

# Intelligent Traffic Big Data Mining and Prediction Model Considering Spatiotemporal Characteristics of Traffic Flow

Shukun Zhou\*

Shandong Hi-speed Construction Management Group Co., Ltd, Jinan, 250014, Shandong, China

\*Corresponding author

**Abstract:** Intelligent transportation is the core solution to alleviating urban traffic congestion and enhancing traffic governance efficiency. Accurately capturing the spatiotemporal coupling characteristics of traffic flow is key to achieving precise traffic flow prediction. Addressing current issues such as the neglect of spatiotemporal correlations in traffic flow prediction models, low efficiency in intelligent transportation big data mining, insufficient prediction accuracy, and weak generalization capabilities, this study focuses on the core spatiotemporal characteristics of traffic flow—time periodicity, volatility, spatial correlation, and aggregation—while incorporating the "5V" features of multi-source heterogeneous big data in intelligent transportation. It conducts research on big data mining and prediction models. First, it identifies and quantifies traffic flow spatiotemporal characteristics to establish a targeted quantitative indicator system. Second, it designs a full-process big data mining framework to complete multi-source data preprocessing and spatiotemporal correlation feature extraction. Building on this, it improves and optimizes the model by introducing a spatiotemporal attention mechanism based on LSTM and GCN algorithms, constructing a traffic flow prediction model that integrates spatiotemporal characteristics. Finally, the model's performance is validated through case studies. The results demonstrate that the proposed model significantly outperforms traditional models, with a minimum MAPE of 4.87%, a single-prediction time of 0.32 seconds, and strong real-time adaptability and scenario compatibility. This study enriches the theory of spatiotemporal analysis and big data integration in traffic flow, providing scientific decision-making support for intelligent transportation management and road network optimization, with important theoretical and practical application value.

**Keywords:** intelligent transportation; traffic flow; spatiotemporal characteristics; big data mining; spatiotemporal attention mechanism

## 1. Introduction

Accelerating the development of a strong transportation system is a national strategic priority. As the core vehicle for high-quality transportation development, smart transportation's digital and intelligent transformation has become a key pathway to alleviating urban traffic congestion and enhancing traffic governance efficiency [1]. With the deep integration of technologies such as the Internet of Things and big data, vast amounts of multi-source data—including fixed monitoring devices, floating cars, and mobile signaling—are accumulated, providing rich data support for traffic flow analysis. However, this also brings governance challenges such as data heterogeneity and low value density. Traffic flow exhibits significant spatiotemporal coupling characteristics, displaying daily and weekly periodicity along with sudden fluctuations over time, while spatially manifesting as road segment correlation, regional clustering, and spatiotemporal heterogeneity. These core features determine that traffic flow prediction must break through the limitations of traditional single-dimensional analysis.

Current research on traffic flow prediction has introduced technologies such as deep learning and graph neural networks, yet it still suffers from notable shortcomings: most models fail to deeply explore the spatiotemporal characteristics of traffic flow, remaining at a simplistic aggregation level and struggling to capture multi-scale spatiotemporal correlations; the integration efficiency of multi-source heterogeneous traffic big data is low, with insufficient accuracy in extracting spatiotemporal correlation features; models exhibit weak generalization capabilities in adapting to dynamic traffic flow changes and handling extreme scenarios, making it difficult to meet the practical demands of real-time smart traffic management[2,3].

To this end, this paper focuses on the core of traffic flow spatiotemporal characteristics, conducting research on intelligent transportation big data mining and predictive models. First, it systematically examines the intrinsic patterns of traffic flow spatiotemporal characteristics and the structural features of intelligent transportation big data, designing targeted data preprocessing and feature mining methods. Building on this foundation, it constructs a traffic flow predictive model integrating spatiotemporal characteristics, validating the model's effectiveness and superiority through practical examples. The research outcomes can enrich the theoretical framework of traffic flow spatiotemporal analysis and big data integration, enhance the accuracy and timeliness of traffic flow prediction, and provide scientific support for optimizing traffic networks, congestion warnings, and the implementation of intelligent transportation systems. This contributes to the transition from "experience-based decision-making" to "data-driven governance" in traffic management.

## **2. Core theoretical foundation and analysis of spatiotemporal characteristics of traffic flow**

This chapter systematically elaborates on the core theoretical support required for the research, focusing on the identification and quantitative analysis of the spatiotemporal characteristics of traffic flow, clarifying the spatiotemporal evolution laws and quantitative representation methods of traffic flow, providing solid theoretical and data support for the subsequent construction of intelligent transportation big data mining and prediction models, and ensuring the scientific and rational nature of the research.

### **2.1 Core theoretical basis**

The theory of intelligent transportation is the core support of this study, which relies on new generation information technologies such as the Internet of Things, big data, and artificial intelligence to achieve real-time perception, accurate analysis, intelligent decision-making, and efficient control of the transportation system[4]. The core connotation is to improve the operational efficiency and service quality of the transportation system through data-driven approaches, among which big data technology and artificial intelligence technology are key supports for achieving traffic flow analysis and prediction. Smart transportation big data has the "5V" characteristics of massive, diverse, real-time, correlated, and low value density. Its core value lies in extracting traffic flow operation rules through multi-source data fusion and mining.

The theories of big data mining and traffic flow prediction are the core methods supporting research. The core technologies of big data mining include data preprocessing, feature extraction, data fusion, etc[5]. Among them, preprocessing technology is used to solve problems such as noise, missing, and redundancy in traffic data, and feature extraction technology is used to mine the spatiotemporal correlation laws behind the data; The theory of traffic flow prediction covers both traditional time series models and modern intelligent prediction models. Traditional models such as ARIMA are suitable for predicting smooth traffic flow, while modern intelligent models such as LSTM and GCN can better capture the nonlinear spatiotemporal correlation characteristics of traffic flow, providing algorithmic references for model construction.

### **2.2 Identification and quantitative analysis of spatiotemporal characteristics of traffic flow**

The spatiotemporal characteristics of traffic flow are the core features of traffic flow operation, and their identification and quantification need to be based on high-quality smart transportation data[6]. Fixed monitoring data and floating vehicle trajectory data of urban core road networks are selected as research objects, and data integrity and accuracy are ensured through preprocessing steps such as data cleaning, deduplication, standardization, and missing value filling. In terms of time characteristics, time series analysis and statistical analysis methods are used to identify significant daily, weekly, and holiday cycles in traffic flow, as well as fluctuations affected by sudden traffic events and weather changes[7]. Quantitative indicators such as cycle intensity and fluctuation coefficient are constructed to accurately describe the evolution law of time characteristics.

In terms of spatial characteristics, based on spatial autocorrelation analysis and clustering analysis methods, the upstream and downstream correlations between road segments, regional traffic flow aggregation, and spatiotemporal heterogeneity are identified, and quantitative indicators such as spatial correlation and aggregation are constructed [8]. Combined with the road network topology matrix, accurate quantification of spatial characteristics is achieved. On this basis, a coupling coordination degree model is adopted to analyze the coupling relationship between the temporal and spatial characteristics of

traffic flow, quantify the degree of spatiotemporal coupling, clarify the mechanism of mutual influence between spatiotemporal characteristics, and provide quantitative basis for subsequent spatiotemporal correlation feature mining and prediction model construction.

### **3. Construction of intelligent transportation big data mining and prediction models**

Based on the recognition and quantitative analysis of the spatiotemporal characteristics of traffic flow in the previous section, this chapter focuses on the deep mining and prediction model construction of smart transportation big data. Combining the spatiotemporal coupling characteristics of traffic flow, a targeted big data mining method is designed to extract high-quality spatiotemporal correlation features, and then construct a traffic flow prediction model that integrates spatiotemporal characteristics. This achieves a complete closed-loop from data value mining to accurate prediction, providing technical support for smart transportation control.

#### ***3.1 Intelligent transportation big data mining considering spatiotemporal characteristics***

Combining the spatiotemporal characteristics of traffic flow with the heterogeneity of intelligent transportation big data, a full process big data mining framework is constructed, which includes data input, preprocessing, feature extraction, anomaly mining, and result verification. The core goal is to extract features that can accurately reflect the spatiotemporal correlation laws of traffic flow and provide high-quality input for prediction models. In the data input stage, this study integrates multi-source heterogeneous data, including fixed monitoring data, floating vehicle trajectory data, and environmental meteorological data. It clarifies the collection frequency, coverage scope and dimensional attributes of each data type to ensure data comprehensiveness.

In the data preprocessing stage, to address the issues of noise, missing values, and redundancy in transportation big data, a missing value filling method based on spatiotemporal correlation is adopted, combined with the Isolation Forest algorithm to remove abnormal data. Normalization is used to eliminate dimensional effects and improve data quality. Feature extraction is the core of mining. Based on the spatiotemporal characteristics of traffic flow, sliding window analysis is used to extract temporal correlation features and capture the periodicity and volatility of traffic flow at different time periods; Using spatial autocorrelation analysis and graph structure modeling, extract spatial correlation features and regional clustering features between road segments; By combining the characteristics of spatiotemporal coupling, a coupling feature extraction algorithm is constructed to quantify the mutual influence of spatiotemporal characteristics and form a complete set of spatiotemporal correlation features. At the same time, the association rule mining algorithm is used to mine abnormal traffic flow data, providing support for traffic congestion warning. Finally, the validity of the mining results is verified through rationality testing.

#### ***3.2 Construction of traffic flow prediction model integrating spatiotemporal characteristics***

With the goal of accurately predicting the operation status of traffic flow, following the principles of pertinence, efficiency, and scalability, a hybrid prediction model that integrates the spatiotemporal characteristics of traffic flow is constructed, taking into account both prediction accuracy and real-time performance. The overall architecture of the model is divided into input layer, feature fusion layer, prediction layer, and output layer. The input layer receives preprocessed raw data and mined spatiotemporal correlation features; The feature fusion layer introduces a spatiotemporal attention mechanism to weight and fuse temporal features, spatial features, and spatiotemporal coupling features, highlighting the influence of key spatiotemporal features and solving the problem of unreasonable weight allocation in traditional feature fusion.

The prediction layer is based on LSTM and GCN algorithms. LSTM is used to capture time series correlations of traffic flow, while GCN is used to mine spatial correlations under road network topology. The algorithm is improved by introducing time period encoding and spatial attention mechanism to enhance the model's ability to capture spatiotemporal coupling characteristics. During the model training phase, the dataset is divided into training, validation, and testing sets in a ratio of 7:2:1. The Adam optimization algorithm is used to adjust the model parameters, and key parameters such as learning rate and iteration times are optimized using cross validation and grid search methods to avoid overfitting and underfitting of the model. The model output layer outputs predicted traffic flow values (traffic volume, vehicle speed) for different time periods in the future. Through model training and optimization, accurate

prediction of traffic flow is achieved, laying the foundation for subsequent instance verification while ensuring that the model can adapt to dynamic changes in traffic flow and meet the practical needs of real-time intelligent traffic control.

#### 4. Example verification and analysis

To verify the effectiveness, superiority, and practicality of the intelligent transportation big data mining method considering the spatiotemporal characteristics of traffic flow and the traffic flow prediction model integrating spatiotemporal characteristics proposed earlier, this chapter selects a typical urban core road network as the case study area and conducts case verification through real intelligent transportation big data. Based on actual traffic flow operation scenarios, this study designs a scientifically reasonable verification scheme to compare and analyze the model's prediction performance. It clarifies the model's adaptability under diverse traffic scenarios, provides data support and a practical foundation for the model's practical application, and ensures that the research results possess reliable engineering application value. This instance verification strictly follows the logical process of "data preparation scheme design result analysis conclusion summary", comprehensively testing the feasibility of big data mining methods and the comprehensive performance of prediction models.

##### 4.1 Example background and data preparation

The case study area selects the core road network of a provincial capital city in eastern China, which covers multiple scenarios such as commercial centers, transportation hubs, and residential areas. The road network density is high, and the traffic flow fluctuates significantly. It has typical spatiotemporal characteristics of traffic flow and can effectively verify the universality and applicability of the model. The total length of the road network in this area is about 32km, including 18 intersections and 26 main road sections, including 3 expressways, 8 main roads, and 15 secondary roads. The road network topology is complex, and traffic congestion is prominent during peak hours. It is highly compatible with the current traffic operation status of the city's core road network and is suitable as an example verification area.

The data is sourced from the intelligent transportation monitoring system in the region, with a collection period of 30 consecutive days and a collection frequency of 5 minutes per time, covering multi-source heterogeneous data. The data overview and preprocessing results are shown in Table 1 below:

*Table 1. Smart transportation monitoring system*

Data type	Data content	Data volume	Preprocessing operation	Processing result
Basic traffic flow data	Traffic volume, speed, and road occupancy rate	17280 records	Missing value filling, outlier removal, standardization	Fill accuracy 98.2%, remove 1276 outliers
Floating car trajectory data	More than 800 taxis and ride hailing services are located and in operation	Over 286000 items	Outlier removal and standardization	Data integrity reaches 99.59%
Environmental meteorological data	Temperature, rainfall, wind speed and other parameters	8640 records	Standardization	Mapping data to the [0,1] interval

##### 4.2 Verification plan design

The core objective of this instance validation is to test the effectiveness of big data mining methods and the prediction accuracy, real-time performance, and generalization ability of prediction models. To this end, a scientifically sound validation plan is designed, with clear validation indicators, comparative models, and validation processes. The validation indicators selected are three commonly used core accuracy indicators in the field of traffic flow prediction, namely mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE). MAE is used to measure the average

deviation between predicted values and true values, RMSE is used to amplify the impact of larger errors, and MAPE is used to intuitively reflect the relative degree of prediction errors. The smaller the values of the three indicators, the higher the prediction accuracy of the model. At the same time, real-time indicators such as model training time and single prediction time are added to verify whether the model meets the requirements of real-time control of intelligent transportation.

To highlight the superiority of the fusion spatiotemporal characteristic prediction model proposed in this article, four commonly used traffic flow prediction models were selected as comparison models, namely: traditional ARIMA time series model (considering only temporal characteristics), single GCN model (considering only spatial characteristics), basic LSTM model (simply capturing temporal correlations), and LSTM-GCN hybrid model (not introducing spatiotemporal attention mechanism and not deeply integrating spatiotemporal coupling characteristics). The validation process is divided into four steps: the first step is to divide the preprocessed dataset into a training set, a validation set, and a testing set in a ratio of 7:2:1. The training set is used for model training, the validation set is used for parameter optimization, and the testing set is used for performance validation; The second step is to use the same training and validation sets to train the model in this paper and four comparison models. The initial parameters are set uniformly (learning rate of 0.001, iteration times of 100, hidden layer nodes of 64), and the parameters of each model are optimized through cross validation and grid search methods; Step three, use the test set to conduct prediction tests on each model, record the prediction results, training time, and single prediction time of each model; Step four, calculate the accuracy and real-time performance indicators of each model, compare and analyze the performance differences of each model, and complete model validation.

### 4.3 Verification results and analysis

This instance validation analyzes the comprehensive performance of the model from three dimensions: prediction accuracy, real-time performance, and scene adaptability. In terms of prediction accuracy, the MAE, RMSE, and MAPE indicators of each model were calculated, and the comparison results are shown in Table 2:

*Table 2. MAE, RMSE, MAPE indicator models*

Predictive model	MAE	RMSE	MAPE(%)
Traditional ARIMA model	18.73	24.52	10.23
Single GCN model	15.46	20.18	8.96
Basic LSTM model	12.38	16.74	7.15
LSTM-GCN hybrid model	10.15	14.26	5.98
Fusion spatiotemporal characteristic model	8.62	11.35	4.87

According to the table data, the accuracy indicators of the prediction model that integrates spatiotemporal characteristics proposed in this paper are significantly better than those of the comparison model, with a MAPE reduction of over 40%, indicating that the proposed model can effectively capture the spatiotemporal coupling characteristics of traffic flow and greatly improve prediction accuracy.

In terms of real-time performance, the comparison of training time and single prediction time for each model is shown in Table 3:

*Table 3. Comparison of training time and single prediction time for each model*

Predictive model	Time	Single prediction time
Traditional ARIMA model	42.5	0.86
Single GCN model	38.9	0.73
Basic LSTM model	33.1	0.51
LSTM-GCN hybrid model	31.2	0.45
Fusion spatiotemporal characteristic model	28.7	0.32

The data shows that the real-time performance of the model in this article is superior to all compared models, and can meet the practical needs of "second level prediction" in real-time control of smart

transportation. In terms of scenario adaptability, four typical scenarios were selected for testing: peak hours, off peak hours, rainy weather, and sudden traffic incidents. The results are shown in Table 4:

Table 4. Testing four typical scenarios

Traffic scene	This article models MAPE	Compare the average MAPE of the model	Precision advantage
Off peak period	4.12	6.85	2.73
Rush hour	5.72	9.36	3.64
Rainy weather	6.35	10.12	3.77
Sudden traffic incidents	7.18	11.54	4.36

From the table, it can be seen that the prediction accuracy of our model is superior to the comparison model in various scenarios, especially in the case of sudden traffic events, where the accuracy advantage is most obvious. This indicates that our model can adapt well to the dynamic changes in traffic flow and has strong scene adaptation and generalization capabilities. In addition, through the analysis of spatiotemporal correlation feature mining results, it has been verified that the big data mining method proposed earlier can effectively extract spatiotemporal correlation rules of traffic flow. The mining results are highly consistent with the actual operating status of traffic flow in the region, further proving the feasibility of the big data mining method.

#### 4.4 Verification conclusion

Based on the results and analysis of this case study, the following core conclusions can be drawn: Firstly, the intelligent transportation big data mining method proposed in this paper, which considers the spatiotemporal characteristics of traffic flow, has high feasibility and can effectively handle multi-source heterogeneous intelligent transportation big data, accurately extract traffic flow temporal correlation features, spatial correlation features, and spatiotemporal coupling correlation features, provide high-quality feature input for the prediction model, and the mining results can truly reflect the operation rules of traffic flow. Secondly, the traffic flow prediction model constructed in this article that integrates spatiotemporal characteristics achieves deep integration of spatiotemporal characteristics by introducing spatiotemporal attention mechanisms and improved algorithms. Compared with traditional prediction models and commonly used hybrid models, it has significant advantages in prediction accuracy, real-time performance, and scene adaptability, and can accurately predict the traffic flow operation status in different scenarios, meeting the practical needs of real-time control of smart transportation. Thirdly, case studies have shown that the model and big data mining methods can adapt to the traffic operation characteristics of urban core road networks, providing scientific decision-making basis for traffic congestion warning, road network optimization scheduling, and travel service improvement, and have strong engineering application value. At the same time, this verification also found the shortcomings of the model: under the special scenarios such as extreme rainstorm and large-scale activities, the prediction accuracy declined slightly, and the adaptability of the model to extreme scenarios still needs to be further optimized, indicating the direction of improvement for subsequent research.

## 5. Conclusion

This article focuses on the research of intelligent transportation big data mining and prediction models that consider the spatiotemporal characteristics of traffic flow. In response to the problems of current traffic flow prediction models ignoring spatiotemporal correlations and low efficiency of big data mining, combined with the spatiotemporal coupling characteristics of traffic flow and the characteristics of intelligent transportation big data, theoretical analysis, method design, model construction, and case verification are completed. The core conclusions, shortcomings, and prospects are as follows.

The core research conclusion of this article is to clarify the evolution law and quantitative method of the spatiotemporal characteristics of traffic flow, identify its temporal periodicity, volatility, spatial correlation, and aggregation, and construct quantitative indicators that can provide support for subsequent research. The second is to propose a big data mining method that adapts to spatiotemporal characteristics, design a full process mining framework, and verify that it can effectively improve data quality, accurately extract spatiotemporal correlation features, and is feasible. The third is to construct a hybrid prediction model that integrates spatiotemporal characteristics, introducing spatiotemporal

attention mechanism optimization algorithm. The example verification shows that its MAPE is as low as 4.87%, and the single prediction time is 0.32 seconds. Its accuracy and real-time performance are significantly better than traditional models, and it is suitable for various traffic scenarios. Fourthly, the research results have enriched the relevant theoretical system, providing decision-making basis for traffic control and possessing engineering application value.

There are shortcomings in this study: the mining of multi-scale spatiotemporal correlations is not comprehensive enough, making it difficult to adapt to extreme scenarios; The interpretability of the model is insufficient, and the efficiency of multi-source data fusion needs to be improved. Future research will focus on optimization: deepening the exploration of multi-scale spatiotemporal characteristics and improving quantitative indicators; Optimize model structure to enhance adaptability and interpretability to extreme scenarios; Improve data fusion algorithms to enhance processing efficiency; Explore model lightweighting and industrial deployment, expand application scenarios, and assist in the high-quality development of smart transportation.

## References

- [1] Xiao G N, Wang Y Q, Cai Z Y. A review of machine learning based traffic flow prediction research [J]. *Highway Transportation Technology*, 2025, 42 (10): 145-160.
- [2] Hou Y, Yin J, Zhang Z H, Lu K K. A spatiotemporal heterogeneous two-stage fusion network for traffic flow prediction [J]. *Journal of South China University of Technology (Natural Science Edition)*, 2025, 53 (05): 82-93.
- [3] Hou Y, Zhou R J, Zhang X. Research on Spatiotemporal Consistent Traffic Flow Prediction Based on Adaptive Dynamic Correlation Matrix [J]. *Journal of Lanzhou Jiaotong University*, 2024, 43 (06): 42-53.
- [4] Pan L H, Yin J L, Zhang R, Xie B H, Zhang L L. Global local spatiotemporal perception model for traffic flow prediction [J]. *Computer Engineering*, 2026, 52 (03): 392-402.
- [5] Yao J F, He R, Shi T T, Wang P, Zhao X M. A review of machine learning based traffic flow prediction methods [J]. *Journal of Transportation Engineering*, 2023, 23 (03): 44-67.
- [6] Guan W S, Xiao J L. A review of traffic flow parameter prediction based on joint spatiotemporal features [J]. *Journal of Shanghai University of Technology*, 2022, 44 (06): 592-602.
- [7] JILANI U, ASIF M, ZIA M Y I, et al. A systematic review on urban road traffic congestion[J]. *Wireless Personal Communications*, 2023, 140: 1-29.
- [8] Zhang Y, Liu X, Wang L. Spatiotemporal Attention-Based Graph Neural Network for Traffic Flow Prediction[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2023, 24(8): 8215-8226.