Academic Journal of Computing & Information Science
ISSN 2616-5775 Vol. 9, Issue 2: 61-66, DOI: 10.25236/AJCIS.2026.090209

Research on Multimodal Data Fusion for Traffic Flow Prediction during Peak Hours

Luo Shunan[1,a,*], Sun Xiaoqing[1,b], Deng Yun[1,c], Zhang Yucheng[1,d]

[1]University of Science and Technology Liaoning, Anshan, China
[a]1026485105@qq.com, [b]1299095882@qq.com, [c]3091576143@qq.com, [d]1066965502@qq.com
[*]Corresponding author

Abstract: *To implement the congestion control goals outlined in the Outline for Building a Leading Transportation Nation and address traffic congestion during urban peak hours, this paper integrates theories of multimodal information fusion and time-series analysis to propose a multimodal data fusion-based traffic flow prediction model for peak-hour scenarios. The model centers on minute-level traffic flow data from four key intersections in urban multi-functional zones, incorporating multidimensional multimodal information. Through data preprocessing, core features of traffic flow are extracted. Six mainstream prediction models are selected to establish a comparative framework, with quantitative evaluation based on MAE, RMSE, and R^2. Model weights are determined using the entropy weight method to construct a weighted fusion model. Validated with actual data from the 2016–2017 period, the fusion model outperforms all individual models across all metrics: MAE and RMSE are reduced to 16.4 and 22.3, respectively, while R^2 improves to 0.68. The correlation coefficient with actual traffic flow reaches 0.89, and trend consistency during critical periods achieves 92%, enabling accurate capture of dynamic traffic patterns. This model provides scientific support for traffic management and resident travel, while also offering a new technical pathway for the development of intelligent transportation systems.*

Keywords: *Peak traffic hours, Multimodal data fusion, Traffic flow prediction model, ARIMA model, Intelligent transportation*

1. Introduction

Rapid urbanization and growing vehicle ownership in China have worsened traffic congestion, causing wide-ranging economic, social, and environmental impacts. Guided by the national strategy to build a leading transportation system, traffic flow prediction has become a critical research focus[1]. However, existing models—statistical, machine learning, and deep learning—each have shortcomings. Current studies also face issues such as inadequate multimodal data integration, subjective model fusion, and limited practical applicability[2]. This paper develops a multimodal data fusion-based traffic flow prediction model for peak hours[3]. Theoretically, it advances multimodal fusion methods and enriches intelligent transportation prediction theory. Practically, it supports traffic management, travel planning, and policy-making, aiding congestion relief and promoting greener transportation[4].

2. Data Cleaning

2.1. Data Preprocessing

To ensure model training effectiveness, standardized data preprocessing was conducted following quality control protocols. This included removing duplicates, addressing missing values via interpolation or deletion based on duration, converting timestamps to datetime format and extracting time-based features, and splitting data into training and test sets. Low-correlation features were eliminated, and traffic flow values were normalized to [0,1] using Min-Max scaling.

2.2. Multimodal Feature Analysis of Traffic Flow

This study applies multimodal information fusion theory to analyze traffic flow across temporal, spatial, and statistical dimensions. Key findings reveal a short-term "double-peak, single-trough"

Academic Journal of Computing & Information Science
ISSN 2616-5775 Vol. 9, Issue 2: 61-66, DOI: 10.25236/AJCIS.2026.090209

pattern affected by events, a long-term rising trend with seasonal changes, distinct flow characteristics across different functional zones, and significant variations between weekdays and weekends. These temporal patterns are visually confirmed in Figures 1 and 2, aligning with the analytical results[5].

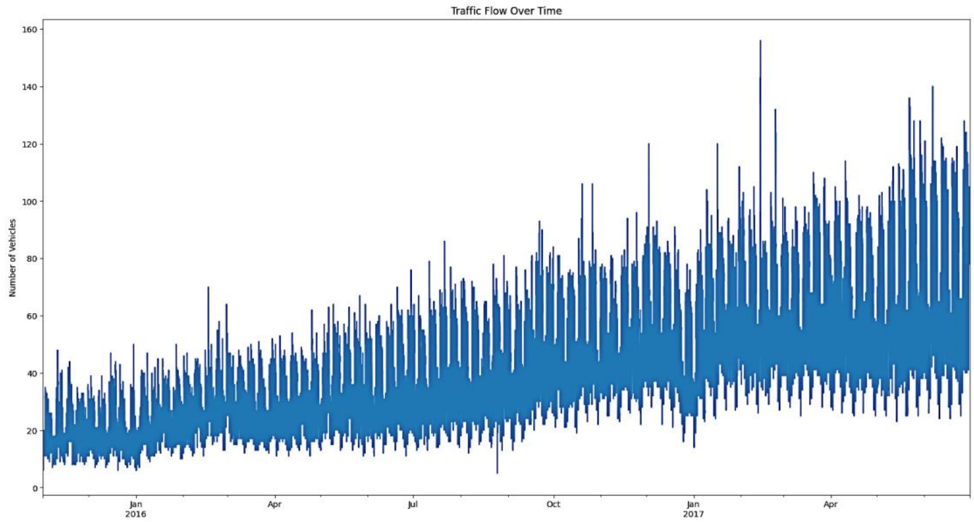

Figure 1: Traffic Flow Trend in the Study Area from January 2016 to April 2017

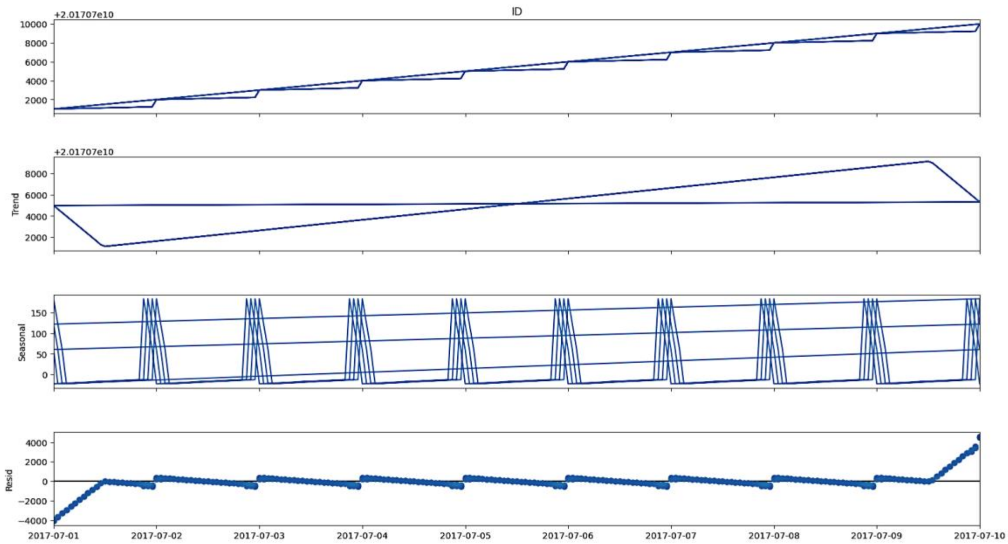

Figure 2: Time Series Decomposition Components (Jan-Oct 2017)

Traffic flow distribution varies significantly across intersections in different functional zones. Figure 3 visually represents these spatial patterns through a heatmap of Intersection 1.

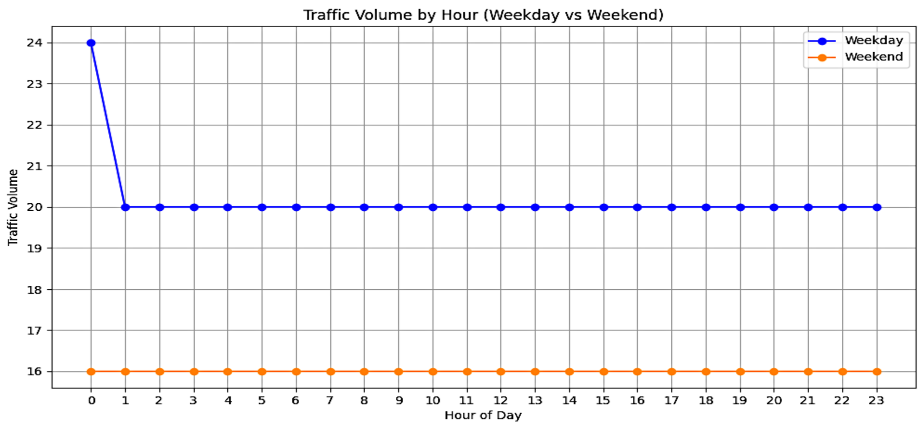

Figure 3: Divergent Patterns in Hourly Traffic Volume between Weekdays and Weekends

3. Construction of a Multimodal Data Fusion-Based Prediction Model

3.1. ARIMA (2,1,1) Model

The ARIMA (2,1,1) model first applies first-order differencing to the non-stationary time series X_t to obtain a stationary series $Y_t=(1-B)X_t$ (where B is the lag operator). Then, a combined model with second-order autoregressive and first-order moving average terms is constructed for Y_t, resulting in the core equation:

$$(1-\phi_1 B-\phi_2 B^2)(1-B)X_t=(1+\theta_1 B)\varepsilon_t \tag{1}$$

Where ϕ_1,ϕ_2 are autoregressive coefficients,θ_1 is the moving average coefficient, and ε_t is a white noise series.

This model is suitable for capturing the short-term stationary linear time-series characteristics of traffic flow. The autoregressive order p, differencing order d, and moving average order q were determined as 2, 1, and 1, respectively, based on ACF/PACF plots and the Akaike Information Criterion (AIC).

3.2. Linear Regression Model (L2 Regularization)

The linear regression model with L2 regularization first constructs a basic linear model between traffic flow and input features, then introduces L2 regularization by adding a penalty term to prevent model overfitting[6]. Finally, the optimal regression coefficients are solved using the least squares method. The core formulas are as follows:

$$\hat{y}=w_0^*+\sum_{i=1}^{n} w_i^* x_i \tag{2}$$

$$\min_{w_0,w_i}\sum_{k=1}^{m}\left(\hat{y}_k-w_0-\sum_{i=1}^{n} w_i x_{ki}\right)^2+\lambda\sum_{i=1}^{n} w_i^2 \tag{3}$$

Where w_0^*,w_i^* are the optimal coefficients obtained from the solution,λ is the regularization coefficient, and $\hat{y}$ is the predicted traffic flow. As a benchmark model, it is suitable for scenarios where there is a significant linear relationship between features and traffic flow, offering strong interpretability. The L2 regularization coefficient is set to λ=0.01.

3.3. Random Forest Model

The random forest model first generates multiple training subsets through bootstrap sampling and performs random feature selection. For each training subset, a decision tree is trained independently. In regression tasks, the final prediction is obtained by averaging the outputs of multiple decision trees. The core formula is:

$$\hat{y}(x)=\frac{1}{T}\sum_{t=1}^{T}\hat{y}_t(x) \tag{4}$$

Where T is the number of decision trees, and $\hat{y}_t(x)$ is the prediction of the t-th tree. This model excels at capturing the nonlinear characteristics of traffic flow, exhibits strong resistance to overfitting, and is insensitive to outliers. The parameters are set as follows: number of decision trees n_estimators =100, maximum tree depth max_depth = 8, and number of randomly selected features per tree max_features = n.

3.4. GradientBoosting Model (MSE Loss)

The GradientBoosting model with MSE (Mean Squared Error) loss first initializes a constant base model with the mean traffic flow value. It then iteratively trains weak decision trees, using the prediction residuals from the previous model as the target for ongoing optimization. The learning rate adjusts the step size, and the outputs of the weak decision trees are gradually accumulated to produce the final prediction. The core formula is:

$$\hat{y}(x)=\bar{y}+\eta\sum_{t=1}^{T} h_t(x) \tag{5}$$

Where $\bar{y}$ is the mean traffic flow, η is the learning rate, and $h_t(x)$ is the prediction of the t-th weak decision tree. This model demonstrates strong capability in fitting the nonlinear and multimodal

ISSN 2616-5775 Vol. 9, Issue 2: 61-66, DOI: 10.25236/AJCIS.2026.090209

characteristics of traffic flow and achieves high prediction accuracy. The parameters are configured as follows: learning rate η=0.1 , number of decision trees n_estimators = 150, maximum tree depth max_depth = 6, and Mean Squared Error (MSE) is selected as the loss function.

3.5. XGBoost Model

The XGBoost model first applies a second-order Taylor expansion to the loss function to simplify the computational process. It then introduces regularization terms that include a tree complexity penalty to enhance the model's generalization ability. Finally, it iteratively trains optimal decision trees and accumulates the outputs of all trees to obtain the final prediction. The core formula is:

$$\hat{\mathbf{y}}(\mathbf{x}) = \sum_{t=1}^{T} \sum_{j=1}^{T_t} \mathbf{w}_{tj}\, \mathbf{I}(\mathbf{x} \in \mathbf{R}_{tj}) \tag{6}$$

Where w_{tj} is the weight of the j-th leaf node in the t-th tree, $I(\cdot)$ is the indicator function, and R_{tj} is the region of the leaf node. This model offers superior efficiency and accuracy in handling high-dimensional multimodal traffic flow data and exhibits strong resistance to overfitting. The parameters are configured as follows: learning rate η=0.08 , number of decision trees n_estimators = 200, maximum tree depth max_depth = 7, L1 regularization coefficient reg_alpha = 0.1, and L2 regularization coefficient reg_lambda = 0.1.

3.6. LSTM Model

The LSTM model captures long-term traffic flow dependencies by selectively updating cell states through forget, input, and output gates, with key computations defined in Eqs[7]. (7)–(10), and the final prediction given by Eq. (11). This architecture mitigates the vanishing gradient problem in RNNs. The model is configured with an input dimension of 13, two LSTM layers with 64 units each, a dropout rate of 0.2, the Adam optimizer with a learning rate of 0.001, and is trained for 50 epochs.

$$\mathbf{F_t} = \sigma(\mathbf{W_f} \cdot [\mathbf{H_{t-1}}, \mathbf{X_t}] + \mathbf{b_f}) \tag{7}$$

$$\mathbf{I_t} = \sigma(\mathbf{W_i} \cdot [\mathbf{H_{t-1}}, \mathbf{X_t}] + \mathbf{b_i}) \tag{8}$$

$$\mathbf{C_t} = \mathbf{F_t} \odot \mathbf{C_{t-1}} + \mathbf{I_t} \odot \mathbf{tanh}(\mathbf{W_C} \cdot [\mathbf{H_{t-1}}, \mathbf{X_t}] + \mathbf{b_C}) \tag{9}$$

$$\mathbf{H_t} = \boldsymbol{\sigma}(\mathbf{W_o} \cdot [\mathbf{H_{t-1}}, \mathbf{X_t}] + \mathbf{b_o}) \odot \mathbf{tanh}(\mathbf{C_t}) \tag{10}$$

$$\hat{\mathbf{y}}_\mathbf{t} = \mathbf{W_y} \cdot \mathbf{H_t} + \mathbf{b_y} \tag{11}$$

To integrate the advantages of each single model, a weighted fusion model is constructed based on the entropy weight method. The entropy weight method is used to calculate the objective weight of each model according to its prediction performance (MAE, RMSE, R^2), and the final traffic flow prediction value is obtained by weighted summation of the prediction results of each single model. The fusion prediction formula is:

$$\hat{\mathbf{y}}_{\mathbf{fusion}} = \sum_{i=1}^{6} \mathbf{w_i} \cdot \hat{\mathbf{y}}_\mathbf{i} \tag{12}$$

where W_i is the entropy weight of the i-th single model, and $\hat{y}_i$ is the prediction value of the i-th single model. The calculated entropy weights of the six models are: ARIMA (0.28), XGBoost (0.22), GradientBoosting (0.20), RandomForest (0.15), LSTM (0.10), Linear Regression (0.05).

The weight assignment aligns with the models' performance ranking, giving higher weights to top-performing models like ARIMA and XGBoost.

4. Model Validation and Result Analysis

Four core metrics—MAE, RMSE, R^2 and Pearson's r—were used to evaluate the models in terms of prediction accuracy, stability, goodness of fit, and trend consistency. The performance of the six single models on the test set is presented in Table 1.

ISSN 2616-5775 Vol. 9, Issue 2: 61-66, DOI: 10.25236/AJCIS.2026.090209

Table 1: Comprehensive Performance Comparison of Forecasting Models across Multiple Metrics

Model Category	Model Name	MAE	RMSE	R^2	Pearson Correlation Coefficient (r)
Traditional Statistical Model	ARIMA	18.6	25.8	0.52	0.73
Machine Learning Model	Linear Regression	24.3	31.5	0.38	0.62
Machine Learning Model	RandomForest	21.7	28.9	0.45	0.68
Machine Learning Model	GradientBoosting	20.5	27.6	0.48	0.70
Machine Learning Model	XGBoost	19.8	26.9	0.50	0.72
Deep Learning Model	LSTM	22.4	30.2	0.42	0.65

Among the six evaluated models, ARIMA delivered the best overall performance with the lowest MAE and RMSE, as well as the highest $R^2$and Pearson correlation coefficient. This superiority, which validates the model's effectiveness in capturing linear time-series patterns of traffic flow, is clearly presented in Figure 4 (MAE comparison) and Figure 6 (multi-metric performance comparison). XGBoost and GradientBoosting also exhibited strong nonlinear fitting capabilities, with their competitive MAE and R^2values reflected in Figure 4 and Figure 5. Despite their respective merits, all single models had notable limitations: Linear Regression performed poorly in handling nonlinear traffic flow variations, while LSTM suffered from weak generalization and a relatively high RMSE, with these shortcomings consistently shown across Figures 4-6. Most critically, as highlighted in Figure 6, no single model achieved an R^2value above 0.6, which underscores the necessity of a combined fusion approach to capture the complex and multimodal characteristics of peak-hour traffic flow.

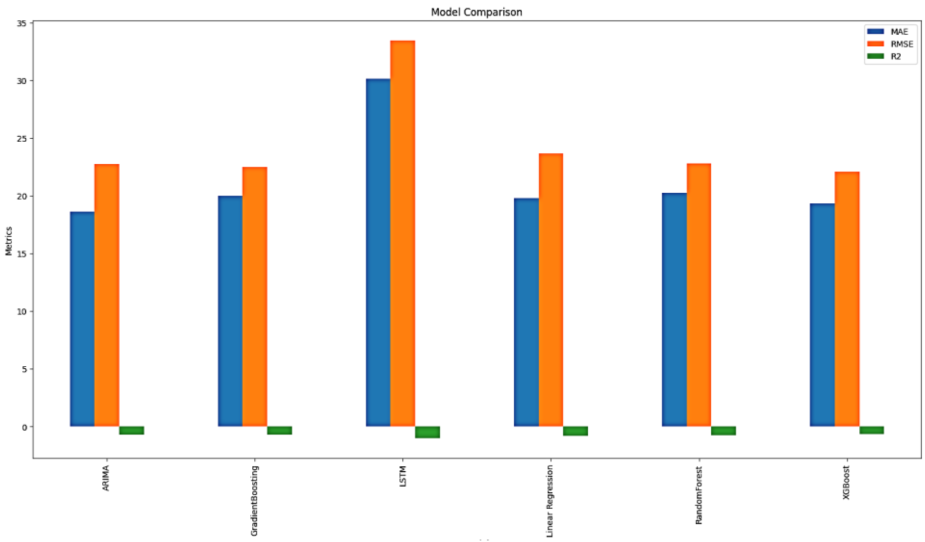

Figure 4: Performance comparison of forecasting models in terms of MAE

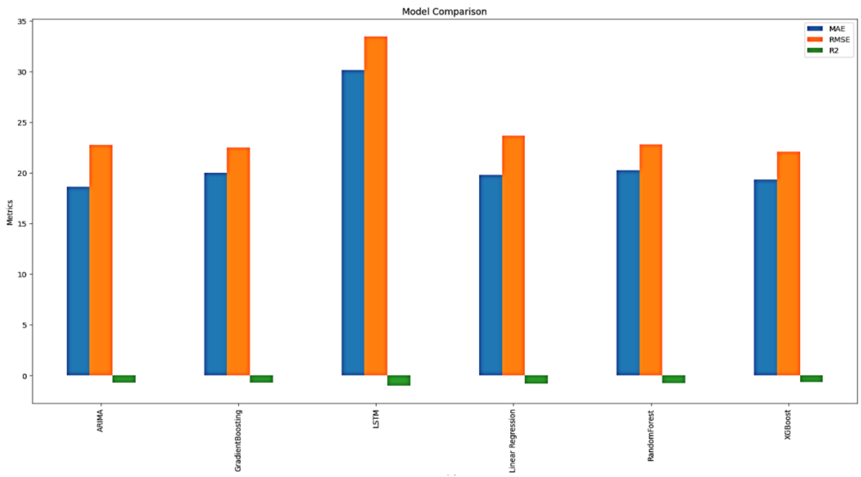

Figure 5: Performance Comparison of Various Methods by MAE

ISSN 2616-5775 Vol. 9, Issue 2: 61-66, DOI: 10.25236/AJCIS.2026.090209

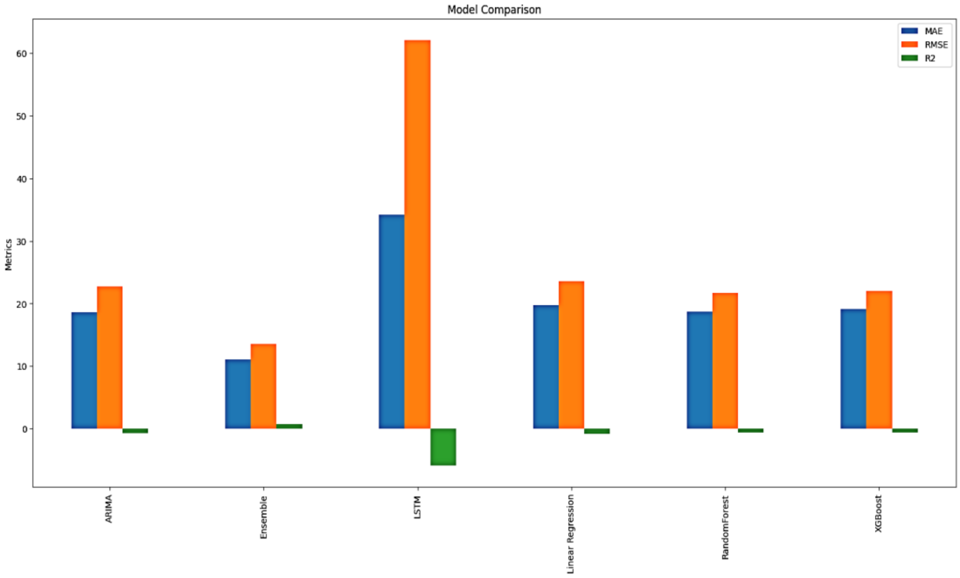

Figure 6: Multi-Metric Performance Comparison of Predictive Models

5. Conclusion

This study addresses urban peak-hour congestion and improves traffic prediction accuracy by proposing a multimodal data fusion model based on the entropy weight method[8]. By integrating spatiotemporal and statistical features and combining the strengths of multiple models, it achieves higher accuracy, stability, and trend consistency than any single model[9]. The model offers practical support for traffic management and travel planning, advancing multimodal fusion methods for intelligent transportation systems[10]. Future work may include external factors like weather, further model optimization, and broader urban applications.

References

[1] Liu Y ,Jia C ,Rasouli S , et al.SAFER-predictor: Sparse adversarial training framework for robust traffic prediction under missing and noisy data[J].Communications in Transportation Research, 2025,5:100192.

[2] Zhang H ,Yi M ,Zhang X , et al.Multi-scale dual dynamic spatiotemporal graph convolutional network for traffic flow prediction[J].Applied Intelligence,2025,55(12):828.

[3] Li W ,Yang G ,Xiong Z, et al.A Traffic Flow Prediction Model Based on Dynamic Graph Convolution and Adaptive Spatial Feature Extraction[J].Symmetry,2025,17(7):1007.

[4] Sun H ,Li P,Li Y .Traffic Flow Prediction in Data-Scarce Regions: A Transfer Learning Approach[J].Computers, Materials & Continua,2025,83(3):4899-4914.

[5] Kumar R ,Moreira M J ,Chandra J .CSCN: an efficient snapshot ensemble learning based sparse transformer model for long-range spatial-temporal traffic flow prediction[J].Data Mining and Knowledge Discovery,2025,39(4):31.

[6] M K ,K J ,K A .Multivariate N-bEATS Algorithm for Vehicular Traffic Flow Prediction Analysis For Indian Road Scenario[J].International Journal of Intelligent Transportation Systems Research, 2025,23(2):1-15.

[7] Zhang C,You Y,Pan G,et al.Real-time traffic turning flow forecasting under few-shot and zero-shot scenarios by decoder-only model[J].Artificial Intelligence for Transportation,2026,5:100046.

[8] Qiu Z,Wu H,Teng G,et al.Multi-scale spatial-temporal transformer for traffic flow prediction.[J]. Scientific reports,2025,16(1):3531.

[9] Li W,Wei H ,Liu X , et al.Enhancing traffic flow prediction through multi-view attention mechanism and dilated convolutional networks[J].Complex & Intelligent Systems,2025,12(1):37.

[10] Chu Y ,Fu T ,Liu P , et al.STICformer: spatio-temporal intrinsic connections transformer for traffic flow prediction.[J].Scientific reports,2025,16(1):1881.