MESIGCN: Multi-scale Equidistant Spatio-temporal Graph Convolution Networks

SI-WEI WEI[1,2], DING-BO HU[1*], XIANG-YUAN XU[1], CHUN-ZHI WANG[1]

[1]School of Computer Science, Hubei University of Technology, Wuhan 430068, China
[2]CCCC Second Highway Consultants Co. Ltd., Wuhan 430056, China
E-MAIL: waosfengw@whut.edu.cn, 102201072@hbut.edu.cn, 102301173@hbut.edu.cn, chunzhiwang@hbut.edu.cn
*Corresponding author: DING-BO HU (Email: 102201072@hbut.edu.cn)

Abstract:

Traffic flow forecasting is essential for Intelligent Transportation Systems (ITS) and is vital for managing and developing urban traffic. Traditional methods like historical averages and time-series analysis have limited accuracy due to their failure to address the non-linear and dynamic nature of traffic. Although recent advancements in machine learning and deep learning, including techniques like KNN, SVM, CNN, and LSTM, have improved prediction capabilities, challenges in incorporating external factors such as weather conditions and Points of Interest (POI) persist.

This paper introduces a novel Multi-scale Equidistant Spatio-temporal Graph Convolution Network (MESIGCN) that significantly enhances traffic forecasting accuracy. By integrating external factors with traffic data as distinct feature channels, this model leverages deep learning to merge these channels and model spatio-temporal dependencies effectively. Experimental results demonstrate that MESIGCN surpasses traditional models in both prediction accuracy and robustness, providing a more effective solution for real-world traffic forecasting challenges.

Keywords:

traffic forecasting, deep learning, external factors

1. Introduction

Traffic flow prediction, a critical component of Intelligent Transport Systems (ITS)[1], plays an essential role in urban transportation management by helping to alleviate congestion and optimize resource allocation and traffic state prediction[2]. It leverages historical data to enhance the controllability and stability of urban traffic, improving efficiency[3]. By anticipating potential traffic scenarios such as congestion, road owners can implement effective resource management strategies in advance, thus improving the utilization of traffic resources, reducing delays, minimizing accidents, and enhancing the efficiency of traffic management[4].

Traditional traffic flow prediction methods primarily rely on statistical analysis of historical traffic data or employ neural networks and deep learning to learn features from traffic data [5]. However, future traffic conditions are influenced not only by historical data but also by external factors like weather and Points of Interest (POIs) [6]. For instance, traffic conditions vary with weather changes, and traffic in busy commercial areas is more complex than in quieter ones [7].

To address these challenges, we propose a Multi-scale Equidistant Spatio-temporal Graph Convolutional Network (MESIGCN) that incorporates external factors for traffic flow prediction. This model uses graph convolution and multiscale isometric convolution to effectively capture spatio-temporal dependencies in traffic features. By integrating external factors such as weather and POIs, MESIGCN enhances its understanding of the factors influencing traffic systems. Compared to traditional models, MESIGCN can model essential factors more broadly and maintain high prediction accuracy under various conditions, significantly improving the accuracy and robustness of traffic flow predictions.

2. Model and method

- Model Structure.The main structure of the MESIGCN model is shown in Figure 1. It integrates all processed features with external factors and consists of two main components: the spatio-temporal feature extraction module and the Multi-scale Isometric Spatio-temporal Map Convolutional Network (MISTMC).

In the spatio-temporal feature extraction module, three parallel components are designed to extract spatio-temporal features, weather features, and POI features, re-

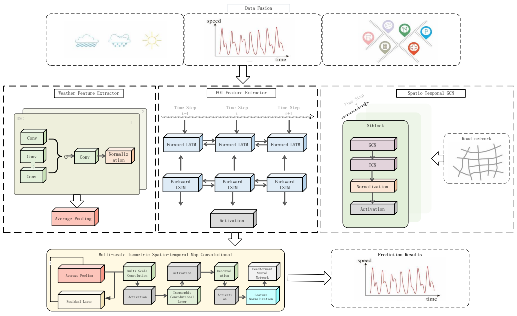

FIGURE 1. The structure of the MESIGCN model, which incorporates external factors of traffic flow (e.g., weather features and POI features). By integrating multi-scale isometric spatio-temporal convolutions and graph neural networks, it effectively captures spatial and temporal dependencies for traffic prediction.

spectively. The MISTMC module captures complex relationships between spatio-temporal features by performing transformations such as decomposition, downsampling, and isometric convolution on the time series data through multi-scale and isometric convolution operations. This enhances the correlation understanding and improves the model's generalization ability.

2.1. Spatio-temporal Feature Extraction Module

The spatio-temporal feature extraction module processes multi-dimensional features affecting traffic flow and consists of three components: the Weather Feature Extractor (WFE), the Spatio Temporal GCN (STG), and the POI Feature Extractor. The outputs of these components are concatenated and passed to subsequent layers, effectively combining spatial, temporal, and external factors to enhance the model's predictive performance and robustness.

- Weather Feature Extractor.The WFE module employs a two-layer Depthwise Separable Convolution (DSC) to extract temporal and feature-dimension information from weather data:

$$\text{DSC} = \sum_{c=1}^{C_{\text{in}}} (\mathbf{X}_{n,c} * \mathbf{k}_c)(l) \cdot w_{c',c} \quad (1)$$

Here, $\mathbf{n}$ is the sample index, c the channel index, and l the timestep index. Each input channel is convolved independently to extract temporal features, which are then linearly combined across channels. After two DSCs with Batch Normalization (BN) and ReLU activation, internal covariate shifts are reduced to accelerate convergence and prevent overfitting:

$$\mathbf{Y}_1 = \text{ReLU}(\text{BN}(\text{DSC}(\mathbf{X}))) \quad (2)$$

$$\mathbf{Y}_2 = \text{ReLU}(\text{BN}(\text{DSC}(\mathbf{Y}_1))) \quad (3)$$

$$\text{WFE} = \mathbf{P}(n, c) = \frac{1}{L} \sum_{l=1}^{L} \mathbf{Y}_2(n, c, l) \quad (4)$$

Adaptive mean pooling adjusts the output to a fixed size, facilitating fusion with other features.

- Spatio Temporal GCN.The STG module extracts spatio-temporal features from the fused data $\mathbf{X} \in \mathbb{R}^{B \times T \times N \times F_{\text{in}}}$, which includes traffic data and the road network ($\mathbf{A}$). Temporal information enriches feature representations and supports multi-task learning. For each timestep t, the data is processed with GCN, followed by Batch Normalization and ReLU activation:

$$\mathbf{Z}_t = \sigma(\mathbf{A}\mathbf{X}_t\mathbf{W}) \quad (5)$$

$$\mathbf{H}_t = \text{ReLU}(\text{BatchNorm}(\mathbf{Z}_t)) \quad (6)$$

This operation forms an STGCINBlock, which is iteratively applied:

$$\mathbf{H}_{t+1} = \text{STGCINBlock}(\mathbf{H}_t, \mathbf{A}) \quad (7)$$

The final output is:

$$\mathbf{STG} = [\mathbf{H}_1^{(L)}, \mathbf{H}_2^{(L)}, \ldots, \mathbf{H}_T^{(L)}] \in \mathbb{R}^{B \times T \times N \times F_{\text{out}}} \quad (8)$$

The STG module effectively captures spatial structures and temporal dynamics, enhancing performance in complex traffic scenarios. Its flexibility allows adjustment of the number of modules based on data complexity.

- POI Feature Extractor.The POI dataset comprises 9 types of POIs. For each road section, the predominant POI type represents the feature vector $\mathbf{x}_t \in \mathbb{R}^F$ for time step t. The input sequence is:

$$\mathbf{X}_{\text{POI}} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T]^\top \in \mathbb{R}^{T \times F} \quad (9)$$

A bidirectional LSTM captures temporal dynamics:

$$\overleftrightarrow{\mathbf{h}} = \text{BiLSTM}(\mathbf{X}_{\text{POI}}) = [\overrightarrow{\mathbf{h}}; \overleftarrow{\mathbf{h}}] \in \mathbb{R}^{2H} \quad (10)$$

The hidden state is mapped to the target feature dimension:

$$\text{POI} = \sigma(\mathbf{W}\overleftrightarrow{\mathbf{h}} + \mathbf{b}) \in \mathbb{R}^{H'} \quad (11)$$

This approach captures POI feature correlations and trends, providing rich and precise representations integrated with traffic and weather data.

2.2. Multi-scale Isometric Spatio-temporal Map Convolutional Networks

Outputs from WFE, STG, and POI are aligned in the time dimension and concatenated to form the multimodal feature sequence $\mathbf{F}_{\text{input}} \in \mathbb{R}^{N \times T \times (3D)}$. The sequence undergoes average pooling cS to capture long-term trends and a residual link ($\mathbf{R}$) to capture short-term volatility:

$$\mathbf{S} = \mathbf{AvgPoul}1\text{D}(\mathbf{F}_{\text{input}}, k) \tag{12}$$

$$\mathbf{R} = \mathbf{F}_{\text{input}} - \mathbf{S} \tag{13}$$

$$\mathbf{C}^{(s)} = \text{ReLU}(\text{Conv1D}(k_s)) \tag{14}$$

Temporal dependencies are captured through multi-scale convolution followed by isometric convolution:

$$\mathbf{I}^{(s)} = \text{ReLU}(\text{Conv1D}(k_s^{(s)})) \tag{15}$$

Features are upsampled to original temporal length and fused:

$$\mathbf{U}^{(s)} = \text{ReLU}(\text{ConvTranspose1D}(\mathbf{I}^{(s)}, k_s)) \tag{16}$$

$$f_{\text{multi}} = \text{Aggregate}(\{\mathbf{U}^{(s)}\}_{s=1}^{S}) \tag{17}$$

$$f_{\text{output}} = \text{LayerNorm}(f_{\text{multi}} + \text{FFN}(f_{\text{multi}})) \tag{18}$$

MISTMC integrates multimodal data (weather, POI, traffic speed) and enhances prediction performance through feature fusion and multi-scale temporal capturing. Residual connections and normalization improve model stability and prevent overfitting.

3. Experiments

3.1. Dataset

- traffic data.In this experiment, the primary source of traffic data is the SZ_taxi dataset, along with its neighborhood matrix. The SZ_taxi dataset consists of taxi travel trajectory data collected every inute by sensors during January 2015, covering 156 major road segments in Luohu District, Shenzhen.
- external factors.In this experiment, the dataset of external factors mainly uses SZ_Weather and SZ_POI. SZ_Weather records five types of weather conditions in Shenzhen in January 2015 at 15-minute intervals, constructing a 156*2976 matrix aligned with the traffic data. The SZ_POI dataset provides point-of-interest (POI) information for selected road sections, divided into 9 categories such as restaurants, living services, and medical services. To integrate external POI factors with the traffic data seamlessly, we extract the POI types that dominate in specific sections and construct them into a 156*1 matrix, simplifying data dimensions and emphasizing key information.

3.2. Experimental Setup

We conducted our experiments on a physical device equipped with a Core i7 8700 CPU, 2070ti GPU with 8GB of memory, and 64GB of RAM. The learning rate was set to 0.001, the batch size was 64, and the training set utilized 80

To evaluate model performance, we employed five metrics: mean absolute error (MAE), root mean square error (RMSE), Accuracy, Coefficient of Determination (R^2), and Explained Variation Score (VAR). MAE and RMSE measure the difference between actual and predicted values, with smaller values indicating better model performance. Accuracy measures prediction correctness, where a higher value signifies better performance. R^2 represents the proportion of variance explained by the model, with values ranging from negative infinity to 1; a value closer to 1 implies a better fit, while negative values indicate poorer performance compared to a simple average. VAR measures how well the model explains the overall variance in the data, also ranging from negative infinity to 1, with values closer to 1 indicating better performance.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{19}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{20}$$

$$\text{Accuracy} = 1 - \frac{\|Y - \hat{Y}\|_F}{\|Y\|_F} \tag{21}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{22}$$

$$\text{VAR} = 1 - \frac{\text{Var}(y - \hat{y})}{\text{Var}(y)} \tag{23}$$

In the formula for the above indicators, $\bar{y}$ represents the average of the true values, y represents the true value, $\hat{y}$

represents the predicted vColor legendalue, Y represents the matrix of true values, $\hat{Y}$ represents the matrix of predicted values, and var represents the variance.

3.3. Baseline

We compared the MESIGCN model against six baseline methods for traffic flow prediction:

SVR[8]: Utilizes Incremental Support Vector Regression for real-time updating and excels in short-term traffic flow prediction.

ARIMA[9]: Autoregressive Integrated Moving Average Model. It integrates three main time - series modeling techniques: Autoregressive (AR), Differencing (I), and Moving Average (MA).

GCN[10]: Enhances Graph WaveNet by introducing learning rate decay, increasing filter numbers, and adding jump connections along with short-term pre-training strategies, significantly improving traffic prediction performance.

DCRNN[11]: Combines graph diffusion convolution with Gated Recurrent Units (GRU) to effectively capture spatiotemporal dependencies in traffic data, demonstrating substantial improvements over other models across various forecasting horizons.

TGCN[12]: Merges the benefits of Graph Convolutional Networks (GCN) and Temporal Recurrent Neural Networks (GRU) to enhance traffic flow prediction accuracy by considering both temporal and spatial dependencies.

AST-GCN[13]: An Attribute-augmented Spatio-Temporal Graph Convolutional Network that boosts prediction accuracy and interpretability by incorporating dynamic and static external factors of the road network.

3.4. Experimental Results and Analysis

- Comparison and Analysis of Model Metrics. Table 1 compares MESIGCN with six baseline models. MESIGCN achieves lower MAE and RMSE, higher Accuracy, R^2, and VAR, demonstrating superior performance, greater prediction precision, and enhanced stability.

Specifically, compared to the best baseline model, AST-GCN, MESIGCN reduces MAE by 1.67%, RMSE by 0.21%, increases Accuracy by 0.07%, and both R^2 and VAR by 0.08%.

Thus, MESIGCN outperforms the six classic traffic flow prediction models across all five metrics, confirming its superior performance.

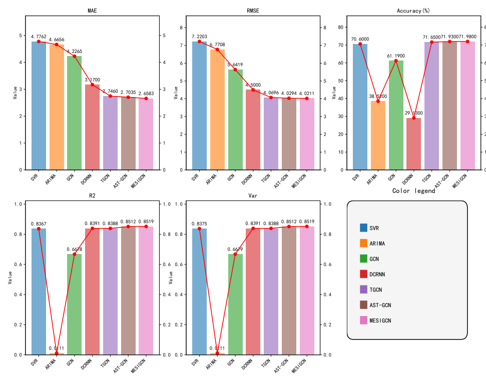

FIGURE 2. Comparison of model performance metrics, including the performance of MESIGCN and all baseline models across five metrics.

TABLE 1. Model Performance Metrics: A Comparison of MAE, RMSE, Accuracy, R², and Variance

	Model						
Metrics	SVR	ARIMA	GCN	DCRNN	TGCN	AST-GCN	MESIGCN
MAE	4.7762	4.6656	4.2265	3.1700	2.7460	2.7035	2.6583
RMSE	7.2203	6.7708	5.6419	4.5000	4.0696	4.0294	4.0211
Accuracy(%)	70.60	38.52	61.19	29.13	71.65	71.93	71.98
R²	0.8367	0.0111	0.6678	0.8391	0.8388	0.8512	0.8519
Var	0.8375	0.0111	0.6679	0.8391	0.8388	0.8512	0.8519

- Performance Analysis Over Different Horizons. Building on the model metrics comparison, this section analyzes MESIGCN against four advanced deep learning models across additional prediction tasks with varying time steps: 6 (30 minutes), 9 (45 minutes), and 12 (60 minutes). Since AST-GCN and MESIGCN rank highest for all metrics, only their performances are compared in Table

For 6 time steps, MESIGCN reduces MAE by 1.48%, RMSE by 0.20%, and increases Accuracy by 0.10%, R^2 by 0.06%, and VAR by 0.04% compared to AST-GCN.

At 9 time steps, MESIGCN decreases MAE by 2.42%, RMSE by 0.68%, and boosts Accuracy by 0.25%, R^2 by 0.25%, and VAR by 0.27%.

For 12 time steps, MESIGCN lowers MAE by 2.62%, RMSE by 0.75%, and enhances Accuracy by 0.31%, R^2 by 0.28%, and VAR by 0.27% relative to AST-GCN. 2. Overall, MESIGCN consistently outperforms baseline models across all time intervals, showing superior stability and performance in both short-term and long-term predictions, with increasing advantages as the prediction horizon extends.

- Analysis of Prediction Results. Figures 3 and 4 illus-

TABLE 2. Comparison of Model Performance Across Different Time Steps

Time	Metric	DCRNN	TGCN	AST-GCN	MESIGCN
15 min	MAE	3.1700	2.7460	2.7035	2.6583
	RMSE	4.5000	4.0696	4.0294	4.0211
	Accuracy(%)	29.13	71.65	71.93	71.98
	R2	0.8391	0.8388	0.8512	0.8519
	Var	0.8391	0.8388	0.8512	0.8519
30 min	MAE	3.2300	2.7470	2.7265	2.6862
	RMSE	4.5600	4.0770	4.0529	4.0448
	Accuracy(%)	29.70	71.59	71.76	71.83
	R2	0.8332	0.8377	0.8494	0.8499
	Var	0.8360	0.8377	0.8495	0.8498
45 min	MAE	3.2700	2.7788	2.7611	2.6944
	RMSE	4.6000	4.1035	4.0822	4.0544
	Accuracy(%)	0.3021	71.41	71.56	71.74
	R2	0.8275	0.8327	0.8473	0.8494
	Var	0.8314	0.8357	0.8474	0.8496
60 min	MAE	3.3100	2.7911	2.7744	2.7016
	RMSE	4.6400	4.1266	4.1001	4.0692
	Accuracy(%)	30.69	71.25	71.43	71.64
	R2	0.8219	0.8339	0.8459	0.8483
	Var	0.8267	0.8340	0.8460	0.8483

trate traffic data for sections 90217–90221 in the SZ_Taxi dataset from January 1st to 7th, including peak traffic during the New Year holiday. MESIGCN's predictions (red lines) closely match the actual traffic data (blue lines) both during peak and non-peak periods over the week and on the specific day of January 3rd.

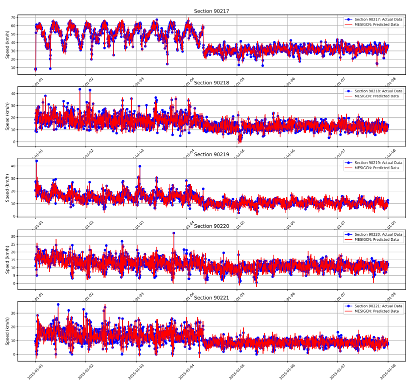

FIGURE 3. The prediction performance of the MESIGCN model on weekly traffic data, showing the speed variation trends of actual and predicted data across multiple sections (e.g., 90217, 90218).

These results demonstrate that MESIGCN effectively predicts traffic flow trends in both daily and weekly scopes. Combined with the metrics in Figure 2, Tables 1 and 2, MESIGCN significantly outperforms classical and ad-

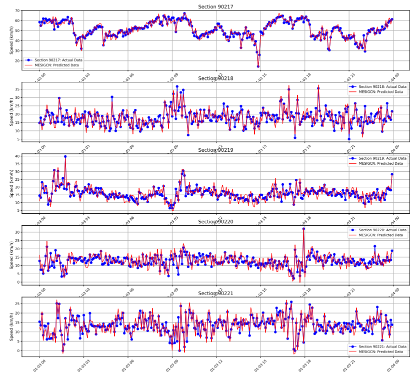

FIGURE 4. The prediction performance of the MESIGCN model on daily traffic data, showing the speed variation trends of actual and predicted data across multiple sections (e.g., 90217, 90218).

vanced baseline models, showcasing high consistency with actual traffic data and strong applicability and stability across various scenarios.

3.5. Ablation Study

- Model Ablation. To evaluate the effectiveness of the MESIGCN model, ablation experiments were conducted on the SZ Taxi dataset. Four model variants were proposed: W/O WFE (removing the weather feature extractor), W/O POI (removing the POI feature extractor), W/O POI & WFE (removing both weather and POI feature extractors), and W/O MISTMC (removing the MISTMC module). The complete model (Base) was compared with these variants in Table 3. The results show that W/O WFE, W/O POI, and W/O POI & WFE all perform worse than the full model across various metrics, indicating that WFE and POI significantly impact MESIGCN's structure. Specifically, W/O POI & WFE exhibits a substantial performance drop compared to MESIGCN, demonstrating that these features are crucial for traffic flow prediction. Additionally, W/O MISTMC also shows performance degradation at different scales, suggesting that this module significantly contributes to the model's performance.
- Dataset Ablation.To assess the impact of external factors on traffic flow prediction, ablation experiments were conducted on the dataset. Three variants were proposed: W/O SZ_Weather (removing weather data), W/O

TABLE 3. Model Ablation Results

Metrics	W/O WFE	W/O POI	W/O POI&WFE	W/O MISTMC	Base
MAE	2.7731	2.7829	2.9194	2.7109	2.6583
RMSE	4.3010	4.2436	4.0344	4.0271	4.0211
Accuracy (%)	71.03	70.89	70.43	71.34	71.98
R^2	0.8508	0.8508	0.8350	0.8515	0.8519
Variance	0.8508	0.8508	0.8361	0.8515	0.8519

SZ_POI (removing POI data), and W/O SZ_POI & SZ_Weather (removing all external factors). The prediction performance of the model with external factors (Base) was compared with these variants in Table 4. Removing either weather or POI data resulted in a slight decrease in prediction performance compared to the model with external factors. However, removing both data types led to a significant decline in prediction performance. This indicates that external factors have a considerable impact on traffic flow prediction, especially when multiple external variables are considered. Therefore, incorporating the influence of external factors will be an important direction for future research in traffic flow prediction.

TABLE 4. Dataset Ablation Results

Metrics	W/O SZ_Weather	W/O SZ_POI	W/O SZ_Weather & SZ_POI	Base
MAE	2.6829	2.6731	2.9194	2.6583
RMSE	4.0348	4.0344	4.2436	4.0211
Accuracy (%)	70.89	71.19	70.43	71.98
R^2	0.8508	0.8508	0.8350	0.8519
Variance	0.8512	0.8512	0.8361	0.8519

4. CONCLUSION

This paper introduces the MESIGCN model to address traffic flow prediction incorporating external factors. MESIGCN employs three parallel deep learning modules to extract features from external factors and traffic data, using graph convolution and multi-scale isotropic convolution to capture spatiotemporal dependencies effectively. Extensive experiments demonstrate MESIGCN's adaptability and robustness in processing multi-feature tasks and its superior performance in traffic flow prediction with external factors. Thanks to its parallel structure, MESIGCN can integrate richer feature representations, showing significant scalability for future applications.

Future research in traffic flow prediction should emphasize the influence of external factors on prediction accuracy, with a focus on how to integrate and process these external features with traffic data.

Acknowledgements

This work is funded by the National Natural Science Foundation of China(62472149) and enterprises research project(Intelligent prediction of 5G transmission network traffic and energy consumption based on neural network)

References

[1] V. C. S. Reddy, S. Ganji, M. M. Nayak, M. M. Yadav, and G. D. Reddy, "Survey on traffic flow prediction for intelligent transportation system using machine learning," World Journal of Advanced Research and Reviews, vol. 17, no. 2, pp. 460–463, 2023.

[2] A. A. Musa, S. I. Malami, F. Alanazi, W. Ounaies, M. Alshammari, and S. I. Haruna, "Sustainable Traffic Management for Smart Cities Using Internet-of-Things-Oriented Intelligent Transportation Systems (ITS): Challenges and Recommendations," Sustainability, vol. 15, no. 13, p. 9859, 2023.

[3] G. Balan and S. Luke, "History-based traffic control," in Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems, 2006, pp. 616–621.

[4] C. Chen, K. Li, S. G. Teo, X. Zou, K. Li, and Z. Zeng, "Citywide traffic flow prediction based on multiple gated spatio-temporal convolutional neural networks," ACM Transactions on Knowledge Discovery from Data (TKDD), vol. 14, no. 4, pp. 1–23, 2020, ACM New York, NY, USA.

[5] 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 Transactions on Knowledge and Data Engineering, vol. 34, no. 4, pp. 1544–1561, 2020.

[6] A. Koesdwiady, R. Soua, and F. Karray, "Improving traffic flow prediction with weather information in connected cars: A deep learning approach," IEEE Transactions on Vehicular Technology, vol. 65, no. 12, pp. 9508–9517, 201

[7] M. Jiang, C. Li, K. Li, and H. Liu, "Destination prediction based on virtual POI docks in dockless

bike-sharing system," IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 3, pp. 2457–2470, 2021.

[8] H. Su, L. Zhang, and S. Yu, "Short-term traffic flow prediction based on incremental support vector regression," in Third International Conference on Natural Computation (ICNC 2007), vol. 1, pp. 640–645, 2007.

[9] M. M. Hamed, H. R. Al-Masaeid, and Z. M. Bani Said, "Short-term prediction of traffic volume in urban arterials," Journal of Transportation Engineering, vol. 121, no. 3, pp. 249–254, 1995.

[10] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," arXiv preprint arXiv:1609.02907, 2016.

[11] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," arXiv preprint arXiv:1707.01926, 2017.

[12] L. Zhao, Y. Song, C. Zhang, Y. Liu, P. Wang, T. Lin, M. Deng, and H. Li, "T-GCN: A temporal graph convolutional network for traffic prediction," IEEE Transactions on Intelligent Transportation Systems, vol. 21, no. 9, pp. 3848–3858, 2019.

[13] J. Zhu, Q. Wang, C. Tao, H. Deng, L. Zhao, and H. Li, "AST-GCN: Attribute-augmented spatiotemporal graph convolutional network for traffic forecasting," IEEE Access, vol. 9, pp. 35973–35983, 2021.