

¹ Learning and Predicting Traffic Conflicts in Mixed Traffic: A ² Spatiotemporal Graph Neural Network with Manifold Similarity ³ Learning

⁴ Zongshi Liu^{a,b,c}, Guojian Zou^{a,b}, Ting Wang^{a,b}, Meiting Tu^{a,b}, Hongwei Wang^{c,*} and Ye Li^{a,b,*}

⁵ *^aThe Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, Shanghai, 201804, PR China*

⁶ *^bCollege of Transportation Engineering, Tongji University, Shanghai 201804, PR China*

⁷ *^cInstitute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), Singapore, 138632, Republic of Singapore*

¹⁰ ARTICLE INFO

¹² Keywords:

¹⁴ Conflict risk modeling
¹⁵ Manifold similarity
¹⁶ Mixed traffic flow
¹⁷ Spatiotemporal characteristics
¹⁸ Adaptive graph networks

ABSTRACT

The coexistence of connected and automated vehicles (CAVs) with human-driven vehicles (HDVs) in mixed traffic scenarios introduces significant uncertainties for real-time safety risk assessment. However, the development of safety-prediction models tailored to CAV or mixed-traffic environments remains relatively limited. To address public safety challenges and fortify the security of transportation systems, it is imperative to develop a safety-prediction model tailored for mixed traffic environments. In this study, we leveraged advanced microscopic simulation techniques to generate realistic mixed traffic environments and introduced a novel framework—the Manifold Similarity Spatiotemporal Graph Network (MS-STGNet) to predict real-time conflict potential on freeways. The MS-STGNet framework comprises four strategically designed modules: a residual convolutional module, a manifold-similarity graph module, a temporal convolution layer, and an adaptive fusion gate mechanism. These components dynamically capture both semantic and physical dependencies within traffic data, seamlessly integrating them into a unified predictive model, yielding precise identification of roadway conflict events. Our novel manifold-similarity module incorporates a broader array of traffic-flow attributes during neighbor selection, thereby reducing the propensity for false-positive conflict event predictions, which ensures the model's robust performance within complex, mixed traffic environments. We evaluated the framework's performance under mixed traffic scenarios with varying penetration rates of CAVs and HDVs. The experimental results demonstrate that MS-STGNet achieves consistently exceptional and stable performance across varying market penetration levels and traffic scenarios. Compared to state-of-the-art baseline models, it delivers higher predictive accuracy and substantially lower false alarm rates. The methodologies and outcomes presented in this study have the potential to be used for real-time mixed traffic control on intelligent highways and crash prevention in real-time crash risk warnings at high-risk locations.

³⁶

³⁷ 1. Introduction

³⁸ Traffic crashes remain a significant global issue, resulting in over 1.19 million deaths annually and imposing an
³⁹ economic burden estimated at USD 1.8 trillion, equivalent to approximately 10% of the global GDP (Organization,
⁴⁰ 2023). Predicting and mitigating these crashes has become a critical focus for researchers, with models evolving from
⁴¹ traditional statistical approaches to advanced machine learning (ML) and deep learning (DL) techniques. Leveraging
⁴² real-time traffic data, researchers aim to assess crash risks within road networks, identifying the potential timing and
⁴³ locations of accidents to enable proactive traffic management strategies (Lu et al., 2021; Wang et al., 2024a). Tradition-
⁴⁴ ally, crash-risk estimation has been performed using collision-based models that depend on police-reported accident
⁴⁵ records. A more effective alternative derives risk estimates from traffic conflict analysis - identifying conflicts as pre-
⁴⁶ cursor events to crashes (Tarko, 2012). By obviating the dependency on the slow accumulation of crash data, this
⁴⁷ methodology enables a proactive framework for traffic safety assessment.

⁴⁸ Traffic accidents are inherently stochastic events, influenced by numerous conditional factors such as road and
⁴⁹ vehicle characteristics, as well as environmental and human elements. These complexities render accidents seemingly

* Corresponding authors: Ye Li, Hongwei Wang.

✉ chuochuoliu@tongji.edu.cn (Z. Liu); 2010768@tongji.edu.cn (G. Zou); 2110763@tongji.edu.cn (T. Wang);
meitingtu@tongji.edu.cn (M. Tu); wang_hongwei@a-star.edu.sg (H. Wang); JamesLI@tongji.edu.cn (Y. Li)
ORCID(s): 0009-0004-4049-1956 (Z. Liu)

random in both time and space (Cai et al., 2021; Li et al., 2024). However, historical accident data reveal a strong correlation between traffic accidents and the operational characteristics of traffic flow. Traffic flow dynamics play a pivotal role in the occurrence, progression, and variability of accident risks, ultimately determining outcomes (Liu et al., 2021). Traffic accidents, therefore, result from a dynamic evolutionary process. The transition of traffic states from non-accident to accident conditions reflects a series of changes most evident in real-time traffic data (Wang et al., 2022; Santos et al., 2022). **Despite these insights, existing predictive models have yet to effectively quantify the dynamic evolution of traffic states.**

The rapid integration of connected and automated vehicles (CAVs) into existing road networks marks a pivotal shift in global transportation systems. CAVs have demonstrated significant potential to enhance traffic safety and efficiency by leveraging onboard sensor data and information obtained through V2X communication from other vehicles and infrastructure equipped with roadside units (RSUs) to regulate driving behavior (Zhou et al., 2024; Ma et al., 2024). While CAVs are expected to greatly improve transportation outcomes, projections indicate that by 2045, only up to 24.8% of vehicles will be CAVs (Bansal and Kockelman, 2017). This suggests that for an extended period, mixed traffic—comprising vehicles with varying levels of longitudinal and lateral control as well as differing communication capabilities—will persist (Liu et al., 2018a; Galvani, 2019; Ahangar et al., 2021).

Research on autonomous vehicles predominantly focuses on developing control algorithms to optimize overall traffic efficiency (Hu and Sun, 2019) and enhance vehicle stability (Zheng et al., 2015; Zhou et al., 2019), robustness (Fiengo et al., 2019), and disturbance resistance (Hou et al., 2024b; Chen et al., 2024a). Despite these advancements, CAVs face additional challenges in mixed traffic environments due to the inherent uncertainties of human driving behaviors, such as longer reaction times and perceptual errors (Ivanchev et al., 2019). These human factors contribute to stop-and-go waves characterized by rapid speed fluctuations, reducing traffic safety and efficiency (Di Vaio et al., 2019). **Significant gaps remain in crash analysis within mixed traffic scenarios, where the interaction between CAVs and human-driven vehicles (HDVs) introduces complex dynamics that are not yet fully understood.** Addressing these gaps is essential for ensuring the safe and efficient integration of CAVs into heterogeneous traffic systems.

Traffic datasets often exhibit a highly imbalanced distribution of accident and non-accident cases. This imbalance poses significant challenges for road accident prediction. Over 70% of accident occurrence and injury severity models fail to address class imbalance, potentially biasing predictions toward the dominant class, such as non-crash events in crash occurrence models (Ali et al., 2024). First, zero inflation is inherent in accident data, as locations with no accidents are far more common than those with accidents. This spatially skewed distribution biases models toward predicting zero crashes, hindering the effective training of predictive algorithms (Wu et al., 2023). Second, even in locations where accidents do occur, they are predominantly minor, resulting in a disproportionate number of low-risk scores across the spatial grid. This skews the narrative and underestimates the severity of less frequent but more serious accidents (Shirazi and Lord, 2019; Saha et al., 2020).

Previous studies have employed methods such as random sampling, matched case-control designs (Ma et al., 2023; Cai et al., 2020; Theofilatos et al., 2019), and fixed time-window approaches (Basso et al., 2021; Abou Elassad et al., 2020) to mitigate data imbalance in accident occurrence models. Random sampling, while straightforward and effective in reducing analyst bias during multiple trials, fails to incorporate prior knowledge that could inform non-accident events. In contrast, matched case-control designs naturally align with the framework of balancing accident and non-accident events. **The challenges posed by extensive zero-accident areas and the predominance of minor accidents in high-accident regions highlight the complexity of developing models capable of accurately predicting accident occurrences across diverse regions** (Wang et al., 2024b).

To bridge these research gaps, we propose innovative solutions to overcome the limitations of existing approaches. First, we constructed a realistically simulated environment to model conflict events under mixed traffic conditions, employing conflict prediction in lieu of traditional crash forecasting to pinpoint roadway segments that pose heightened safety challenges. Second, we propose MS-STGNet, a spatiotemporal graph neural network that fuses physical adjacency and semantic features for traffic conflict prediction in mixed CAV–HDV traffic. The framework intentionally builds on mature components (e.g., residual CNN and TCN) to ensure robustness in this new application setting, while introducing a manifold-similarity graph as a physically meaningful prior for adaptive adjacency, which has not been explored in existing mixed-traffic conflict prediction models. The framework incorporates four key components: **1)** A residual convolutional network to extract geographical features in interconnected areas of the land space. **2)** A manifold similarity graph module to capture spatial semantic features in regions. **3)** A temporal convolutional network to model temporal dependencies in traffic flow data, extending spatial features into spatiotemporal representations. **4)** An adaptive fusion gate mechanism combines geographical and semantic spatiotemporal features to generate final

¹ predictions. The contributions of this study are summarized as follows:

- ² 1. A realistically mixed traffic environment has been established to explore the microscopic interactions that may lead to conflict events between CAVs and HDVs. By calibrating the parameters of the car-following model and incorporating heterogeneous Cooperative Adaptive Cruise Control (CACC) platooning, we ensured that the simulated driving behavior closely aligns with real-world driving patterns.
- ³ 2. Our framework introduces a residual convolutional module, temporal convolutional layers, and an adaptive fusion gating mechanism, and integrates them into a unified predictive architecture. This approach enhances the ability of our model to capture and synthesise the intricate dynamics between spatial and temporal points in traffic data.
- ⁴ 3. In MS-STGNet, a manifold similarity graph module has been developed and implemented. By leveraging a similarity matrix derived from traffic state data within the manifold space, we provide prior knowledge regarding the evolution of traffic states. The manifold-similarity module incorporates a broader array of traffic-flow attributes during neighbor selection and uses a pre-computed manifold similarity matrix as an interpretable structural prior, thereby reducing the propensity for false-positive conflict-event predictions.
- ⁵ 4. The performance of MS-STGNet was evaluated on simulated traffic datasets. The experimental results demonstrated the effectiveness and superiority of MS-STGNet in terms of prediction accuracy and its capability to capture traffic conflict events.

⁶ The remainder of this paper is organized as follows. Section 2 mainly reviews the relevant literature. Section 3 states
⁷ the preliminary. Section 4 establishes the simulation environment. Section 5 proposes the MS-STGNet framework and
⁸ Section 6 conducts the experiments. Finally, we conclude the paper in Section 7 and discuss further research.

⁹ 2. Related work

¹⁰ 2.1. Mixed traffic flow modeling for traffic safety

¹¹ Exploring the impact of mixed traffic flow modeling on safety is critical for identifying the key factors required
¹² to accurately simulate the driving behaviors of CAVs and HDVs. Existing studies commonly adopt longitudinal car-
¹³ following models such as Cooperative Adaptive Cruise Control (CACC), Adaptive Cruise Control (ACC) developed
¹⁴ by the PATH laboratory (Milanés et al., 2013; Milanés and Shladover, 2014), and the Intelligent Driver Model (IDM)
¹⁵ (Treiber et al., 2000) to represent the dynamics of CAVs, autonomous vehicles (AVs), and HDVs in mixed traffic
¹⁶ environments (Liu et al., 2018a; Zhou and Zhu, 2020; Yao et al., 2023; Chen et al., 2024b). These models are typically
¹⁷ implemented in microscopic traffic simulation tools such as VISSIM, SUMO, and CARLA to evaluate the safety
¹⁸ implications of different CAV market penetration rates (MPRs) and traffic demand levels. In general, simulation-based
¹⁹ studies report reductions in rear-end and lane-changing conflicts and increases in average travel speeds as CAV/AV
²⁰ penetration increases (Mousavi et al., 2021; Tan et al., 2023). However, several works also highlight that, without
²¹ advanced V2X communication frameworks and richer behavior modeling, the safety benefits tend to be modest and
²² context-dependent (Tarko, 2021). These findings underscore the importance of integrating realistic vehicle behavior
²³ models and communication schemes into mixed-traffic safety assessment frameworks.

²⁴ A notable gap in these studies is the insufficient distinction between CAVs and HDVs, particularly in behavioral
²⁵ characteristics such as prolonged reaction times and perceptual uncertainties associated with human drivers, which are
²⁶ often oversimplified in HDV modeling (Gu et al., 2022). While analyses of macroscopic traffic characteristics (e.g.,
²⁷ fundamental diagram parameters) may not introduce significant biases, neglecting these distinctions can substantially
²⁸ impact the evaluation of microscopic traffic characteristics, especially those related to safety-critical features (Garg and
²⁹ Bouroche, 2023). In addition, existing conflict or crash prediction models have been rarely tested for their performance
³⁰ in mixed traffic scenarios, leaving a significant gap in understanding their applicability and effectiveness under such
³¹ complex conditions (Hou et al., 2024a).

³² 2.2. Spatial-temporal safety prediction with learning-based model

³³ Predicting traffic accidents has long been a critical topic in mobility management research. Early studies predom-
³⁴ inantly employed traditional statistical methods such as regression models (Caliendo et al., 2007; Bergel-Hayat et al.,
³⁵ 2013), Bayesian networks (Martin et al., 2009; Hossain and Muromachi, 2012), and tree-based algorithms (Wang et al.,

1 2010; Lin et al., 2015). These approaches provided initial insights into accident patterns, particularly in small geo-
 2 graphical areas, but their ability to capture nonlinear relationships and dynamic dependencies between road segments
 3 was limited (Zhang et al., 2014a). Moreover, they often analyzed accident data in isolation, neglecting critical interde-
 4 pendencies between locations, which restricted their applicability to citywide analyses with large datasets (Wang et al.,
 5 2021).

6 With the advent of deep learning, researchers began exploring models that jointly capture spatial and temporal
 7 patterns. Convolutional neural networks (CNNs) have been widely used to detect spatial structures (Chen et al., 2018;
 8 Hu et al., 2020), while recurrent neural networks (RNNs) and their variants model temporal dependencies (Sameen
 9 and Pradhan, 2017; Yuan et al., 2019). Hybrid frameworks such as Long Short-Term Memory (LSTM) networks and
 10 ConvLSTM-based architectures further advanced citywide accident prediction by integrating spatial and temporal fac-
 11 tors. For example, Ren et al. (2018) used LSTM networks to incorporate temporal influences across multiple locations,
 12 and Bao et al. (2019) developed a spatiotemporal convolutional LSTM network (STCL-Net) that effectively captured
 13 the spatiotemporal dependencies of urban road networks. However, these grid-based methods often overlooked de-
 14 tailed urban geo-semantic information, such as complex road network semantics and intersection configurations.

15 To overcome these limitations, graph-based deep learning methods have emerged, leveraging the inherent graph
 16 structure of road networks to model spatial relationships. Graph convolutional networks (GCNs) (Zhou et al., 2020;
 17 Trirat et al., 2023), graph attention networks (GATs) (Huang et al., 2019; Wang et al., 2023), and spatiotemporal graph
 18 neural networks (ST-GNNs) (Yu et al., 2021) have proven effective in integrating spatial and temporal dynamics by
 19 representing road segments as nodes and their connections as edges. Several studies have pioneered these advance-
 20 ments. Zhou et al. (2020) introduced the Differential Time-Varying Graph Neural Network (DTGN), integrating spa-
 21 tiotemporal correlations with a data augmentation strategy to address zero inflation in accident data. Yu et al. (2021)
 22 proposed a spatiotemporal graph convolutional network featuring a three-layer structure that independently processes
 23 the road graph, spatiotemporal data, and embeddings, and tackled zero inflation by undersampling to balance risky and
 24 non-risky segments.

25 Recent work has further integrated probabilistic frameworks into graph-based models to explicitly account for un-
 26 certainty in accident risk. Gao et al. (2024) incorporated Zero-Inflated Tweedie Distributions (ZITD) into an ST-GNN
 27 model, parameterizing accident risk with components for mean, variance, and zero inflation to better handle highly
 28 imbalanced and long-tailed data. Trirat et al. (2023) proposed a multi-view graph neural network that incorporates
 29 both dynamic and static similarity information, providing a more adaptive representation of traffic accidents under
 30 dynamic geographical semantics and structural alignment. Their model employs a Huber loss to robustly adapt to
 31 zero inflation. Although spatiotemporal GNNs and attention-based adaptive graphs have significantly improved traffic
 32 prediction and safety modelling, their applications to real-time conflict prediction in mixed CAV-HDV traffic remain
 33 limited, and most adaptive adjacency mechanisms are learned purely from instantaneous node embeddings without an
 34 explicit traffic-state prior, which motivates our manifold-similarity-based graph design in the following sections.

35 In summary, despite significant advancements and promising results in traffic safety prediction, existing research
 36 has yet to fully address the uncertainty associated with predicting accident occurrences and assessing accident risk.
 37 Many models overlook the underlying spatial correlations and the inherent dynamic interactions within road networks.
 38 Specifically, the transition of traffic states from non-risky to risky is a dynamic evolutionary process, which is crucial
 39 for reliable safety prediction but remains insufficiently explored in current studies. Moreover, the use of traffic-conflict
 40 data in place of crash records for safety forecasting constitutes an emerging trend that has been scarcely addressed in
 41 the existing literature (Ali et al., 2023). In addition, existing spatiotemporal graph-based safety models typically define
 42 spatial dependencies through fixed adjacency matrices or adaptive attention mechanisms in the original feature space,
 43 and rarely exploit manifold-based traffic-state similarity as an explicit prior, particularly in mixed CAV-HDV traffic
 44 environments.

45 2.3. Manifold learning in traffic state modelling

46 Traffic states (e.g., free flow, congestion, bottleneck distributions) can be viewed as a dynamic system whose intrin-
 47 sic structure is often embedded nonlinearly in high-dimensional space. Traditional distance metrics, such as Euclidean
 48 distance and Manhattan distance, operate within high-dimensional linear spaces and are susceptible to the "curse of
 49 dimensionality," making it difficult to accurately capture the intrinsic geometric properties of high-dimensional traffic
 50 data (Liu et al., 2022; Wang et al., 2024c). As an alternative, manifold distance measures the geometric path length
 51 along the surface of the manifold, providing a more accurate representation of the dynamic evolution and intrinsic
 52 similarity of traffic systems. Specifically, manifold distance assumes that the traffic state data are distributed on a

Table 1

Summary of key aspects considered in related studies on traffic safety prediction.

Study category	Mixed CAV-HDV traffic	Traffic-conflict-based surrogate-safety indicator	Deep spatiotemporal neural model for safety	Graph-based representation	Manifold-based traffic state similarity prior	Explicit treatment of class imbalance/zero inflation
Mixed-traffic microsimulation safety studies						
(e.g., Liu et al. (2018a); Zhou and Zhu (2020); Yao et al. (2023); Chen et al. (2024b))	✓	✓	✗	✗	✗	✗
Grid-based deep learning crash-risk prediction						
(e.g., Ren et al. (2018); Bao et al. (2019); Chen et al. (2018); Hu et al. (2020))	✗	✗	✓	✗	✗	✓
ST-GNN-based accident risk models						
(e.g., Zhou et al. (2020); Yu et al. (2021); Gao et al. (2024); Trirat et al. (2023))	✗	✗	✓	✓	✗	✓
Manifold learning for traffic state/safety modelling						
(e.g., Wang et al. (2009); Su et al. (2020); Seoa (2023); Liu et al. (2022))	✗	✗	✓	✗	✓	✓
Our framework (MS-STGNet)	✓	✓	✓	✓	✓	✓

1 low-dimensional manifold embedded within high-dimensional space. By calculating the shortest path length between
 2 traffic states in the manifold space, it effectively characterizes the true evolutionary trajectory of the system (Yousaf
 3 et al., 2020; Liu et al., 2018b).

4 Early studies have applied manifold learning to various traffic-related tasks. For example, Wang et al. (2009) pro-
 5 posed a cooperative traffic state recognition method based on manifold learning that preserves the geometric structure
 6 of high-dimensional data, and Lu et al. (2012) introduced a graph embedding algorithm that balances local manifold
 7 structures and global discriminative information for traffic sign recognition. Manifold techniques have also been used
 8 to identify moving vehicle trajectories and collective behavior patterns. Lee et al. (2012) projected trajectory features
 9 onto a 2D manifold and clustered them into a small number of Gaussian components, while Yang and Zhou (2011)
 10 combined Local Linear Embedding (LLE) and Principal Component Analysis (PCA) to capture local and global fea-
 11 tures of traffic parameter data. In addition, Zhang et al. (2014b) employed weighted Euclidean distance based on
 12 traffic-parameter similarity to classify traffic states.

13 Recent studies have begun to explicitly model traffic flow on low-dimensional manifolds. For example, Su et al.
 14 (2020) used a convolutional variational auto-encoder to extract low-dimensional manifold representations of daily
 15 urban traffic flow and showed that clustering in this latent space reveals meaningful traffic patterns. Seoa (2023)
 16 applied Uniform Manifold Approximation and Projection (UMAP), a non-linear dimension-reduction method based
 17 on manifold learning, to obtain two-dimensional embeddings of large-scale network traffic states, demonstrating that
 18 the learned manifold coordinates intuitively capture different congestion regimes. In the field of traffic safety, Liu
 19 et al. (2022) incorporated manifold characteristics of traffic flow into a transfer-learning-based highway crash risk
 20 evaluation model and reported improved discrimination between high- and low-risk traffic states compared with models
 21 that rely solely on Euclidean features. These studies indicate that manifold-based representations can provide a more
 22 faithful description of the dynamic evolution and similarity of traffic systems than conventional distance measures in
 23 the original feature space.

24 Existing studies indicate correlations between traffic flow data at each collection point, especially concerning multi-
 25 source fluctuations, warranting further investigation. However, current accident prediction research rarely considers
 26 the manifold characteristics of traffic states. Additionally, few studies have attempted to integrate the concept of state
 27 transitions in manifold learning into deep learning frameworks, and, to the best of our knowledge, none has embedded
 28 manifold-based traffic-state similarity into a spatiotemporal graph neural network for real-time conflict prediction in
 29 mixed CAV-HDV traffic.

30 As shown in Table 1, existing studies usually cover only a subset of these aspects, leaving important gaps such
 31 as the lack of conflict-based indicators, manifold-based state similarity, or explicit treatment of class imbalance. In

¹ contrast, the proposed MS-STGNet jointly incorporates all these aspects within a single framework, providing a more
² comprehensive solution for real-time safety prediction in mixed CAV-HDV traffic.

³ 3. Preliminary

⁴ 3.1. Traffic network graph

⁵ Road networks can be conceptualized as connected and directed topological structures within a physical space.
⁶ To effectively undertake the essential preliminary work for traffic conflict modeling, it is imperative to map the road
⁷ network of the geographical area into a logical space interpretable by computational systems. Based on the arrangement
⁸ of loop detectors within the simulation environment, the study area can be partitioned into $L \times S$ grids, determined by
⁹ the number of lanes and the lengths of the segments. The input road network can be defined as $G = \{V, E, A\}$ where
¹⁰ V represents the set of nodes, defined as $V = \{0, 1, \dots, N\}$, with N being the total number of nodes. In this study,
¹¹ the entire network is partitioned into 4×27 grids (based on the detector setting), which represent 4×27 nodes. The
¹² set of edges E signifies the connections between these nodes. A is the adjacency matrix, represents the proximity of
¹³ the nodes, and is expressed as $A \in \mathbb{R}^{N \times N}$.

¹⁴ 3.2. Embedding

¹⁵ Timestamps play a critical role in modeling processes. In this study, timestamp embedding specifically includes
¹⁶ hour embedding and day-of-week embedding. Traditional timestamps are typically represented as integers, where iden-
¹⁷ tical numerical values are often misinterpreted as contributing equally to conflict prediction. For instance, Saturday is
¹⁸ represented as the integer 5, and 6:00 AM is also represented as the integer 5, leading to equivalent significance in the
¹⁹ input variables. This study addresses the issue by mapping integers to one-hot vectors, which are subsequently trans-
²⁰ formed into high-dimensional embeddings by applying two-dimensional convolutional neural networks (2-D CNNs)
²¹ (Zou et al., 2023b; Wang et al., 2024d). As a result, hour embedding can be represented as $\mathbf{H}_t \in \mathbb{R}^{N \times d}$, and day-of-
²² week embedding can be represented as $\mathbf{W}_t \in \mathbb{R}^{N \times d}$, where d denotes the dimensionality of the input, set to 64 in this
²³ study. The 2D-CNNs transformation process can be expressed as:

$$\mathbf{F}_{l,t} = \sigma(\mathbf{K}_{l,t} \odot \mathbf{F}_{l-1,t} + \mathbf{b}_{l,t}) \quad (1)$$

²⁴ where $\mathbf{F}_{l,t}$ represents the feature output at layer l and time step t , $\mathbf{K}_{l,t} \in \mathbb{R}^{1 \times 1 \times d}$ denotes the convolution kernel with a
²⁵ size of 1×1 ; \odot indicates the convolution operation; $\mathbf{b}_{l,t} \in \mathbb{R}^d$ is the bias term; and σ is the ReLU activation function.

²⁶ 3.3. Problem definition

²⁷ The real-time conflict risk analyses were proposed to establish the relationships between conflict occurrence proba-
²⁸ bility and pre-conflict traffic operational conditions. Based on historical input sequences, including traffic flow, speed,
²⁹ and occupancy, combined with the traffic network graph, a model is established to predict the likelihood of conflict
³⁰ occurrences within future time steps, as shown below:

$$(\hat{Y}_{t_p+1}, \hat{Y}_{t_p+2}, \dots, \hat{Y}_{t_p+M}) = f_{\Theta}(X_{t_1}, X_{t_2}, \dots, X_{t_p}; G) \quad (2)$$

³¹ where $X = \{X_{t_1}, \dots, X_{t_p}\} \in \mathbb{R}^{T \times W \times H \times d_x}$ represents the historical observations of $W \times H$ grids; $\hat{Y} =$
³² $\{\hat{Y}_{t_p+1}, \dots, \hat{Y}_{t_p+M}\} \in \mathbb{R}^{T \times W \times H \times 1}$ represents the prediction of conflict occurrence at the next time slot; Θ is learn-
³³ able parameters.

³⁴ 4. Simulation and data generation

³⁵ The open-source platform Simulation of Urban MObility (SUMO) is utilized to perform day-long traffic simulations
³⁶ on the target road segment, generating data for the proposed model.

³⁷ 4.1. Simulation network

³⁸ To evaluate and implement the proposed modeling framework, we conduct simulations on a four-lane highway
³⁹ stretching over 14 km, which also has been used for model calibration. This choice is consistent with the calibration of

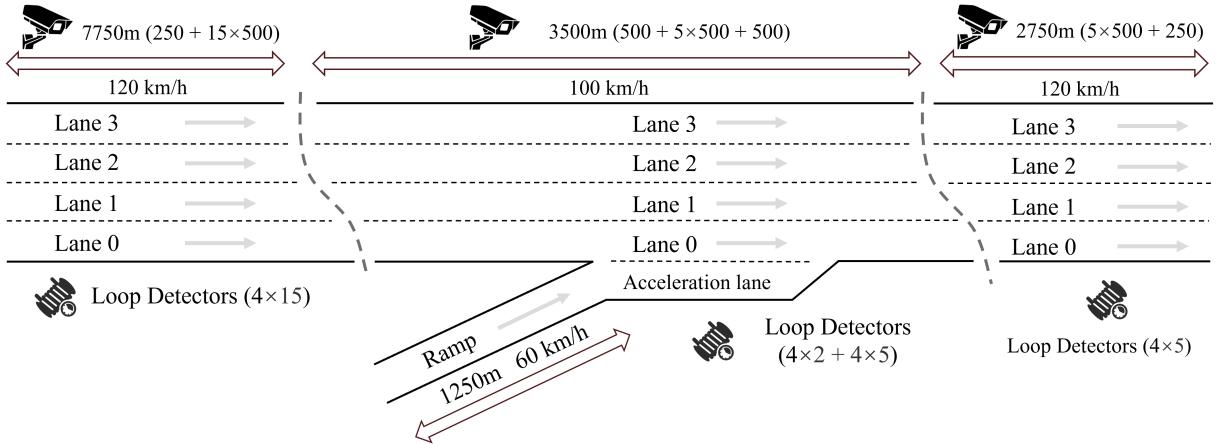


Fig. 1. The silhouette of the simulation road network.

the Enhanced Intelligent Driver Model (EIDM), whose parameters are estimated from the highD dataset of naturalistic trajectories on multilane highways. A segment of 14 km provides sufficient distance for vehicles to accelerate, cruise, and interact, so that stable traffic states and realistic conflict events can emerge without being dominated by boundary effects. The main road is segmented into three parts measuring 7,750 m, 3,500 m, and 2,750 m, with speed limits set at 120 km/h, 100 km/h, and 120 km/h, respectively. In addition to the upstream and downstream trunk links, this section includes connections to five on-ramps, featuring a 250-meter-long acceleration lane running parallel, to mimic real freeway operations and to increase the complexity of traffic interactions, thereby generating more representative conflict-prone situations (as shown in Fig.1).The on-ramp has a designated speed limit of 60 km/h.

Two types of detectors are installed on the main roads: Virtual surveillance cameras and Loop detectors. The virtual cameras monitor the entire main road, capturing detailed information for every vehicle passing through these sections. The data were collected at a frame interval of 0.2 s, providing high-resolution trajectory details for conflict analysis. The average distance between loop detectors is spaced at 500-meter intervals, each lane has a detector to gather traffic flow, speed, and occupancy data within localized zones in 30 s collection intervals (resulting in a total of 4×27 detectors). Additionally, 250-meter buffer zones are established at both the start and end of the road segment to exclude data from statistical analysis.

4.2. Car-Following model and lane-change model

Traffic simulations rely on car-following model and lane-change model to accurately represent the longitudinal and lateral movements of vehicles. In this study, the Enhanced Intelligent Driver Model (EIDM) (Salles et al., 2020), an improved version of the commonly used Intelligent Driver Model (IDM), is chosen to model the car-following behavior of HDVs. For CAVs, the PATH CACC model is applied (Milanés et al., 2013; Milanés and Shladover, 2014; Makridis et al., 2020). Lane-changing behavior is simulated using SUMO's default LC2013 model for HDVs. For CAVs, the Plexe extension in SUMO facilitates platoon-specific lane-changing maneuvers.

Table 2

Calibrated parameters for EIDM model.

Parameter (unit)	Car			Truck		
	Bounds	Mean	Std.	Bounds	Mean	Std.
length (m)	[3.6, 5.9]	4.7	0.4	[4.0, 23.2]	14.6	3.9
maxSpeed (m/s)	[33, 45]	36	4.7	[26, 28]	27	1.9
decel (m/s^2)	[4.5, 5.5]	5.0	2.0	[2.6, 3.4]	3.0	2.0
accel (m/s^2)	[2.0, 3.5]	2.5	2.0	[1.0, 1.4]	1.2	2.0
tau (s)	[0.5, 5.8]	1.5	1.0	[0.5, 8.1]	2.1	1.6
minGap (m)	[2.5, 3.5]	3.0	1.0	[4.0, 5.7]	4.5	1.0

4.3. Model parameter calibration

Real-world traffic exhibits variability and diversity in driving behaviors. To replicate these characteristics in the simulations, we utilized the parameter distribution calibrated by Liu et al. (2024) for the EIDM model. Vehicle parameters are assigned individually using a distribution generator, ensuring unique behavior for each vehicle while collectively representing real-world traffic conditions from a statistical perspective. Liu et al. (2024) performed EIDM calibration using the HighD dataset. By extracting vehicle trajectories from SUMO and computing their symmetric mean absolute percentage error (SMAPE) against corresponding HighD dataset trajectories, these values were ranked to produce statistical descriptions. Model parameters were iteratively adjusted until the third quartile (Q3) of SMAPE fell below 10%. The calibration process and calibrated vehicle parameters are shown in Fig.2 and Table2 respectively. For CACC, researchers in the PATH project calibrated the model using real experimental data (Makridis et al., 2020). The calibrated CACC model successfully replicates the car-following dynamics observed in real-world CAV platoons.

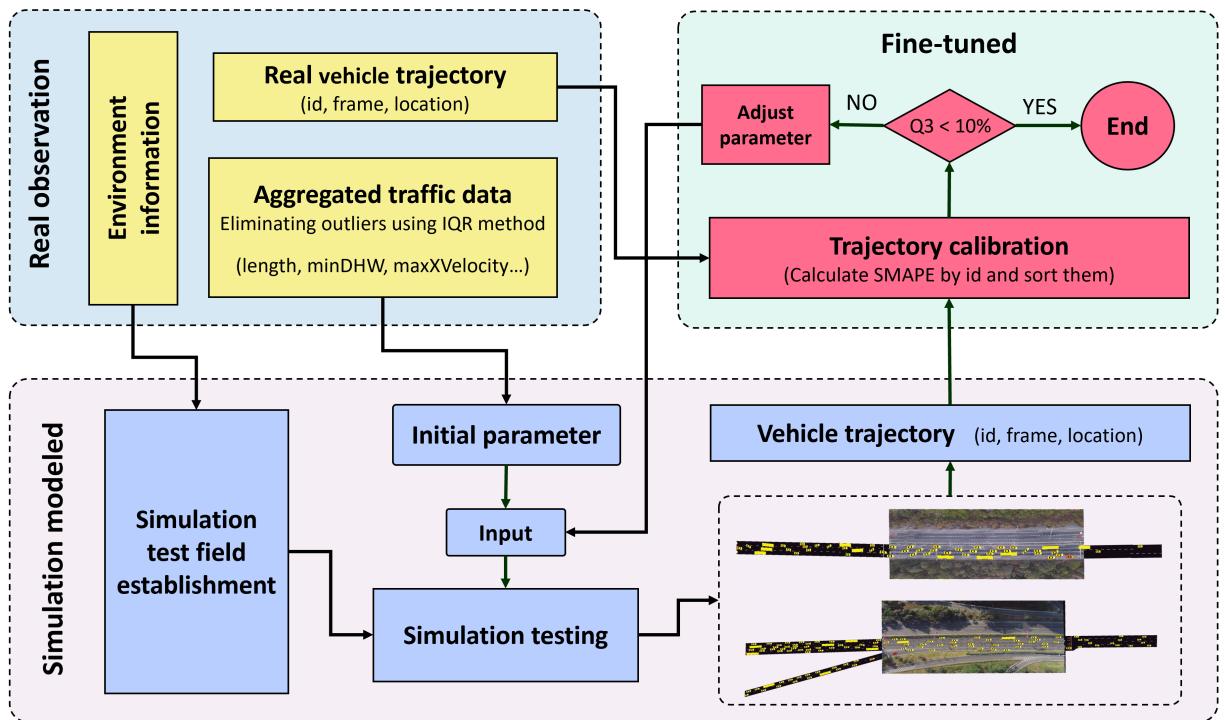


Fig. 2. The process of calibrating the EIDM model.

4.4. Conflict definition based on safety surrogate measures

Given the limited availability of field data for mixed traffic, numerous studies examining the safety impacts of CAVs have utilized surrogate safety measures (SSM) to assess safety risks in mixed traffic scenarios (Zhang et al., 2020; Papadoulis et al., 2019). In this study, we considered two widely adopted SSMs for rear-end crash analysis to quantify the traffic conflicts and provide an indication of how close a vehicle is to being involved in a collision: Time-to-Collision (TTC), and Deceleration Rate to Avoid a Collision (DRAC). On the other hand, for lateral maneuvers, we employ the Distance Differential Ratio (DDR) to quantify the risk associated with lateral movement.

(1) Time-to-Collision (TTC) measures the time remaining until a potential collision occurs if both the leading and following vehicles maintain their current speeds and trajectories (Vogel, 2003).

$$TTC_i(t) = \begin{cases} \frac{x_{i-1}(t) - x_i(t) - L_{i-1}}{v_i(t) - v_{i-1}(t)}, & \text{if } v_i(t) > v_{i-1}(t) \\ \infty, & \text{otherwise} \end{cases} \quad (3)$$

(2) Deceleration rate to avoid a crash (DRAC) refers to the minimum rate at which a following vehicle must decelerate to align its speed with that of the leading vehicle (Fu and Sayed, 2021). (Lu et al., 2021).

$$DRAC_i(t) = \begin{cases} \frac{(v_i(t) - v_{i-1}(t))^2}{(x_i(t) - x_{i-1}(t) - L_{i-1})}, & \text{if } v_i(t) > v_{i-1}(t) \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

1 (3) Distance Differential Ratio (DDR) concentrate on the critical instant when a vehicle completes its lane change
 2 and assess the safety by examining its spatial gap to both the leading and trailing vehicles in the target lane (Fu and
 3 Sayed, 2021).

$$DDR = \min\left(\frac{d_f - d_f^*}{d_f}, \frac{d_l - d_l^*}{d_l}\right) \quad (5)$$

4 For the above formulas, where $x_i(t)$ and $x_{i-1}(t)$ are the longitudinal location of the leader and follower at timestamp
 5 t , respectively, while $v_i(t)$ and $v_{i-1}(t)$ are the corresponding speeds. L_{i-1} is the length of the preceding vehicle $i - 1$.
 6 d_l and d_f denote the longitudinal distances from the subject vehicle to its immediate leader and follower, respectively,
 7 measured immediately after completing the lane-change. Correspondingly, d_l^* (d_f^*) specifies the minimum safe gap that
 8 must be maintained to the nearest leading (following) vehicle in order to satisfy the prescribed deceleration constraints.
 9 A smaller TTC / DDR or a larger DRAC indicates a more hazardous situation. Predefined thresholds are essential to
 10 detect potential traffic conflicts. A traffic conflict is identified when the TTC and DDR drop below the set threshold or
 11 the DRAC exceeds it. Referring to previous studies, this study establishes the TTC threshold for conflict identification
 12 at 2s, the DDR threshold is set to be -0.12, and the DRAC threshold at 2 m/s² (Yang et al., 2021; Li et al., 2017a; Zhang
 13 et al., 2020).

14 5. Methodology

15 5.1. Model architecture overview

16 The architecture of the MS-STGNet model proposed in this study is illustrated in Fig.3, comprising four main components:
 17 the residual convolutional module, the manifold-similarity graph module, the TCN layer, and the fusion gate
 18 mechanism. Initially, the spatial dependency among road segments within the study area is modeled using a residual
 19 convolutional network. On the other hand, traffic flow, speed, and occupancy data from the road network are input into
 20 the manifold-similarity graph module, where manifold distance is computed to characterize the traffic state similarity
 21 between different road segments and the evolutionary trajectories of traffic states. Subsequently, the features captured
 22 by these two modules are processed through a specially designed temporal convolutional network (TCN) to extract
 23 their respective temporal dependencies, thereby forming comprehensive spatiotemporal feature information. Finally,
 24 a fusion gate mechanism autonomously integrates the spatiotemporal features from both components, producing the
 25 final output. Further details on each component will be provided in the subsequent sections.

26 5.2. Residual convolutional module

27 Traffic states demonstrate significant geographical spatial dependencies within road networks. Adjacent areas are
 28 inherently linked by road segments. Traffic propagation between neighboring regions introduces causality, particu-
 29 larly in the context of traffic incidents. For example, the traffic dynamics in the target region are influenced by inflows
 30 from its neighboring regions, which may exacerbate traffic congestion. Elevated traffic volumes in these areas substan-
 31 tially increase the likelihood of accidents, such as crashes and casualties. Furthermore, adjacent regions often share
 32 comparable environmental conditions, such as weather patterns and road infrastructure designs, further reinforcing
 33 their interdependence. To model these intricate spatial relationships, a residual convolutional network (ResNet) is pro-
 34 posed. This network captures spatial dependencies ranging from localized interactions to global patterns by employing
 35 stacked residual blocks. Each residual block integrates two 2D convolutional layers and a shortcut connection. The
 36 transformation process within the k -th residual block at time step t is defined as:

$$\mathbf{H}_{\text{res}}^{t,k} = \mathcal{F}_k(\mathbf{H}_{t,k-1}) + \mathcal{R}_k(\mathbf{H}_{t,k-1}) \quad (6)$$

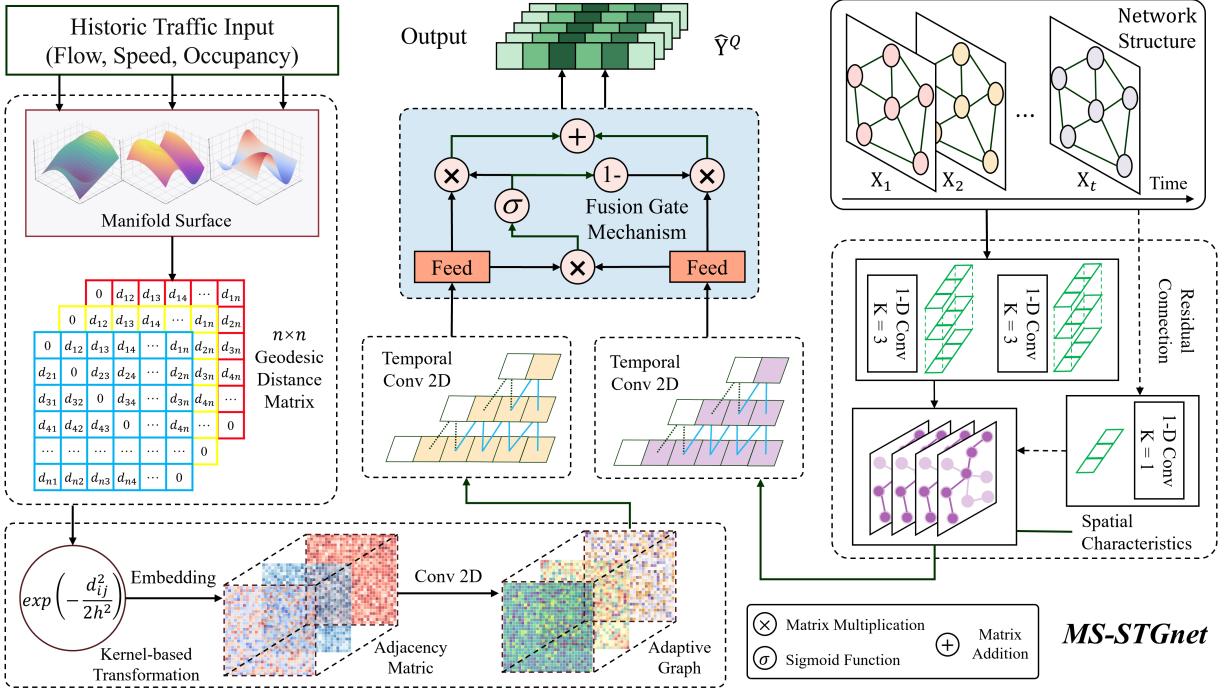


Fig. 3. Framework overview of MS-STGNet.

- 1 where $\mathcal{F}_k(\cdot)$ represents 2D convolutional transformation capturing local dependencies, and $\mathcal{R}_k(\cdot)$ denotes shortcut
2 (residual) connection ensuring feature propagation. Specifically, $\mathcal{F}_k(\cdot)$ can be expressed as:

$$\mathcal{F}_k(\mathbf{X}_t) = \sigma\left(\mathbf{W}_k^{(1)} \circledast \sigma\left(\mathbf{W}_k^{(0)} \circledast \mathbf{X}_t + \mathbf{b}_k^{(0)}\right) + \mathbf{b}_k^{(1)}\right) \quad (7)$$

- 3 where \circledast indicates the convolution operation; $\mathbf{W}_k^{(0)}$ and $\mathbf{W}_k^{(1)}$ represent convolution kernels for the first and second
4 layers (3×3); $\mathbf{b}_k^{(0)}$ and $\mathbf{b}_k^{(1)}$ denote bias terms; and σ is the ReLU activation function. $\mathcal{R}_k(\cdot)$ can be expressed as:

$$\mathcal{R}_k(\mathbf{X}_{t,k-1}) = \mathbf{W}_{k,\text{res}} * \mathbf{X}_{t,k-1} + \mathbf{b}_{k,\text{res}} \quad (8)$$

- 5 where $\mathbf{W}_{k,\text{res}}$ is 1×1 kernel for dimensional alignment; $\mathbf{b}_{k,\text{res}}$ is residual bias. The initial input is $\mathbf{X} \in \mathbb{R}^{T \times W \times H \times d_x}$,
6 and the output is $\mathbf{H}_{\text{res}} \in \mathbb{R}^{T \times W \times H \times d}$.

5.3. Manifold-similarity graph module

7 Although the residual convolutional network (ResNet) is specifically designed to capture spatial dependencies
8 among physically connected regions, its modeling capability is limited in certain cases. For instance, some regions
9 may lack direct road segment connections, while others, despite being geographically distant, exhibit high correlations
10 or shared characteristics. This limitation is particularly evident in traffic conflict analysis, where upstream and
11 downstream road segments of a conflict site may display similar traffic characteristics due to the incident. Such con-
12 straints hinder the ability of ResNet to comprehensively model spatial dependencies in these complex scenarios. To
13 address these challenges, a novel methodology has been proposed to reconstruct the relationships between regions
14 within a non-Euclidean space, which integrates three innovative techniques—similarity matrices, adaptive graphs, and
15 bidirectional random walks—to extract deep and semantic spatial features effectively, as illustrated in Fig.4.

16 Specifically, we utilize predefined similarity matrices to encode the semantic spatial dependencies between valid
17 regions. However, solely relying on prior knowledge imposes limitations on uncovering latent spatial correlations
18 embedded within the data. Adaptive graphs are integrated to dynamically capture global spatial relationships across
20 valid regions to address this problem. These adaptive graphs are initialized using predefined similarity matrices and

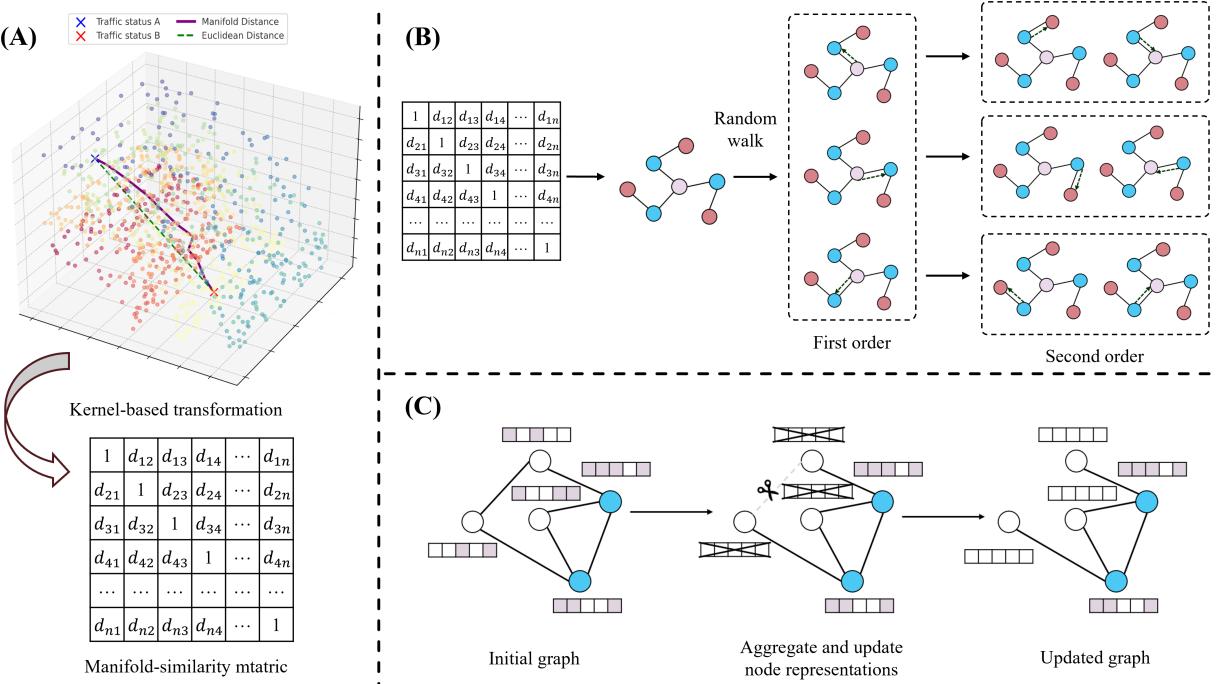


Fig. 4. Processing of self-adaptive graph based on manifold similarity. **(A)** The transformation process from manifold distance to similarity matrix. **(B)** Two-step random walk process. **(C)** Adaptive graph update process.

1 iteratively refined during the training process. Moreover, modeling deep spatial dependencies and intricate inter-
2 relations among regions proves insufficient with single-directional and first-order graph structures. Consequently, we
3 incorporates multi-order bidirectional random walks, enabling the aggregation and refinement of node representations
4 by leveraging information from higher-order and bidirectional neighboring regions.

5.3.1. Manifold similarity graph

6 Assume the traffic state dataset is defined as $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$, where each data point $\mathbf{x}_i \in \mathbb{R}^d$ represents a
7 d -dimensional feature vector. The weighted adjacency graph of n traffic state data points can be expressed as $G =$
8 (V, E, W) , where $V = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ represents the set of nodes. E denotes the set of edges, indicating whether
9 points \mathbf{x}_i and \mathbf{x}_j are adjacent. The weight w_{ij} represents the Euclidean distance between traffic state points \mathbf{x}_i and \mathbf{x}_j .
10 If the two points are neighbors, the distance is preserved; otherwise, it is set to ∞ , indicating no direct connection. The
11 calculation formula is:

$$w_{ij} = \begin{cases} \|\mathbf{x}_i - \mathbf{x}_j\|_2, & \text{if } \|\mathbf{x}_i - \mathbf{x}_j\|_2 < \varepsilon \text{ or } \mathbf{x}_j \in \text{kNN}(\mathbf{x}_i) \\ \infty, & \text{otherwise} \end{cases} \quad (9)$$

12 For any two points \mathbf{x}_i and \mathbf{x}_j , the shortest path length $d_M(\mathbf{x}_i, \mathbf{x}_j)$ on the graph is computed using Dijkstra's algorithm
13 as an approximation of their geodesic distance:

$$d_M(\mathbf{x}_i, \mathbf{x}_j) = \min_{\text{path in } G} \sum_{(k,l) \in \text{path}} w_{kl} \quad (10)$$

14 Take the traffic conflict data as an example, Fig.5 illustrates the speed variation curve over 24 hours within a
15 measured area. Under normal traffic conditions, the speed remains relatively stable, whereas traffic conflicts cause
16 significant fluctuations in the speed curve. During the traffic conflict period, the regions enclosed by boxes represent
17 the conflict states identified using manifold distance. In contrast, conflict and non-conflict states distinguished based
18 on Euclidean distance are shown in blue and red, where blue represents normal traffic conditions, and red indicates
19 conflict states. It is evident that the Euclidean distance metric measures the absolute error between speed values,
20 leading to misclassifications of traffic conflict states.

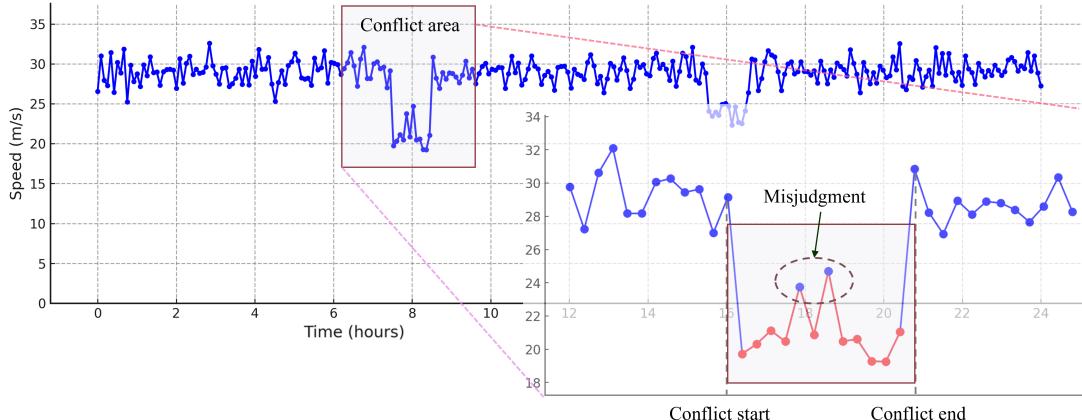


Fig. 5. Metrics for traffic conflicts based on Manifold distance and Euclidean distance.

For valid regions, direct connections between units may not always exist. However, certain grids exhibit shared characteristics, such as comparable traffic accident risks or similar geographical contexts. To capture the evolutionary patterns of traffic states across different regions, three similarity matrices are constructed: flow, speed, and occupancy similarity graphs. These matrices enable the establishment of relationships among non-connected units. By computing the manifold distances between traffic-state vectors across all road segments, we obtain an $n \times n$ geodesic distance matrix. This distance matrix is then converted into a similarity matrix using a Gaussian kernel with bandwidth h . The bandwidth h is automatically selected by minimizing the corrected Akaike Information Criterion (AICc) via a golden-section search. The detailed expressions of the kernel function and the AICc objective are provided in Appendix B (Eqs. (B.1)–(B.2)).

Conceptually, the proposed manifold-similarity graph plays a role that is related to, but distinct from, the adaptive adjacency mechanisms used in STGAT-type models. In conventional STGAT, edge weights are learned solely from instantaneous node features via attention, and the adjacency matrix is dynamically reconstructed at each time step. In MS-STGNet, the adjacency structure is instead initialized from manifold distances computed over historical traffic states, which encode long-term traffic-flow evolution and physically meaningful similarity between spatiotemporal patterns. The subsequent adaptive update in MSGNet refines this manifold-based prior rather than discarding it. This separation between a manifold-informed prior graph that reflects the geometric structure of traffic dynamics and a lightweight adaptive refinement brings two benefits: it constrains the learned graph to remain consistent with empirical traffic-state geometry, and it limits the additional per-iteration cost compared with fully attention-based dynamic graphs, keeping the overall complexity comparable to that of standard STGNN models.

5.3.2. Adaptive graph and bidirectional random walks

To incorporate potential spatial correlations into our framework, we construct three adaptive graphs by initializing the weights between nodes using similarity matrices. Singular Value Decomposition (SVD) is employed for graph initialization, and the resulting singular components are used to define an initial graph representation. We then introduce learnable left and right transformation matrices, \mathbf{M}_{lt} and \mathbf{M}_{rt} , which operate on the truncated singular vectors and singular values. A nonlinear mapping with ReLU activation and a row-wise softmax is applied to obtain a normalized adaptive adjacency matrix $\tilde{\mathbf{A}}^*$ that balances flexibility and interpretability. The complete mathematical formulation of this SVD-based initialization and adaptive update, including the definitions of \mathbf{M}_{lt} , \mathbf{M}_{rt} , and $\tilde{\mathbf{A}}^*$, is given in Appendix B (Eqs. (B.3)–(B.5)).

The process of aggregating and updating node representations adopts a multi-order bidirectional random walk (Li et al., 2017b). This approach iteratively accumulates high-order neighborhood information through forward and backward similarity matrices, as expressed below:

$$\mathbf{Z}_t^* = \sum_{k=1}^K \left((\mathbf{P}_f^*)^k \mathbf{X}^t \mathbf{W}_{k,1} + (\mathbf{P}_b^*)^k \mathbf{X}^t \mathbf{W}_{k,2} + \tilde{\mathbf{A}}^* \mathbf{X}^t \mathbf{W}_{k,3} + \mathbf{b}_k \right) \quad (11)$$

where $(\mathbf{P}_f^*)^k$ indicate forward k -th order random walk transition, \mathbf{P}_f^* describes the influence of the target node on its

1 neighboring nodes, expressed as: $\mathbf{P}_f^* = \tilde{\mathbf{A}}^*$; Similarly, $(\mathbf{P}_b^*)^k$ represent backward k -th order random walk transition,
 2 \mathbf{P}_b^* describes the influence of the neighboring nodes on the target node, expressed as: $\mathbf{P}_b^* = \tilde{\mathbf{A}}_\star^\top; \mathbf{W}_{k,1}, \mathbf{W}_{k,2}$ and $\mathbf{W}_{k,3}$
 3 are learnable weights for the k -th order neighbors; \mathbf{b}_k is bias term. The final semantic spatial features are aggregated
 4 by summing contributions from all similarity graphs:

$$\mathbf{H}_{MS}^t = \sum_{* \in \{F, S, O\}} \mathbf{Z}_t^* \quad (12)$$

5 where F, S, O represent the flow, speed, and occupancy graphs, respectively. The initial input is $\mathbf{X} \in \mathbb{R}^{T \times W \times H \times d_x}$,
 6 and the output is $\mathbf{H}_{MS} \in \mathbb{R}^{T \times W \times H \times d}$.

7 5.4. Temporal convolutional network (TCN) layer

8 Both long-term and short-term temporal observations play a crucial role in characterizing traffic conflicts. Long-
 9 term observations capture the distribution of conflicts over identical target periods in historical records, whereas short-
 10 term observations delineate the recurrent patterns and trends of conflicts—thereby posing a significant challenge for
 11 prediction models that emphasize tail-period dynamics (Bai et al., 2018). To address this issue, we devise a Temporal
 12 Convolutional Network (TCN) underpinned by dilated causal convolutional operators to extract temporal dependencies
 13 separately from heterogeneous long- and short-term sequences. Specifically, the convolution operation at time t within
 14 a dilated causal 1D-CONV layer with a dilation factor d is defined by Eq. 13.

$$\mathbf{f}_t^{l,k} * \mathbf{H}_{t-d \cdot m} = \sum_{m=0}^{C-1} \mathbf{f}^{l,k}(m) \cdot \mathbf{H}_{t-d \cdot m} \quad (13)$$

15 where C is the number of channels; d is the dilation factor; m indexes the dilation intervals; and $\mathbf{f}^{l,k} \in \mathbb{R}^C$ denotes
 16 the 1D convolution kernel of the l -th TCN layer and the k -th output channel. Each residual block comprises two 1D-
 17 CONV layers, and a skip connection is introduced by adding a block's input with its output. This converts a regular
 18 TCN block into a residual TCN block whose output is as per the given equation.

$$\begin{cases} \mathbf{H}_t^{(k)} = \text{ReLU}\left(\mathbf{f}_k^{(1)} * \text{ReLU}\left(\mathbf{f}_k^{(0)} * \mathbf{H}_t^{(k)} + \mathbf{b}_{t,k}^{(0)}\right) + \mathbf{b}_{t,k}^{(1)}\right) \\ \mathbf{H}_t^{(k)} = \mathbf{H}_t^{(k)} + \mathbf{W}_{B,k}^{(0)} * \mathbf{H}_t^{(k-1)} + \mathbf{b}_{B,k}^{(0)} \end{cases} \quad (14)$$

19 where $\mathbf{f}_k^{(0)}$ and $\mathbf{f}_k^{(1)}$ are also 1D convolution kernels, corresponding to the first and second dilated convolutions in the
 20 k -th residual TCN block, respectively. They are specific instances of the generic kernel $\mathbf{f}^{l,k}$ defined in Eq.(18), but we
 21 use superscripts (0) and (1) to distinguish the two convolutional layers within each block; $\mathbf{b}_{t,k}^{(0)}$ and $\mathbf{b}_{t,k}^{(1)}$ represent the
 22 learnable biases; and $*$ is the convolution operator. In this study, multiple temporal blocks based on TCN are employed
 23 to extract temporal features from the output results of the residual convolutional module and the manifold similarity
 24 graph module. These temporal blocks are designed to capture both short-term and long-term temporal dependencies,
 25 enabling the model to effectively learn time-series patterns within the input data.

$$\mathbf{H}_t = \text{Stack}(\mathbf{H}_t^{(1)}, \mathbf{H}_t^{(2)}, \dots, \mathbf{H}_t^{(L)}) \quad (15)$$

26 where L represents the total number of temporal blocks. The output at the last time slot are $\mathbf{H}_{\text{resT}} \in \mathbb{R}^{W \times H \times d}$ and
 27 $\mathbf{H}_{\text{MST}} \in \mathbb{R}^{W \times H \times d}$, respectively.

28 5.5. Adaptive channel fusion gate

29 When integrating two different spatiotemporal feature representations, directly combining them with equal weight-
 30 ing may fail to effectively capture heterogeneous characteristics, such as the differences between static factors (e.g.,
 31 road distribution) and dynamic factors (e.g., traffic mobility) (Zou et al., 2023a). To achieve dynamic weighted fusion,
 32 an Adaptive Channel Fusion Gate (ACFG) mechanism is designed, which dynamically assigns weights based on the
 33 semantic importance of the features. The ACFG performs weighted fusion of the two feature representations through
 34 a dynamically generated weight matrix. The formula is as follows:

$$\mathbf{H} = \Phi \odot \mathbf{H}_{\text{resT}} + (1 - \Phi) \odot \mathbf{H}_{\text{MST}} \quad (16)$$

1 where Φ is the dynamic weight matrix, representing the importance of \mathbf{H}_{resT} . Its values are constrained within the
 2 range $[0, 1]$; \odot denotes the element-wise (Hadamard) product; and \mathbf{H} is the fused feature matrix. The weight matrix
 3 Φ is generated based on the input features through the following computation:

$$\Phi = \sigma \left(\mathbf{W}_\Phi^{(0)} * \mathbf{H}_{\text{resT}} + \mathbf{W}_\Phi^{(1)} * \mathbf{H}_{\text{MST}} + \mathbf{b}_\Phi \right) \quad (17)$$

4 where $\sigma(\cdot)$ denotes the sigmoid activation function, which maps the input to a range between 0 and 1; $\mathbf{W}_\Phi^{(0)}, \mathbf{W}_\Phi^{(1)} \in$
 5 $\mathbb{R}^{1 \times 1 \times d}$ are learnable convolutional filters; and \mathbf{b}_Φ is a learnable bias term. The fused feature matrix, denoted as \mathbf{H}' ,
 6 undergoes a nonlinear transformation to extract higher-level features:

$$\mathbf{H}' = \text{ReLU} \left(\mathbf{W}_H^{(1)} * \text{ReLU} \left(\mathbf{W}_H^{(0)} * \mathbf{H} + \mathbf{b}_H^{(0)} \right) + \mathbf{b}_H^{(1)} \right) \quad (18)$$

7 where $\mathbf{W}_H^{(0)}, \mathbf{W}_H^{(1)} \in \mathbb{R}^{1 \times 1 \times d}$ are convolutional filters applied to the input feature matrix; $\mathbf{b}_H^{(0)}, \mathbf{b}_H^{(1)}$ are bias terms.

8 5.6. Loss function

9 Class imbalance is a prevalent challenge in traffic conflict classification tasks, particularly in scenarios involving
 10 minority classes with limited sample sizes. This imbalance leads to models disproportionately favoring the majority
 11 classes (non-conflict class), thereby diminishing their performance in accurately identifying minority classes (conflict
 12 class). This study amalgamates two well-established loss functions—Focal Loss and Label Distribution Aware Margin
 13 (LDAM) Loss—within a unified framework to mitigate these limitations (Sadi et al., 2022). This combination simulta-
 14 neously optimizes decision boundary margins for minority classes and prioritizes hard-to-classify samples, enhancing
 15 model robustness and overall classification performance. The formula can be expressed as:

$$\begin{cases} \mathcal{L}_{\text{Focal}} = -\alpha_t(1 - p_t)^\gamma \log(p_t) \\ \mathcal{L}_{\text{LDAM}} = -\log \frac{\exp(z_y - \Delta_y)}{\exp(z_y - \Delta_y) + \sum_{j \neq y} \exp(z_j)}, \quad \Delta_y = \frac{S}{n_y^\sigma} \\ \text{Loss}(\mathbf{Y}, \hat{\mathbf{Y}}) = \alpha \cdot \mathcal{L}_{\text{LDAM}} + \beta \cdot \mathcal{L}_{\text{Focal}} \end{cases} \quad (19)$$

16 by introducing two adjustable hyperparameters α and β , the LMF loss function can dynamically balance the contribu-
 17 tions of Focal and LDAM Losses, adapting to diverse datasets and task requirements.

18 After undergoing dynamic weighted fusion and nonlinear transformation, the resulting feature matrix $\mathbf{H}' \in$
 19 $\mathbb{R}^{W \times H \times d}$ encapsulates the essential characteristics of both feature representations. This fused matrix ensure that
 20 the model effectively integrates both static and dynamic properties. Such an approach enhances the flexibility and
 21 applicability of feature fusion, enabling the model to adaptively combine complementary information from diverse
 22 sources.

23 6. Experiments

24 6.1. Data preparation

25 In this study, the total simulation time was set to 500 hours. Traffic flow, speed, and occupancy data collected
 26 by loop detectors at unit time intervals were used as model inputs, as detailed in Table 3. To accurately simulate
 27 realistic traffic conditions, the total hourly traffic volume was randomly sampled for each hour within the range of
 28 2,500 to 10,000 vehicles. Additionally, three distinct traffic volume ranges were defined: 2,500-4,000; 4,000-7,500; and
 29 7,500-10,000, ensuring a balanced proportion of samples across these ranges during random sampling. Furthermore,
 30 five different market penetration rates (MPRs)—10%, 30%, 50%, 70%, and 90%—were employed to reflect the model’s
 31 performance under varying mixed traffic conditions.

32 The loop detector data were collected at 30-second intervals and aggregated into 5-minute granularity, resulting
 33 in 6,000 time slices. Data from 5 to 20 minutes before a conflict were identified as potential precursors for predicting
 34 conflicts (Li et al., 2020; Kamel et al., 2023, 2024). Consequently, the target time step for model training was set to 1,
 35 with the most recent 4 preceding time steps (20 minutes) used as inputs for conflict prediction. During the simulation,
 36 this process generated a total of $4 \times 27 \times 6,000 = 648,000$ traffic data samples. The number of samples labeled as
 37 traffic conflicts was 24,087 (10% MPR), 24,561 (30% MPR), 19,432 (50% MPR), and 17,146 (70% MPR), and 16,580

Table 3

Variable descriptive statistics.

Variable (Unit)	Description	Distribution
Volume (vehicles)	Volume in five minutes	Min: 1.00, mean: 81.25, max: 230.00
Speed_mean (mph)	Average speed of the current segment in five minutes	Min: 8.74, mean: 65.24, max: 120.00
Occupancy_mean (%)	Average lane occupancy in five minutes	Min: 0.50, mean: 10.08, max: 73.58

1 (90% MPR). This yielded a conflict-to-non-conflict sample ratio of approximately 1:26, 1:25, 1:32, 1:36, and 1:38,
 2 highlighting the presence of significant zero-inflation in the data. This imbalance underscores the applicability of our
 3 proposed MS-STGNet model in handling rare-event scenarios effectively.

4 **6.2. Experimental setup**

5 PyTorch framework are utilized to construct all experiments, and the training, validation, and testing process is
 6 executed on a platform with Intel(R) Xeon(R) Gold 6336Y CPU and NVIDIA RTX 4090 GPU-24 GB card. The
 7 dataset is split into training, validation, and testing sets in a 6:2:2 ratio. During the training process, the maximum
 8 number of epochs is set to 200, with a batch size of 32 and a learning rate of 0.0005. The Adam optimizer is employed
 9 to update model weights. Model performance is evaluated on the validation set after each epoch, and the weights are
 10 saved whenever a reduction in loss is observed. Additionally, an early stopping mechanism with a patience value of
 11 10 is applied to mitigate overfitting. If the validation loss remains unchanged for 10 consecutive epochs, the training
 12 process terminates early. To reduce the impact of randomness and evaluate the stability of each method, all models are
 13 trained and evaluated five times with different random seeds orders. The detailed parameter settings in each module
 14 are summarized in Table 4.

15 **6.3. Evaluation metrics**

16 To evaluate the classification performance of the MS-STGNet model, we employed metrics commonly used in
 17 conflict risk analysis, including recall, accuracy, and false alarm rate (FAR) (Li et al., 2020). Additionally, the area
 18 under the ROC curve (AUC) was used to assess the performance of the binary classifier; the G-mean (geometric mean)
 19 served as an indicator of a model's performance on the minority class. The descriptions are as follows:

$$\text{Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}} \quad (20)$$

$$\text{Accuracy} = \frac{\text{True Positives (TP)} + \text{True Negatives (TN)}}{\text{Total Samples}} \quad (21)$$

$$\text{FAR} = \frac{\text{False Positives (FP)}}{\text{True Negatives (TN)} + \text{False Positives (FP)}} \quad (22)$$

$$\text{G-Mean} = \sqrt{\text{Recall} \times (1 - \text{FAR})} \quad (23)$$

$$\text{AUC} = \int_0^1 \text{Recall}(\text{FAR}) d(\text{FAR}) \quad (24)$$

24 The model's estimated outputs are transformed into posterior probabilities of conflict occurrence using the sigmoid
 25 function, with values ranging from 0 to 1. To evaluate the classification accuracy of the model, a threshold (or cutoff
 26 point) must be selected for binary classification. In this study, a fixed threshold of 0.5 was adopted, a widely used
 27 standard in the literature (Abdel-Aty and Pande, 2005; Yu et al., 2020; Jiang et al., 2020).

Table 4

Model hyperparameter.

Hyperparameter type	Hyperparameter	Values
Embedding/Feed/Residual	Channel	64/64
	Fitter size	1 × 1 / 1 × 1
	Number of layers	2
MSGNet	Feed	1
	Bidirectional walks orders	2
ResNet	Channel	64/64
	Fitter size	3 × 3 / 3 × 3
	Number of layers	2
	Padding	1 × 1
	Residual	1
TCN	Channel	64/64/64
	Fitter size	2 × 2 / 2 × 2 / 2 × 2
	Number of blocks	3
	Dilation size	0/2/4
Adaptive fusion gate	Padding	1/1/1
	Channel	64
	Fitter size	1 × 1
FC	Number of layers	1
	Channel	64/1
	Fitter size	1 × 1 / 1 × 1
Loss function	Number of layers	2
	α_t	0.5
	γ	2
	S	3
	σ	0.7
	α	0.5
Other hyperparameters	β	0.5
	Training optimizer	Adam
	Decay rate	0.9
	Batch size	32
	Learning rate	0.0005
	Dropout	0
	Epochs	200
	Patience	10

6.4. Baseline models for conflict prediction

In evaluating the proposed model, we carefully selected baseline models that allow for a comprehensive assessment of the proposed model's capabilities. Referring to the review by Ali et al. (2024) on accident prediction studies, current machine learning models for accident prediction can be broadly categorized into three types: traditional machine learning models, neural network-based models, and graph-structured models. From each category, we selected two representative models as our baseline, including: Support Vector Machines (SVM) and XGBoost for traditional machine learning methods; CNN and LSTM-CNN for recent neural network-based architectures; Spatiotemporal Graph Convolutional Networks (STGCN) and Spatiotemporal Graph Attention Networks (STGAT) for advanced spatiotemporal graph deep learning techniques. The details of the baseline models are as follows:

- **SVM:** A supervised statistical learning method applied to predict accident occurrence, accident frequency and injury severity (Yu and Abdel-Aty, 2013, 2014).
- **XGBoost:** A gradient boosting framework learns from weak classifiers and adjusts/increases the weight of in-

1 correctly classified samples, demonstrates superior performance in predicting accident severity (Goswamy et al.,
 2 2023).

- 3 • **CNN:** A deep learning approach focused on capturing spatial patterns, particularly effective for grid-like structured data. Hu et al. (2020) indicates that CNN can properly identify the important features contributing to risk level decisions such as signal light, traffic flow and vehicle start/brake frequency.
- 4 • **LSTM-CNN:** A hybrid model that combines the temporal sequence modeling capability of LSTMs with the spatial feature extraction strengths of CNNs. This hybrid architecture has been demonstrated to achieve superior predictive performance in accident detection and model transferability (Zhang and Abdel-Aty, 2022).
- 5 • **STGCN:** A graph-based deep learning model leverages graph convolution to capture spatial dependencies and 1D convolution to model temporal correlations effectively, It has been proven to more effectively capture the spatiotemporal patterns in traffic accident data for crash prediction (Yu et al., 2021).
- 6 • **STGAT:** An attention-based graph model that dynamically balances the importance of spatial and temporal interactions to enhance feature representation. Additionally, the attention mechanism models the influence of various factors on traffic accident occurrences, enabling the identification of key variables contributing to crashes
 7 (Wu et al., 2023).

16 6.5. Performance comparison

17 Table 5 summarizes the performance metrics of the proposed MS-STGNet and all baseline models for conflict prediction under five different CAV penetration rates (10%, 30%, 50%, 70%, and 90%). Improvements over suboptimal models are indicated by upward arrows. Overall, MS-STGNet demonstrates superior performance across most comparison metrics. These enhancements highlight the significant impact of incorporating the manifold similarity matrix and the carefully designed modules within the model on the accuracy and stability of conflict prediction outcomes. To further assess cross-run stability, each entry in Table 5 is reported as the mean \pm standard deviation over five independent runs with different random seeds. Statistical tests across the five independent runs show that the improvements of all reported metrics and penetration-rate scenarios are statistically significant at the 5% level ($p < 0.05$).

18 Traffic conflict prediction remains a significant challenge, particularly in distinguishing between non-conflict and conflict states. Traditional machine learning algorithms, such as SVM and XGBoost, struggle with this task compared to deep learning approaches. For example, under a 30% penetration rate, the recall rates of SVM and XGBoost were 19 23% and 17.8% lower, respectively, than those of the proposed MS-STGNet. Additionally, their false alarm rates 20 increased by 27.9% and 20.0%, AUC values decreased by 24.3% and 18.9%, and accuracy was reduced by 26.4% and 21 20.5%. These results emphasize the importance of extracting nonlinear correlations for traffic conflict prediction.

22 The introduction of deep learning methods significantly improved model performance. CNN and LSTM-CNN 23 outperformed SVM and XGBoost across all metrics, demonstrating the importance of capturing spatial dependencies 24 and temporal correlations in conflict prediction. However, deep learning methods relying on CNNs to capture spatial 25 dependencies face a notable limitation: they cannot model spatial similarities in unconnected grid fields. This highlights 26 the advantage of leveraging graph neural networks (GNNs), such as STGCN and STGAT, to model semantic 27 spatial dependencies, further enhancing performance. For instance, under a 30% penetration rate, STGAT and STGCN 28 improved recall rates by 4.0% and 3.9%, reduced false alarm rates by 2.4% and 3.0%, increased AUC values by 2.9% 29 and 1.5%, and improved accuracy by 4.4% and 3.2%, respectively, compared to LSTM-CNN. These results underscore 30 the advanced capability of utilizing the inherent graph structure of road networks to extract spatial dependencies related 31 to conflict risks. GNNs are particularly well-suited for capturing complex relationships between road segments, 32 integrating heterogeneous road features, and learning network-wide patterns while retaining local details. Comparatively, GAT-based models often outperform GCN models by incorporating predefined adjacency matrices embedded 33 with spatial proximity and contextual similarity, better representing spatial dependencies.

34 Building on prior advancements in graph-based models, the proposed MS-STGNet model demonstrated robust performance 35 across all penetration rate scenarios. For instance, under a 50% penetration rate, MS-STGNet outperformed 36 the next-best models by 4.9% in recall, reduced false alarm rates by 3.3%, improved AUC by 5.3%, and increased 37 accuracy by 3.3%. Notably, as shown in Table 5, MS-STGNet achieved a significant reduction in false alarm rates, 38 with improvements of 23.9%, 24.0%, and 23.8% under 50%, 70%, and 90% penetration rates, respectively. This improvement 39 can be attributed to the manifold similarity module, which reduces misjudgments in conflict-prone areas of 40 traffic flow—a point further analyzed in subsequent sections.

Table 5

Performance of Different Models on Datasets.

Penetration rates	Metric	SVM	XGBoost	CNN	LSTM-CNN	STGCN	STGAT	MS-STGNet
10%	Recall	0.531 ± 0.049	0.577 ± 0.037	0.713 ± 0.031	0.726 ± 0.024	0.766 ± 0.011	0.782 ± 0.019	0.797 ^{11.92%} ± 0.014
	False alarm rate	0.440 ± 0.047	0.413 ± 0.038	0.206 ± 0.029	0.201 ± 0.017	0.175 ± 0.012	0.165 ± 0.009	0.150 ^{19.09%} ± 0.009
	AUC	0.588 ± 0.049	0.632 ± 0.039	0.758 ± 0.034	0.769 ± 0.021	0.790 ± 0.020	0.807 ± 0.024	0.824 ^{12.11%} ± 0.016
	Accuracy	0.581 ± 0.039	0.652 ± 0.055	0.788 ± 0.029	0.803 ± 0.030	0.830 ± 0.014	0.829 ± 0.017	0.855 ^{13.01%} ± 0.011
	G-mean	0.543 ± 0.067	0.581 ± 0.053	0.745 ± 0.037	0.769 ± 0.026	0.793 ± 0.021	0.803 ± 0.016	0.820 ^{12.12%} ± 0.017
	Recall	0.578 ± 0.040	0.630 ± 0.036	0.742 ± 0.022	0.738 ± 0.028	0.777 ± 0.016	0.778 ± 0.013	0.808 ^{13.86%} ± 0.010
30%	False alarm rate	0.417 ± 0.049	0.338 ± 0.054	0.194 ± 0.024	0.173 ± 0.013	0.143 ± 0.013	0.149 ± 0.015	0.138 ^{13.50%} ± 0.013
	AUC	0.596 ± 0.047	0.650 ± 0.034	0.757 ± 0.027	0.781 ± 0.025	0.796 ± 0.020	0.810 ± 0.017	0.839 ^{13.58%} ± 0.007
	Accuracy	0.592 ± 0.065	0.651 ± 0.048	0.781 ± 0.033	0.789 ± 0.018	0.821 ± 0.023	0.833 ± 0.021	0.856 ^{12.76%} ± 0.015
	G-mean	0.592 ± 0.041	0.644 ± 0.046	0.773 ± 0.039	0.775 ± 0.020	0.815 ± 0.014	0.816 ± 0.015	0.831 ^{11.84%} ± 0.012
	Recall	0.564 ± 0.047	0.593 ± 0.054	0.767 ± 0.028	0.788 ± 0.031	0.803 ± 0.023	0.828 ± 0.019	0.877 ^{15.92%} ± 0.009
	False alarm rate	0.417 ± 0.047	0.332 ± 0.039	0.181 ± 0.021	0.165 ± 0.019	0.139 ± 0.018	0.138 ± 0.015	0.105 ^{123.91%} ± 0.017
50%	AUC	0.580 ± 0.044	0.672 ± 0.042	0.790 ± 0.036	0.802 ± 0.026	0.823 ± 0.022	0.833 ± 0.024	0.886 ^{16.36%} ± 0.013
	Accuracy	0.590 ± 0.035	0.650 ± 0.053	0.794 ± 0.032	0.830 ± 0.015	0.852 ± 0.019	0.857 ± 0.016	0.890 ^{13.85%} ± 0.008
	G-mean	0.563 ± 0.050	0.642 ± 0.045	0.789 ± 0.027	0.813 ± 0.030	0.826 ± 0.014	0.843 ± 0.011	0.887 ^{15.22%} ± 0.011
	Recall	0.571 ± 0.038	0.617 ± 0.051	0.759 ± 0.025	0.749 ± 0.028	0.770 ± 0.014	0.789 ± 0.018	0.816 ^{13.42%} ± 0.012
	False alarm rate	0.427 ± 0.050	0.329 ± 0.047	0.168 ± 0.030	0.170 ± 0.022	0.141 ± 0.017	0.125 ± 0.012	0.095 ^{124.00%} ± 0.010
	AUC	0.576 ± 0.037	0.673 ± 0.050	0.782 ± 0.023	0.781 ± 0.029	0.816 ± 0.020	0.822 ± 0.015	0.860 ^{14.62%} ± 0.015
70%	Accuracy	0.589 ± 0.046	0.668 ± 0.057	0.809 ± 0.035	0.801 ± 0.011	0.830 ± 0.013	0.836 ± 0.020	0.898 ^{17.42%} ± 0.007
	G-mean	0.588 ± 0.051	0.639 ± 0.045	0.802 ± 0.028	0.795 ± 0.027	0.811 ± 0.022	0.828 ± 0.009	0.860 ^{13.86%} ± 0.014
	Recall	0.597 ± 0.053	0.622 ± 0.050	0.782 ± 0.034	0.770 ± 0.021	0.783 ± 0.023	0.809 ± 0.010	0.819 ^{11.24%} ± 0.008
	False alarm rate	0.388 ± 0.038	0.352 ± 0.041	0.170 ± 0.024	0.147 ± 0.027	0.130 ± 0.016	0.122 ± 0.022	0.093 ^{123.77%} ± 0.011
	AUC	0.595 ± 0.040	0.658 ± 0.031	0.793 ± 0.038	0.786 ± 0.012	0.821 ± 0.011	0.832 ± 0.013	0.860 ^{13.37%} ± 0.009
	Accuracy	0.591 ± 0.063	0.682 ± 0.053	0.812 ± 0.020	0.835 ± 0.024	0.857 ± 0.028	0.873 ± 0.018	0.896 ^{12.63%} ± 0.016
90%	G-mean	0.600 ± 0.048	0.635 ± 0.035	0.798 ± 0.031	0.802 ± 0.018	0.822 ± 0.019	0.839 ± 0.025	0.863 ^{12.86%} ± 0.013
	Recall	0.597 ± 0.053	0.622 ± 0.050	0.782 ± 0.034	0.770 ± 0.021	0.783 ± 0.023	0.809 ± 0.010	0.819 ^{11.24%} ± 0.008
	False alarm rate	0.388 ± 0.038	0.352 ± 0.041	0.170 ± 0.024	0.147 ± 0.027	0.130 ± 0.016	0.122 ± 0.022	0.093 ^{123.77%} ± 0.011
	AUC	0.595 ± 0.040	0.658 ± 0.031	0.793 ± 0.038	0.786 ± 0.012	0.821 ± 0.011	0.832 ± 0.013	0.860 ^{13.37%} ± 0.009
	Accuracy	0.591 ± 0.063	0.682 ± 0.053	0.812 ± 0.020	0.835 ± 0.024	0.857 ± 0.028	0.873 ± 0.018	0.896 ^{12.63%} ± 0.016
	G-mean	0.600 ± 0.048	0.635 ± 0.035	0.798 ± 0.031	0.802 ± 0.018	0.822 ± 0.019	0.839 ± 0.025	0.863 ^{12.86%} ± 0.013

Because the task is a binary conflict/non-conflict prediction problem on a large-scale dataset, the standard deviations across runs are generally small for all models. Nevertheless, the reported mean \pm standard deviation helps to reveal relative robustness: MS-STGNet maintains consistent advantages over STGCN and STGAT across different penetration rates, and in most cases exhibits comparable or slightly lower variation in key metrics. This indicates that the improvements of MS-STGNet are not due to a single favourable initialization but are reproducible under different random seeds. In addition to the simulation-based evaluation on the 14 km mixed CAV–HDV corridor, we further conduct supplementary tests on real-world freeway traffic from the FHWA NGSIM program, to examine the transferability of the proposed framework (see Appendix D). The results show that MS-STGNet still outperforms the baseline models and effectively controls false alarms, providing supporting evidence for its robustness and extensibility beyond the original simulated setting.

These empirical results also clarify how MS-STGNet differs in practice from STGAT-type adaptive graph models. Although both approaches employ graph-based representations, STGAT relies on feature-driven attention to construct adjacency at each time step, which can be sensitive to local fluctuations in highly imbalanced conflict datasets. By contrast, MS-STGNet constrains the adaptive graph updates within a manifold-similarity prior derived from historical traffic states. As the market penetration of CAVs increases and pronounced speed separation emerges, this manifold-informed prior helps the model avoid spuriously high conflict probabilities in non-conflict regions, leading to consistently lower false alarm rates and more stable performance across all penetration scenarios. In this sense, our findings are consistent with previous studies showing that graph-based spatiotemporal models such as STGCN and STGAT outperform traditional machine-learning and sequence models in traffic prediction tasks, while further extending them by explicitly incorporating a manifold-based state similarity prior into the adaptive graph learning process. At the same time, our results complement recent manifold-learning approaches for traffic state analysis by demonstrating that manifold-informed similarity can be embedded into deep spatiotemporal graph networks to improve conflict prediction in mixed CAV–HDV freeway traffic.

6.6. Computation cost

In real-world deployment, predictive accuracy is the primary requirement for traffic safety applications, while the hardware cost of the deployed model constitutes a secondary but still crucial consideration for practical implementation. To highlight the computational overhead of different approaches, Table 6 reports three indicators under five CAV penetration-rate scenarios: GPU-MUT (peak GPU memory usage during training), GPU-MUI (peak GPU memory usage during inference), and the number of trainable parameters. For the classical machine-learning baselines (SVM and XGBoost), GPU-based indicators are omitted (“–”) because they are trained and executed on CPU and their memory footprint is negligible compared with deep models in our setting.

Several observations can be made from Table 6. First, among the deep learning baselines, STGCN consistently has the largest parameter count and highest GPU memory usage, with STGAT slightly smaller but still noticeably heavier than CNN and LSTM-CNN. For example, at a 50% penetration rate, STGCN and STGAT require 479,816 and 426,572 parameters, respectively, and their GPU-MUT values reach 4,497 MiB and 4,681 MiB. By contrast, the proposed MS-STGNet uses fewer parameters than both graph-based baselines (395,428 at 50% penetration) and reduces peak GPU memory by roughly 10–15% in training (e.g., 4,059 MiB versus 4,497–4,681 MiB) and 15–25% in inference (e.g., 2,710 MiB versus 3,216–3,587 MiB), while still incorporating a manifold-similarity module and adaptive fusion. Compared with CNN and LSTM-CNN, MS-STGNet understandably incurs moderately higher GPU memory usage due to the additional graph operations, but remains in the same order of magnitude and does not introduce prohibitive overhead.

Overall, these results indicate that MS-STGNet achieves superior predictive performance (as shown in Table 5) with a computational cost that is only modestly higher than conventional CNN-based models and clearly lower than that of STGCN and STGAT. This suggests that the proposed architecture strikes a reasonable balance between accuracy and efficiency, making it suitable for deployment in practical mixed CAV–HDV conflict prediction systems. We do not report wall-clock training or inference time, as such measurements are highly dependent on specific hardware, software environments, and background system load; instead, we focus on parameter counts and GPU memory usage, which provide hardware-agnostic indicators of computational complexity.

Table 6

The computational performance of different models on dataset.

Penetration rates	Metric	SVM	XGBoost	CNN	LSTM-CNN	STGCN	STGAT	MS-STGNet
10%	GPU-MUT	—	—	4,333MiB	4,443MiB	5,574MiB	5,802MiB	5,031MiB
	GPU-MUI	—	—	2,283MiB	2,799MiB	4,446MiB	3,986MiB	3,359MiB
	Parameters	—	—	298,742	346,251	594,758	528,759	490,154
30%	GPU-MUT	—	—	4,419MiB	4,530MiB	5,684MiB	5,917MiB	5,130MiB
	GPU-MUI	—	—	2,328MiB	2,854MiB	4,534MiB	4,065MiB	3,425MiB
	Parameters	—	—	304,621	353,064	606,462	539,164	499,800
50%	GPU-MUT	—	—	3,496MiB	3,584MiB	4,497MiB	4,681MiB	4,059MiB
	GPU-MUI	—	—	1,842MiB	2,258MiB	3,587MiB	3,216MiB	2,710MiB
	Parameters	—	—	241,008	279,335	479,816	426,572	395,428
70%	GPU-MUT	—	—	3,085MiB	3,162MiB	3,968MiB	4,130MiB	3,581MiB
	GPU-MUI	—	—	1,625MiB	1,992MiB	3,165MiB	2,838MiB	2,391MiB
	Parameters	—	—	212,656	246,474	423,370	376,390	348,909
90%	GPU-MUT	—	—	2,983MiB	3,058MiB	3,837MiB	3,994MiB	3,463MiB
	GPU-MUI	—	—	1,572MiB	1,927MiB	3,061MiB	2,744MiB	2,312MiB
	Parameters	—	—	205,636	238,338	409,394	363,965	337,392

6.7. Ablation study

Ablation experiments were conducted to systematically evaluate the importance of specific components within the model. This process involved progressively removing these components to assess their impact on overall performance. Table 7 and Appendix B provide a detailed statistical representation of the ablation experiment results. The configurations in the table are described as: **w/o Manifold-similarity**: removes the traffic state matrix based on manifold distances and replaces it with a similarity computation using the Jensen-Shannon divergence method (Lin, 1991). **w/o Adaptive Graphs**: eliminates the adaptive correlation matrix used for dynamic graph construction. **w/o MSG**: ablates the entire Manifold-Similarity Graph module, removing its contribution entirely. **w/o ResNet**: removes the Residual Convolutional module. **w/o TCN**: excludes the Temporal Convolutional Network, replacing it with Gated Recurrent Units (GRU) and attention mechanisms for temporal feature extraction. **w/o Adaptive Fusion Gate**: removes the Adaptive Fusion Gate mechanism and uses a simple addition operation to combine semantic spatiotemporal features with geographical spatiotemporal features. The results analyses are primarily summarized as follows:

- Across different scenarios, the removal of the manifold similarity module (i.e., the **w/o Manifold-similarity** and **w/o MSG** variants) significantly degraded model performance. This further underscores the critical role of the proposed manifold similarity approach in traffic conflict prediction. Traditional methods (such as Jensen-Shannon divergence method in **w/o Manifold-similarity**) for measuring similarity using traffic flow, speed, and occupancy data suffer from substantial limitations. The stop-and-go wave phenomena prevalent in traffic flows make it challenging for models to distinguish between different traffic states. Moreover, these errors are amplified as the penetration rate increases, indicating that the sparsity of traffic conflict data exacerbates the robustness challenges in identifying rare events. The proposed model architecture effectively addresses these deficiencies, enhancing the model's ability to capture nuanced traffic dynamics and improving its robustness in predicting small-sample events. This highlights the importance of incorporating advanced similarity metrics, such as the manifold similarity matrix, in traffic conflict prediction tasks.
- We utilized the manifold similarity method to measure the similarity of node attributes and establish proximity relationships between nodes. While this approach is highly effective, it also has limitations. Predefined graphs

Table 7

Performance comparison in ablation experiments.

Penetration rates	Metric	w/o Manifold- similarity	w/o Adaptive Graphs	w/o MSG	w/o ResNet	w/o TCN	w/o Adaptive Fusion Gate	MS- STGNet
10%	Recall	0.775	0.787	0.788	0.791	0.797	0.790	0.794
	False alarm rate	0.165	0.160	0.167	0.157	0.153	0.150	0.149
	AUC	0.794	0.762	0.781	0.814	0.829	0.801	0.822
	Accuracy	0.821	0.827	0.805	0.837	0.857	0.847	0.852
	G-mean	0.804	0.813	0.810	0.817	0.822	0.820	0.822
30%	Recall	0.781	0.785	0.773	0.804	0.817	0.792	0.806
	False alarm rate	0.158	0.153	0.161	0.143	0.140	0.148	0.139
	AUC	0.801	0.813	0.784	0.829	0.842	0.826	0.837
	Accuracy	0.826	0.837	0.814	0.845	0.870	0.852	0.859
	G-mean	0.811	0.819	0.806	0.828	0.841	0.819	0.833
50%	Recall	0.842	0.856	0.837	0.864	0.868	0.870	0.874
	False alarm rate	0.132	0.128	0.135	0.112	0.117	0.115	0.107
	AUC	0.854	0.867	0.848	0.879	0.877	0.881	0.884
	Accuracy	0.863	0.869	0.857	0.886	0.894	0.890	0.892
	G-mean	0.843	0.854	0.842	0.877	0.870	0.873	0.884
70%	Recall	0.776	0.792	0.772	0.808	0.805	0.812	0.814
	False alarm rate	0.135	0.115	0.137	0.105	0.107	0.101	0.096
	AUC	0.819	0.834	0.813	0.849	0.843	0.847	0.859
	Accuracy	0.867	0.884	0.859	0.884	0.876	0.895	0.900
	G-mean	0.817	0.837	0.816	0.853	0.848	0.857	0.858
90%	Recall	0.787	0.790	0.775	0.810	0.808	0.814	0.818
	False alarm rate	0.134	0.120	0.132	0.103	0.105	0.107	0.095
	AUC	0.827	0.831	0.810	0.852	0.847	0.854	0.861
	Accuracy	0.871	0.886	0.863	0.877	0.877	0.891	0.897
	G-mean	0.826	0.834	0.821	0.852	0.851	0.853	0.860

are insufficient to capture comprehensive spatial dependency information, and their indirect relevance to the prediction task can introduce significant biases. Removing the adaptive graph component from MS-STGNet (i.e., the **w/o Adaptive Graphs** variant) resulted in a performance decline across all metrics. These results suggest that the adaptive graph compensates for the weaknesses of predefined similarity matrices and provides valuable insights that could benefit other traffic prediction tasks. Additionally, the bidirectional relationships between nodes are a critical factor. In the **w/o MSG** variant, removing the bidirectional random walks led to a notable performance drop. This indicates that the similarity from region i to its neighbor j is not necessarily identical to the similarity from j to i .

- It is worth noting that replacing TCN with GRU and attention mechanisms (i.e., the **w/o TCN** variant) did not yield better performance compared to MS-STGNet. For instance, under the 30% penetration rate scenario, removing TCN surprisingly improved model performance. However, in the 50% and 70% penetration rate scenarios, TCN consistently outperformed the alternative, indicating that the design of TCN provides stronger stability and adaptability across broader scenarios. Additionally, using the adaptive fusion gate mechanism to determine the weights assigned to each module in the final prediction is critical. The overall model performance declined when the adaptive fusion gate was replaced with a simple addition operation (i.e., the **w/o Adaptive Fusion**

1 Gate variant). This underscores the importance of effectively combining heterogeneous spatiotemporal features
 2 in traffic conflict prediction.

3 6.8. Posterior probability analyses

4 To gain a deeper understanding of the model's ability to discriminate between conflict and non-conflict events, and
 5 to examine how this capability evolves with varying market penetration rates, Fig.6 presents scatter plots of posterior
 6 probabilities. In each panel, negative and positive samples occupy two vertical bands (negative samples on the left;
 7 positive samples on the right). Within each band, correctly classified instances and misclassifications are denoted by
 8 blue and red markers, respectively, with a dashed horizontal line at the 0.5 threshold. Consequently, the proportion
 9 of markers in the upper-right quadrant corresponds to the recall rate, while that in the upper-left quadrant denotes the
 10 false alarm rate (FAR). Given that conflict risk analysis demands high sensitivity at low FARs (Hossain et al., 2019), an
 11 optimal model will maximize the density of points in the upper-right region while minimizing those in the upper-left.
 12 More precisely, superior discriminative performance is reflected by a pronounced separation in posterior-probability
 13 distributions between the negative and positive classes. Above and to the right of the scatter plot, histograms depict
 14 the marginal distributions of predicted probabilities for negative (top) and positive (right) samples, respectively. For
 15 comparison, we include the STGCN and STGAT architectures, both of which demonstrated strong performance in our
 16 preliminary experiments, alongside our proposed MS-STGNet model.

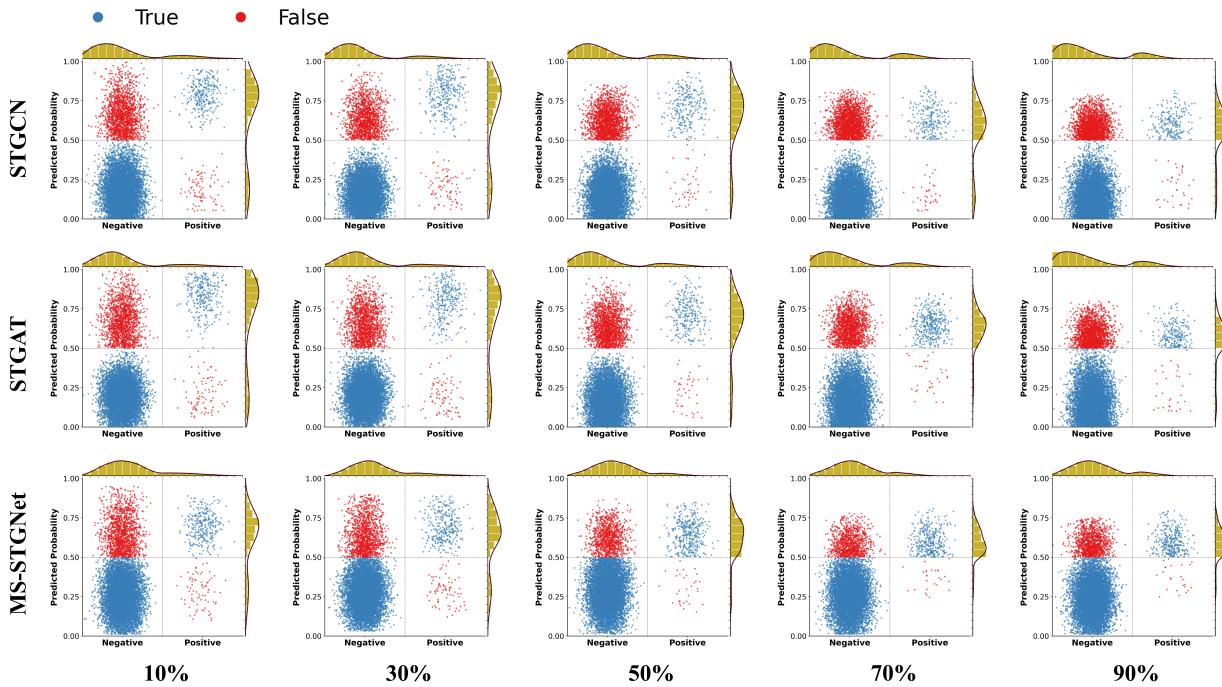


Fig. 6. Predicting probability values for STGCN, STGAT, and MS-STGNet under different market penetration rates.

17 From the results illustrated in Fig.6, it is apparent that at lower market-penetration levels (10% and 30%), the pos-
 18 terior probability estimates for negative and positive samples are distinctly dispersed on opposite sides of the decision
 19 threshold, thereby facilitating precise class separation. Under these conditions, the marginal distributions of predicted
 20 probabilities for both classes remain largely consistent across the STGCN, STGAT, and MS-STGNet models. How-
 21 ever, as penetration increases to 50%, performance begins to diverge: MS-STGNet exhibits a markedly lower density
 22 of false positives compared with the other two architectures—a trend that persists at higher penetration rates (70% and
 23 90%). These findings indicate that our proposed framework achieves a reduced FAR, corroborating the characteristics
 24 identified in earlier sections and validating the incorporation of the manifold-similarity module within our predictive
 25 model design.

26 Additionally, we observed that as market penetration increases, the probability distributions produced by the
 27 STGCN and STGAT models become increasingly skewed towards lower values, indicating a growing tendency to

1 classify samples as belonging to the negative class. We attribute this effect to the rising ratio of non-conflict to conflict
 2 events: as penetration rates climb, the prevalence of negative instances increases, exacerbating dataset imbalance and
 3 biasing model outputs downward. Although this pattern is also evident in our MS-STGNet framework, MS-STGNet
 4 yields more balanced probability distributions across varying penetration levels, and thus across differing class propor-
 5 tions. Consequently, its 0.5 decision threshold more reliably separates conflict and non-conflict events, demonstrating
 6 superior stability.

7 In summary, the proposed MS-STGNet framework not only enhances the accuracy of traffic-conflict prediction
 8 but also delivers superior robustness on imbalanced datasets, thereby yielding a marked improvement in overall model
 9 performance.

10 6.9. Spatiotemporal heat map analysis

11 In Section 5.3.1, we observed that using absolute speed error to assess traffic state similarity can inadvertently
 12 amplify model misclassifications. To further examine our model's predictive accuracy and stability under varying
 13 traffic conditions, Fig. 7 visualizes the speed heatmap of different roadway segments at multiple market penetration
 14 levels. We also compared the predicted result between STGCN, STGAT, and our proposed MS-STGNet. It is important
 15 to note that the conflict predictions correspond to segments where the output of the sigmoid function exceeds 0.5,
 16 indicating areas predicted to experience conflicts in this study.

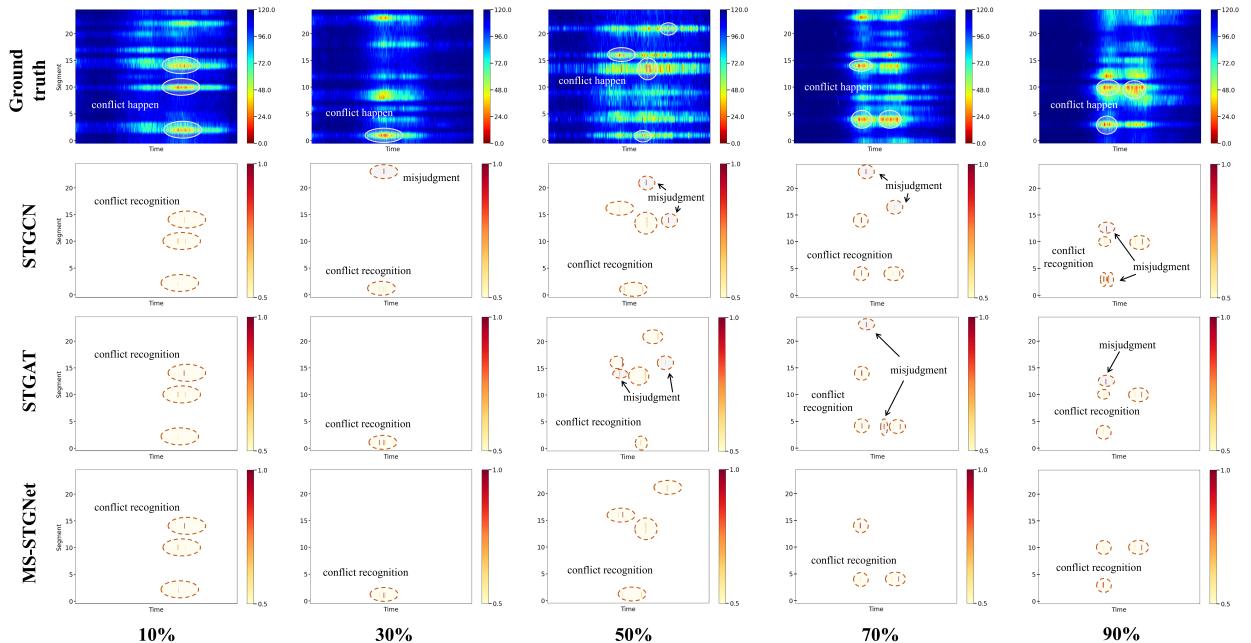


Fig. 7. Speed heatmap and conflict recognition result for STGCN, STGAT, and MS-STGNet under different market penetration rates.

17 At low penetration rates (10% and 30%), all three architectures yield nearly identical predictions, with STGCN ex-
 18 hibiting only a few false positives. However, as penetration climbs to 50%, 70%, and 90%, STGCN and STGAT mani-
 19 fest a pronounced increase in false alarms, and even a small number of missed conflict events. In contrast, MS-STGNet
 20 maintains consistently strong performance across every penetration scenario, reliably capturing and identifying traffic
 21 conflicts in heterogeneous traffic environments. The speed-fluctuation plots delineate a clear trajectory of traffic-state
 22 transitions, demonstrating that the progression from non-conflict to conflict conditions is inherently gradual. The ac-
 23 companying heatmaps further accentuate the characteristic stop-and-go oscillatory pattern—alternating between high
 24 and low speeds, that injects noise into the evolution of traffic states, thereby elevating the risk of misclassification by
 25 predictive models, a challenge previously identified. While both STGCN and STGAT architectures exhibit instability
 26 under these transitional dynamics, our proposed MS-STGNet framework integrates a manifold-similarity module to
 27 attenuate such perturbations, significantly reducing the incidence of erroneous predictions, particularly in high penetra-

tion scenarios. Despite the noisy fluctuations, MS-STGNet consistently localizes conflict events with high precision, underscoring its superior conflict detection capability and robustness within mixed-traffic environments.

In traffic conflict prediction, recall and false alarm rates are particularly critical metrics as they directly impact the practical utility and safety of predictive models. The experimental results highlight the high predictive performance of MS-STGNet and its practical value in real-world applications. Additionally, the comparison across the three penetration rate scenarios reveals a more pronounced improvement in false alarm rates under high penetration rates. This may be due to the increased regularity of traffic flow with more CAVs, although a more likely reason is the varying sparsity of event samples across the scenarios.

6.10. Impact of speed separation on conflict prediction

In our preceding analysis, we observed that as market penetration increases, disparities in false alarm rate performance across models become increasingly pronounced. We posited that employing absolute speed error to assess traffic-state similarity under high penetration may inadvertently amplify misclassifications. To further substantiate this hypothesis, we selected a segment of approximately 250 meters of an on-ramp merging scenario to illustrate the position-velocity trajectories of vehicles from both the HDV and CAV groups (as shown in Fig.8). Compared to the main highway, the merging scenario on the ramp exhibits more pronounced fluctuations and oscillations in vehicle speed, which facilitates a clearer observation of the differences between the two groups. It is evident that the CACC platoon formed by CAVs exhibits smooth trajectory profiles, with anticipatory deceleration upon obstacle approach. In contrast, HDV trajectories display markedly greater oscillation amplitudes and frequencies, indicative of systemic instability. These speed perturbations propagate and evolve over time and distance, culminating in pronounced speed separation within the traffic flow, usually represented as enlarged absolute speed errors. The resultant velocity differentials arising from HDV–CAV interactions readily predispose the predictive models to erroneous conflict judgments, thereby corroborating our earlier analysis and assertions.

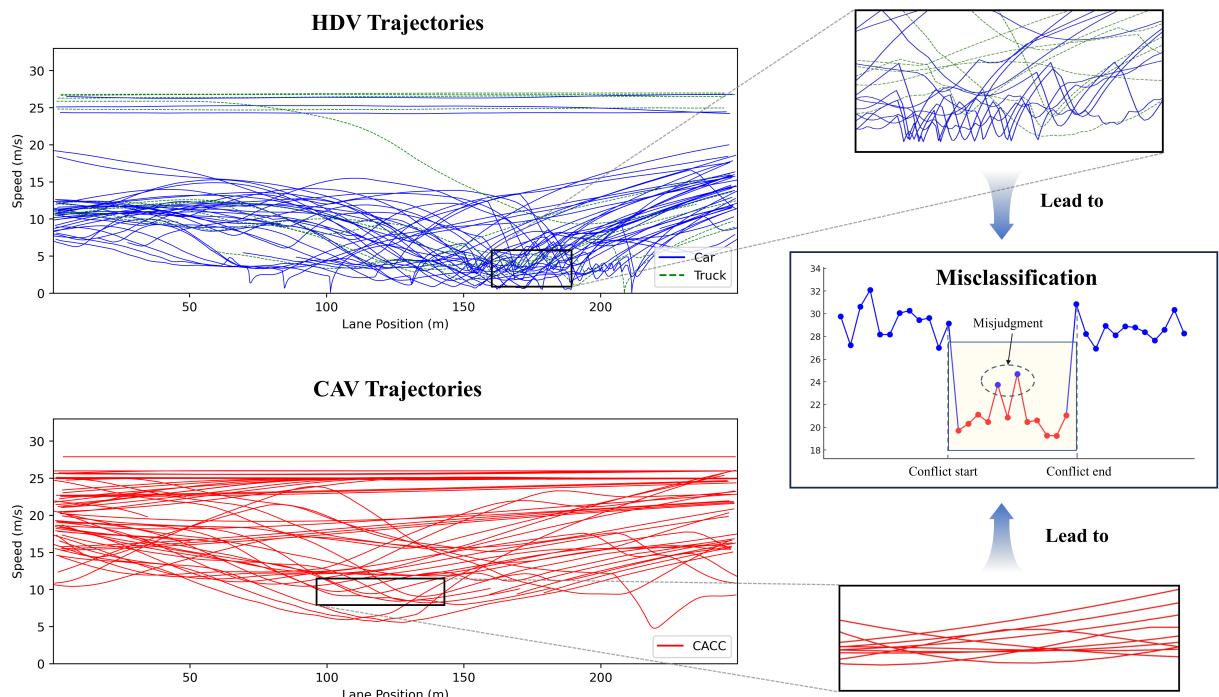


Fig. 8. Position-speed profiles of two types of vehicles.

Fig.9 highlights the speed-position trajectories along a merge ramp segment under a high penetration scenario (70%), alongside the corresponding risk prediction outputs of the three models. We subdivide this segment into three zones: upstream of the merge, within the merge, and downstream of the merge. Compared to the mainline, merge ramp environments are inherently more intricate, with intensified vehicle interactions that readily induce velocity sep-

aration. As depicted, vehicles traversing the merge zone frequently deviate from the linear free-flow regime, exhibiting stochastic decelerations and accelerations that oscillate between high-speed and low-speed clusters. This dual stream phenomenon generates unstable disturbance regions, which pose significant challenges for conflict prediction algorithms. A comparison against ground-truth conflict events reveals that not all perturbation zones correspond to actual conflict risks. In both the pre-merge and post-merge regions, vehicle speeds remain relatively uniform, and all three architectures—STGCN, STGAT, and MS-STGNet—produce risk estimates that closely align with observed outcomes. However, within the merge zone itself, as oscillatory amplitudes intensify, STGCN and STGAT manifest pronounced false positives, assigning elevated risk scores to nonconflict areas. In contrast, MS-STGNet maintains a precise risk delineation throughout.

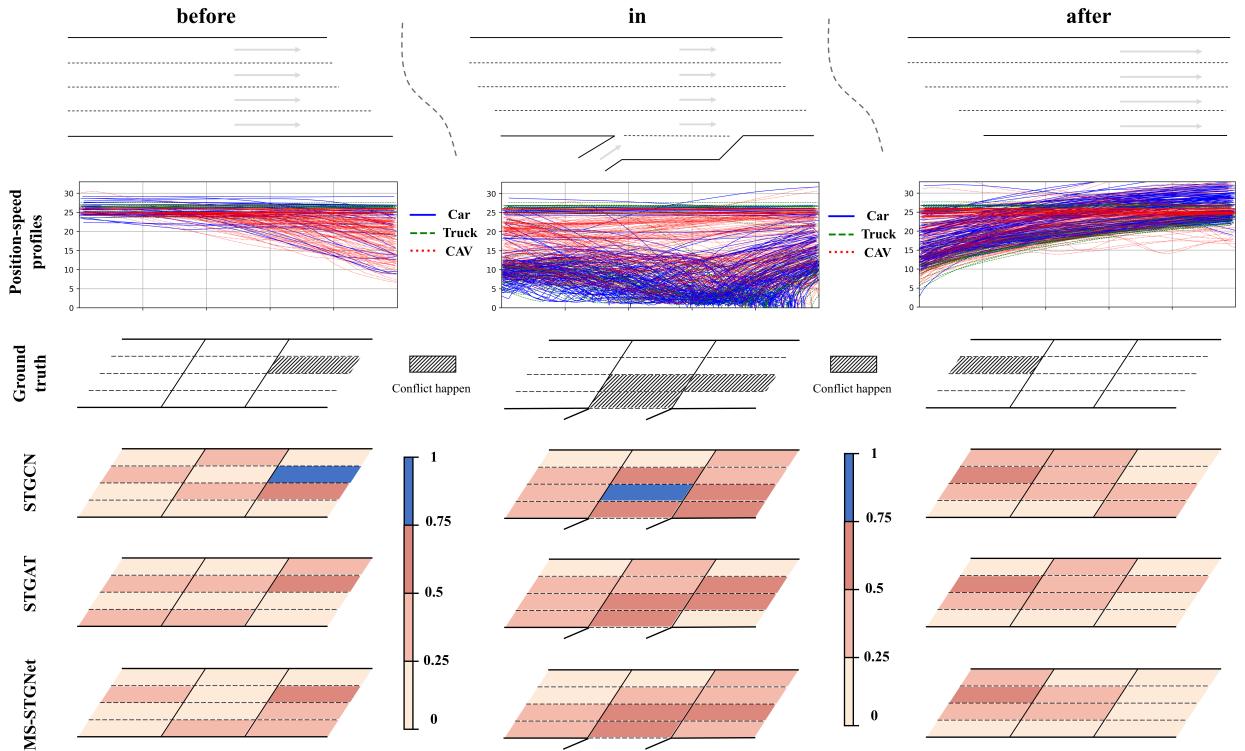


Fig. 9. Vehicle position-speed trajectories and predicting probability values for STGCN, STGAT, and MS-STGNet in (**before**) pre-merging segment, (**in**) merging segment, (**after**) post-merging segment.

Beyond penetration rates, we also examined how traffic volume and the resulting speed dispersion patterns affect conflict risk and model behavior. In the simulation, different representative demand levels were considered over a total of 500 hours, covering low-, medium-, and high-volume conditions. The supplementary trajectory plots in Appendix C (Figures C1–C3) show that as traffic volume increases, pronounced speed oscillations emerge along the segment and become more frequent and severe. This indicates that, even under mixed CAV–HDV conditions, higher demand intensifies vehicle interactions and amplifies the likelihood of conflicts, which supports our use of traffic state variations as predictors of conflict occurrence. A closer inspection of these trajectories further highlights the role of different vehicle classes and CAV penetration as key traffic features. The green and blue trajectories representing HDVs exhibit larger amplitude and higher-frequency speed fluctuations than the red trajectories representing CAVs, reflecting more aggressive driving behavior and delayed responses in the human-driven fleet. Heavy vehicles (trucks) introduce additional instability due to their limited acceleration and deceleration capabilities and larger size, which force surrounding vehicles to adjust their speeds more frequently and create pronounced perturbation zones. As CAV penetration increases, these unstable zones shrink and the gaps between high-speed and low-speed vehicle clusters are gradually bridged by heterogeneous CACC queues, leading to smoother trajectories and reduced speed dispersion. Combined with the segment-level risk profiles in Fig. 9, these observations indicate that CAV penetration rate, traffic volume, and the resulting speed separation patterns are among the most influential traffic features for conflict prediction

1 in the proposed framework: MS-STGNet is particularly effective at aligning its predicted risk with these underlying
 2 speed dispersion structures, while STGCN and STGAT tend to generate spurious conflict probabilities in disturbance
 3 zones.

4 7. Conclusion

5 Real-time conflict analysis provides valuable insights into crash precursors and supports the implementation of
 6 proactive traffic safety management strategies. To obtain a better conflict risk prediction performance, tremendous
 7 efforts have been made using various operational sensing data and advanced modeling techniques. However, due to
 8 the inherent complexity of conflict modeling, the application of advanced machine learning methods remains in its
 9 infancy. Addressing this challenge requires algorithms and architectures capable of handling such complexities while
 10 delivering high predictive accuracy.

11 In this study, we propose a novel Manifold Similarity-based Multi-Graph Spatiotemporal Network (MS-STGNet)
 12 for conflict prediction in mixed traffic environments. The proposed framework demonstrates strong performance in
 13 scenarios where CAVs and HDVs coexist, while also exhibiting a degree of generalizability to non-mixed traffic con-
 14 ditions. Specifically, the model incorporates the following components: A residual convolutional network to extract
 15 geographical features in interconnected areas of the land space. A manifold similarity graph module to capture spa-
 16 tial semantic features in regions. A temporal convolutional network to model temporal dependencies in traffic flow
 17 data, extending spatial features into spatiotemporal representations. An adaptive fusion gate mechanism combines
 18 geographical and semantic spatiotemporal features to generate final predictions. As a main contribution of the study,
 19 we introduce a manifold similarity method to model the similarity of traffic states. Historical traffic flow data are ag-
 20 ggregated into traffic state vectors, and manifold distances are used to calculate similarities between these vectors. The
 21 similarity matrix is then integrated into the graph network as prior knowledge, imposing a layer of physical constraints
 22 on the deep learning outcomes. This ensures that the predicted transition of traffic states from conflict to non-conflict
 23 aligns with inherent spatiotemporal patterns, which is crucial for accurately identifying different traffic states and re-
 24 ducing misclassifications in conflict prediction.

25 The proposed MS-STGNet was evaluated using simulation datasets across different market penetration rates (10%,
 26 30%, 50%, 70%, and 90%). In future scenarios where CAVs and HDVs coexist in mixed traffic environments, exploring
 27 novel approaches to evaluate real-time conflict risks is imperative. Simulations were conducted using SUMO and its
 28 extension, Plexe, with model parameters calibrated based on the HighD dataset. Conflict events were identified from
 29 vehicle trajectories and categorized under the five penetration rates to form the datasets. The experimental results
 30 demonstrate that: **1)** MS-STGNet outperforms baseline models in traffic conflict prediction, particularly excelling in
 31 reducing false alarm rates. This highlights the superiority of the manifold similarity module in capturing transitions in
 32 traffic states. **2)** Across varying market penetration rates and sample-balance conditions, the MS-STGNet framework
 33 consistently delivers robust performance, effectively counteracting the zero-inflation phenomenon inherent in traffic-
 34 event datasets. **3)** Within complex traffic scenarios or under conditions of pronounced state volatility, MS-STGNet
 35 precisely identifies traffic conflict occurrences and delineates their associated risk profiles along roadway segments. **4)**
 36 Ablation studies emphasize the positive contributions of each model component to the overall predictive quality.

37 The proposed framework has several practical implications. It can be embedded as a safety prediction component
 38 in CAV cloud management systems for freeway corridors and urban expressways, integrated into freeway traffic man-
 39 agement centers and ramp control or variable speed limit systems to support mixed CAV-HDV operations, and used
 40 within regional expressway operation platforms to provide real-time conflict or crash risk warnings at bottlenecks and
 41 merging/diverging areas, thereby enhancing the safety management and visualization of freeway networks. The limi-
 42 tations of this study are summarized as follows: **1)** The model is calibrated and evaluated in a microscopic simulation
 43 of a four-lane freeway segment with motorized traffic only. Although the simulation is grounded in highD trajectory
 44 data, we do not yet validate MS-STGNet on large-scale field observations of mixed CAV-HDV traffic, and the direct
 45 transferability of the results to urban or suburban road networks with signalised intersections, pedestrians, and non-
 46 motorised vehicles is therefore limited. **2)** The current experiments focus on a single 14 km corridor with specific
 47 demand patterns; additional facilities and more diverse demand scenarios would further test the generalizability of the
 48 framework. **3)** The predefined manifold similarity matrix remains static over time, preventing the model from captur-
 49 ing previously unseen traffic state transitions unless it is retrained. **4)** The proposed framework currently focuses on
 50 binary conflict/non-conflict prediction. Although the sigmoid activation in the output layer produces continuous risk
 51 scores in the [0,1] range, we do not explicitly model or evaluate graded levels of conflict severity (e.g., minor versus

severe conflicts). Moving forward, future works contain: **1)** Collecting or leveraging emerging mixed CAV–HDV field datasets with continuous monitoring, so as to retrain and validate MS-STGNet under real-world conditions and assess its scalability. **2)** Developing online or adaptive manifold-learning strategies to update similarity matrices in real time. **3)** Exploring scalable pretraining and training strategies on larger and more diverse networks, including freeway corridors and urban expressways with additional contextual variables such as weather conditions, pavement friction, and points of interest (POIs). **4)** Extending MS-STGNet from binary conflict detection to graded or ordinal conflict severity prediction by combining continuous risk scores with appropriate severity labels.

8 Acknowledgement

This research was supported by the project of the National Key R&D Program of China (No. 2018YFB1601301), the National Natural Science Foundation of China (No. 71961137006, NO. 52302441), the Science and Technology Commission of Shanghai Municipality (No. 22dz1207500), and the China Scholarship Council (No.202406260279).

12 Appendix A. Visualization of the learned manifold-similarity matrices

Appendix A presents the learned manifold-similarity matrices for flow, occupancy, and speed, denoted by **Matrices**^(flow), **Matrices**^(occupancy), and **Matrices**^(speed), respectively. Each matrix is of size 108×108 ; for readability, each matrix lists the top-left 5×5 block together with the last row and last column, with ellipses indicating continuation to the full size.

$$\text{Matrices}^{(\text{flow})} = \begin{bmatrix} 1.000 & 0.277 & 0.268 & 0.274 & 0.745 & \cdots & 0.686 \\ 0.277 & 1.000 & 0.701 & 0.285 & 0.689 & \cdots & 0.279 \\ 0.268 & 0.701 & 1.000 & 0.707 & 0.693 & \cdots & 0.699 \\ 0.274 & 0.285 & 0.707 & 1.000 & 0.688 & \cdots & 0.759 \\ 0.745 & 0.689 & 0.693 & 0.688 & 1.000 & \cdots & 0.696 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0.686 & 0.279 & 0.699 & 0.759 & 0.696 & \cdots & 1.000 \end{bmatrix}, \quad \text{Matrices}^{(\text{flow})} \in \mathbb{R}^{108 \times 108}$$

$$\text{Matrices}^{(\text{occupancy})} = \begin{bmatrix} 1.000 & 0.365 & 0.316 & 0.276 & 0.353 & \cdots & 0.250 \\ 0.365 & 1.000 & 0.367 & 0.327 & 0.302 & \cdots & 0.314 \\ 0.316 & 0.367 & 1.000 & 0.390 & 0.283 & \cdots & 0.358 \\ 0.276 & 0.327 & 0.390 & 1.000 & 0.237 & \cdots & 0.499 \\ 0.353 & 0.302 & 0.283 & 0.237 & 1.000 & \cdots & 0.016 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0.250 & 0.314 & 0.358 & 0.499 & 0.016 & \cdots & 1.000 \end{bmatrix}, \quad \text{Matrices}^{(\text{occupancy})} \in \mathbb{R}^{108 \times 108}$$

$$\text{Matrices}^{(\text{speed})} = \begin{bmatrix} 1.000 & 0.602 & 0.491 & 0.523 & 0.600 & \cdots & 0.396 \\ 0.602 & 1.000 & 0.537 & 0.566 & 0.561 & \cdots & 0.441 \\ 0.491 & 0.537 & 1.000 & 0.503 & 0.450 & \cdots & 0.402 \\ 0.523 & 0.566 & 0.503 & 1.000 & 0.474 & \cdots & 0.512 \\ 0.600 & 0.561 & 0.450 & 0.474 & 1.000 & \cdots & 0.344 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0.396 & 0.441 & 0.402 & 0.512 & 0.344 & \cdots & 1.000 \end{bmatrix}, \quad \text{Matrices}^{(\text{speed})} \in \mathbb{R}^{108 \times 108}$$

17 Appendix B. Detailed formulation of manifold-based similarity and adaptive adjacency

18 B.1. Manifold similarity kernel and bandwidth selection

Given the geodesic distances d_{ij} on the traffic-state manifold, we convert them into a similarity matrix \mathbf{W} using a Gaussian kernel:

$$W_{ij} = \exp\left(-\frac{d_{ij}^2}{2h^2}\right), \quad (\text{B.1})$$

1 where d_{ij} represents the manifold distance between traffic states i and j ; \exp is the exponential function e^x ; and
 2 h denotes the kernel bandwidth. The bandwidth h is selected by minimizing the corrected Akaike Information
 3 Criterion (AICc) of the resulting model:

$$f(h) = 2k - 2 \ln(\mathcal{L}(h)) + \frac{2k(k+1)}{n-k-1}, \quad (\text{B.2})$$

4 where n is the sample size, k is the number of free parameters, and $\mathcal{L}(h)$ denotes the likelihood function under band-
 5 width h .

6 B.2. SVD-based initialization and adaptive adjacency

7 To incorporate potential spatial correlations into our framework, we construct three adaptive graphs by initializing
 8 the weights between nodes using similarity matrices. Singular Value Decomposition (SVD) is employed for graph
 9 initialization (Guo et al., 2015; Zou et al., 2024), and \mathbf{A}^* can be expressed as the product of three distinct matrices, as
 10 follows:

$$\mathbf{A}^* = \mathbf{U}^* \boldsymbol{\Sigma}^* \mathbf{V}^{*\top} \quad (\text{B.3})$$

11 where \mathbf{U}^* and \mathbf{V}^* represent orthogonal matrices representing the left and right singular vectors, respectively. $\boldsymbol{\Sigma}^*$ is a
 12 diagonal matrix containing singular values. The graph initialized through SVD decomposition provides only a static
 13 representation and cannot adapt to the dynamic changes in the data. Therefore, the weight matrix of the adaptive graph,
 14 \mathbf{A}^* , needs to be optimized through a learnable function:

$$\mathbf{A}^* = \text{ReLU}(\mathbf{M}_{lt} \mathbf{M}_{rt}) \quad (\text{B.4})$$

15 where \mathbf{M}_{lt} and \mathbf{M}_{rt} are the core learnable parameter matrices, which play a crucial role in dynamically modeling the
 16 weight relationships between nodes in the graph. \mathbf{M}_{lt} is the left transformation matrix, designed to encode a linear
 17 transformation of the input features or spatial dependency information. It operates as a critical step in updating the
 18 representation of node relationships by applying a transformation to the input data, expressed as: $\mathbf{M}_{lt} = \mathbf{W}_{lt} (\hat{\mathbf{U}}_\star \hat{\boldsymbol{\Sigma}}_\star)$.
 19 Similarly, \mathbf{M}_{rt} is the right transformation matrix, responsible for adjusting or aggregating the information encoded in
 20 \mathbf{M}_{lt} , expressed as: $\mathbf{M}_{rt} = \mathbf{W}_{rt} (\hat{\boldsymbol{\Sigma}}_\star \hat{\mathbf{V}}_\star^\top)$. The ReLU function is applied to introduce nonlinearity and ensure that the
 21 weights remain non-negative. Subsequently, the softmax function is used to normalize the weights of each node, ensur-
 22 ing that their sum equals 1. This normalization guarantees a balanced distribution of information during transmission,
 23 preventing any single node from dominating the interaction:

$$\tilde{\mathbf{A}}^* = \mathbf{I}_N + \text{softmax}(\text{ReLU}(\mathbf{M}_{lt} \mathbf{M}_{rt})) \quad (\text{B.5})$$

24 where \mathbf{I}_N is the identity matrix.

25 Appendix C. Supplementary vehicle position–speed trajectories

26 In this appendix, we provide additional vehicle position–speed trajectory plots for three representative demand lev-
 27 els, corresponding to low-, medium-, and high-volume conditions. For each traffic volume, the trajectories are shown
 28 separately for the pre-merging, merging, and post-merging segments, with different colors indicating HDVs, CAVs,
 29 and heavy vehicles (trucks). These plots illustrate how increasing traffic volume and changes in vehicle composition
 30 lead to more pronounced speed oscillations and perturbation zones, complementing the case study around Fig. 9 in the
 31 main text and supporting the discussion in Section 6.10 on the impact of traffic volume, CAV penetration, and speed
 32 separation on conflict risk.

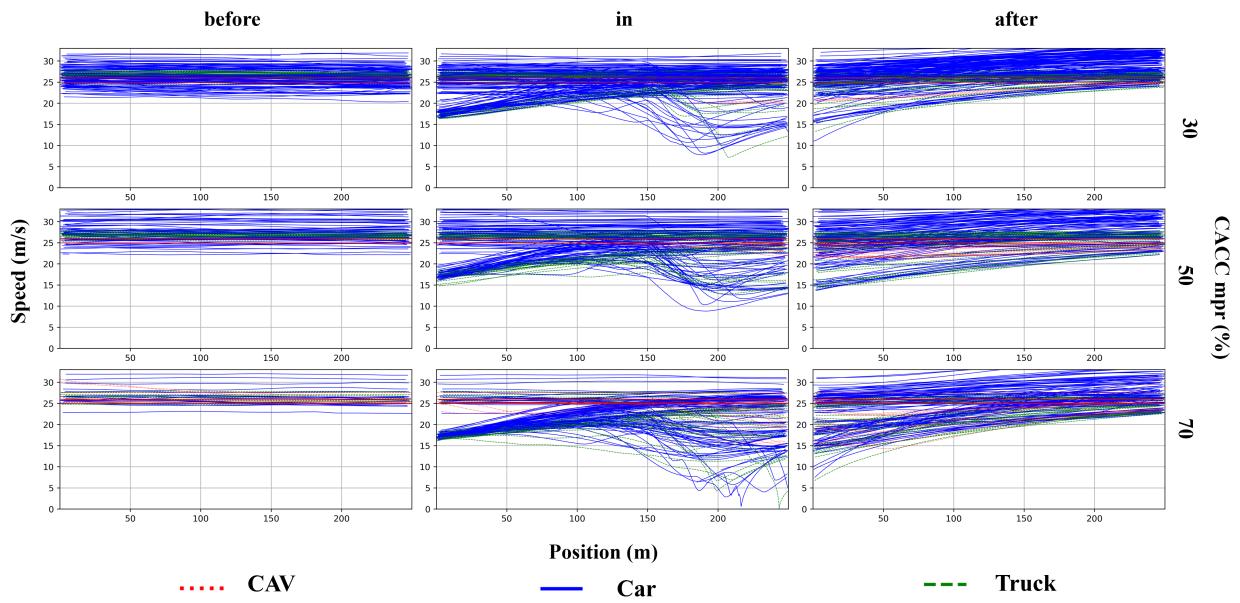


Fig. C1. Vehicle position-speed trajectories at different penetration rates with a traffic volume of 3000 vehicles/hour. **(before)** pre-merging segment. **(in)** merging segment. **(after)** post-merging segment.

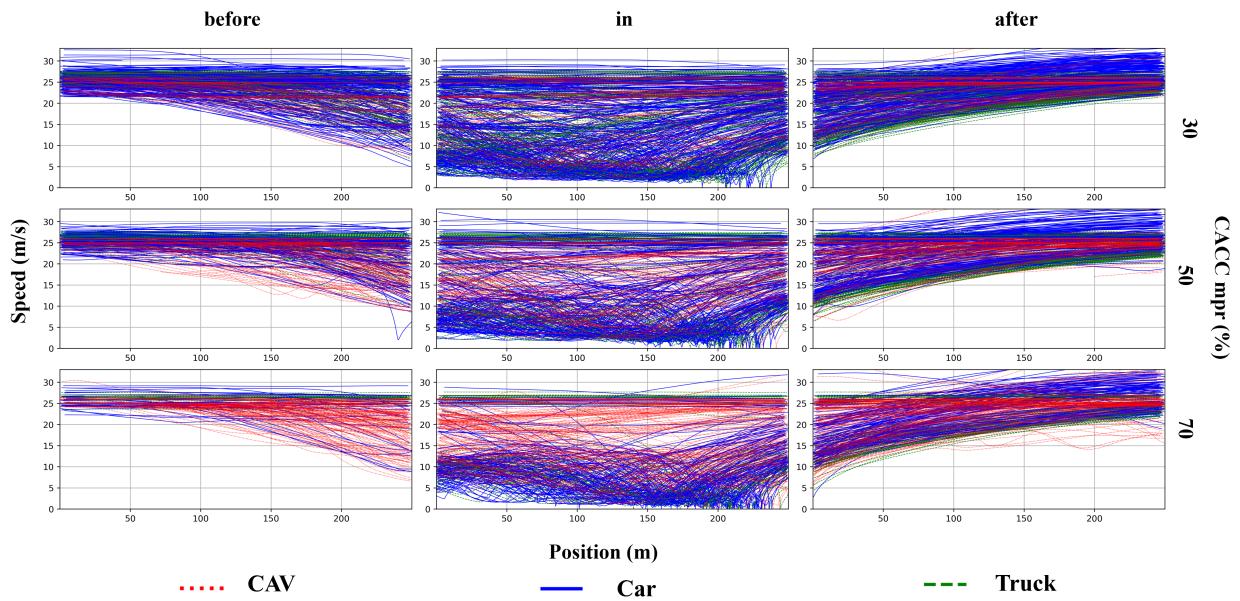


Fig. C2. Vehicle position-speed trajectories at different penetration rates with a traffic volume of 6000 vehicles/hour. **(before)** pre-merging segment. **(in)** merging segment. **(after)** post-merging segment.

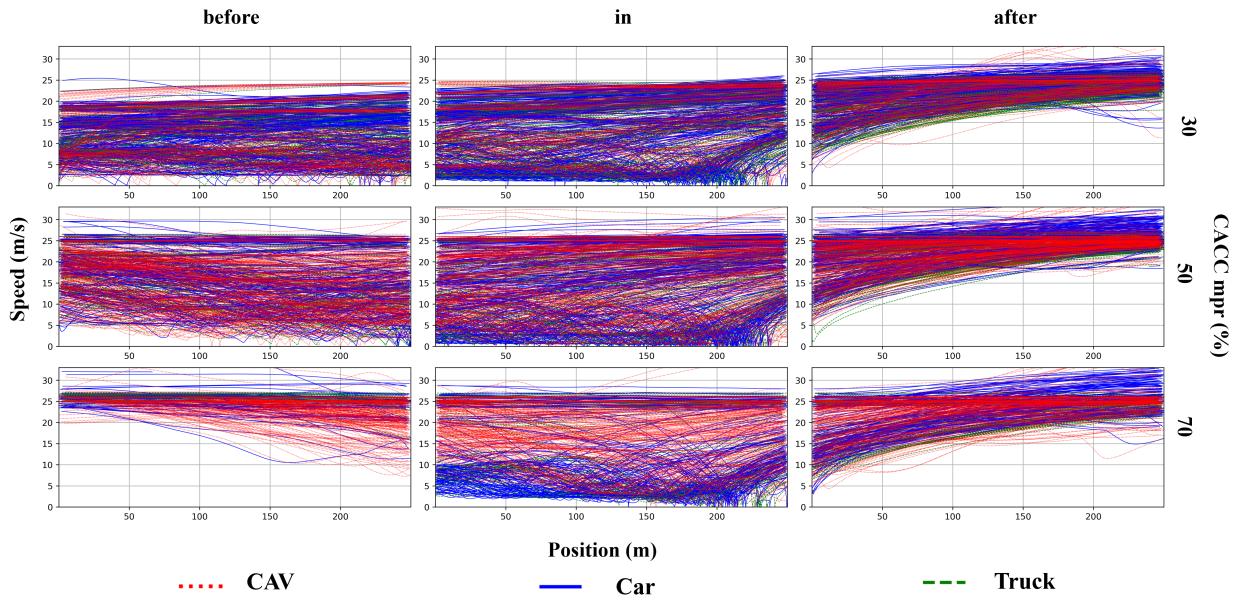


Fig. C3. Vehicle position-speed trajectories at different penetration rates with a traffic volume of 9000 vehicles/hour. **(before)** pre-merging segment. **(in)** merging segment. **(after)** post-merging segment.

Appendix D. Performance on NGSIM dataset

To further evaluate the scalability of the model and its adaptability to real-world conditions, we tested the MS-STGNet model and the baseline models on a real-world trajectory dataset. Specifically, we employed the trajectory data from the Next Generation Simulation (NGSIM) program of the U.S. Federal Highway Administration (FHWA, 2006), which has been widely used for the analysis of longitudinal car-following, lane-changing maneuvers, and traffic flow characteristics (Chen et al., 2021; Zong et al., 2024). The highway subset of NGSIM comprises the I-80 and US-101 datasets, both collected using cameras mounted on the rooftops of adjacent high-rise buildings. The I-80 dataset covers an approximately 500 m freeway segment with six unidirectional lanes, with a total observation duration of 45 minutes (Fig. D1(a)). The US-101 trajectory dataset was collected on a roughly 640 m freeway segment with six lanes, with the same 45-minute observation period (Fig. D1(b)). Because the dimensions of the traffic statistics and the road segment lengths differ from those of our own dataset, we adopted a slightly modified data-processing procedure. The spatial aggregation interval was set to 100 m, and traffic flow, speed, and occupancy were aggregated over 3-minute windows. The indicators used to quantify traffic conflicts were kept consistent with the original study (TTC, DRAC, DDR), and the input-output configurations of all models were aligned with the original setting. The experimental results are reported in Table D1.

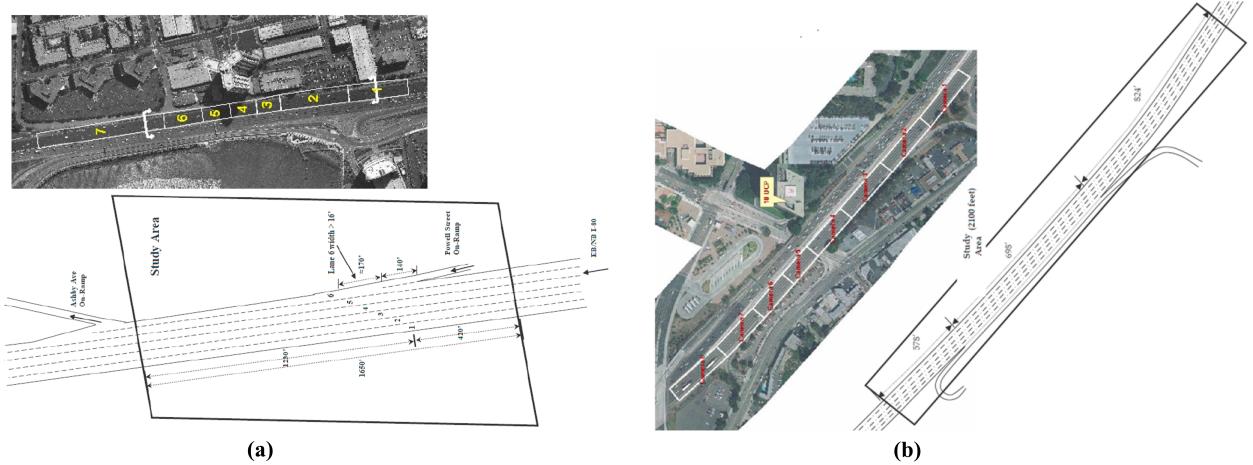


Fig. D1. Collection Scenarios of open-source datasets **(a)** NGSIM, I-80; **(b)** NGSIM US-101.

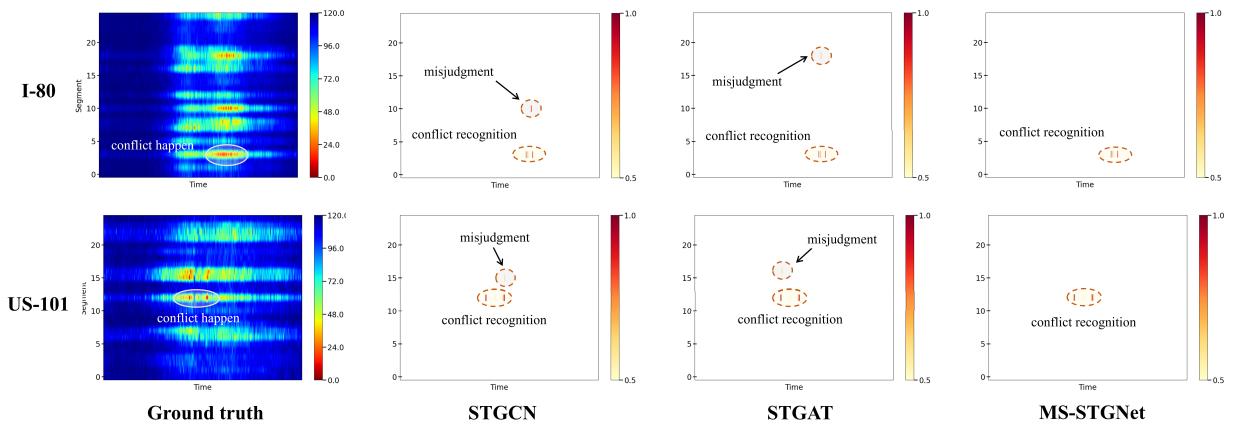
As shown in Table D1, the proposed MS-STGNet still exhibits consistently superior performance on both the I-80 and US-101 sub-datasets. On I-80, MS-STGNet achieves a recall of 76.6% compared with 75.3% for STGAT and 74.0% for STGCN, while reducing the false alarm rate from 18.0% (STGAT) to 17.1% and improving AUC, accuracy, and G-mean from 77.6%, 81.0%, and 77.4% to 79.6%, 84.0%, and 80.4%, respectively. A similar pattern is observed on US-101, where MS-STGNet improves recall from 74.3% to 76.2%, decreases the false alarm rate from 20.4% to 16.9%, and further increases AUC, accuracy, and G-mean from 78.2%, 80.0%, and 76.8% to 80.2%, 83.2%, and 79.5%. These consistent gains over both traditional machine-learning methods (SVM, XGBoost) and strong deep and graph-based baselines (CNN, LSTM-CNN, STGCN, STGAT) confirm that MS-STGNet maintains strong scalability and robustness when transferred to real-world freeway traffic conditions.

Table D1

Performance on NGSIM dataset.

Sub-datasets	Metric	SVM	XGBoost	CNN	LSTM-CNN	STGCN	STGAT	MS-STGNet
I-80	Recall	0.502	0.570	0.692	0.709	0.740	0.753	0.766 _{+11.73%}
	False alarm rate	0.464	0.436	0.224	0.220	0.195	0.180	0.171 _{+15.00%}
	AUC	0.557	0.614	0.740	0.749	0.768	0.776	0.796 _{+2.58%}
	Accuracy	0.562	0.628	0.774	0.779	0.819	0.810	0.840 _{+12.56%}
	G-mean	0.512	0.565	0.721	0.738	0.778	0.774	0.804 _{+13.34%}
US-101	Recall	0.492	0.560	0.682	0.699	0.730	0.743	0.762 _{+12.56%}
	False alarm rate	0.469	0.436	0.232	0.219	0.182	0.204	0.169 _{+17.14%}
	AUC	0.547	0.604	0.730	0.739	0.758	0.782	0.802 _{+12.56%}
	Accuracy	0.552	0.618	0.764	0.769	0.809	0.800	0.832 _{+12.84%}
	G-mean	0.502	0.555	0.711	0.728	0.764	0.768	0.795 _{+13.52%}

From the results reported in Table D1, we can again observe that MS-STGNet achieves the most pronounced improvement in terms of reducing the false positive rate. Accordingly, we further visualized the spatiotemporal speed heatmaps and the corresponding conflict detection outcomes (Fig. D2). The visualizations indicate that, under real-world testing conditions, the three architectures yield almost identical predictions of traffic conflicts. Nevertheless, within the stop-and-go waves, a small number of false positives can still be observed for the STGAT and STGCN models. This finding further corroborates that, by incorporating flow similarity to suppress noise, the MS-STGNet framework attains robust predictive performance with strong adaptability and reliability across diverse traffic scenarios.

**Fig. D2.** Speed heatmap and conflict recognition result for STGCN, STGAT, and MS-STGNet on NGSIM dataset.

References

Abdel-Aty, M., Pande, A., 2005. Identifying crash propensity using specific traffic speed conditions. Journal of safety Research 36, 97–108.

- 1 Abou Elassad, Z.E., Mousannif, H., Al Moatassime, H., 2020. A real-time crash prediction fusion framework: An imbalance-aware strategy for
2 collision avoidance systems. *Transportation research part C: emerging technologies* 118, 102708.
- 3 Ahangar, M.N., Ahmed, Q.Z., Khan, F.A., Hafeez, M., 2021. A survey of autonomous vehicles: Enabling communication technologies and
4 challenges. *Sensors* 21, 706.
- 5 Ali, Y., Haque, M.M., Mannering, F., 2023. Assessing traffic conflict/crash relationships with extreme value theory: Recent developments and
6 future directions for connected and autonomous vehicle and highway safety research. *Analytic methods in accident research* 39, 100276.
- 7 Ali, Y., Hussain, F., Haque, M.M., 2024. Advances, challenges, and future research needs in machine learning-based crash prediction models: A
8 systematic review. *Accident Analysis & Prevention* 194, 107378.
- 9 Bai, S., Kolter, J.Z., Koltun, V., 2018. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv*
10 preprint arXiv:1803.01271 .
- 11 Bansal, P., Kockelman, K.M., 2017. Forecasting americans' long-term adoption of connected and autonomous vehicle technologies. *Transportation*
12 Research Part A: Policy and Practice 95, 49–63.
- 13 Bao, J., Liu, P., Ukkusuri, S.V., 2019. A spatiotemporal deep learning approach for citywide short-term crash risk prediction with multi-source
14 data. *Accident Analysis & Prevention* 122, 239–254.
- 15 Basso, F., Pezoa, R., Varas, M., Villalobos, M., 2021. A deep learning approach for real-time crash prediction using vehicle-by-vehicle data.
16 *Accident Analysis & Prevention* 162, 106409.
- 17 Bergel-Hayat, R., Debbah, M., Antoniou, C., Yannis, G., 2013. Explaining the road accident risk: Weather effects. *Accident Analysis & Prevention*
18 60, 456–465.
- 19 Cai, Q., Abdel-Aty, M., Yuan, J., Lee, J., Wu, Y., 2020. Real-time crash prediction on expressways using deep generative models. *Transportation*
20 research part C: emerging technologies 117, 102697.
- 21 Cai, Y., Dai, L., Wang, H., Chen, L., Li, Y., Sotelo, M.A., Li, Z., 2021. Pedestrian motion trajectory prediction in intelligent driving from far shot
22 first-person perspective video. *IEEE Transactions on Intelligent Transportation Systems* 23, 5298–5313.
- 23 Caliendo, C., Guida, M., Parisi, A., 2007. A crash-prediction model for multilane roads. *Accident Analysis & Prevention* 39, 657–670.
- 24 Chen, C., Fan, X., Zheng, C., Xiao, L., Cheng, M., Wang, C., 2018. Sdcae: Stack denoising convolutional autoencoder model for accident risk
25 prediction via traffic big data, in: 2018 sixth international conference on advanced cloud and big data (CBD), IEEE. pp. 328–333.
- 26 Chen, K., Luo, Y., Zhu, M., Yang, H., 2024a. Human-like interactive lane-change modeling based on reward-guided diffusive predictor and planner.
27 *IEEE Transactions on Intelligent Transportation Systems* , early access.
- 28 Chen, Q., Gu, R., Huang, H., Lee, J., Zhai, X., Li, Y., 2021. Using vehicular trajectory data to explore risky factors and unobserved heterogeneity
29 during lane-changing. *Accident Analysis & Prevention* 151, 105871.
- 30 Chen, X., Tiu, P., Zhang, Y., Zhu, M., Zheng, X., Wang, Y., 2024b. Improving car-following control in mixed traffic: A deep reinforcement learning
31 framework with aggregated human-driven vehicles, in: 2024 IEEE Intelligent Vehicles Symposium (IV), IEEE. pp. 627–632.
- 32 Di Vaio, M., Fiengo, G., Petrillo, A., Salvi, A., Santini, S., Tufo, M., 2019. Cooperative shock waves mitigation in mixed traffic flow environment.
33 *IEEE Transactions on Intelligent Transportation Systems* 20, 4339–4353.
- 34 FHWA, U., 2006. Next generation simulation (ngsim).
- 35 Fiengo, G., Lui, D.G., Petrillo, A., Santini, S., Tufo, M., 2019. Distributed robust pid control for leader tracking in uncertain connected ground
36 vehicles with v2v communication delay. *IEEE/ASME Transactions on Mechatronics* 24, 1153–1165.
- 37 Fu, C., Sayed, T., 2021. Comparison of threshold determination methods for the deceleration rate to avoid a crash (drac)-based crash estimation.
38 *Accident Analysis & Prevention* 153, 106051.
- 39 Galvani, M., 2019. History and future of driver assistance. *IEEE Instrumentation & Measurement Magazine* 22, 11–16.
- 40 Gao, X., Jiang, X., Haworth, J., Zhuang, D., Wang, S., Chen, H., Law, S., 2024. Uncertainty-aware probabilistic graph neural networks for road-level
41 traffic crash prediction. *Accident Analysis & Prevention* 208, 107801.
- 42 Garg, M., Bourouche, M., 2023. Can connected autonomous vehicles improve mixed traffic safety without compromising efficiency in realistic
43 scenarios? *IEEE Transactions on Intelligent Transportation Systems* 24, 6674–6689.
- 44 Goswamy, A., Abdel-Aty, M., Islam, Z., 2023. Factors affecting injury severity at pedestrian crossing locations with rectangular rapid flashing
45 beacons (rrfb) using xgboost and random parameters discrete outcome models. *Accident Analysis & Prevention* 181, 106937.
- 46 Gu, Z., Wang, Z., Liu, Z., Saberi, M., 2022. Network traffic instability with automated driving and cooperative merging. *Transportation Research*
47 Part C: Emerging Technologies 138, 103626.
- 48 Guo, Q., Zhang, C., Zhang, Y., Liu, H., 2015. An efficient svd-based method for image denoising. *IEEE transactions on Circuits and Systems for*
49 *Video Technology* 26, 868–880.
- 50 Hossain, M., Abdel-Aty, M., Quddus, M.A., Muromachi, Y., Sadeek, S.N., 2019. Real-time crash prediction models: State-of-the-art, design
51 pathways and ubiquitous requirements. *Accident Analysis & Prevention* 124, 66–84.
- 52 Hossain, M., Muromachi, Y., 2012. A bayesian network based framework for real-time crash prediction on the basic freeway segments of urban
53 expressways. *Accident Analysis & Prevention* 45, 373–381.
- 54 Hou, K., Zheng, F., Liu, X., 2024a. Enhancing mixed traffic safety assessment: A novel safety metric combined with a comprehensive behavioral
55 modeling framework. *Accident Analysis & Prevention* 208, 107766.
- 56 Hou, K., Zheng, F., Liu, X., Fan, Z., 2024b. Cooperative vehicle platoon control considering longitudinal and lane-changing dynamics. *Transport-*
57 *metrica A: transport science* 20, 2182143.
- 58 Hu, J., Huang, M.C., Yu, X., 2020. Efficient mapping of crash risk at intersections with connected vehicle data and deep learning models. *Accident*
59 *Analysis & Prevention* 144, 105665.
- 60 Hu, X., Sun, J., 2019. Trajectory optimization of connected and autonomous vehicles at a multilane freeway merging area. *Transportation Research*
61 Part C: Emerging Technologies 101, 111–125.
- 62 Huang, Y., Bi, H., Li, Z., Mao, T., Wang, Z., 2019. Stgat: Modeling spatial-temporal interactions for human trajectory prediction, in: Proceedings
63 of the IEEE/CVF international conference on computer vision, pp. 6272–6281.

- 1 Ivanchev, J., Eckhoff, D., Knoll, A., 2019. System-level optimization of longitudinal acceleration of autonomous vehicles in mixed traffic, in: 2019
 2 IEEE Intelligent Transportation Systems Conference (ITSC), IEEE. pp. 1968–1974.
- 3 Jiang, F., Yuen, K.K.R., Lee, E.W.M., 2020. A long short-term memory-based framework for crash detection on freeways with traffic data of
 4 different temporal resolutions. *Accident Analysis & Prevention* 141, 105520.
- 5 Kamel, A., Sayed, T., Fu, C., 2023. Real-time safety analysis using autonomous vehicle data: a bayesian hierarchical extreme value model.
 6 *Transportmetrica B: Transport Dynamics* 11, 826–846.
- 7 Kamel, A., Sayed, T., Kamel, M., 2024. Real-time combined safety-mobility assessment using self-driving vehicles collected data. *Accident
 8 Analysis & Prevention* 199, 107513.
- 9 Lee, G., Mallipeddi, R., Lee, M., 2012. Identification of moving vehicle trajectory using manifold learning, in: Neural Information Processing:
 10 19th International Conference, ICONIP 2012, Doha, Qatar, November 12–15, 2012, Proceedings, Part IV 19, Springer. pp. 188–195.
- 11 Li, P., Abdel-Aty, M., Yuan, J., 2020. Real-time crash risk prediction on arterials based on lstm-cnn. *Accident Analysis & Prevention* 135, 105371.
- 12 Li, S., Pu, Z., Cui, Z., Lee, S., Guo, X., Ngoduy, D., 2024. Inferring heterogeneous treatment effects of crashes on highway traffic: A doubly robust
 13 causal machine learning approach. *Transportation research part C: emerging technologies* 160, 104537.
- 14 Li, Y., Lu, J., Xu, K., 2017a. Crash risk prediction model of lane-change behavior on approaching intersections. *Discrete Dynamics in Nature and
 15 Society* 2017, 7328562.
- 16 Li, Y., Yu, R., Shahabi, C., Liu, Y., 2017b. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *arXiv preprint
 17 arXiv:1707.01926*.
- 18 Lin, J., 1991. Divergence measures based on the shannon entropy. *IEEE Transactions on Information theory* 37, 145–151.
- 19 Lin, L., Wang, Q., Sadek, A.W., 2015. A novel variable selection method based on frequent pattern tree for real-time traffic accident risk prediction.
 20 *Transportation Research Part C: Emerging Technologies* 55, 444–459.
- 21 Liu, H., Kan, X.D., Shladover, S.E., Lu, X.Y., Ferlis, R.E., 2018a. Modeling impacts of cooperative adaptive cruise control on mixed traffic flow in
 22 multi-lane freeway facilities. *Transportation Research Part C: Emerging Technologies* 95, 261–279.
- 23 Liu, Q., Cai, Y., Jiang, H., Lu, J., Chen, L., 2018b. Traffic state prediction using isomap manifold learning. *Physica A: Statistical Mechanics and
 24 its Applications* 506, 532–541.
- 25 Liu, Q., Gao, C., Wang, H., Cai, Y., Chen, L., Lv, C., 2024. Learning from trajectories: How heterogeneous cacc platoons affect the traffic flow in
 26 highway merging area. *IEEE Transactions on Vehicular Technology* .
- 27 Liu, Q., Li, C., Jiang, H., Nie, S., Chen, L., 2022. Transfer learning-based highway crash risk evaluation considering manifold characteristics of
 28 traffic flow. *Accident Analysis & Prevention* 168, 106598.
- 29 Liu, Z., Cai, Y., Wang, H., Chen, L., Gao, H., Jia, Y., Li, Y., 2021. Robust target recognition and tracking of self-driving cars with radar and camera
 30 information fusion under severe weather conditions. *IEEE Transactions on Intelligent Transportation Systems* 23, 6640–6653.
- 31 Lu, K., Ding, Z., Ge, S., 2012. Sparse-representation-based graph embedding for traffic sign recognition. *IEEE Transactions on Intelligent Trans-
 32 portation Systems* 13, 1515–1524.
- 33 Lu, Q.L., Yang, K., Antoniou, C., 2021. Crash risk analysis for the mixed traffic flow with human-driven and connected and autonomous vehicles,
 34 in: 2021 ieee international intelligent transportation systems conference (itsc), IEEE. pp. 1233–1238.
- 35 Ma, X., Lu, J., Liu, X., Qu, W., 2023. A genetic programming approach for real-time crash prediction to solve trade-off between interpretability and
 36 accuracy. *Journal of Transportation Safety & Security* 15, 421–443.
- 37 Ma, Y., Li, Y., Zheng, Z., Huang, H., 2024. Developing merging policies for cavs: A policy training framework combining human experience with
 38 reinforcement learning. *IEEE Transactions on Intelligent Vehicles* , early access.
- 39 Makridis, M., Mattas, K., Ciuffo, B., Re, F., Kriston, A., Minarini, F., Rognelund, G., 2020. Empirical study on the properties of adaptive cruise
 40 control systems and their impact on traffic flow and string stability. *Transportation research record* 2674, 471–484.
- 41 Martin, J.E., Rivas, T., Matías, J., Taboada, J., Argüelles, A., 2009. A bayesian network analysis of workplace accidents caused by falls from a
 42 height. *Safety Science* 47, 206–214.
- 43 Milanés, V., Shladover, S.E., 2014. Modeling cooperative and autonomous adaptive cruise control dynamic responses using experimental data.
 44 *Transportation Research Part C: Emerging Technologies* 48, 285–300.
- 45 Milanés, V., Shladover, S.E., Spring, J., Nowakowski, C., Kawazoe, H., Nakamura, M., 2013. Cooperative adaptive cruise control in real traffic
 46 situations. *IEEE Transactions on intelligent transportation systems* 15, 296–305.
- 47 Mousavi, S.M., Osman, O.A., Lord, D., Dixon, K.K., Dadashova, B., 2021. Investigating the safety and operational benefits of mixed traffic
 48 environments with different automated vehicle market penetration rates in the proximity of a driveway on an urban arterial. *Accident Analysis
 49 & Prevention* 152, 105982.
- 50 Organization, W.H., 2023. Global status report on road safety 2023. World Health Organization.
- 51 Papadoulis, A., Quddus, M., Imprailou, M., 2019. Evaluating the safety impact of connected and autonomous vehicles on motorways. *Accident
 52 Analysis & Prevention* 124, 12–22.
- 53 Ren, H., Song, Y., Wang, J., Hu, Y., Lei, J., 2018. A deep learning approach to the citywide traffic accident risk prediction, in: 2018 21st International
 54 Conference on Intelligent Transportation Systems (ITSC), IEEE. pp. 3346–3351.
- 55 Sadi, A.A., Chowdhury, L., Jahan, N., Rafi, M.N.S., Chowdhury, R., Khan, F.A., Mohammed, N., 2022. Lmfloss: A hybrid loss for imbalanced
 56 medical image classification. *arXiv preprint arXiv:2212.12741*.
- 57 Saha, D., Alluri, P., Dumbaugh, E., Gan, A., 2020. Application of the poisson-tweedie distribution in analyzing crash frequency data. *Accident
 58 Analysis & Prevention* 137, 105456.
- 59 Salles, D., Kaufmann, S., Reuss, H.C., 2020. Extending the intelligent driver model in sumo and verifying the drive off trajectories with aerial
 60 measurements, in: SUMO Conference Proceedings, pp. 1–25.
- 61 Sameen, M.I., Pradhan, B., 2017. Severity prediction of traffic accidents with recurrent neural networks. *Applied Sciences* 7, 476.
- 62 Santos, K., Dias, J.P., Amado, C., 2022. A literature review of machine learning algorithms for crash injury severity prediction. *Journal of safety
 63 research* 80, 254–269.

- 1 Seoa, T., 2023. Understanding large-scale traffic flow using model-based and data-driven dimension reduction: with covid-19 and olympic-
2 paralympic case study. EU Science Hub , 124.
- 3 Shirazi, M., Lord, D., 2019. Characteristics-based heuristics to select a logical distribution between the poisson-gamma and the poisson-lognormal
4 for crash data modelling. *Transportmetrica A: Transport Science* 15, 1791–1803.
- 5 Su, M.T., Zheng, J., Zhang, Z.P., 2020. Clustering mining of urban traffic flow based on cvae. *Journal of Traffic and Logistics Engineering* Vol 8.
- 6 Tan, H., Zhao, F., Zhang, W., Liu, Z., 2023. An evaluation of the safety effectiveness and cost of autonomous vehicles based on multivariable
7 coupling. *Sensors* 23, 1321.
- 8 Tarko, A.P., 2012. Use of crash surrogates and exceedance statistics to estimate road safety. *Accident Analysis & Prevention* 45, 230–240.
- 9 Tarko, A.P., 2021. A unifying view on traffic conflicts and their connection with crashes. *Accident Analysis & Prevention* 158, 106187.
- 10 Theofilatos, A., Chen, C., Antoniou, C., 2019. Comparing machine learning and deep learning methods for real-time crash prediction. *Transportation*
11 research record 2673, 169–178.
- 12 Treiber, M., Hennecke, A., Helbing, D., 2000. Congested traffic states in empirical observations and microscopic simulations. *Physical review E*
13 62, 1805.
- 14 Trirat, P., Yoon, S., Lee, J.G., 2023. Mg-tar: Multi-view graph convolutional networks for traffic accident risk prediction. *IEEE Transactions on*
15 *Intelligent Transportation Systems* 24, 3779–3794.
- 16 Vogel, K., 2003. A comparison of headway and time to collision as safety indicators. *Accident analysis & prevention* 35, 427–433.
- 17 Wang, B., Lin, Y., Guo, S., Wan, H., 2021. Gsnet: Learning spatial-temporal correlations from geographical and semantic aspects for traffic accident
18 risk forecasting, in: Proceedings of the AAAI conference on artificial intelligence, pp. 4402–4409.
- 19 Wang, L., Ren, Y., Jiang, H., Cai, P., Fu, D., Wang, T., Cui, Z., Yu, H., Wang, X., Zhou, H., et al., 2024a. Accidentgpt: A v2x environmental
20 perception multi-modal large model for accident analysis and prevention, in: 2024 IEEE Intelligent Vehicles Symposium (IV), IEEE. pp. 472–
21 477.
- 22 Wang, L., Wang, K., Ma, W., Abdel-Aty, M., Li, L., 2022. Real-time safety analysis for expressways considering the heterogeneity of different
23 segment types. *Journal of safety research* 80, 349–361.
- 24 Wang, Q., Wang, S., Zhuang, D., Koutsopoulos, H., Zhao, J., 2024b. Uncertainty quantification of spatiotemporal travel demand with probabilistic
25 graph neural networks. *IEEE Transactions on Intelligent Transportation Systems* 25, 8770–8781.
- 26 Wang, Q., Zhang, K., Zhu, C., Zhou, Y., 2023. A multi-regional spatio-temporal network for traffic accident risk prediction. *Engineering Letters*
27 31.
- 28 Wang, T., Ngoduy, D., Li, Y., Lyu, H., Zou, G., Dantsuji, T., 2024c. Koopman theory meets graph convolutional network: Learning the complex
29 dynamics of non-stationary highway traffic flow for spatiotemporal prediction. *Chaos, Solitons & Fractals* 187, 115437.
- 30 Wang, T., Ngoduy, D., Zou, G., Dantsuji, T., Liu, Z., Li, Y., 2024d. Pi-stgnet: Physics-integrated spatiotemporal graph neural network with
31 fundamental diagram learner for highway traffic flow prediction. *Expert Systems with Applications* 258, 125144.
- 32 Wang, W., Jiang, X., Xia, S., Cao, Q., 2010. Incident tree model and incident tree analysis method for quantified risk assessment: an in-depth
33 accident study in traffic operation. *Safety Science* 48, 1248–1262.
- 34 Wang, W., Pan, L., Liu, B., 2009. Synergetic method of traffic state recognition based on manifold learning, in: 2009 IEEE International Conference
35 on Automation and Logistics, IEEE. pp. 587–591.
- 36 Wu, M., Jia, H., Luo, D., Luo, H., Zhao, F., Li, G., 2023. A multi-attention dynamic graph convolution network with cost-sensitive learning approach
37 to road-level and minute-level traffic accident prediction. *IET Intelligent Transport Systems* 17, 270–284.
- 38 Yang, D., Ozbay, K., Xie, K., Yang, H., Zuo, F., Sha, D., 2021. Proactive safety monitoring: A functional approach to detect safety-related anomalies
39 using unmanned aerial vehicle video data. *Transportation research part C: emerging technologies* 127, 103130.
- 40 Yang, S., Zhou, W., 2011. Anomaly detection on collective moving patterns: Manifold learning based analysis of traffic streams, in: 2011 IEEE
41 Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing, IEEE.
42 pp. 704–707.
- 43 Yao, Z., Wu, Y., Wang, Y., Zhao, B., Jiang, Y., 2023. Analysis of the impact of maximum platoon size of cavs on mixed traffic flow: An analytical
44 and simulation method. *Transportation Research Part C: Emerging Technologies* 147, 103989.
- 45 Yousaf, M., Rehman, T.U., Jing, L., 2020. An extended isomap approach for nonlinear dimension reduction. *SN Computer Science* 1, 160.
- 46 Yu, L., Du, B., Hu, X., Sun, L., Han, L., Lv, W., 2021. Deep spatio-temporal graph convolutional network for traffic accident prediction. *Neuro-*
47 computing
- 48 Yu, R., Abdel-Aty, M., 2013. Utilizing support vector machine in real-time crash risk evaluation. *Accident Analysis & Prevention* 51, 252–259.
- 49 Yu, R., Abdel-Aty, M., 2014. Analyzing crash injury severity for a mountainous freeway incorporating real-time traffic and weather data. *Safety*
50 *science* 63, 50–56.
- 51 Yu, R., Wang, Y., Zou, Z., Wang, L., 2020. Convolutional neural networks with refined loss functions for the real-time crash risk analysis. *Trans-*
52 *portation research part C: emerging technologies* 119, 102740.
- 53 Yuan, J., Abdel-Aty, M., Gong, Y., Cai, Q., 2019. Real-time crash risk prediction using long short-term memory recurrent neural network. *Trans-*
54 *portation research record* 2673, 314–326.
- 55 Zhang, C., Yan, X., Ma, L., An, M., 2014a. Crash prediction and risk evaluation based on traffic analysis zones. *Mathematical Problems in*
56 *Engineering* 2014, 987978.
- 57 Zhang, J., Wu, K., Cheng, M., Yang, M., Cheng, Y., Li, S., 2020. Safety evaluation for connected and autonomous vehicles' exclusive lanes
58 considering penetrate ratios and impact of trucks using surrogate safety measures. *Journal of advanced transportation* 2020, 5847814.
- 59 Zhang, L., Jia, Y., Niu, Z., Liao, C., 2014b. Traffic state classification based on parameter weighting and clustering method. *Journal of Transportation*
60 *Systems Engineering and Information Technology* 14, 147–151.
- 61 Zhang, S., Abdel-Aty, M., 2022. Real-time crash potential prediction on freeways using connected vehicle data. *Analytic methods in accident*
62 *research* 36, 100239.
- 63 Zheng, Y., Li, S.E., Wang, J., Cao, D., Li, K., 2015. Stability and scalability of homogeneous vehicular platoon: Study on the influence of information

- 1 flow topologies. IEEE Transactions on intelligent transportation systems 17, 14–26.
- 2 Zhou, J., Zhu, F., 2020. Modeling the fundamental diagram of mixed human-driven and connected automated vehicles. Transportation research
3 part C: emerging technologies 115, 102614.
- 4 Zhou, R., Zhang, G., Huang, H., Wei, Z., Zhou, H., Jin, J., Chang, F., Chen, J., 2024. How would autonomous vehicles behave in real-world crash
5 scenarios? Accident Analysis & Prevention 202, 107572.
- 6 Zhou, Y., Wang, M., Ahn, S., 2019. Distributed model predictive control approach for cooperative car-following with guaranteed local and string
7 stability. Transportation research part B: methodological 128, 69–86.
- 8 Zhou, Z., Wang, Y., Xie, X., Chen, L., Liu, H., 2020. Riskoracle: A minute-level citywide traffic accident forecasting framework, in: Proceedings
9 of the AAAI conference on artificial intelligence, pp. 1258–1265.
- 10 Zong, R., Wang, Y., Ding, J., Deng, W., 2024. Statistical risk and performance analyses on naturalistic driving trajectory datasets for traffic modeling.
11 World Electric Vehicle Journal 15, 77.
- 12 Zou, G., Lai, Z., Ma, C., Li, Y., Wang, T., 2023a. A novel spatio-temporal generative inference network for predicting the long-term highway traffic
13 speed. Transportation research part C: emerging technologies 154, 104263.
- 14 Zou, G., Lai, Z., Ma, C., Tu, M., Fan, J., Li, Y., 2023b. When will we arrive? a novel multi-task spatio-temporal attention network based on
15 individual preference for estimating travel time. IEEE Transactions on Intelligent Transportation Systems 24, 11438–11452.
- 16 Zou, G., Lai, Z., Wang, T., Liu, Z., Li, Y., 2024. Mt-stnet: A novel multi-task spatiotemporal network for highway traffic flow prediction. IEEE
17 Transactions on Intelligent Transportation Systems 25, 8221–8236.