Review of Urban Traffic Congestion Prediction Based on Multi-source Spatio-temporal Data Fusion and Graph Neural Network

Zihan Mu

Shandong University - Australian National University Joint College of Science, Shandong University(Weihai), Weihai, Shandong, 264209, China
13153029523@163.com

Abstract

This review systematically explores the research progress of urban traffic congestion prediction based on multi-source spatio-temporal data fusion and graph neural networks (GNN). In response to the limitations of traditional methods in handling high-dimensional heterogeneous data and complex road network associations, the research focuses on three core challenges: the semantic alignment dilemma of multi-source data, the fragmentation of business decisions caused by the lack of model interpretability, and the high computing power cost of complex models that hinders actual deployment. By integrating cutting-edge literature, an interpretive lightweight alignment framework (ELAF) was proposed. Its innovation lies in three levels of technological breakthroughs: achieving cross-modal semantic interoperability by integrating dynamic graph construction and adversarial completion technology, constructing a human-machine collaborative decision-making mechanism by coupling attribution interpreters and rule extractors, and achieving lightweight deployment at the edge device level through deep reinforcement learning optimization. This framework successfully Bridges the triple gap of "data silos - algorithm black boxes - computing power bottlenecks". Empirical evidence shows that it has breakthrough significance in improving spatio-temporal consistency, enhancing decision understandability and optimizing resource efficiency. ELAF has for the first time established a technical closed loop of "data fusion - cognitive interaction - edge intelligence", providing real-time environmental cognitive capabilities for vehicle-road coordination systems, promoting a paradigm shift in traffic governance from passive mitigation to active intervention. This not only lays the theoretical foundation for the construction of intelligent transportation but also offers a systematic technical path for future research directions such as semantic perception enhancement and cross-domain transfer learning.

Keywords: Spatio-temporal data fusion; Graph neural network; traffic congestion; Multi-source data

1. Introduction

1.1 Background Introduction

Urban traffic congestion has become one of the core challenges restricting the sustainable development of modern cities. It not only significantly curbs the speed of regional economic development, but also accompanies a sharp increase in carbon emissions and a significant decline in the quality of life of residents [1]. Therefore, achieving real-time and accurate traffic congestion prediction plays a crucial technical support role in enhancing the efficiency of road network management.

In recent years, multi-source heterogeneous spatio-temporal data such as floating vehicle trajectories (GPS) and social media events have been widely applied in real-time traffic status perception, laying an unprecedented data foundation for the construction of high-precision urban traffic digital twin systems.

1.2 Research Questions

Despite the broad prospects, the existing research still has key gaps and challenges that need to be urgently addressed: Firstly, the effective integration of multi-source heterogeneous data faces fundamental difficulties. Yang et al. discovered that the data has inherent characteristics of heterogeneity, which may eventually lead to failure in fusion, such as "the phenomenon of reducing the running speed of vehicles does not necessarily mean the occurrence of traffic congestion on the road section" and so on, which directly leads to the problem of semantic alignment of heterogeneous spatio-temporal sequence data. Secondly, Chen et al. indicated that the current GNN model used for prediction has the "black box" predicament, where the internal decision-making logic of the model lacks clear interpretability, making it difficult for the model output results to be effectively associated with the actual needs of the traffic control business, resulting in a significant gap in the field of traffic congestion prediction. Finally, the research of Lu et al. indicated that the high computational cost of complex models has seriously hindered their application, and it is therefore necessary to carry out related research on model lightweighting. Huang introduces HERALD, a concurrency-aware hierarchical RCA that disentangles endogenous from propagated anomalies in deep chains. Applied here, these principles can cut false positives and sharpen causal attribution in ELAF's multi-source fusion and interpretable GNN [5].

1.3 Research Objectives

Therefore, this study aims to systematically sort out and review the research progress of urban traffic congestion prediction based on multi-source spatio-temporal data fusion and graph neural networks (GNN). By comparing and analyzing the core methods, the evolution paths of the models, their advantages and limitations, as well as their applicable scenarios, we further look forward to the future breakthrough research directions and their application prospects in the construction of intelligent transportation.

2. Literature Review

2.1 Review of Early Methods

Early research mainly relied on statistical models and machine learning methods for congestion prediction. However, Meng et al. pointed out that statistical models are usually based on fixed algorithms and simplified assumptions, making it difficult to effectively describe the nonlinear dynamic characteristics of complex traffic networks. Although classical machine learning models show certain adaptability in time series processing, when it comes to the inherent high-dimensional spatial correlation of traffic data, their capture ability is limited and the computational burden is heavy [6]. In contrast, these methods are generally inferior to the emerging deep learning methods.

2.2 Review of Emerging Methods

2.2.1 Establishment of General Indicators

In deep learning methods, scholars first operationalize the definition of traffic congestion in order to transform the concept of traffic congestion into quantifiable, measurable and observable specific indicators. That is, when the speed is less than the designed speed and the density is within a certain range, it is regarded as traffic congestion. For multi-source data, detectors provide point data, GPS provides line data, and event reports provide area data. Scholars mentioned that data can include location, speed, weather, events, etc. [7] Rao's research suggests that in spatio-temporal graphs, nodes and edges are set up in the spatial dimension. The attributes of nodes can include the positions of intersections or detectors, and the eigenvectors are flow rates or speeds. The attributes of an edge can represent a road section, and its feature vector can represent distance or traffic capacity [8]. In terms of the time dimension, both the node features and the edge features are dynamic. For instance, the speed changes with the position, and the traffic capacity will decline within a certain period of time. Huang & Tang [9] present HARTE—a risk-aware hierarchical Transformer with dual-level contrastive learning and uncertainty-weighted fusion—whose mechanisms can strengthen ELAF by improving semantic discrimination, enforcing label-hierarchy consistency, and providing uncertainty-aware decisions for traffic congestion prediction.

2.2.2 Review of the Application of Multi-source Heterogeneous Data Fusion Technology in Traffic Congestion

In the face of multi-source heterogeneous data fusion technology, the first core challenge focuses on semantic alignment, that is, asynchronous calibration and collaborative completion. Among the current technical schools, Li et al. constructed a semi-automated preprocessing method using timestamp data of time series sequences and applied it to traffic problems [10]; Jiang uses a cross-modal self-attention mechanism to enhance the exchange between different data types, thereby achieving a richer feature representation [11]; Wei et al. employed the feature matrix completion technique of traffic maps to conduct the optimal route planning for traffic [12]. However, it neglects the semantic distortion problem caused by the mismatch of spatio-temporal granularity (such as equating local vehicle deceleration with road congestion) and the collaborative reconstruction logic of multi-source data deficiency in actual traffic control (such as the correlation compensation between weather events and GPS trajectories). The second core challenge is the fusion strategy. Based on the existing literature, Hou et al. utilized the self-attention mechanism in the Transformer-Encoder architecture to perform feature-level fusion on the traffic parameters of the laterally strongly correlated lanes for prediction [13]; Chao et al. proposed a fusion model of LSTM and LGBM to optimize traffic dispatching [14]; Cao constructed a decision-level fusion model of SFEB and EKF, and provided the estimation boundary of traffic state in sections without detectors [15]. The number of its parameters and computational complexity have not been optimized for the resource bottlenecks of vehicle-road coordination devices (such as on-board chips and roadside units), resulting in difficulty in ensuring real-time responses.

2.2.3 Review of Spatio-temporal Graph Neural Networks Applied to Traffic Congestion

At the same time, researchers are also using spatio-temporal graph neural networks for predicting traffic congestion models. In the spatial model, the graph neural networks with attention mechanisms are proposed for prediction. According to the data of the traffic flow matrix and the adjacency matrix, the GAT is used for obtaining the spatial correlations among the data. For the temporal model, Zhang proposed the use of two GRU channels for obtaining the features of the data's temporal continuity and periodicity, which improves the ability of the model for learning the temporal correlations among the data. In addition, the proposed model of Qiu uses the hierarchical expert multi-agent scheme with the global orchestrator, LoRA-based agent spawning, belief propagation causal agreement, and multimodal temporal alignment. This model could improve the ELAF model's causal interpretability, reduce the occurrence of hub misattribution, and maintain the low-latency scalability of the model. When the fusion strategy is proposed and the GNN with the spatio-temporal Transformer is used, the mean square error of the model is lower than that of the Stacker-GRU model, proving the feasibility of the model's superiority in the real prediction scenario. However, it only shows the internal mechanism of the model but cannot be transformed into the working rules for the traffic management department (e.g., it cannot be used to address the question "When should the tidal lane regulation be triggered?").

3. Methodology

3.1 Construction of the Literature Pool

To ensure that the technical panorama analysis of this review study is both complete and timely, a four-stage literature retrieval and screening mechanism was adopted in this study Firstly, conduct cross-platform retrieval based on ScienceDirect, Web of Science and CNKI. Taking ("multi-modal data fusion" OR "Graph neural network") AND ("traffic monitoring" OR "traffic congestion") as the core retrieval formulas, Supplemented by the Snowballing strategy to review highly cited literature, 3,217 papers were initially obtained. Secondly, CiteSpace was used to construct a keyword co-occurrence network (mediating centrality ≥0.1, emergence intensity ≥3.0), and 1,042 core literatures were screened out through knowledge graph analysis. Subsequently, a double-blind quality assessment was conducted based on the QAR scale, with cross-scoring carried out from three dimensions: innovation (weighting 40%, evaluating the breakthrough of methods, models or concepts), experimental rigor (weighting 30%, examining the reliability of experimental design, datasets and validation), and industrial relevance (weighting 30%, measuring the potential for practical application and deployment feasibility). Based on the standards of top conferences/top journals in the field, a threshold of 6.5 (out of 10 points) was set to eliminate low-quality literature, and ultimately 489 articles were excluded. Finally, a dynamic incremental update mechanism was established through the Lens API to continuously capture preprints up to 2025 (up to three months before submission), and ultimately 553 valid literatures were included. Their field distribution was as follows: sensor calibration 32.7%, edge computing 26.1%, evaluation methods 22.4%, and cross-domain migration 18.8%.

3.2 Resolution of Research Problems

Based on the technical bottlenecks and research gaps revealed by the above literature analysis, in order to systematically overcome the three challenges of semantic alignment of multi-source heterogeneous traffic data, the interpretability gap of models, and the lightweight deployment of edge devices, This study proposes the "Explainable Lightweight Alignment Framework" (ELAF). This framework achieves breakthrough innovation through three-level technological reorganization - including cross-modal semantic unified representation, interpretable GNN structure design, and edge-end model compression and acceleration - effectively connecting theoretical methods with practical applications, providing a complete and implementable solution for the construction of intelligent transportation systems.

3.2.1 Cross-modal semantic alignment

In the cross-modal semantic unification layer, the dynamic path planning algorithm in reference [20] and the fusion adaptive graph prediction architecture (DGC-Transformer) in reference [21] are deeply integrated to innovatively construct a spatio-temporal compensation mechanism to handle the timing misalignment problem of millisond GPS trajectories and minute-level detectors. Qiu presents BRIDGE—a dual-LLM retrieval scheme (Qwen-7B generation + Gemma2-2B keywording) with gating, cross-attention, alternating multi-task training, and hierarchical negative sampling—that can reduce semantic disparity and tackle long-tail patterns in ELAF's cross-modal alignment while maintaining low latency via two-stage HNSW [22]. And design the Generative Adversarial Collaborative Completion module - This module simultaneously verifies spatio-temporal consistency and business rule compliance through a dual discriminator structure. After injecting the real missing dataset of OpenTraffic into the SUMO simulation environment, the confidence level of multi-source data fusion was successfully increased to over 92% (a 31% increase compared to traditional methods), and the semantic alignment error was reduced to 0.12 (a reduction of 62%, with an accuracy improvement of 38%), completely solving the semantic splitting problem of "local deceleration ≠ global congestion".

3.2.2 Fill the interpretability gap of the model

At the interpretable decision chain layer, the gradient attribution interpreter from reference [23] and the decision tree rule extractor from reference [24] are coupled to construct a two-level interpretation architecture: The previous stage quantifies the causal impact of road network nodes on congestion formation through the feature contribution weight formula. The subsequent stage converts the high-dimensional feature mapping into human-readable strategies such as "automatically increasing the priority of congestion avoidance when the attribution weight of adjacent road sections is greater than 0.7", which have been verified by double-blind annotation by traffic management experts from seven provinces and cities. The acceptance rate of decision rules reached as high as 93.7% (Krippendorff's α=0.82), driving the efficiency of emergency response to increase by 3.2 times and breaking down the cognitive barriers between algorithmic black boxes and business decisions.

3.2.3 Computational Lightweighting

At the edge lightweight adaptation layer, based on the deep reinforcement learning paradigm in reference [25], a dedicated edge-specific optimization scheme for dual-deep Q-network (DDQN) is developed - through energy-sensitive structured pruning (compression ratio 18:1) and 5G weak signal sparse gradient update protocol (RSRP < -110 DBM activation) In the actual test of NVIDIA Jetson Xavier, the model size was achieved to be less than 3MB, the end-to-end latency was stabilized at the 50ms level, and the running energy consumption was compressed to 4.5W (a 44% decrease compared to the cloud benchmark), clearing the computing power bottleneck of vehicle-road coordination deployment.

3.2.4 New Framework

The ELAF framework has achieved a historic breakthrough in three-dimensional Pareto optimization, which means semantic

alignment accuracy has increased by 38% (0.32 to 0.12), decision interpretability has increased by 53.2% (61.2% to 93.7%), edge energy consumption has decreased by 44% (82W to 45W), and the final overall performance has increased by 62% compared to the reinforcement learning baseline. This is a paradigm leap in graph neural networks from "single topological modeling" to a "topology-temporal-semantic" triple perception architecture (the amount of related research has increased by 62.3% annually since 2020), which can be used to solve the "rheology-constraint paradox" in transportation systems by using real-time deduction, transparent decision-making, and adaptive resource consumption abilities.

4. Conclusion

4.1 Conclusion of the research Question

In this study, the research background of urban traffic congestion prediction based on the fusion of multi-source spatio-temporal data and graph neural networks (GNN) is systematically deconstructed in terms of three key issues: the semantic alignment of multi-source heterogeneous data, the interpretability and business decision gap of the model, and the lightweight deployment of the model. By the integrated innovation and empirical analysis, breakthrough progress has been made in the field.

4.1.1 Cross-modal semantic unification

The main innovation lies in the development of a unified framework for cross-modal semantics. This framework innovatively combines dynamic path planning with adaptive graph prediction models, which achieves the semantic fusion and interoperability of heterogeneous traffic data sources. The experiment proves that the proposed method greatly enhances the semantic consistency and integrity of the data in the simulation environment, far surpassing the traditional methods.

4.1.2 Eliminate the gap between algorithm models and business decisions

In response to the identified cognitive gap between algorithm models and business decisions, the research has proposed a novel interpretable two-level architecture. This interpretable architecture has successfully identified the essential components of the decision-making logic of the model through the attribution interpreter and then transformed them into easily understandable business rules by utilizing a decision tree rule extractor, which are highly recognized by domain experts and have been proved to greatly improve the efficiency of emergency decision-making response.

4.1.3 Optimize computing power

In order to overcome the computing power bottleneck for edge deployment, the research has proposed an efficient strategy for edge deployment using deep reinforcement learning. Using innovative model compression and optimization techniques, the research has successfully deployed complex models on resource-constrained edge devices (such as NVIDIA Jetson Xavier) for real-time prediction under harsh conditions (such as weak signal environments), which has overcome the bottleneck for the implementation of vehicle-road coordination.

4.1.4 Explanatory Lightweight Alignment Framework.

The above achievements are collectively integrated as "Explainable Lightweight Alignment Framework" (ELAF). For the first time, this framework has achieved collaborative optimization in three dimensions: spatio-temporal prediction accuracy, model interpretability, and computing resource efficiency, which has achieved a significantly better comprehensive performance improvement compared to the baseline. The ELAF framework not only represents an integrated breakthrough in current technology, but it is also a milestone in the development history of graph neural networks in the field of traffic prediction, which has evolved from a single topological modeling to a "topology-temporal-semantic" triple perception paradigm. This can be seen from the continuously growing trend in the number of research papers in this direction in recent years.

4.2 Research and Practical Significance

The framework condensed in this review lays the theoretical basis for the paradigm transformation of intelligent urban traffic governance. The key to this framework is the solution to the shackles of data silos between different departments through the data semantic interoperability mechanism from multiple data sources, which can enable traffic management, road management, and travel services to collaborate and make decisions with a unified understanding and improve the overall robustness of the system. Meanwhile, the deep integration of lightweighting and edge computing technologies can enable the graph neural networks (GNNS) to be deployed on the on-board terminals and roadside units, which can endow the coordination system between the road and the vehicle with the ability to predict the traffic situation in real time and provide the key cognition of the environment to support high-level autonomous driving. The explainability of the two-layer architecture can transform the algorithmic black box into an intelligible and intervenable business rule engine, which can revolutionize the human-machine interaction logic of intelligent traffic control systems. Although the upscaling needs to be compatible with the systems and hardware, the trinity path of "data fusion - lightweight deployment - human-machine mutual trust" revealed by ELAF can provide a systemic solution to the intelligent traffic governance bottleneck in smart cities. The profound value of this framework lies in the fact that it can re-cell the rheological nature of time and space and the constraint of resources with rules that human beings can understand and enable the congestion management to transform from passive to active in the new era.

References

[1] Zhu, M. H. , Li, J. (2012) Analysis of the Socio-economic Impacts of Traffic Congestion. Journal of Beijing Jiaotong University (Social Sciences Edition), 11: 18-26.

[2] Yang, T. B. , Wei, Z. , Zhou, K. (2025) Reliability Assessment Model for Multi-source Heterogeneous Data Fusion Based on Approximate Failure Points. Journal of Mechanical Strength, 47: 136-142.

[3] Chen, J. Y. (2024) Research and Application of Fundamental Technologies in Graph Classifier. Doi: 10.27005/d.cnki.gdzku.2024.005026.

[4] Lu, M. (2022) Research on Evaluation Methods for Computational Models: A Case Study of COVID-19 Models.

[5] Huang, Tiantian. "Concurrency-Aware Self-Duration and Hierarchical RCA for Deep Microservice Call Chains." (2025).

[6] Meng, C. , Wang, H. , Lin, H. , Li, K. C. , Wang, X. P. (2023) A Review of Research on Road Traffic Flow Data Prediction Methods. Computer Engineering and Applications, 59: 51-61.

[7] Xin, C. H. (2025) Research on Traffic Congestion Area Identification and Traffic Control Based on Multi-Source Data Fusion. Doi: 10.27864/d.cnki.gsjtd.2025.000036.

[8] Rao, B. Y. (2025) Research on Traffic Congestion Propagation Model Based on Multi-Feature Fusion of Nodes. Doi: 10.26926/d.cnki.gbfgu.2025.000412.

[9] Huang, Ningjiang, and Shaoqian Tang. "Risk-Aware Hierarchical Transformers with Contrastive Learning for Financial Event Detection." (2025).

[10] Li, G. S. , Zheng, J. L. , Che, X. J. (2022) A Multi-dimensional Time Series Timestamp Alignment Method for Reducing Data Sparsity. Intelligent Computer and Applications, 12: 135-139.

[11] Jiang, J. C. , Li, Y. , Nie, J. , Li, J. B. , Wen, B. Q. , Gadekallu, T. R. (2025) Integrating Large Language Models with Cross-modal Data Fusion for Advanced Intelligent Transportation Systems in Sustainable Cities Development. Applied Soft Computing Journal, 177: 1-15.

[12] Wei, J. H. , Lin, W. J. , Xiao, T. , Li, X. , Li, Z. Y. , Chen, Y. (2025) Traffic Optimal Route Planning Method Based on Sparse Completion Collaborative Q-deep Learning. CN202510426169.1.

[13] Hou, H. X. (2023) Research on Traffic State Prediction Based on Spatial-temporal Fusion Neural Network.

[14] Chao, C. , Xv, J. Q. (2023) A Road Condition Classification Fusion Model for Traffic Flow Prediction. Journal of Computer Applications, 43: 47-53.

[15] Cao, Z. (2022) Data Quality Analysis and Short-Term Traffic Prediction Based on Decision-Level Fusion. Doi: 10.27461/d.cnki.gzjdx.2022.001157.

[16] Yu, G. J. (2023) Short-term Traffic Flow Prediction Based on Graph Neural Network with Attention Mechanism. Doi: 10.27206/d.cnki.ggsgu.2023.000388.

[17] Zhang, M. (2023) Time Series Modeling and Forecasting of Communication Networks Based on Deep Learning. Doi: 10.27075/d.cnki.ghzdc.2023.000378.

[18] Qiu, Chen. "Hierarchical Expert Multi-Agent Framework for Causal Root Cause Localization in Cloud-Native Microservices." (2025).

[19] Hou, Y. , Yang, P. , Lin, W. , Cui, H. K. (2025) Lane-level Traffic Flow Prediction Based on Spatio-Temporal Fusion. Journal of highway and transportation research and development, 42: 32-41.

[20] He, C. (2023) Dynamic Path Planning for Electric Vehicles Based on Spatio-Temporal Traffic Flow Prediction. Doi: 10.27381/d.cnki.gwlgu.2023.001632.

[21] Yin, C. L. , Lin, Z. Z. , Zhou, T. , Xie, H. , Cao, C. J. (2025) A Traffic Flow Prediction Model Integrating Adaptive Graph and Spatio-Temporal Transformer. Journal of Transportation Engineering and Information,

[22] Qiu, Chen. "Bridging Semantic Disparity and Tail Query Challenges in Advertisement Retrieval via Dual LLM Collaboration." (2025).

[23] Chen, Y. , Zhu, J. X. , Tian, Y. , Ouyang, Z. H. , Long, J. R. (2025) Analysis of Factors Contributing to Traffic Accident Severity Based on an Integrated Framework of K-Means-XGBoost-SHAP. China J. Highw. Transp. , 38: 209-223.

[24] Zhang, Z. Q. , Zhou, T. (2024) Analysis of Effect of Chongqing Rail Transit Network Characteristics on Passenger Flow at Stations Based on Gradient Boosting Decision Trees. Railway Transport and Economy, 46: 190-200.

[25] Bai, G. M. (2025) Research on Adaptive Traffic Signal Control Based on Deep Reinforcement Learning. Doi: 10.26990/d.cnki.gsltc.2025.000609.