



DSTGN-ExpertNet: A Deep Spatio-Temporal Graph Neural Network for High-Precision Traffic Forecasting

Seyyed Ahmad Edalatpanah^{1*}, Javad Pourqasem²¹ Department of Applied Mathematics, Ayandegan Institute of Higher Education, 4681853617 Tonekabon, Iran² Department of Computer Engineering, Urmia University, 5756151818 Urmia, Iran

* Correspondence: Seyyed Ahmad Edalatpanah (s.a.edalatpanah@aihe.ac.ir)

Received: 01-05-2025

Revised: 02-10-2025

Accepted: 02-16-2025

Citation: S. A. Edalatpanah and J. Pourqasem, "DSTGN-ExpertNet: A Deep Spatio-Temporal Graph Neural Network for high-precision traffic forecasting," *Mechatron. Intell Transp. Syst.*, vol. 4, no. 1, pp. 28–40, 2025. <https://doi.org/10.56578/mits040103>.



© 2025 by the author(s). Licensee Acadlore Publishing Services Limited, Hong Kong. This article can be downloaded for free, and reused and quoted with a citation of the original published version, under the CC BY 4.0 license.

Abstract: Accurate traffic prediction is essential for optimizing urban mobility and mitigating congestion. Traditional deep learning models, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), struggle to capture complex spatiotemporal dependencies and dynamic traffic variations across urban networks. To address these challenges, this study introduces DSTGN-ExpertNet, a novel Deep Spatio-Temporal Graph Neural Network (DSTGNN) framework that integrates Graph Neural Networks (GNNs) for spatial modeling and advanced deep learning techniques for temporal dynamics. The framework employs a Mixture of Experts (MoE) approach, where specialized expert models are dynamically assigned to distinct traffic patterns through a gating network, optimizing both prediction accuracy and interpretability. The proposed model is evaluated on large-scale real-world traffic datasets from Beijing and New York, demonstrating superior performance over conventional methods, including Spatio-Temporal Graph Convolutional Networks (ST-GCN) and attention-based models. With a mean absolute error (MAE) of 1.97 on the BikeNYC dataset and 9.70 on the TaxiBJ dataset, DSTGN-ExpertNet achieves state-of-the-art accuracy. These findings highlight the potential of GNN-based frameworks in revolutionizing traffic forecasting and intelligent transportation systems (ITS).

Keywords: Traffic prediction; ST-GNN; GNNs; ITS; Deep learning; Traffic flow forecasting; Urban mobility; Real-time traffic analytics

1 Introduction

Intelligent Transport Systems (ITS) [1], aim to provide precise and sophisticated traffic forecasts, guaranteeing easy and effective transportation and controls. Since ITS has been the subject of research for many years and is still developing, the discipline has experienced significant expansion and advancement. As a result, there is a vast body of literature on traffic prediction, reflecting the ongoing efforts to improve and refine predictive models for better transportation outcomes.

Short-term traffic prediction [2] enhances road management by reducing delays, incidents, and unexpected events through real-time data. This paper reviews literature on model/data driven approaches in short-term traffic predictions. It begins with an analysis of real-time traffic data collection methods used as input for predictive algorithms. The study then describes key traffic prediction [3] outputs based on available input variables and discusses standard metrics for assessing prediction accuracy. Lastly, it offers information on the effectiveness of the existing data-driven and model-driven techniques that provide real-time short-term traffic projections.

Road transportation [4], as the most extensive and complex nonlinear component of traffic management, requires accurate traffic predictions for the effective functioning of ITS. However, selecting the right prediction technique remains challenging for transportation departments, ensuring users can effectively utilize the forecasted information. This paper [5] surveys the latest forecasting technologies, offering insights into the core concepts behind various prediction approaches to aid in understanding and improving ITS applications.

Traffic prediction [6] is essential for addressing global traffic congestion, which leads to issues like longer journey duration and more fuel usage. Both new research difficulties and notable gains in traffic prediction are brought about by the incorporation of modern technology into transportation systems. This paper offers a thorough analysis of

traffic prediction techniques [7], emphasizing new developments and promising prospects in AI-based approaches. Given their recent success and promise, special emphasis is paid to multivariate traffic time series modelling for handling open research challenges in traffic prediction.

The increase in traffic-related information and its applications has led to a greater focus on traffic prediction. Spatial-temporal prediction [8] has a variety of uses, including as urban planning and climate forecasting. For example, businesses might better arrange taxis to fulfill commute demands by using precise taxi demand forecast [9]. Simulating intricate geographical and temporal connections is the primary issue in traffic prediction. Based on two important findings, this research presents a unique Spatial-Temporal Dynamic Network (STDN) paradigm [10]: (1) Temporal dependencies show substantial periodicity but are not exactly periodic because of dynamic temporal changes, and (2) spatial relationships between sites are dynamic. Figure 1 illustrates the DSTGN-ExpertNet enhances real-time traffic prediction by using Graph Neural Networks for spatial modelling and specialized expert models for varying traffic patterns, showing superior accuracy and interpretability in large-scale urban datasets.

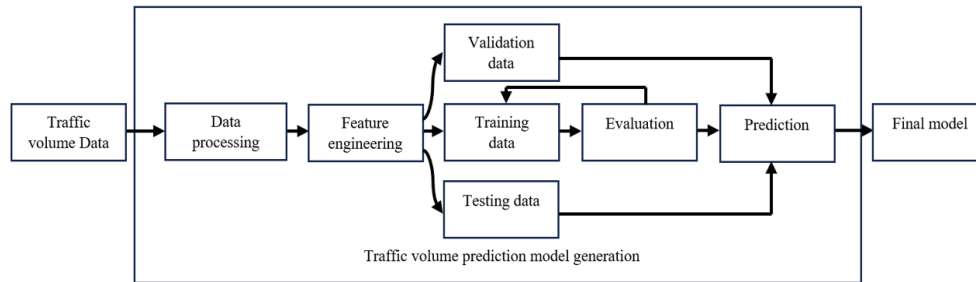


Figure 1. Block diagram for traffic prediction

2 Literature Review

Ali and Mahmood [11] proposed LSTM Neural Networks because they work best with sequential data, making usable for short-term traffic flow predictions. The most popular dataset is PeMS. For increased accuracy, contemporary ITS increasingly include weather, traffic feelings from social media, temporal, and geographical data. This is ITS's first review on deep learning.

Korkmaz and Erturk [12] introduced an in-depth review which listed important studies, publications, authors, and developments that have impacted predicting lengths of traffic incidents using statistical and machine learning (ML) techniques. It addresses unobserved heterogeneity and unpredictability while analyzing novel approaches, data kinds, and significant variables. The study also employs VOSviewer® to visualize knowledge mapping from 2010 to 2022. The paper's contributions include comparing previous studies, identifying key conceptual features in analysis and prediction, and exploring future trends. The study notably found lesser usages of crowdsourcing, social media, and textual data.

Shi et al. [13] focuses on the prediction of network traffic and critically analyzes the decomposition and optimization methods applied in predictive models. It describes model parameters, datasets, and assessment standards, emphasizing how well Variational Mode Decomposition (VMD) and Particle Swarm Optimization (PSO) handle prediction problems. According to the investigation, the best optimization algorithm and decomposition method for improving the precision and rate of convergence in network traffic forecast models are PSO and VMD, respectively.

Kolekar et al. [14] developed with an emphasis on AI tactics, new trends, databases, and research difficulties in order to forecast the behavior of traffic actors in intelligent cars. With an analysis of peer-reviewed publications from 2011 to 2021 and a focus on five major research issues, it is the first comprehensive literature assessment in this field. The findings highlight that AI-based solutions, utilizing advanced input representations like traffic rules and road geometry, have shown significant success in predicting vehicle behavior, particularly in complex driving scenarios.

Wang et al. [7] presented LibCity, unified and extensible libraries for traffic predictions with easy-to-use programming framework and a dependable experimental instrument. LibCity facilitates extensive experimentation by providing 29 spatial-temporal datasets and 42 replicated traffic prediction models. By isolating implementation details, the library's standardized model interfaces, which are based on uniform data formats, make it easier to design new models. The study creates a performance leaderboard for four traffic prediction tasks and presents repeatability comparisons to verify its efficacy.

Irawan et al. [5] predicted traffic flows using Intelligent Transport Systems (ITS) which divided predictions into short and long term. While Long-Term Prediction depends on time series data and has trouble with exceptional events like accidents, Short-Term Prediction employs real-time data and may be redundant in ordinary traffic. In

order to develop a more successful traffic prediction model, the author suggests a novel approach that blends time series analysis and dynamic real-time prediction.

The use of the Adaptive Neuro-Fuzzy Inference System (ANFIS) in intelligent transport systems was emphasized by Stojčić [15], who also recommended its application in traffic and transportation. The papers have been divided into seven categories: vehicle routing, traffic signal management at intersections, vehicle steering and control, safety, fuel consumption, engine performance, and emissions modeling, traffic congestion prediction, and other applications. Each sub-area is assessed using a tabular summary of the input and output variables, with the results explained in the third part.

Xiao et al. [16] suggested a crucial element for maritime traffic networks, emphasizing the most recent developments in vessel motion forecasting to improve situation awareness and pattern extraction from marine traffic data mining. It highlights how knowledge of traffic patterns is crucial for the development of knowledge-based forecasting methods and offers insightful information for a range of applications. The paper emphasizes how important advanced marine traffic studies are to enhancing intelligence and safety, especially when combined with big data, IoT, AI, and knowledge engineering.

Shaik et al. [17] suggested usages of numerous neural network models, such as the radial basis functions (RBFs), RNNs, CNNs, single-layer perceptrons (SLPs), and multilayer perceptrons (MLPs) for predicting severity of injuries received in traffic accidents. It discusses possible future directions and challenges, as well as providing an overview of the models' inputs (independent variables), outputs (dependent variables), and performance evaluation methodologies.

Ashwini and Sumathi [18] suggested a number of data sources for traffic prediction, including cutting-edge ones like cellular network and social media data. In addition to doing a comparison analysis based on quality factors such data correctness, dependability, preprocessing complexity, data sufficiency, infrastructure cost, and maintenance overhead, it describes the contributions of each data source to traffic forecast.

3 Problem Formulation

Traffic prediction [19] relies on historical traffic data from multiple road segments or monitoring devices to forecast future traffic conditions, including flow and speed. Initially, the challenges associated with traffic prediction are defined as follows:

Definition 1 (Traffic Prediction): Formally, the anticipated future signals may be described as $Y = yt + 1$ given a set of fully observed time series signals $X = x_1, x_2, \dots, x_t$ of all the road segments or sensors. Traffic networks may be represented as directed graphs using structured time-series data as roadways include directional information.

The essential terms used in directed graph-based traffic prediction will be outlined in the following sections.

Definition 2 (Directed Graph for Traffic Network): $G = (V, E, W)$ is a weighted directed graph that can be used to model a traffic network. N is the number of nodes, and $W \in RN \times N$ is the weighted adjacency matrices of G that shows the nodes' proximities, such as separations between nodes. In traffic networks, V stands for node collections, like sensors, that are able to keep an eye on certain road segments. If v_i and v_j are linked, then $\varepsilon(i, j) = 1$; if not, then $\varepsilon(i, j) = 0$. This is because E is the connection between the nodes. Specifically, the edge is $w(i, j)$, weights between v_i and v_j . Figure 2 illustrates the framework of the proposed ST-ExpertNet for traffic prediction.

Definition 3 (Traffic Prediction on Graphs): Let $x_t \in RN \times P$ be a graph signal noticed at times t , given a $G.P.$ indicates attribute counts noticed by nodes, such as flow, vehicle speed, etc. The following formula can be used to forecast $yt + 1$:

$$y_{t+1} = f([x_{t-\bar{h}+1}, \dots, x_t], G) \quad (1)$$

Definition 4 (Diffusion Convolution): The following is the definition of diffusion convolutions over graph signals $x \in RN \times P$ and filters Θ :

$$x \star \theta = \sum_{k=0}^{K-1} (\theta_{k,1} (D_O^{-1}W)^k + \theta_{k,2} (D_I^{-1}W^T)^k) x \quad (2)$$

where x represents diffusion convolutions, K represents diffusion steps while learnt parameters of Θ for the graph's two directions are $\theta \in RK \times 2$; $DI = \text{diag}(WT)$ is the in-degree diagnosis; $1 \in RN$ indicates the all-one vector; and $DO = \text{diag}(W1)$ is the out-degree diagonal matrix. $D_o^{-1}W, D_I^{-1}W^T$ represents transition matrices of diffusions and reversals [20], respectively.

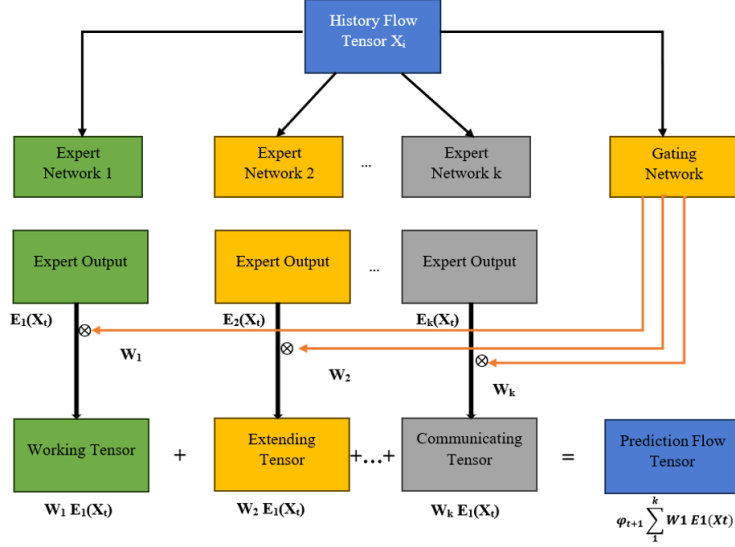


Figure 2. The framework of our ST-ExpertNet for traffic prediction

4 Proposed Methodology

Traffic monitoring today employs a vast amount of hardware to collect traffic data, and the amount of traffic data grows as the problem of traffic prediction gains traction. We are searching for efficient computational models for analyzing traffic data. There are now various models for analysing and forecasting traffic flow. A model called GraphSAGE learns nodes effectively by aggregating nearby nodes. It excels in drug detection, system recommendations, and classification jobs due to its excellent computing efficiency. Good performance. Nonetheless, it concentrates on data items with strong order features and does not include time series. Traffic flow data exhibits significant time series patterns, but the possible link between time and location is more complex. As a result, GraphSAGE is unable to effectively address these issues. This research introduces a novel ST-GNN for learning and forecasting traffic flow. This technique mimics the concept of the GraphSAGE model by generating a directed graph that depicts the spatiotemporal interaction between routes. The Skip-gram [3] approach applies unsupervised learning to a particular road segment to predict future traffic conditions. This model’s distinctive characteristic is that Graph SAGE is an inductive graph learning model that, after sampling nodes and aggregating neighboring data, handles the final goal (like node classification). Instead of using the final node embedding itself, Graph SAGE uses the node attributes (text, node degree, node attribute description, etc.) to determine the node embedding pattern (i.e., function). This enables it to cope with unseen nodes and even observe unique network topologies by using the proper learning node embedding mode, which sets it apart from the generic matrix decomposition-based graph embedding approach.

4.1 The Define of STGNN

A Spatiotemporal Graph Neural Network (STGNN) framework has been specifically developed to capture the spatiotemporal relationships between roads and traffic flow, particularly under complex road conditions. Given the intricate nature of road networks and their strong correlation with real-world traffic challenges, the road variations depicted in Figure 3 provide a practical and meaningful foundation for this study.

By transforming the road structure into an undirected graph, the problem of traffic flow forecasting is reframed as learning the correlations between nodes within the network. Our primary objective is to uncover implicit patterns in traffic flow variations across different road segments over time. By analyzing historical traffic data and the underlying relationships between road sections, we can effectively anticipate future traffic flow patterns.

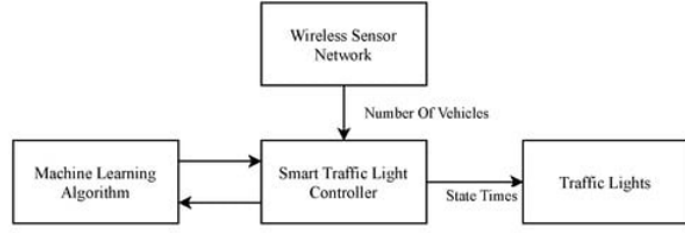
Figure 3 illustrates the road network structure, where a highway is divided into seven segments or routes: A, B, C, D, E, F, and G. The network contains two junctions, one of which is paved. Notably, routes C, F, and G originate from the same straight road, whereas routes B and D form another straight connection. Additionally, routes A and E are linked by a common road.

The interactions between routes are complex and interdependent. For instance, routes C, B, D, and G can directly influence one another. However, the degree of influence varies—for example, the impact between routes B and F is significantly weaker than that between routes F and G. While routes B and F exhibit a stronger correlation, routes F and G are relatively more distantly connected in terms of influence.

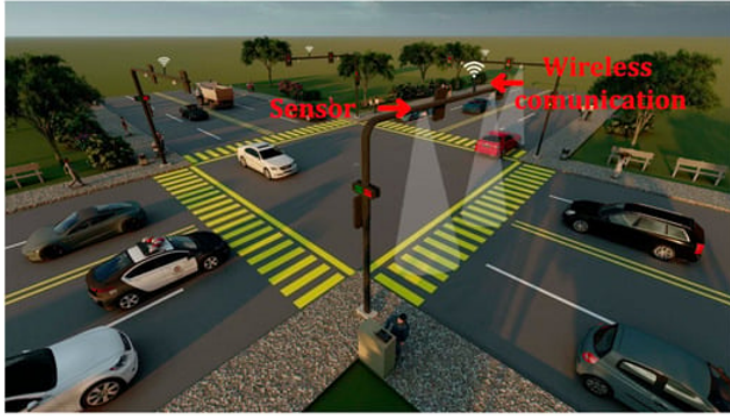
This intricate interplay between road segments highlights the need for advanced spatiotemporal modeling,

enabling more accurate and reliable traffic flow predictions.

We depict road links as graph relationships, as seen in Figure 3, where all alignments are corresponding nodes in the network; all routes are corresponding nodes in the graph; and the lines linking the routes are nodes in directed graphs. The following guidelines describe road structures:



(a)



(b)

Figure 3. Proposed usage. (a) Block diagram; (b) Representation in real-world scenario

$$R = G(N, E_s, E_e) \quad (3)$$

where N represents all routes, E_s represents starting points of routes, E_e represents destinations, and R represents road structures [21]. Moreover, N contains two types of nodes: embedded representations of the same routes at different times and embedding nodes of routes.

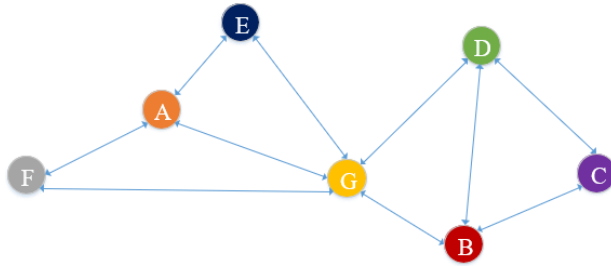


Figure 4. Graph structure representation between different road sections

Figure 4 shows the spatiotemporal relationship between road segments.

Node E1 represents embedded representations of routes E at given times E and E1, with connections established between them to enhance functionality. The same is true for G and G1, A and A1, and so on. The following equation can be used to illustrate the link between nodes:

$$G_{st} := (R, R_t) = (G(N, E_s, E_e), G_t(N_t, E_{st}, E_{et})) \quad (4)$$

where nodes are represented by N_t , such as E1, A1, G1, D1, and so on, and R_t contains embedded representations of routes at various time periods. G_{st} stands for spatio-temporal direct graphs. R may be determined using Eq. (4). Nodes can be connected in both directions in the same amount of time. Depending on the period, nodes are either unidirectional or unjoinable. G_t 's start node is represented by E_{st} and its end node by E_{et} . Although E_1 and A_1 cannot be physically connected, it can be observed that E_t , A_t , and E are related in both directions. The structures of N and N_t are the same.

As shown in Figure 5, we can define $N = \{A, B, C, D, E, F, G\}$, $N_1 = \{A1, B1, C1, D1, E1, F1, G1\}$. We can observe that $A \leftrightarrow A1$, it means A is the start node and $A1$ is the end node when $A \rightarrow A1$. Similarly, $A \leftrightarrow F$, $A \leftrightarrow E$ and $A \leftrightarrow G$. In addition, we can find that $A1 \leftrightarrow A$, $A1 \leftrightarrow E1$, $A1 \leftrightarrow G1$ and $A1 \leftrightarrow F1$. But as we can see, $A1$ is not connected to E ; instead, it is expressed as $A1 = E$. Those associated with $A1$ are denoted by $A1 = G$ and $A1 = F$. The road segment E and the E node are then shown at different times in Figure 5, along with the relationships between the nodes at different times. The embedding of road segment $E1$ at time $t1$ is shown in the following notation:

$$F = \{N1, N2, N3, \dots, Ni\}$$

$$Ni = (VN, Ti), (0 < i, Ti = Ti - 1) \quad (5)$$

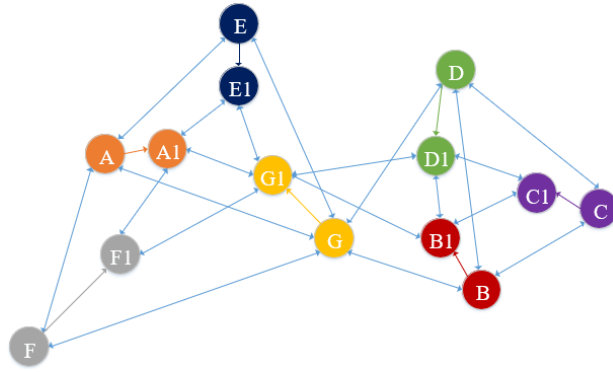


Figure 5. Graphic representation of spatio-temporal relationship of road sections

The node embedding vectors for all time periods are represented by F , the embedding vectors are Ni , the traffic flow properties of routes N are implied by VN , and the values of VN vary between Ti and $Ti-1$. Although Ti and $Ti-1$ have the same value, Ni and $Ni-1$ are not the same. According to Eq. (3), the letters $E1, E2, E3, E4, Ei$, and Ep stand for the route E embedding throughout different time periods. The node that represents the future time period is specifically Ep .

$E1 \rightarrow E2, E2 \rightarrow E3, E3 \rightarrow E4, E4 \rightarrow Ei, Ei \rightarrow Ep$, but $Ep \nrightarrow Ei, Ei \nrightarrow E4, E4 \nrightarrow E3, E3 \nrightarrow E2, E2 \nrightarrow E1$, as seen in subgraph (a) of Figure 6. The serialization of nodes at different times is shown in subgraph (b) of Figure 6; the node that is closest in time, Ep , is the one where we choose to predict the data. We may be able to get the desired node Ep from nodes E and Ei .

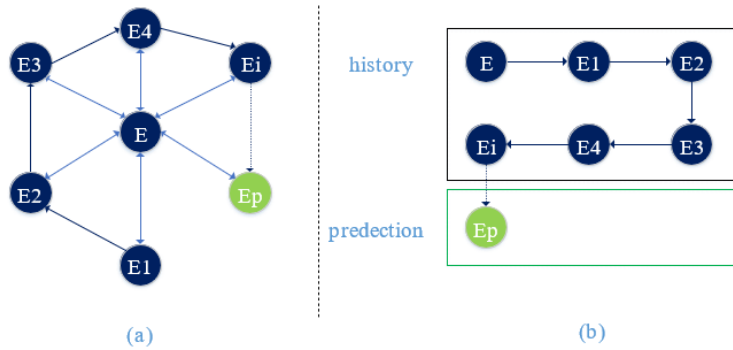


Figure 6. Temporal graph structure for a single node. (a) Graphic representation of spatio-temporal relationship of a road section. (b) Representation of traffic flow at different time periods on a road section

4.2 The Prediction Process of STGNN

The ST-GNN is primarily made up of two kinds of node embeddings, as shown in Figure 7: The traffic flow node embedding of the road segments at various periods and the node embedding of the directed graph initialization of the link between the beginning road segments.

The segment node must be embedded before a multi-segment segment may be created as a directed graph. Determine the new node A's connectable neighbors first. The value of k determines the maximum distance between the neighbors.

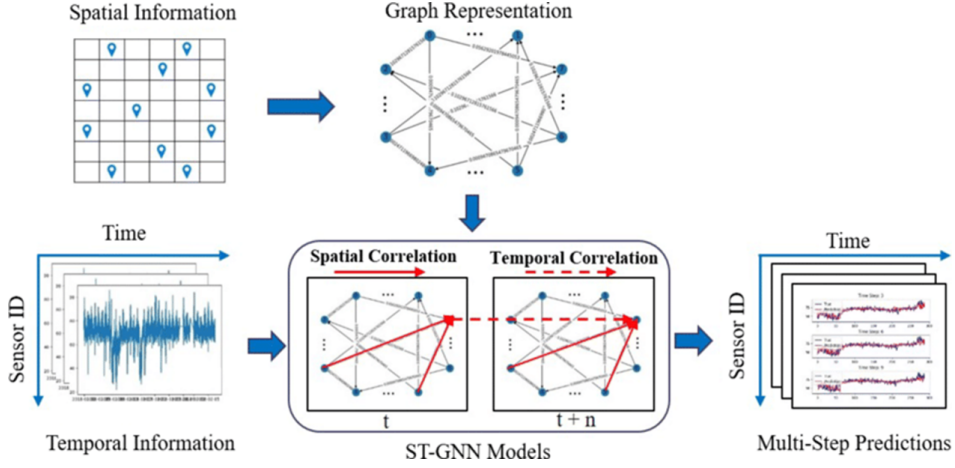


Figure 7. A general pipeline of ST-GNN models for traffic prediction

The nearest effective neighbor of A is chosen when $k = 1$. These neighbor nodes also have their own neighbor nodes. When $k = 2$, they select the neighbor nodes of node A's effective neighbors. As needed, they are free to change the value of k. Finally, in order to obtain information on node A, they combine all of the nodes they have selected.

Different nodes can be used to depict the road section's traffic flow at various periods. Assume that at time t_i , node A_{t_i} reflects node A's state. A_{t_i} has a direct relationship with node A and is solely connected to the node of t_{i-1} . Assign k to 1. In order to embed the node A_{t_i} , first ascertain whether there are any additional nodes in the time period at t_{i-1} . If the neighbor of the node is connected to the node, choose it. Similarly, after setting $k = 2$ if required, find the neighbor nodes of the selected neighbor nodes, and so on. Ultimately, the aggregate of these neighbor nodes yields A_{t_i} 's representation vector.

5 Result and Discussion

DSTGN-ExpertNet [22] was evaluated and compared using TaxiBJ and BikeNYC-I real-world public crowd flow datasets. Both quantitative and qualitative experiments were conducted to show the effectiveness and superiority of our proposed method.

5.1 Data Sets

TaxiBJ: Four time periods are used to gather trajectory GPS data [23] in Beijing taxicabs: 01/07/2013 - 30/10/2013, 01/03/2014 - 30/06/2014, 01/03/2015 - 30/06/2015, and 01/11/2015 - 10/04/2016. Beijing's map is divided into 32×32 identical areas, and metadata factors such as temperature, wind speed, weather, and holiday information are also saved in this collection.

BikeNYC-I: These datasets were released in NYC-Bike between July 1, 2016, and August 29, 2016, and April 1, 2014, and September 30, 2014, respectively. The datasets published from various sources are distinguished using I and II.

5.2 Evaluation Metric

The outcomes of suggested algorithms and current techniques have been evaluated using metrics such as Mean Squared Error (MSE), Root Mean Square Error (RMSE), and MAE. The following is the definition of these metrics:

$$MSE = \frac{1}{N} \sum_{i=1}^N \left(\hat{Y}_{t+1} - Y_{t+1}^i \right)^2 \quad (6)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{Y}_{t+1}^i - Y_{t+1}^i)^2} \quad (7)$$

$$MAE = \frac{1}{N} \sum_i^n |\hat{Y}_{t+1}^i - Y_{t+1}^i| \quad (8)$$

$$MAPE = \frac{1}{N} \sum_i^n \left| \frac{\hat{Y}_{t+1}^i - Y_{t+1}^i}{\hat{Y}_{t+1}^i} \right| \quad (9)$$

where \hat{Y}_{t+1}^i and Y_{t+1}^i represent prediction and real values of regions i for time intervals $t + 1$ and where N implies sample counts.

5.3 Methods for Comparison

The settings for each method are specified, with the highest performance achieved on the validation set reported. Additionally, the CNN, ConvLSTM, and ESOM-ResNet [24] architects were incorporated into our model to evaluate the broad applicability of ESOM-Expert Net [25]. The following list includes the comprehensive implementations for each ESOM-Expert Net and the experiment setup based on the benchmark work [26] in traffic flow prediction:

Historical average (HA): By averaging the value of past flow at the same location within the same relative time period, the historical average forecasts future flow.

ConvLSTM: The fully connected LSTM (FC-LSTM) is extended by convolutional LSTM [27], which preserves its benefits. Furthermore, it is possible to capture both temporal and spatial dependence thanks to the convolutional processes preserved in input-to-state and state-to-state transitions. The ConvLSTM layers maintain a 3×3 kernel size and 32 filters. The ReLU operation is selected as the final output after four ConvLSTM layers are employed.

ST-ResNet: ST-ResNet extends the ResNet [28] into a traffic prediction framework. Furthermore, the three ResNet blocks handle the hour, day, and week patterns, respectively. The closeness, period, and trend sequence lengths are set to $lc = 3$, $lp = 1$, and $lt = 1$. In ST-ResNet, every block has three residual units with $32 \ 3 \ 3$ kernel filters.

CNN-ExpertNet [29]: CNNs are included in ST Expert Net with 10 expert networks having the same architectures as pure CNNs.

ExpertNet-ST-ResNet [30]: ST-ResNet designs are utilized, and only three expert networks are used to compare performance with the original ST-ResNet. The ST-ResNet-ExpertNet does not demonstrate that its three ST-ResNet blocks are comparable to those of ST-ResNet; rather, it shows that the three ResNets are.

Table 1 presents the parameter settings and compares the performance of DSTGN-ExpertNet with various baseline methods. For the TaxiBJ and BikeNYC datasets, as shown in Figure 8, DSTGN-ExpertNet achieves an MSE of 275.130 for TaxiBJ and 20.043 for BikeNYC, demonstrating a significant improvement over the best-performing baseline models.

The performance of the proposed approach, in comparison to competing methods, is shown in Figure 9. When compared to the TaxiBJ (14.123%) and BikeNYC (3.890%) datasets, the RMSE results for traffic flow prediction are evaluated, with the proposed method demonstrating the best performance among the baseline approaches.

As shown in Figure 10, DSTGN-ExpertNet achieves a MAE of 9.700 for TaxiBJ and 1.970 for BikeNYC, demonstrating a significant improvement over the baseline methods in terms of MAE performance.

Table 1. Effectiveness evaluation on TaxiBJ and BikeNYC-I

Models	TaxiBJ			BikeNYC		
	MSE	RMSE	MAE	MSE	RMSE	MAE
HistoricalAverage	3025.123	56.005	26.465	230.723	16.767	4.778
CopyYesterday	2015.256	42.704	24.004	210.541	15.121	4.601
ConvLSTM	470.458	20.247	12.816	43.341	8.112	2.512
CNN-ExpertNet	370.636	19.510	10.323	38.001	6.362	2.144
ST-ResNet-ExpertNet	340.258	17.153	10.183	36.658	6.063	2.101
ESCOM-ExpertNet	310.158	15.153	10.045	23.001	4.004	2.001
DSTGN-ExpertNet	275.130	14.123	9.700	20.043	3.890	1.970

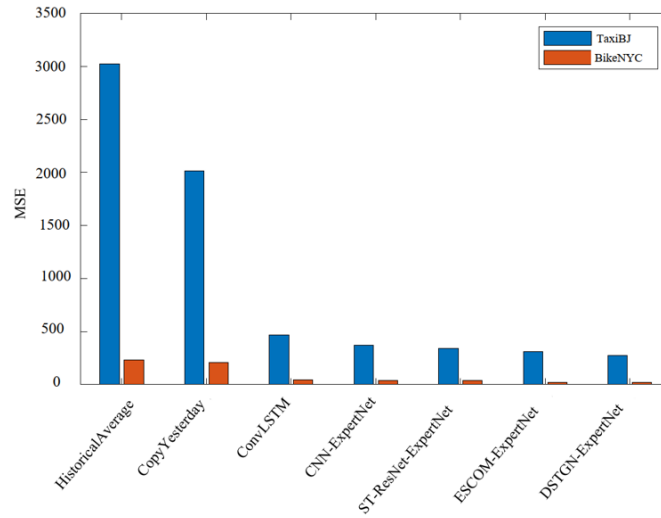


Figure 8. Parameter analysis of TaxiBJ & BikeNYC-MSE

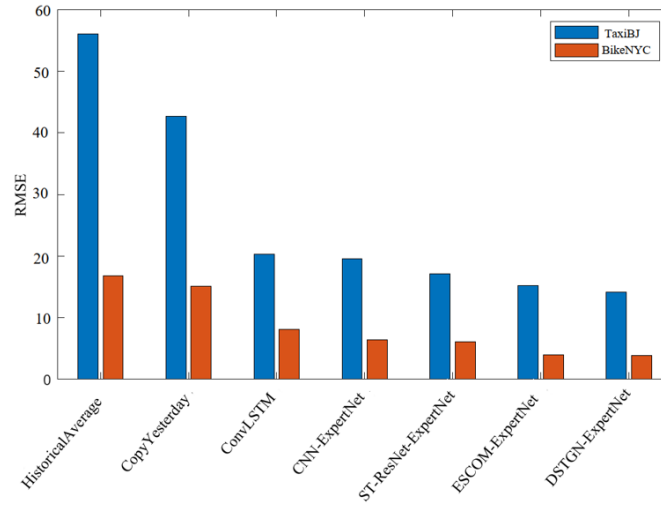


Figure 9. Parameter analysis of TaxiBJ & BikeNYC – RMSE

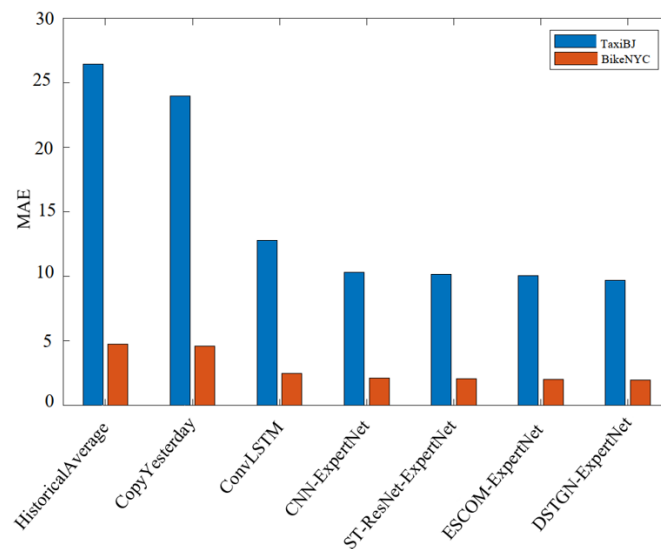


Figure 10. Parameter analysis of TaxiBJ & BikeNYC – MAE

5.4 Parameter Analysis

5.4.1 Ablation study

Ablation tests were conducted on the TaxiBJ dataset using distinct masks for G_s and G_t . Masking G_s is achieved by removing G_s and incorporating self-attention directly from the expert’s output in the model. Conversely, G_t can be masked simply by discarding it.

Table 2 presents the results of the ablation study for the TaxiBJ and BikeNYC datasets. As shown in Figure 11, the performance is compared with several models and evaluated using metrics such as MAE and Mean Absolute Percentage Error (MAPE). The results indicate that BikeNYC yields superior prediction accuracy from the expert models.

Table 2. Ablation study results for TaxiBJ and BikeNYC datasets

Models	TaxiBJ		BikeNYC	
	MSE	MAPE	MSE	MAPE
CNN-ExpertNet	340.614	5.18%	112.717	4.49%
CNN-ExpertNet:Gs	450.734	6.48%	130.002	5.05%
CNN-ExpertNet:Gt	430.237	6.13%	143.523	5.15%
ST-ResNet-ExpertNet	340.258	5.05%	110.435	4.05%
ST-ResNet-ExpertNet:Gs	347.201	5.24%	125.173	4.15%
ST-ResNet-ExpertNet:Gt	342.434	5.18%	154.110	5.23%
ConvLSTM-ExpertNet	330.250	4.95%	108.445	3.95%
ConvLSTM-ExpertNet:Gs	343.748	5.67%	110.318	4.24%
ConvLSTM-ExpertNet:Gt	330.004	5.07%	116.742	4.52%
GNN-ExpertNet:Gs	353.004	5.03%	125.670	4.35%
GNN-ExpertNet:Gt	347.790	5.25%	156.800	5.19%

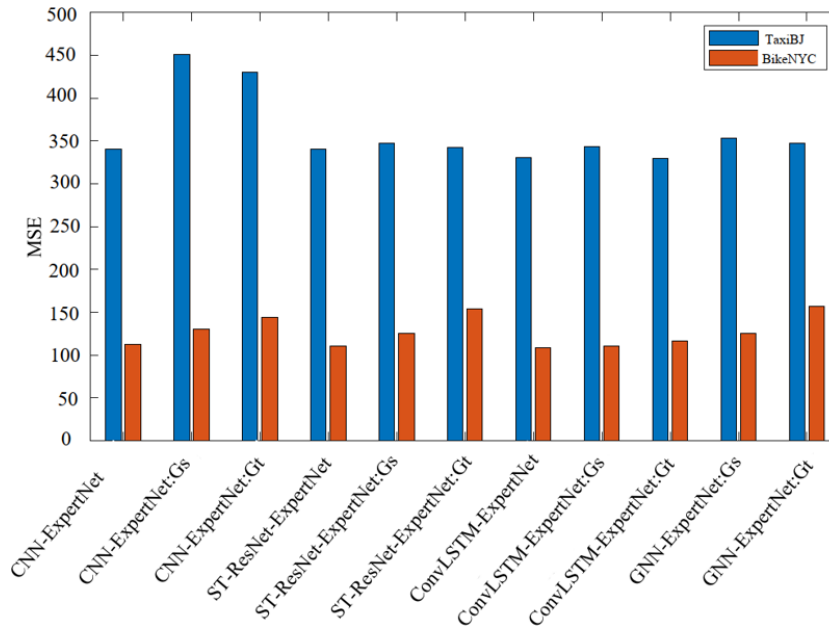


Figure 11. Sequence length of TaxiBJ & BikeNYC - MSE

Figure 12 presents the results of the ablation study for the TaxiBJ and BikeNYC datasets. A comparison with several models is made, and the performance is evaluated using metrics such as MAE and MAPE. The results indicate that BikeNYC achieves superior traffic prediction accuracy.

5.4.2 Parameters search

Sensitivities of three key hyperparameters in CNN-ExpertNet training on four datasets: the number of experts K , the punishment intensity of λ_{er} , and the punishment intensity of λ_{eid} . Every experiment will be conducted three times, and the variance and mean have been computed and shown.

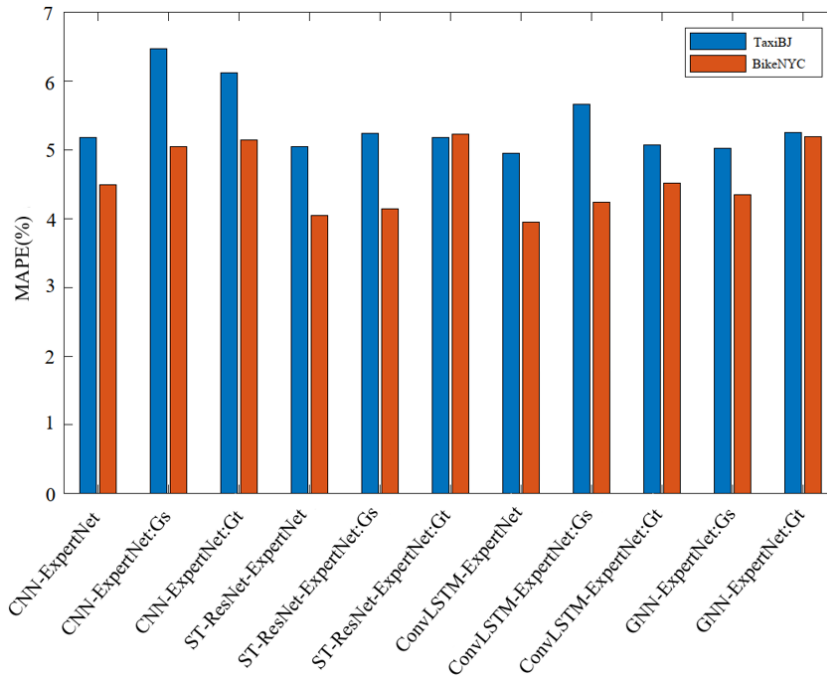


Figure 12. Sequence length of TaxiBJ & BikeNYC - MAPE

6 Conclusion and Future Work

DSTGN-ExpertNet, as proposed in this study, addresses the limitations inherent in conventional traffic prediction models, representing a significant advancement in the field. By leveraging advanced deep learning techniques to capture temporal dynamics and incorporating GNNs to model spatial relationships, DSTGN-ExpertNet effectively handles the complex and diverse traffic patterns characteristic of urban environments. The model architecture is supported by a dynamic gating network, which allocates resources based on real-time data, while simultaneously utilizing a range of expert models, each designed to specialise in distinct traffic patterns. This innovative approach ensures more accurate and adaptable traffic predictions.

Extensive evaluations conducted on large-scale traffic datasets from Beijing and New York demonstrate that DSTGN-ExpertNet outperforms existing methods, including attention-based models and ST-GCN, in both predictive accuracy and interpretability. Moreover, the framework not only enhances the accuracy of traffic forecasts but also provides deeper insights into the underlying traffic dynamics, offering valuable support for urban traffic management and planning. The results underscore the potential of this model to contribute significantly to the optimisation of ITS and urban mobility solutions.

Data Availability

The data used to support the research findings are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflict of interest.

References

- [1] S. Suhas, V. V. Kalyan, M. Katti, B. A. Prakash, and C. Naveena, "A comprehensive review on traffic prediction for intelligent transport system," in *2017 International Conference on Recent Advances in Electronics and Communication Technology (ICRAECT)*, Bangalore, India, 2017, pp. 138–143. <https://doi.org/10.1109/ICRAECT.2017.33>
- [2] J. Barros, M. Araujo, and R. J. Rossetti, "Short-term real-time traffic prediction methods: A survey," in *2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, Budapest, Hungary, 2015, pp. 132–139. <https://doi.org/10.1109/MTITS.2015.7223248>
- [3] K. Lee, M. Eo, E. Jung, Y. Yoon, and W. Rhee, "Short-term traffic prediction with deep neural networks: A survey," *IEEE Access*, vol. 9, pp. 54 739–54 756, 2021. <https://doi.org/10.1109/ACCESS.2021.3071174>
- [4] S. George and A. K. Santra, "Traffic prediction using multifaceted techniques: A survey," *Wirel. Pers. Commun.*, vol. 115, no. 2, pp. 1047–1106, 2020. <https://doi.org/10.1007/s11277-020-07612-8>

- [5] K. Irawan, R. Yusuf, and A. S. Prihatmanto, "A survey on traffic flow prediction methods," in *2020 6th International Conference on Interactive Digital Media (ICIDM)*, Bandung, Indonesia, 2020, pp. 1–4. <https://doi.org/10.1109/ICIDM51048.2020.9339675>
- [6] M. Shaygan, C. Meese, W. Li, X. G. Zhao, and M. Nejad, "Traffic prediction using artificial intelligence: Review of recent advances and emerging opportunities," *Transp. Res. Part C Emerg. Technol.*, vol. 145, p. 103921, 2022. <https://doi.org/10.1016/j.trc.2022.103921>
- [7] J. Wang, J. Jiang, W. Jiang, C. Li, and W. X. Zhao, "Libcity: An open library for traffic prediction," in *29th International Conference on Advances in Geographic Information Systems*, Beijing, China, 2021, pp. 145–148. <https://doi.org/10.1145/3474717.3483923>
- [8] Y. Li, W. Zhao, and H. Fan, "A spatio-temporal graph neural network approach for traffic flow prediction," *Mathematics*, vol. 10, no. 10, p. 1754, 2022. <https://doi.org/10.3390/math10101754>
- [9] J. Simeunović, B. Schubnel, P. J. Alet, and R. E. Carrillo, "Spatio-temporal graph neural networks for multi-site PV power forecasting," *IEEE Trans. Sustain. Energy*, vol. 13, no. 2, pp. 1210–1220, 2021. <https://doi.org/10.1109/TSTE.2021.3125200>
- [10] M. Khodayar and J. Wang, "Spatio-temporal graph deep neural network for short-term wind speed forecasting," *IEEE Trans. Sustain. Energy*, vol. 10, no. 2, pp. 670–681, 2018. <https://doi.org/10.1109/TSTE.2018.2844102>
- [11] U. Ali and T. Mahmood, "Using deep learning to predict short term traffic flow: A systematic literature review," in *Intelligent Transport Systems—From Research and Development to the Market Uptake: First International Conference, INTSYS 2017*, Hyvinkää, Finland, 2018, pp. 90–101. https://doi.org/10.1007/978-3-319-93710-6_11
- [12] H. Korkmaz and M. A. Erturk, "Prediction of the traffic incident duration using statistical and machine-learning methods: A systematic literature review," *Technol. Forecast. Soc. Change*, vol. 207, p. 123621, 2024. <https://doi.org/10.1016/j.techfore.2024.123621>
- [13] J. Shi, Y. B. Leau, K. Li, Y. J. Park, and Z. Yan, "Optimization and decomposition methods in network traffic prediction model: A review and discussion," *IEEE Access*, vol. 8, pp. 202 858–202 871, 2020. <https://doi.org/10.1109/ACCESS.2020.3036421>
- [14] S. Kolekar, S. Gite, B. Pradhan, and K. Kotecha, "Behavior prediction of traffic actors for intelligent vehicle using artificial intelligence techniques: A review," *IEEE Access*, vol. 9, pp. 135 034–135 058, 2021. <https://doi.org/10.1109/ACCESS.2021.3116303>
- [15] M. Stojčić, "Application of ANFIS model in road traffic and transportation: A literature review from 1993 to 2018," *Oper. Res. Eng. Sci. Theor. Appl.*, vol. 1, no. 1, pp. 40–61, 2018. <https://doi.org/10.31181/oresta19012010140s>
- [16] Z. Xiao, X. Fu, L. Zhang, and R. S. M. Goh, "Traffic pattern mining and forecasting technologies in maritime traffic service networks: A comprehensive survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 5, pp. 1796–1825, 2019. <https://doi.org/10.1109/TITS.2019.2908191>
- [17] M. E. Shaik, M. M. Islam, and Q. S. Hossain, "A review on neural network techniques for the prediction of road traffic accident severity," *Asian Transp. Stud.*, vol. 7, p. 100040, 2021. <https://doi.org/10.1016/j.eastsj.2021.100040>
- [18] B. P. Ashwini and R. Sumathi, "Data sources for urban traffic prediction: A review on classification, comparison and technologies," in *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*, Thoothukudi, India, 2020, pp. 628–635. <https://doi.org/10.1109/ICISS49785.2020.9316096>
- [19] C. Chen, K. Li, S. G. Teo, X. Zou, K. Wang, J. Wang, and Z. Zeng, "Gated residual recurrent graph neural networks for traffic prediction," *Proc. AAAI Conf. Artif. Intell.*, vol. 33, no. 1, pp. 485–492, 2019. <https://doi.org/10.1609/aaai.v33i01.3301485>
- [20] G. Jin, Y. Liang, Y. Fang, Z. Shao, J. Huang, J. Zhang, and Y. Zheng, "Spatio-temporal graph neural networks for predictive learning in urban computing: A survey," *IEEE Trans. Knowl. Data Eng.*, vol. 36, no. 10, pp. 5388–5408, 2023. <https://doi.org/10.1109/TKDE.2023.3333824>
- [21] H. Yao, X. Tang, H. Wei, G. Zheng, and Z. Li, "Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction," *Proc. AAAI Conf. Artif. Intell.*, vol. 33, no. 1, pp. 5668–5675, 2019. <https://doi.org/10.1609/aaai.v33i01.33015668>
- [22] D. A. Tedjopurnomo, Z. Bao, B. Zheng, F. M. Choudhury, and A. K. Qin, "A survey on modern deep neural network for traffic prediction: Trends, methods and challenges," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 4, pp. 1544–1561, 2020. <https://doi.org/10.1109/TKDE.2020.3001195>
- [23] T. Otsoshi, Y. Ohsita, M. Murata, Y. Takahashi, K. Ishibashi, and K. Shiimoto, "Traffic prediction for dynamic traffic engineering," *Comput. Netw.*, vol. 85, pp. 36–50, 2015. <https://doi.org/10.1016/j.comnet.2015.05.001>
- [24] M. Akhtar and S. Moridpour, "A review of traffic congestion prediction using artificial intelligence," *J. Adv. Transp.*, vol. 2021, no. 1, p. 8878011, 2021. <https://doi.org/10.1155/2021/8878011>

- [25] A. Boukerche and J. Wang, "Machine learning-based traffic prediction models for intelligent transportation systems," *Comput. Netw.*, vol. 181, p. 107530, 2020. <https://doi.org/10.1016/j.comnet.2020.107530>
- [26] M. F. Iqbal, M. Zahid, D. Habib, and L. K. John, "Efficient prediction of network traffic for real-time applications," *J. Comput. Netw. Commun.*, vol. 2019, no. 1, p. 4067135, 2019. <https://doi.org/10.1155/2019/4067135>
- [27] N. Ramakrishnan and T. Soni, "Network traffic prediction using recurrent neural networks," in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Orlando, FL, USA, 2018, pp. 187–193. <https://doi.org/10.1109/ICMLA.2018.00035>
- [28] J. Xu, D. Deng, U. Demiryurek, C. Shahabi, and M. Van der Schaar, "Mining the situation: Spatiotemporal traffic prediction with big data." *IEEE J. Sel. Top. Signal Process.*, vol. 9, no. 4, pp. 702–715, 2015. <https://doi.org/10.1109/JSTSP.2015.2389196>
- [29] Y. Chen, W. Wang, and X. M. Chen, "Bibliometric methods in traffic flow prediction based on artificial intelligence," *Expert Syst. Appl.*, vol. 228, p. 120421, 2023. <https://doi.org/10.1016/j.eswa.2023.120421>
- [30] P. Cao, F. Dai, G. Liu, J. Yang, and B. Huang, "A survey of traffic prediction based on deep neural network: Data, methods and challenges," in *Cloud Computing: 11th EAI International Conference, Virtual Event, 2021*, pp. 17–29. https://doi.org/10.1007/978-3-030-99191-3_2