Integrating GCN, BiLSTM, and Attention for Accurate Short-Term Traffic Forecasting on Urban Road Networks

Xuanhao Tian

School of Artificial Intelligence, Xidian University, Xi'an 710126, China
23009200444@stu.xidian.edu.cn

Abstract. Urban traffic congestion has become a global challenge, causing economic losses, environmental pollution, and reduced quality of life. Accurate short-term traffic flow prediction is essential for optimizing traffic management, improving travel efficiency, and supporting the development of intelligent transportation systems. When applied to large-scale traffic systems, conventional statistical models frequently fail to account for the highly nonlinear and interdependent spatio-temporal patterns. This paper presents a GCN-BiLSTM-Attention model for short-term traffic flow prediction. This model integrates graph convolution to capture spatial dependencies, bidirectional LSTM for temporal modeling, and an attention mechanism to enhance interpretability. Evaluated on the METR-LA dataset with 207 traffic sensors, the model achieves an MAE of 1.96 mph, RMSE of 4.25 mph, and ±10% accuracy of 91.58%. A comprehensive data preprocessing pipeline—including anomaly removal, imputation, and normalization—ensures high-quality input. Attention weight analysis shows a focus on recent time steps, aligning with traffic dynamics. Results demonstrate the model's effectiveness in learning complex spatio-temporal patterns and its potential for deployment in intelligent transportation systems.

Keywords: Graph Convolutional Network, Bidirectional Long Short-Term Memory, Attention Mechanism, Spatio-Temporal Modeling.

1 Introduction

High levels of traffic are putting increasing pressure on today's cities, leading to heavy congestion that threatens economic performance and environmental integrity. As smart infrastructure becomes more common, the need to predict traffic flow accurately has grown. Accurate traffic forecasting supports real-time traffic control, dynamic routing, and congestion mitigation strategies. Short-term traffic forecasting can ease urban congestion and contribute to smart traffic management [1].

Earlier traffic forecasting methodologies were based on traditional statistical approaches such as HA, ARIMA models, and Kalman filters [2]. These approaches were initially attractive due to their ease of implementation and low computational cost.

S. Zhang (ed.), *Proceedings of the 2025 International Conference on Electronics, Electrical and Grid Technology (ICEEGT 2025)*, Advances in Engineering Research 292,
https://doi.org/10.2991/978-94-6463-986-5_55

However, they struggle to generalize in modern, large-scale transportation networks. Their dependency on data stationarity and linear assumptions limits their capacity to capture the complex nonlinear nature of urban traffic states. Furthermore, these methods fail to effectively model interdependencies among road network sections, which are non-Euclidean and graph-like in nature.

A new age in traffic prediction has been brought about by deep learning technologies, which provide strong instruments for identifying intricate patterns in massive amounts of traffic data [3]. In contrast to conventional models, deep learning is capable of efficiently modeling nonlinear dependencies across geographical and temporal dimensions and automatically extracting hierarchical features. When it comes to identifying temporal patterns in sequential traffic observations, RNNs and Long Short-term Memory (LSTM) networks have demonstrated excellent performance [4]. Graph-based traffic network structure hinders convolutional neural networks (CNNs), which are excellent at identifying spatial patterns in grid-like data [5].

Graph Convolutional Networks (GCNs) address these structural constraints by generalizing convolution operations to the graph domain, modeling connectivity between traffic monitoring sensors. The fusion of GCN architectures and LSTM units creates powerful spatio-temporal learning ability, where graph convolution learns spatial correlations and LSTM learns temporal correlations [6]. This hybrid model, known as GCN-LSTM or T-GCN, yields significant improvement over traditional statistical models and individual deep learning components [7].

Although existing studies have demonstrated GCN-LSTM model effectiveness, there remains a lack of fully open implementations detailing its architecture and preprocessing pipeline [8]. This work provides a concise PyTorch-based implementation to facilitate reproducibility and future improvements.

The paper builds a straightforward yet effective GCN-BiLSTM model using PyTorch for short-term traffic forecasting with the METR-LA dataset, a benchmark consisting of real traffic sensor data from Los Angeles highways. The model uses graph convolution to simulate how traffic sensors are spatially dependent while using LSTM units for temporal correlation in time series. To increase the model's ability to concentrate on time steps that are informative, an attention mechanism is incorporated, enabling dynamic weighting of temporal features and improving prediction accuracy and interpretability [9, 10].

This is how the rest of the paper is structured. The methods, including datasets, preprocessing procedures, and model architectures, are presented in Section 2. Experimental results are reported in Section 3. Discussion and perspectives are given in Section 4. The article is concluded and next directions are outlined in Section 5.

2 Methods

2.1 Data Source and Description

The study employs the publicly available METR-LA traffic dataset, which consists of real-time traffic speed recordings from 207 loop detectors deployed across major roadways in Los Angeles. The dataset includes both the temporal speed measurements and a spatial adjacency matrix representing the pairwise connectivity among sensors. The data spans approximately four months, with a sampling frequency of every five minutes, totaling 34272-time steps.

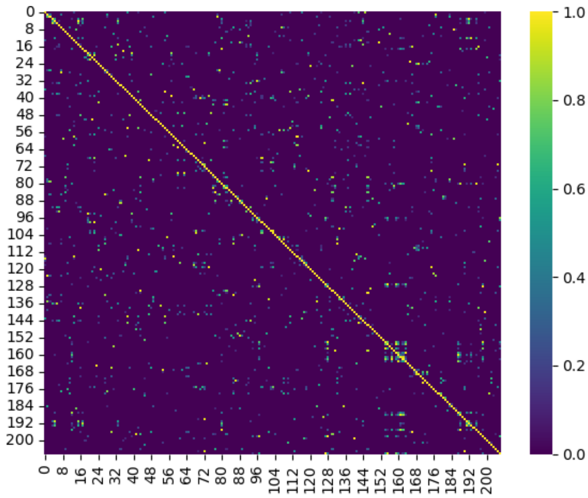

Fig. 1. Heatmap of the adjacency matrix among sensors. (Picture credit: Original)

Figure 1 illustrates the adjacency matrix of the 207 sensors, which is sparse and symmetric, indicating that most sensors do not share direct spatial connections. The presence of localized bright regions suggests areas of strong spatial correlation, likely due to geographical proximity or shared traffic patterns. This spatial structure is well-suited for capturing localized dependencies using graph neural networks. Figure 2 shows the variation of traffic speed over time.

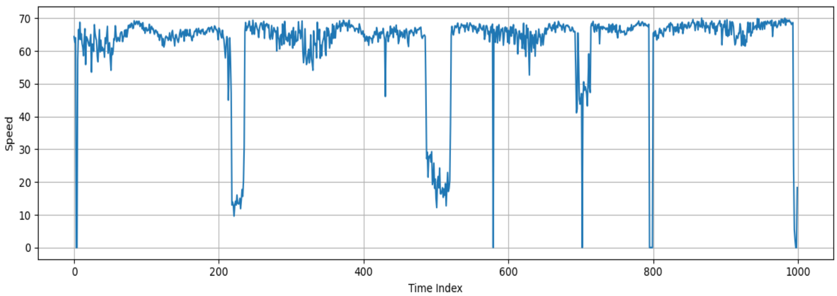

Fig. 2. Sensor 0 - Traffic speed over time. (Picture credit: Original)

To enhance the model's ability to learn from traffic flow data, a systematic data cleaning and preprocessing procedure was applied to the raw sensor records. First, system failures were identified by detecting time steps during which all sensors simultaneously recorded zero values. To avoid misclassifying periods of naturally low traffic, nighttime hours (00:00–04:00) were excluded from this check.

Next, the paper detected and removed various types of anomalous values, including isolated zeros surrounded by high-speed readings, unrealistic speeds (below 2 mph or above 70 mph), and statistical outliers identified using Z-score, interquartile range (IQR), and percentile-based methods. In addition, abrupt spikes in speed over short time windows were also filtered out to eliminate isolated temporal anomalies.

The missing values resulting from the cleaning process were imputed using a multi-stage filling strategy. This involved linear interpolation and limited forward/backward filling for short gaps, followed by using the historical average at the same time of day to preserve temporal traffic patterns.

Finally, the cleaned data were normalized to the [0,1] range using Min-Max scaling, and a sliding window approach was employed to generate supervised learning samples, where each input sequence is paired with corresponding prediction targets for model training. A summary of dataset statistics and preprocessing outcomes is presented in Table 1.

Table 1. Dataset Statistics and Preprocessing Summary

Project	Value	Description
Number of time steps	34,272	Sampled every 5 minutes
Number of sensors	207	
Zero-value ratio (before)	8.11%	
Zero-value ratio (after)	3.44%	After removal of anomalous values
Total number of anomalies removed	609,113	
Average speed after imputation	58.16 mph	Minimum 0.00, maximum 70.00
Number of sliding window samples	34,259	12-step input, 2-step output
Normalization range	[0, 1]	Min-Max scaling applied

2.2 Proposed Method

The paper proposes a GCN-BiLSTM-Attention model for short-term traffic forecasting. This model integrates spatial and temporal dependencies with a dynamic attention mechanism. The structure is as follows:

Graph Convolution Layers (GCN). Two GCN layers extract spatial dependencies based on a normalized adjacency matrix $\hat{A} = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$, where $\boldsymbol{A}$ is the adjacency matrix and $\boldsymbol{D}$ is the degree matrix.

Temporal Modeling (BiLSTM). The time-series outputs of the GCN layers are reshaped and passed to a bidirectional LSTM to capture forward and backward temporal dependencies.

Attention Mechanism. An attention layer assigns dynamic weights across time steps, allowing the model to emphasize informative historical inputs.

Prediction LayerThe attention-weighted context vector is passed through a fully connected layer to produce the final prediction of shape (B, 2, 207), as shown in Figure 3.

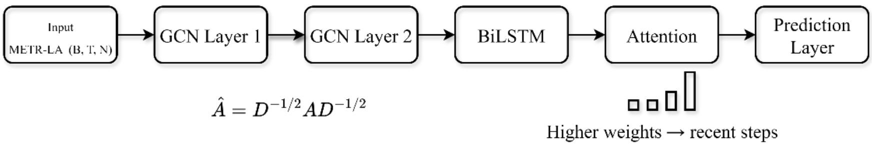

Fig. 3. GCN-BiLSTM-Attention model framework. (Picture credit: Original)

The training process employs L1 loss as the objective function and Adam as the optimization algorithm. The hyperparameters are set as follows. The learning rate is fixed at 0.0005, while the batch size is configured to 32. Training is allowed to proceed for up to 50 epochs, with early stopping triggered if no improvement is observed within 5 consecutive epochs. Training and validation losses are tracked at each epoch to assess performance and mitigate overfitting.

2.3 Performance Metrics

To evaluate the model, five performance indicators are employed:

- Mean Absolute Error (MAE) is defined as the mean value of the absolute differences between predicted outputs and their corresponding ground truth values, serving as an indicator of prediction accuracy.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (1)$$

- Root Mean Squared Error (RMSE) is computed as the square root of the mean squared error, which reflects the scale of prediction deviations.

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

- R-squared Score (R^2) indicates how well predictions approximate actual values; higher is better.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y}_i)^2} \quad (3)$$

- ±10% Accuracy (Acc@10%) the proportion of predicted values falling within ±10% of the ground truth.

$$Accuracy_{10\%} = \frac{1}{n}\sum_{i=1}^{n} 1\left[\left|\frac{y_i-\hat{y}_i}{y_i}\right| \le 0.1\right] \quad (4)$$

All predictions are inverse-transformed from the normalized scale before metric computation.

3 Results

3.1 Model Training and Overall Performance

As illustrated in Figure 4, both validation and training loss values are tracked throughout 50 epochs to demonstrate the learning process.

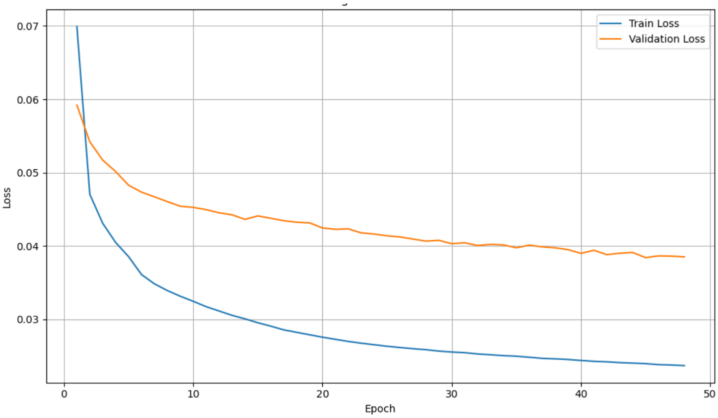

Fig. 4. Loss curves for training and validation throughout 50 epochs. (Picture credit: Original)

Training and validation losses both drop significantly throughout the first five epochs, indicating the model's quick early convergence. The training loss continues to decrease steadily throughout the training process, reaching approximately 0.138 by epoch 30. The validation loss stabilizes around 0.155 after epoch 10. The consistent trend between training and validation losses indicates good generalization capability without significant overfitting.

Table 2 summarizes model performance across 207 sensors. The model achieves an MAE of 1.96 mph and RMSE of 4.25 mph. R^2 reaches 0.8696, and 91.58% of predictions fall within a 10% relative error—demonstrating strong predictive capability for real-world applications

Table 2. Overall Model Performance

Metric	**Value**
MAE	1.96 mph
RMSE	4.25 mph
MSE	18.09
R^2	0.8696
Accuracy@10%	91.58%

3.2 Sensor-level Prediction Analysis

Sensor 12 is selected to illustrate the model's performance on individual sensors. As shown in Figure 5, the model accurately captures daily traffic patterns, including weekday peaks and smoother weekend flows.

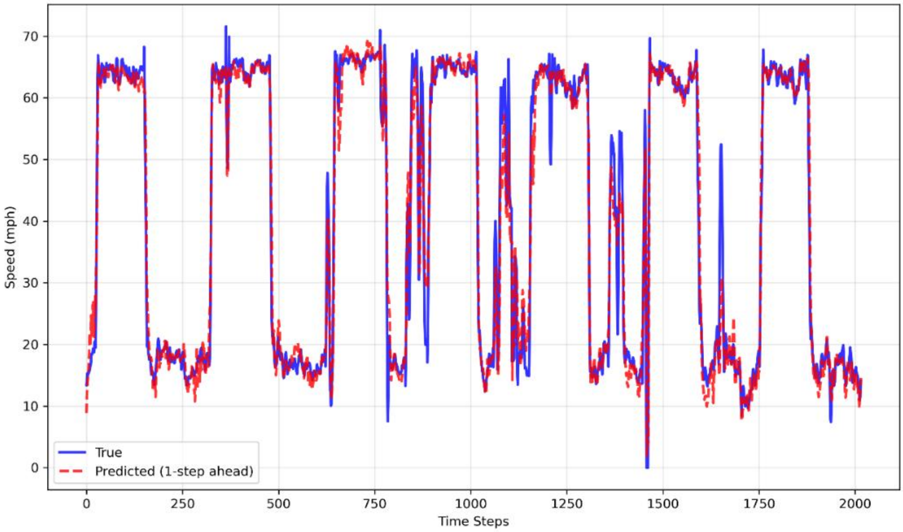

Fig. 5. Predicted vs. True speeds for sensor 12 (1-step ahead forecast) (Picture credit: Original)

Table 3 reports a high R^2 of 0.9577, with the mean prediction deviating only 1.35 mph from the true mean speed. Figure 6 and Figure 7 further validate this accuracy: the prediction scatter plot shows strong linear correlation, while the error distribution approximates a zero-centered normal curve. Around 80% of predictions fall within ±3 mph.

Table 3. Sensor 12 Performance Metrics

Metric	Value
MAE	3.17mph
RMSE	4.67mph
R^2	0.9577
Mean True Speed	41.55mph
Mean Predicted Speed	42.90mph

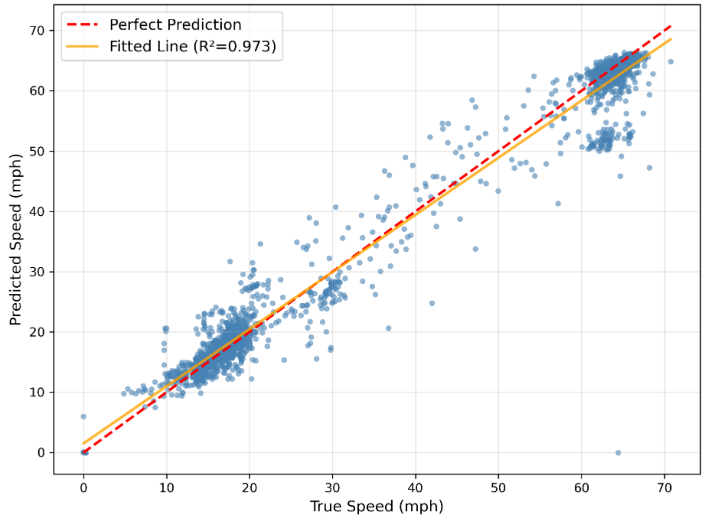

Fig. 6. Predicted vs. True speeds for sensor 12 (Picture credit: Original)

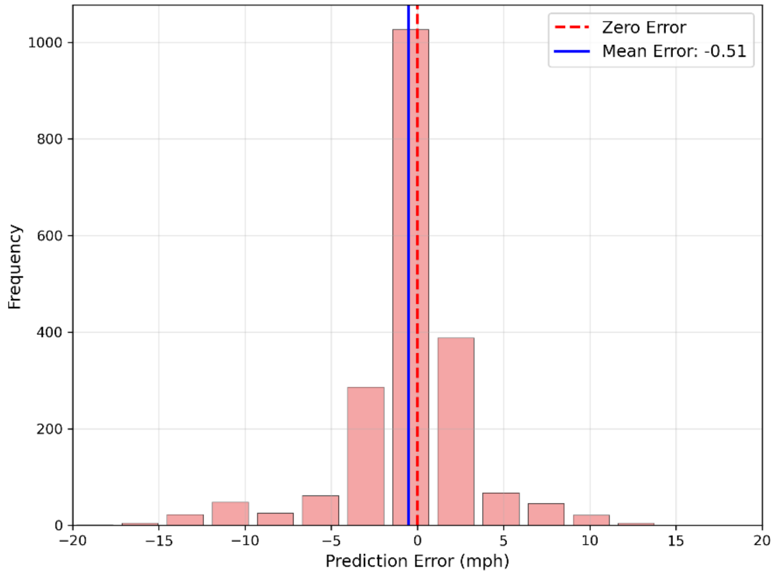

Fig. 7. Prediction error distribution for sensor 12 (Picture credit: Original)

3.3 Temporal Attention Analysis

Figure 8 visualizes attention weights assigned to past time steps. The model places highest weight on the most recent two steps (steps 11 and 12), aligning with the intuition that recent traffic states are most relevant for short-term prediction. Earlier steps still receive non-zero weights, indicating the model also captures longer-term trends.

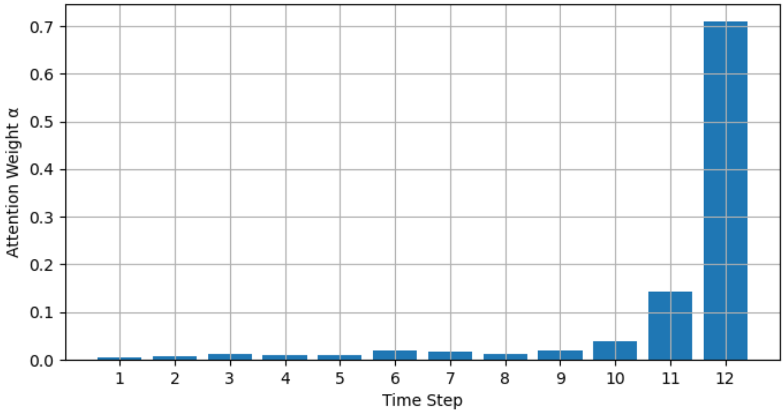

Fig. 8. Temporal attention distribution for one step (Picture credit: Original)

4 Discussion

4.1 Model Performance Analysis

The proposed GCN-BiLSTM-Attention model demonstrates strong predictive performance with an overall MAE of 1.96 mph and R^2 of 0.8696 across all 207 sensors. These findings show that the model well depicts the intricate spatiotemporal dynamics of traffic flow in cities. The ±10% accuracy of 91.58% suggests high reliability for practical traffic management applications, where predictions within acceptable error bounds are crucial for decision-making.

The superior performance on representative sensors like Sensor 12 ($R^2 = 0.9577$) can be attributed to several factors: the sensor exhibits highly regular temporal patterns with clear daily and weekly cycles, which align well with the BiLSTM-Attention mechanism's ability to capture long-term dependencies; the sensor's position in the graph topology provides rich spatial information for the GCN layers; and the data quality shows fewer anomalies, facilitating more effective learning.

4.2 Component Contribution Analysis

The three-component architecture demonstrates effective integration of spatial and temporal modelling. The GCN layers successfully extract spatial dependencies from the traffic network, as evidenced by the model's ability to capture coordinated traffic patterns across connected sensors. The bidirectional LSTM component effectively models temporal sequences, capturing both forward and backward temporal dependencies that are essential for understanding traffic flow dynamics.

The attention mechanism proves particularly valuable, as shown in the temporal attention analysis. The mechanism appropriately assigns higher weights to recent time steps (with the most recent step receiving ~70% attention weight), which aligns with traffic flow theory where immediate past conditions are most predictive of near-future states. This automatic weighting eliminates the need for manual feature engineering and enhances model interpretability.

4.3 Model Limitations and Challenges

Despite strong overall performance, the model exhibits varying accuracy across different sensors, highlighting the challenge of heterogeneous traffic patterns in urban networks. Some sensors show fewer regular patterns or lower data quality, leading to reduced prediction accuracy. This suggests that model performance is influenced by both the inherent characteristics of traffic flow at specific locations and the quality of sensor data. To address these challenges, future work could incorporate sensor-specific calibration techniques and dynamic weighting strategies to account for varying data reliability. Additionally, integrating external data sources such as weather conditions, road incidents, and special events may help capture more comprehensive traffic dynamics, thereby enhancing prediction accuracy in complex urban environments.

4.4 Practical Implications

The model is appropriate for real-world traffic control applications because to its high accuracy and dependability. The short-term prediction capability (5-15 minutes ahead) aligns well with adaptive traffic signal control and route guidance systems. The attention mechanism's interpretability provides traffic managers with insights into which historical periods most influence current predictions, supporting decision-making processes.

The efficient processing of large-scale traffic networks (207 sensors) demonstrates the model's scalability for metropolitan-level deployment. The robust performance across various traffic conditions suggests potential for integration into intelligent transportation systems for proactive traffic management and congestion mitigation.

5 Conclusion

In summary, this study introduces a GCN-BiLSTM-Attention model for short-term traffic flow prediction, integrating spatial dependencies, temporal patterns, and attention mechanisms. The experimental results on the METR-LA dataset demonstrate strong performance, achieving an overall MAE of 1.96 mph, R^2 of 0.8696, and ±10% accuracy of 91.58% across 207 traffic sensors.

The model appropriately emphasizes recent historical information, according to the attention mechanism analysis, with the most recent time step receiving approximately 70% attention weight, aligning with traffic flow theory and enhancing interpretability. The comprehensive data preprocessing strategy contributes significantly to the model's robust performance.

The proposed model demonstrates strong potential for intelligent transportation systems with its accurate predictions and computational efficiency. The results confirm the effectiveness of combining graph neural networks with sequential modeling and attention mechanisms for traffic prediction, contributing to data-driven traffic management solutions. In future work, the model could be extended by incorporating real-time external data sources such as weather conditions, road incidents, and event schedules to further improve prediction accuracy in dynamic environments. Additionally, exploring adaptive learning strategies and transfer learning could enable the model to generalize across different cities or road networks, supporting broader deployment in intelligent transportation systems.

References

1. Guo, S., Lin, Y., Feng, N., Song, C., Wan, H.: Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In: AAAI Conference on Artificial Intelligence, vol. 33, pp. 922–929 (2019).
2. Xu, D., Wang, Y., Jia, L., Qin, Y.: Real-time road traffic state prediction based on ARIMA and Kalman filter. Front. Inf. Technol. Electron. Eng. 18(3), 373–385 (2017).
3. Lee, K., Eo, M., Jung, E., Yoon, Y., Rhee, W.: Short-term traffic prediction with deep neural networks: A survey. IEEE Access 9, 54739–54776 (2021).
4. Han, X., Gong, S.: LST-GCN: Long short-term memory embedded graph convolution network for traffic flow forecasting. Electronics 11(14), 2230 (2022).
5. Li, Z., Yu, H., Shahabi, C., Liu, Y.: A hybrid deep learning approach with GCN and LSTM for traffic flow prediction. In: IEEE Intelligent Transportation Systems Conference, pp. 1929–1934 (2019).
6. Ghosh, A., Giacobbe, M., Rafiq, M.T., Puliafito, A., Giorgi, R.: Enhancing traffic prediction with spatio-temporal deep learning: A GCN-LSTM hybrid model. In: Mediterranean Conference on Embedded Computing (2025).
7. Zhao, L., Song, Y., Zhang, C., Liu, Y., Wang, P.: T-GCN: A temporal graph convolutional network for traffic prediction. IEEE Trans. Intell. Transp. Syst. 21(9), 3848–3858 (2020).
8. Cui, Z., Henrickson, K., Ke, R., Wang, Y.: Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting. IEEE Trans. Intell. Transp. Syst. 21(11), 4883–4894 (2020).
9. Zhang, H., Yang, G., Yu, H., Zheng, Z.: Kalman filter-based CNN-BiLSTM-ATT model for traffic flow prediction. Comput. Mater. Contin. 76(1), 1047–1063 (2023). https://doi.org/10.32604/cmc.2023.039274
10. Liu, H., Zhu, C., Zhang, D., Li, Q.: Attention-based spatial-temporal graph convolutional recurrent networks for traffic forecasting. arXiv:2302.12973 (2023).