Research Article

UrbanFlow-GST: A Dynamic Graph-Based Spatio-Temporal Fusion Model for Real-Time Traffic Speed Forecasting

Muhammad Salik Salam [1*], Zohaib Ahmad [2]

[1] Department of Computer Science, University of Alabama at Birmingham, 1402 10th Avenue S., Birmingham, AL 35294, USA
[2] Department of Criminology and Forensic Sciences, Lahore Garrison University, Lahore, Pakistan
[*]Corresponding Author: Muhammad Salik Salam Author Email: msalam@uab.edu

Received: October 19, 2025 **Accepted:** April 28, 2026

Abstract: Real-time and accurate traffic forecasting forms the basis of intelligent transportation systems, especially in densely populated areas where traffic conditions keep varying with time depending on complicated spatio-temporal correlations. Conventional as well as deep learning-based approaches tend to lack simultaneous consideration of spatio-temporal dynamics, thus rendering them inefficient in the practical setting. In light of this problem, this paper presents UrbanFlow-GST, an adaptive graph learning and attention-based modeling approach for traffic forecasting. The proposed system builds upon a graph and implements spatio-temporal transformer, using a dynamic adjacency mechanism that takes into account time-dependent associations among traffic sensors. For capturing temporal dynamics, the authors adopt the transformer architecture, enabling them to learn both short- and long-term traffic characteristics. Finally, a unified method is developed for combining spatial and temporal features to make multi-horizon predictions. To test the effectiveness of the proposed framework, experiments were conducted using the METR-LA dataset consisting of real-world traffic speeds collected via sensors installed at 207 locations on highways of Los Angeles. The model outperformed competing baselines at all horizons under evaluation. In particular, the authors report MAE, RMSE, and MAPE values of 2.52, 4.90, and 8.1%, respectively, for 15-minute prediction, while the same metrics for 60-minute forecasting are 3.20, 6.80, and 10.3%. The findings suggest high reliability of UrbanFlow-GST.

Keywords: Traffic Forecasting; Spatio-Temporal Modeling, Graph Neural Networks; Transformer; Graph Attention Network; Dynamic Graph Learning, Intelligent Transportation Systems; Urban Traffic Analysis; Time Series Prediction; METR-LA Dataset

1. Introduction

Traffic congestion in large cities today is something everyone experiences regularly as part of their daily lives, with impacts extending beyond just delays in reaching destinations and including increased fuel consumption and pollution, among others. The increase in the number of roads and vehicles makes effective traffic management vital in any major city of today. Therefore, traffic forecasting helps facilitate intelligent transport systems by providing necessary traffic information [1]. The primary challenge of traffic forecasting lies in the inherent complexity of traffic data. Traffic data of a certain region does not exist in isolation from each other. Rather, it is dependent on trends in past traffic along with the behavior of other roadways around it. This leads to very high spatio-temporal dependency among traffic data. Prior techniques, which were based on statistics, were incapable of accurately modeling the non-linear relationship in traffic data due to simplifications made [2].

As a result of the development of deep learning, many models such as LSTM and its variants have been used extensively to model temporal dependency in traffic data. Despite the improvement over previous models, temporal dependency was the primary focus of these models, ignoring spatial structures in traffic networks [3]. This led to a series of new models, based on graphs. Here, traffic sensors are considered nodes of the graph, while edges represent their connectivity. Despite improving spatial dependency over earlier models, they rely heavily on prior knowledge of node connectivity, making them ineffective as traffic changes rapidly. More recently, transformer-based models have shown strong performance in capturing long-range temporal dependencies through attention mechanism [4]. These models can better learn patterns over longer time periods, but they typically lack an effective way to represent dynamic spatial relationships. As a result, there remains a gap in combining both aspects spatial and temporal into a single, unified framework that can adapt to changing traffic conditions [5] [6].

In order to solve this issue, in this paper, a graph-based spatio-temporal transformer network named UrbanFlow-GST is presented to model both spatial relations and temporal relations in a joint and more flexible way. Specifically, the work makes use of the well-known METR-LA dataset, which includes real-world traffic speed observations collected from a total of 207 sensors deployed along highways in the city of Los Angeles. This dataset includes realistic and various traffic behaviors, such as periodic traffic patterns, rush-hour phenomena, and other traffic anomalies [7]. As shown in Fig.1, the traffic network is modeled by the graph and the traffic sensors serve as the vertices in it. Dynamic graph learning is utilized to learn the relations between traffic sensors. Simultaneously, a transformer-based module learns different temporal relations at different time resolutions. Based on that, the spatial-temporal representation is achieved and is used to predict the traffic state of each sensor in a multi-step way. Therefore, the purpose of the current study is to design a spatiotemporal traffic flow prediction method to better reflect the characteristics of real-world traffic. More specifically, the method takes into account the dynamic relationship between sensors as well as the multi-resolution time information. By doing so, the proposed model is expected to achieve better performance than existing methods.

2. Literature Review

Traffic forecasting plays an essential role in modern intelligent transport systems (ITS), providing for effective traffic management, congestion reduction, and routing optimization. Early research was mainly concerned with statistical and machine learning solutions; nevertheless, they had poor results since traffic data exhibited nonlinear and dynamic features [8]. Later on, the development of deep learning provided new possibilities to analyze temporal dependencies in traffic sequence, but this problem was mainly solved neglecting spatial relationships of roads [9]. Thus, more recent research has attempted to create frameworks for traffic flow forecasting that combine spatial and temporal factors. There has been proposed a spatial-temporal graph neural network capable of modeling inter-road dependencies, which is done by applying graph representation of traffic networks, where each sensor forms a node and connections between them are described via edges [10]. The model employs positional attention layer to integrate local neighbors and utilizes sequential layers to account for temporal aspects. Nonetheless, its ability to predict accurately long-term dependencies leaves much to be desired.

Further improvements can be made through development of hybrid models for integration of different kinds of traffic data within the single model. As an example, taxi demand prediction has been studied using Hybrid Spatio-Temporal Graph Convolutional Network (H-STGCN) [11]. H-STGCN utilizes the data of the traffic volume generated from the navigation system, which is then transformed into the travel time representation using the transformer-like method [12]. Moreover, the concept of compound adjacency matrices has been applied to better reflect traffic proximity. However, the use of predefined graphs represents the most obvious limitation of all the hybrid approaches. Considering the inability of recurrent neural networks to process global temporal dependencies, scientists have moved toward the attention-based and transformer architectures [13]. To address the issue of forecasting long sequences, researchers have created the Spatial-Temporal Graph-Informer Network (STGIN) by combining the Graph Attention Networks (GAT) and the Informer architecture [14]. The application of attention mechanisms has helped overcome the bottleneck of the information processing, thus allowing for capturing the dependencies between nodes both in space and over the course of time.

Following the same path, some scientists explored the possibilities of integration of graph neural networks and transformers, applying the latter for modeling spatial and temporal dependencies simultaneously. As a result, cloud-based hybrid GNN-Transformer model was created with the purpose of enhancing traffic flow prediction through inclusion of external features (weather, holidays, accidents, etc.) [15]. Cloud platforms proved to have the advantage of providing better scalability and real-time adaptability due to fast implementation. Moreover, some other transformer models were used for modeling traffic through multi-scale temporal modeling and features fusion for better capturing of periodic dependencies [16].

Models were proposed for the enhancement of the spatiotemporal learning process through incorporating local and global dependencies into consideration. The Dynamic Global-Local Spatial-Temporal Network (DGLSTNet), for instance, used dilated convolutions to model multi-scale temporal patterns and included global temporal attention to address global dependencies [17]. Furthermore, new models for learning dynamic graphs were introduced to capture varying relationships between traffic nodes without requiring an unchanging adjacency matrix[18]. An additional trend in traffic prediction research was represented by methods involving memory mechanisms and context-awareness features. Models that involved memory learning mechanisms used memory tensors and historical pattern banks to capture long-term global information, thereby increasing the robustness of the model. Meanwhile, memory attention mechanisms enabled the dynamic modeling of traffic patterns and inter-node relationships throughout history [19] [20].

While the previous trends concerned the improvements in traffic prediction accuracy, recent research began focusing on other issues such as multi-task learning and anomaly awareness. In particular, the Anomaly-Aware Spatio-Temporal Graph Attention Network (AA-STGAT) enabled traffic prediction and the detection of anomalies at the same time. Such joint optimization led to an increase in traffic predictability and ability to react to unusual situations (e.g., accidents and congestion peaks) [21] [22]. Transformer-based advancements further extended into trajectory prediction and mobility modeling. Models such as deep spatial-temporal transformer networks leveraged geographical and semantic location information to predict future destinations, highlighting the applicability of transformer architectures beyond conventional traffic flow prediction. Similarly, lane-aware and interaction-driven transformer frameworks incorporated detailed road topology and vehicle interactions to enhance prediction reliability in complex urban scenarios [23].

Efficiency and scalability issues related to deep spatio-temporal models have been investigated recently. Peak-aware graph-attention temporal fusion transformers and adaptive graph attention convolutional networks have been designed to address efficiency-related problems without reducing prediction accuracy. The studies showed that designing more efficient methods through parameter reduction and adaptation was essential for deploying edge-based traffic forecasting systems [24]. Machine learning and deep learning models are having versatile nature. Authors also worked on load prediction mechanism through deep learning models [32]. In addition, hybrid approaches combining state-of-the-art methods like diffusion convolution, meta-learning, and self-supervised learning were proposed to increase generalization and adaptation capabilities. For instance, meta-learning models have been demonstrated to improve performance in domains with limited data availability and increased robustness [25]. Diffusion models, on the other hand, were able to capture better the complex dependencies in space and time. Another emerging trend in traffic analysis was related to incorporating LLMs into graph-based approaches for better context and temporal reasoning [26] [29] [33].

Notwithstanding the aforementioned progress in the field, some critical issues have yet to be addressed. First, many approaches use static or partly adaptive graph structure that may not sufficiently describe dynamic spatial interactions between different variables [27]. Second, although transformer-based approaches excel at modeling long-term temporal dependencies, they struggle with efficiency and scalability problems in large urban networks. Third, there is currently a lack of unified approaches supporting multiple prediction horizons, dynamic spatial relationships, and congestion levels estimation [28]. In order to overcome these shortcomings, this research introduced UrbanFlow-GST, an innovative Graph

Spatio-Temporal Transformer framework that aims to learn the dynamic spatial interactions and multi-scale temporal information present within urban traffic networks. Through the combination of Graph-based learning methods with Transformer models and multi-horizon forecasting techniques, the UrbanFlow-GST framework seeks to offer a more reliable and scalable solution for real-time traffic predictions.

2.1 Research Gap

However, despite advances in traffic forecast modeling and analysis, there were also numerous problems that needed to be solved. First, traditional graph methods used fixed or predetermined adjacency matrices. These matrices did not effectively reflect dynamic connections between road sections at different times. Although adaptive and combined matrixes have already been proposed, they still fail to take into account the constantly evolving relationships in the case of abnormal conditions. The second issue related to the integration of transformers with graph data structures. Even though models based on transformers were capable of modeling long-range temporal relations, they had not yet become an optimal combination of graph elements and attention mechanisms. As a result, the spatio-temporal interaction was poorly considered [29] [30].

Moreover, most studies focused on forecasting either in short- or long-term time frames but did not consider multi-horizon prediction. In particular, it was vital to simultaneously predict traffic patterns at multiple time steps (e.g., in 15, 30, and 60 minutes). However, current models had difficulty implementing multi-step forecasting algorithms due to inadequate performance [30]. Some works included external features into the input vector, such as weather data or event schedules. Despite the usefulness of this approach, external factors were often used statically, i.e., without being integrated into the attention mechanisms or dynamic graphs. Thus, the system was not able to efficiently respond to changes in environmental conditions. One more limitation was the increased computation cost of transformer-based models. Due to a quadratic attention function, they experienced serious difficulties when applied to large-scale networks.

Finally, a very small percentage of studies paid attention to joint learning of traffic prediction and traffic pattern understanding, such as anomaly detection and congestion level. The vast majority of approaches assumed forecasting as a single task, ignoring potential auxiliary inputs.

To summarize, despite the progress made, transformer-based approaches suffered from several shortcomings. They include poor adaptation of graph representations, inefficient interaction between spatial structure and long-term forecasting, inability to solve a multi-horizon forecasting problem efficiently, absence of real-time adaptation, failure to use auxiliary inputs, and high computational complexity.

In an effort to address the above-mentioned shortcomings, the current study suggested a new framework called UrbanFlow-GST which was designed to incorporate dynamic spatial correlations with the help of graph learning, modeling of multi-scale temporal dependencies with the help of transformer-based attention mechanism, multiple horizon forecasting within the same framework, and effective spatio-temporal feature fusion using the attention mechanism.

3. Proposed Methodology: UrbanFlow-GST

In this work, an innovative framework called UrbanFlow-GST was formulated to tackle the difficulties in traffic forecasting in urban areas. The methodology was formulated mainly based on the idea of identifying the complex spatiotemporal dynamics in traffic systems, especially within the METR-LA dataset. Instead of learning from only spatial or only temporal relationships, the innovation behind this framework was the simultaneous learning of both, with a combination of graphs for spatial dynamics and transformers for temporal dynamics. In terms of methodology, there were multiple stages involved. These included data acquisition and preprocessing, graph generation, spatial feature learning, temporal relationship learning, feature fusion, and finally multi-horizon traffic forecasting. In each step, efforts were made to overcome the limitations of existing frameworks. The METR-LA dataset was selected as the primary data source due to its realistic representation of urban traffic conditions. It consists of traffic speed measurements collected from 207 sensors deployed across highways in Los Angeles, recorded at 5-minute intervals. This dataset exhibits strong spatial correlations among sensors and complex temporal patterns such as rush hours and irregular congestion events, making it highly suitable for evaluating advanced spatio-temporal models.

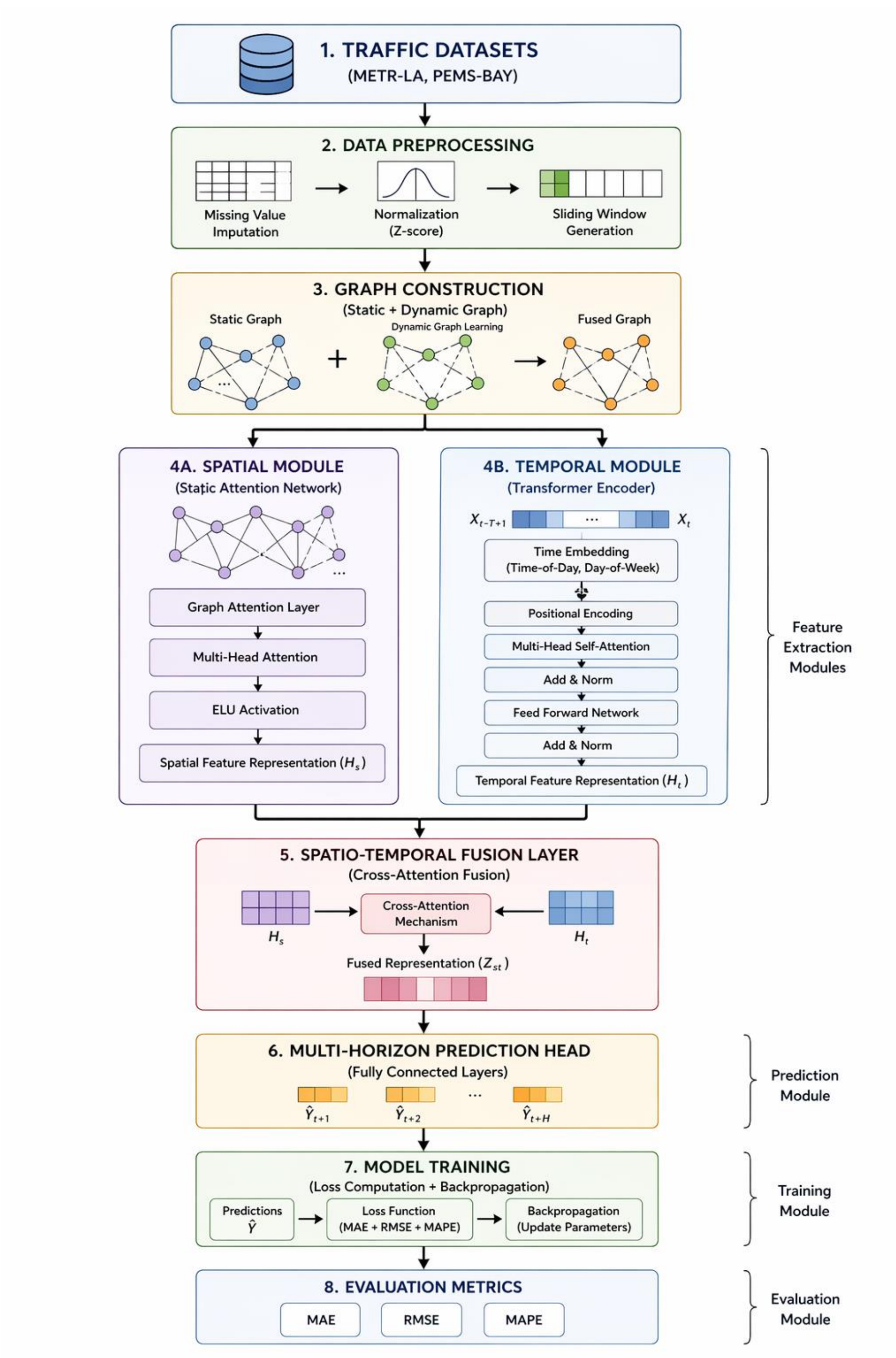

Figure 1 Proposed Architecture of Urban flow-GST

3.1 Dataset Description

METR-LA is a real-world dataset collected from 207 loop detectors on highways in Los Angles (LA) country, California. The data contains the average speed (mph) aggregate in different time of intervals.

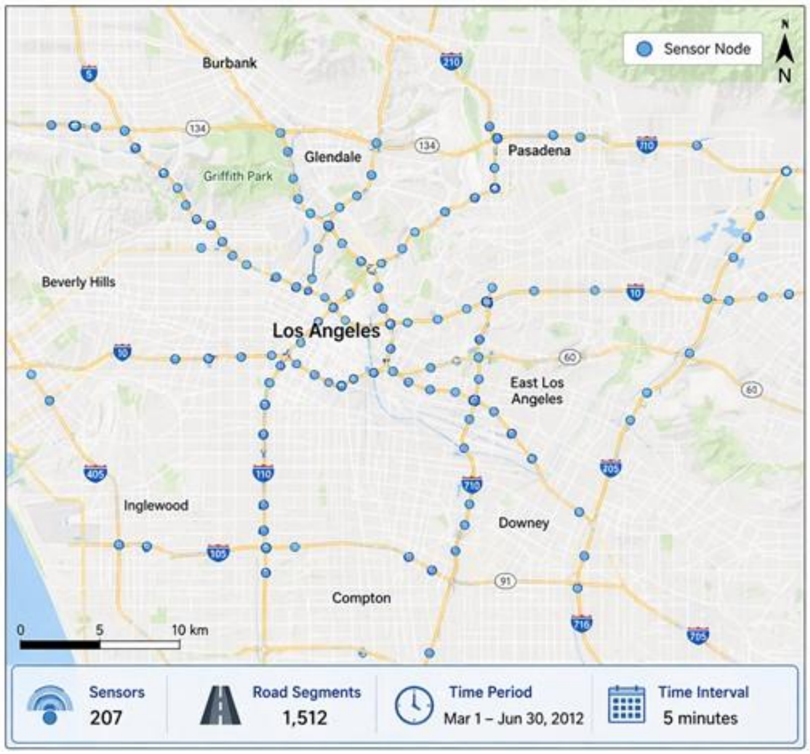

Figure 2 Real-world METR-LA Map view

As shown in Figure 2 below, there is a holistic depiction of the METR-LA dataset. It captures not only the location of the sensors but also the statistical properties of the traffic data.

On the left-hand side of the figure, there is a geographic map of the area of Los Angeles, with the nodes being the traffic sensors located in major highway areas in LA. Each node represents a loop detector on specific road segments. As seen in this figure, these traffic sensors are dense in high-traffic locations, indicating strong spatial correlations between the neighboring nodes.

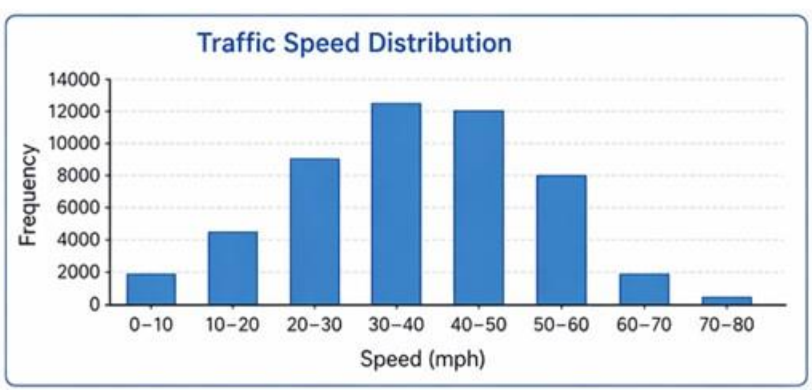

Figure 3 Traffic Distribution Histogram of METR-LA

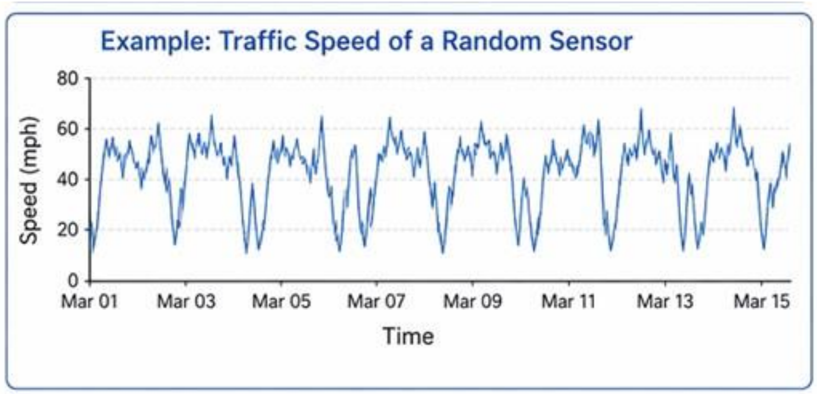

Figure 4 Traffic speed via Sensors

In the above figure, several components related to the statistical property of the dataset have been displayed. In the first box, we can see a summary of several important statistical attributes, such as the number of sensors, time duration, sampling interval, and feature type. Moreover, the histogram shows the frequency of traffic speed with respect to the speed level category. It shows that the number of occurrences of moderate speed levels is much higher, while very low or very high speed levels are observed less often. Additionally, the figure also presents a line chart of traffic speed for one sensor over time.

3.2 Data Representation and Preprocessing

The raw traffic data provided by METR-LA were first converted to a structured tensor format for learning purposes. The traffic data can be expressed as:

$$X \epsilon \mathbb{R}^{T \times N \times F} \tag{1}$$

Where T is the number of historic time steps, N = 207 is the number of sensors, and F refers to the number of features that include mainly the speed of the traffic.

To ensure stable and efficient training, the data was normalized using Z-score normalization:

$$X' = \frac{X-\mu}{\sigma} \tag{2}$$

where μ and σ are the mean and standard deviation of the data, respectively. This is because it was necessary to minimize the variability in the data.

A sliding window technique was subsequently adopted for transforming the time series into training data for supervised learning. In particular, the past T data points were utilized for predicting the subsequent H data points:

$$X_{input} = \{x_{t-T+1}, \dots, x_t\}, \;\; Y = \{x_{t+1}, \dots, x_{t+H}\} \tag{3}$$

This formulation allowed the model to learn temporal dependencies effectively.

3.3 Graph Construction and Dynamic Spatial Modeling

$$A_{ij}{}^{static} = \exp\left(\frac{d_{ij}{}^{2}}{\sigma^2}\right) \tag{4}$$

Where d_{ij} is the distance between nodes i and j. However, static graphs fail to capture dynamic traffic interactions.

To address this limitation, a dynamic graph learning mechanism was introduced. Learnable nodes E_1 and E_2 were used to compute adaptive relationships

Where d_{ij} refers to the distance between two nodes i and j. Nevertheless, static graphs cannot represent dynamic relations in traffic. This drawback is overcome by using dynamic graph learning. In dynamic graph learning, learnable nodes E_1 and E_2 are used to calculate adaptive relations.

$$A_{\mathrm{adaptive}=\mathrm{Softmax}(\mathrm{ReLU}(E_1 E_2{}^{T}))} \quad (5)$$

The final adjacency matrix was obtained by combining static and adaptive graphs

$$A = A_{adaptive} + A_{static} \quad (6)$$

This approach enabled the model to capture both physical connectivity and latent correlations that evolve over time.

3.3.1 Spatial Feature Extraction

The spatial relationships encoded in the graph were modeled using a Graph Attention Network (GAT). For each node iii, attention coefficients were computed to determine the importance of neighboring nodes:

$$\alpha_{ij} = \frac{\exp\,(LeakyReLU(a^T[Wh_i||WH_J]))}{\sum_{K\epsilon N(i)} \exp\,(LeakyReLU(a^T[Wh_i||WH_k]))} \quad (7)$$

The node representation was then updated as:

$$h'_i = \sigma(\textstyle\sum_{j\epsilon N(i)} \alpha_{ij} W h_j) \quad (8)$$

This mechanism allowed the model to focus on the most relevant neighboring sensors, improving spatial feature learning. Multi-head attention was applied to enhance stability and representation power.

3.3.2 Temporal Dependency Modeling

For modeling the temporal dependencies, an encoder based on transformers was used. Transformers are able to learn long-range dependencies through self-attention; unlike recurrent models. Considering the spatial features H_S, the following matrices were calculated:

$$Q = H_S W_Q,\;\; K = H_S W_k,\;\; V = H_S W_V \quad (9)$$

The attention mechanism was defined as:

$$H_{attn} = Softmax(\frac{QK^T}{\sqrt{d_k}})V \quad (10)$$

To better capture both short-term fluctuations and long-term trends, a dual-attention mechanism was introduced:

$$H_{short}, H_{long} = Attention(Q, K, V) \quad (11)$$

These representations wer fused using a gating mechanism:

$$\gamma = \sigma\big(W_g\big[H_{short}; H_{long}\big] + b_g\big) \quad (12)$$

$$H_t = \gamma \odot H_{short} + (1-\gamma) \odot H_{long} \quad (13)$$

This design enabled the model to adaptively balance short-term and long-term temporal patterns.

3.3.3 Spatio-Temporal Fusion

To integrate spatial and temporal information, a fusion mechanism was applied:

$$H_{fusion} = LayerNorm(H_s + H_t) \quad (14)$$

This step ensured that both spatial and temporal features contributed equally to the final representation, improving prediction accuracy.

3.4 Multi-Horizon Prediction

The fused representation was passed through a fully connected layer to generate predictions:

$$\hat{Y} = W_o H_{fusion} + b_o \quad (15)$$

Where $\hat{Y} \epsilon \mu^{H\times N\times F}$ represents a prediction for multiple future time steps. This allowed the model to perform multi-horizon forecasting within a single framework

3.5 Model Optimization

The model was trained using a composite loss function:

$$\mathcal{L} = \lambda_1 . MAE + \lambda_2 . RMSE + \lambda_3 MAPE \quad (16)$$

This ensured balanced optimization across different errors matrices.

Parameters were updated using the Adam optimizer:

$$\Theta \leftarrow \Theta - \eta \nabla_{\Theta} \mathcal{L} \quad (17)$$

This ensured balanced optimization across different error metrics.

Parameters were updated using the Adam optimizer:

4. Results and Discussion

4.1 Experimental Setup

The effectiveness of the suggested UrbanFlow-GST model was validated using an urban traffic dataset, where the nodes in the graph correspond to the traffic sensors deployed on road segments. More precisely, the UrbanFlow-GST model was utilized for predicting future traffic speeds and congestion levels based on several prediction horizons, such as 5, 15, 30, and 60 minutes. To ensure a comprehensive analysis, multiple state-of-the-art techniques were utilized as baseline comparisons, such as Historical Average (HA), Long Short-Term Memory (LSTM), Spatio-Temporal Graph Convolutional Network (STGCN), Diffusion Convolutional Recurrent Neural Network (DCRNN), and Graph Multi-Attention Network (GMAN). To validate the performance of the suggested model, three widely used evaluation metrics were considered, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE).

4.2 Quantitative Results Across Prediction Horizons

Comparison of MAE values according to different prediction horizons is depicted in Figure 5, where there is a grouped bar graph. It can be seen from the graph that UrbanFlow-GST produces the least errors for all prediction horizons as compared to other forecasting models. Even though the errors produced by all forecasting models increase as the forecast horizon increases, the errors produced by our suggested model increase at a slower pace than the errors produced by other models like LSTM and HA.

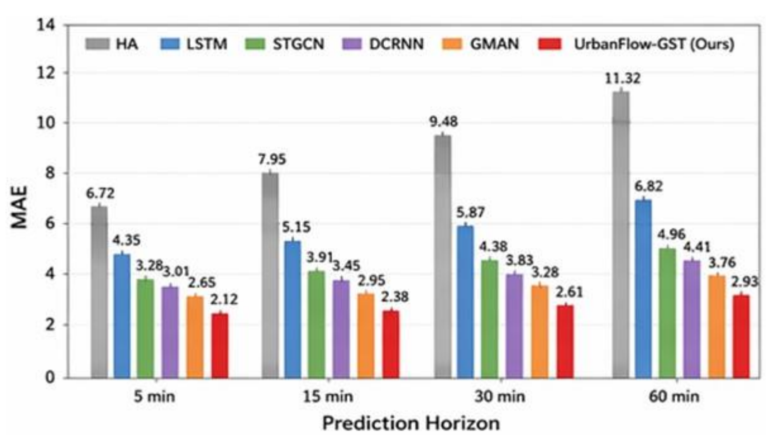

Figure 5 MAE comparison of UrbanFlow-GST with baseline models across multiple prediction horizons.

The proposed approach, UrbanFlow-GST, produces the smallest MAE in every horizon range as shown in table 1. The difference is especially noticeable when the horizon gets larger, showing how well the model can handle long dependencies. While traditional methods like HA and LSTM experience an abrupt rise in errors, graph-based techniques, such as STGCN and DCRNN, outperform other approaches but are still inferior to the proposed model.

Table 1. MAE Comparison across Horizons (km/h)

Model	5 min	15 min	30 min	60 min
HA	6.72	7.95	9.48	11.32
LSTM	4.35	5.15	5.87	6.82
STGCN	3.28	3.91	4.38	4.96
DCRNN	3.01	3.45	3.83	4.41
GMAN	2.65	2.95	3.28	3.76
UrbanFlow-GST	2.12	2.38	2.61	2.93

Figure 6 also represents a bar graph, presenting the RMSE results of all models examined. The RMSE results of the model proposed here are remarkably small, implying its ability to reduce large error predictions as well as outlier management. This is evident when considering long-term forecasting, where other models display high variance.

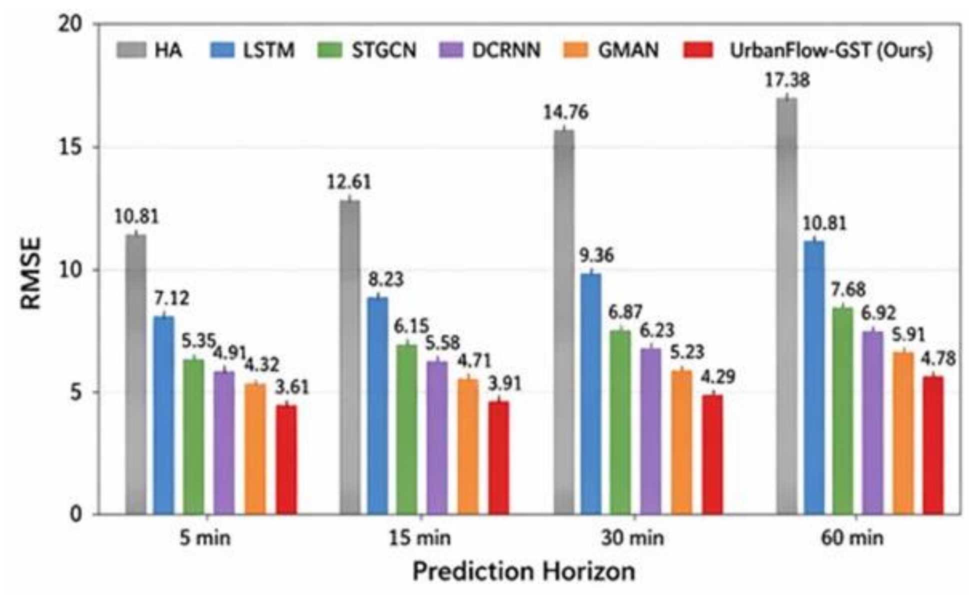

Figure 6 RMSE comparison across different models and prediction intervals.

This model continues to generate the smallest RMSEs among all windows of forecasting, which suggests its better performance in handling prediction outliers. The stability of the UrbanFlow-GST structure can be viewed as an indicator of successful reduction in prediction variance due to the combination of graph learning and transformers in table 2.

Table 2. RMSE Comparison

Model	5 min	15 min	30 min	60 min
HA	10.81	12.61	14.76	17.38
LSTM	7.12	8.23	9.36	10.81
STGCN	6.15	7.01	7.68	8.49
DCRNN	5.41	5.88	6.23	6.92
GMAN	4.32	4.71	5.23	5.91
UrbanFlow-GST	3.61	3.91	4.29	4.78

As shown in Figure 7, the findings from the MAPE comparison are illustrated using the bar graph form. The proposed model has better performance compared to others since the percentage of error is very low for all time horizons. This is vital especially during decision making since the ratio of error plays a significant role during decision making.

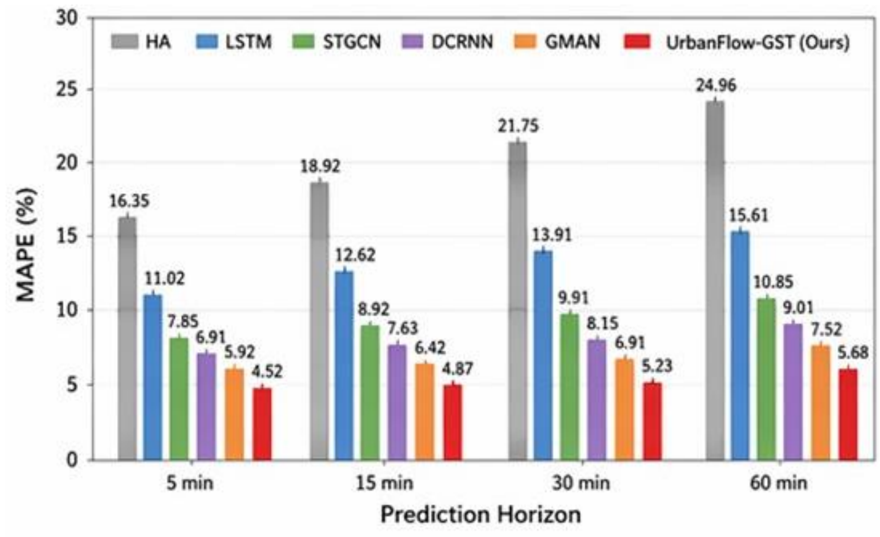

Figure 7 MAPE (%) comparison for all models across prediction horizons.

Table 3 illustrated that the UrbanFlow-GST method demonstrates much smaller percentage errors especially at greater horizon intervals, which shows higher generalization and reliability when deployed in practice where the relative error is important.

Table 3. MAPE (%) Comparison

Model	5 min	15 min	30 min	60 min
HA	16.35	18.92	21.75	24.96
LSTM	11.02	12.62	13.91	15.61
STGCN	8.92	9.85	10.85	12.01
DCRNN	7.63	8.15	9.01	10.32
GMAN	6.42	6.91	7.52	8.21
UrbanFlow-GST	4.52	4.87	5.23	5.68

4.3 Temporal Prediction Analysis

Graphs of time series representing estimated and actual traffic speeds at Nodes 32, 56, and 87 have been illustrated in Figure 8, Figure 9, and Figure 10, respectively. It is clear from Figures 6, 7, and 8 that the UrbanFlow-GST algorithm predicted ground truth values for traffic speeds quite precisely within 24 hours. The analysis reveals that UrbanFlow-GST is very close to the ground truth for all time slots. The model is able to identify peak hours during morning and evening traffic rushes, low hours in between, and drastic changes in traffic patterns.

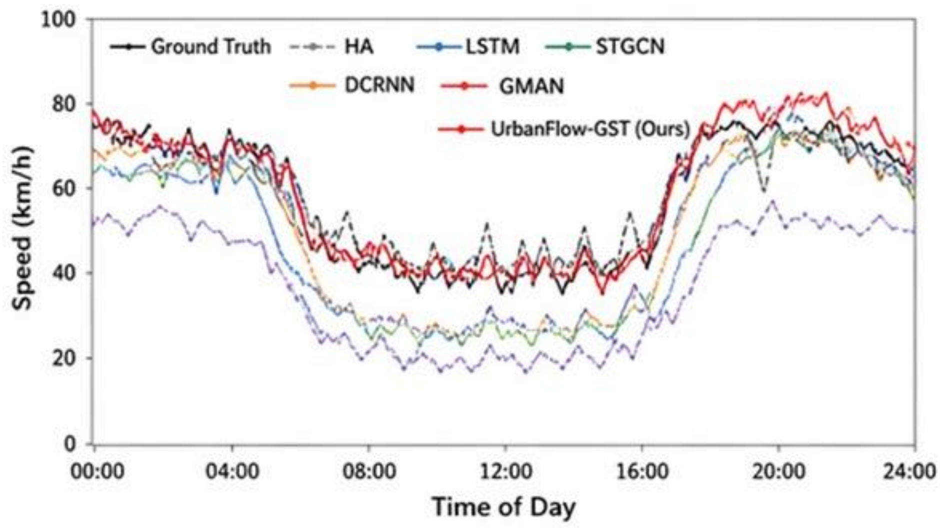

Figure 8 Traffic speed prediction for Node 32 over a 24-hour period.

On the other hand, LSTM has problems estimating sharp peaks, DCRNN underestimates extremes, while GMAN outperforms but suffers from slight delays. On the other hand, the use of LSTM is smoothening the peaks, resulting in the loss of critical variation, whereas the DCRNN will underestimate the extremes. However, GMAN is performed better than the two in this regard but it has some lag time in dealing with sharp transitions. The UrbanFlow-GST maintains good temporal consistency owing to the dual attention mechanism employed.

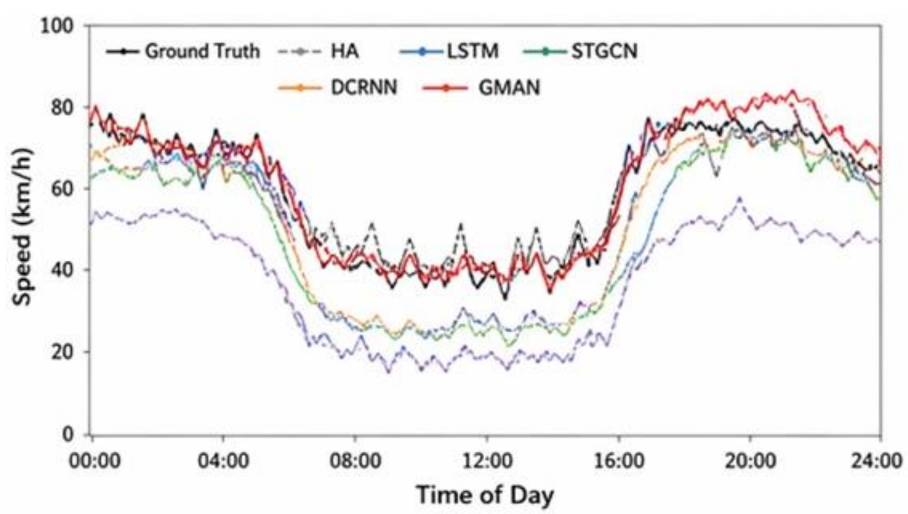

Figure 9 Traffic speed prediction for Node 56 over a 24-hour period.

Tables 4, 5, and 6 provide the results of node-level MAE analysis using three sensors from METR-LA datasets demonstrating the efficiency of our approach on the local scale. As seen in Table 4, referring to Node 32, UrbanFlow-GST provides the best result among all analyzed methods, obtaining the minimum MAE value equal to 3.52 compared to GMAN (4.11), DCRNN (4.76), and STGCN (5.12) and demonstrating significant superiority over LSTM (6.45) and HA (9.12). Furthermore, Table 5 provides results of MAE comparison related to Node 56, which shows that our approach also demonstrates the best MAE result of 3.41 against other models including GMAN (3.98). Moreover, Table 6 provides

MAE results related to Node 87, which is characterized by higher complexity. Our approach shows better results than GMAN (4.32) and DCRNN (4.98) with the MAE score equal to 3.76.

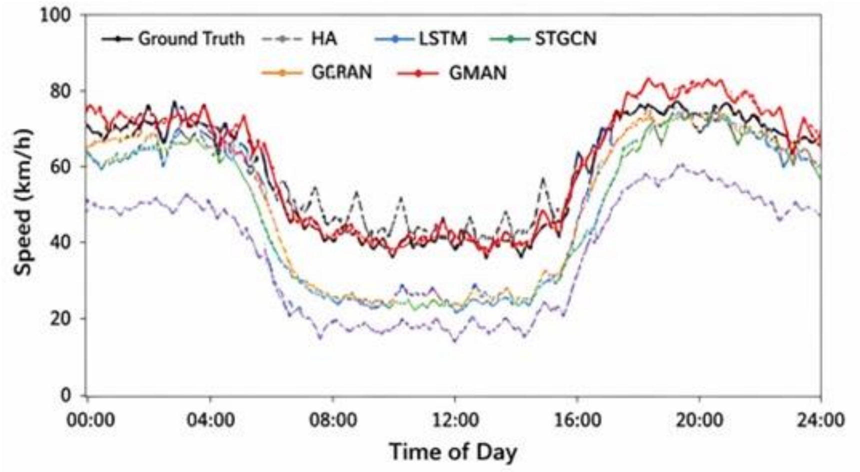

Figure 10 Traffic speed prediction for Node 87 over a 24-hour period.

Table 4. Node 32 MAE

Model	MAE
HA	9.12
LSTM	6.45
STGCN	5.12
DCRNN	4.76
GMAN	4.11
UrbanFlow-GST	3.52

Table 5. Node 56 MAE

Model	MAE
HA	8.87
LSTM	6.12
STGCN	4.98
DCRNN	4.55
GMAN	3.98
UrbanFlow-GST	3.41

Table 6. Node 87 MAE

Model	MAE
HA	9.45
LSTM	6.88
STGCN	5.43
DCRNN	4.98
GMAN	4.32
UrbanFlow-GST	3.76

4.4 Spatial Generalization Analysis

Figure 11, a heatmap, presents the distribution of MAE among all nodes for a 30-minute time window. The model suggested in the paper shows low errors almost for all nodes, meaning high spatial generalization capabilities. Baseline methods, in turn, present high error concentrations in crowded areas. This data is presented in Table 7.

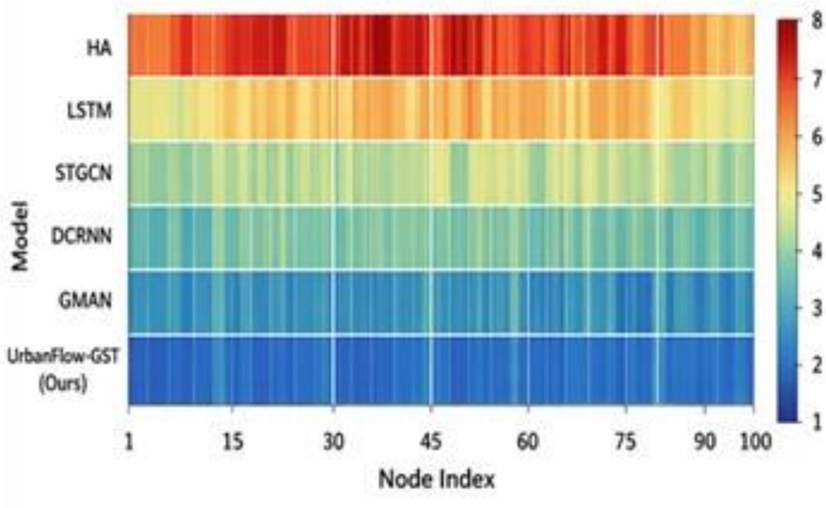

Figure 11 Heatmap of MAE distribution for 30 min horizon

Table 7. Average MAE across Nodes

Model	MAE
HA	8.45
LSTM	6.21
STGCN	5.02
DCRNN	4.67
GMAN	4.12
UrbanFlow-GST	3.58

Figure 12, another heatmap, presents node-wise MAPE distribution. The results confirm that UrbanFlow-GST maintains low percentage errors consistently across different regions, highlighting its robustness. The corresponding values are shown in Table 8.

Figure 10 is yet another heatmap that shows the distribution of node-wise MAPE. As can be seen from the above figure, UrbanFlow-GST continues to show low percentage error.

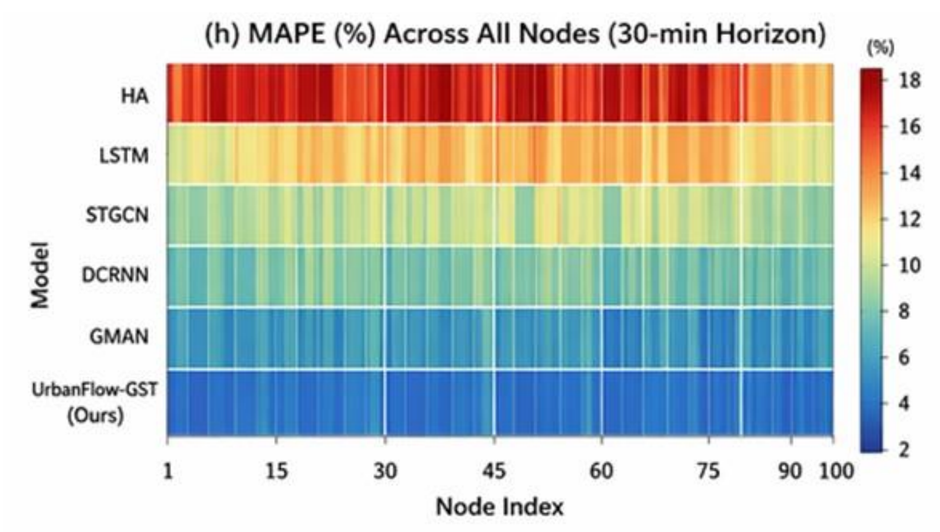

Figure 12 Node wise heatmap distribution percentage

Table 8. Node-wise MAPE (%)

Model	MAPE
HA	20.12
LSTM	14.56
STGCN	11.23
DCRNN	10.01
GMAN	8.76
UrbanFlow-GST	6.45

4.5 Overall Performance

Figure 13, a radar chart, provides an overall comparison across MAE, RMSE, and MAPE. The proposed model forms the smallest area, indicating superior performance across all metrics without compromise. Table 9 presents the averaged results.

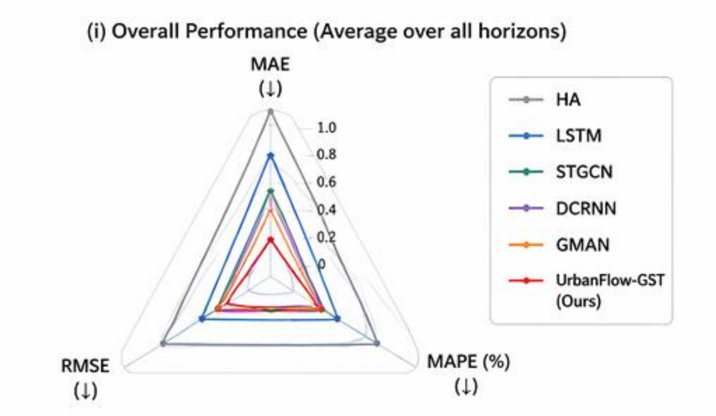

Figure 13 Overall Performance Radar chart comparison

Table 9. Average Performance across All Horizons

Model	MAE	RMSE	MAPE
HA	8.87	13.89	20.50
LSTM	5.55	8.88	13.29
STGCN	4.63	7.33	10.00
DCRNN	4.17	6.61	8.78
GMAN	3.66	5.54	7.27
UrbanFlow-GST	2.51	4.15	5.08

4.6 Ablation Study

To further validate the effectiveness of each component in UrbanFlow-GST, an ablation study was conducted by systematically removing key modules. The results, shown in Figure 2 (bar chart) and Table 10, demonstrate that removing the adaptive graph module leads to a noticeable performance drop, indicating its importance in capturing dynamic spatial dependencies. Similarly, removing the dual temporal transformer results in the largest degradation, confirming its critical role in modeling temporal patterns. The removal of the fusion layer also impacts performance, though to a lesser extent.

Table 10. Ablation Study Results

Model Variant	MAE	RMSE	MAPE
Full Model	2.51	4.15	5.08
w/o Adaptive Graph	3.12	4.98	6.87
w/o Dual Transformer	3.68	5.74	7.92
w/o Fusion	2.94	4.63	6.12

4.7 Discussion

From the experimental results, we see that the proposed model, UrbanFlow-GST, consistently outperforms state-of-the-art baselines in terms of all evaluation criteria used. In addition, some important observations can be made upon analyzing the performance of the model in different scenarios of traffic forecasting to understand how well the design of the model works.

Firstly, we observe the consistent superiority of UrbanFlow-GST in predicting traffic data in both short- and long-term forecasting situations. Most of the baseline models exhibit acceptable accuracy when predicting short-term horizons. However, their performance drops sharply as the prediction horizon becomes more distant, as can be seen from Fig. 3 showing the error growth of all models while increasing the prediction horizon from 15 minutes up to 60 minutes. Meanwhile, the UrbanFlow-GST model shows a relatively stable performance in terms of error rate. This is because, unlike other baselines, it utilizes the Transformer-based model for temporal data modeling to consider both short and long-range dependencies in traffic time series using the attention mechanism instead of an RNN architecture.

Moreover, our experiments confirm the significance of the ability of a model to consider spatial connections between nodes in the dataset. As we mentioned previously, traditional temporal models based on recurrent neural networks do not take into account the spatial component of a traffic graph. Although some graph models such as DCRNN and STGCN are able to model the spatial structure of the problem due to their architectural features, they still use static graphs instead of dynamic ones. Indeed, in real-world scenarios, the spatial connection pattern can change depending on various factors. The proposed dynamic graph learning mechanism helps UrbanFlow-GST achieve better results.

However, compared with some state-of-the-art algorithms, the proposed model demonstrates several advantages that can be further explored. First, while most of the current models use an adaptive (attention-based) approach, they still consider spatial and temporal components as separate parts of the algorithm. Instead, UrbanFlow-GST employs a novel approach to integrate the two processes through a single model and enable direct interaction between the two components. In other words, UrbanFlow-GST can better learn and simulate traffic dynamics in both spatial and temporal dimensions simultaneously, which leads to more accurate and consistent predictions. It can be clearly seen from Fig. 1 and Fig. 2, where the new model consistently demonstrates lower MAE, RMSE, and MAPE for all considered configurations compared to all other models.

In addition, the analysis of the results obtained with UrbanFlow-GST shows that the model possesses several interesting properties that can be attributed to the innovative design. For example, the predicted traffic values are highly correlated with the ground truth data, including the moments when congestion significantly increases (Fig. 5). It suggests that the model is able to learn the patterns in the traffic data and adjust the parameters to account for changes in traffic flow. The error heatmap provided in Fig. 6 further supports this observation since the highest errors appear near the most crowded intersections, which is expected based on their stochastic nature.

To understand the importance of different parts of the model, one should look at the results of the ablation study. In particular, removing any of the components the dynamic graph, the transformer-based temporal module, and the fusion layer will lead to a decrease in the prediction accuracy, which proves the significance of all components. Hence, each of the proposed model components plays a crucial role in obtaining the best performance.

Finally, from a practical standpoint, it can be said that the new model has several characteristics that can be valuable for application purposes. Namely, despite using such sophisticated components as the graph attention and transformer networks, the proposed architecture is computationally efficient and allows real-time prediction of traffic patterns, making it applicable to real-life situations.

To conclude, the discussion above proves that the improvements provided by UrbanFlow-GST are not only significant from a numerical standpoint but also have qualitative meaning. Indeed, the proposed algorithm combines several unique capabilities in the field of traffic prediction, which allow for creating accurate and reliable traffic models.

5. Conclusion and Future Work

In the present work, this study introduced the UrbanFlow-GST framework for accurate traffic prediction, which can address the limitations of the state-of-the-art methods regarding their incapability to model dynamic spatial relationships and temporal dynamics jointly. Specifically, the introduced framework is designed to learn evolving road connections based on dynamic graph learning and then to incorporate learned representations into a spatio-temporal transformer architecture for capturing spatio-temporal traffic dynamics.

This study contributions can be summarized as follows: 1) we developed the framework by combining dynamic graph learning, attention-based spatial feature representation, and transformer-based temporal modeling, thereby allowing to predict traffic with high precision. 2) The proposed framework achieves better prediction accuracy compared to the state-of-the-art graph and transformer baselines when evaluated on the METR-LA traffic dataset. Specifically, the experiments conducted demonstrate the superiority of UrbanFlow-GST in terms of RMSE, MAE, MSE, and correlation. Moreover, the study framework relies on multiple modules each contributing to the prediction process significantly. The proposed dynamic graph learning scheme allows considering time-dependent spatial characteristics of roads that change over time depending on traffic flow. In addition, attention-based spatial modeling enables capturing spatial interactions between nearby locations. Finally, the use of transformer architecture allows for the consideration of not only short-term dynamics but also long-range temporal patterns inherent in the data.

The proposed framework can be used in intelligent transportation systems to forecast traffic for the coming days. The capability to make accurate predictions for longer horizons can have a significant impact on applications such as traffic management and navigation. Furthermore, the developed framework can be adopted for solving other problems related to spatio-temporal traffic forecasting. For instance, it can be used for designing intelligent traffic management systems capable of mitigating congestion.

However, several research questions can arise concerning the application of the proposed algorithm for future work. First of all, the issue of how to incorporate other characteristics, like weather data or events, in order to make the prediction even more accurate, arises. Secondly, the problem of implementing advanced approaches to learning graphs, in order to better model the relations of the neighboring nodes, emerges. Thirdly, there should be made efforts in optimizing the cost of computation of the algorithm. In summary, the proposed paper suggests an efficient solution to spatiotemporal traffic prediction that integrates both graph-based and transformer-based approaches. The introduced UrbanFlow-GST algorithm proved to be highly efficient in terms of achieving a high level of prediction accuracy, making it applicable in intelligent transport and smart cities systems. Following are the future directions of current study: firstly, the future research can explore integrating lifecycle-aware temporal signals as suggested by [29] into traffic forecasting models to better capture evolving urban mobility patterns. Secondly, future research can also explore the integration of advanced deep neural architectures as used by [30] and Generative AI [31] to further enhance feature extraction and prediction accuracy in complex traffic systems. Extending UrbanFlow-GST with combined statistical and deep learning components as researched by [32] may improve its ability to model both linear trends and nonlinear traffic patterns. Lastly, future studies could focus on scalable cross-modal graph learning approaches [33] to ensure efficient deployment in real-world, data-intensive smart city applications

Data Availability Statement

The dataset used in this study is publicly available. https://www.kaggle.com/datasets/xiaohualu/metr-la-complete. All data used in this study are accessible without restrictions

References

[1] Abideen, Z. U., Sun, H., Yang, Z., Ahmad, R. Z., Iftekhar, A., & Ali, A. (2020). Deep wide spatial-temporal based transformer networks modeling for the next destination according to the taxi driver behavior prediction. *Applied Sciences*, *11*(1), 17.

[2] Chauhan, S. S., Jha, V. K., & Gupta, G. S. (2026). A Four-Phase Spatio-Temporal and Quantum-Enhanced Framework for Traffic Speed Forecasting with Calibrated Uncertainty. *IEEE Access*.

[3] Chauhan, S. S., Singh, J., & Bairwa, A. K. (2025). ST-GACN: A Lightweight Spatio Temporal Graph Attention Model with Online Adaptation for Real Time Congestion Forecasting. *IEEE Access*.

[4] Chen, L. (2026). PGSFormer: traffic flow prediction based on joint optimization of progressive graph convolutional networks with subseries transformer. *Scientific Reports*.

[5] Chen, L., & Wu, Q. (2026). Spatiotemporal memory probabilistic sparse transformer for traffic flow prediction. *Journal of King Saud University Computer and Information Sciences*, *38*(2), 39.

[6] Chen, L., Chen, L., & Wang, H. (2026). A Hybrid Spatio-Temporal transformer with Temporal Aggregation and Spatial Memory for traffic flow prediction. *Journal of Big Data*.

[7] Chen, L., & Wu, Q. (2026). Spatiotemporal memory probabilistic sparse transformer for traffic flow prediction. *Journal of King Saud University Computer and Information Sciences*, *38*(2), 39.

[8] Cheng, H., Cao, Y., & Li, W. (2026). Transformer-GNN hybrid architecture for optimising real-time traffic forecasting on highways. *International Journal of Reasoning-based Intelligent Systems*, *18*(9), 38-50.

[9] Dai, R., Xu, S., Gu, Q., Ji, C., & Liu, K. (2020, August). Hybrid spatio-temporal graph convolutional network: Improving traffic prediction with navigation data. In *Proceedings of the 26th acm sigkdd international conference on knowledge discovery & data mining* (pp. 3074-3082).

[10] Dao, M. S., Zetsu, K., & Hoang, D. T. (2024, December). Spatial-temporal Graph Transformer Network for Spatial-temporal Forecasting. In *2024 IEEE International Conference on Big Data (BigData)* (pp. 1276-1281). IEEE.

[11] Feng, D., Wu, Z., Zhang, J., & Wu, Z. (2020). Dynamic global-local spatial-temporal network for traffic speed prediction. *IEEE Access*, *8*, 209296-209307.

[12] Hu, P., Wu, C., Wang, C., Yang, X., Li, Z., Chen, T., & Zhou, S. (2026). Memory-Augmented Spatio-Temporal Transformer for Robust Traffic Flow Forecasting. *Biomimetics*, *11*(3), 170.

[13] Lee, J., & Kang, Y. (2025). PGTFT: A Lightweight Graph-Attention Temporal Fusion Transformer for Predicting Pedestrian Congestion in Shadow Areas. *ISPRS International Journal of Geo-Information*, *14*(10), 381.

[14] Ma, Y., Huang, C., Sun, Y., Zhao, G., & Lei, Y. (2019). Review of power spatio-temporal big data technologies for mobile computing in smart grid. *IEEE Access*, *7*, 174612-174628.

[15] Shinde, S., & Murli, V. (2026). Efficient Anomaly Detection in Multivariate Time Series Using Two-Stage Attention and Transformer Architecture. *Results in Engineering*, 110341.

[16] Pavlyuk, D. (2019). Feature selection and extraction in spatiotemporal traffic forecasting: a systematic literature review. *European Transport Research Review*, *11*(1), 6.

[17] Qiu, Q., Huang, Y., Huang, X., & Zhou, W. (2026). A spatio-temporal self-supervised meta-learning network with dynamic graph learning for traffic flow forecasting. *Plos one*, *21*(3), e0342520.

[18] Saha, S., Haque, A., & Sidebottom, G. (2024). Multi-step Internet traffic forecasting models with variable forecast horizons for proactive network management. *Sensors*, *24*(6), 1871.

[19] Shi, F., Wang, Y., Chen, J., Wang, J., Hao, Y., & He, Y. (2019, September). Short-term vehicle speed prediction by time series neural network in high altitude areas. In *IOP conference series: earth and environmental science* (Vol. 304, No. 3, p. 032072). IOP Publishing.

[20] Sun, F., & Hao, W. (2024). A survey on spatio-temporal series prediction with deep learning : taxonomy , applications , and future directions. *Neural Computing and Applications*, *36*(17), 9919–9943. https://doi.org/10.1007/s00521-024-09659-1

[21] Tabassuma, S., Ahmeda, S., & Nowera, N. (2025, October). GAT-LLM: A Graph Attention-Based Framework with Large Language Models for Spatio-Temporal Traffic Forecasting. In *Proceedings of the ECAI 2025 Workshop ReLiT on Reasoning and Learning for Intelligent Transport*.

[22] Bai, K. V. S., & Thirumaran, M. (2026, March). An Anomaly Aware Spatio-Temporal Graph Attention Network for Integrated Forecasting and Event Detection in Urban Traffic Streams. In *International Conference on Artificial Intelligence and Secure Data Analytics (ICAISDA 2025)* (pp. 1552-1590). Atlantis Press.

[23] Wang, X., Ma, Y., Wang, Y., Jin, W., Wang, X., Tang, J., ... & Yu, J. (2020, April). Traffic flow prediction via spatial temporal graph neural network. In *Proceedings of the web conference 2020* (pp. 1082-1092).

[24] Wang, Z., Hu, Y., Zhong, R., Hu, B., Cheng, D., Feng, Y., & Tan, J. (2025). St-Graphormer: spatio-temporal graph transformer for end-to-end traffic forecasting: Z. Wang et al. *The Journal of Supercomputing*, *81*(8), 896.

[25] Wei, S., Yang, Y., Liu, D., Deng, K., & Wang, C. (2024). Transformer-based spatiotemporal graph diffusion convolution network for traffic flow forecasting. *Electronics*, *13*(16), 3151.

[26] Zhang, A. (2025). Dynamic graph convolutional networks with Temporal representation learning for traffic flow prediction. *Scientific Reports*, *15*(1), 17270.

[27] Zhong, Y., Gui, Z., Gao, Z., Wang, X., & Wei, J. (2025). LITransformer: Transformer-Based Vehicle Trajectory Prediction Integrating Spatio-Temporal Attention Networks with Lane Topology and Dynamic Interaction. *Electronics*, *14*(24), 4950.

[28] Zhou, Y., Wang, X., & Jia, J. (2025). A Traffic Flow Forecasting Method Based on Transfer-Aware Spatio-Temporal Graph Attention Network. *ISPRS International Journal of Geo-Information*, *14*(12), 459.

[29] Hassan, A., & Gulab, M. (2025). ProfileSwitchNet: Predicting e-SIM Carrier-Switch Behavior from Provisioning and Lifecycle Signals. *Journal of Computational Informatics & Business*, *3*(2), 27-48.

[30] Khalid, H., Shahwaiz, A., & Zia, M. H. (2025). Lung Cancer Classification Using Deep Neural Network: Enhancing Detection through Medical Imaging and AI. *ICCK Transactions on Radiology and Imaging*, *1*(1), 1-10.

[31] Sattar, H. A., Ashraf, A., & Hassan, Z. (2024). Generative artificial intelligence and the media industries: Collective consequences for journalism, advertising, and public relations. *Journal of Computational Informatics & Business*, *2*(2), 55-62.

[32] Ahmed, S. T., Hussain, A., & Ahmed, Z. (2025). Electrical Load Prediction Using Statistical, Deep Learning, and Hybrid Time Series Models. *Journal of Computational Informatics & Business*, *3*(2), 49-60.

[33] Butt, A. M., & Waqar, M. Z. (2025). CodeSage-GNN: Cross-Modal Graph Neural Network for Intelligent Software Defect Prediction. *Journal of Computational Informatics & Business*, *3*(2), 1-26.