Li et al. [17] propose diffusion convolutional recurrent neural networks (DCRNN) that model traffic flow spatio-temporal dependencies through directed graph diffusion processes, combined with a bidirectional random walk to capture complex network spatial associations. The short-term traffic flow prediction model based on LSTM-RNN, proposed by Wang et al. [18], can effectively process time series data for urban expressways. Zhang et al. [19] proposed a channel-based attention separation convolutional neural network (CAS-CNN), which is based on channel attention mechanisms and separate CNNs for short-term traffic flow OD prediction. However, traditional RNN series models often face problems such as gradient disappearance or gradient explosion when dealing with long-term dependencies. They also ignore spatial correlations and lack spatial modelling capabilities, making it difficult for the model to capture long-term correlation features and spatio-temporal dependencies in sequences.

Traffic flow prediction involves not only temporal evolution but also spatial correlations between different areas within the road network. Therefore, the construction of traffic flow prediction models requires the systematic integration of the spatio-temporal dependency characteristics inherent in the road network topology to characterise the dynamic evolution of traffic flow accurately. With the continuous development of graph convolutional network (GCN) [20] and transformer [21] models, an increasing number of research efforts are incorporating these methods into traffic flow prediction, aiming to comprehensively consider the spatial and temporal correlations of traffic data. As proposed by the team from Peking University [22], the STGCN model can effectively extract spatio-temporal characteristics from traffic networks, which combines graph convolution and gated temporal convolution units. Zhao et al. [23] propose a time graph convolution network (T-GCN) model to derive spatio-temporal correlations from traffic data. Zhang [24] proposed the dynamic graph convolution temporal representation learning model (DGCN-TRL), which captures the spatio-temporal dependencies of traffic flow by constructing a time series convolution framework, and designed a mask time representation learning module to extract long-term periodic features. Chen et al. [25] proposed a multi-scale spatio-temporal graph neural network for urban traffic flow forecasting, enhancing prediction accuracy through multi-scale feature decomposition and a unified spatio-temporal encoding mechanism. To address the limitation of existing methods in failing to holistically integrate spatio-temporal dependencies. Deng [26] proposed the spatio-temporal kernel graph convolutional network (STK-GCN), which leverages specially designed spatio-temporal kernels to construct spatial and temporal matrices. Within an encoder-decoder framework, this design enables the simultaneous modelling of spatial and temporal dependencies, thereby improving the effectiveness of traffic flow prediction. Reza et al. [27] proposed a transformer model based on multi-head attention mechanisms, aimed at more effectively capturing spatio-temporal correlations in traffic flow data and enhancing predictive performance. Ma et al. [28] proposed the spatio-temporal transformer (STTLM) to achieve traffic flow prediction, decoupling spatiotemporal features through a pure transformer architecture that integrates multi-layered embedding mechanisms. Zhou et al. [29] designed a fusion model,

MCST-Transformer, which combines the advantages of multi-channel structures and attention mechanisms, simultaneously models spatio-temporal associations and multidimensional features in traffic data, thereby effectively enhancing the accuracy of traffic flow prediction. Liu et al. [30] proposed the implicit information embedding and enhanced spatio-temporal multi-head attention transformer (IEEAFormer). By integrating implicit information from traffic data through an embedding layer, time-context-aware self-attention and a parallel spatial self-attention architecture, the model effectively captures multi-scale spatio-temporal dependencies.

However, while GCN excels at capturing the local topological structure of road networks and spatial correlations between nodes, its modelling capabilities in the temporal dimension remain relatively limited. This makes it difficult to effectively capture the dynamic evolution trends of traffic flow over medium- and long-term time scales, thereby weakening the model's ability to represent global temporal dependencies. In contrast, the transformer architecture demonstrates significant advantages in modelling long-range dependencies and capturing global context through its self-attention mechanism. Nevertheless, it exhibits insufficient sensitivity to local fine-grained features and short-term disturbance patterns – micro-level variations that play a crucial role in characterising sudden fluctuations in traffic flow and achieving high-precision predictions.

Therefore, an increasing number of studies are exploring the combination of GCN and transformer to build powerful spatio-temporal prediction frameworks [31-32]. For example, Lang et al. [33] proposed the generalised spatio-temporal regression graph convolutional transformer (GSTRGCT), which utilises dynamic adaptive spatial structure learning and periodic dependency modelling for traffic prediction. Hu et al. [34] proposed the GCN-Transformer model for load forecasting at electric vehicle battery swapping stations, enabling precise multi-dimensional analysis. Zhao et al. [35] proposed a hybrid architecture that effectively integrates graph convolutional networks and the transformer to enhance spatio-temporal sequence forecasting. The proposed model achieved superior performance in both traffic flow prediction and air quality monitoring tasks. Remmouche et al. [36] proposed a long-term spatio-temporal graph attention network (LSTGAN), which integrates convolutional temporal encoding, spatial attention and graph convolutional modules to achieve superior prediction performance, compared to existing baseline models across multiple standard datasets. Zhang et al. [37] proposed a deep learning model integrating transformers with multi-image convolutions to address the shortcomings of traditional methods in long-term forecasting and spatial correlation modelling. He et al. [38] proposed the multi-level causal attention spatio-temporal transformer (MLCAFormer), which captures multi-scale temporal dependencies through a hierarchical temporal causal attention mechanism while enhancing spatial correlation learning via node identity-aware spatial attention.

Despite significant advances in spatio-temporal forecasting models in recent years, there remain notable shortcomings in modelling coupled dynamic spatio-temporal dependencies. On the one hand, existing methods generally rely on predefined adjacency matrices to characterise spatial relationships between road segments, making it difficult to dynamically reflect the evolving spatial correlations driven by real-time traffic conditions. This limits the models' ability to express irregular, time-varying spatial dependency structures. On the other hand, some models employ fixed structures or simple weighting mechanisms when integrating spatio-temporal features, lacking explicit modelling of the deep interaction mechanisms between spatial topology and temporal dynamics. Such superficial fusion strategies often overlook nonlinear, high-order coupling effects present in spatio-temporal dimensions, making it difficult for models to capture the underlying spatio-temporal synergistic patterns within complex transportation systems. Therefore, the aforementioned shortcomings prevent the model from fully demonstrating its intended modelling and predictive capabilities when confronted with the dynamic and complex nature of real traffic systems.

To address the aforementioned issues, this paper systematically improves existing temporal and spatial prediction models from the perspectives of temporal modelling and structural fusion, proposing a multi-level dynamic GCN-Transformer framework (DGTFormer) for in-depth modelling of traffic flow spatial and temporal interactions. The model integrates the advantages of graph convolutional networks in modelling local spatial topology with transformer's ability to capture global temporal dependencies, and innovatively designs spatio-temporal dynamic graph convolutional network, temporal transformer encoder and spatio-temporal gated fusion mechanism, to capture spatial and temporal dependencies in the traffic data, and to realise the in-depth interaction and self-adaptive fusion of the temporal and spatial characteristics of the traffic flow information, which significantly improves the prediction performance and generalisation ability.

The main contributions of this research can be summarised as follows:

- 1) We introduce an adaptive dynamic graph convolutional network that effectively models deep, dynamically changing spatial dependency structures. At the same time, by jointly introducing the temporal dilated convolutional network, the evolutionary relationships between continuous time frames in the traffic road network topology are further revealed, providing a detailed description of the dynamic changes in road network topology.
- 2) We construct a temporal transformer encoder for temporal feature learning, leveraging the transformer's autocentric mechanism and its ability to capture long-distance dependencies, to efficiently model critical temporal dependencies in complex traffic flow sequences, thereby enhancing perception of global temporal features.
- 3) We propose a new multi-level dynamic GCN-Transformer framework, which achieves organic collaboration among modules, enhancing the model's comprehensive perception of dynamic spatial structures and temporal change characteristics. By designing fusion mechanisms, the model can jointly model the multiscale spatio-temporal interactions between traffic flows and road network structures, achieving deep fusion of spatio-temporal feature information and significantly enhancing predictive outcomes.

The subsequent section of this article is organised as follows: In the "Methodology" section, we provide a detailed introduction to the modelling methods for traffic flow spatio-temporal information, thereby constructing the overall architecture of the DGTFormer long- and short-term traffic flow combination prediction model. In the "Experiments and result analysis" section, we designed a series of experiments to verify the effectiveness and superiority of the proposed model. We conducted an in-depth analysis and discussion of the experimental results. Finally, in the "Conclusions" section, we summarise our entire work.

2. METHODOLOGY

In this section, we propose a novel DGTFormer prediction model, which is a spatio-temporal modelling framework specifically designed for traffic flow prediction tasks. In order to effectively capture the heterogeneous spatial and temporal dependency features in traffic data, DGTFormer adopts a dual-stream architecture design to achieve spatio-temporal feature decoupling modelling, thereby enhancing the model's temporal and spatial expression capabilities. We also provide a detailed description of the structure and functionality of the three core components in DGTFormer, including the spatio-temporal dynamic graph convolutional network, temporal transformer encoder and spatio-temporal gated fusion mechanism, further explaining how they work together in the overall model to jointly improve the model's predictive performance.

2.1 Traffic flow prediction model

Efficient and accurate traffic flow prediction is of great significance for urban traffic planning and management. To address the complex temporal and spatial correlation issues in traffic flow prediction, this paper proposes a traffic flow temporal and spatial interaction prediction model, based on a multi-level dynamic graph convolution and transformer fusion architecture, named DGTFormer. As shown in *Figure 1*, this figure depicts the overall structural framework of the prediction model DGTFormer. The DGTFormer prediction model mainly consists of three key functional modules: spatio-temporal dynamic graph convolutional network, temporal transformer encoder and spatio-temporal gated fusion mechanism. Each module plays a different role in modelling the complex spatial and temporal dependencies of traffic data, effectively extracting spatial and temporal features from dynamic spatio-temporal traffic data, thereby significantly improving the accuracy of traffic flow prediction.

In this model architecture, the spatio-temporal dynamic graph convolutional network and the temporal transformer encoder are each composed of N stacked identical spatio-temporal layers. This layered design facilitates multi-level feature extraction, enabling the network to represent changes in traffic dynamics better across different periods and temporal evolution. A spatio-temporal dynamic graph convolutional network is used to capture the spatial topology of road networks and their dynamic changes over time, enhancing the model's ability to perceive spatial correlations. The temporal transformer encoder focuses on capturing long-term temporal dependencies and dynamic temporal patterns in traffic flow sequences, thereby effectively modelling the evolution of traffic conditions over time. The spatio-temporal gated fusion mechanism dynamically weights and fuses spatial and temporal feature information to enhance the interaction and coupling between spatio-temporal information, thereby effectively improving the model's accuracy in predicting future traffic conditions.

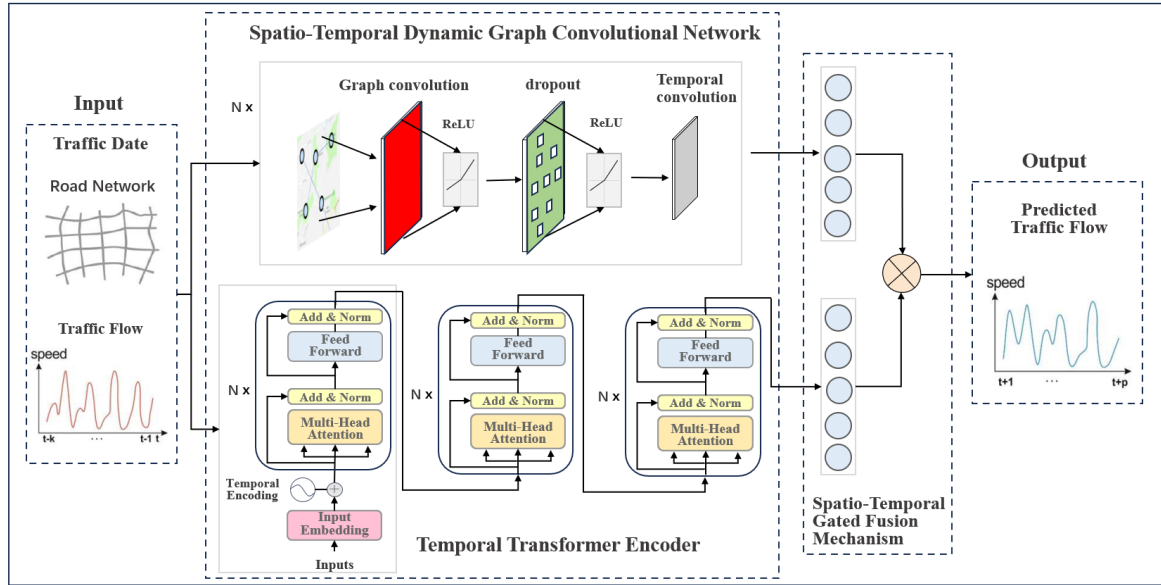


Figure 1 – DGTFFormer predictive model overall framework

The spatio-temporal dynamic graph convolutional network comprises two core components: adaptive dynamic graph convolutional network (ADGCN) and temporal dilated convolutional network (TDCN), which collaboratively model the dynamic and complex evolution of spatial structures over time in traffic networks. Among these, the ADGCN component excels in dynamically understanding the topological complexity of the graphics and can deeply portray the potential structural changes in the target road network, focusing on capturing the propagation paths and structural dynamics of the traffic flow at the spatial level [39]. While the TDCN component enables the network to adapt to variations in graph data characteristics across different periods, it provides a detailed portrayal of the dynamic evolution of the graph structures [40].

The temporal transformer encoder comprises multiple encoders with the same structure, each retaining key components such as multi-head self-attention and feedforward neural networks, thereby effectively enhancing the expressiveness and stability of temporal modelling. By deeply modelling historical traffic data, the temporal transformer encoder can fully tap into the inherent temporal correlations in the evolution of traffic flows, thereby efficiently extracting key temporal features.

Due to the adoption of a spatio-temporal decoupled modelling architecture, modules have a limited perception of each other’s outputs during the feature extraction phase, which may lead to insufficient coupling between spatial and temporal features. For this reason, this paper designs a spatio-temporal gated fusion mechanism to achieve deep interaction and fusion between spatial and temporal features, thereby enhancing overall predictive performance. It introduces a spatio-temporal cross-attention mechanism, which first interacts with the extracted spatial and temporal feature information to achieve deep coupling of spatio-temporal information. Then, a gating mechanism is employed to dynamically fuse the spatio-temporal features, after which the integrated representations are fed into a multilayer perceptron (MLP) to generate accurate predictions of future traffic states.

2.2 Spatio-temporal dynamic graph convolutional network

In terms of spatial dimensions, this paper proposes a spatio-temporal dynamic graph convolutional network aimed at capturing the complex spatial dependencies among nodes in traffic road networks and their evolving topology over time. This module can adaptively respond to topological changes in road networks caused by events such as road closures and construction, as well as their spatial dependencies.

We first provide the initial topology graph $G = (V, E, A)$ as the model input, where V represents the set of nodes, E represents the edge set connecting nodes with spatial connectivity, and A represents the adjacency matrix, reflecting the network’s spatial topology. At each time step t , the graph convolutional network receives the corresponding graph structure input and extracts spatial structure information under each time step through graph convolution operations. The specific calculations are as follows:

$$H_{\text{ANGCN},t}^{l+1} = F(H_t^l, A_t, W_t^{l+1}) = \sigma \left(\tilde{D}_t^{-\frac{1}{2}} \tilde{A}_t \tilde{D}_t^{-\frac{1}{2}} H_t^l W_t^{l+1} \right) \quad (1)$$

where: $H_t^{l+1} \in \mathbb{R}^{N \times d}$ represents the embedded output at the first layer of time step t ; $H_t^l \in \mathbb{R}^{N \times d}$ represents the original feature matrix of the input; A_t represents the normalised adjacency matrix; D_t represents the degree matrix of the graph; $W_t^{l+1} \in \mathbb{R}^{D \times D}$ represents the weight matrix updated at time step t ; σ represents the Sigmoid activation function.

In a graph convolutional network, the graph convolution operation enhances the local perception of node representations by modelling the adjacency between nodes to achieve an updated representation of the features of neighbouring nodes. As shown in Figure 2, each node generates a new node representation by weighted aggregation of the features of neighbouring nodes based on the adjacency relationship while retaining its own features. After the L -layer graph convolution module, the model can extract and fuse the topological structure information of the multi-order neighbours at each time step, so that each node representation not only retains its own features, but also significantly enhances the contextual relevance to its local nodes.

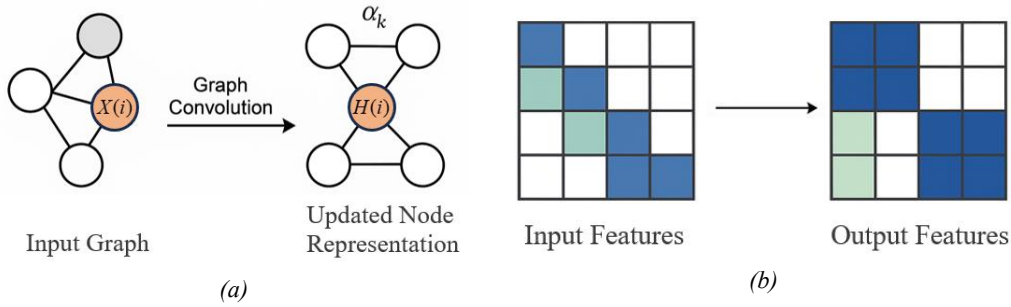


Figure 2 – Schematic diagram of the convolution operation process; a) Feature aggregation process; b) Feature matrix evolution diagram

Traditional graph convolutional network models assume that graph structures remain fixed, making it difficult to characterise dynamic changes in road networks. For this reason, we innovatively introduced a GRU gating mechanism into the graph convolutional network structure to achieve timing updates for graph convolution weights. This mechanism adaptively updates spatial weight parameters based on historical states and current inputs, dynamically adjusting the model to respond to structural changes. The weight update for the $l + 1$ layer at time step t is as follows:

$$Z_t^{l+1} = \text{sigmoid}(U_Z^{l+1}H_t^l + V_Z^{l+1}W_{t-1}^{l+1} + B_Z^{l+1}) \tag{2}$$

$$R_t^{l+1} = \text{sigmoid}(U_R^{l+1}H_t^l + V_R^{l+1}W_{t-1}^{l+1} + B_R^{l+1}) \tag{3}$$

$$\tilde{W}_t^{l+1} = \tanh(U_W^{l+1}H_t^l + V_W^{l+1}(R_t^{l+1} \circ W_{t-1}^{l+1}) + B_W^{l+1}) \tag{4}$$

$$W_t^{l+1} = G(H_t^l, W_{t-1}^{l+1}) = (1 - Z_t^{l+1}) \circ W_{t-1}^{l+1} + Z_t^{l+1} \circ \tilde{W}_t^{l+1} \tag{5}$$

The weight $l + 1$ at time step t is calculated by the weight update function G , which combines the input H_t^l with the weight W_{t-1}^{l+1} from the previous moment. Z_t^{l+1} , R_t^{l+1} and \tilde{W}_t^{l+1} are respectively the update gate, reset gate and candidate weights. The update process is controlled by update gate $Z_t \in [0,1]^{N \times N}$, dynamically blending the weights from the previous time step with the new candidate weights. Through the update mechanism, the network can dynamically adjust weights based on changes in input data, thereby achieving adaptive updates to the weights.

Additionally, to model the sequence evolution process of road networks more effectively over time dimensions, we introduce the temporal dilated convolutional network (TDCN) structure. TDCN can focus on local temporal characteristics under the backdrop of topological changes in road networks by introducing multiscale expanded convolutional kernels, effectively uncovering traffic flow propagation patterns triggered by topological shifts between nearby sections. For the output feature $H_{\text{TDCN},t}^{l+1}(n)$ of the $l + 1$ layer and n th segment node at time step t , it can be expressed as:

$$H_{\text{TDCN},t}^{l+1}(n) = \sum_{k=0}^{K-1} W_k^{l+1} \cdot H_{t-r_{l+1},k}^l(n) + b^{l+1} \tag{6}$$

where: $H_{TDCN,t}^{l+1}(n)$ represents the output characteristics of node n at time step t in $l + 1$ layer; $H_{t-r_{l+1},k}^l(n)$ represents the historical characteristics of the previous layer at the k th time step (according to vacancy rate r_{l+1} jumps); W_k^{l+1} represents the k th time convolution kernel in $l + 1$ layer; b^{l+1} represents $l + 1$ layer bias; r_{l+1} represents the vacancy rate in $l + 1$ layer; K represents convolutional kernel size.

To deeply analyse predictive target road network structure data and learn the temporal and spatial characteristics at different levels at different time steps, we combine the spatial graph convolutional network module (GCN + GRU) with the temporal dilated convolutional network module (TDCN) to form a unified spatio-temporal dynamic graph convolutional network. As shown in Figure 3, the GCN module aggregates spatial information of neighbouring nodes from bottom to top, while the GRU updates weight parameters from left to right over time dimensions, enabling joint modelling of spatial structure and time evolution to extract network spatial features.

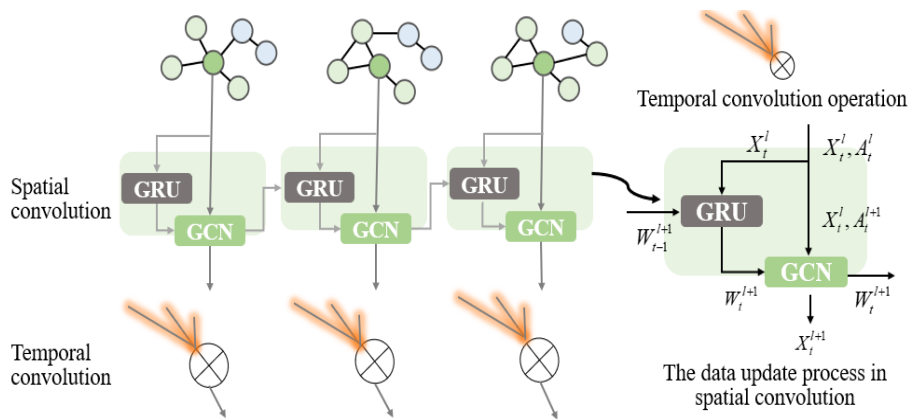


Figure 3 – Spatiotemporal dynamic GCN basic architecture

By interacting with and extracting temporal and spatial features of the road network, the final output’s feature representation integrates global topology with local evolutionary patterns, enabling precise modelling of dynamic changes in the road network topology. The final output result can be represented as:

$$H_{GCN,t}^{l+1} = \sigma(H_{ANGCN,t}^l + H_{TDCN,t}^l) \tag{7}$$

2.3 Temporal transformer encoder

In terms of temporal dimensions, we designed a temporal transformer encoder based on attention mechanisms, using an improved transformer architecture for temporal feature extraction. Compared to traditional transformer encoder-decoder architectures, we use a transformer encoder structure, enhancing the modelling capabilities for traffic flow data on different time scales through multi-layer stacking (as shown in Figure 4). This approach significantly simplifies the architecture and reduces computational complexity, thereby more efficiently extracting traffic flow timing features.

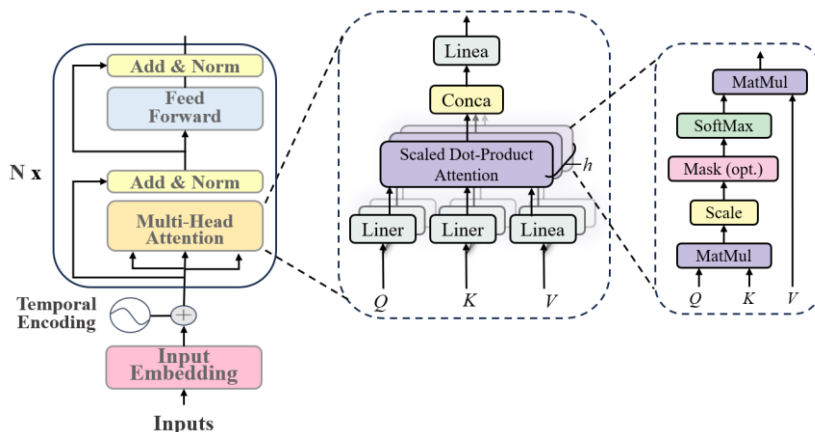


Figure 4 – Temporal transformer encoder main structure

The multi-head self-attention mechanism is a core component of the transformer module, modelling global dependencies across different time steps using multi-head attention structures to generate an effective global field of perception and enhance the extraction of global temporal features. The formula for calculating bullish attention is as follows:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (8)$$

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O \quad (9)$$

$$\text{where head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \quad (10)$$

For each attention head_{*i*}, the original input matrices Q , K , and V are mapped by parameterising the learnable matrix W^O , and utilising the three projection matrices $W_i^Q \in \mathbb{R}^{d_{\text{model}} \times d_k}$, $W_i^K \in \mathbb{R}^{d_{\text{model}} \times d_k}$ and $W_i^V \in \mathbb{R}^{d_{\text{model}} \times d_k}$, respectively, and then the self-attention operation is performed before splicing the outputs of each head to enhance the feature fusion by linear transformation. Meanwhile, residual linkage with layer normalisation is used inside the module to ensure gradient stability and training depth.

To supplement the temporal order information, we introduce temporal embedding before the input of the traffic flow sequence, and map the temporal index pos to the temporal vector by position encoding to incorporate semantic information for the temporal information of the sequence data, which is as follows:

$$PE(pos, 2i) = \sin(pos/10000^{2i/d_{\text{model}}}) \quad (11)$$

$$PE(pos, 2i + 1) = \cos(pos/10000^{2i/d_{\text{model}}}) \quad (12)$$

This results in a temporal embedding matrix $E_T \in \mathbb{R}^{|V| \times |T|}$, which provides location awareness for each time step and aids the prediction task.

2.4 Spatio-temporal gated fusion mechanism

Upon completion of spatial feature extraction from the road network and temporal feature modelling of traffic flow, we design a spatio-temporal gated fusion mechanism to achieve a deeper and effective interactive fusion of spatiotemporal features, as well as a final and accurate prediction of traffic state. Firstly, a spatiotemporal cross-attention mechanism is introduced to focus on the information flow between spatial and temporal characteristics of traffic flows, where the spatial feature $H_S \in \mathbb{R}^{N \times d_s}$ extracted by GCN and the temporal feature $H_T \in \mathbb{R}^{N \times d_t}$ extracted by transformer are mapped to a unified dimension d (i.e. $H'_S \in \mathbb{R}^{N \times d}$ and $H'_T \in \mathbb{R}^{T \times d}$), which are used as each other key-value pairs for spatial-guided temporal attention, and temporal-guided spatial attention, respectively, which in turn interact with each other's extracted feature information:

Spatially guided temporal attention:

$$Q_T = H'_T W_Q^T, K_S = H'_S W_K^S, V_S = H'_S W_V^S \quad (13)$$

$$A_{T \leftarrow S} = \text{Softmax}\left(\frac{Q_T K_S^T}{\sqrt{d}}\right) \quad (14)$$

$$\tilde{H}_T = A_{T \leftarrow S} V_S + H'_T \quad (15)$$

Temporally guided spatial attention:

$$Q_S = H'_S W_Q^S, K_T = H'_T W_K^T, V_T = H'_T W_V^T \quad (16)$$

$$A_{S \leftarrow T} = \text{Softmax}\left(\frac{Q_S K_T^T}{\sqrt{d}}\right) \quad (17)$$

$$\tilde{H}_S = A_{S \leftarrow T} V_T + H'_S \quad (18)$$

To achieve adaptive saliency weighting, a gated fusion mechanism is introduced to dynamically weigh the importance of the output spatial features \tilde{H}_S and temporal feature \tilde{H}_T , among the spatio-temporal features, so

that the spatial and temporal features are combined in an optimal ratio. This step controls the output by the gated weights g , which are computed as follows:

$$H_f = g \circ \tilde{H}_S + (1 - g) \circ \tilde{H}_T \quad (19)$$

The fusion feature H_f contains more comprehensive information about both the spatial structure of the road network and the temporal characteristics of the traffic flow. Finally, this fusion feature is mapped to the output layer through the MLP to perform high-precision prediction and output the prediction result with the following expression:

$$\hat{Y} = \text{MLP}(H_f) = W_2(\text{ReLU}(W_1 H_f + b_1)) + b_2 \quad (20)$$

This mechanism can fully exploit the complex spatio-temporal dependencies of traffic data to achieve highly accurate and robust dynamic traffic state prediction.

3. EXPERIMENTS AND RESULT ANALYSIS

In this section, we conducted a series of comparative experiments based on three real-world datasets (PeMSD4, PeMSD7 and PeMSD8) to comprehensively validate the predictive performance of the proposed DGTFormer model. At the same time, we introduced four evaluation metrics – RMSE, MAE, MAPE and R^2 to perform a quantitative analysis of the experimental results. The experimental setup primarily consists of the following three parts: 1) Introducing six representative baseline models for comparative experiments, to evaluate the overall predictive capability of DGTFormer; 2) Designing ablation experiments, to analyse the specific contributions of each component module in the DGTFormer framework; 3) Conducting an effectiveness analysis of the GRU component, to further evaluate the adaptability and effectiveness of the GRU module in the face of dynamic changes in road network topology.

3.1 Datasets description

In this paper, three publicly available real datasets, PeMSD4, PeMSD7 and PeMSD8, are selected for experimental analysis and experimental validation of the proposed DGTFormer model. *Table 1* shows the basic statistical information of these three datasets.

Table 1 – Description of the datasets

Typology	Datasets	Sensor set	Frontier set	Timeframe
Highway traffic flow	PeMSD4	307	340	2018-01-01-2018-02-28
	PeMSD7	883	1,025	2012-05-01-2012-06-30
	PeMSD8	170	295	2016-07-01-2016-08-31

These datasets are derived from the Los Angeles Area Freeway Traffic Monitoring System (LAFTS) in California, with PeMSD4, PeMSD7 and PeMSD8 corresponding to traffic flow monitoring data from three different regions of California, respectively. These data are collected continuously by sensors deployed across the roadway network at a frequency of once every 30 seconds, and then the data samples are aggregated at 5-minute intervals. Therefore, each traffic node will correspond to 12 time series observations in each hour, providing high-frequency, high-quality experimental data support for subsequent spatio-temporal modelling and multi-step forecasting tasks.

3.2 Baselines

In order to comprehensively evaluate the performance advantages of the proposed DGTFormer model in the traffic flow prediction tasks, six representative baseline models are introduced in this paper for comparative experiments. The selected baseline models cover three categories of mainstream methods. The first category is traditional time series methods, including ARIMA [4] and LSTM [41]. The second category encompasses deep learning temporal modelling approaches, such as the transformer [21] and CNN-Transformer [42]. The third category includes spatio-temporal prediction models, represented by T-GCN [23] and STGCN [22]. This selection encompasses multiple representative approaches, including classical sequence models, deep learning-based sequence models and spatio-temporal fusion models, enabling a multidimensional, multi-level systematic comparison.

3.3 Settings and hyperparameters

All experiments are conducted on a system equipped with an Intel i5-13500HX CPU, 16.0 GB RAM, and an NVIDIA RTX 4050 GPU, based on Python 3.12. To ensure the representativeness of the experimental data, this study only used traffic flow data from weekdays for analysis. We used a grid search strategy to determine the optimal hyperparameter combination on the validation set, trained the model using the Adam optimiser, and set the training cycle to 100 epochs. The experiment employed a sliding window mechanism, utilising historical traffic flow data from the previous 12 time steps to predict traffic conditions for the next 3, 6 and 12 time steps. To ensure temporal consistency between model training and evaluation, all datasets were divided into training (60%), validation (20%) and testing (20%) sets based on time order, with a batch size of 32.

3.4 Evaluation metrics

The experiments in this paper adopt four evaluation metrics to comprehensively evaluate the prediction performance of all the models: mean absolute error (MAE), root mean squared error (RMSE), mean absolute percentage error (MAPE) and the coefficient of determination R^2 . Specifically, MAE measures the average absolute deviation between the predicted values and the true values, reflecting the overall error level of the prediction results. RMSE assigns a higher weight to larger errors, which makes it more sensitive to the volatility of prediction deviations. MAPE eliminates the influence of the scale and magnitude, and is suitable for the relative comparison of errors under different scales. R^2 , the coefficient of determination, evaluates the goodness of fit between predicted and actual values, indicating the degree of linear correlation. The formulas for the above indicators are shown between *Equations 21 and 24*:

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{Y} - Y_i| \quad (21)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{Y} - Y_i)^2} \quad (22)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{\hat{Y} - Y_i}{\hat{Y}} \right| \times 100\% \quad (23)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (\hat{Y} - Y_i)^2}{\sum_{i=1}^N (Y_i - \bar{Y})^2} \quad (24)$$

where: N denotes the total number of traffic flow data samples; \hat{Y} denotes the predicted value; Y_i denotes the true value of the sample and \bar{Y} denotes the average of the true value.

3.5 Prediction results and analysis

The prediction performance of the proposed DGTFormer interactive fusion model, which is compared with various baseline methods on the highway traffic flow datasets (PeMSD4, PeMSD7 and PeMSD8), is shown in *Tables 2-4*. As shown by the experimental results in the table, the DGTFormer model consistently achieves superior performance compared to baseline methods across all the prediction time steps (15, 30 and 60 minutes) and evaluation metrics (RMSE, MAE and MAPE). In particular, compared with the current state-of-the-art baselines, our approach achieves performance improvements of up to 8.61% and 9.15% in RMSE and MAE, respectively. The results fully validate the outstanding performance of the proposed model in the task of traffic flow prediction under large-scale road networks, further demonstrating its generalisation ability and robustness in spatio-temporal dependence modelling.

For the traditional statistical model ARIMA, it performs the worst in the experiments due to its strong a priori assumptions on the model structure, limited modelling capability, and difficulty in portraying the complex nonlinear spatio-temporal dependencies in the traffic flow. Especially in the long-term prediction (60-minute) task, its RMSE indicator rises significantly, which further validates the limitations of statistical models in dealing with complex traffic dynamics.

The single sequence models, LSTM and transformer, have some improvement compared with the traditional models, but due to the lack of modelling capability of spatial topology, they fail to effectively capture the spatial correlation features of traffic flow, and the overall performance still lags behind the spatio-temporal map model. The hybrid structural model CNN-Transformer is between the spatio-temporal graph model and the pure sequence model in terms of prediction performance, which indicates that the local feature extraction ability of CNN can supplement the expression ability of the transformer to a certain extent, but it still fails to adequately capture the complex spatio-temporal relationships across nodes and scales in the traffic flow.

The spatio-temporal graph models T-GCN and STGCN effectively integrate the spatial correlation information between nodes through the graph convolution mechanism, which significantly improves the prediction performance. However, T-GCN is difficult to adapt to the dynamic changes of traffic flow due to its static graph convolution structure, and its performance is still limited in complex temporal patterns. Although STGCN introduces spatio-temporal convolution blocks to jointly model spatio-temporal features, the local limitation of the convolutional operation leads to a performance bottleneck in long-term prediction tasks.

In contrast, the DGTFormer model proposed in this paper integrates the spatial structure modelling capability of a graph convolutional network and the strong temporal modelling advantage of the transformer, and introduces the spatio-temporal cross-attention mechanism and the gating fusion mechanism, which constructs a unified spatio-temporal interaction fusion framework. After experimental verification, the model shows superior performance in both long- and short-term prediction tasks, with more stable prediction performance and higher accuracy, which fully reflects its strong generalisation and modelling ability in complex traffic scenarios.

Table 2 – Model evaluation metrics on the dataset PeMSD4

Metrics		RMSE			MAE			MAPE(%)		
		15min	30min	60min	15min	30min	60min	15min	30min	60min
PeMSD4	ARIMA	15.11	17.25	22.09	12.68	15.03	19.33	14.58	15.85	17.73
	LSTM	14.40	14.82	15.58	13.88	14.34	14.77	12.85	13.53	14.02
	Transformer	13.78	14.87	15.32	12.69	13.95	14.19	12.76	13.62	13.94
	CNN-Transformer	13.25	14.17	15.29	12.36	13.30	13.85	12.35	13.11	13.79
	T-GCN	12.37	13.42	13.82	12.69	12.82	12.98	10.72	11.15	11.68
	STGCN	12.31	13.39	13.58	12.31	12.37	12.54	10.51	10.64	10.97
	DGTFormer	11.93	12.36	13.01	11.75	11.81	12.40	8.59	9.30	10.56

Table 3 – Model evaluation metrics on the dataset PeMSD7

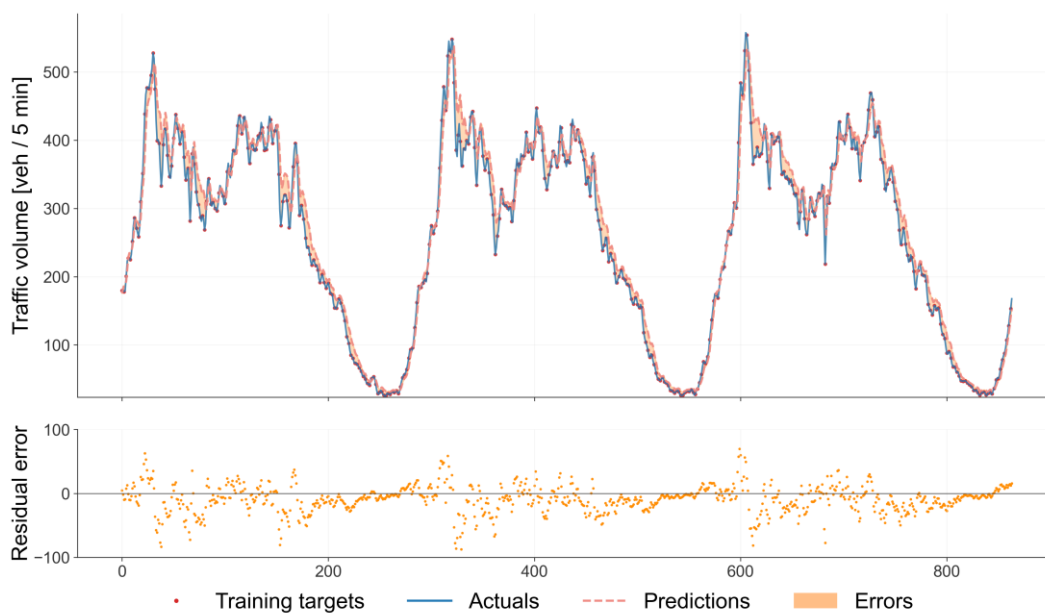
Metrics		RMSE			MAE			MAPE(%)		
		15min	30min	60min	15min	30min	60min	15min	30min	60min
PeMSD7	ARIMA	17.50	19.86	25.98	15.09	16.58	19.20	14.22	16.51	17.10
	LSTM	15.57	16.43	17.03	14.86	15.62	16.77	13.59	13.94	14.31
	Transformer	15.93	16.41	16.72	14.94	15.57	15.71	11.14	10.97	11.36
	CNN-Transformer	14.29	14.65	15.94	13.24	14.08	14.68	10.72	10.47	11.94
	T-GCN	13.88	14.21	15.01	12.73	12.99	13.62	9.37	10.24	10.96
	STGCN	13.85	14.06	14.36	12.67	12.83	13.07	9.19	9.90	10.83
	DGTFormer	12.26	12.85	13.14	11.51	11.90	12.19	9.23	9.58	10.71

Table 4 – Model evaluation metrics on the dataset PeMSD8

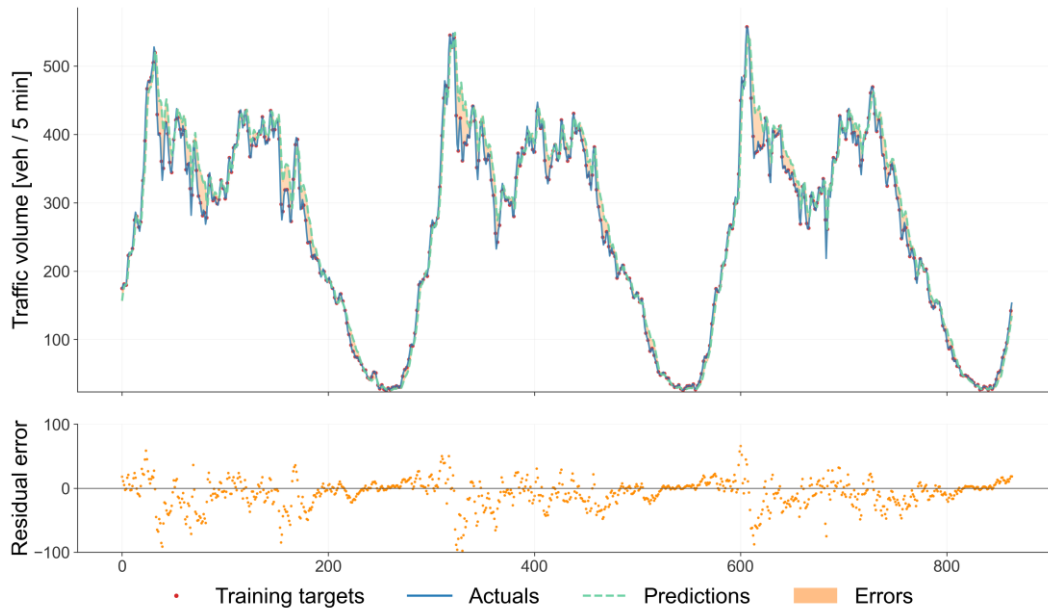
Metrics		RMSE			MAE			MAPE(%)		
		15min	30min	60min	15min	30min	60min	15min	30min	60min
PeMSD8	ARIMA	15.63	16.19	21.94	12.81	14.86	19.05	14.24	15.36	16.90
	LSTM	14.22	14.58	15.54	13.05	13.79	14.37	11.71	13.02	13.60
	Transformer	13.31	14.19	14.71	12.94	13.58	13.80	11.93	12.27	12.87
	CNN-Transformer	13.96	14.11	14.50	12.88	13.54	13.61	10.21	10.33	11.54
	T-GCN	12.90	13.17	13.73	12.32	13.58	14.13	9.46	9.92	10.83
	STGCN	12.57	13.16	13.38	12.25	12.37	12.92	9.25	9.81	9.75
	DGTFormer	12.30	12.80	13.18	11.56	11.97	12.50	8.89	9.15	9.20

To further verify the prediction ability of the proposed model at the single-node level, we randomly select a sensor node in the target road network, predict the traffic flow of the node on the PeMSD4 dataset, and visualise and analyse the prediction results under multiple prediction horizons. The experimental results are shown in Figures 5a–5c. From the figure, it can be observed that the model is able to fit the real traffic flow trend more accurately and shows good prediction performance under different prediction time steps (15, 30 and 60 minutes). In particular, for the 15-minute prediction task, the prediction curve of the model is highly consistent with the real value, which can closely track the fluctuation of the actual traffic flow, indicating that the model has a very high accuracy and responsiveness in short-term prediction.

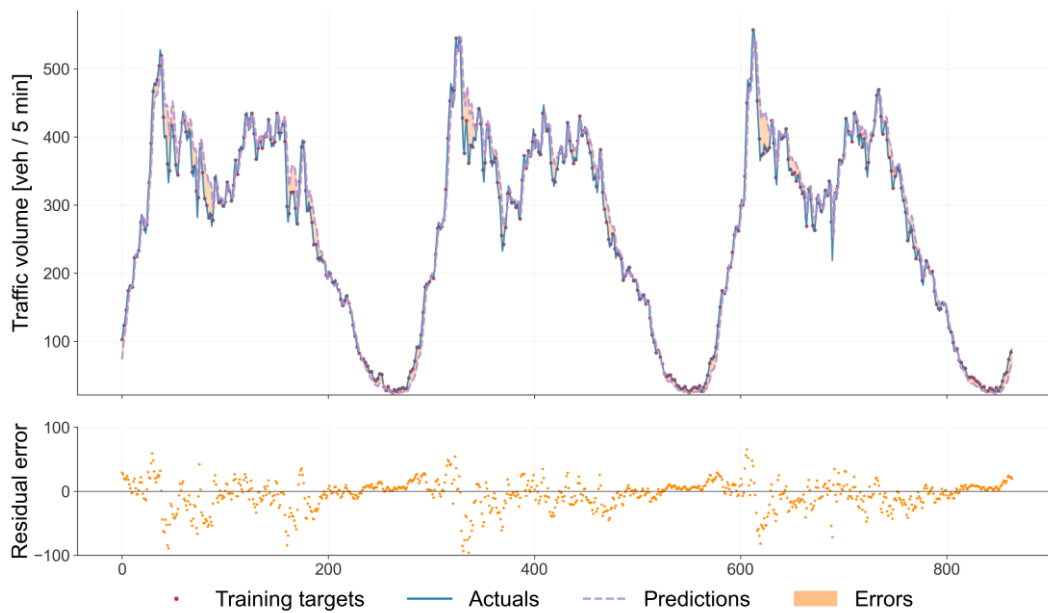
As the forecasting time span increases, the model’s prediction error rises slightly, as evidenced by the deviation of the prediction curve from the true value in the 30-minute and 60-minute prediction time steps. This trend is consistent with the commonly observed error accumulation phenomenon in time series prediction tasks. However, benefiting from the introduction of the transformer encoding structure and the multi-head self-attention mechanism, the proposed model effectively captures long-range temporal dependencies. As a result, even in medium- and long-term forecasting scenarios, the model maintains strong fitting capability. The predicted curves generally follow the overall trends of actual traffic flow, with relatively small error deviations, demonstrating the model’s robust temporal modelling capacity and high resilience to prediction drift.



(a) 15-minute



(b) 30-minute



(c) 60-minute

Figure 5 – A figure with three subplots; a) Comparison of time series predictions at 15-minute time steps; b) Comparison of time series predictions at 30-minute time steps; c) Comparison of time series predictions at 60-minute time steps

Meanwhile, in order to reflect more intuitively the fitting ability of the model in different prediction time ranges, this paper uses scatter plots to visualise the prediction results and introduces the coefficient of determination R^2 as an indicator of the model’s explanatory power. The results are shown in Figures 6a–6c. The figure shows the prediction performance of DGTFormer under 15-minute, 30-minute and 60-minute prediction tasks, respectively. As shown in Figures 6a–6c, DGTFormer exhibits excellent forecasting performance in different prediction time steps (15, 30 and 60 minutes), and the coefficients of determination are always maintained at a high level of 0.9497, 0.9369 and 0.9071, which reflects its high accuracy in short-term prediction and strong generalisation and robustness in medium- and long-term forecasting tasks. This verifies the effectiveness and robustness of the proposed model in modelling the complex spatio-temporal dependencies of traffic flow.

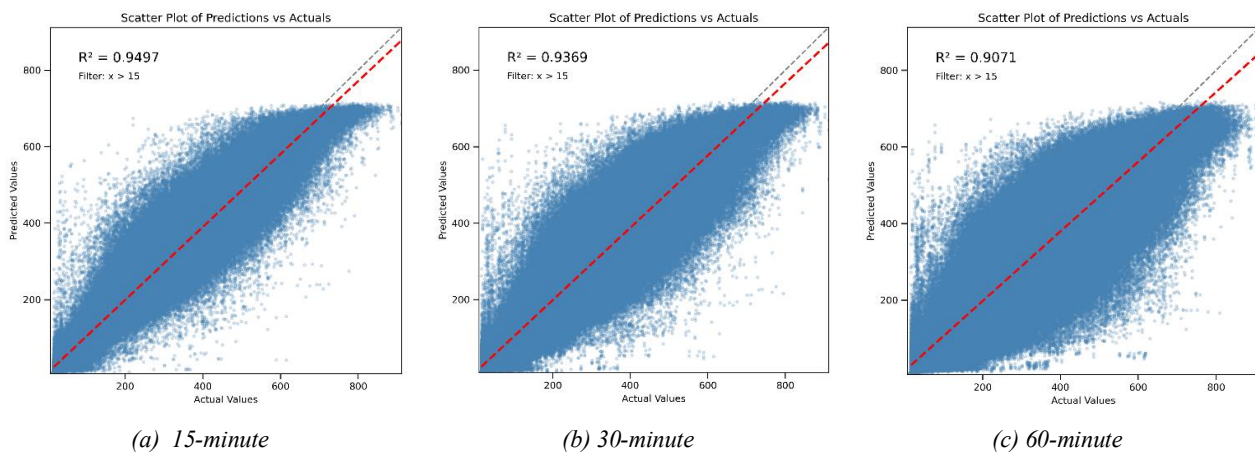


Figure 6 – A figure with three subplots; a) Prediction scatter plot at the 15-minute time step; b) Prediction scatter plot at the 30-minute time step; c) Prediction scatter plot at the 60-minute time step

3.6 Ablation study

In order to further validate the contribution of each key component in the model to the impact of model prediction performance, this paper designs and implements a series of systematic ablation experiments on the PeMSD4 dataset. By gradually removing the core components in the DGTFFormer model, the impact of each component on the model’s performance is analysed. In this paper, we specifically constructed one of the following four variants of the model for comparison:

- 1) DGTFFormer-T: Removal of the temporal transformer encoder from the complete DGTFFormer architecture to evaluate its role in modelling temporal dependencies.
- 2) DGTFFormer-G: Remove the spatio-temporal dynamic graph convolutional network to analyse the contribution of spatial dependency modelling capabilities to prediction performance.
- 3) DGTFFormer-F: Remove the spatio-temporal gated fusion mechanism that integrates the spatio-temporal cross-attention mechanism, and only use simple weighted summing and splicing for spatio-temporal feature fusion to test the effect of fusion strategy on the modelling effect.
- 4) DGTFFormer-E: Removal of the temporal location coding embedding module and analysis of the role and necessity of explicit time-aware mechanisms in time modelling.

These four variant models are analysed in comparison with the DGTFFormer model designed in this paper for ablation experiments, and all the variant models are trained and tested under the same experimental setup as DGTFFormer, and the experimental results are shown in Figures 7a–7c.

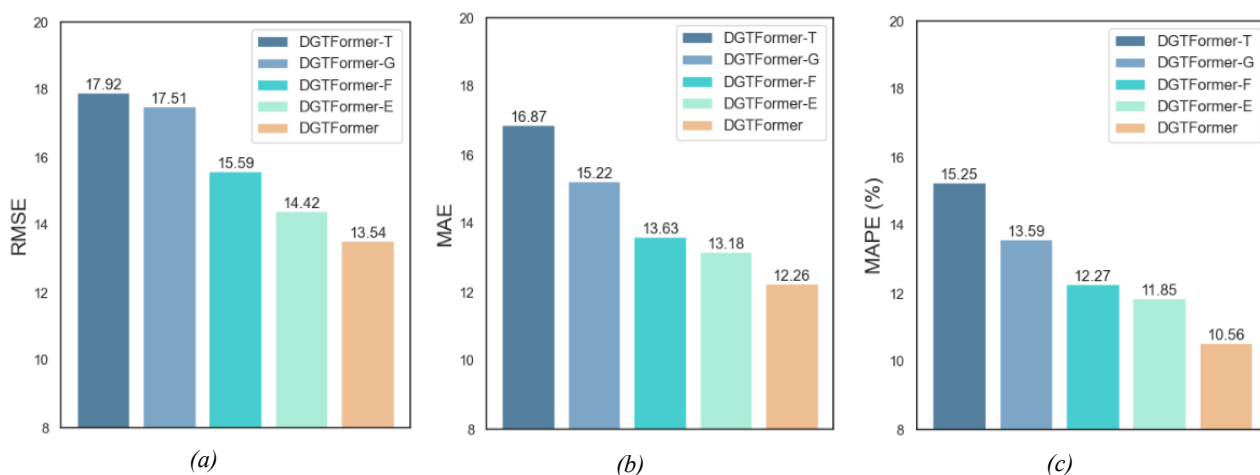


Figure 7 – A figure with three subplots; a) RMSE-based performance comparison; b) MAE-based performance comparison; c) MAPE-based performance comparison

The experimental results show that the prediction performance of both DGTFormer-T and DGTFormer-G models is significantly lower than that of the full model, especially DGTFormer-T with the temporal transformer module removed. This result fully indicates that the temporal transformer plays a key role in capturing long-term dependencies. Meanwhile, the model’s ability to model spatial structure and temporal dynamics is further verified, which is of great significance in improving the modelling effect of complex spatio-temporal dependencies. Secondly, the performance of DGTFormer-F and DGTFormer-E also decreases compared with the full model, indicating that the gated fusion module of the spatio-temporal cross-attention mechanism and the temporal location encoding embedding are also important for the model performance. The former helps to fuse multi-scale spatio-temporal features more efficiently, while the latter enhances the model’s ability to perceive and understand the sequence of events in a time series, which is particularly important for improving prediction accuracy.

In summary, the results of the four ablation experiments fully validate the rationality and necessity of the constituent modules in the DGTFormer architecture and provide theoretical and practical support for high-precision traffic flow prediction.

3.7 GRU effectiveness analysis

In order to further verify the adaptive ability of the GRU weight updating mechanism introduced in the model to cope with the topological changes of the road network, this paper designs and carries out a node masking experiment on the PeMSD4 dataset. Specifically, the experiment randomly blocks 5% of the nodes to simulate the real-life scenarios of local network unavailability or failure due to construction, traffic accidents and other emergencies, thus triggering dynamic changes in the network topology. *Figures 8a–8c* show the adjacency weight matrix during the process of road network topology changes, illustrating the evolution of the road network topology structure at different time steps ($t_1=0$, $t_2=10$, $t_3=20$). The depth of colour in the matrix blocks reflects the strength of the connection weights between nodes, intuitively demonstrating the impact of dynamic topology perturbations on the graph structure.

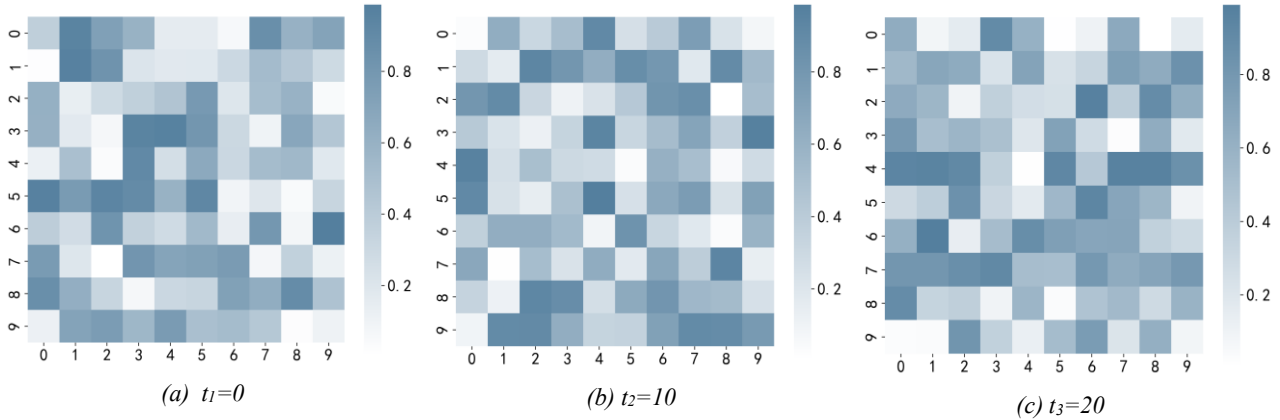


Figure 8 – A figure with three subplots; a) Heatmap of the adjacency weight matrix representing the initial graph topology at $t_1 = 0$; b) Heatmap of the adjacency weight matrix representing the graph topology at $t_2 = 10$; c) Heatmap of the adjacency weight matrix representing the graph topology at $t_3 = 20$

In this paper, a systematic comparison is made with the full model proposed in this paper (denoted as Ours-GRU) by constructing a disturbed experimental scenario with 5% node shielding and designing comparative experiments in which the GRU module in the model is replaced with StaticGCN (static graph convolutional network) and MLPUpdate (multilayer perceptron update module), respectively. The experiments continue to evaluate the degree of performance degradation of each scheme under topological perturbation conditions in terms of three metrics, namely, RMSE, MAE and MAPE. The experimental results are shown in *Figure 9*.

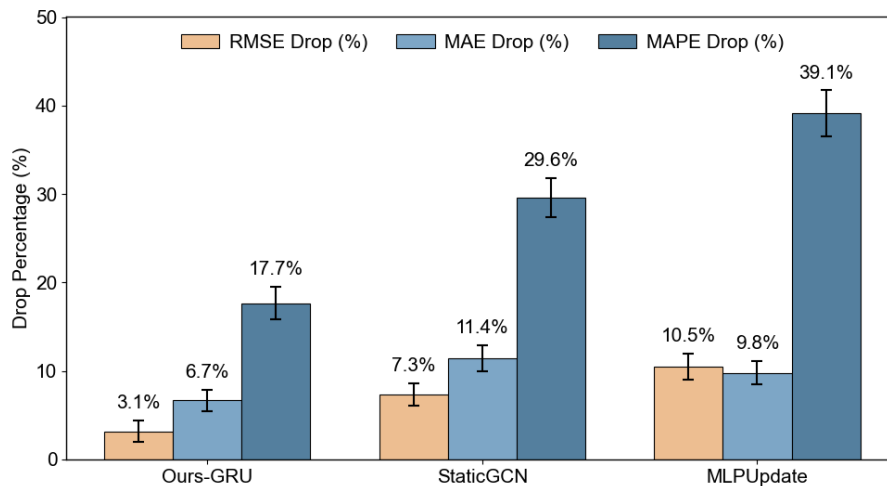


Figure 9 – Comparison of prediction performance degradation in node masking scenarios

From the experimental results, it can be seen that the prediction performance of each model is degraded to different degrees under the condition of perturbation of the road network topology. Among them, the StaticGCN and MLPUpdate models are particularly degraded under the node masking scenario, showing poor robustness. In comparison, the Ours-GRU model still maintains the optimal performance under the node masking condition, showing good perturbation resistance. This is because the local topological breaks induced by node masking propagate layer by layer in the network through the graph convolution operation, leading to global feature distortion and thus affecting the overall prediction accuracy of the model.

The GRU module introduced in Ours-GRU can effectively alleviate this problem, which drives the adaptive adjustment of the adjacency matrix through the dynamic state update mechanism, compensates for the information loss caused by missing nodes to a certain extent, suppresses the hierarchical accumulation of errors, and strengthens the model's adaptive ability to topological perturbations. Specifically, the performance degradation of Ours-GRU in the disturbance scenario is minimised, and the RMSE and MAE only increase by 3.1% and 6.7%, respectively, which is significantly better than the other comparative methods. This result fully verifies the effectiveness of the dynamic structure learning mechanism of GRU in improving the robustness of the model, and also highlights its potential to cope with sudden topological changes in complex traffic scenarios.

4. CONCLUSION

This study proposes DGTFormer, a high-precision deep learning model framework for traffic flow prediction. The model integrates a spatio-temporal dynamic graph convolutional network and a temporal transformer encoder. It achieves decoupled modelling of temporal and spatial aspects by building a spatio-temporal dual-stream architecture. This setup effectively captures complex dependencies in urban traffic flows over time and space. It also enhances both long-term and short-term prediction accuracy. The model incorporates a spatio-temporal gated fusion mechanism to deeply integrate spatial and temporal information, further improving its ability to represent dynamic changes in road topology and traffic flow patterns. The DGTFormer framework achieves high prediction accuracy and provides an efficient, robust modelling paradigm for urban traffic flow prediction. The specific research conclusions are as follows:

- 1) Our research institute developed the DGTFormer model, which combines the ability of dynamic graph convolution to model local spatial features with the advantage of the transformer in modelling global temporal series. This effectively enhances the model's capabilities in spatio-temporal dependency modelling. The model can adaptively capture dynamic changes in traffic network structures and learn complex nonlinear spatio-temporal dependency relationships, balancing short- and long-term prediction requirements.
- 2) The experimental results demonstrate that the DGTFormer model has outstanding predictive capabilities for three real-world traffic datasets: PeMSD4, PeMSD7 and PeMSD8. Systematic comparative experiments show that this model significantly outperforms existing mainstream methods in key evaluation metrics, such as RMSE, MAE and MAPE. Compared with the most advanced baseline models currently available, the DGTFormer model achieves performance improvements of up to 8.61% and 9.15%

in RMSE and MAE, respectively. Furthermore, the coefficient of determination, R^2 , consistently exceeds 0.9 across various prediction time steps, confirming the DGTFormer model's exceptional predictive accuracy and generalisation capabilities.

- 3) Through comparative experiments, we verified that incorporating a spatial weight update mechanism based on GRU into the DGTFormer model dynamically adjusts the adjacency strength between nodes based on historical traffic conditions. This makes the model more accurately depict the evolution of the topological structure of the traffic network over time. This suggests that the mechanism is crucial for improving the model's ability to adapt to non-stationary traffic patterns.

The experimental results of the DGTFormer model framework on real traffic datasets indicate that this method has excellent predictive performance and dynamic adaptability for traffic flow prediction tasks. However, the current research has limitations and requires further exploration. Currently, the model relies primarily on historical traffic flow parameter data for modelling and has not yet considered the impact of external environmental factors, such as weather conditions and holidays, on dynamic changes in traffic flow. Future research could incorporate multi-source, heterogeneous traffic data to develop a more comprehensive feature representation system, thereby improving the model's applicability and reliability in real-world scenarios. Additionally, future work can further explore the dynamic evolution laws of traffic flow from a local, micro perspective in terms of temporal feature modelling. Designing a multi-scale temporal feature extraction mechanism could enhance the model's ability to capture micro-dynamic changes and macro-evolutionary laws, aiming to achieve more accurate prediction results.

ACKNOWLEDGEMENTS

The research in this article was funded by the Key R&D Program in Guangxi (AB24010221, AB24010124), National Natural Science Foundation of China, grant No. 52462044 and 52202383.

REFERENCES

- [1] Zhang K, et al. A deep learning based multitask model for network-wide traffic speed prediction. *Neurocomputing*. 2020;396:438-450. DOI: [10.1016/j.neucom.2018.10.097](https://doi.org/10.1016/j.neucom.2018.10.097).
- [2] Han SY, et al. Adaptation scheduling for urban traffic lights via FNT-based prediction of traffic flow. *Electronics*. 2022;11(4):658. DOI: [10.3390/electronics11040658](https://doi.org/10.3390/electronics11040658).
- [3] Chen W, et al. Dynamic Path Planning Based on Traffic Flow Prediction and Traffic Light Status. *International Conference on Algorithms and Architectures for Parallel Processing 2023 Oct 20. Singapore: Springer Nature Singapore*. 2023. p. 419-438. DOI: [10.1007/978-981-97-0834-5_24](https://doi.org/10.1007/978-981-97-0834-5_24).
- [4] Siroos S, et al. Ensemble of ARIMA: combining parametric and bootstrapping technique for traffic flow prediction. *Transport metrika A Transport Science*. 2020;16(3):1552-1573. DOI: [10.1080/23249935.2020.1764662](https://doi.org/10.1080/23249935.2020.1764662).
- [5] Pei-Lin N, Zhi YU, Zhao-Cheng. Constrained kalman filter combined predictor for short-term traffic flow. *Journal of Traffic and Transportation Engineering*. 2008;8(5):86-90.
- [6] Guo HF, Fang LJ, Yu L. A short-term traffic flow prediction model based on fuzzy Kalman filtering. *Journal of Zhejiang University of Technology*. 2013;41(2):218-221.
- [7] Wang Z, et al. A novel hybrid method for achieving accurate and timeliness vehicular traffic flow prediction in road networks. *Computer Communications*. 2023;209:378-386. DOI: [10.1016/j.comcom.2023.07.019](https://doi.org/10.1016/j.comcom.2023.07.019).
- [8] Mrudula ST, et al. Internet of things and optimized knn based intelligent transportation system for traffic flow prediction in smart cities. *Measurement: Sensors*. 2024;35:101297. DOI: [10.1016/j.measen.2024.101297](https://doi.org/10.1016/j.measen.2024.101297).
- [9] CHEN Z, et al. Short-term traffic state prediction approach based on FCM and random forest. *Journal of Electronics and Information Technology*. 2018;40 (8):1879-1886. DOI: [10.11999/JEIT171090](https://doi.org/10.11999/JEIT171090).
- [10] Xia Y, Chen J. Traffic flow forecasting method based on gradient boosting decision tree. *2017 5th International Conference on Frontiers of Manufacturing Science and Measuring Technology (FMSMT 2017)*. Atlantis Press. 2017. p. 413-416. DOI: [10.2991/fmsmt-17.2017.87](https://doi.org/10.2991/fmsmt-17.2017.87).
- [11] CUI Jian-xun, et al. Review on short-term traffic flow prediction methods based on deep learning. *Journal of Traffic and Transportation Engineering*. 2024;24(2):50-64. DOI: [10.19818/j.cnki.1671-1637.2024.02.003](https://doi.org/10.19818/j.cnki.1671-1637.2024.02.003).
- [12] Y. Lv, et al. Traffic Flow Prediction With Big Data: A Deep Learning Approach. *IEEE Transactions on Intelligent Transportation Systems*. 2015;16(2):866-875. DOI: [10.1109/TITS.2014.2345663](https://doi.org/10.1109/TITS.2014.2345663).

- [13] He Y, et al. In-depth insights into the application of Recurrent Neural Networks (RNNs) in traffic prediction: A comprehensive review. *Algorithms*. 2024;17(9):398. DOI: [10.3390/a17090398](https://doi.org/10.3390/a17090398).
- [14] Song Ruirong, Wang Binjun, Tong Xin. Traffic flow forecasting model based on multi dimensional spatial and temporal characteristics. *Science Technology and Engineering*. 2021;21(31):13439-13446.
- [15] Shao HX, Soong BH. Traffic flow prediction with long short-term memory networks(lstms). *Proceedings of 2016 IEEE Region 10 Conference. Singapore: IEEE*. 2016:2986-2989. DOI: [10.1109/TENCON.2016.7848593](https://doi.org/10.1109/TENCON.2016.7848593).
- [16] Luo W, Dong B, Wang Z. Short-term traffic flow prediction based on CNN-SVR hybrid deep learning model. *Journal of Transportation Systems Engineering and Information Technology*. 2017;17(5):68-74.
- [17] Li Y, et al. Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting. *In International Conference on Learning Representations (ICLR)*, 2018. DOI: [10.48550/arXiv.1707.01926](https://doi.org/10.48550/arXiv.1707.01926).
- [18] Wang X, Xu L. Short-term traffic flow prediction based on deep learning. *Journal of Transportation Systems Engineering and Information Technology*. 2018;18(1):81-88. DOI: [10.16097/j.cnki.1009-6744.2018.01.012](https://doi.org/10.16097/j.cnki.1009-6744.2018.01.012).
- [19] Zhang J, et al. Short-term origin-destination demand prediction in urban rail transit systems: A channel-wise attentive split-convolutional neural network method. *Transportation Research Part C: Emerging Technologies*. 2021;124:102928. DOI: [10.1016/j.trc.2020.102928](https://doi.org/10.1016/j.trc.2020.102928).
- [20] Pareja A, et al. EvolveGCN: Evolving Graph Convolutional Networks for Dynamic Graphs. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2020;34(04):4542-4549. DOI: [10.1609/aaai.v34i04.5984](https://doi.org/10.1609/aaai.v34i04.5984).
- [21] Vaswani A, et al. Attention is all you need. *Advances in Neural Information Processing Systems 30 (NIPS 2017)*. Long Beach, CA, USA: Curran Associates, Inc.; 2017. p. 5998-6008. DOI: [10.48550/arXiv.1706.03762](https://doi.org/10.48550/arXiv.1706.03762).
- [22] Yu B, Yin H, Zhu Z. Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting. *Proceedings of the 27th International Joint Conference on Artificial Intelligence*; 2018. DOI: [10.24963/ijcai.2018/505](https://doi.org/10.24963/ijcai.2018/505).
- [23] Zhao L, et al. T-GCN: A temporal graph convolutional network for traffic prediction. *IEEE transactions on intelligent transportation systems*. 2019;21(9):3848-3858. DOI: [10.1109/TITS.2019.2935152](https://doi.org/10.1109/TITS.2019.2935152).
- [24] Zhang A. Dynamic graph convolutional networks with Temporal representation learning for traffic flow prediction. *Scientific Reports*. 2025;15(1):1-17. DOI: [10.1038/s41598-025-01696-7](https://doi.org/10.1038/s41598-025-01696-7).
- [25] Chen, Hui, et al. Multi-scale spatio-temporal graph neural network for urban traffic flow prediction. *Scientific Reports*. 2025;15(1):26732. DOI: [10.1038/s41598-025-11072-0](https://doi.org/10.1038/s41598-025-11072-0).
- [26] Deng H. Traffic-Forecasting Model with Spatio-Temporal Kernel. *Electronics*. 2025;14(7): 1410. DOI: [10.3390/electronics14071410](https://doi.org/10.3390/electronics14071410).
- [27] Reza S, Marta C, José J M. A multi-head attention-based transformer model for traffic flow forecasting with a comparative analysis to recurrent neural networks. *Expert Systems with Applications*. 2022;202:117275. DOI: [10.1016/j.eswa.2022.117275](https://doi.org/10.1016/j.eswa.2022.117275).
- [28] Ma J, Zhao J, Hou Y. Spatial–Temporal Transformer Networks for Traffic Flow Forecasting Using a Pre-Trained Language Model. *Sensors*. 2024;24(17):5502. DOI: [10.3390/s24175502](https://doi.org/10.3390/s24175502).
- [29] Zhou CH, Lin PQ. Traffic volume prediction method based on multi-channel transformer. *Comput. Appl. Res*. 2023;40:435-439. DOI: [10.19734/j.issn.1001-3695.2022.06.0306](https://doi.org/10.19734/j.issn.1001-3695.2022.06.0306).
- [30] Liu, S., Wang, X. An improved transformer based traffic flow prediction model. *Scientific Reports*. 2025;15(1):8284. DOI: [10.1038/s41598-025-92425-7](https://doi.org/10.1038/s41598-025-92425-7).
- [31] Xing Z, et al. Spatial linear transformer and temporal convolution network for traffic flow prediction. *Scientific Reports*. 2024;14(1):4040. DOI: [10.1038/s41598-024-54114-9](https://doi.org/10.1038/s41598-024-54114-9).
- [32] Xu M, et al. Spatial-temporal transformer networks for traffic flow forecasting. *arxiv preprint arxiv:2001.02908*; 2020. DOI: [10.48550/arXiv.2001.02908](https://doi.org/10.48550/arXiv.2001.02908).
- [33] Xiong L, et al. Generalized spatial–temporal regression graph convolutional transformer for traffic forecasting. *Complex Intell. Syst*. 2024;10:7943–7964. DOI: [10.1007/s40747-024-01578-x](https://doi.org/10.1007/s40747-024-01578-x).
- [34] Hu X, et al. GCN-Transformer-Based Spatio-Temporal Load Forecasting for EV Battery Swap Stations under Differential Couplings. *Electronics*. 2024;13(17): 3401. DOI: [10.3390/electronics13173401](https://doi.org/10.3390/electronics13173401).
- [35] Zhao Y, Lu J. Spatiotemporal Sequence Prediction Based on Spatiotemporal Self-Attention Mechanism. *International Journal of Computers Communications & Control (December)*. 2024;19(6). DOI: [10.15837/ijccc.2024.6.6771](https://doi.org/10.15837/ijccc.2024.6.6771).
- [36] Remmouche B, et al. Long-Term Spatio-Temporal Graph Attention Network for Traffic Forecasting. *Expert Systems with Applications*. 2025;128244. DOI: [10.1016/j.eswa.2025.128244](https://doi.org/10.1016/j.eswa.2025.128244).
- [37] Zhang J, et al. Spatio-temporal transformer and graph convolutional networks based traffic flow prediction. *Scientific Reports*. 2025;15(1):24299. DOI: [10.1038/s41598-025-10287-5](https://doi.org/10.1038/s41598-025-10287-5).

- [38] He H, et al. Spatio-temporal transformer traffic prediction network based on multi-level causal attention. *PLoS One*. 2025;20(9):e0331139. DOI: [10.1371/journal.pone.0331139](https://doi.org/10.1371/journal.pone.0331139).
- [39] Chen D, et al. Two-stream spatio-temporal GCN-transformer networks for skeleton-based action recognition. *Scientific Reports*. 2025;15(1):4982. DOI: [10.1038/s41598-025-87752-8](https://doi.org/10.1038/s41598-025-87752-8).
- [40] Lv Z, et al. ST-TDCN: A two-channel tree-structure spatial-temporal convolutional network model for traffic velocity prediction. *Expert Systems with Applications*. 2024;257: 125053. DOI: [10.1016/j.eswa.2024.125053](https://doi.org/10.1016/j.eswa.2024.125053).
- [41] Lu H, et al. A temporal-aware LSTM enhanced by loss-switch mechanism for traffic flow forecasting. *Neurocomputing*. 2021;427:169-178. DOI: [10.1016/j.neucom.2020.11.026](https://doi.org/10.1016/j.neucom.2020.11.026).
- [42] Li S, Yang L, Zhao X. Prediction Algorithm of Short-term Traffic Congestion in Urban Area based on Deep Learning. *Science Technology and Engineering*. 2023;23(25):10866-10878.